ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Evolution of gene expression across brain regions in behaviourally divergent deer mice

Andreas F. Kautt | **Jenny Chen** | **Caitlin L. Lewarch** | **Caroline Hu** | **Kyle Turner** | **Jean-Marc Lassance** | **Felix Baier** | **Nicole L. Bedford** | **Andres Bendesky** | **Hopi E. Hoekstra**

Department of Organismic & Evolutionary Biology, Department of Molecular & Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts, USA

Correspondence
Hopi E. Hoekstra, Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, MA, USA.
Email: hoekstra@oeb.harvard.edu

Present address
Caitlin L. Lewarch, Quiver Bioscience, Cambridge, Massachusetts, USA
Caroline Hu, Department of Integrative Sciences and Biological Arts, Massachusetts College of Art and Design, Boston, Massachusetts, USA
Kyle Turner, Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
Jean-Marc Lassance, GIGA Institute, University of Liege, Liege, Belgium
Felix Baier, Neural Systems Department, Max Planck Institute for Brain Research, Frankfurt, Germany
Nicole L. Bedford, Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming, USA
Andres Bendesky, Department of Ecology, Evolution & Environmental Biology, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, New York, USA

Abstract

The evolution of innate behaviours is ultimately due to genetic variation likely acting in the nervous system. Gene regulation may be particularly important because it can evolve in a modular brain-region specific fashion through the concerted action of *cis-* and *trans-*regulatory changes. Here, to investigate transcriptional variation and its regulatory basis across the brain, we perform RNA sequencing (RNA-Seq) on ten brain subregions in two sister species of deer mice (*Peromyscus maniculatus* and *P. polionotus*)—which differ in a range of innate behaviours, including their social system—and their F_1 hybrids. We find that most of the variation in gene expression distinguishes subregions, followed by species. Interspecific differential expression (DE) is pervasive (52–59% of expressed genes), whereas the number of DE genes between sexes is modest overall (~3%). Interestingly, the identity of DE genes varies considerably across brain regions. Much of this modularity is due to *cis*-regulatory divergence, and while 43% of genes were consistently assigned to the same gene regulatory class across subregions (e.g. conserved, *cis-* or *trans-*regulatory divergence), a similar number were assigned to two or more different gene regulatory classes. Together, these

Andreas F. Kautt and Jenny Chen contributed equally.

results highlight the modularity of gene expression differences and divergence in the brain, which may be key to explain how the evolution of brain gene expression can contribute to the astonishing diversity of animal behaviours.

1 | INTRODUCTION

Differences in innate behaviours are observed even among closely related species, from feeding and reproductive behaviour in *Drosophila* (Markow & O'Grady, 2008), to sleep behaviour in cavefish (Duboué et al., 2011), to social behaviours in mammals (Young et al., 1999). Changes that affect gene expression in the nervous system are thought to be particularly relevant in the evolution of behavioural divergence (Alaux et al., 2009; Barrett et al., 2013; Pizzollo et al., 2022). Thus, profiling gene expression changes across species can be an important first step in elucidating the molecular mechanisms underlying the evolution of behaviours.

Deer mice of the genus *Peromyscus* occupy almost every terrestrial ecological niche in North America, likely contributing to the large amount of interspecific variation in naturally occurring behaviours (Bedford & Hoekstra, 2015; Dewey & Dawson, 2001). For example, two sister species, *P. maniculatus* and *P. polionotus*, diverged from each other around 1.8 million years ago (Schenk et al., 2013) (Figure 1a) but nonetheless vary in a number of innate behaviours (e.g. burrowing (Dawson et al., 1988), infant vocalization (Jourjine et al., 2023) and thermoregulatory nesting (Lewarch & Hoekstra, 2018)). In particular, these two species lie on opposite ends of the monogamy-promiscuity spectrum and, among other reproductive behaviours, exhibit differences in the degree of sexual dimorphism in parental care (Bendesky et al., 2017; Foltz, 1981). Because these two species can be interbred to produce F_1 hybrids, this system provides an opportunity to not only compare interspecific differences in gene expression but also to better understand the regulatory mechanisms underlying those differences.

Comparing species expression differences to allelic expression differences in the F_1 hybrids allows one to dissect the *cis*- (i.e. genetic variants residing on the same chromatid) and *trans*-regulatory (i.e. genetic variants affecting diffusible elements) components of expression divergence because both alleles in F_1 hybrids share the same *trans*-regulatory environment (Goncalves et al., 2012; McManus et al., 2010; Tirosh et al., 2009; Verta & Jones, 2019; Wittkopp et al., 2008). *Cis*-regulatory changes are often presumed to have local, tissue-specific effects and are expected to be major contributors to the evolution of gene expression compared to *trans*-regulatory changes, which potentially have pleiotropic, broad effects (Signor & Nuzhdin, 2018). While a few studies have used F_1 hybrids to profile regulatory changes in the brains of behaviourally divergent species (Benowitz et al., 2020; Reuveni et al., 2018; Wang et al., 2019), they have typically focused on only one or two brain regions of interest. Given that the brain is a heterogeneous organ composed of functionally specialized subregions composed of transcriptionally distinct cell clusters (Zhang et al., 2023), this raises the question: if a gene is differentially expressed in one region of the brain, is it typically also differentially expressed in other regions, and do we observe the expected predominance of *cis*-regulatory changes underlying this modularity in gene expression?

To address this question and gain a better understanding of the evolution of gene regulation in the brain, we investigate gene expression in both females and males of *P. maniculatus* and *P. polionotus* as well as their F_1 hybrids using RNA-Seq. We compare variation in gene expression in the whole brain as well as across ten distinct subregions of the brain: the amygdala, cerebellum, cortex, hindbrain, hippocampus, hypothalamus, midbrain, septum, striatum and the thalamus (Figure 1b, Figure S1). These subregions form fundamental systems of the mammalian brain, including the limbic system (i.e. amygdala, hypothalamus, hippocampus, septum), which is involved in emotion, behaviour, and long-term memory, the mesolimbic pathway (i.e. cortex, thalamus, septum, striatum), which regulates motivation and reward processing, and motor control regions (i.e. midbrain, hindbrain, cerebellum), which control animal movement. Using these data, we (i) quantify gene expression differences across the brain, (ii) screen for differentially expressed genes between sexes and species, and (iii) investigate the regulatory mechanisms driving this gene expression divergence. Importantly, in addition to detecting and characterizing DE genes per se, we also (iv) assess the extent of shared versus region-specific gene expression differences and their underlying regulatory divergence across the brains of these two deer mice species.

2 | MATERIALS AND METHODS

2.1 | Animal husbandry

Peromyscus maniculatus bairdii and *P. polionotus subgriseus* animals were originally acquired from the *Peromyscus* Genetic Stock Center (Columbia, SC, USA). To generate F_1 hybrids, we crossed female *P. maniculatus* and male *P. polionotus*, since hybrids derived from reciprocal crosses are inviable (Vrana et al., 2000; Watson, 1942). We weaned animals at 21–23 days of age into single-sex, same-strain groups. We maintained animals on a 16 h:8 h light:dark cycle at 22°C, housed them in standard mouse cages with corncob bedding and provided them with food and water ad libitum. Animal husbandry
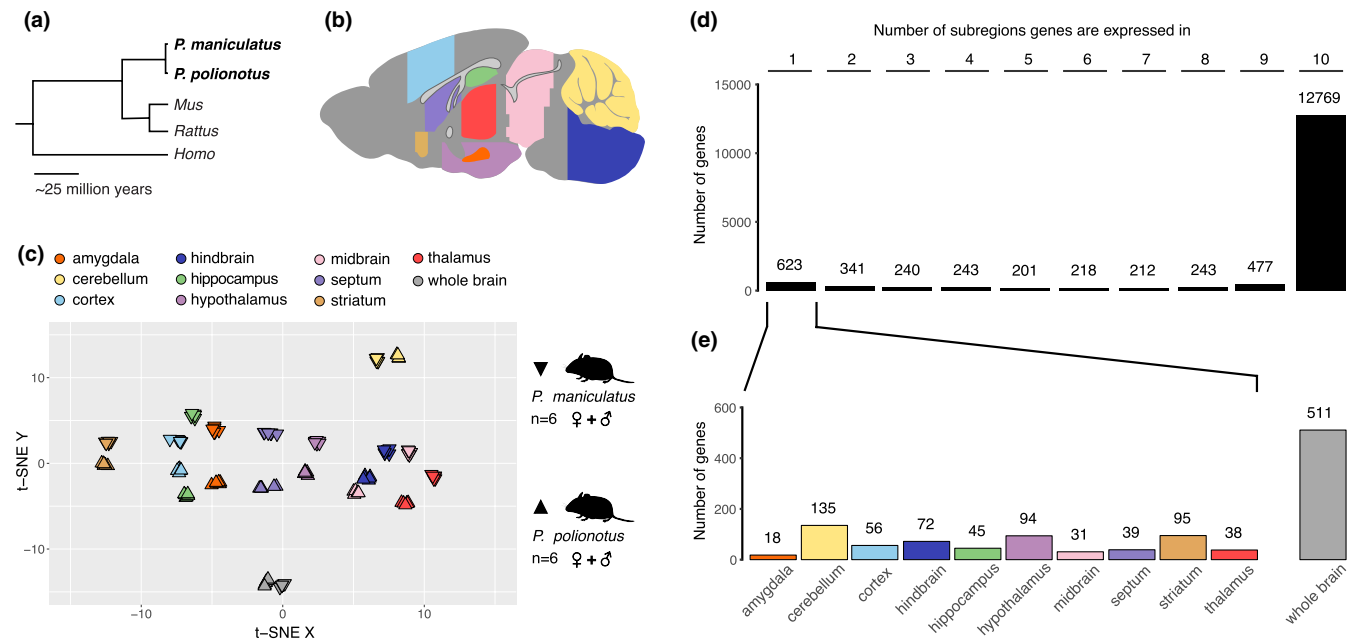
FIGURE 1 Gene expression across the brain of *Peromyscus maniculatus* and *Peromyscus polionotus* mice. (a) Phylogenetic relationships of the two focal *Peromyscus* species (in bold) with house mice (*Mus*), rats *(Rattus)* and humans (*Homo*). (b) Schematic of a sagittal brain section highlighting the locations of the 10 (sub)region dissections (colour) used in this study, as well as whole brain (grey). For details on dissections, see Section 2 and Table S2. (c) t-distributed stochastic neighbour embedding (t-SNE) plot of the overall variation in gene expression. *P. maniculatus* samples are depicted as downward-facing triangles, *P. polionotus* samples as upward-facing triangles. (d) The number of genes (n = 16,078) expressed in one or multiple brain regions. (e) Distribution of privately expressed genes across the 10 (sub)regions; 511 genes are expressed in the whole brain samples but not in any of the 10 sampled (sub)regions.

and experimental procedures were approved by the Harvard University Faculty of Arts and Sciences Institutional Animal Care and Use Committee (protocol 27-15-3).

2.2 | Tissue collection, RNA extraction, library preparation and sequencing

2.2.1 | Whole brain data

For the whole brain dataset, a cohort of virgin male *P. maniculatus* (n = 5) and *P. polionotus* (n = 6) animals were used. The two species of *Peromyscus* become sexually mature by around 50 days (Clark, 1938), live up to 3–5 years in captivity (Sumner, 1922), and do not show signs of aging (e.g. lack of grooming, fat build-up around face and neck, difficulty moving) in the first two years of life. Therefore, we used animals that were 66–90 days old (median: 81 days)—a range well within the healthy, adult period of the *Peromyscus* life cycle. Animals were euthanized by CO_2 inhalation 7–10 h before the start of the dark period. Brains were then rapidly dissected, flash-frozen and stored at −70°C. We homogenized the samples in TRIzol Reagent (Invitrogen, 15,596,026) using a Bio-Gen PRO200 homogenizer (PRO Scientific, Oxford, USA), and extracted RNA using a Direct-zol™ RNA MiniPrep Plus kit (Zymo Research, R2070), including a DNase I treatment. We measured RNA concentrations using a Qubit RNA BR Assay kit (Thermo Fisher Scientific, Q10210) and assessed RIN scores using an Agilent TapeStation 2200 (Agilent Technologies, Santa Clara, CA).

We prepared each sequencing library from 1 μg of total RNA using a PrepX PolyA mRNA Isolation Kit (Wafergen, 40,047) on the Apollo 324 NGS Library Prep System (Takara Bio USA Holdings, Mountain View, CA). We added Illumina indices and amplified libraries with 12 cycles of PCR and cleaned them using PCRClean DX beads (Aline Biosciences, Woburn, MA). To quantify library concentrations, we used the Quant-iT BR dsDNA kit (ThermoFisher Scientific, Q33130). We then assessed library fragment size distributions using an Agilent TapeStation 2200 D1000 tape. To further quantify the libraries, we used a Kapa Library Quantification Kit for Illumina Platforms (Kapa Biosystems, Wilmington, MA). In total, we pooled 33 libraries to 20 nM and sequenced them on 5 lanes of an Illumina HiSeq 2500 2 × 150 bp chemistry. Additional information on samples and sequencing batches is provided in Table S1 and Figure S2B.

2.2.2 | Subregion data

For the brain subregion dataset, a cohort of 82–85-day-old virgin male and female *P. maniculatus*, *P. polionotus* and their F_1 hybrids (see Table S1 for sample sizes) were euthanized by CO_2 inhalation followed by cervical dislocation 3–10 h before the onset of the dark period. One cohort was used for dissection of the cerebellum, hippocampus, and hindbrain samples. In these samples, after the brain was removed, a dorsal-to-ventral incision was made posterior to the cerebrum to release the hindbrain and cerebellum. The

cerebellar peduncles were cut, and the hindbrain and cerebellum were placed in tubes and flash frozen in liquid nitrogen. We note that our dissection strategy may have inadvertently captured part of the midbrain in the hindbrain samples. The Petri dish with the remaining brain was then placed on ice, and the two cerebral hemispheres were cut apart. For each hemisphere separately, the hippocampus was severed from the dorsal fornix and gently scooped out of the hemisphere with a brush in ventral-to-dorsal direction (Lu et al., 2001), and then flash frozen. For a second cohort of animals, we dissected the brains and then immediately submerged them into embedding compound for cryosectioning (Tissue-Tek O.C.T.) and froze them on dry ice. We sectioned brains on a coronal plane at 150 μm using a cryostat at −20°C (Leica CM3050 S, Leica Biosystems Inc., Buffalo Grove, IL) and stored the sections at −70°C until we dissected the medial prefrontal cortex, septum, striatum, hypothalamus, amygdala, thalamus and midbrain with sample corers according to anatomical landmarks (Figure S1, Table S2). The sample cores were immediately transferred into homogenization buffer. After subregion dissection, we homogenized samples for 10 seconds in Maxwell RSC Homogenization Buffer (Promega, Madison, USA) with 2% 1-thioglycerol using a Bio-Gen PRO200 homogenizer (PRO Scientific, Oxford, USA) and then extracted RNA using the Maxwell RSC simplyRNA Tissue Kit (Promega). We quantified extracts using a Quant-IT RNA Assay Kit, assayed for purity on a NanoDrop ND-1000, and tested for RNA integrity on a TapeStation 4200.

We prepared RNA-Seq libraries using an mRNA HyperPrep kit (Kapa Biosystems, Wilmington, USA) automated on a BioMek FXp (Beckman Coulter, Brea, USA). Post-capture, we fragmented mRNA for 6 min at 94°C for a target insert size of 200–300 base pairs. We amplified dual-indexed adapter-ligated libraries in 11 PCR cycles and then sequenced the amplified libraries on a NovaSeq 6000 2×150 bp on two S2 flow cells.

2.3 | Trimming and read mapping

Using *cutadapt v.2.3* (Martin, 2011), we trimmed very low-quality bases (quality score < 10) and applied a minimum length cut-off of 25 bp. We then mapped trimmed reads of *P. maniculatus* and *P. polionotus* individuals to their respective reference genome, Pman_2.1 (GCA_003704035.1) or Ppol_1.3.3 (GCA_003704135.2), and in-house annotations generated with *Comparative Annotation Toolkit* (Fiddes et al., 2018) using the GENCODE v15 primary-assembly annotation of *Mus musculus* (GRCm38 mm10) as reference. To capture allele-specific expression in F_1 hybrids, we mapped their reads to a diploid reference genome created by concatenating the two species assemblies and annotations. This competitive mapping strategy in F_1 hybrids, in combination with *MMSEQ* (see below), has been shown to be robust for estimating allele-specific expression in F_1 hybrids, including in different subspecies of *Mus* (Goncalves et al., 2012; Perez et al., 2015). We performed read mapping with *STAR v.2.7.0e*

(Dobin et al., 2013) in a transcriptome-aware manner to more accurately map reads spanning splice junctions. On average, we successfully mapped 30.18 ± 5.44 (SD), 31.58 ± 6.99 (SD) and 63.06 ± 15.34 (SD) million read pairs per sample for *P. maniculatus*, *P. polionotus* and their F_1 hybrids, respectively (Table S3).

2.4 | Gene expression estimation

We estimated gene expression with *RSEM v1.3.1* (Li & Dewey, 2011). We retained only genes present in both species' annotations for downstream analyses ($n = 33,836$). By performing a blast search against an NCBI-generated annotation of *P. maniculatus* (Annotation Release 100), we identified 825 genes annotated by NCBI as pseudogenes. We excluded these genes from all analyses, together with any gene whose orthologue was flagged as pseudogene in the *Mus musculus* annotation (GENCODE v15).

2.5 | t-SNE/PCA/distance-based clustering

To investigate the main axes of variation in gene expression in our dataset, we used three different approaches. Prior to these analyses, we log-transformed and normalized count data using the regularized log-transformation function (rlog) implemented in *DESeq2* (Love, Huber, & Anders, 2014). First, we used t-distributed stochastic neighbour embedding (t-SNE) implemented in *Rtsne* (Krijthe, 2015). Second, we performed a principal component analysis (PCA) on centred and scaled data. Finally, we performed hierarchical clustering based on Euclidean distance. A single F_1 hybrid hypothalamus sample was identified as an outlier in all three approaches, and therefore we excluded this sample from all downstream analyses.

2.6 | Differential expression (DE) analyses and multivariate adaptive shrinkage

We tested for differential gene expression (1) between sexes within species and (2) between the two species independent of sex. Briefly, for each comparison, we used *DESeq2* to obtain log_2 fold changes (effect size estimates) together with their associated per-gene standard errors and then used multivariate adaptive shrinkage (MASH) (Urbut et al., 2019) to adjust these effect size estimates and assess their statistical significance (see Appendix S1: Materials A for details).

To test if the number of DE genes was biased towards a species or sex, we performed exact binomial tests in *R*, assuming an equal probability of success ($p = .5$) and correcting for multiple testing using the Bonferroni–Holm approach. In analyses comparing the concordance between individual subregion and whole brain differential expression calls, we used *DESeq2* and called statistically differentially expressed genes ($p < .05$) in each subregion and the whole

brain separately, correcting for multiple testing using the Benjamini–Hochberg procedure.

2.7 | Gene regulation analyses

For these analyses, we estimated gene expression with *bam2hits* and *MMSEQ* (Turro et al., 2011). We assigned genes to five gene regulatory classes based on the individual expression estimates in both parental species and alleles in F_1 hybrids: "conserved", "compensatory", "*cis*", "*trans*" and "*cis & trans*" using *MMDIFF* (Turro et al., 2014). To compare the relative support (i.e. posterior likelihood) for each of the five models for any given gene, we used a polytomous model comparison approach (see Turro et al., 2014 for details) using a flat prior probability of 0.2 for each model. Design matrices for all models included sex as a covariate to account for potential confounding effects. We performed all gene regulation analyses separately for each subregion. We evaluated the performance and robustness of our model selection approach using a shuffling and a bootstrapping approach (see Appendix S1: Materials B). Based on these analyses, we decided to use a posterior probability threshold of 0.75, which resulted in almost no false positives in the shuffled controls and more than 95% concordant assignments in our bootstrap approach (Figure S3).

After statistical assignment via model comparisons, we further subdivided genes in the "*cis & trans*" class into the "*cis + trans*" and "*cis × trans*" classes (Landry et al., 2005) based on the ratio of weighted log-fold changes (Shen et al., 2014) in the parental species to weighted log-fold changes of alleles in F_1 hybrids.

Because all male F_1 hybrids carry exclusively the *P. maniculatus* X chromosome, we did not consider genes on the X chromosome in the gene regulation analyses. Furthermore, we only kept genes expressed at a minimum μ (equivalent to RPKM; Turro et al., 2011) of one in at least three animals. We omitted between 2 and 21 genes per subregion due to model convergence issues, together with an additional 138 genes that are subject to genomic imprinting (Perez et al., 2015). Finally, we performed clustering of categorical data in heatmaps of inter-subregion regulation change based on Gower distance and a divisive algorithm.

2.8 | Gene enrichment analyses

We tested if the function of sets of genes of interest were over-represented with *WebGestaltR* (Liao et al., 2019). Applying a false discovery rate of 0.05 and using *Mus musculus* as a reference, we queried the non-redundant gene ontology databases as well as the Mammalian Phenotype Ontology, the KEGG, Panther and Reactome pathway databases. If indicated, we reduced the redundancy of results to a maximum of 10 sets by applying *WebGestaltR's* weighted set cover function. As background, we only retained genes that were expressed in our data set and passed the same filtering thresholds as the genes of interest (foreground) on an analysis-by-analysis basis.

3 | RESULTS

3.1 | Main axes of variation in gene expression separate subregions and species

As a first step, we characterized overall patterns of gene expression across subregions in both species. We also characterized gene expression in whole brain samples for comparison with subregions. Overall, we found 16,078 genes (excluding pseudogenes) are expressed in the brain, using a threshold of at least one count per million reads (CPM) in at least three samples in any given subregion. t-distributed stochastic neighbour embedding (t-SNE) showed that samples cluster consistently both by subregion and species (Figure 1c). This finding was also supported by a principal component analysis (PCA) (Figure S2A): three axes of the PCA explain most of the variation between whole brain and subregions (PC1, 17.85%), across subregions (PC2, 14.95%) and between cerebellum and all other subregions (PC4, 12.08%), and an additional axis (PC3) explains variation between species (13.59%). Finally, the whole brain and cerebellum gene expression profiles were the most distinct in a hierarchical, distance-based clustering approach (Figure S2B).

3.2 | Most genes are expressed in all subregions while few are subregion-specific

Given that variation in gene expression was primarily driven by differences among subregions, we asked whether brain subregions are defined by subregion-specific gene expression and what the functions of such genes would be. We found that each individual subregion expressed between 13,755 and 14,429 genes (of the 16,078 genes that met our expression threshold). Of these, 12,769 genes (82.03% of genes included in this analysis) were expressed in all ten subregions and only 623 genes (4.00%) were expressed exclusively in one subregion (Figure 1d). The top three most enriched terms of these privately expressed genes represented specialized functions: "Hormone ligand-binding receptors", "Glycoprotein hormones" and "Peptide hormone biosynthesis" (all three in the Reactome pathway database). The highest number of privately expressed genes was found in the cerebellum and the lowest in the amygdala (Figure 1e). An additional 511 genes were captured only in the whole brain samples and are likely regionally restricted to areas not captured by our subregion sampling.

3.3 | Whole brain samples only partially capture subregion-specific expression divergence between species

Because comparative studies often use whole brain dissections rather than specific subregions, we investigated the extent to which region-specific differential expression between species is reflected in a sample generated from the entire brain. In total, we found 12,624 genes (78.5% of the 16,078 genes included in this

analysis) were significantly differentially expressed (DE) (local false sign rate <0.05) in at least one subregion or the whole brain, and 3462 genes (21.5%) were significantly DE in all subregions and the whole brain. We found that 58.4–63.3% of DE calls were significant (FDR <0.05) and in the same direction between the whole brain and any single subregion (Figure S4). In 10.0–19.5% of cases, there was a significant difference in the respective subregion but not the whole brain, which appears to be driven by localized DE that is not detectable (too diluted) in a whole brain sample (see Appendix S1: Materials C for additional analyses). We also found 16.2–22.7% of cases where a DE call was in the whole brain but not in any given subregion, which could reflect localized DE in a region not sampled in our subregion dissections or differences in subregion size, which would only be apparent in whole brain comparisons. Only in 0.8–2.2% cases were mismatches due to a "sign flip" where the direction of expression differences ($\log_2$ fold changes) was inverted yet significant in both the whole brain and a specific subregion. Taken together, these results suggest that whole brain samples capture most, but not all, of the subregion-specific DE genes.

3.4 | Differential gene expression is pervasive between species

Given that each subregion shows a distinct transcriptomic profile, we wondered whether gene expression divergence between species also happens in a subregion-specific manner. To this end, we compared differential gene expression (DE) between the two species across subregions. Instead of analysing each subregion separately, we used multivariate adaptive shrinkage to account for correlations in effect sizes among subregions and to improve power (Urbut et al., 2019). Within subregions, the number of DE genes ranged from 52.0 to 59.3% of expressed genes, with no bias towards either species (exact binomial test; all $p_{adj} > .05$) (Figure 2a). To focus on genes that most likely have biologically relevant levels of differential expression, we applied a conservative effect-size filter of a minimum two-fold change in DE between species. This filter reduced the number to 2919 genes (18.2% of all expressed genes). Most of these genes—especially the most differentially expressed ones—were consistently differentially expressed in the same direction across brain regions (Figure 2b). However, 628 of these (21.5% of the 2919 genes) showed a sign flip in $\log_2$ fold changes across subregions (Figure 2b), indicating considerable subregion-specific variation in differential expression. To quantitatively assess the correlations in species-biased gene expression among subregions and the whole brain, we calculated the proportion of sharing by magnitude (i.e. the proportion of genes significantly DE in at least one of the two tested subregions that exhibited a $\log_2$ fold change within a factor of 0.5 of each other) among all pairwise comparisons. The proportion of sharing ranged from 0.51 to 0.86 (Figure 2c). The highest correlations were found among subregions such as midbrain and

hypothalamus (0.86) and amygdala and septum (0.85). By contrast, correlations between subregions and the whole brain showed the lowest levels of pairwise sharing with any of the subregions, followed by the cerebellum.

3.5 | Lack of support for universal transcriptomic mechanism underlying monogamy

Motivated by the difference in mating systems in our two focal species (*P. maniculatus* = promiscuous; *P. polionotus* = monogamous), we wanted to test whether the same set of genes are identified as DE in our data as a set of previously identified mating system-related genes: previous work comparing whole brain gene expression between pairs of promiscuous and monogamous species across four major vertebrate clades proposed a set of 41 candidate genes associated with monogamous mating systems (Young et al., 2019). Importantly, this inference was based on species pairs encompassing several different conditions (e.g. laboratory-bred vs. wild-caught, virgin vs. sexually reproductive) and thought to represent steady-state differences in gene expression related to different mating systems. Of the candidate genes identified in Young et al., (2019), 38 were present in our *Peromyscus* transcriptome and 30 of the 38 were also DE across our focal species. Of these 30 genes, one was reported to be up-regulated in promiscuous species, while the remaining 29 were up-regulated in monogamous species. In our data, however, only 15 of 30 (50%) were DE in the same direction (in any subregion) (Figure S5), consistent with the number of direction-matching DE genes expected by chance. Thus, we do not find evidence for the proposed universal gene set underlying monogamy in our data.

3.6 | Sex-biased gene expression is often species-specific

In addition to mating system differences, the two focal *Peromyscus* species also differ in their degree of sexually dimorphic reproductive behaviours, including parental care, which is performed by females of both species but, in males, is largely restricted to *P. polionotus* and limited in *P. maniculatus* (Bendesky et al., 2017). Thus, we were also interested in identifying genes with a species-specific sex bias, thereby characterizing how sex-biased gene expression evolves across our focal species. We note that our whole brain samples were collected exclusively from males and therefore not suited for these analyses.

In total, we found 461 significantly DE genes between females and males in at least one of the 20 species and subregion combinations, corresponding to 3.0% of all 15,503 genes in this analysis. While 203 out the 461 genes were shared (i.e. showed a sex-bias in at least one subregion) between the two species, 150 genes were sex-biased only in *P. polionotus* and 108 only in *P. maniculatus*. The number of sex-biased genes ranged from 118 to 205 with
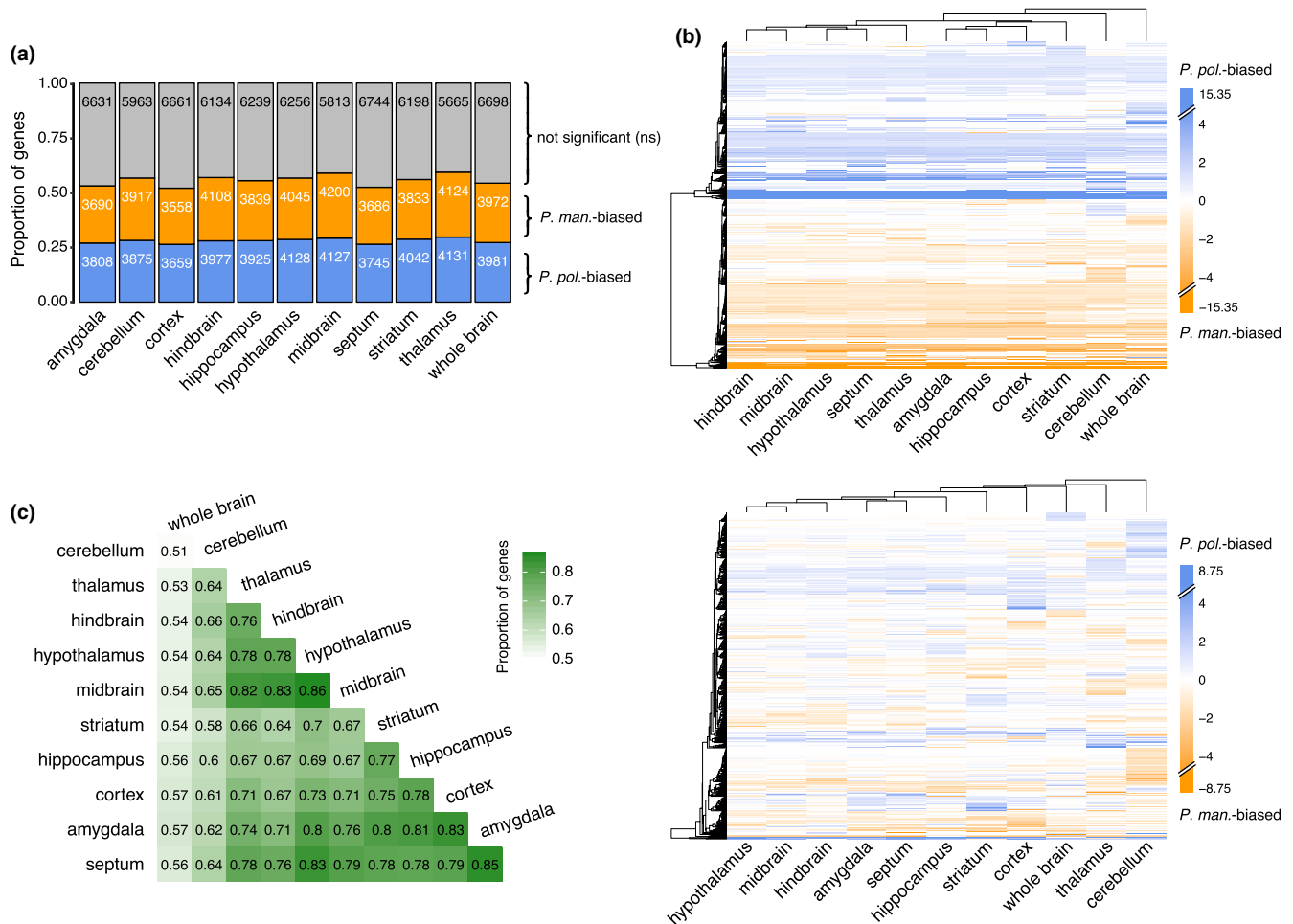
FIGURE 2 Species-biased gene expression across subregions and the whole brain. (a) Proportion and number of significantly differentially expressed (DE) genes biased towards either species across 10 subregions and the whole brain. (b) Adjusted log$_2$ fold change (LFC) estimates (after multivariate adaptive shrinkage) across subregions and the whole brain of 2919 genes that are significantly differentially expressed and exhibited a log$_2$ fold change >1 in at least one subregion (top); subset of 628 genes that show a sign flip in LFC estimates across subregions (bottom). Note that maximum colour intensity was cut off at ±5 LFC. Maximum and minimum LFC estimates are indicated in the legend. (c) Proportion of significantly DE genes in any pairwise comparison that exhibited a shared signal, defined as exhibiting a log$_2$ fold change in the same direction within a factor of 0.5 of each other.

a significantly higher number of male-biased than female-biased genes in seven out of 20 species-subregion comparisons (exact binomial test; $p_{adj} < .05$; Bonferroni–Holm correction). Only the hippocampus of *P. polionotus* exhibited the reverse pattern of a significant excess of female-biased genes over male-biased genes (Figure 3a). Given the reduction of behavioural sexual dimorphism in *P. polionotus*, we were curious as to whether *P. polionotus* exhibited fewer sex-biased genes compared to *P. maniculatus*. However, the total number of sex-biased genes (male-biased plus female-biased) was lower in *P. polionotus* than *P. maniculatus* in only the septum; and in fact was higher in *P. polionotus* in the cerebellum and hippocampus (exact binomial test; $p_{adj} < .05$; Bonferroni–Holm correction) (Figure 3a).

We next characterized the evolution of sex-biased genes across the two species. Per subregion, between 28.4 and 41.3% of sex-biased genes were shared between the two species (Figure 3b). The

log$_2$ fold change distributions of these shared sex-biased genes were overall very similar in both species across subregions (Figure 3c). Many previously known sex-biased genes were differentially expressed between females and males in both species across all subregions (e.g. the X-linked Lysine-specific demethylase 5C [Kdm5c]) (Naqvi et al., 2019), or only in specific subregions (e.g., oestrogen receptor 1 [Esr1] in the hypothalamus) (Xu et al., 2012) (Figure 3d). However, we also detected several genes that were sex-biased in only one species and not the other in both the monogamous and the promiscuous species. Some of the genes were consistently sex-biased in all subregions of one species (e.g. aldehyde dehydrogenase 1 family, member B1 [Aldh1b1] in *P. polionotus*), whereas others were so in only specific subregions (e.g. Lysophosphatidic acid receptor [Lpar1] in the hypothalamus of *P. polionotus*) (Figure 3d). These data emphasize that sex-specific DE is only partially shared between species.
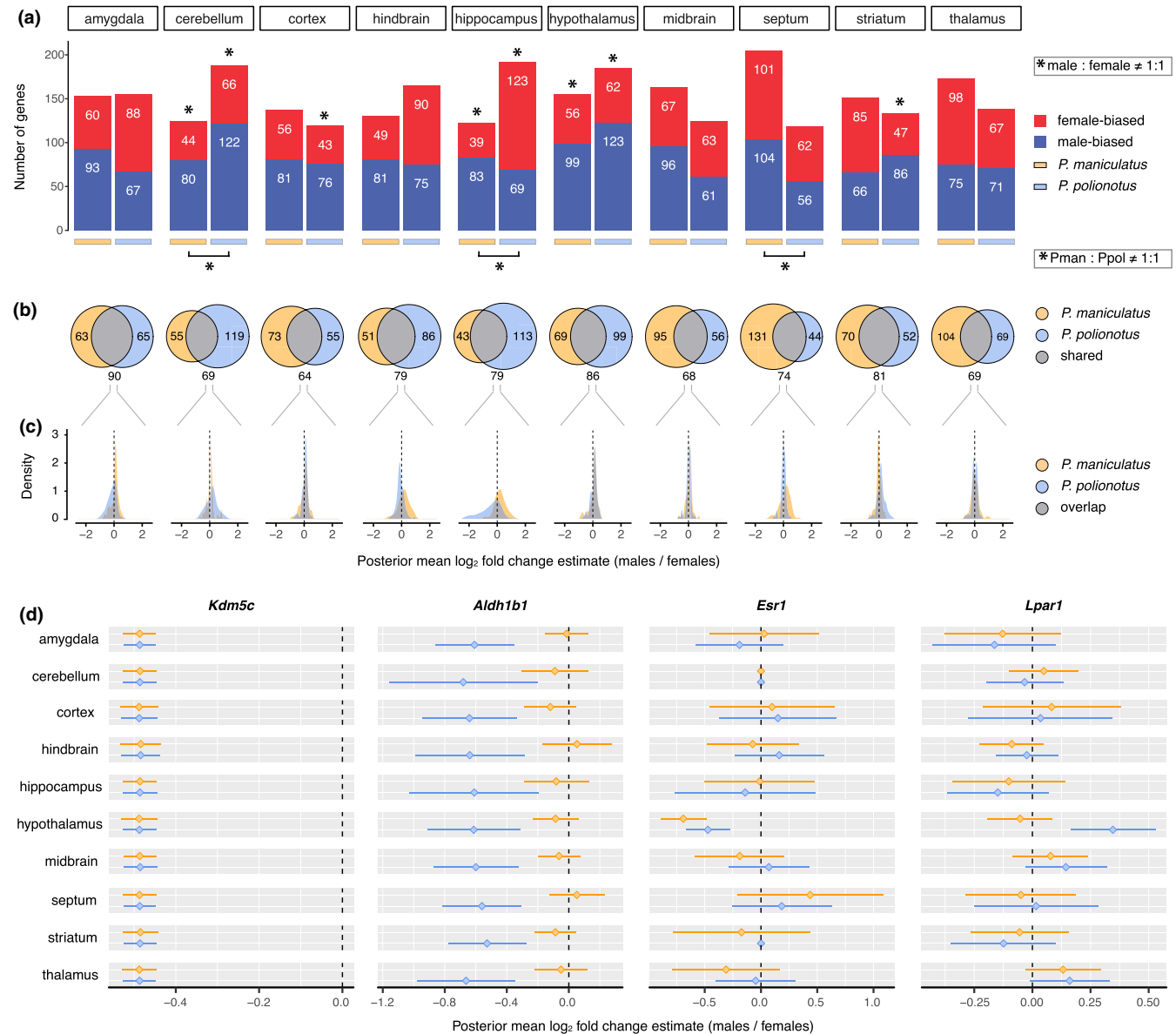
FIGURE 3 Shared and species-specific sex-biased gene expression. (a) Number of significantly differentially expressed female- or male-biased genes by species and subregion. Asterisks indicate statistically significant differences between males and females within species (on top of bars) and between species (below bars) based on binomial tests after Bonferroni–Holm family-wise error rate correction ($p < .05$). (b) Number of shared and species-specific sex-biased genes. (c) Distribution of $\log_2$ fold changes for shared sex-biased genes in both species. (d) Exemplary sex-biased genes, either sex-biased in both species across all subregions (*Kdm5c*), only one species across all subregions (*Aldh1b1*), in both species in only one specific subregion (*Esr1*) or one species in only one specific subregion (*Lpar1*). Diamonds and error bars indicate posterior mean $\log_2$ fold change estimates (after multivariate adaptive shrinkage) with 95% confidence intervals (CI) (mean ± 1.96 SD). Estimates of zero with a CI of zero indicate genes that were not expressed in a specific subregion.

3.7 | Differential gene expression between species is mostly due to *cis*-regulatory divergence

We next investigated the regulatory bases underlying gene expression divergence between species by leveraging allele-specific expression patterns in F_1 hybrids. In F_1 hybrids, chromosomes from both species are present in the same nucleus and thus subjected to the same *trans* environment. Allelic expression differences in F_1 hybrids are, therefore, due to *cis*-regulatory changes that have evolved across species. The *trans*-regulatory component can be inferred by

comparing F_1 expression to parental expression levels (Figure S6A). We assigned genes to the five following regulatory classes using a Bayesian regression framework followed by polytomous model comparisons: "conserved" (i.e. no regulatory changes), "compensatory" (i.e. no expression differences across parental species, but evidence of regulatory changes from F_1 hybrids), purely "*cis*", purely "*trans*" and "*cis* & *trans*" (Figure S6B). The "*cis* & *trans*" class was further subdivided post hoc: *cis*- and *trans*-regulatory effects that act synergistically in driving gene expression divergence were classified as "*cis* + *trans*" and *cis*- and *trans*-regulatory effects that act in

opposite directions were classified as *"cis×trans"* (Figure 4a). These regulatory categories have been suggested to be indicative of selection pressures on gene expression across the genome (see below). We quantified the relative proportions of gene regulatory classes across subregions both for functional genes and pseudogenes, which should be no longer under selective constraint, to further test for evidence of selection. We note that we are unable to determine whether any given gene has been under selection but are only able to test whether the overall gene regulatory class is enriched for certain modes of evolution.

The "conserved" and "compensatory" classes contain non-DE genes between species and comprise 14.2–22.8% and 27.4–30.9% of all functional genes across subregions, respectively (Figure 4c). While the lack of regulatory divergence in the conserved class could be due to stabilizing selection or neutral evolution, compensatory gene regulation is often assumed to be due to stabilizing selection (Landry et al., 2005). The fact that pseudogenes are depleted for these two classes with 2.9–7.4% and 12.4–16.7%, respectively (Figure S7F), lends evidence to the inference that the expression of a large swathe of genes is being kept constant at least in part due to stabilizing selection.

The other regulatory classes, *"cis"*, *"trans"*, *"cis+trans"* and *"cis×trans"*, which encompass genes that show expression differences in the parental species, account together for 45.8–58.2% of genes across subregions (Figure 4c) – a range close to the 52.0–59.3% observed in our DE analysis using a different methodological approach and considering only the parental species (Figure 2a). While most of these genes exhibited modest expression differences (log$_2$ fold change [LFC] >−3 and <3) (Figure 4b), most of the large expression differences (LFC <−3 or >3) were due purely to *cis*-regulatory divergence (Figure S7A). This pattern was consistent across individual subregions (Figure S7B–D). Purely *cis*-regulatory divergence was also the most common regulatory class underlying expression differences with 30.7–35.0% compared to only 4.0–7.2% purely *trans*-regulatory divergence across subregions (Figure 4c). While DE due to purely *cis*- or purely *trans*-regulatory divergence could be due to divergent selection or neutral evolution, an accumulation of mutations driving gene expression divergence between species in the same direction (*"cis+trans"*) is often interpreted as evidence for divergent selection (Fraser et al., 2010; Orr, 1998; Verta & Jones, 2019). This class accounted for 0.4–1.6% of genes. Taken together, these analyses demonstrate that most of gene expression
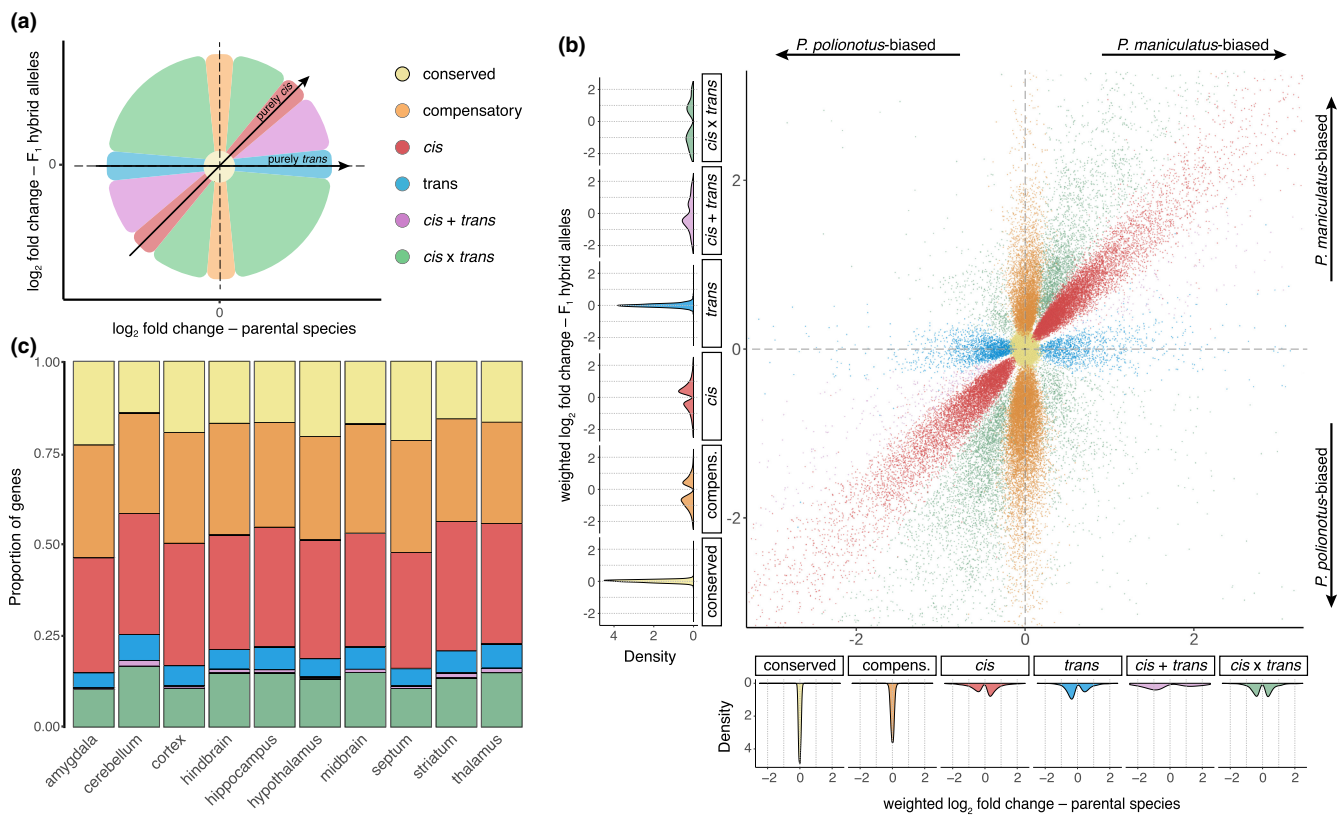

FIGURE 4 Assignment of genes to gene regulatory classes. (a) Schematic of gene regulatory classes in relation to expression ratios in parental species versus F$_1$ hybrid alleles. (b) Empirically determined weighted gene expression log$_2$ fold changes in parental species versus log$_2$ fold changes of alleles in F$_1$ hybrids across the entire data set. Each dot is a single gene colour-coded by its inferred gene regulatory class (applying posterior probability cut-off of 0.75). Side-panels provide distribution of weighted log$_2$ fold change estimates for parental species (bottom) and alleles in F$_1$ hybrids (left) by gene regulatory class. Genes with stark expression differences (log$_2$ fold change of more than plus/minus three) were omitted to enhance readability (see Figure S7A for full distribution). (c) Proportion of genes assigned to different gene regulatory classes across subregions. Genes that could not be assigned (at this threshold) were omitted for clarity from all panels (see Figure S7E for proportion of unassigned genes).

divergence between these two species is due to *cis*-regulatory differences.

To quantitatively assess whether expression in F_1 hybrids is regulated in an additive, dominant, or transgressive (over- or underdominant) way (Gibson et al., 2004; Landry et al., 2005), we also classified genes in an orthologous manner by their overall expression levels in parental species versus F_1 hybrids (Figure S8A,B). Consistent with a study in three-spined stickleback fish (Verta & Jones, 2019), we found genes in the "*cis*" and "*cis + trans*" classes to be expressed in a slightly more additive way than genes in the "*trans*" and "*cis × trans*" classes (Figure S8C). The "*cis × trans*" class also contained many genes that showed transgressive expression patterns (Figure S8C); such misexpression of genes in hybrids due to unbalanced *cis*- and *trans*-regulatory mechanism could contribute to hybrid incompatibilities and thereby speciation or species maintenance (McGirr & Martin, 2019).

3.8 | High degree of modularity in gene regulation across the brain

Having quantified the relative sizes of gene regulatory classes among subregions, we were motivated to assess how consistent

gene regulation was for any given gene across subregions. In other words, how frequently is a gene assigned to the same gene regulatory class across all ten subregions. In total, 6600 genes (43.1% of the 15,308 genes included in this analysis after filtering; see Methods for details) were exclusively assigned to a single regulatory class (Figure 5a). Out of these, 1696 genes could only be assigned unambiguously in a single subregion. Disregarding these genes leaves 4904 genes (32.0%) that could be assigned in more than one subregion and were consistently assigned to the same regulatory class. Interestingly, just as many genes, 4975 (32.5%), were assigned to two different regulatory classes. A smaller, but considerable, number of 1171 (7.6%) genes was assigned to three different classes, and a fraction of genes, 109 (0.7%) and 6 (<0.1%), were even assigned to four or five different classes, respectively (Figure 5a). The remaining 2447 genes (16.0%) could not be assigned to any regulatory class in any single subregion at the applied posterior probability threshold of 0.75.

Considering only genes that were consistently assigned to one class, the "*cis*" class was again the largest, followed by the "compensatory", "conserved", "*cis × trans*", "*trans*" and "*cis + trans*" classes (Figure 5b), akin to the frequencies observed when we considered subregions separately (Figure 4c). Genes consistently assigned to the "compensatory" and "conserved" classes were
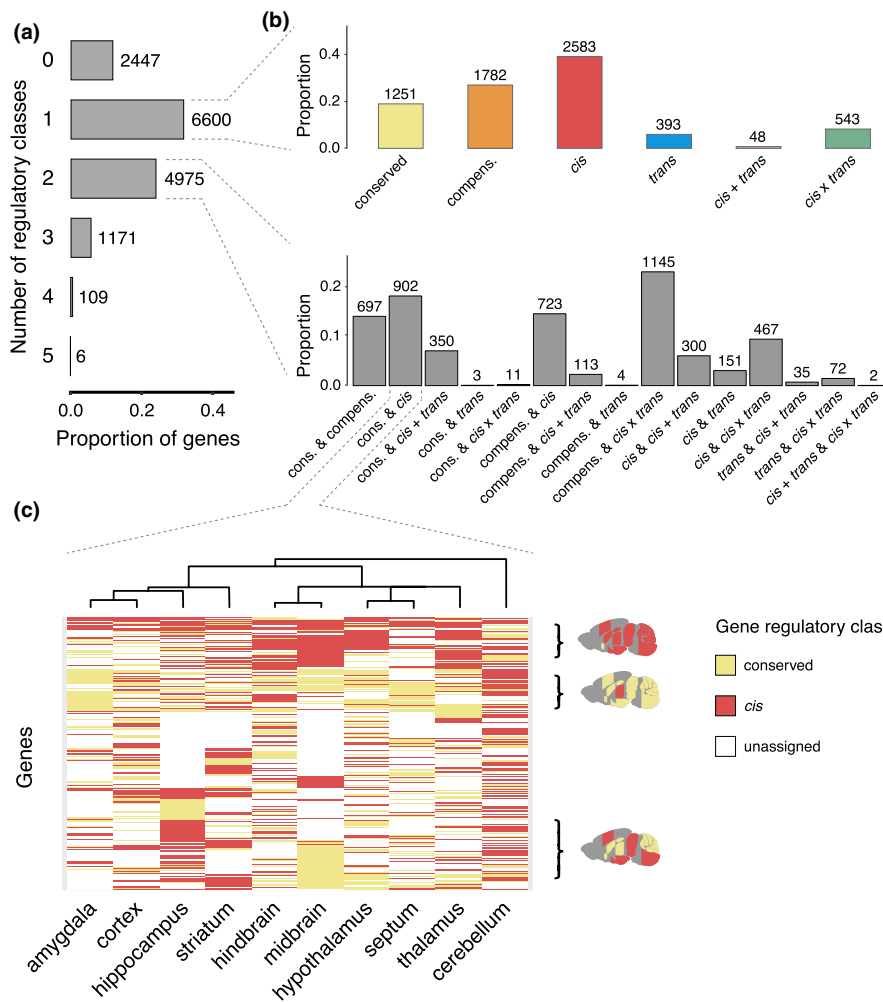

FIGURE 5 Shared versus modular gene regulation across subregions of the brain. (a) Proportion and number of genes that were assigned to either none, only one or up to 5 different regulatory classes across subregions. (b) Breakdown of the proportion of genes consistently assigned to one of the 6 classes (top) or one of the 15 two-class combinations for genes consistently assigned to only one or exactly two different classes (bottom). (c) Regulatory class assignment across subregions of all genes (rows) that exhibited both conserved gene expression and differential expression due to *cis*-regulatory divergence in at least one subregion. Some genes showed differential expression due to *cis*-regulatory divergence in all but one subregion (where gene expression was conserved) and vice versa or a mix of regulation patterns, indicated by brain icons on the right. Subregions and genes were clustered based on Gower distance. Genes that could not be assigned at the applied posterior probability threshold of 0.75 in any given subregion are shown in white.

enriched for the terms "preweaning lethality" and "abnormal survival" (Mammalian Phenotype Ontology database), indicating that they might be partially under strong selective constraint. Genes consistently in the "*cis* × *trans*" class, in which *cis*- and *trans*-regulatory differences act in a compensatory but not completely balanced way were enriched for terms involved in nuclear processes, including "single-stranded RNA binding" and "translation factor activity, RNA binding" (Molecular Function GO database), and "metabolism of RNA" (Reactome pathway). In contrast, genes assigned to the "*cis*" class were enriched for the term "receptor regulator activity" (Molecular Function GO database). The two smallest classes, "*trans*" and "*cis* + *trans*" were not statistically significantly enriched for any terms.

Of the genes that were assigned to two different classes, the combination of "compensatory" and "*cis*×*trans*" was the largest (Figure 5b), which is not unexpected given that the "compensatory" class can be considered a special (completely balanced) case of "*cis*×*trans*" regulatory divergence. In contrast, genes that are differentially expressed between the parental species in some subregions but not others imply that regulation is mediated by subregion-specific transcription factors. Focusing on the "conserved & *cis*" class, the second largest class overall, we found a broad spectrum of regulation patterns, ranging from genes being affected by *cis*-regulatory divergence in most subregions to the opposite extreme of genes whose regulation is conserved in almost all subregions (Figure 5c). Based on the (categorical) assignment of genes in this subset the cerebellum was again the most distinct subregion, standing apart from one cluster comprising the amygdala, cortex, hippocampus, and striatum and another cluster comprising the hindbrain, midbrain, hypothalamus, septum, and thalamus (Figure 5c). In conclusion, we find that the evolution of gene expression divergence between the two focal species happens to a considerable extent through subregion-specific mechanisms across the brain.

4 | DISCUSSION

The evolution of gene expression in the brain and how it contributes to the evolution of behaviour remains poorly understood. Here, we focused on two closely related yet behaviourally divergent species and compared gene expression across brain regions, between species, and between sexes as a first step towards understanding how their brain transcriptomes evolve. We find that slightly more of the variation in gene expression is across brain subregions rather than between species, echoing similar patterns found in comparative gene expression studies of different tissues (Brawand et al., 2011; Merkin et al., 2012). This suggests an evolutionarily conserved transcriptional program for each subregion of the brain that may reflect the specialized functions of each subregion. Region-specific genes are enriched for both hormone synthesis and hormone receptors which may reflect the important nature of hormone signalling for carrying out specific brain functions. However, we find only a small number of region-specific genes, suggesting that these specialized functions arise from the coordination of many genes acting in concert rather than through single genes.

Overall, subregion transcriptomes form clusters that reflect their developmental history. Cortex, hippocampus, amygdala, and striatum form one cluster, while midbrain, hypothalamus, and thalamus form another, with septum falling in between. Cerebellum is the most distinct subregion of all. The clustering of hindbrain with midbrain in our data may reflect difficulties in dissecting pure subregion samples or batch effects. These associations are also reflected in differential expression (DE) patterns between species, implying the presence of shared regulatory programs across subregions.

Importantly, even though our data imply conserved transcriptional profiles for each subregion, we also detect pervasive DE between species: most of the brain transcriptome is differentially expressed in at least one subregion. This is in line with expectations, given extensive variation in gene expression observed even among strains of a single species (Nadler et al., 2006). We did not find any bias of gene upregulation in any one species or any specific subregion. Given that our two focal species differ in their mating system, we were interested in testing whether our data would support previous work that proposed a set of candidate genes robustly associated with monogamous mating systems across vertebrates (Young et al., 2019). Our results do not provide evidence for this universal transcriptomic signature, consistent with a re-analysis of Young et al., 2019 that did not find evidence of transcriptome-wide parallel evolution in the repeated evolution of monogamy (Jiang & Zhang, 2019). We note, however, that behavioural transitions may induce gene expression differences in the brain (Hu et al., 2022; Ray et al., 2016), which could neither have been captured by our data nor those of Young et al., 2019. Therefore, future studies may reveal DE changes related to monogamy initiated by reproductive events.

Sex-specific behaviours such as parental care behaviour are also known to correlate with mating system. In *Peromyscus* deer mice, large differences in parental behaviour are observed between our two focal species, but the differences are more pronounced in fathers than mothers (Bendesky et al., 2017). Motivated by this observation, we looked at differential expression across the sexes, but did not find obvious patterns of increased sex bias in one species or the other, neither in overall number nor in expression magnitude. Overall, we found that sex-specific DE is only partially shared between species and we find considerable turnover of sex-biased genes, even across short evolutionary timescales. However, we note that our bulk RNA-sequencing strategy may not provide the spatial and cellular resolution to accurately characterize functionally important sex-biases in expression that may be restricted to very specific cell types (Kim et al., 2019; Xu et al., 2012). We believe this avenue of research may be of interest for future studies.

Finally, we sought to understand the regulatory mechanisms underlying the evolution of gene expression, leveraging F_1 hybrids to differentiate *cis*- and *trans*-regulatory factors and assign genes to gene regulatory classes (e.g. *cis*, *trans*, *cis* & *trans*, etc.). Almost half the transcriptome is conserved in expression due to lack of regulatory differences or compensatory gene regulation. This suggests that

a large portion of the transcriptome may be under stabilizing selection and explains the conserved transcriptional programs we found for each brain subregion. Of the genes differentially expressed between species, we found that the majority is driven by *cis*-regulatory factors. The predominance of *cis*-regulatory changes in expression evolution between species has been found repeatedly across many animal taxa including Mexican cavefish (Leclercq et al., 2022), stickleback (Verta & Jones, 2019), *Mus* (Reuveni et al., 2018) and *Drosophila* (Benowitz et al., 2020). These observations and our data are consistent with the prediction that selection on gene expression would be driven primarily from *cis*- rather than *trans*-regulatory evolution due to the reduced pleiotropy of *cis*-acting changes (which tend to occur in promoters or enhancers of target genes) compared to *trans*-regulatory changes (which tend to affect transcription factors that have downstream effects on large networks of genes) (Signor & Nuzhdin, 2018). While many genes were assigned exclusively to a single regulatory class, we find over 6000 genes that are evolving via two or more different regulatory classes across subregions of the brain, suggesting that region-specific regulatory evolution is widespread. This set of genes may include those that are undergoing region-specific selection and may be of interest for future follow-up studies interested in behaviours likely mediated by region-specific gene functions.

Our study highlights the complexity of gene expression evolution in the brain. Even between closely related species, we observe widespread DE. This DE is driven by a combination of correlated changes across several or all subregions and region-specific regulatory changes, predominantly in *cis*. We hypothesize that this generates modularity, which may be key in providing the evolutionary substrate for the myriad of behavioural phenotypes displayed by animals. Having extensively characterized gene expression differences and their regulatory basis across subregions, our study lays the groundwork for future investigations that seek to understand the mechanistic basis of behavioural evolution, in deer mice and beyond.

AUTHOR CONTRIBUTIONS

CLL, CH, AB and HEH conceived of the study. FB, KT and AB set up crosses to generate F_1 hybrids. CLL, CH, NLB and FB performed tissue dissections. KT generated libraries for sequencing. AFK, JC, CLL and JML analysed data. CLL, CH, AB and HEH acquired funding. AFK, JC and HEH wrote the manuscript with contributions from CLL, CH and KT. All authors approved of the final version of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw sequence reads and gene expression estimates have been deposited on SRA (PRJNA1027331) and GEO (GSE245182). Individual gene expression levels and gene regulation information can be accessed via https://peromyscus.shinyapps.io/brain-rnaseq-atlas/. Custom code to reproduce these results has been uploaded to GitHub (https://github.com/akautt/Pero_brain_RNAseq).

ORCID

Andreas F. Kautt https://orcid.org/0000-0001-7792-0735
Andres Bendesky https://orcid.org/0000-0002-4220-3653

REFERENCES

Alaux, C., Sinha, S., Hasadsri, L., Hunt, G. J., Guzmán-Novoa, E., DeGrandi-Hoffman, G., Uribe-Rubio, J. L., Southey, B. R., Rodriguez-Zas, S., & Robinson, G. E. (2009). Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proceedings of the National Academy of Sciences*, *106*(36), 15400–15405. https://doi.org/10.1073/pnas.0907043106

Barrett, C. E., Keebaugh, A. C., Ahern, T. H., Bass, C. E., Terwilliger, E. F., & Young, L. J. (2013). Variation in vasopressin receptor (Avpr1a) expression creates diversity in behaviors related to monogamy in prairie voles. *Hormones and Behavior*, *63*(3), 518–526. https://doi.org/10.1016/j.yhbeh.2013.01.005

Bedford, N. L., & Hoekstra, H. E. (2015). The natural history of model organisms: Peromyscus mice as a model for studying natural variation. *eLife*. https://doi.org/10.7554/eLife.06813

Bendesky, A., Kwon, Y.-M., Lassance, J.-M., Lewarch, C. L., Yao, S., Peterson, B. K., He, M. X., Dulac, C., & Hoekstra, H. E. (2017). The genetic basis of parental care evolution in monogamous mice. *Nature*, *544*(7651), 434–439. https://doi.org/10.1038/nature22074

Benowitz, K. M., Coleman, J. M., Allan, C. W., & Matzkin, L. M. (2020). Contributions of cis- and trans-regulatory evolution to transcriptomic divergence across populations in the Drosophila mojavensis larval brain. *Genome Biology and Evolution*, *12*(8), 1407–1418. https://doi.org/10.1093/gbe/evaa145

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. https://doi.org/10.1038/nature10532

Clark, F. H. (1938). Age of sexual maturity in mice of the genus Peromyscus. *Journal of Mammalogy*, *19*(2), 230–234. https://doi.org/10.2307/1374620

Dawson, W. D., Lake, C. E., & Schumpert, S. S. (1988). Inheritance of burrow building in Peromyscus. *Behavior Genetics*, *18*(3), 371–382. https://doi.org/10.1007/BF01260937

Dewey, M. J., & Dawson, W. D. (2001). Deer mice: The Drosophila of North American mammalogy. *Genesis*, *29*(3), 105–109. https://doi.org/10.1002/gene.1011

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Duboué, E. R., Keene, A. C., & Borowsky, R. L. (2011). Evolutionary convergence on sleep loss in cavefish populations. *Current Biology*, 21(8), 671–676. https://doi.org/10.1016/j.cub.2011.03.020

Fiddes, I. T., Armstrong, J., Diekhans, M., Nachtweide, S., Kronenberg, Z. N., Underwood, J. G., Gordon, D., Earl, D., Keane, T., Eichler, E. E., Haussler, D., Stanke, M., & Paten, B. (2018). Comparative annotation toolkit (CAT)—Simultaneous clade and personal genome annotation. *Genome Research*, 28(7), 1029–1038. https://doi.org/10.1101/gr.233460.117

Foltz, D. W. (1981). Genetic evidence for long-term monogamy in a small rodent, *Peromyscus polionotus*. *The American Naturalist*, 117(5), 665–675. https://doi.org/10.1086/283751

Fraser, H. B., Moses, A. M., & Schadt, E. E. (2010). Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proceedings of the National Academy of Sciences*, 107(7), 2977–2982. https://doi.org/10.1073/pnas.0912245107

Gibson, G., Riley-Berger, R., Harshman, L., Kopp, A., Vacha, S., Nuzhdin, S., & Wayne, M. (2004). Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics*, 167(4), 1791–1799. https://doi.org/10.1534/genetics.104.026583

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., Brazma, A., Odom, D. T., & Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, 22(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Hu, C. K., York, R. A., Metz, H. C., Bedford, N. L., Fraser, H. B., & Hoekstra, H. E. (2022). Cis-regulatory changes in locomotor genes are associated with the evolution of burrowing behavior. *Cell Reports*, 38(7), 110360. https://doi.org/10.1016/j.celrep.2022.110360

Jiang, D., & Zhang, J. (2019). Parallel transcriptomic changes in the origins of divergent monogamous vertebrates? *Proceedings of the National Academy of Sciences*, 116(36), 17627–17628. https://doi.org/10.1073/pnas.1910749116

Jourjine, N., Woolfolk, M. L., Sanguinetti-Scheck, J. I., Sabatini, J. E., McFadden, S., Lindholm, A. K., & Hoekstra, H. E. (2023). Two pup vocalization types are genetically and functionally separable in deer mice. *Current Biology*, 33(7), 1237–1248.e4. https://doi.org/10.1016/j.cub.2023.02.045

Kim, D.-W., Yao, Z., Graybuck, L. T., Kim, T. K., Nguyen, T. N., Smith, K. A., Fong, O., Yi, L., Koulena, N., Pierson, N., Shah, S., Lo, L., Pool, A.-H., Oka, Y., Pachter, L., Cai, L., Tasic, B., Zeng, H., & Anderson, D. J. (2019). Multimodal analysis of cell types in a hypothalamic node controlling social behavior. *Cell*, 179(3), 713–728.e17. https://doi.org/10.1016/j.cell.2019.09.020

Krijthe, J. H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation [Computer software]. https://github.com/jkrijthe/Rtsne

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., & Hartl, D. L. (2005). Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics*, 171(4), 1813–1822. https://doi.org/10.1534/genetics.105.047449

Leclercq, J., Torres-Paz, J., Policarpo, M., Agnès, F., & Rétaux, S. (2022). Evolution of the regulation of developmental gene expression in blind Mexican cavefish. *bioRxiv*. https://doi.org/10.1101/2022.07.12.499770

Lewarch, C. L., & Hoekstra, H. E. (2018). The evolution of nesting behaviour in Peromyscus mice. *Animal Behaviour*, 139, 103–115. https://doi.org/10.1016/j.anbehav.2018.03.008

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323.

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, 47(W1), W199–W205. https://doi.org/10.1093/nar/gkz401

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). https://doi.org/10.1186/s13059-014-0550-8

Lu, L., Airey, D. C., & Williams, R. W. (2001). Complex trait analysis of the hippocampus: Mapping and biometric analysis of two novel gene loci with specific effects on hippocampal structure in mice. *Journal of Neuroscience*, 21(10), 3503–3514. https://doi.org/10.1523/JNEUROSCI.21-10-03503.2001

Markow, T. A., & O'Grady, P. (2008). Reproductive ecology of drosophila. *Functional Ecology*, 22(5), 747–759. https://doi.org/10.1111/j.1365-2435.2008.01457.x

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17, 10–12. https://doi.org/10.14806/ej.17.1.200

McGirr, J. A., & Martin, C. H. (2019). Hybrid gene misregulation in multiple developing tissues within a recent adaptive radiation of Cyprinodon pupfishes. *PLoS One*, 14(7), e0218899. https://doi.org/10.1371/journal.pone.0218899

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in drosophila revealed by mRNA-seq. *Genome Research*, 20(6), 816–825. https://doi.org/10.1101/gr.102491.109

Merkin, J., Russell, C., Chen, P., & Burge, C. B. (2012). Evolutionary dynamics of gene and Isoform regulation in mammalian tissues. *Science*, 338(6114), 1593–1599. https://doi.org/10.1126/science.1228186

Nadler, J. J., Zou, F., Huang, H., Moy, S. S., Lauder, J., Crawley, J. N., Threadgill, D. W., Wright, F. A., & Magnuson, T. R. (2006). Large-scale gene expression differences across brain regions and inbred strains correlate with a behavioral phenotype. *Genetics*, 174(3), 1229–1236. https://doi.org/10.1534/genetics.106.061481

Naqvi, S., Godfrey, A. K., Hughes, J. F., Goodheart, M. L., Mitchell, R. N., & Page, D. C. (2019). Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science*, 365(6450), aaw7317. https://doi.org/10.1126/science.aaw7317

Orr, H. A. (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, 149(4), 2099–2104.

Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro, S. W., Needleman, L. A., Ho-Shing, O., Choi, J. J., Zirlinger, M., Chen, S.-K., Liu, J. S., & Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *eLife*, 4, e07860. https://doi.org/10.7554/eLife.07860

Pizzollo, J., Zintel, T. M., & Babbitt, C. C. (2022). Differentially active and conserved neural enhancers define two forms of adaptive non-coding evolution in humans. *Genome Biology and Evolution*, 14(8), evac108. https://doi.org/10.1093/gbe/evac108

Ray, S., Tzeng, R.-Y., DiCarlo, L. M., Bundy, J. L., Vied, C., Tyson, G., Nowakowski, R., & Arbeitman, M. N. (2016). An examination of dynamic gene expression changes in the mouse brain during pregnancy and the postpartum period. *G3: Genes, Genomes, Genetics*, 6(1), 221–233. https://doi.org/10.1534/g3.115.020982

Reuveni, E., Getselter, D., Oron, O., & Elliott, E. (2018). Differential contribution of cis and trans gene transcription regulatory mechanisms in amygdala and prefrontal cortex and modulation by social stress. *Scientific Reports*, 8(1), 6339. https://doi.org/10.1038/s41598-018-24544-3

Schenk, J. J., Rowe, K. C., & Steppan, S. J. (2013). Ecological opportunity and incumbency in the diversification of repeated continental Colonizations by Muroid rodents. *Systematic Biology*, 62(6), 837–864. https://doi.org/10.1093/sysbio/syt050

Shen, S. Q., Turro, E., & Corbo, J. C. (2014). Hybrid mice reveal parent-of-origin and cis- and trans-regulatory effects in the retina. *PLoS One*, *9*(10), e109382. https://doi.org/10.1371/journal.pone.0109382

Signor, S. A., & Nuzhdin, S. V. (2018). The evolution of gene expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Sumner, F. B. (1922). Longevity in Peromyscus. *Journal of Mammalogy*, *3*(2), 79–81. https://doi.org/10.2307/1373298

Tirosh, I., Reikhav, S., Levy, A. A., & Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*, *324*(5927), 659–662. https://doi.org/10.1126/science.1169766

Turro, E., Astle, W. J., & Tavaré, S. (2014). Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics*, *30*(2), 180–188. https://doi.org/10.1093/bioinformatics/btt624

Turro, E., Su, S.-Y., Gonçalves, Â., Coin, L. J., Richardson, S., & Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biology*, *12*(2), R13. https://doi.org/10.1186/gb-2011-12-2-r13

Urbut, S. M., Wang, G., Carbonetto, P., & Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics*, *51*(1), 187–195. https://doi.org/10.1038/s41588-018-0268-8

Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife*, *8*, 43785. https://doi.org/10.7554/eLife.43785

Vrana, P. B., Fossella, J. A., Matteson, P., Rio, T., O'Neill, M. J., & Tilghman, S. M. (2000). Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in Peromyscus. *Nature Genetics*, *25*(1), 120–124. https://doi.org/10.1038/75518

Wang, Q., Jia, Y., Wang, Y., Jiang, Z., Zhou, X., Zhang, Z., Nie, C., Li, J., Yang, N., & Qu, L. (2019). Evolution of cis- and trans-regulatory divergence in the chicken genome between two contrasting breeds analyzed using three tissue types at one-day-old. *BMC Genomics*, *20*(1), 933. https://doi.org/10.1186/s12864-019-6342-5

Watson, M. L. (1942). Hybridization experiments between *Peromyscus polionotus* and *Peromyscus maniculatus*. *Journal of Mammalogy*, *23*(3), 315–316. https://doi.org/10.2307/1375001

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between *drosophila*

species. *Nature Genetics*, *40*(3), 346–350. https://doi.org/10.1038/ng.77

Xu, X., Coats, J. K., Yang, C. F., Wang, A., Ahmed, O. M., Alvarado, M., Izumi, T., & Shah, N. M. (2012). Modular genetic control of sexually dimorphic behaviors. *Cell*, *148*(3), 596–607. https://doi.org/10.1016/j.cell.2011.12.018

Young, L. J., Nilsen, R., Waymire, K. G., MacGregor, G. R., & Insel, T. R. (1999). Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. *Nature*, *400*, 766–768. https://doi.org/10.1038/23475

Young, R. L., Ferkin, M. H., Ockendon-Powell, N. F., Orr, V. N., Phelps, S. M., Pogány, Á., Richards-Zawacki, C. L., Summers, K., Székely, T., Trainor, B. C., Urrutia, A. O., Zachar, G., O'Connell, L. A., & Hofmann, H. A. (2019). Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proceedings of the National Academy of Sciences*, *116*(4), 1331–1336. https://doi.org/10.1073/pnas.1813775116

Zhang, M., Pan, X., Jung, W., Halpern, A., Eichhorn, S. W., Lei, Z., Cohen, L., Smith, K. A., Tasic, B., Yao, Z., Zeng, H., & Zhuang, X. (2023). A molecularly defined and spatially resolved cell atlas of the whole mouse brain. *bioRxiv*. https://doi.org/10.1101/2023.03.06.531348

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.