Comparative neurotranscriptomics reveal widespread species differences associated with bonding

Joel A. Tripp[1†], Alejandro Berrio[1,2†], Lisa A. McGraw[3], Mikhail V. Matz[1], Jamie K. Davis[4], Kiyoshi Inoue[3], James W. Thomas[5], Larry J. Young[3†] and Steven M. Phelps[1*†]

Abstract

Background: Pair bonding with a reproductive partner is rare among mammals but is an important feature of human social behavior. Decades of research on monogamous prairie voles (*Microtus ochrogaster*), along with comparative studies using the related non-bonding meadow vole (*M. pennsylvanicus*), have revealed many of the neural and molecular mechanisms necessary for pair-bond formation in that species. However, these studies have largely focused on just a few neuromodulatory systems. To test the hypothesis that neural gene expression differences underlie differential capacities to bond, we performed RNA-sequencing on tissue from three brain regions important for bonding and other social behaviors across bond-forming prairie voles and non-bonding meadow voles. We examined gene expression in the amygdala, hypothalamus, and combined ventral pallidum/ nucleus accumbens in virgins and at three time points after mating to understand species differences in gene expression at baseline, in response to mating, and during bond formation.

Results: We first identified species and brain region as the factors most strongly associated with gene expression in our samples. Next, we found gene categories related to cell structure, translation, and metabolism that differed in expression across species in virgins, as well as categories associated with cell structure, synaptic and neuroendocrine signaling, and transcription and translation that varied among the focal regions in our study. Additionally, we identified genes that were differentially expressed across species after mating in each of our regions of interest. These include genes involved in regulating transcription, neuron structure, and synaptic plasticity. Finally, we identified modules of co-regulated genes that were strongly correlated with brain region in both species, and modules that were correlated with post-mating time points in prairie voles but not meadow voles.

* Correspondence: sphelps@utexas.edu
†Joel A. Tripp and Alejandro Berrio contributed equally to this work.
†Larry J. Young and Steven M. Phelps contributed equally to this work.
[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA
Full list of author information is available at the end of the article

Conclusions: These results reinforce the importance of pre-mating differences that confer the ability to form pair bonds in prairie voles but not promiscuous species such as meadow voles. Gene ontology analysis supports the hypothesis that pair-bond formation involves transcriptional regulation, and changes in neuronal structure. Together, our results expand knowledge of the genes involved in the pair bonding process and open new avenues of research in the molecular mechanisms of bond formation.

Keywords: Vole, Pair bond, RNA-sequencing, Amygdala, Hypothalamus, Nucleus accumbens, Ventral pallidum

Background

The profound social attachments humans form with one another are a defining feature of our species. While many species develop parent-offspring or mate pair bonds, these attachments are especially strong and long-lasting in humans, particularly those between reproductive partners [1]. In fact, the social bonds formed by human partners are quite uncommon among mammals: only 9% of mammalian species display social monogamy, including just 6% of rodents, which are often used as model species to investigate the neural basis of behavior [2]. Importantly, commonly used laboratory species including mice and rats are not monogamous and lack the ability to form lasting pair bonds. A better understanding of the mechanisms underlying the formation of partner bonds requires use of a species that forms such bonds.

The voles of the genus *Microtus* (Fig. 1a, b) have been developed into powerful models for better understanding the neural mechanisms underlying attachment and pair bonding [5]. Typically, the focal species of these studies is the prairie vole (*Microtus ochrogaster*, Fig. 1a), which has become famous for the tendency for males and females to form long-lasting, socially monogamous bonds. Field studies of prairie vole space use and behavior demonstrated a socially monogamous mating system [6, 7] (though extra-pair fertilization does occur and some individuals adopt more promiscuous mating tactics [8, 9]). In the lab, receptive virgin prairie voles will mate with novel opposite-sex conspecifics and, over a period of several hours of mating and other affiliative behaviors, will form a pair bond characterized by selective affiliation and increased aggression towards intruders [3, 4, 6, 7].

Prairie voles are often contrasted with closely related species which do not form these bonds [3, 10–14], including the meadow vole (*Microtus pennsylvanicus*, Fig. 1b). Field studies of meadow voles are consistent with a promiscuous mating system and in the lab, they do not show the selective affiliation with mating partners characteristic of pair bonds [12, 15, 16]. Comparative studies have been particularly useful in uncovering the features of the prairie vole brain that facilitate bond formation. These include neuroendocrine signaling in the brain (particularly via the nonapeptides oxytocin [OT] and arginine-vasopressin [AVP]) which conveys social salience, mechanisms of social recognition and memory, and dopaminergic reward pathways [17, 18]. Current hypotheses propose that pair bonding results from synaptic plasticity that links the neural encoding of partner cues with the reward system, leading to persistent reinforcement of the partner that results in selective affiliation [17].

While significant progress has been made in understanding the neural mechanisms underlying vole pair-bond formation, previous studies have mainly focused on a relatively small number of neuromodulatory pathways, including the nonapeptides and dopamine reward systems mentioned above, as well as opioid signaling in the brain [17, 18]. In order to identify additional genes that may be playing a significant role in bond formation, it is necessary to take a broader perspective and examine changes to global gene expression in regions with a known role in bonding.

In this study, we used a comparative approach, taking advantage of the close evolutionary relationship [19] and stark contrast in bonding behavior between prairie and meadow voles to better understand the molecular mechanisms that support pair-bond formation and identify novel candidate genes for future study. Our study focused on three regions (Fig. 1c): the amygdala (AMY), hypothalamus (HT), and a region inclusive of the ventral pallidum and nucleus accumbens (VP/NAc). Each of these regions is proposed to have a critical function in the development of a pair bond [17]. The rodent AMY receives olfactory information important for social recognition [20, 21], encodes valence [22], and plays an important role in social memory [23, 24]. The HT contains several nuclei that are involved in regulating social behavior, including populations that produce the neuropeptides OT and AVP, which signal social salience, influence social memory, and regulate related behaviors [12, 25–28]. Finally, signals of social salience converge with dopaminergic signaling from the ventral tegmental area in the NAc, resulting in disinhibition of the VP, which influences behavioral responses to rewarding stimuli [29–32]. While several other regions, such as the ventral tegmental area, hippocampus, and cortical regions also have important functions in the development and maintenance of pair bonds [17, 18], our focal regions represent key nodes involved in critical aspects of bonding: social context, social memory, and reward.
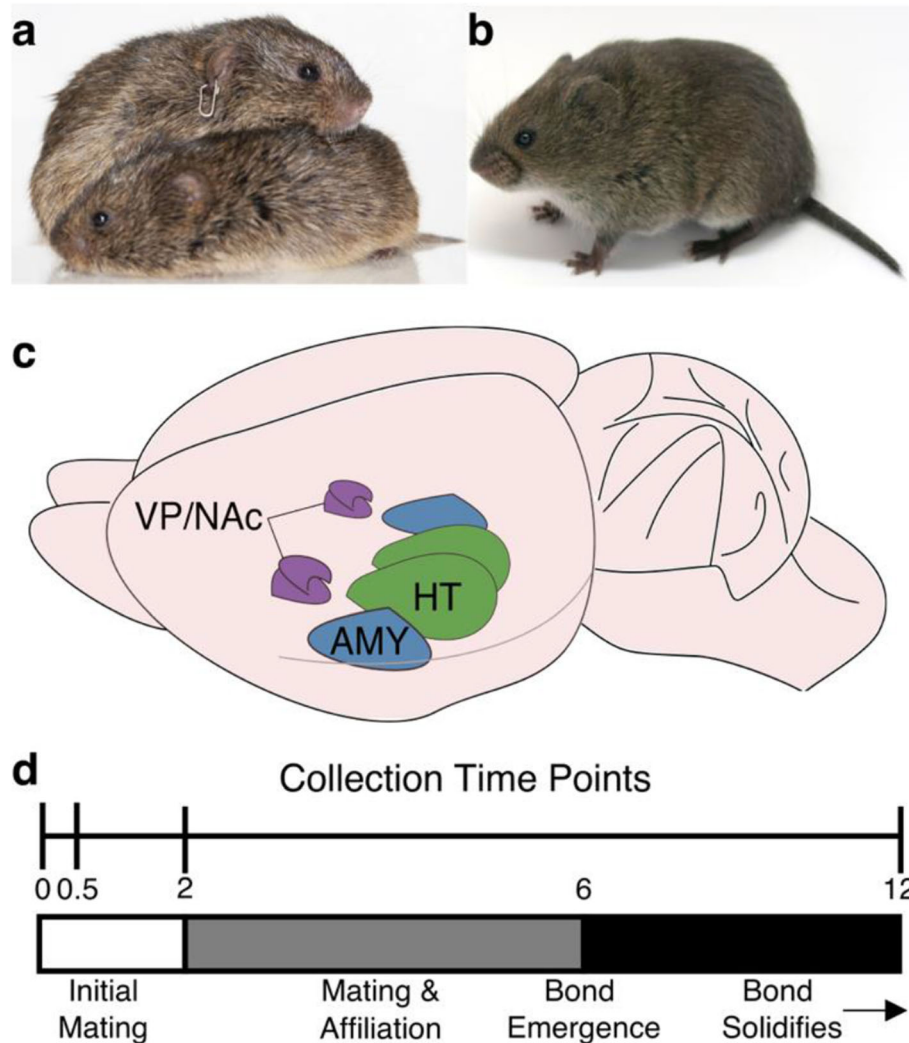
Fig. 1 Experiment overview. **a** Pair of prairie voles (*Microtus ochrogaster*) huddling (Photo credit: Aubrey M. Kelly) **b** A single adult meadow vole (*Microtus pennsylvanicus*) (Photo credit: Beery Lab) **c** Drawing of vole brain. Regions of interest for this study are highlighted, including the amygdala (AMY, blue), hypothalamus (HT, green), and ventral pallidum and nucleus accumbens (VP/NAc, purple). **d** Above: Experiment Timeline. Virgin animals (time 0) were collected prior to mating. Mated animals were collected at either 0.5, 2, or 12 h after first intromission. Below: Timeline of prairie vole bond formation. Unfamiliar opposite sex conspecifics will mate shortly after introduction. Bonds become detectable at 6 h after onset of mating and strengthen up to 24 h after mating onset [3, 4]

We used RNA-sequencing to quantify gene expression before and at three time points (0.5, 2, and 12 h) after mating (Fig. 1d) in each of these regions across both prairie and meadow voles. By observing changes in gene expression in these regions across time relative to mating and across pair-bonding and non-bonding species, we sought to identify genes involved in the process of pair-bond formation, either by conferring the capacity to bond before mating or changing their expression to support bonding after mating. We find gene categories that differ across species at baseline (pre-mating) and across brain regions, as well as genes that differ in expression post-mating between prairie and meadow voles. Further, using gene network analysis, we identify gene modules

that have expression patterns strongly related to specific brain regions, as well as modules that change in expression in response to mating. Together, these results provide further support for the current models of mammalian pair bonding and provide additional candidate genes that may underlie the formation of monogamous bonds.

Results

In order to identify the factors (e.g. species, brain region, time relative to mating) most strongly associated with gene expression in our samples, we first performed hierarchical clustering based on Poisson distance [33], followed by principal component analysis (PCA). In both

cases, the primary factor influencing gene expression was species, followed by brain region (Fig. 2). Hierarchical clustering based on Poisson distance resulted in a tree with two major branches representing samples from prairie and meadow voles. These branches were then split corresponding to brain region. In prairie voles, AMY and VP/NAc clustered more closely to each other, while in meadow voles AMY and HT were more closely clustered (Fig. 2a). This result was also supported by PCA, which revealed a first principal component (PC1) that explained 39% of the variance in the data and split samples by species, followed by a second component (PC2) that explained 34% of the variance and primarily separated samples by brain region (Fig. 2b).

Differential gene expression
Species comparison of virgins
To identify genes that may confer the capacity to bond in prairie but not meadow voles prior to mating, we used a subset of the data containing only samples collected from virgins (AMY, HT, and VP/NAc samples from $n = 4$ individuals of each species). We tested for the effect of species using the likelihood ratio test implemented by DESeq2 [34]. This test compares two generalized linear models fit to the data, a full model including all terms and a reduced model with one or more terms omitted. The test determines if the effect of the terms omitted in the reduced model is significantly greater than zero.
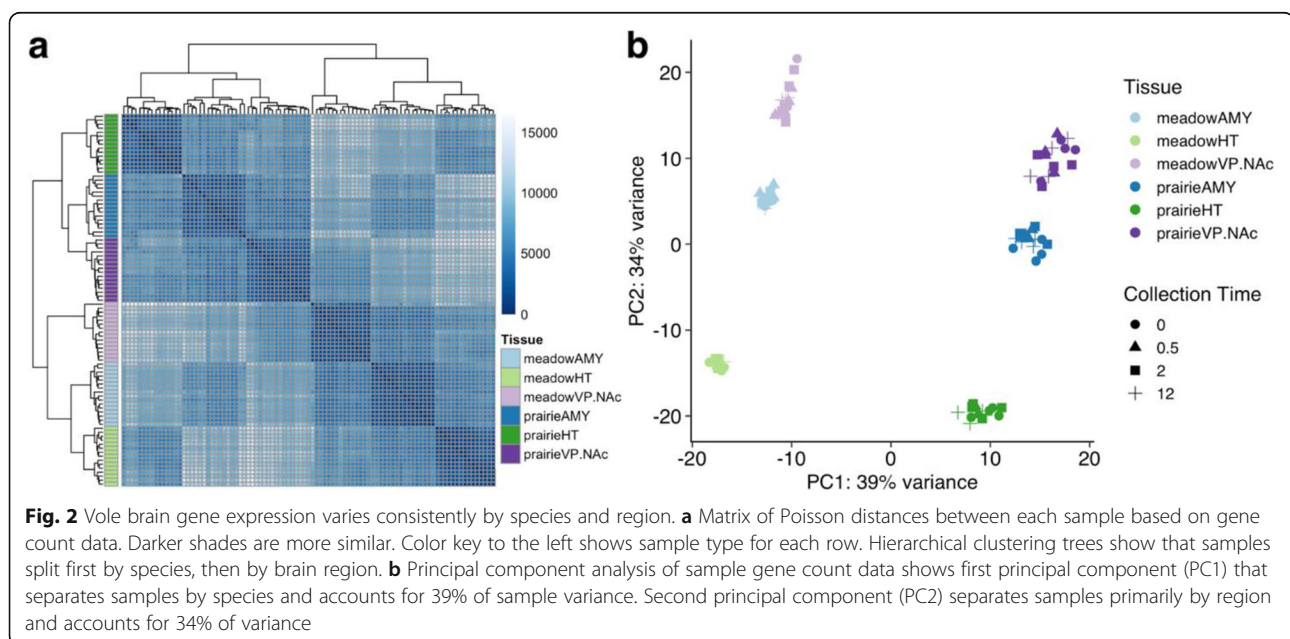
To determine the effect of species, we applied the full model ~region + species and reduced model ~region. 8377 of 12,672 genes were significantly differentially expressed (FDR < 0.1) in this comparison (Fig. 3a, Supplementary Table 1). In order to identify the types of

genes that differed in expression across species, we used the Mann-Whitney U (MWU) test (see Methods for details) to identify enriched gene ontology (GO) categories [36, 37]. The MWU test of GO term enrichment revealed 48 significantly enriched (FDR < 0.1) terms in the contrast between prairie and meadow vole virgins (Fig. 3b). This included 10 terms in the category Molecular Function, 11 in Biological Process, and 27 in the category Cellular Component. Enriched terms in prairie voles were largely associated with cell structure and extracellular space, while those enriched in meadow voles were associated with the ribosome and translation, mitochondrial function, and metabolic and biosynthetic processes.

Comparison of regions
As sample region was the second largest factor explaining gene expression data, we next compared expression across brain regions. To understand how gene expression differed across our focal brain regions, we used the entire dataset and applied the likelihood ratio test with the full model ~species + region + time and the reduced model ~species + time. Overall, 10,891 of 12,687 total genes significantly differed (FDR < 0.1) in expression on the basis of sample region (Fig. 4a-c, Supplementary Table 2).

Thirty-two GO terms were significantly enriched (MWU FDR < 0.1) in the contrast between AMY and HT, including two in the category Biological Process and 30 under Cellular Component (Fig. 4d). Terms enriched in AMY were largely associated with neuronal structure and the synapse, while those enriched in HT were associated with membranous structures including endoplasmic reticulum, Golgi apparatus, intracellular vesicle, and

Fig. 2 Vole brain gene expression varies consistently by species and region. **a** Matrix of Poisson distances between each sample based on gene count data. Darker shades are more similar. Color key to the left shows sample type for each row. Hierarchical clustering trees show that samples split first by species, then by brain region. **b** Principal component analysis of sample gene count data shows first principal component (PC1) that separates samples by species and accounts for 39% of sample variance. Second principal component (PC2) separates samples primarily by region and accounts for 34% of variance

Fig. 3 (See legend on next page.)

whole membrane. Twelve GO terms were enriched in the contrast between AMY and VP/NAc (Fig. 4e). These included six in the category Molecular Function, two under Biological Process, and four in Cellular Component. Categories enriched in AMY were largely associated with neuropeptide signaling, including terms related to neuropeptide hormones and receptors, as well as extracellular space. Terms enriched in VP/NAc

include *RNA polymerase II repressing transcription factor binding*, *endocytotic vesicle*, and *chromosomal part*. Finally, 42 terms were significantly enriched in the contrast between HT and VP/NAc, including three under Molecular Function, three in the category Biological Process, and 36 in the category Cellular Component (Fig. 4f). GO terms enriched in HT were largely associated with neuropeptide signaling and extracellular space,

Fig. 4 Vole brain gene expression varies by tissue type. **a-c** Volcano plots showing significance ($-\log_{10}$(p-value)) and magnitude of difference ($\log_2$ fold change, LFC) in expression for each gene. Each point represents a single gene. Orange colored points pass significance threshold of FDR < 0.1. Darker shaded points pass FDR cutoff and have LFC > 2. **a** Contrast between AMY and HT. Genes with positive LFC values are enriched in AMY, genes with negative LFC values are enriched in HT. **b** Contrast between AMY and VP/NAc. Genes with positive LFC values are enriched in AMY, genes with negative LFC values are enriched in VP/NAc **c** Contrast between HT and VP/NAc. Genes with positive LFC values are enriched in HT, genes with negative LFC values are enriched in VP/NAc. **d-f** Enriched GO terms. Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of "good" genes that have raw p-value< 0.05 compared to total number of genes in the category. P-values are corrected using Benjamini-Hochberg false discovery rate procedure [35]. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category. **d** Contrast between AMY and HT. Red terms enriched in AMY, blue terms enriched in HT. **e** Contrast between AMY and VP/NAc. Red terms enriched in AMY, blue terms enriched in VP/NAc. **f** Contrast between HT and VP/NAc. Red terms enriched in HT, blue terms enriched in VP/NAc

while those enriched in VP/NAc were related to synaptic structures as well as the nucleus, transcription, and translation.

Species comparison across time

We next sought to determine the influence of mating and, in prairie voles, pair-bond formation on gene expression in our focal regions. In order to identify how gene expression changes differently across species following mating, we divided the full dataset into subsets containing all samples from a single region. We then tested these regional datasets with the likelihood ratio test, applying the full model ~species + time + species: time and the reduced model ~species + time. This comparison allowed us to identify those genes that exhibit species-specific patterns of expression following mating.

In the AMY, we found four genes that were significantly differentially expressed (FDR < 0.1) across species over mating time points (Fig. 5a, Table 1, Supplementary Table 3). The gene with the smallest adjusted *p*-value (FDR = $6.17*10^{-5}$) in this model was *Npas4*, an activity-dependent transcription factor involved in synapse formation [38] (Fig. 5d). We found 33 significantly enriched GO terms between prairie and meadow voles (MWU FDR < 0.1), including two under Molecular Function and 31 in the category Cellular Component (Fig. 5g). Enriched terms in prairie voles were related to peptide receptors and the plasma membrane. Enriched terms in meadow voles were associated with ribosomal and mitochondrial function.

In the HT, 31 genes were significantly differentially expressed (FDR < 0.1) across species over mating time points (Fig. 5b, Supplementary Table 4), of which 25 were annotated (Table 2). These included water channel-encoding gene *Aqp3* [39] (Fig. 5e), which had the lowest adjusted *p*-value (FDR = 0.0089713). Sixty-five GO terms were significantly enriched (MWU FDR < 0.1) in the contrast between prairie and meadow voles. Among these were two in the category Molecular Function, 15 in Biological Process, and 48 in Cellular Component (Fig. 5h). Enriched terms in prairie voles were typically associated with neuronal development and cell structure. Terms enriched in meadow voles included many related to ribosomal and mitochondrial function, metabolic and biosynthetic processes, and vesicular transport.

In the VP/NAc we found 21 genes that were significantly differentially expressed over mating time points (Fig. 5c, Supplementary Table 5), 19 of which were annotated (Table 3). Among these, *Spata16*, a gene with known function in spermatogenesis [40] (Fig. 5f) had the lowest adjusted p-value (FDR = 0.00318981). In addition, we found 30 significantly enriched (MWU FDR < 0.1) GO terms between prairie and meadow voles, which

included five under Molecular Function, three under Biological Process, and 22 in Cellular Component (Fig. 5i). Only two of these terms were enriched in prairie voles: *cluster of actin-based cell projections* and *extracellular region*. Many of the terms enriched in meadow voles were associated with translation, protein localization, mitochondria, and the endosome.

Genes shared across comparisons

Several genes were differentially expressed in multiple comparisons. In region-by-region comparisons across species in virgin animals, the largest proportion of differentially expressed genes were significantly different in all three regions (Fig. 6a). Additionally, 1206 genes were differentially expressed in both HT and VP/NAc, 395 in AMY and VP/NAc, and 409 in AMY and HT. Relatively few genes were uniquely differentially expressed in a single region in this comparison, just 2499 compared to a total of 5031 genes that were differentially expressed in two or all three regions.

While not as large a proportion as in virgins, several genes were also differentially expressed in multiple regions in our species:time comparisons (Fig. 6b-h). The gene *Nfkbia*, which encodes NF-Kappa-B Inhibitor Alpha, was significant in our models of both AMY and VP/NAc (Fig. 6e). Five genes were significant in the species:time comparison for both HT and VP/NAc. These included *Pxn*, which encodes paxillin, an adaptor protein that interacts with the cytoskeleton and signaling molecules to facilitate neurite outgrowth and long-term potentiation [41, 42], the circadian clock gene *Per1* [43], *Cables1* which encodes a kinase-binding protein that promotes neurite outgrowth [44], as well as two unannotated genes (Fig. 6f-h). Each of these genes was consistently upregulated in one species over the other across tissues, with the exception of *Per1*, which nevertheless showed consistent patterns of expression over time by tissue and species (Fig. 6g).

Finally, many of the genes that significantly differed across species over post-mating time points were also differentially expressed across species in virgins (Fig. 6c, d). Three of the four genes differentially expressed in AMY at post-mating time points (all upregulated in prairie voles) were also upregulated in prairie vole virgins. For the HT, of 31 genes significant for the species: time interaction, 15 were upregulated in prairie voles in virgins and after mating and eight were upregulated in meadow voles as virgins and after mating. In the VP/NAc, 15 genes were upregulated in prairie voles as virgins and after mating, while two genes were upregulated in meadow voles in both comparisons.

Correlated gene networks

We next sought to determine if co-regulated groups of genes varied in expression related to our variables of
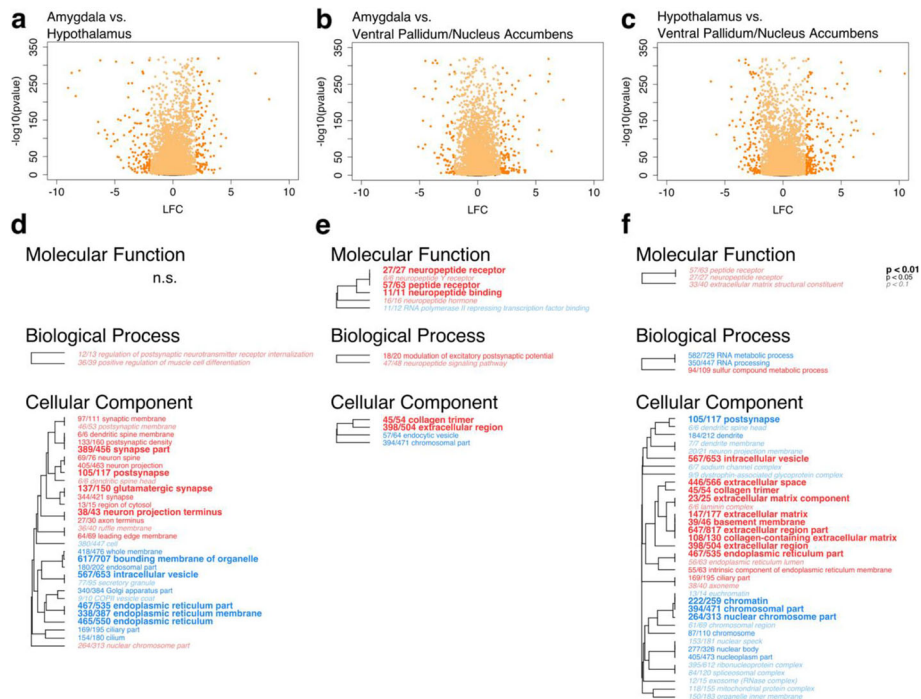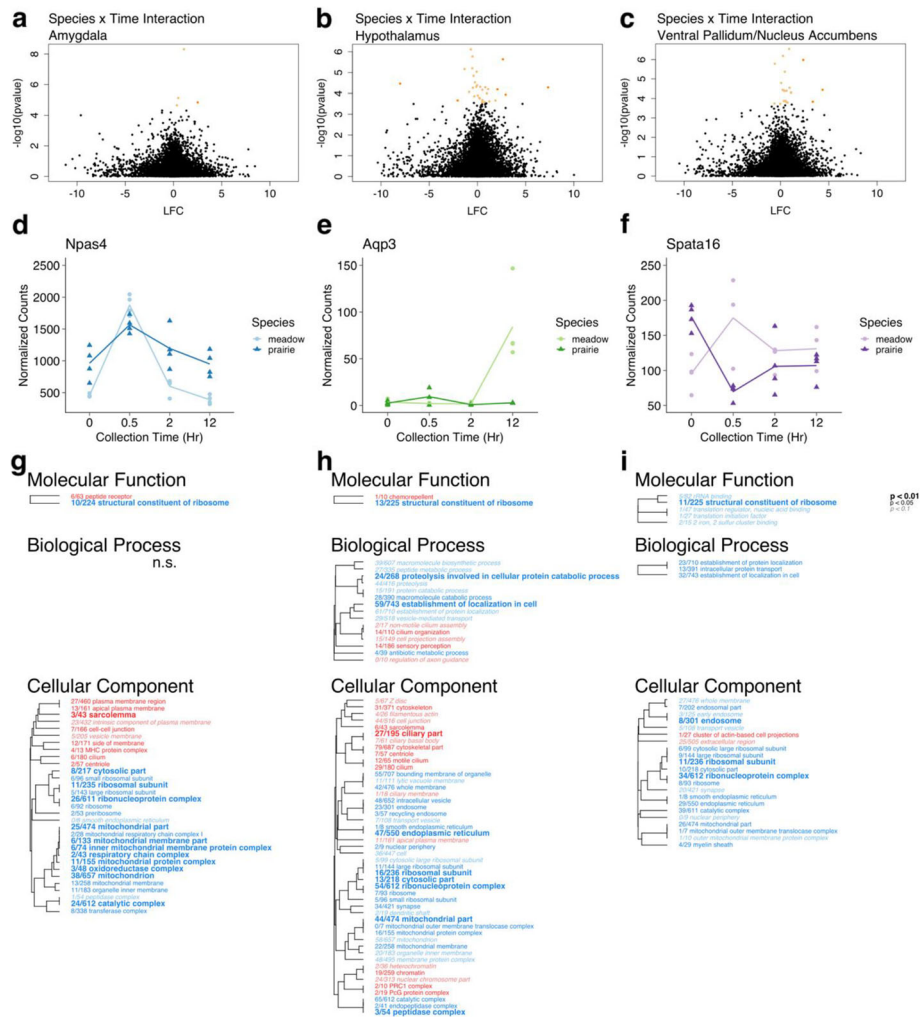
Fig. 5 Vole brain gene expression varies over time by species. **a-c** Volcano plots showing significance ($-\log_{10}$(p-value)) and magnitude of difference ($\log_2$ fold change, LFC) in expression for each gene in species:time interaction in **a** AMY, **b** HT, and **c** VP/NAc. Each point represents a single gene. Orange colored points pass significance threshold of FDR < 0.1. Darker shaded points pass FDR cutoff and have LFC > 2. Genes with positive LFC values are enriched in prairie voles, genes with negative LFC values are enriched in meadow voles. **d-f** Pattern of expression over time for genes with lowest adjusted p-value (FDR) in each regional model. Plots show normalized counts for samples collected at each pre-mating (0 h) and post-mating (0.5, 2, or 12 h) collection point. Each point represents one sample. Darker shades and triangles represent prairie vole samples. Lighter shades and circles represent meadow vole samples. **d** *Npas4* in AMY samples. **e** *Aqp3* in HT samples. **f** *Spata16* in VP/NAc samples. **g-i** Enriched GO terms for species:time interaction in **g** AMY, **h** HT, and **i** VP/NAc. Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of "good" genes that have raw p-value< 0.05 compared to total number of genes in the category. P-values are corrected using Benjamini-Hochberg false discovery rate procedure [35]. Red terms enriched in prairie voles, blue terms enriched in meadow voles. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category

Table 1 Differentially expressed genes in amygdala over time across species

Gene ID	Description	LFC	FDR
Npas4	Neuronal PAS domain containing protein 4A	1.07764799	6.17E-05
Adamts1	A disintegrin and metalloproteinase with thrombospondin motifs 1[a]	0.49359749	0.04640262
Cacna1s	Voltage-dependent L-type calcium channel subunit alpha-1S[a]	2.5010691	0.06006207
Nfkbia	NF kappa B inhibitor alpha	0.34440018	0.07022008

FDR False Discovery Rate, *LFC* $\log_2$ fold change; Genes with positive LFC have greater total expression in prairie voles, those with negative LFC have greater total expression in meadow voles. See Supplementary Table 3 for full model results. [a]Annotation obtained from UniprotKB (uniprot.org)

Tripp *et al. BMC Genomics*   (2021) 22:399

Page 9 of 22

Table 2 Differentially expressed genes in hypothalamus over time across species

Gene ID	Description	LFC	FDR
Aqp3	Aquaporin-3[a]	−0.6826921	0.0089713
Hsp90ab1[a]	Heat shock protein HSP 90 beta	−0.5144606	0.00903347
NA	Guanylate binding protein 6	−1.3369143	0.01596747
Serpine1	Plasminogen activator inhibitor 1[a]	−0.5338288	0.01596747
Ttr	Transthyretin	−0.080078	0.01792675
Prph	Peripherin	0.29781725	0.04287813
Frmpd1	FERM and PDZ domain containing protein 4	0.81441894	0.04287813
Slc6a18	Sodium-dependent neutral amino acid transporter B(0)AT3[a]	7.32918535	0.04287813
Wrap53	Telomerase Cajal body protein 1[a]	−0.2558135	0.04287813
Slc25a4[a]	ADP/ATP translocase 1	−8.0122925	0.04287813
Per1[a]	Period circadian protein homolog 1	−0.1178588	0.04287813
Cables1	CDK5 and ABL1 enzyme substrate 2	1.11126916	0.04423695
Pxn[a]	Paxillin	0.5140673	0.04423695
Hif1a[a]	Hypoxia inducible factor 1 alpha	0.97379217	0.0588332
Erich2	Glutamate rich protein 2	2.92340668	0.06351925
Dnah8	Dynein heavy chain 8, axonemal[a]	0.02817643	0.06693531
Ccdc189	Coiled coil domain containing protein 189	−0.3298147	0.07033001
Foxo1	Forkhead box protein O1[a]	0.17775249	0.07965468
Elmo2	Engulfment and cell motility protein 2[a]	0.2559439	0.08211435
St6galnac2	Alpha N acetylgalactosaminide alpha-2,6-sialyltransferase 2	1.87200042	0.09137815
Ovol1	Putative transcription factor Ovo like 1	−2.0478213	0.09137815
Ccdc13[a]	Coiled coil domain containing protein 13	0.52641328	0.09137815
Micalcl[a]	MICAL C-terminal-like protein	1.23718647	0.09701189
Rasgrf2	Ras-specific guanine nucleotide-releasing factor 2[a]	0.74713328	0.09959283
Riox2	Ribosomal oxygenase 2[a]	0.48958706	0.09959283

FDR False Discovery Rate, *LFC* log$_2$ fold change, *NA* no annotation. Genes with positive LFC have greater total expression in prairie voles, those with negative LFC have greater total expression in meadow voles. Unannotated genes excluded. See Supplementary Table 4 for full model results. [a]Annotation obtained from UniprotKB (uniprot.org)

interest (sample region, mating status [virgin vs. mated], or time of collection after mating onset). Using the R package WGCNA [45], we constructed gene expression networks for prairie and meadow voles, and detected modules of genes with correlated expression patterns (Fig. 7, see Supplementary Tables 6 and 7 for module assignments and Supplementary Tables 8 and 9 for Pearson r and *p*-values for significant correlations). For prairie voles, we detected 16 modules ranging in size from 72 to 1707 genes as well as 392 that were unassigned to any module. In meadow voles we detected 12 modules ranging in size from 167 to 1577 genes and 517 genes that were unassigned. We then correlated these modules to sample traits including brain region (AMY, HYP, or VP/NAc), mating status (mated vs. virgin), and collection time points, including time as a continuous variable and each post-mating time point individually.

Module correlation with sample traits

For both species, the traits most strongly correlated with modules were sample brain region (Fig. 7a, b). In prairie voles (Supplementary Table 8), the green module was most strongly positively correlated with AMY, while the greenyellow module was strongly negatively correlated with the region. The brown, magenta, tan, cyan, pink, black, red, salmon, and turquoise modules were also significantly correlated with AMY. For HT, the turquoise module showed the strongest positive correlation, and yellow had the strongest negative correlation. Modules blue, purple, greenyellow, magenta, pink, green, red, and salmon were also significantly correlated with HT. Finally, the magenta module was most strongly positively, and red module most strongly negatively correlated with VP/NAc. The purple, yellow, greenyellow, tan, pink, green, black, and turquoise modules were also significantly correlated with this region.

Table 3 Differentially expressed genes in ventral pallidum/nucleus accumbens over time across species

Gene ID	Description	LFC	FDR
Spata16	Spermatogenesis-associated protein 16[a]	0.88879921	0.00318981
Dhx35	Probable ATP dependent RNA helicase DHX35	0.2499954	0.00366884
Mccc2	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial[a]	−0.6041633	0.00462278
Pxn[a]	Paxillin	0.43629243	0.00956998
Sike1	Suppressor of IKBKE 1	0.06660377	0.03030073
Rbm3	RNA binding protein 3	0.38113582	0.03897143
P4ha2	Prolyl 4-hydroxylase subunit alpha-2[a]	0.93405983	0.03897143
Best3	Bestrophin 2	4.3486464	0.03897143
Per1[a]	Period circadian protein homolog 1	0.21428454	0.03897143
Ciart	Circadian associated transcriptional repressor	0.31693109	0.03897143
Agt	Angiotensinogen[a]	0.61071671	0.03897143
Prss56	Transmembrane protease serine 3	1.07993336	0.04014217
Cables1	CDK5 and ABL1 enzyme substrate 2	1.09018372	0.04014217
Sgk1	Serine/threonine-protein kinase Sgk1[a]	0.34324947	0.09405603
Nfkbia	NF kappa B inhibitor alpha	0.53419379	0.09405603
Adamts14	NA	3.33669686	0.09496818
Csrnp1	Cysteine/serine rich nuclear protein 3	0.59391793	0.09496818
Lrrc6	Protein tilB homolog[a]	0.71670559	0.09496818
Smoc1	SPARC related modular calcium binding protein 2	−0.6408534	0.09995594

FDR False Discovery Rate, *LFC* $\log_2$ fold change, *NA* no annotation. Genes with positive LFC have greater total expression in prairie voles, those with negative LFC have greater total expression in meadow voles. Unannotated genes excluded. See Supplementary Table 5 for full model results. [a]Annotation obtained from UniprotKB (uniprot.org)

In meadow voles (Supplementary Table 9), the darkorchid module was most strongly positively, and tomato most strongly negatively correlated with AMY. The ghostwhite, orange, seagreen, darkred, plum, gold, and navyblue modules were also significantly correlated with AMY. Modules seagreen and darkred were most positively and negatively correlated with HT, respectively. The orange, tomato, darkorchid, plum, deeppink, and gold modules also had significant correlations with HT. Finally, the gold module had the strongest positive correlation with VP/NAc and orange module had the strongest negative correlation. Modules ghostwhite, seagreen, darkorchid, darkred, and navyblue were each also significantly correlated with VP/NAc.

In general, module correlations with mating status and collection time points were weaker than those with brain regions (Fig. 7a, b). For prairie voles (Fig. 7a, Supplementary Table 8) three modules, cyan, pink, and brown were significantly positively correlated with having mated. The blue, salmon, and midnightblue modules were negatively correlated with having mated, indicating a positive association with virgin animals. Only the grey module containing unassigned genes was significantly correlated with overall time of collection in the experiment. This relationship appears to be driven by a significant negative correlation with the 12 h collection time

point. However, two modules were significantly correlated with specific collection time points. The cyan module was positively associated with the 0.5 h time point and the blue module was significantly negatively correlated with this time.

Module correlations with mating status and postmating time points were weaker in meadow voles than in prairie voles (Fig. 7b, Supplementary Table 9). No modules were significantly correlated with mating status, and only the grey module containing unassigned genes was correlated with collection time or individual time points. It was negatively correlated with time overall and the 12 h collection time point, and positively correlated with the 0.5 h collection time point.

Module associations across species
To identify relationships between modules across species, we clustered prairie and meadow vole modules on the basis of Euclidean distance (Fig. 7c) and Pearson correlation (Fig. 7d). Clustering revealed eight pairs of modules that were closely related across species. These included the brown prairie vole and navyblue meadow vole modules, prairie magenta and meadow gold, prairie green and meadow darkorchid, prairie yellow and meadow darkred, prairie black and meadow ghostwhite, prairie red and meadow orange, prairie turquoise and
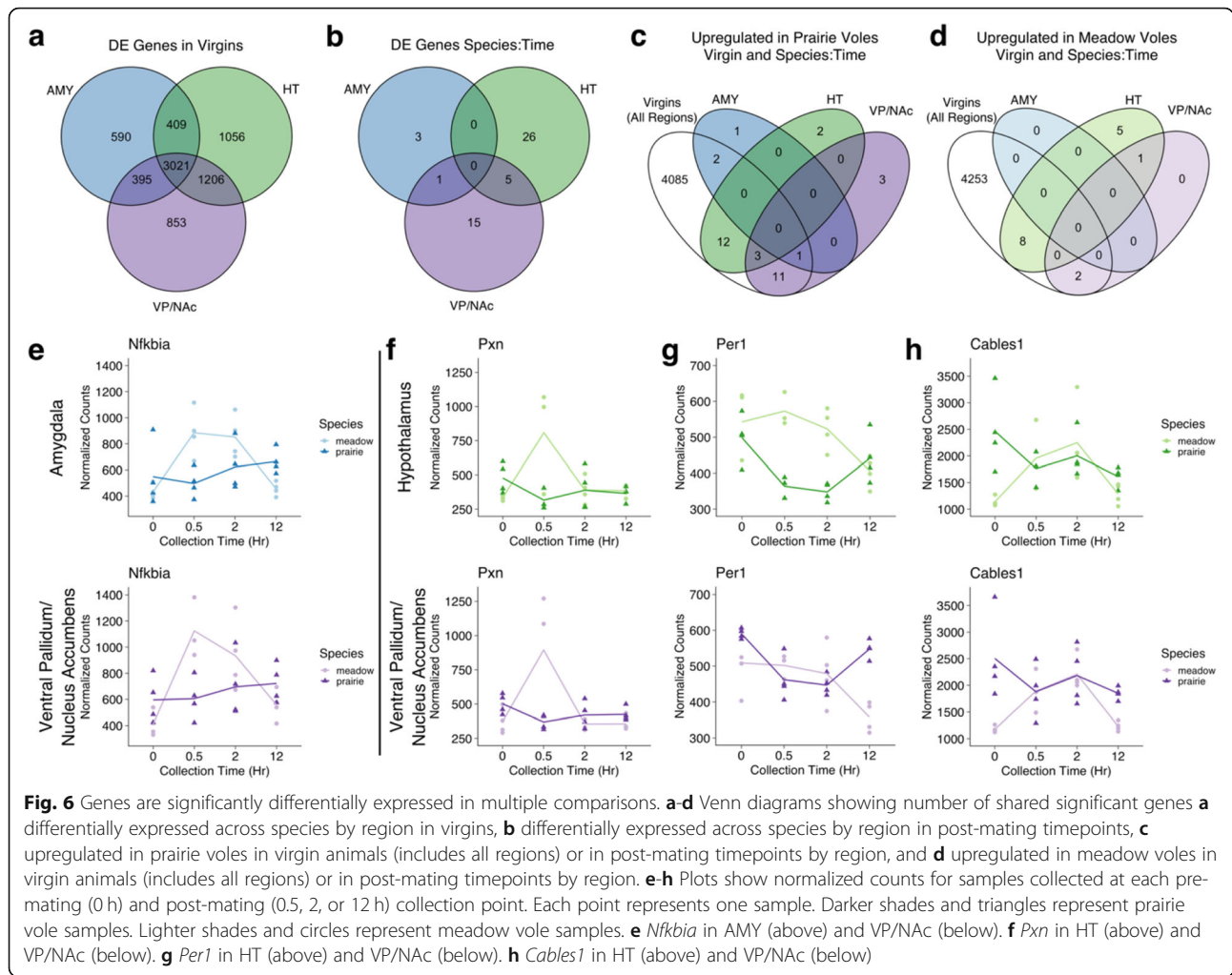
Fig. 6 Genes are significantly differentially expressed in multiple comparisons. **a-d** Venn diagrams showing number of shared significant genes **a** differentially expressed across species by region in virgins, **b** differentially expressed across species by region in post-mating timepoints, **c** upregulated in prairie voles in virgin animals (includes all regions) or in post-mating timepoints by region, and **d** upregulated in meadow voles in virgin animals (includes all regions) or in post-mating timepoints by region. **e-h** Plots show normalized counts for samples collected at each pre-mating (0 h) and post-mating (0.5, 2, or 12 h) collection point. Each point represents one sample. Darker shades and triangles represent prairie vole samples. Lighter shades and circles represent meadow vole samples. **e** *Nfkbia* in AMY (above) and VP/NAc (below). **f** *Pxn* in HT (above) and VP/NAc (below). **g** *Per1* in HT (above) and VP/NAc (below). **h** *Cables1* in HT (above) and VP/NAc (below)
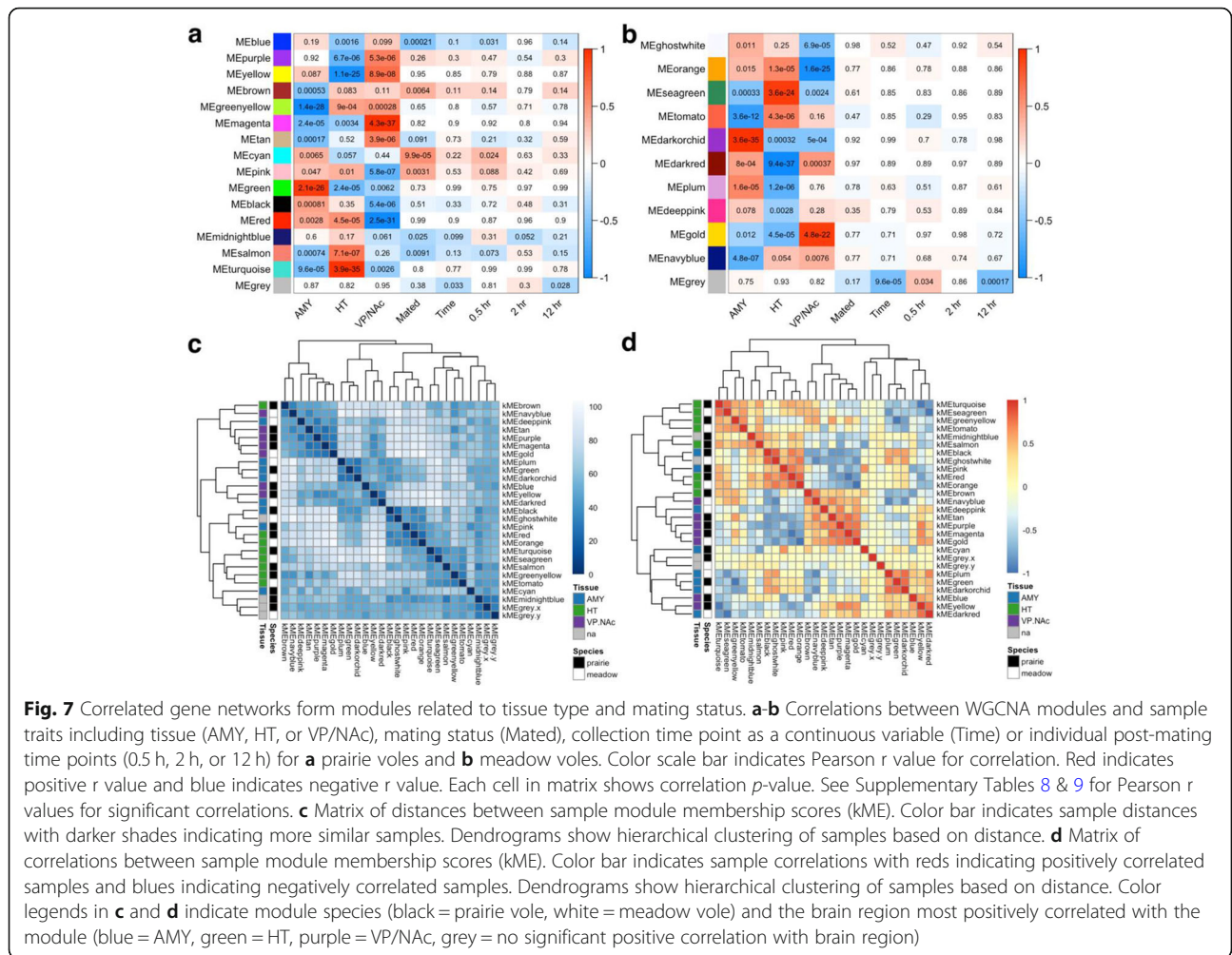
meadow seagreen, and prairie greenyellow and meadow tomato. Clustering by correlation also found one pair of prairie vole modules, midnightblue and salmon, that were most closely related.

Of these closely related modules, the majority were strongly associated with individual brain regions. The green/darkorchid modules were the most strongly positively correlated with AMY in both species, while the greenyellow/tomato modules were the most negatively correlated with AMY. Modules turquoise/seagreen were most highly positively correlated with HT and yellow/darkred were most strongly negatively correlated with it. Finally, the gold/magenta modules were most strongly associated and red/orange modules most strongly negatively associated with VP/NAc.

We used GO term analysis on pairs of modules that were strongly positively correlated with each region to identify enriched gene categories. The prairie vole green module, which was strongly correlated with AMY had 85 significantly enriched GO terms (MWU FDR < 0.1), including four in the category Molecular Function, 42

under Biological Process, and 39 in the category Cellular Component (Supplementary Fig. 1a). In meadow voles, the AMY-associated darkorchid module had 53 significantly enriched (MWU FDR < 0.1) GO terms (Supplementary Fig. 1b). These included one term in Molecular Function, *cation channel activity*, 27 terms related to Biological Process, and 25 terms in the category Cellular Component. In both species, terms related to neuronal projections, ion channels, plasma membrane and extracellular space, receptors, and synaptic transmission were enriched. The terms *behavior* and *cognition* were also enriched in both the prairie vole green module and meadow vole darkorchid module. However, the terms *associative learning* and *learning* were uniquely enriched in the prairie vole green module.

The prairie vole module turquoise and meadow vole module seagreen were each strongly associated with HT. 33 GO terms were significantly enriched in the turquoise module (MWU FDR < 0.1), including 12 in the category Molecular Function, nine in Biological Process, and 12 in Cellular Component (Supplementary Fig. 2a). In the

Fig. 7 Correlated gene networks form modules related to tissue type and mating status. **a-b** Correlations between WGCNA modules and sample traits including tissue (AMY, HT, or VP/NAc), mating status (Mated), collection time point as a continuous variable (Time) or individual post-mating time points (0.5 h, 2 h, or 12 h) for **a** prairie voles and **b** meadow voles. Color scale bar indicates Pearson r value for correlation. Red indicates positive r value and blue indicates negative r value. Each cell in matrix shows correlation *p*-value. See Supplementary Tables 8 & 9 for Pearson r values for significant correlations. **c** Matrix of distances between sample module membership scores (kME). Color bar indicates sample distances with darker shades indicating more similar samples. Dendrograms show hierarchical clustering of samples based on distance. **d** Matrix of correlations between sample module membership scores (kME). Color bar indicates sample correlations with reds indicating positively correlated samples and blues indicating negatively correlated samples. Dendrograms show hierarchical clustering of samples based on distance. Color legends in **c** and **d** indicate module species (black = prairie vole, white = meadow vole) and the brain region most positively correlated with the module (blue = AMY, green = HT, purple = VP/NAc, grey = no significant positive correlation with brain region)

meadow vole seagreen module, 11 terms were enriched (MWU FDR < 0.1), all under Cellular Component (Supplementary Fig. 2b). In both species, several terms related to cilia were enriched. Additionally, in prairie voles, terms relating to neuropeptide signaling, transcription factors, extracellular space were enriched along with the term *feeding behavior.*

No GO terms were significantly enriched in the prairie vole magenta module, which was strongly correlated with VP/NAc. However, 12 terms were significantly enriched (MWU FDR < 0.1) in the related meadow vole gold module (Supplementary Fig. 3). These included five in the category Molecular Function and six in the category Cellular Component. In addition, one term, *regulation of postsynaptic neurotransmitter receptor internalization* was enriched under the category Biological Process. Overall, most enriched terms related to the cell membrane, synaptic transmission, or enzymatic function.

Modules associated with mating
Finally, to identify gene categories associated with mating, we looked for GO term enrichment among the

modules most strongly correlated with mating status in prairie voles (Supplementary Fig. 4). We used the MWU test to identify significantly enriched terms in the cyan and pink modules, which were the most significantly positively correlated with having mated, and the blue and salmon modules, which were the two most significantly negatively correlated with having mated (thus positively associated with virgin animals).

In addition to mating status, these modules also had significant correlations with brain regions (Fig. 7a, Supplementary Table 8). The cyan module was positively correlated with AMY, pink module was positively correlated with AMY and HT, and negatively correlated with VP/NAc. Module blue was negatively correlated with HT and the salmon module was positively correlated with HT and negatively correlated with AMY.

Both the cyan and pink modules, each had significantly enriched terms (MWU FDR < 0.5) related to translation, ribosomes, and mitochondria (Supplementary Fig. 4a, b). In addition, terms associated with phosphatase binding were enriched in the cyan module and several terms

related to ubiquitination, catabolic and metabolic processes, protein localization, and the ESCRT complex were enriched in the pink module. Terms significantly enriched (MWU FDR < 0.5) in the blue included several related to peptidyl-amino acid modification and chromatin modification (Supplementary Fig. 4c). Those enriched in the salmon module (MWU FDR < 0.5) were largely associated with neuronal development and axon guidance, as well as extracellular space (Supplementary Fig. 4d).

Discussion

Pair bonding is an important social behavior of many animal species, including humans [1, 6, 46, 47]. Significant efforts have been made to understand the neural and hormonal mechanisms underlying the formation of pair bonds; however, these studies have focused on a relatively small number of genes and their products [17]. A critical next step in understanding the mechanisms of bond formation is more fully accounting for the genes underlying pair-bond formation. While other studies have made comparisons of neural gene expression in closely-related monogamous and non-monogamous species [48], to our knowledge, ours is the first to make comparisons across species during an actively-developing bond.

In this study, we used RNA-sequencing to observe changes in gene expression in three regions that have critical roles in pair-bond formation including social memory, social context, and reward. We compared two species, the bond-forming prairie vole and the non-bonding meadow vole, before and at three time points after mating. By using a comparative approach, we sought to differentiate those genes which change in expression in response to mating, but do not play a role in bond formation, from the genes that may be critical to pair-bond development. We first identified differences in the gene categories expressed across species in virgins, as well as differences in the categories of genes expressed across our focal regions. Next, we identified those genes that differed in expression post-mating across species in each focal region. Finally, we identified gene expression networks in each species and related modules of genes with correlated expression to sample features. This allowed us to identify modules strongly associated with each brain region, as well as modules which changed in expression in response to mating in prairie voles.

Pre-mating differences and the capacity to bond

Our results emphasize the importance of pre-mating differences in gene expression that confer the capacity to bond in prairie voles but not meadow voles. We found that gene expression count data were most strongly associated with sample species (Fig. 2), and the vast

majority of genes included in this study were differentially expressed across prairie and meadow vole virgins (Fig. 3). This is consistent with prior studies that have investigated the neural mechanisms of pair bonding through comparisons between monogamous and promiscuous vole species. For example, the neuropeptide AVP plays an important role in the development of partner preference and aggression toward unfamiliar conspecifics that characterize pair-bond formation in prairie voles [26]. Expression of the V1a vasopressin receptor in the VP is necessary for bond formation in male voles [49]. Meadow voles naturally have very low expression of this receptor in the VP, however, driving its expression in this brain region results in increased partner preference, a hallmark of bond formation, in meadow voles [12].

Similarly, several other studies focused on individual candidate genes or proteins found species differences in expression of OT receptor, mu-opioid receptor, corticotropin-releasing factor receptors, estrogen receptor alpha, and neuropeptide Y [50–54]. Gene expression differences important for behavior have also been found across bonding and non-bonding species in other taxa, including variation in *Avp* (gene encoding arginine-vasopressin) expression related to parental care in *Peromyscus* mice [55] and variation in OT receptor expression across pair-bonding versus solitary species of butterflyfish [56]. Together, these and the present study show that species-level differences in brain gene expression play a large role in determining the capacity to form pair bonds.

While the effects of vasopressin and its receptor are dramatic, it remains likely that many genes underlie the species differences in bonding capacity. Our current study has identified several gene categories that differed in expression across species in virgins. GO terms enriched in prairie voles were largely related to cell structure, the nucleus and extracellular space. These include *microtubule organizing center part*, *centriole*, *chromatin*, and *cell-cell junction*. Enriched in meadow voles were terms associated with the ribosome and translation, mitochondrial function, and metabolic and biosynthetic processes including *translation initiation factor*, *translation regulator*, *peptide metabolic process*, *macromolecule biosynthetic process*, *ribosomal subunit*, and *mitochondrial part*. These results indicate that gene expression differences across virgin prairie and meadow voles are largely associated with genes involved in cell structure (enriched in prairie voles) and protein synthesis (enriched in meadow voles).

Function of critical bonding-related regions

While it was not the central focus of this study, given that we found sample brain region to be the second

strongest factor associated with gene expression, we explored the types of genes that differed between our focal brain regions. Unsurprisingly, given its role in social memory [23], when comparing the AMY to HT, we found enrichment of terms related to synaptic signaling including *regulation of postsynaptic neurotransmitter receptor internalization*, *postsynapse*, and *glutamatergic synapse* (Fig. 4d). In comparison to the VP/NAc, many enriched GO terms in AMY were related to neuropeptide signaling (Fig. 4e). Neuropeptide signaling, particularly by OT, is necessary for the AMY's function in social recognition and memory. The medial AMY expresses OT receptor and receives OT innervation from the paraventricular nucleus of the HT in both mice and prairie voles [23, 29]. OT knockout mice are unable to recognize familiar conspecifics, but this deficit is rescued by site-specific injection of OT into the medial AMY [23]. Further, site-specific injection of an OT antagonist into the medial AMY also disrupted social recognition in wild type mice. Beyond OT, the AMY expresses many other neuropeptides and their receptors [57]. Genes in the neuropeptide-related categories enriched in AMY included those encoding galanin, neuropeptide FF, urocortin-3 along with receptors for neuropeptide Y, galanin, somatostatin, and neuropeptide FF along with several others. Our results suggest that there are a variety of neuropeptide systems that are likely at play in AMY function.

Several synapse-related terms were also enriched in the prairie vole green and meadow vole darkorchid WGCNA modules which were strongly correlated with the AMY (Supplementary Fig. 1). Additionally, the terms *cognition* and *behavior* were enriched in both modules, while *learning* and *associative learning* were enriched only in the prairie vole green module. Individual recognition is an essential aspect of bonding behaviors [17], one presumably mediated by these suites of genes. Taken together, these results point to the active role of the AMY in navigating social interactions, with a particular importance in prairie voles for developing social memory for a new partner during bond formation.

The gene categories we found to be enriched in the HT are also consistent with known functions of this region, a diverse structure important for sensing and producing hormonal signals [58]. In the contrast between HT and VP/NAc, we found enrichment of the terms *peptide receptor* and *neuropeptide receptor* (Fig. 4f). Similarly, these terms along with several others related to hormone signaling were enriched in the prairie vole turquoise module which was strongly correlated to HT (Supplementary Fig. 2a). Somewhat surprisingly, we did not find these terms to be enriched in the meadow vole seagreen module which was also strongly correlated with HT (Supplementary Fig. 2b), though we would expect

genes related to hormone and peptide signaling to be just as important in that species as they are in prairie voles. We interpret this as reflecting the well-known role of the hypothalamus in neuroendocrine coordination [58]. Another set of GO terms that were enriched in HT compared to both AMY and VP/NAc as well as in both the prairie vole turquoise and meadow vole seagreen modules were several terms related to cilia (Fig. 4d, f; Supplementary Fig. 2). These genes are likely related to the ependymal cells that lie along the surface of the ventral part of the third ventricle, which is flanked on either side by the HT [59]. These cilia play an important role in regulating the flow of cerebrospinal fluid through this ventricle. Other groups of terms that were enriched in the HT and related modules included several related to the endoplasmic reticulum and the extracellular region. The endoplasmic reticulum is involved in protein synthesis, including of neuropeptides [60], so enrichment of this category may be related to neuropeptide synthesis and release.

Many of the terms enriched in the VP/NAc compared to either the AMY or HT (Fig. 4e, f)—or in the meadow vole gold module (Supplementary Fig. 3), which was strongly correlated to VP/NAc—were related to the nucleus, chromosomes, and transcription/translation. In addition, there were several enriched terms related to synaptic function including *postsynapse* and terms related to dendritic spines in both the comparison with the HT and in the meadow vole gold module and the term *regulation of postsynaptic neurotransmitter receptor internalization* in the gold module. Together, these enriched terms paint a picture of a transcriptionally actively region with a potentially important role for the formation new synapses or modulation of existing ones.

Post-mating gene expression in support of bond formation

Pair-bond formation is a major life history transition that is dependent on mechanisms of social reward and memory [17]. The necessity of these systems in bonding is the reason we chose to focus on the AMY, HT, and VP/NAc as regions of interest. Prior studies have shown that the AMY is necessary for social recognition and encodes valence [22]. The HT is a diverse structure that has several functions related to social behavior [61]. Finally, the VP/NAc form a circuit important for action selection. Both dopaminergic and OT inputs to NAc are necessary for pair-bond formation in prairie voles [31, 32]. Activation of the D2 dopamine receptor in NAc reduces activity of inhibitory neurons in that region, which is proposed to then disinhibit VP neurons which guide behavioral responses to rewarding stimuli, such as mating [17, 62].

To better understand the molecular mechanisms at work in these regions in response to mating, and during pair-bond formation in prairie voles, we first focused on the genes in each region that were most significantly differentially expressed (lowest FDR) in our model testing for the effect of species:time interaction. In the AMY, this was *Npas4*, which peaked in expression for both species at the 0.5 h post-mating time point but was expressed at higher levels in prairie voles pre-mating and at later post-mating time points (Fig. 5d). *Npas4* is a transcription factor that plays an important role in the development of inhibitory synapses in response to excitatory inputs by regulating expression of activity-dependent genes [38]. In mice, social encounters drive increased *Npas4* expression in the hippocampus and knock out of *Npas4* results in alterations of social behavior, including reduced time investigating novel conspecifics [63]. Knockout animals in that study also showed deficits in learning and memory tasks.

Mating is a highly-socially salient behavior that increases neural activity in the AMY [64] so it is unsurprising that *Npas4* peaks in expression shortly after mating onset. However, the fact that it differs across species both before mating, and at later time points following mating onset suggest that it may play a particularly important role for AMY function in prairie voles compared to meadow voles. It may be that AMY activity in prairie voles is simply higher in the absence of mating or other social cues, driving *Npas4* expression more strongly. Alternatively, it may be expressed at constitutively higher levels in prairie voles in the absence of strong activation so that neurons in the region are prepared to develop inhibitory synapses in response to strong excitatory inputs. Either way, this result points to the importance of maintaining excitation/inhibition balance in the AMY for prairie voles.

In the HT, the most strongly differentially expressed gene across species over post-mating time points was *Aqp3* (Fig. 5e), a gene encoding a water/glycerol transporting channel [39] that was lowly expressed in both species until the 12 h time point, where it increased in expression in meadow voles but not prairie voles. In rats, *Aqp3* is expressed in astrocytes and neurons of several regions, and weakly expressed in ependymal cells, but no expression was found in either the supraoptic or suprachiasmatic nuclei of the hypothalamus [65]. Expression of *Aqp3* increased in response to stroke in rats [66] and is elevated in pyramidal cells of cortical tissue from human patients with edema [67]; however, the reason why this gene increases in expression in the HT of male meadow voles 12 h after beginning to mate with a novel female is not immediately clear.

The gene that was most strongly differentially expressed in the VP/NAc between species across post-

mating time points was *Spata16* (Fig. 5f). This gene showed an interesting opposing pattern of expression across species, starting at high levels and reducing quickly following mating in prairie voles and beginning at low levels of expression and peaking shortly after mating in meadow voles. *Spata16* is a gene important for spermatogenesis [40], and to our knowledge it has not so far been studied for its function or localization in the brain.

In addition to the genes most strongly differentially expressed in our model testing the effect of the species: time interaction, we found four genes that significantly differed across species in multiple brain regions (Fig. 6e-h). Among these were *Nfkbia*, which encodes an inhibitor of the transcription factor nuclear factor kappa B [68]; *Pxn*, which acts as an adaptor protein that interacts with the cytoskeleton and signaling molecules that regulate cell adhesion to the extracellular matrix [69]; the circadian clock gene *Per1* [43]; and *Cables1*, which encodes a protein that promotes neurite outgrowth in neurons and plays a role in regulating cell death [70].

Interestingly, though the absolute level of expression varied across regions within a species, the pattern of expression of these genes was largely similar across our focal regions. This suggests that whatever function they serve in response to mating is consistent across the brain regions. Further, these genes tended to increase after mating in meadow voles, then return to baseline, while maintaining relatively more consistent expression across all time points in prairie voles. This suggests that they are responsive to mating in the promiscuous species but are not necessarily important for bonding.

Alternatively, it is possible that the changes in expression over time we observe in some or all of these genes are related to time itself, rather than time relative to mating. This is particularly true for circadian-related genes *Per1*, which has opposing patterns of expression in prairie and meadow voles (Fig. 6g), and *Ciart*, though this caveat applies more broadly, as nearly half of protein-coding mammalian genes show circadian variation in their expression [71]. If this is the case, their variation may reflect differences in circadian cycles of gene expression across species. Prairie voles show ultradian patterns in cardiovascular measures [72] and in activity in the field [73]. This is also true of common voles (*Microtus arvalis*) [74], though daily rhythms of meadow voles and potential differences between prairie and meadow voles remain to be investigated. Given that in this study we collected prairie and meadow voles at similar time points, our results are unlikely to be an artifact of timing of tissue collection. Moreover, none of our gene ontology analyses identified enrichment of circadian-related terms. While some genes (e.g. *Per1*, *Ciart*) may be influenced by species differences in

circadian rhythms, on balance, species differences in circadian function do not seem to be major predictors of gene expression in our data. Future investigations can more directly test for circadian differences between prairie and meadow voles and their effects on brain gene expression and bonding behavior.

In addition to these genes that were differentially expressed across species, we also examined modules that were significantly correlated with mating status to identify enriched gene categories associated with these modules. The cyan and pink modules were each positively correlated with having mated, and enriched terms in these modules included many related to translation, ribosomes, and mitochondria. Terms related to catabolic and metabolic processes were also enriched in the pink module. The terms enriched in these modules are likely associated with transcriptional and energetic demands associated with the response to a novel social encounter and initial mating activity. For example, an earlier study has found that 30 min of mating elicits activity-dependent expression of the immediate early genes Fos and Egr-1 in dopaminergic neurons of the AMY in male prairie voles [75]. Consistent with this interpretation, the post-mating time point these modules were most strongly related to was 0.5 h (Fig. 7a). Cyan was significantly positively correlated with this point (r = 0.33, *p* = 0.024) while pink trended toward a significant positive correlation (r = 0.25, *p* = 0.088).

Similar to the modules positively associated with mating, the modules negatively correlated with mating (and so positively correlated with virgin status) were most negatively associated with the 0.5 h post-mating onset time point (Fig. 7a). The blue module was significantly negatively correlated with this time (r = − 0.32, *p* = 0.031) and salmon module trended toward a negative correlation (r = − 0.26, *p* = 0.073). Many of the terms enriched in the blue module were related to chromosomes and chromatin modifications, suggesting that activity of these genes is reduced in early mating. In the salmon module, most enriched terms were associated with neural development and axon guidance. As the salmon module was positively correlated with HT, this result suggests that during early mating there is a reduction in the formation or growth of axons from neurons in this region. Together, these results suggest that during very early mating in prairie voles, expression of genes involved in synaptic plasticity and activity of epigenetic modifications are actually reduced compared to the virgin state, perhaps due to the energetic and protein synthesis demands occurring at this time. Though it remains possible that the products of these genes are still present and active during early mating.

Limitations and future studies
One limitation of the current study is the fact that annotations for the genes encoding OT, OT receptor, and the

V1a AVP receptor are absent from the currently available annotated prairie vole genome. While their absence is unfortunate, as these genes would have served as useful positive controls, given the fact that they have been a major focus of studies of mating and pair bonding in voles and other taxa [5, 9, 17, 56, 76–82], it is unlikely that the current study would reveal any novel function for these genes. So, although their inclusion in the annotated genome would have been useful, their absence does not impede the key goals of this study.

While this study identified several gene categories that differ between virgin prairie and meadow voles and across regions in both species, as well as intriguing candidate genes that vary over time after mating across species, we are not able to tie these genes and GO categories to specific behaviors. Future studies that include more detailed behavioral observation will be useful for correlating expression of genes of interest to specific mating-related or other prosocial behaviors that voles engage in during bond formation. Additional manipulative studies will be useful in determining which gene expression differences between prairie and meadow voles.

Additionally, our GO term enrichment analyses comparing across species often found significant enrichment of terms related to transcription and translation as well as metabolic activity. While this may be indicative of real differences in these processes between prairie and meadow voles, it is important to consider that this may instead be an artifact of the sequencing and analysis process. That is, due to the enrichment of genes in other categories (e.g. cell structure and neuronal development) in prairie voles, the relative abundance of "housekeeping" genes involved in protein synthesis and metabolism appears to be reduced. Nevertheless, our results still emphasize the importance of brain transcriptional regulation in supporting prairie vole bond formation.

Finally, this study investigated gene expression differences in males only. A benefit of this approach is reducing the number of animals needed to obtain a sample size with appropriate statistical power, while limiting the number of samples and factors needed to analyze a large dataset. Further, most of the neural mechanisms of prairie vole pair bonding investigated so far are shared across the sexes [17]; however, sex differences in the neurohormonal control of pair bonding do exist, such as dimorphic effects of stress on bond formation [83]. Future studies using RNA-sequencing or a candidate gene approach should investigate whether our results also hold true for female voles, or if there are important sex differences in brain gene expression facilitating bond formation.

Conclusions
Our results emphasize the importance of pre-mating differences in gene expression that confer the ability to pair

bond in prairie voles but not in non-bonding species such as meadow voles. In addition, they support the hypothesis that pair-bond formation relies on transcriptional regulation and alterations of neuronal structure in regions including the amygdala, hypothalamus, ventral pallidum, and nucleus accumbens that links neural encoding of partner cues with the reward system, resulting in reinforcement of the partner that leads to selective affiliation (Fig. 8). Finally, we identify several intriguing candidate genes that may play important roles in bond formation following co-habitation and mating. Together, our results should help broaden the scope of research on the neural and molecular mechanisms of pair-bond formation by providing novel candidates to investigate in future manipulative studies.
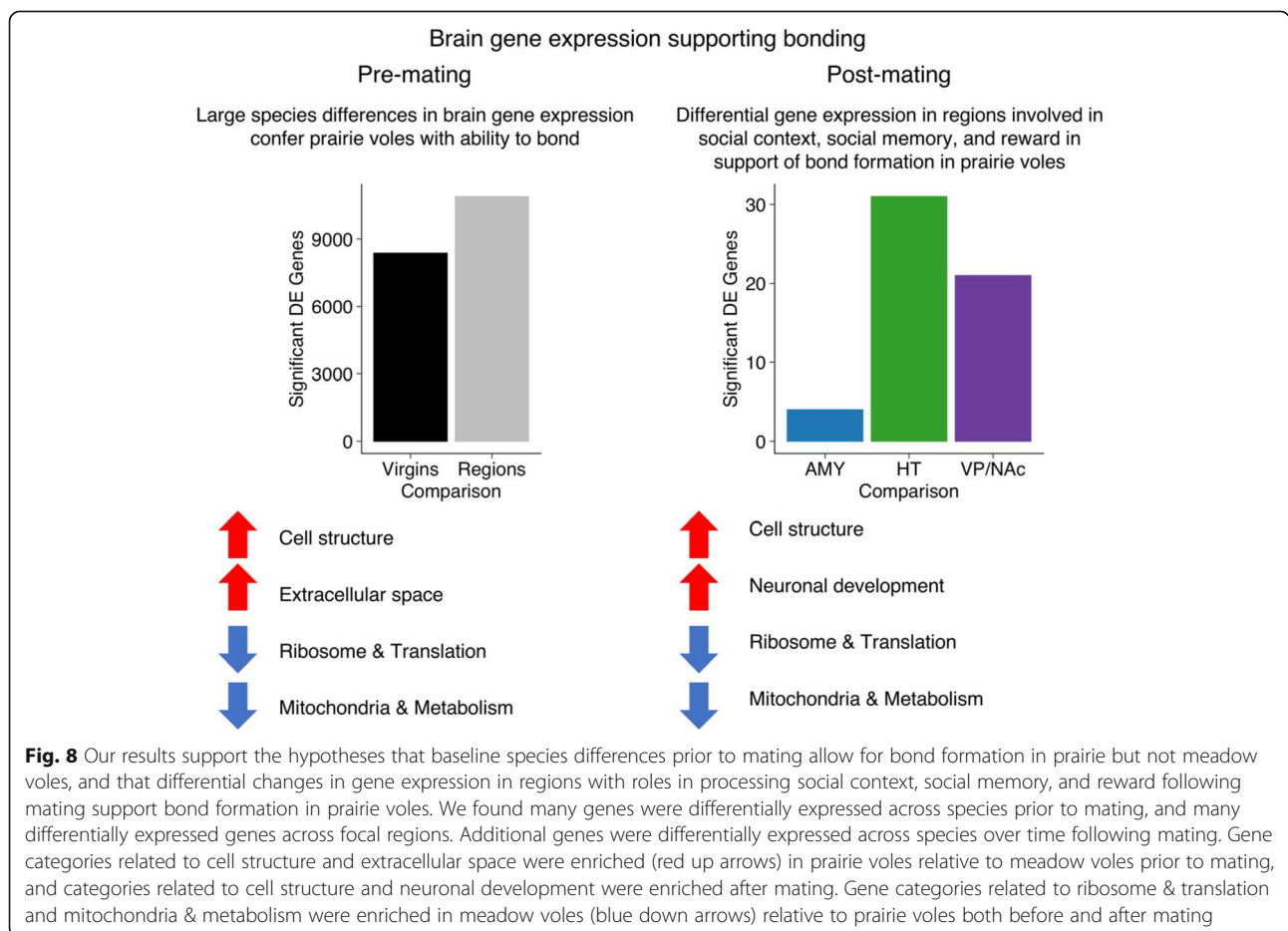
Methods

Animals

All animals used in this study were taken from laboratory breeding colonies at Emory University, derived from field-captured voles collected in Illinois. Prairie voles and meadow voles were housed separately in same-sex groups of two to three voles per cage from postnatal day

21. Housing consisted of a ventilated 36 cm × 18 cm × 19 cm plexiglass cage filled with Bed-o'-Cobs laboratory animal bedding (The Andersons Inc., Maumee, Ohio) under a 14/10 h light/dark cycle (lights on 7:00 AM-9:00 PM) at 22 °C with access to food (rabbit diet; LabDiet, St. Louis, Missouri) and water ad libitum. All experiments were performed in accordance with relevant guidelines and regulations and were approved by the Emory University Institutional Animal Care and Use Committee.

Tissue collection

All animals were 60–90 days old and sexually naïve at the time of the experiment and tissues were collected from males only. Voles were euthanized by isoflurane inhalation overdose followed immediately by rapid decapitation as recommended by the American Veterinary Medical Association and approved by the Emory Institutional Animal Care and Use Committee. Tissues for the virgin group (time point 0) were collected from sexually naïve males without exposure to females ($n = 4$ per species). For tissues from mated animals, single male voles were paired with a single unrelated female of the same

Fig. 8 Our results support the hypotheses that baseline species differences prior to mating allow for bond formation in prairie but not meadow voles, and that differential changes in gene expression in regions with roles in processing social context, social memory, and reward following mating support bond formation in prairie voles. We found many genes were differentially expressed across species prior to mating, and many differentially expressed genes across focal regions. Additional genes were differentially expressed across species over time following mating. Gene categories related to cell structure and extracellular space were enriched (red up arrows) in prairie voles relative to meadow voles prior to mating, and categories related to cell structure and neuronal development were enriched after mating. Gene categories related to ribosome & translation and mitochondria & metabolism were enriched in meadow voles (blue down arrows) relative to prairie voles both before and after mating

species primed with estradiol benzoate (Sigma Aldrich, St. Louis, MO, USA, BP958) for two days prior to pairing to ensure sexual receptivity. Pairs were observed and the time of first intromission was recorded. Males were collected at 30 min (n = 4 per species), 2 h (n = 4 per species), and 12 h (n = 4 per species) following the first intromission. To reduce the possible influence of circadian effects on gene expression, prairie and meadow voles were collected at similar times for each post-mating time point. Virgin animals were collected at the same times as postmating animals. Following collection, animals were euthanized using isoflurane, and whole brains were harvested and dissected on a block of dry ice. The AMY, HT, and VP/NAc were collected from each brain and stored in RNAlater (Applied Biosystems AM7020).

RNA extraction and sequencing
Total RNA was extracted from each region from all individuals using TRIzol (Sigma-Aldrich) following manufacturer's protocol. An additional DNAse treatment (Purelink) was included to eliminate any contaminating DNA. RNA quantity and quality were assessed using a BioAnalyzer (Agilent) and quantified by spectrophotometry. cDNA was prepared using Superscript reverse transcriptase II (Invitrogen) and purified with QIAquick PCR purification kit. After checking cDNA quantity and quality in the BioAnalyzer, libraries were prepared using Illumina's TruSeq Sample Prep Kit with starting amount of 1.25 μg cDNA. Libraries were sequenced as 50 bp single-end reads at the Emory University genomics core facility using Illumina HiScan, and HiSeq 2000 machines. Read files were trimmed and cleaned using fastq tools.

Differential gene expression
Reads were aligned to an annotated prairie vole genome (MicOch1.0, INSDC Assembly GCA_000317375.1, Full genebuild annotation by Ensembl released February 2017) using STAR aligner (v2.7.2b) [84]. The R package arrayQualityMetrics [85] was used to assess sample quality and identify outliers. One sample was removed for having low read counts and two removed as outliers resulting in 47 prairie vole (*n* = 16 AMY samples, *n* = 15 HT samples, n = 16 VP/NAc samples) and 46 meadow vole (n = 16 AMY samples, n = 15 HT samples, n = 15 VP/NAc samples) samples. Genes with less than one count per sample on average were removed prior to analyses.

The dataset was characterized and differential gene expression was analyzed with the R package DESeq2 [34] following the workflow provided by the package authors [86]. For analyses of the full dataset, a DESeq dataset was created with the model ~species + region + time. Poisson distance was calculated using the PoiClaClu package [33]. Variance stabilizing transformation and

principal component analysis (PCA) were conducted using DESeq2 functions. A likelihood ratio test with the reduced model ~species + time was used to test for genes that were differentially expressed across regions. To identify genes that were differentially expressed across species before mating, the full model ~species + region was applied to count data from both species from the virgin time point only, and the likelihood ratio test was used with the reduced model ~region. To test for differentially expressed genes across species at the virgin time point within each region, the data were subset by region and a likelihood ratio test was used with the full model ~species and reduced model ~ 1. To test for species differences in gene expression within a region following mating, the data were subset by region and a likelihood ratio test was used with the full model ~species + time + species:time and reduced model ~species + time. This model tests whether species differ in gene expression at any time point after time 0 (virgin animals) [86]. Additional annotation information for genes of interest (Tables 1, 2 and 3) was obtained from the UniProtKB (uniprot.org) manually annotated, reviewed (Swiss-prot) database. Manually annotated gene IDs and descriptions were derived from mouse (*Mus musculus*) annotations.

Gene correlation networks
We used the R package WGCNA [45] to identify networks of genes with correlated expression. Count data were divided by species, and separate networks were created for prairie voles and meadow voles. Lowly expressed genes (mean of less than 10 counts per sample) were removed and count data were transformed using DESeq2 variance-stabilizing transformation [34]. We identified signed modules (i.e. modules consist of genes with strong positive correlation to one another) of correlated genes within networks for each species and found correlations between module eigengenes and sample traits including brain region, mating status (coded 1 for any post-mating time point, 0 for virgin), time of collection (0 [virgin],0.5, 2, or 12 h), and individual postmating collection time points. Following this, we compared species networks by calculating module membership score (kME) [45] for each gene in each module for each species, then merged the kME datasets. We used R package pheatmap [87] to create heatmaps based on module distances and Pearson correlations and used hierarchical clustering by the function hclust with method "average" to visualize relationships of modules across species.

Gene ontology term enrichment
To test for enrichment of GO terms among genes associated with our differential expression comparisons or

WGCNA modules of interest, we used the Mann-Whitney U (MWU) test [36] implemented by the GO_MWU R script [37] (Available at https://github.com/z0on/GO_MWU). Rather than test for enrichment among genes determined to be significant by an FDR cutoff, the MWU test determines if GO categories are enriched by up- or down-regulated genes, or within modules, compared to all other categories using continuous measures. For tests of enrichment among differential expression comparisons, we used signed $-\log_{10}$ of the p-value as the measure of comparison to identify terms associated with genes most strongly differentially expressed across species or brain regions [37]. The $-\log_{10}$ of the p-value from the model was calculated for each gene, then "signed" by multiplying by -1 if the $\log_2$ fold change (LFC) for the contrast of interest was negative [37]. For tests of enrichment in WGCNA modules of interest, we used gene kME values for genes included in the module, and all genes not included in the module were assigned a value of zero and adjusted p-value was determined by 100 permutations where significance measures are randomly shuffled among genes [37].

Data analysis and visualization

All analyses were conducted using R (v4.0.0). Details of packages and scripts used for differential expression, gene correlation network, and GO term enrichment analyses are described in the relevant sections above. Plots were made using DESeq2 [34] and WGCNA [45] as well as the packages ggplot2 [88], pheatmap [87], and RColorBrewer [89].

Abbreviations
AMY: Amygdala; AVP: Arginine-vasopressin; FDR: False discovery rate; HT: Hypothalamus; LFC: Log (base 2) fold change; ME: Module eigengene; MWU: Mann-Whitney U test; OT: Oxytocin; VP/NAc: Ventral pallidum/nucleus accumbens

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07720-0.

Additional file 1: Supplementary Table 1. Results of DESeq2 comparison of virgin samples across species. Positive log2FoldChange values indicate higher expression in prairie voles and negative log2FoldChange values indicate higher expression in meadow voles.

Additional file 2: Supplementary Table 2. Results of DESeq2 comparison of gene expression across brain regions.

Additional file 3: Supplementary Table 3. Results of DESeq2 comparison of species x time interaction in amygdala. Positive log2FoldChange values indicate higher overall expression in prairie voles and negative log2FoldChange values indicate higher overall expression in meadow voles.

Additional file 4: Supplementary Table 4. Results of DESeq2 comparison of species x time interaction in hypothalamus. Positive log2FoldChange values indicate higher overall expression in prairie voles and negative log2FoldChange values indicate higher overall expression in meadow voles.

Additional file 5: Supplementary Table 5. Results of DESeq2 comparison of species x time interaction in ventral pallidum/nucleus accumbens. Positive log2FoldChange values indicate higher overall expression in prairie voles and negative log2FoldChange values indicate higher overall expression in meadow voles.

Additional file 6: Supplementary Table 6. WGCNA module assignments for all genes in prairie voles, along with module membership scores (MM) and module membership p-values (p.MM) for each module.

Additional file 7: Supplementary Table 7. WGCNA module assignments for all genes in meadow voles, along with module membership scores (MM) and module membership p-values (p.MM) for each module.

Additional file 8: Supplementary Table 8. Statistics for prairie vole WGCNA modules with significant correlations.

Additional file 9: Supplementary Table 9. Statistics for meadow vole WGCNA modules with significant correlations.

Additional file 10: Supplementary Fig. 1. Expression pattern of significant genes is consistent across regions. Plots show counts for samples collected at each pre-mating (0 h) and post-mating (0.5, 2, or 12 h) collection point. Each point represents one sample. Darker shades and triangles represent prairie vole samples. Lighter shades and circles represent meadow vole samples. a. *Nfkbia* in AMY (above) and VP/NAc (below). b. *Pxn* in HT (above) and VP/NAc (below). c. *Per1* in HT (above) and VP/NAc (below). d. *Cables1* in HT (above) and VP/NAc (below). **Supplementary Fig. 2.** Gene ontology analysis for amygdala gene-expression modules. Enriched GO terms in a. prairie vole green module and b. meadow vole darkorchid module. Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of genes from the category that are included in the module of interest. FDR determined by 100 permutations where significance measures are randomly shuffled among genes. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category. **Supplementary Fig. 3.** Gene ontology analysis for hypothalamic gene-expression modules. Enriched GO terms in a. prairie vole tuquoise module and b. meadow vole seagreen module. Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of genes from the category that are included in the module of interest. FDR determined by 100 permutations where significance measures are randomly shuffled among genes. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category. **Supplementary Fig. 4.** Gene ontology analysis for ventral pallidum/nucleus accumbens gene-expression modules. Enriched GO terms in meadow vole gold module. (There were no significant enrichments for the prairie vole magenta module.) Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of genes from the category that are included in the module of interest. FDR determined by 100 permutations where significance measures are randomly shuffled among genes. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category. **Supplementary Fig. 5.** Effects of mating on gene ontology categories. Enriched GO terms in modules most strongly positively (a-b) and negatively (c-d) correlated with mating status in prairie voles. a. cyan module, b. pink module c. blue module, d. salmon module. Hierarchical clustering tree shows relationship between GO categories based on shared genes. Branches with length of zero are subsets of one another. Fractions preceding GO terms indicate proportion of genes from the category that are included in the module of interest. FDR determined by 100 permutations where significance measures are randomly shuffled among genes. Bold text indicates adjusted $p < 0.01$, plain text indicates adjusted $p < 0.05$, and italicized text indicates adjusted $p < 0.1$ for term. n.s. indicates no significant terms in the category.

Declarations

Ethics approval and consent to participate
All experiments were approved by the Institutional Animal Care and Use Committee of Emory University. All procedures were carried out in accordance with ARRIVE guidelines.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests

Author details
[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA. [2]Present Address: Department of Biology, Duke University, Durham, NC 27708, USA. [3]Center for Translational Social Neuroscience, Department of Psychiatry and Behavioral Sciences, Yerkes National Primate Research Center, Emory University, Atlanta, GA 30329, USA. [4]Centers for Disease Control and Prevention, Atlanta, GA 30333, USA. [5]National Institutes of Health Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, MD, USA.

References
1. Fletcher GJO, Simpson JA, Campbell L, Overall NC. Pair-bonding, romantic love, and evolution: the curious case of *Homo sapiens*. Perspect Psychol Sci. 2015;10(1):20–36. https://doi.org/10.1177/1745691614561683.
2. Lukas D, Clutton-Brock TH. The evolution of social monogamy in mammals. Science. 2013;341(6145):526–30. https://doi.org/10.1126/science.1238677.
3. Insel TR, Preston S, Winslow JT. Mating in the monogamous male: behavioral consequences. Physiol Behav. 1995;57(4):615–27. https://doi.org/10.1016/0031-9384(94)00362-9.
4. Williams JR, Catania KC, Carter CS. Development of partner preferences in female prairie voles (*Microtus ochrogaster*): the role of social and sexual experience. Horm Behav. 1992;26(3):339–49. https://doi.org/10.1016/0018-506X(92)90004-F.
5. McGraw LA, Young LJ. The prairie vole: an emerging model organism for understanding the social brain. Trends Neurosci. 2010;33(2):103–9. https://doi.org/10.1016/j.tins.2009.11.006.
6. Carter CS, Getz LL. Monogamy and the prairie vole. Sci Am. 1993;268(6):100–6. https://doi.org/10.1038/scientificamerican0693-100.
7. Getz LL, Carter CS, Gavish L. The mating system of the prairie vole, *Microtus ochrogaster*: field and laboratory evidence for pair-bonding. Behav Ecol Sociobiol. 1981;8(3):189–94. https://doi.org/10.1007/BF00299829.
8. Ophir AG, Wolff JO, Phelps SM. Variation in neural V1aR predicts sexual fidelity and space use among male prairie voles in semi-natural settings. Proc Natl Acad Sci U S A. 2008 Jan 29;105(4):1249–54. https://doi.org/10.1073/pnas.0709116105.
9. Okhovat M, Berrio A, Wallace G, Ophir AG, Phelps SM. Sexual fidelity trade-offs promote regulatory variation in the prairie vole brain. Science. 2015;350(6266):1371–4. https://doi.org/10.1126/science.aac5791.
10. Pierce JD, Ferguson B, Dewsbury DA. Conspecific preferences in prairie voles, *Microtus ochrogaster*, and meadow voles, *M pennsylvanicus*. Bull Psychon Soc. 1989;27(3):267–70. https://doi.org/10.3758/BF03334603.
11. Gruder-Adams S, Getz LL. Comparison of the mating system and paternal behavior in *Microtus ochrogaster* and *M. pennsylvanicus*. J Mammal. 1985;66(1):165–7. https://doi.org/10.2307/1380976.
12. Lim MM, Wang Z, Olazábal DE, Ren X, Terwilliger EF, Young LJ. Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. Nature. 2004;429(6993):754–7. https://doi.org/10.1038/nature02539.
13. Wang Z, Young LJ, Liu Y, Insel TR. Species differences in vasopressin receptor binding are evident early in development: comparative anatomic studies in prairie and montane voles. J Comp Neurol. 1997;378(4):535–46. https://doi.org/10.1002/(SICI)1096-9861(19970224)378:4<535::AID-CNE8>3.0.CO;2-3.
14. Young LJ, Winslow JT, Nilsen R, Insel TR. Species differences in V1a receptor gene expression in monogamous and nonmonogamous voles: behavioral consequences. Behav Neurosci. 1997;111(3):599–605. https://doi.org/10.1037/0735-7044.111.3.599.
15. Madison DM. Space use and social structure in meadow voles, *Microtus pennsylvanicus*. Behav Ecol Sociobiol. 1980;7(1):65–71. https://doi.org/10.1007/BF00302520.
16. Boonstra R, Xia X, Pavone L. Mating system of the meadow vole, *Microtus pennsylvanicus*. Behav Ecol. 1993;4(1):83–9. https://doi.org/10.1093/beheco/4.1.83.
17. Walum H, Young LJ. The neural mechanisms and circuitry of the pair bond. Nat Rev Neurosci. 2018;19(11):643–54. https://doi.org/10.1038/s41583-018-0072-6.
18. Ophir AG. Navigating monogamy: Nonapeptide sensitivity in a memory neural circuit may shape social behavior and mating decisions. Front Neurosci. 2017;11(JUL):397.
19. Jaarola M, Martínková N, Gündüz I, Brunhoff C, Zima J, Nadachowski A, et al. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Mol Phylogenet Evol. 2004;33(3):647–63. https://doi.org/10.1016/j.ympev.2004.07.015.
20. Pro-Sistiaga P, Mohedano-Moriano A, Ubeda-Bañon I, Arroyo-Jimenez MDM, Marcos P, Artacho-Pérula E, et al. Convergence of olfactory and vomeronasal projections in the rat basal telencephalon. J Comp Neurol. 2007;504(4):346–62. https://doi.org/10.1002/cne.21455.
21. Bielsky IF, Young LJ. Oxytocin, vasopressin, and social recognition in mammals. Peptides. 2004;25(9):1565–74. https://doi.org/10.1016/j.peptides.2004.05.019.
22. Beyeler A, Chang CJ, Silvestre M, Lévêque C, Namburi P, Wildes CP, et al. Organization of valence-encoding and projection-defined neurons in the basolateral amygdala. Cell Rep. 2018;22(4):905–18. https://doi.org/10.1016/j.celrep.2017.12.097.
23. Ferguson JN, Aldag JM, Insel TR, Young LJ. Oxytocin in the medial amygdala is essential for social recognition in the mouse. J Neurosci. 2001;21(20):8278–85. https://doi.org/10.1523/JNEUROSCI.21-20-08278.2001.
24. Ferguson JN, Young LJ, Insel TR. The neuroendocrine basis of social recognition. Front Neuroendocrinol. 2002;23(2):200–24. https://doi.org/10.1006/frne.2002.0229.
25. Gobrogge KL, Liu Y, Young LJ, Wanga Z. Anterior hypothalamic vasopressin regulates pair-bonding and drug-induced aggression in a monogamous rodent. Proc Natl Acad Sci U S A. 2009;106(45):19144–9. https://doi.org/10.1073/pnas.0908620106.
26. Winslow JT, Hastings N, Carter CS, Harbaugh CR, Insel TR. A role for central vasopressin in pair bonding in monogamous prairie voles. Nature. 1993;365(6446):545–8. https://doi.org/10.1038/365545a0.
27. Ferguson JN, Young LJ, Hearn EF, Matzuk MM, Insel TR, Winslow JT. Social amnesia in mice lacking the oxytocin gene. Nat Genet. 2000;25(3):284–8. https://doi.org/10.1038/77040.
28. Johnson ZV, Walum H, Jamal YA, Xiao Y, Keebaugh AC, Inoue K, et al. Central oxytocin receptors mediate mating-induced partner preferences and enhance correlated activation across forebrain nuclei in male prairie voles. Horm Behav. 2016;79:8–17. https://doi.org/10.1016/j.yhbeh.2015.11.011.

29. Ross HE, Cole CD, Smith Y, Neumann ID, Landgraf R, Murphy AZ, et al. Characterization of the oxytocin system regulating affiliative behavior in female prairie voles. Neuroscience. 2009;162(4):892–903. https://doi.org/10.1016/j.neuroscience.2009.05.055.

30. Love TM. Oxytocin, motivation and the role of dopamine. Pharmacol Biochem Behav. 2014;119:49–60. https://doi.org/10.1016/j.pbb.2013.06.011.

31. Aragona BJ, Liu Y, Curtis JT, Stephan FK, Wang Z. A critical role for nucleus accumbens dopamine in partner-preference formation in male prairie voles. J Neurosci. 2003;23(8):3483–90. https://doi.org/10.1523/JNEUROSCI.23-08-03483.2003.

32. Liu Y, Wang ZX. Nucleus accumbens oxytocin and dopamine interact to regulate pair bond formation in female prairie voles. Neuroscience. 2003; 121(3):537–44. https://doi.org/10.1016/S0306-4522(03)00555-4.

33. Witten DM. Classification and clustering of sequencing data using a poisson model. Ann Appl Stat. 2011;5(4):2493–518.

34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.

35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1): 289–300.

36. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 2005;3(6):0976–85.

37. Wright RM, Aglyamova GV, Meyer E, Matz MV. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. BMC Genomics. 2015;16(1):1–12.

38. Lin Y, Bloodgood BL, Hauser JL, Lapan AD, Koon AC, Kim TK, et al. Activity-dependent regulation of inhibitory synapse development by Npas4. Nature. 2008;455(7217):1198–204. https://doi.org/10.1038/nature07319.

39. Verkman AS, Mitra AK. Structure and function of aquaporin water channels. Am J Physiol Physiol. 2000;278(1):F13–28. https://doi.org/10.1152/ajprenal.2000.278.1.F13.

40. De Braekeleer M, Nguyen MH, Morel F, Perrin A. Genetic aspects of monomorphic teratozoospermia: a review. J Assist Reprod Genet. 2015;32(4): 615–23. https://doi.org/10.1007/s10815-015-0433-2.

41. Schaller MD. Paxillin: A focal adhesion-associated adaptor protein. Oncogene. 2001;20(44 REV. ISS. 5):6459–72.

42. López-Colomé AM, Lee-Rivera I, Benavides-Hidalgo R, López E. Paxillin: a crossroad in pathological cell migration. J Hematol Oncol. 2017;10(1):1–15.

43. Masubuchi S, Kataoka N, Sassone-Corsi P, Okamura H. Mouse period1 (mPER1) acts as a circadian adaptor to entrain the oscillator to environmental light/dark cycles by regulating mPER2 protein. J Neurosci. 2005;25(19):4719–24. https://doi.org/10.1523/JNEUROSCI.4761-04.2005.

44. Zukerberg LR, Patrick GN, Nikolic M, Humbert S, Wu C-L, Lanier LM, et al. Cables links Cdk5 and c-Abl and facilitates Cdk5 tyrosine phosphorylation, kinase upregulation, and neurite outgrowth. Neuron. 2000;26(3):633–46. https://doi.org/10.1016/S0896-6273(00)81200-3.

45. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559. https://doi.org/10.1186/1471-2105-9-559.

46. Nowicki JP, O'Connell LA, Cowman PF, Walker SPW, Coker DJ, Pratchett MS. Variation in social systems within Chaetodon butterflyfishes, with special reference to pair bonding. PLoS One. 2018;13(4):1–24.

47. Brown JL, Morales V, Summers K. A key ecological trait drove the evolution of biparental care and monogamy in an amphibian. Am Nat. 2010;175(4): 436–46. https://doi.org/10.1086/650727.

48. Young RL, Ferkin MH, Ockendon-Powell NF, Orr VN, Phelps SM, Pogany A, et al. Conserved transcriptomic profiles underpin monogamy across vertebrates. Proc Natl Acad Sci. 2019;116(4):1331–6. https://doi.org/10.1073/pnas.1813775116.

49. Lim MM, Young LJ. Vasopressin-dependent neural circuits underlying pair bond formation in the monogamous prairie vole. Neuroscience. 2004;125(1): 35–45. https://doi.org/10.1016/j.neuroscience.2003.12.008.

50. Inoue K, Burkett JP, Young LJ. Neuroanatomical distribution of μ-opioid receptor mRNA and binding in monogamous prairie voles (*Microtus ochrogaster*) and non-monogamous meadow voles (*Microtus pennsylvanicus*). Neuroscience. 2013;244:122–33. https://doi.org/10.1016/j.neuroscience.2013.03.035.

51. Insel TR, Shapiro LE. Oxytocin receptor distribution reflects social organization in monogamous and polygamous voles. Proc Natl Acad Sci. 1992;89(13):5981–5. https://doi.org/10.1073/pnas.89.13.5981.

52. Lim MM, Nair HP, Young LJ. Species and sex differences in brain distribution of corticotropin-releasing factor receptor subtypes 1 and 2 in monogamous and promiscuous vole species. J Comp Neurol. 2005;487(1):75–92. https://doi.org/10.1002/cne.20532.

53. Cushing BS, Wynne-Edwards KE. Estrogen receptor-α distribution in male rodents is associated with social organization. J Comp Neurol. 2006;494(4): 595–605. https://doi.org/10.1002/cne.20826.

54. Hostetler CM, Hitchcock LN, Anacker AMJ, Young LJ, Ryabinin AE. Comparative distribution of central neuropeptide y (NPY) in the prairie (*Microtus ochrogaster*) and meadow (*M. pennsylvanicus*) vole. Peptides. 2013; 40:22–9. https://doi.org/10.1016/j.peptides.2012.12.008.

55. Bendesky A, Kwon Y, Lassance J, Lewarch CL, Yao S, Peterson BK. The genetic basis of parental care evolution in monogamous mice. Nature. 2017;544(7651):434–9. https://doi.org/10.1038/nature22074.

56. Nowicki JP, Pratchett MS, Walker SPW, Coker DJ, O'Connell LA. Gene expression correlates of social evolution in coral reef butterflyfishes. Proceedings Biol Sci. 2020;287(1929):20200239.

57. Neugebauer V, Mazzitelli M, Cragg B, Ji G, Navratilova E, Porreca F. Amygdala, neuropeptides, and chronic pain-related affective behaviors. Neuropharmacology. 2020;170(March):108052. https://doi.org/10.1016/j.neuropharm.2020.108052.

58. Saper CB, Lowell BB. The hypothalamus. Curr Biol. 2014;24(23):R1112.

59. Eichele G, Bodenschatz E, Ditte Z, Günther A-K, Kapoor S, Wang Y, Westendorf C. Cilia-driven flows in the brain third ventricle. Phil Trans R Soc B. 2019;375:20190154. https://doi.org/10.1098/rstb.2019.0154.

60. Mains RE, Eipper BA. The neuropeptides. In: basic neurochemistry: molecular, cellular and medical aspects. 6th ed. Philadelphia: Lippincott-Raven; 1999.

61. Chen P, Hong W. Neural circuit mechanisms of social behavior. Vol. 98, Neuron. Elsevier Inc.; 2018. 16–30.

62. Humphries MD, Prescott TJ. The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog Neurobiol. 2010;90(4):385–417. https://doi.org/10.1016/j.pneurobio.2009.11.003.

63. Coutellier L, Beraki S, Ardestani PM, Saw NL, Shamloo M. Npas4: A neuronal transcription factor with a key role in social and cognitive functions relevant to developmental disorders. PLoS One. 2012;7(9):e46604.

64. Yamaguchi T, Wei D, Song SC, Lim B, Tritsch NX, Lin D. Posterior amygdala regulates sexual and aggressive behaviors in male mice. Nat Neurosci. 2020; 23:1111–24.

65. Yang M, Gao F, Liu H, Yu WH, He GQ, Zhuo F, et al. Immunolocalization of aquaporins in rat brain. J Vet Med Ser C Anat Histol Embryol. 2011;40(4): 299–306. https://doi.org/10.1111/j.1439-0264.2011.01070.x.

66. Yang M, Gao F, Liu H, Yu WH, Sun SQ. Temporal changes in expression of aquaporin3, −4, −5 and −8 in rat brains after permanent focal cerebral ischemia. Brain Res. 2009;1290:121–32. https://doi.org/10.1016/j.brainres.2009.07.018.

67. Faragó N, Kocsis ÁK, Braskó C, Lovas S, Rózsa M, Baka J, et al. Human neuronal changes in brain edema and increased intracranial pressure. Acta Neuropathol Commun. 2016;4(1):78. https://doi.org/10.1186/s40478-016-0356-x.

68. Hatakeyama S, Kitagawa M, Nakayama K, Shirane M, Matsumoto M, Hattori K, et al. Ubiquitin-dependent degradation of IκBα is mediated by a ubiquitin ligase Skp1/Cul 1/F-box protein FWD1. Proc Natl Acad Sci U S A. 1999;96(7): 3859–63. https://doi.org/10.1073/pnas.96.7.3859.

69. Turner CE. Molecules in focus: Paxillin. Int J Biochem Cell Biol. 1999;30:955–9.

70. Tsuji K, Mizumoto K, Yamochi T, Nishimoto I, Matsuoka M. Differential effect of ik3-1/cables on p53- and p73-induced cell death. J Biol Chem. 2002; 277(4):2951–7. https://doi.org/10.1074/jbc.M108535200.

71. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. Proc Natl Acad Sci U S A. 2014;111(45):16219–24. https://doi.org/10.1073/pnas.1408886111.

72. Lewis R, Curtis JT. Male prairie voles display cardiovascular dipping associated with an ultradian activity cycle. Physiol Behav. 2016;156:106–16. https://doi.org/10.1016/j.physbeh.2016.01.012.

73. Wallace G, Elden M, Boucher R, Phelps S. An automated radio-telemetry system (ARTS) for monitoring small mammals. bioRxiv. 2021. https://doi.org/10.1101/2021.03.06.434221.

74. Gerkema MP, van der Leest F. Ongoing ultradian activity rhythms in the common vole, *Microtus arvalis*, during deprivations of food, water and rest. J Comp Physiol A. 1991;168(5):591–7. https://doi.org/10.1007/BF00215081.

75. Northcutt KV, Lonstein JS. Social contact elicits immediate-early gene expression in dopaminergic cells of the male prairie vole extended olfactory amygdala. Neuroscience. 2009;163(1):9–22. https://doi.org/10.1016/j.neuroscience.2009.06.018.

76. Donaldson ZR, Young LJ. Oxytocin, vasopressin, and the neurogenetics of behavior. Science. 2008;322(5903):900–4. https://doi.org/10.1126/science.1158668.

77. Garrison JL, Macosko EZ, Bernstein S, Pokala N, Albrecht DR, Bargmann CI. Oxytocin/vasopressin-related peptides have an ancient role in reproductive behavior. Science. 2012;338(6106):540–3. https://doi.org/10.1126/science.1226201.

78. Pitkow L, Sharer C, Ren X, Insel TR, Terwiliger EF, Young LJ. Facilitation of affiliation and pair-bond formation by vasopressin receptor gene transfer into the ventral forebrain of a monogamous vole. J Neurosci. 2001;21(18):7392–6. https://doi.org/10.1523/JNEUROSCI.21-18-07392.2001.

79. Young LJ, Nilsen R, Waymire K. Increased affiliative response to vasopressin in mice expressing the V1a receptor from a monogamous vole. Nature. 1999;400(August):1998–2000.

80. Kelly AM, Goodson JL. Hypothalamic oxytocin and vasopressin neurons exert sex-specific effects on pair bonding, gregariousness, and aggression in finches. Proc Natl Acad Sci U S A. 2014 Apr 22;111(16):6069–74. https://doi.org/10.1073/pnas.1322554111.

81. Gil M, Bhatt R, Picotte KB, Hull EM. Sexual experience increases oxytocin receptor gene expression and protein in the medial preoptic area of the male rat. Psychoneuroendocrinology. 2013 Sep;38(9):1688–97. https://doi.org/10.1016/j.psyneuen.2013.02.002.

82. Barrett CE, Keebaugh AC, Ahern TH, Bass CE, Terwilliger EF, Young LJ. Variation in vasopressin receptor (Avpr1a) expression creates diversity in behaviors related to monogamy in prairie voles. Horm Behav. 2013;63(3):518–26. https://doi.org/10.1016/j.yhbeh.2013.01.005.

83. DeVries AC, DeVries MB, Taymans SE, Carter CS. The effects of stress on social preferences are sexually dimorphic in prairie voles. Proc Natl Acad Sci. 1996;93(21):11980–4. https://doi.org/10.1073/pnas.93.21.11980.

84. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

85. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics - a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25(3):415–6. https://doi.org/10.1093/bioinformatics/btn647.

86. Love MI, Anders S, Kim V, Huber W. RNA-Seq workflow: Gene-level exploratory analysis and differential expression [version 2; peer review: 2 approved]. F1000Research. 2016;4:1070.

87. Kolde R. pheatmap: Pretty Heatmaps. R Package. 2019. Available from: https://cran.r-project.org/package=pheatmap

88. Wickham H. ggplot2: elegant graphics for data analysis. New York: Spring-Verlag; 2016. Available from: https://ggplot2.tidyverse.org. https://doi.org/10.1007/978-3-319-24277-4.

89. Neuwirth E. RColorBrewer: ColorBrewer palettes. R package. 2014; Available from: https://cran.r-project.org/package=RColorBrewer.

Publisher's Note