Single nucleus RNA-sequencing reveals transcriptional synchrony across different relationships

Liza E. Brusman[1], Allison C. Fultz[2], Robin D. Dowell[1,3], Mary A. Allen*[1,3], Zoe R. Donaldson*[1,2]

*These authors contributed equally

[1]Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, 80309

[2]Department of Psychology and Neuroscience, University of Colorado Boulder, 80309

[3]Biofrontiers Institute, University of Colorado Boulder, 80309

Correspondence: Zoe.Donaldson@colorado.edu; Mary.A.Allen@colorado.edu

Keywords: prairie vole, social bond, nucleus accumbens, synchrony, single nucleus RNA-sequencing

Short title: Cellular transcription and social bonds

Abstract

Relationships are shaped by reciprocal interaction and feedback between individuals. As relationships mature, pairs share common goals, improve their ability to work together, and experience coordinated emotions. However, the neural underpinnings responsible for this unique, pair-specific experience remain largely unexplored. Here, we used single nucleus RNA-sequencing to examine the transcriptional landscape of the nucleus accumbens (NAc) in socially monogamous prairie voles in long-term peer or mating-based relationships. We identified cell type-specific transcriptional differences between relationship types, including proportional differences in subpopulations of medium spiny neurons and module-based gene expression differences in interneurons. We also identified five gene modules correlated with different facets of social preference behavior. Finally, we show that, regardless of relationship type, prairie vole pairs exhibit transcription-based synchrony at the level of individual cells. Together, our results are consistent with a model in which a subset of gene expression changes promote relationship type-appropriate behaviors, while other non-overlapping gene expression changes support the social behaviors that are common across affiliative relationships. In addition, the similarity of gene expression observed across partners suggests an important role for the pair-specific social environment in shaping the NAc transcriptional landscape. This represents an emergent cellular property of social bonds that provides a potential biological mechanism by which shared social experience reinforces and strengthens relationships.

Main

Social bonds are shaped through the continuous integration of shared experiences, creating a unique "relationship canon", composed of these experiences, social norms, and idiosyncrasies that sum to a shared understanding of each other's mental state. This complex learning process allows individuals to tune their responses to their partner's actions. The resulting mutual understanding is critical for successfully navigating real-time interactions and to achieve long-term collaboration; it develops trust, enables partners to work together to achieve common goals, and leads to emotional coordination and empathy.

How a pair is able to coordinate their behavior has been a long-standing question in social neuroscience. Non-invasive human research has repeatedly highlighted interbrain neural synchrony as an emergent property of social interaction[1]. Functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy, electroencephalography or magnetoencephalography used to record neural activity, or its proxies, in two or more individuals simultaneously has shown that interbrain synchrony scales with relational closeness, and is associated with enhanced empathy, liking, rapport, and prosocial behavior[2–6]. While neural synchrony may initially arise from real-time interpretation of the same stimuli, its strengthening between individuals in a close relationship likely reflects aligned cognitive processing, even in the absence of shared cues[7].

In the absence of shared cues, neural synchrony and organized intra-pair behavior must rely on pre-existing, common biological substrates that prime neurons to fire synchronously. This underlying state is likely achieved through shared prior experiences that have altered cell states in the same way. Thus, we would expect this shared biological state to emerge as a product of social experience with another individual. Indeed, in pairs of fighting betta fish, brain-wide gene transcription becomes more correlated between partners as the duration of fighting increases[8]. Critically, this transcriptional synchrony is evident only between fighting partners; cross-pair comparisons showed that it is not attributable to the duration or intensity of fighting.

In the present study, we investigated if a similar pair-specific transcriptional state exists in the context of different types of social bonds, perhaps facilitating a pair's ability to work together effectively. While prior studies have primarily focused on cortical regions, limited fMRI evidence has demonstrated neural synchrony in the striatum between individuals when they have congruent emotional responses to the same piece of music or speech[9,10]. To investigate whether transcription-based synchrony exists in the nucleus accumbens (NAc), a ventral striatal region extensively implicated in social bonds, we performed single nucleus RNA-sequencing (snRNA-seq) of the NAc in peer- and mate-bonded prairie voles, which form affiliative relationships in both contexts, akin to human friendships and marital relationships, respectively[11,12]. While there are transcriptional differences between pairing types, we found that voles were more transcriptionally similar to a partner they were living with than other voles regardless of sex or pairing type. We refer to this as transcription-based synchrony. Our data are the first to identify a potential cell-level molecular mechanism underlying organized intra-pair patterns of neural activity (synchrony) and behavior, providing novel insights into how pair-specific behaviors may be reflected a cellular level.

Prairie vole NAc transcriptional landscape across relationship types

The nucleus accumbens (NAc) is involved in emotional and reward processing[13]. Neuromodulatory signaling within this region is integral for bond formation, which results in subsequent long-term changes in gene expression and physiology[14–17]. Here, we used snRNA-seq to identify cell-type specific changes that support long-term social bonds. Prairie voles form

peer and mating-based relationships akin to friendships and romantic relationships[11,12]. While these two different types of relationships share many behavioral aspects, some behaviors exhibited are unique to the relationship type (eg. reproductive behaviors in mate pairs). To examine transcriptional differences and similarities across different types of relationships, we paired female and male prairie voles with a novel same-sex partner (peer-paired) or novel opposite-sex partner (mate-paired). Partner preference tests performed after 14 days of pairing on 20 vole pairs (4FF, 5MM, 11FM). Of the 40 voles, 39 displayed a partner preference (>2/3 time huddling with partner vs. novel; Fig S1). This is consistent with prior reports in which a small subset of voles do not form a partner preference. There were no group differences in partner preference metrics based on sex or pairing type (Fig S1). These results are consistent with prior reports that prairie voles will form selective affiliative relationships with same- or opposite-sex species members, providing a platform to identify the unique and shared features of different types of social bonds.

We first mapped the transcriptional landscape of the prairie vole NAc in different pairing types. We collected tissue 16 days post-pairing (2 days post-behavior testing). 39 out of 40 tissue samples passed quality control and were merged to create a combined dataset of 142,488 single nuclei transcriptome profiles that passed quality filtering (Table S1). The 39 remaining animals were comprised of 8 peer-paired females, 10 peer-paired males, 11 mate-paired females, and 10 mate-paired males. In pair-wise analyses, the unpaired animal (one female from a mate pair) was excluded. We identified 15 transcriptionally-defined cell type clusters in the NAc based on known markers of cell types in this region (Fig 1A, B). Medium spiny neurons (MSNs), the primary neuronal cell type within the NAc, were identified based on their combined expression of the dopamine receptors *Drd1a* and *Drd2*, the opioid receptor *Oprm1*, and opioid ligands *Pdyn* and *Penk*. This resulted in five MSN clusters we termed MSN-Drd1Pdyn, MSN-Drd1PdynOprm1, MSN-Drd1Penk, MSN-Drd2Penk, and MSN-Drd2NoPenk. Five interneuron clusters were identified based on their expression of only *Gad1/Syt* (Int-GABAergic), *Igfbpl1/Dlx2* (Int-Dlx2Immature), *Sst/Npy* (Int-SstNpy), *Vip/Kit* (Int-Pvalb), and *Chat* (Int-Cholinergic). Glia were identified based on their expression of *Mog* (MatureOligos), *Olig2/Pdgfra* (oligodendrocyte progenitor cells, or OPCs), *Gja1* (Astrocytes), *Aif1* (Microglia), and *Vim* (RadialGlia-LikeCells). These clusters are largely consistent with those identified in snRNA-seq datasets from other species including mice[18], rats[19], and humans[20], suggesting overall cell-type conservation in the NAc. This is congruent with this brain region playing a conserved role in learning and reward across species.

Having identified the major cell classes that comprise the NAc, we next asked whether cell-type proportions differed by sex or pairing condition. For most clusters (13/15), the proportion of cells in each cluster did not differ between sexes or pairing types (Fig S2). However, mate-paired females had a higher proportion of MSN-Drd1Pdyn and MSN-Drd2Penk neurons than peer-paired females (GLM with estimated marginal means, FDR corrected: MSN-Drd1Pdyn: $T = 2.93$, $p = 0.024$; for MSN-Drd2Penk: $T = 2.80$, $p = 0.33$), and peer-paired males had higher proportion of MSN-Drd1Pdyn neurons than peer-paired females ($T = 2.44$, $p = 0.041$, Fig 1C, D, E). This proportional shift is unlikely to be due to social experience-based neurogenesis; rather, it is likely that shifts in neuronal gene expression led to different classification of these cells. This is consistent with prior work indicating that dopamine D1-class receptors and the kappa opioid receptor ligand *Pdyn* are upregulated in the NAc upon bonding in prairie voles, which may contribute to mate bond maintenance through enhanced selective aggression[21,22]. Such changes likely reflect the brain's need to adopt a different behavioral repertoire in different relationship types.

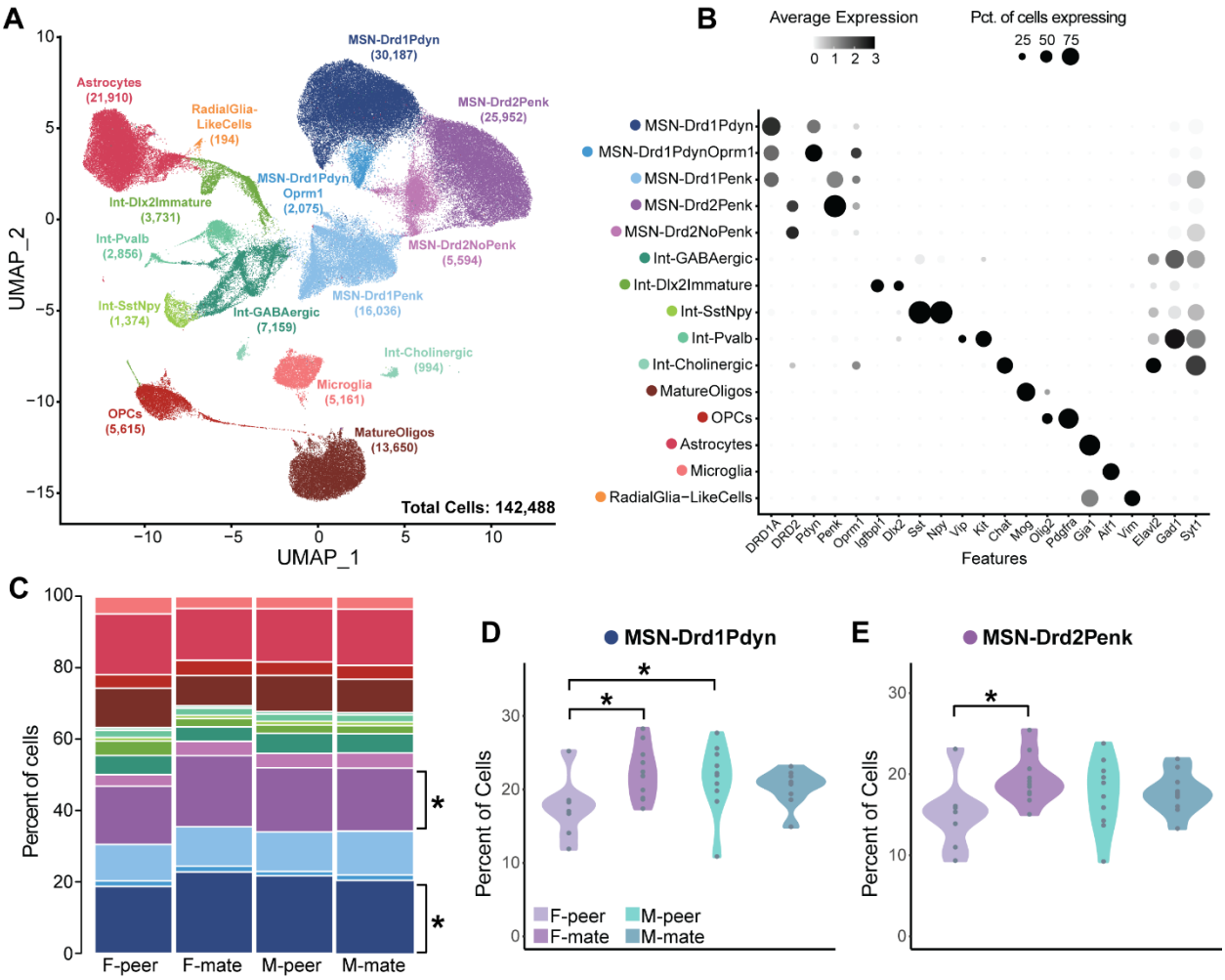

Figure 1. Prairie vole NAc transcriptional landscape across relationship types. A. UMAP clustering of NAc cells revealed 15 clusters. **B.** Dotplot indicating expression of marker genes. **C.** Stacked bar plot of cell type proportions calculated per experimental group. **D, E.** Cell type proportions differ between groups in MSN-Drd1Pdyn and MSN-Drd2Penk clusters. *p < 0.05.

Gene expression patterns underlying different relationship types and individual differences in relationship behaviors

In a separate line of inquiry, we identified gene sets that differed between groups. We used Hotspot, a computational algorithm that groups genes into modules based on similar expression patterns across cells[23,24]. Using a minimum threshold of 30 genes per module, we identified 23 modules in total that range in size from 31-415 genes (Fig S3, Table S2). These coexpression modules represent the broad cell states and dynamics of the NAc, capturing major cell types and their functions. Accordingly, mapping expression of these modules onto clusters revealed that many modules were highly enriched in a single cell cluster (eg. Module-1 in microglia) while others were expressed across clusters of a given cell type (eg. Module-20 across all clusters, Fig S3). Thus, we used this platform to holistically examine potential gene expression changes that are informative to but not constrained by cell type clusters. We found that the majority (20/23) modules did not differ in their expression based on sex or relationship type (Fig S4),

providing a handful of modules that may underpin relationship differences across females and males and between pairing types. One of these was observed in Module-4, which is localized to radial glia-like cells and is a very small cell cluster not detected in all animals. Thus we did not pursue additional inquiry into this module.

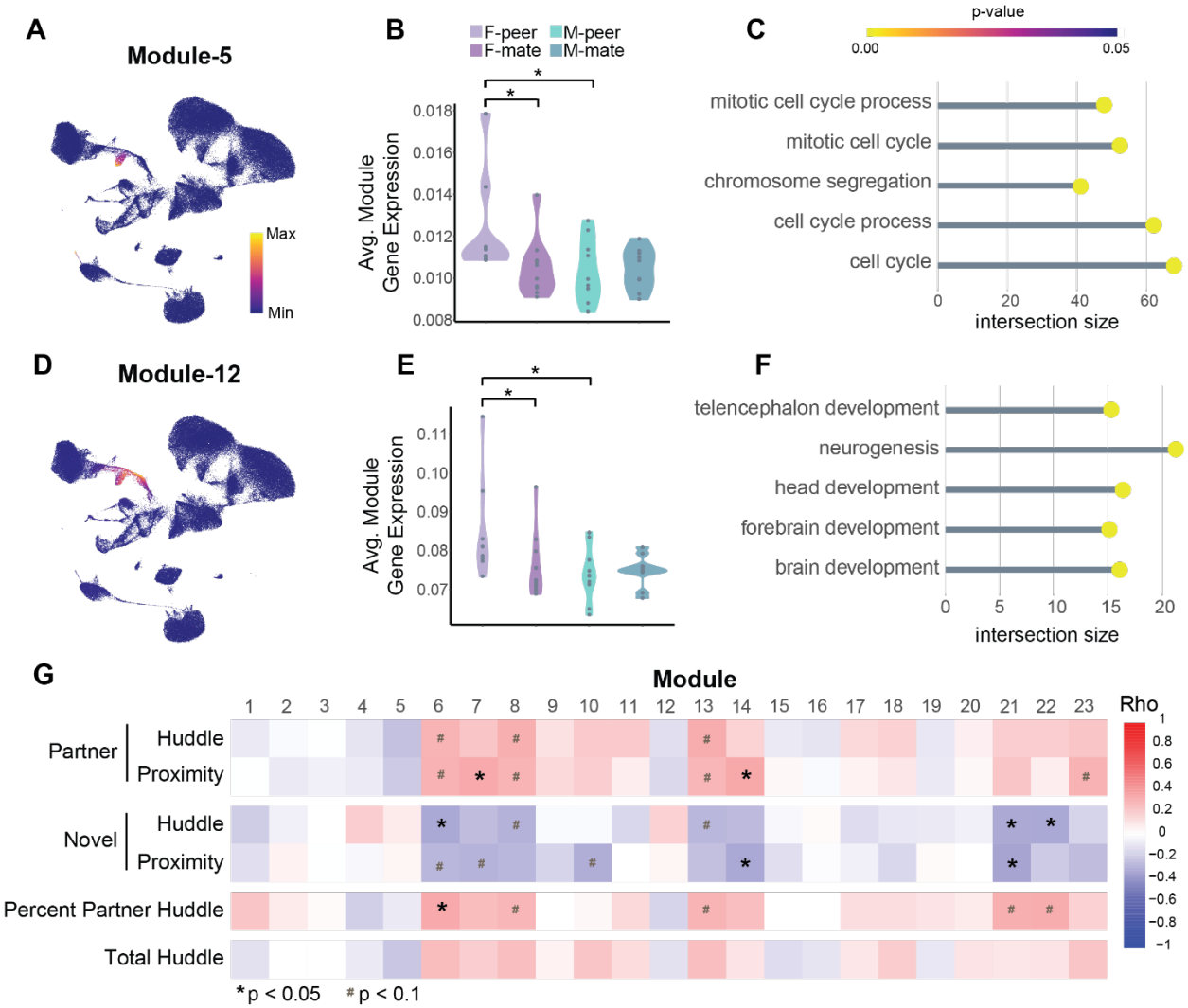

Figure 2. Gene expression patterns underlying different relationship types and individual differences in social preference behavior. A, D. Hotspot module score per cell for Module-5, and Module-12 overlaid on UMAP. **B, E.** Average module gene expression (mean(log(counts))) for Module-5 and Module-12 differs between groups. **C, F.** Top GO terms for genes in each module. Intersection size indicates number of genes in the module also in that GO term. **G.** Relationship between average module gene expression and behavioral metrics within the partner preference test independent of relationship type. *p<0.05, #p<0.1.

We found that the average (mean(log(counts))) expression of genes in Module-5 and Module-12, which are expressed in Dlx2+ immature interneurons, also had higher expression in peer-paired females compared to mate-paired females and peer-paired males (GLM with estimated marginal means, FDR corrected: Module-5 peer vs. mate females: T = 2.71, p = 0.021; peer

169 females vs. peer males: T = 2.84, p = 0.021; Module-12 peer vs. mate females: T = 2.52, p =
170 0.033; peer females vs. peer males: T = 2.94, p = 0.023; Fig 2A, B, D, E). Module-5 is
171 associated with GO terms related to the cell cycle, and Module-12 is associated with GO terms
172 related to neurodevelopment (Fig 2C, F). Together, these modules may reflect proliferation of
173 neural progenitors, neurogenesis and/or maturation of these immature neurons. In humans,
174 retrospective birth dating of neurons that incorporated ^{14}C from nuclear bombs has revealed
175 continuous generation of striatal interneurons, and migration of neural progenitors from the
176 subventricular zone (SVZ) to the striatum and their subsequent maturation into inhibitory
177 interneurons has been observed in mice and rats[25,26]. Additionally, in female prairie voles,
178 neurogenesis has been observed in the striatum[25,27]. Finally, work in mice suggests that
179 postnatal maturation of interneurons can shape dopamine release dynamics that are relevant to
180 motivated behavior[28], potentially linking interneuron maturation and shifts in dopamine dynamics
181 that occur during in pair bonding[29]. Together, our data suggest that cell proliferation and/or
182 maturation may be important factors for pair bonding and the emergence of relevant behaviors
183 in different relationship types.

184 Next, we asked whether module gene expression was associated with individual differences in
185 bonding behaviors independent of relationship type (Fig 2G). Our behavioral data are limited to
186 the partner preference test where we can examine time spent in contact or near either the
187 partner or novel animal, or a combination of the two as an overall metric for sociality (Fig S1).
188 We found that five of 23 modules were correlated with social preference behavior while none
189 were correlated with total interaction/huddle. This suggests a relationship between gene
190 expression and specific aspects of partner preference independent of overall sociality. In most
191 instances, module gene expression was correlated with partner affiliation or novel rejection, but
192 not both, suggesting a dissociation of the pro-and anti-social behaviors that comprise behavioral
193 selectivity within relationships. Of note, Module-14 is the only module that is positively
194 correlated with partner interaction *and* negatively with novel interaction, consistent with
195 previously documented broad roles for these genes, including *Drd1*, *Drd2*, *Pdyn*, and *Penk*, in
196 both partner-directed affiliation and rejection of novel individuals[21,22,29,30]. Modules-6, -21, and -
197 22 were also negatively correlated with novel proximity or novel huddle. Module-6 is enriched in
198 OPCs and is associated with GO terms indicating a role in oligodendrocyte maturation and
199 myelination. Modules-21 and -22 are enriched in Drd1Penk MSNs and Sst/Npy+ interneurons,
200 respectively, and are broadly associated with GO terms related to cell-cell signaling. Finally,
201 Module-7 was positively correlated with partner interaction. Module 7 is enriched in all MSN
202 clusters and is associated with GO terms related to neuronal activity. These findings elucidate
203 the relationship between gene expression and individual differences in behavior as they relate
204 to social preference.

205 The modules that exhibit differences based on relationship type do not overlap with those that
206 correlate with individual differences in preference behavior. This is consistent with a model in
207 which changes in the expression of a subset of genes facilitates relationship-appropriate
208 behaviors, such as the onset of mating and reproduction exclusively in mating-based pairs.
209 Other genes contribute to behaviors that are common across affiliative relationships but which
210 vary in their expression across individuals, including preference to interact with a bonded
211 partner and rejection of unknown individuals. Thus, these complementary gene expression
212 patterns likely act in concert to support the behavioral state transition that accompanies bond
213 formation.

214 **Transcription-based synchrony within pairs**

215 Social bonds, unlike other behavioral changes, are characterized not only by changes in gene
216 expression, but also by emergent dyadic properties including synchronization of neural activity

217 that scales with relational closeness[2,5]. Here, we tested the hypothesis that pairs would exhibit
218 transcription-based synchrony via three different approaches. We began by examining pairwise
219 relationships in Hotspot module gene expression. We found that, regardless of pairing type, two
220 of our Hotspot modules, Module-6 (expressed in OPCs, Fig 3A) and Module-11 (expressed in
221 interneurons, Fig 3D), exhibited pairwise expression similarity using multiple metrics. First,
222 Module-6 and Module-11 average gene expression is correlated between partners (Module-6:
223 Rho = 0.61, p = 0.0057; Module-11: Rho = 0.59, p = 0.0075, Fig 3B, 3E). Second, using the
224 expression of all of the genes within each module to calculate the Euclidean distance between
225 partners, we found that for these two modules, partners were closer together in Euclidean space
226 than non-partners regardless of relationship type (Wilcoxon rank sum tests: Module-6: W =
227 9781, p = 0.00017 ; Module-11: W = 8803, p = 0.0083, Fig 3C, F, Fig S6). Finally, when we
228 ranked the animals by their module gene expression and calculated the rank-based distance
229 between partners, we found that partners were closer together in the ranking than would be
230 expected by chance (One-sample t-tests vs. null (expected) value: Module-6: T = -3.46, p =
231 0.0028; Module-11: T = -2.73, p = 0.014). In all instances, there was no difference in the degree
232 of pairwise similarity (via correlation, Euclidean gene expression distance, or rank) based on
233 relationship type (Figs S5, S6, S7). GO terms for Module-6 (enriched in OPCs), include terms
234 related to oligodendrocyte differentiation and myelination, and GO terms for Module-11
235 (enriched in interneurons) include terms related to synaptic signaling and ion transport. This
236 suggests that oligodendrocyte maturation, subsequent myelination by mature oligodendrocytes,
237 and/or synaptic pruning by oligodendrocytes could be sculpting similar population-level neural
238 activity between partners. Among the genes in Module-6, *Dscam* stands out. This cell adhesion
239 molecule identified as one of 24 genes whose differential expression is conserved in
240 taxonomically diverse monogamous species when compared to closely related non-
241 monogamous species[31]. Further, modulation of neural activity by oligodendrocytes may be the
242 driver behind similar synaptic signaling between partners in interneurons. As interneurons
243 coordinate synchronous firing amongst neurons within the NAc, they also provide an ideal
244 candidate population to orchestrate interbrain neural synchrony at a population level. Together,
245 these data indicate that specific gene networks in OPCs and interneurons are synchronized
246 between partners and that oligodendrocyte maturation and interneuron signaling may be drivers
247 of pair-wise transcriptional similarity regardless of pairing type.

248 Given that partners exhibited correlated expression of Modules-6 and -11, we asked whether
249 the average expression of these modules within a pair was tied to their mutual behavior. We
250 found that pairwise Module-6 expression was correlated with average partner huddle within a
251 pair (Rho = 0.51, p = 0.024), but Module-11 expression was not correlated with any metrics of
252 partner preference (Fig 3J, K). How might synchronized gene expression of Module-6 in OPCs
253 contribute to predictable pairwise behavior, such as the average amount of partner-directed
254 huddling? While OPCs are unlikely to shape neuronal activity on the fast timescales that drive
255 many behavioral decisions, their coordinated activity across two brains may result in a similar
256 neurocellular state through myelination and synaptic pruning. Arguably, such effects will be
257 observed more readily in aggregate behavior over time, such as partner-directed huddling
258 across multiple hours in the partner preference test. Correspondingly, interneurons are
259 potentially poised to contribute to behavior on much faster timescales not examined in the
260 present study, which may explain the lack of correlation observed for Module-11. Supporting
261 this, the biological pathways most strongly implicated by genes in Module-11 include potassium
262 channels (Ingenuity Pathway Analysis, Qiagen; $p = 1.3e^{-11}$), which are likely to affect electrical
263 properties of these neurons.

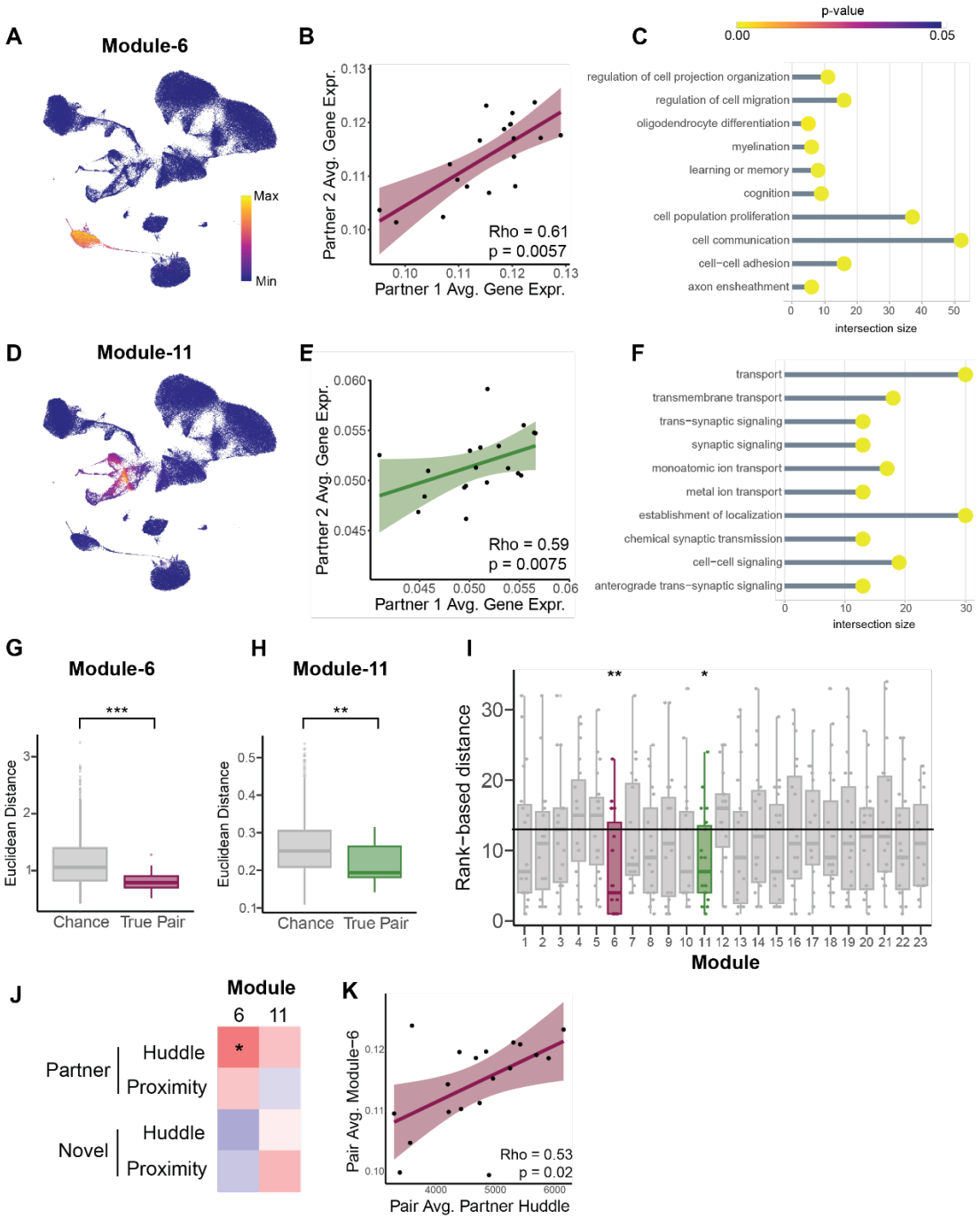

Figure 3. Transcriptional synchrony within pairs. A, D. Hotspot module score per cell for Module-6 and Module-11 overlaid on UMAP. **B, E.** Spearman correlation between partners for average module gene expression in Module-6 and Module-11. **C, F.** Selected GO terms for modules 6 and 11, prioritized based on relevance for neural function. **G, H.** Counts for genes in each module were used to calculate Euclidean distance between partners for that module and compared to all non-true pair distances (Chance). **I.** Rank-based distance between partners for each module (dotted line at y =13 represent chance distance). **J,K.** Within a pair, average Module-6 gene expression is correlated with average partner huddle behavior. *p<0.05, **p<0.005, ***p<0.0005

As an orthogonal approach that leverages the unique advantages of our cellular-resolution dataset, we trained support vector machines (SVMs) to identify the animal from which individual cells originated based on the cell's gene expression. We found that, in every cluster, our SVMs are most likely to classify cells correctly as the animal that they originated from (e.g. self, Fig 4A, left). The mean percent of self classifications ranged from 24.9% to 72.2% across clusters, while random chance would predict correct self classification only 2.6% of the time. While the classifiers always performed better than chance ($p < 0.001$ for chance vs. self in all clusters), they also performed better on some clusters than others. This may be due to there being fewer cells in these clusters (and therefore less training information), which is suggested by a correlation between number of cells in the training dataset and self prediction accuracy (Spearman Rho = 0.88, $p = 7.5 \times 10^{-5}$, data not shown), and it may also be due to some cell types having more homogenous gene expression across all animals.

When cells were mis-classified by the SVM, they were much more likely to be classified as the partner animal than any other non-partner animal (Fig 4A, left, $p < 10^{-5}$ for all partner vs. other comparisons) regardless of sex or relationship type (Fig S9 A, B, C). Mean partner classifications ranged from 7.4% to 25.3% across clusters, still significantly greater than chance (2.6%) albeit lower than self. Most importantly, this was not due to cohort effects, providing evidence that partners are more transcriptionally similar than non-partners.

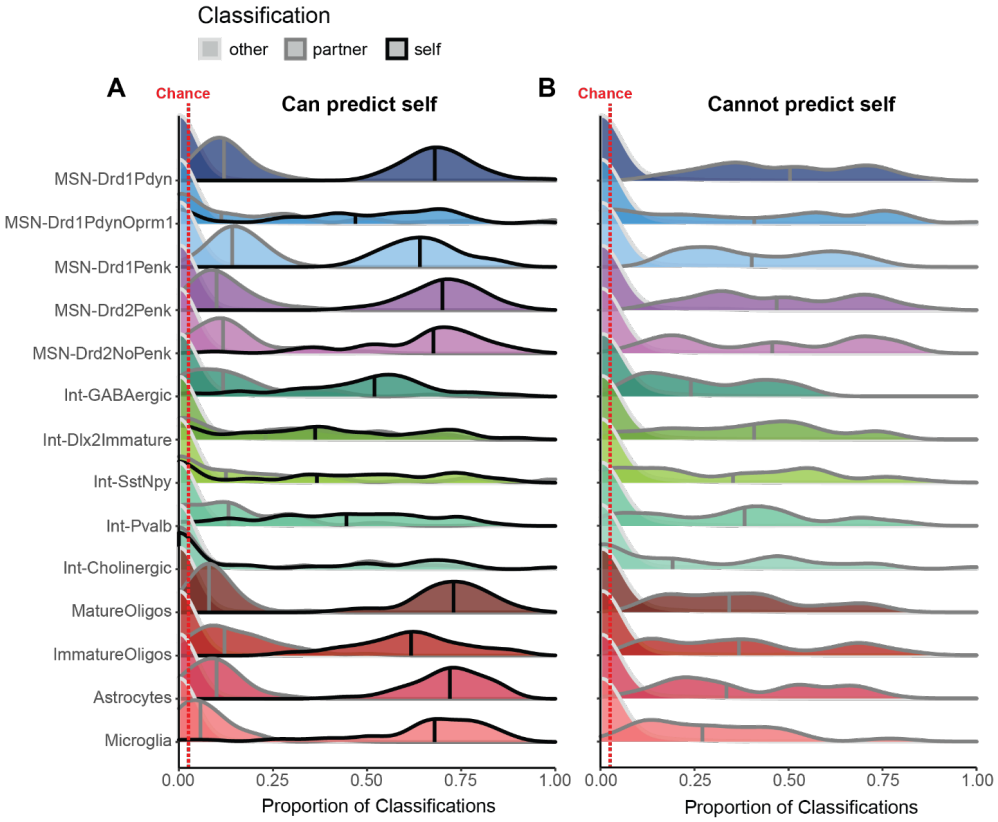

Figure 4. Predicted cell identity reveals similarity between partners. A. An SVM was trained for each cluster to predict the animal identity of individual cells. The SVM predominantly classified cells correctly (self, black outline), with the most common misclassification assigned to their partner (dark grey outline), followed by any other animal (light grey outline). **B.** SVMs were iteratively trained for each cluster by holding one animal out from the training set, and then testing the SVM on the held-out animal. The held-out animal was most frequently classified as their partner (dark grey outline) rather than any other animal (light grey outline).

To further leverage the strengths of machine learning, we tasked an SVM with classifying cells when the correct choice (self) was not available. To that end, we iteratively trained SVMs on cellular-level gene expression data from all animals with one animal held out, and then tested our SVMs on cells only from the held-out animal, thereby forcing the SVMs to choose an animal other than the self. We predicted that, in this scenario, cells originating from the held out animal would most frequently be predicted to be that animal's partner. Strikingly, when the SVMs were unable to classify the cells as the self, they were most likely to classify the cells as originating from the partner animal (Fig 4A, right, $p < 10^{-16}$ for partner vs. other in all clusters). The mean percent of partner classifications ranged from 27.3% to 48.5% across clusters. While this was lower than accurate self classification, it dramatically exceeded random chance, which would predict a partner classification only 2.7% of the time ($p < 10^{-5}$ for chance vs. partner in all clusters). These data indicate that SVMs represent a powerful tool for inference of cell identity and as a novel method for interrogating transcriptional similarity between samples.

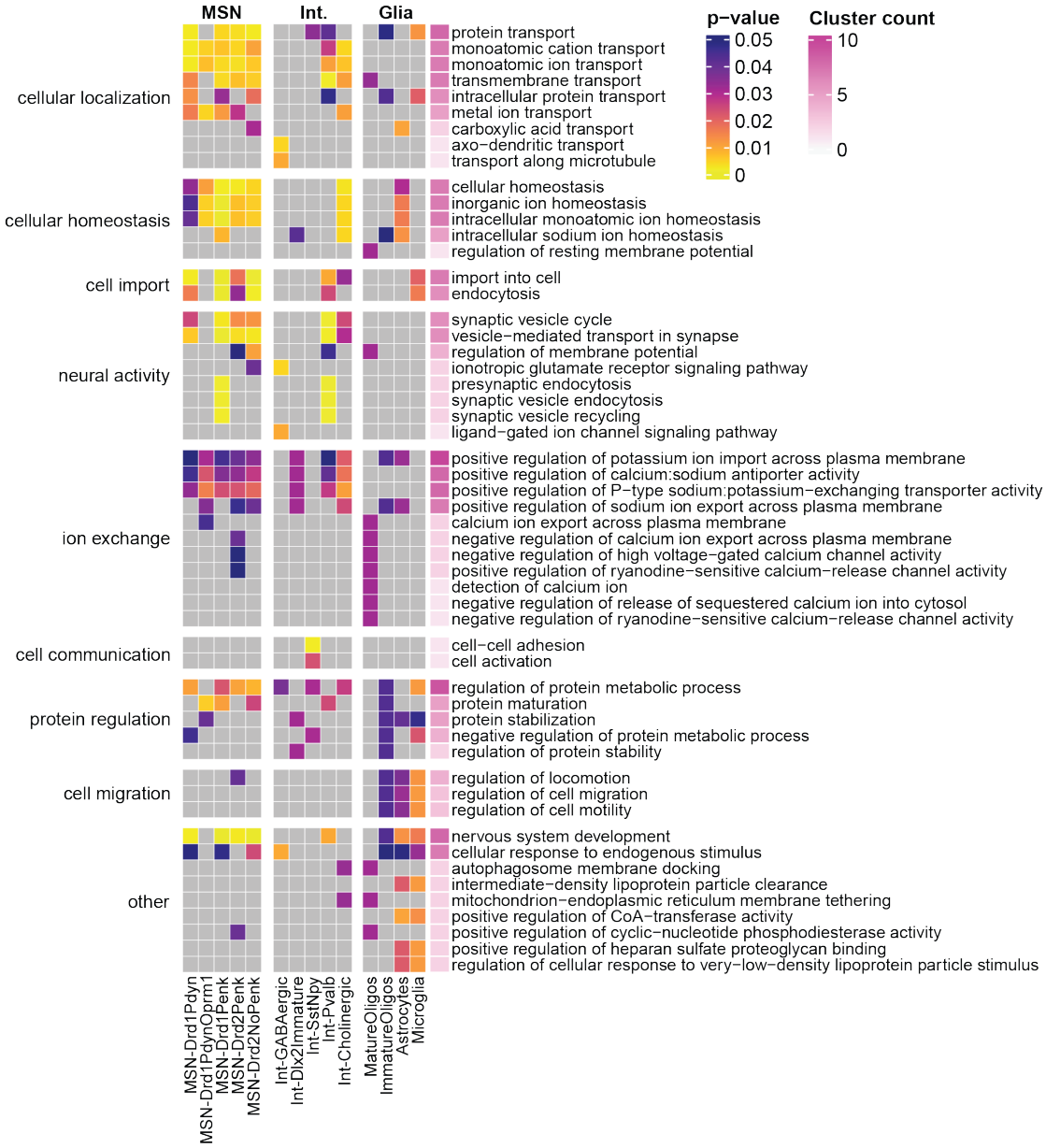

Figure 5. GO terms for genes correlated between partners in each cluster.

To determine which genes are driving transcriptional synchrony between partners, we focused on genes that were highly expressed within each cluster (avg expression > 1 count/cell) and performed Spearman correlations between partners. As a control, we iteratively shuffled partner identity 1000 times via bootstrapping. Correlated genes were defined as those having a true p-value < 0.05 and a Spearman Rho value > 2 standard deviations from the mean of 1000 shuffled pairings. We identified between 5 and 42 correlated genes per cluster (1.7%-29.2% of the genes that met the expression cutoff, Table S3). We found that the GO terms predicted from correlated genes identified across neuronal and glial clusters were broadly associated with cellular functions that may be important for information processing within the brain. These include terms related to cellular localization, cellular homeostasis, cell import, ion exchange, and protein regulation (Fig 5A). In addition, we identified terms related to cell communication and neural activity in neuronal clusters, and terms related to cell migration that were largely restricted to glial clusters (Fig 5A). Notably, 21.09% of correlated genes are identified as syndromic autism spectrum disorders (ASD) genes[32] (Table S3). This is 3.57 times greater than the representation of ASD genes among all genes detected in our dataset. Disruption of these genes in mice routinely yields altered social behavior[32]. Taken together, our data highlight that transcription-based synchrony is not unique to one cell type or process and involves genes with known effects on behavior. This is consistent with transcription-based synchrony leading to a distributed state that performs many interconnected functions.

Discussion

Using single nucleus transcriptomics in prairie voles, we provide new insights into the cellular basis of different types of social bonds. Our data indicate that there are differences in cellular gene expression that underlie different social relationships and individual differences in social behavior, and that pairs exhibit similarity in cellular transcriptional states regardless of relationship type, which we refer to as transcription-based synchrony. Differences in gene expression based on relationship type may enable relationship-appropriate behaviors, such as the onset of reproductive behaviors and mate guarding only in mating-based pairs. In contrast, gene expression patterns that correlate with social preference behaviors across peer and mate bonds may support the shared behavioral features of affiliative relationships. Transcription-based synchrony, which was robustly evident independent of relationship type or sex, may serve as a pair-specific mechanism that contributes to the organization of successful intra-pair behavior. This is consistent with a model in which a relationship type-specific transcriptional state is complemented by a pair-specific transcriptional state, enabling behaviors that are both appropriate for the relationship type and specific to a pair.

We provide compelling evidence for transcription-based synchrony at the cellular level, but the factors that facilitate this shared transcriptional state and/or those that contribute to variation in synchrony remain unknown. One hypothesis is that a shared physical environment drives transcription-based synchrony. All animals share colony room conditions (temperature and humidity, food, water), and pairs additionally share non-social objects (nesting material, plastic houses) in the cage. Supporting a broad role for shared environment in establishing similar transcriptional states that scale with degree of environmental overlap, when our SVM misclassified cells as a non-partner animal, it was most likely to assign the cell to an animal collected in the same behavioral cohort (but not the same sequencing cohort). However, even after analyzing the SVM results within cohorts, animals were still more likely to be classified as their partner than another animal. This indicates that, while environment may have some impact on transcriptional synchronization, pair-specific experiences are the primary driver of this phenotype.

The microbiome provides a particularly intriguing potential mechanism for driving transcription-based synchrony. In humans and other animals, sharing a house or a burrow leads to microbial similarity[33–35]. While the microbiota-gut-brain axis has not been fully explored in the context of social bonding, the microbiome itself can shape social behavior. Antibiotic treatment can cause changes in social behavior and fecal microbiota transfers from humans with depression or social anxiety into rodents induces corresponding behavioral changes in these animals[36–38]. One mechanism that has been proposed to mediate this interaction is microbe-driven epigenetic regulation in the brain[39]. Future work may test this by providing antibiotics to one member of a pair or by using fecal transplants across animals living in separate cages.

Yet the above environmental and microbial factors alone are unlikely to fully explain the emergence of transcription-based synchrony in all contexts. Brain-wide transcriptomes become correlated in pairs of fighting betta fish within 60 minutes, a duration too short to be explained by emerging microbiome similarity[8]. As in our study, cross-pair comparisons were also used as a control. In addition, if environment or microbiome alone were the primary driving factor of transcriptional synchrony, we would expect correlation across pairs to be the same whereas we observe substantial variation in pairwise synchrony levels. This raises the intriguing possibility that the most salient and dynamic features of an environment - namely social interactions with a cohoused partner - are a primary and heterogeneous driver of transcription-based synchrony. Future work examining this will look at transcription-based synchrony across the duration of a bond to determine how these properties emerge over time and examine facets of pair behavior as they correlate with levels of transcription-based synchrony.

Similar to interbrain synchrony at the level of neural activity, the functional effects of transcription-based synchrony remain untested. However, it is intriguing to consider how similarity in gene expression profiles across two individuals may contribute to successful relationships. Individuals in long-term, meaningful relationships often predict the actions of their partner, quickly infer their mental state, and choose an appropriate behavioral response. Among humans, this can even lead to an overlapping sense of self with the other individual and enables behavior coordination at many levels, including being able to “finish each other’s sentences.” Gene expression shapes individual cellular and circuit properties of a brain. At a physiological level, similar transcriptional profiles may provide the common biological substrate that facilitates enhanced interbrain synchrony. At a broader level, this alignment of neural activity may lead to a shared perception of reality, enhancing the likelihood of correctly predicting and responding to the mental state of another individual and ultimately cementing bonds.

While there has been substantial research linking transcriptional state to behavior, here we provide a new layer of understanding by delineating transcriptional changes that may enable relationship-appropriate behaviors as well as the emergence of transcription-based synchrony within pairs. As a result, pairs not only engage in behaviors that facilitate fitness, they come to understand and infer a partner’s mental state, enabling behavioral coordination. The latter represents a putative biological substrate underlying the neural and behavioral organization of relationships that enables the emergent properties of social bonds and makes a relationship more than the sum of its parts.

Author contributions

LEB performed all experiments, analyzed data, and wrote manuscript. ACF helped with behavioral experiments. RDD contributed analytical expertise and revised manuscript. MAA contributed to data analysis, reviewed code, and revised manuscript. ZRD helped conceive experiments and co-wrote the manuscript.

Acknowledgments

We thank Jay Hesselberth and his lab for allowing us to use their space to perform the nuclei isolations. We also thank the Genomics Core at CU Anschutz for sequencing our samples. We thank Jessica Abazaris and the rest of the animal care staff at the University of Colorado Boulder for their excellent care of the voles, and the voles themselves for their sacrifice. We thank Julie Sadino for her extensive feedback throughout each stage of this project, Zachary Johnson for discussion single nucleus data, and Catherine Peña for feedback on initial drafts of the manuscript. Deena Walker helped provide Ingenuity Pathway Analysis. This work was supported by NIH awards 1F31MH132278-01A1, 5T32GM008759-20, and 1T32GM142607-01 (to LEB), Whitehall and Dana Foundation awards, and DP2MH119421, UF1NS122124, R01MH125423, IOS-2045348, U01NS131406 (to ZRD), NIH award R01HL156475 (to RDD and MAA) and the Anna and John J. Sie Foundation/Global Down Syndrome Foundation (to MAA).

Ethics

All animal procedures were carried out in accordance with standard ethical guidelines (National Institutes of Health Guide for the Care and Use of Laboratory Animals) and approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Colorado Boulder.

Conflicts

The authors declare the following competing interests: MAA and RDD have a patent for measuring transcription factor activity from eRNA activity. RD was a co-founder of Arpeggio Biosciences. All other authors have no competing interests.

Data Availability

All sequencing data is deposited on GEO (GSE255620) and will be made available upon publication.

Methods

Animals

Prairie voles were bred in-house in a colony descended from wild animals collected in Illinois, USA. Animals were weaned at 21 days and housed in same-sex groups of 2-4 animals in standard static rodent cages (19l x 10.5w x 5 in. h) with ad-libitum access to rabbit chow (5326-3, PMI Lab Diet) and water. Sunflower seeds and dried fruit treats were given to supplement the diet, and cotton nestlets, PVC tubes, and plastic houses were given for enrichment. All voles were between 77-154 days old at the start of the experiment. On day one of the experiment, animals were co-housed with either a same-sex (peer-paired) or opposite-sex (mate-paired) animal in smaller standard static rodent cages (11l x 8w x 6.5h) with ad-libitum access to food and water, cotton nestlets and plastic houses. These pairings resulted in 11 female/male pairs, 4 female/female pairs, and 5 male/male pairs. The colony room was kept at 23-26C with a 14:10 light:dark cycle. All procedures were performed during the light phase and approved by the University of Colorado Institutional Animal Care and Use Committee.

Partner preference test

Partner preference tests were performed on day 14 of the experiment. Both partners within a pair were tested consecutively, with the order randomized. Briefly, partner preference tests were performed in a three-chamber plexiglass arena (76 cm long x 20 cm wide x 30 cm tall). Opaque dividers were placed between the chambers of the arena to prevent animals from seeing each other at the start of the test. Partner and novel animals were tethered to opposite ends of the chamber using tethers consisting of a wall-attached eye bolt with a chain of fishing swivels. Animals were briefly anesthetized with isoflurane and a zip tie was used around the animal's neck to attach them to the tether. Rabbit chow and water were given for the duration of the test. At the start of the test, the experimental animal was placed in the center chamber and the dividers were removed, allowing the experimental animal to freely roam the arena for 2.5 hours. The tests were recorded using overhead cameras (Panasonic WVCP304).

The videos were analyzed using TopScan software v3.0.0 using the parameters in Ahern et al[40]. Behavior was analyzed using a custom Python script developed in-house (https://github.com/donaldsonlab/CleversysSummaryRearranger) to calculate time spent huddling with the partner and time spent huddling with the novel. The partner preference score was calculated as (partner huddle time/[partner huddle time + novel huddle time]) *100%.

Tissue harvest

On day 16 of the experiment, animals were rapidly decapitated using a guillotine and brains were placed immediately on dry ice before moving to -80C. Uteri of female animals were removed and dissected to assess pregnancy status. Brains were later sectioned on a cryostat until the front of the NAc was reached (coordinates). A 2 mm diameter tissue punch was used to punch the NAc ~1mm deep on each side of the brain. Punches were stored separately for each animal in eppendorf tubes at -80C until nuclei dissociation.

Single nuclei isolation

Brain punches were placed on wet ice until barely thawed. 0.5 mL Complete Buffer HB (250mM sucrose, 25mM KCl, 5mM MgCl2, 20mM Tricine-KOH, 0.042% BSA, 0.06 U/ul RNAsin, 0.15mM spermine, 0.5mM spermidine, 1mM DTT, cOmplete EDTA-free protease inhibitor cocktail tablet) was then added to each tube with brain punches and then Complete Buffer HB and punches were transferred to 2 mL dounce homogenizer. Samples were dounced with a tight pestle 20

476 times. 16 uL 5% IGEPAL CA-630 in Buffer HB (250mM sucrose, 25mM KCl, 5mM MgCl2,
477 20mM Tricine-KOH) was added to the dounce homogenizer and the samples were dounced 10
478 more times. The samples were then filtered through a Flowmi 40 um strainer into a 2 mL LoBind
479 Eppendorf tube. The volume of flowthrough was measured and an equal volume of Working
480 Solution (50% iodixanol (OptiPrep), 25mM KCl, 5mM MgCl2, 20mM Tricine-KOH, 0.042% BSA,
481 0.064 U/ul RNAsin) was added to the homogenate and mixed by gentle pipetting using a wide-
482 bore P1000 pipette tip. First, 50% iodixanol was made by diluting 60% iodixanol (OptiPrep) in
483 Working Solution. Then, 30% and 40% iodixanol solutions were made by diluting the 50%
484 iodixanol/Working Solution in Complete Buffer HB. In 0.8 uL ultracentrifuge tubes, 160 uL 30%
485 iodixanol was carefully layered on top of 80 uL 40% iodixanol and the interface was labeled with
486 pen. Approximately 560 ulof filtered sample was carefully layered on top of the iodixanol
487 gradient using a 23G angled needle attached to a 1 mL insulin syringe. The samples were
488 centrifuged at 10,000g for 30 min in a swinging bucket ultracentrifuge at 4C with the
489 acceleration set to slow and the brake set to slow. When the spin was finished, the majority of
490 the layer above the interface was carefully removed using a 23G angled needle attached to a 1
491 mL insulin syringe. Approximately 70 uL of nuclei were collected at the 30%-40% iodixanol
492 interface using a 14G blunt end needle attached to a 1 mL insulin syringe and put in a LoBind
493 eppendorf tube. To check the concentration of nuclei, 10 uL trypan blue was added to 10 uL
494 nuclei and nuclei were examined using a brightfield microscope at 10X magnification. Samples
495 were diluted to 320-2020 nuclei/uL in Complete Buffer HB before loading onto 10X Genomics
496 Chromium controller.

497 Sequencing

498 Samples were sequenced using the Chromium next GEM Single Cell 5' kit v2 from 10X
499 Genomics. Single nuclei suspensions were loaded onto the 10X Genomics Chromium controller
500 at concentrations between 320-2020 nuclei/ul, aiming to capture 5,000 nuclei/sample at a read
501 depth of 40,000 reads/cell. Single nuclei RNA-seq (snRNA-seq) libraries were prepared
502 according to the manufacturer's instructions. Library quality was assessed using the Agilent
503 High Sensitivity D5000 ScreenTape System and were subsequently sequenced using paired-
504 end sequencing on an Illumina NovaSeq6000.

505 Analysis

506 Sequence alignment and UMI counting:

507 Raw reads were aligned and counted using the Cellranger v3.1.0 analysis pipeline. Briefly, a
508 Cellranger-compatible genome was generated via the mkref function using the published prairie
509 vole genome (MicOch1.0). Cellranger mkfastq was then used to create FASTQ files from the
510 raw BCL files. Across samples, we had an average percent read alignment of 80.3%. Cellranger
511 count was then used to align, filter, and count the reads. The mean number of nuclei per sample
512 was 4,862 and the mean read depth was 37,745 reads/nucleus. Samples with a mean less than
513 20,000 reads/nucleus were excluded from downstream analysis (1/40 samples).

514 Sample integration and clustering

515 We used Seurat v4.3.0 in R v4.2.2 to analyze our single nuclei data. First, the filtered feature
516 barcode matrices were used to create Seurat objects with sample-associated metadata. For
517 each sample, cells were filtered to have >200 genes detected per cell and <5% reads from
518 mitochondrial genes. The 39 samples were then transformed using Seurat's SCTransform
519 function and integrated using Seurat's SCT Integration. This left 178,020 nuclei. Following
520 integration, potential effects of the sequencing cohort and the behavior cohort were regressed

521 out. Cells were re-clustered, and clusters were removed that expressed *Slc17a7* (excitatory
522 neurons) or *Pecam1* (endothelial cells). These are two molecular markers consistent with cell
523 types adjacent to, but not found within, the NAc[41]. Cells were clustered again, and one
524 remaining cluster was removed that expressed *Trh* (hypothalamus). Finally, this yielded 142,488
525 nuclei included in the final clustering analysis.

526 Using the SCT-normalized counts, dimensionality reduction was performed using PCA. The top
527 50 PCs were used for UMAP clustering to generate the final cell clusters. 15 clusters were
528 identified and their biological identities were determined by the expression of known marker
529 genes for cell types in the NAc.

530 We then calculated the percentage of cells in each cluster for each animal. To compare groups,
531 we created generalized linear mixed models with sex and pairing type as fixed effects using the
532 glmmTMB package (v1.1.8) in R (v4.2.2). For post-hoc comparisons, we used the emmeans
533 package (v1.8.5) in R with $p < 0.05$ set as our significance threshold for all experiments.

534 Hotspot analysis

535 We used the Hotspot (v1.1.1) analysis pipeline[23] in Python (v3.11.3) to group genes into gene
536 modules based cross-correlations in gene expression across cells. First, we converted our
537 Seurat object to a SingleCellExperiment object. Next, we created the Hotspot object using the
538 Seurat SCT counts matrix, the PCA previously calculated by Seurat, and a negative binomial
539 model. We then computed the neighborhood using an unweighted graph and 30 nearest
540 neighbors. We computed the autocorrelations for each gene to determine the top 3000 most
541 variable genes, and then computed pairwise local correlations between these genes. Finally, we
542 created our modules using these local correlations, with a minimum module size of 30 genes.
543 This left us with 23 modules ranging in size from 31-415 genes. We then calculated module
544 scores for each module in each cell to determine module enrichment across cell populations.

545 Downstream of the module creation, we calculated the average (mean) gene expression for all
546 genes in each module across all cells for each animal. To determine group differences, we
547 generated generalized linear mixed models using glmmTMB (v1.1.8) with sex and pairing type
548 as fixed effects and did post-hoc testing using emmeans (v1.8.5) as stated above.

549 To determine whether partners were more similar than non-partners, we began by calculating a
550 Spearman correlation for the average module gene expression between partners. We then used
551 the SCT counts for the genes in each module to find the distance between partners in Euclidean
552 space. As a control, we found the distance between all non-true pairs (chance) and compared
553 the true partner distances to that distribution. Finally, for each module, we ranked the animals
554 according to their average module gene expression and found the rank-based distance between
555 partners. With 38 paired animals, the expected rank distance between any two animals by
556 chance is 13. We performed one-sample t-tests for each module against a null value of 13.

557 To determine the roles of the genes in each module, we performed gene ontology (GO) analysis
558 using the gprofiler2 package (v0.2.1) in R. The prairie vole gene names were mapped onto the
559 ENSG namespace prior to running GO analysis via gprofiler2. All GO p-values were FDR
560 corrected.

561 Support vector machine analysis

562 We used the e1071 package (v1.7-13) in R (v4.2.2) to analyze our data using support vector
563 machines (SVMs). For each cluster, we downsampled our cells to 200 cells per animal if there
564 were >200 cells for that animal in that cluster. For our SVMs that were able to classify the

animal as itself, we split the cells into a training set and a test set, where 75% of the cells were randomly allocated to the training set and the remaining 25% to the test set. We then trained the SVM for each cluster to classify cells based on animal identity using the scaled Pearson residuals of the gene expression of the top 3,000 variable genes as determined by Seurat. For each animal, we found the percentage of cells from the test set that were classified as the animal from which they originated (self), the partner of the animal from which they originated (partner), or any other animal (other). We then compared the distributions of “self”, “partner”, and “other” classifications.

To train our SVMs that were not provided with the option to classify as self, we downsampled our cells as mentioned above, and then trained a separate SVM for each animal for each cluster. For each SVM, we held a single animal out of the training set, and then tested the SVM on only the held-out animal. This forced the SVM to classify each cell as an animal other than the animal from which it originated. We then calculated the percentage of cells in the test set classified as the partner or as other and compared these groups using generalized linear mixed models with estimated marginal means as post-hoc tests.

Correlated genes

To find which genes were correlated between partners, for each cluster, we subset only genes with > 1 average count per cell. Then, we ran Spearman correlations between partners on each of these genes. We bootstrapped our correlations by comparing them to a distribution of 1000 iterations of within-cohort shuffled pairings. Our correlated genes were defined as having a Spearman p-value of < 0.05 and for Rho to be > 2 standard deviations from the mean of 1000 pair-shuffled correlations. We then used gprofiler2 as stated above to generate GO terms for these gene lists.

References

1. Kingsbury, L. & Hong, W. A Multi-Brain Framework for Social Interaction. *Trends Neurosci.* (2020) doi:10.1016/j.tins.2020.06.008.
2. Pan, Y., Cheng, X., Zhang, Z., Li, X. & Hu, Y. Cooperation in lovers: An fNIRS-based hyperscanning study. *Hum. Brain Mapp.* **38**, 831–841 (2016).
3. Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y. & Feldman, R. Brain-to-Brain Synchrony during Naturalistic Social Interactions. *Sci. Rep.* **7**, 17060 (2017).
4. Czeszumski, A. *et al.* Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Front. Hum. Neurosci.* **14**, (2020).
5. Djalovski, A., Dumas, G., Kinreich, S. & Feldman, R. Human attachments shape interbrain synchrony toward efficient performance of social goals. *NeuroImage* **226**, 117600 (2021).
6. Hu, Y. *et al.* Inter-brain synchrony and cooperation context in interactive decision making. *Biol. Psychol.* **133**, 54–62 (2018).
7. Gugnowska, K. *et al.* Endogenous sources of interbrain synchrony in duetting pianists. *Cereb. Cortex* **32**, 4110–4127 (2022).
8. Vu, T.-D. *et al.* Behavioral and brain- transcriptomic synchronization between the two opponents of a fighting pair of the fish Betta splendens. *PLOS Genet.* **16**, e1008831 (2020).
9. Nummenmaa, L. *et al.* Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci.* **109**, 9599–9604 (2012).
10. Sachs, M. E., Habibi, A., Damasio, A. & Kaplan, J. T. Dynamic intersubject neural synchronization reflects affective responses to sad music. *NeuroImage* **218**, 116512 (2020).
11. Goodwin, N. L., Lopez, S. A., Lee, N. S. & Beery, A. K. Comparative role of reward in long-term peer and mate relationships in voles. *Horm. Behav.* **111**, 70–77 (2019).
12. Lee, N. S., Goodwin, N. L., Freitas, K. E. & Beery, A. K. Affiliation, Aggression, and Selectivity of Peer Relationships in Meadow and Prairie Voles. *Front. Behav. Neurosci.* **13**, (2019).
13. Salgado, S. & Kaplitt, M. G. The Nucleus Accumbens: A Comprehensive Review. *Stereotact. Funct. Neurosurg.* **93**, 75–93 (2015).
14. Sadino, J. M. *et al.* Prolonged partner separation erodes nucleus accumbens transcriptional signatures of pair bonding in male prairie voles. *eLife* **12**, e80517 (2023).
15. Duclot, F., Sailer, L., Koutakis, P., Wang, Z. & Kabbaj, M. Transcriptomic Regulations Underlying Pair-bond Formation and Maintenance in the Socially Monogamous Male and Female Prairie Vole. *Biol. Psychiatry* **91**, 141–151 (2022).
16. Aragona, B. J. & Wang, Z. The Prairie Vole (Microtus ochrogaster): An Animal Model for Behavioral Neuroendocrine Research on Pair Bonding. *ILAR J.* **45**, 35–45 (2004).
17. Walum, H. & Young, L. J. The neural mechanisms and circuitry of the pair bond. *Nat. Rev. Neurosci.* **19**, 643–654 (2018).
18. Avey, D. *et al.* Single-Cell RNA-Seq Uncovers a Robust Transcriptional Response to Morphine by Glia. *Cell Rep.* **24**, 3619-3629.e4 (2018).
19. Savell, K. E. *et al.* A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci. Adv.* **6**, eaba4221 (2020).
20. Tran, M. N. *et al.* Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. *Neuron* (2021) doi:10.1016/j.neuron.2021.09.001.
21. Aragona, B. J. *et al.* Nucleus accumbens dopamine differentially mediates the formation and maintenance of monogamous pair bonds. *Nat. Neurosci.* **9**, 133–139 (2006).
22. Resendez, S. L. *et al.* Dopamine and opioid systems interact within the nucleus accumbens to maintain monogamous pair bonds. *eLife* **5**, e15325 (2016).
23. DeTomaso, D. & Yosef, N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446-456.e9 (2021).

24. Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428-4447.e28 (2022).
25. García-González, D. *et al.* Neurogenesis of medium spiny neurons in the nucleus accumbens continues into adulthood and is enhanced by pathological pain. *Mol. Psychiatry* **26**, 4616–4632 (2021).
26. Dayer, A. G., Cleaver, K. M., Abouantoun, T. & Cameron, H. A. New GABAergic interneurons in the adult neocortex and striatum are generated from different precursors. *J. Cell Biol.* **168**, 415–427 (2005).
27. Fowler, C. D., Liu, Y., Ouimet, C. & Wang, Z. The effects of social environment on adult neurogenesis in the female prairie vole. *J. Neurobiol.* **51**, 115–128 (2002).
28. McGuirt, A. F., Post, M. R., Pigulevskiy, I., Sulzer, D. & Lieberman, O. J. Coordinated Postnatal Maturation of Striatal Cholinergic Interneurons and Dopamine Release Dynamics in Mice. *J. Neurosci. Off. J. Soc. Neurosci.* **41**, 3597–3609 (2021).
29. Pierce, A. F. *et al.* Nucleus accumbens dopamine release reflects the selective nature of pair bonds. *Curr. Biol. CB* S0960-9822(23)01741–4 (2024) doi:10.1016/j.cub.2023.12.041.
30. Wang, Z. *et al.* Dopamine D2 receptor-mediated regulation of partner preferences in female prairie voles (Microtus ochrogaster): A mechanism for pair bonding? *Behav. Neurosci.* **113**, 602–611 (1999).
31. Young, R. L. *et al.* Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci.* **116**, 1331–1336 (2019).
32. Simons Foundation for Autism Research Initiative. SFARI Gene 2023 Q4 release.
33. Dill-McFarland, K. A. *et al.* Close social relationships correlate with human gut microbiota composition. *Sci. Rep.* **9**, 703 (2019).
34. Valles-Colomer, M. *et al.* The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* **614**, 125–135 (2023).
35. Raulo, A. *et al.* Social networks strongly predict the gut microbiota of wild mice. *ISME J.* **15**, 2601–2613 (2021).
36. Ritz, N. L. *et al.* Social anxiety disorder-associated gut microbiota increases social fear. *Proc. Natl. Acad. Sci.* **121**, e2308706120 (2024).
37. Kelly, J. R. *et al.* Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *J. Psychiatr. Res.* **82**, 109–118 (2016).
38. Bercik, P. *et al.* The Intestinal Microbiota Affect Central Levels of Brain-Derived Neurotropic Factor and Behavior in Mice. *Gastroenterology* **141**, 599-609.e3 (2011).
39. Stilling, R. M., Dinan, T. G. & Cryan, J. F. Microbial genes, brain & behaviour – epigenetic regulation of the gut–brain axis. *Genes Brain Behav.* **13**, 69–86 (2014).
40. Ahern, T. H., Modi, M. E., Burkett, J. P. & Young, L. J. Evaluation of two automated metrics for analyzing partner preference tests. *J. Neurosci. Methods* **182**, 180–188 (2009).
41. Reiner, B. C. *et al.* Single nucleus transcriptomic analysis of rat nucleus accumbens reveals cell type-specific patterns of gene expression associated with volitional morphine intake. *Transl. Psychiatry* **12**, 1–11 (2022).