Behavioral Genomics: Towards a molecular characterization of individual variation in mammalian social behavior

Matthew D MacManes[1,2*] Anna Geraghty [1,4] Eileen A Lacey[1,3,4]

1 *University of California, Berkeley. Berkeley, CA 94720*
2 *California Institute of Quantitative Biology*
3 *Museum of Vertebrate Zoology*
4 *Department of Integrative Biology*

∗ Corresponding author: macmanes@gmail.com, Twitter: @PeroMHC

Abstract

Elucidating the genetic mechanisms that underlie complex phenotypes is a central problem in modern evolutionary biology. For behavioral biologists, the ability to link allelic differences or variation in gene expression to the occurrence of specific behavioral traits promises to create significant new opportunities to explore the proximate and ultimate bases for variation in animal behavior. While much progress has been made, mostly in social Hymenoptera, on how allelic differences may lead to interspecific differences in social behavior, whether or not genetics underlie individual variation in behavior is unknown. Novel high-throughout sequencing techniques provide a newfound ways in which studies of complex traits, particularly in non-model organisms may proceed. Specifically, the fact that whole transcriptomes can be obtained has allowed us to begin to think about systems and phenotypes where no obvious set of candidate genes exists. Here, we generate a large Illumina dataset consisting of transcriptomes from 10 social and 10 solitary individuals. Testing for differential expression did not reveal significant differences in expression between the two groups. While further study is needed, these results provide tantalizing support for the notion that gene expression may not underlie individual differences in behavior, at least in the tuco-tuco, an established model for the study of social behavior.

Introduction

Elucidating the genetic and genomic mechanisms that underlie complex phenotypes remains a central problem in modern evolutionary biology. In particular, understanding how patterns of allelic variation or gene expression lead to observable differences in phenotypic traits remains a substantial challenge for studies of most organisms. While substantial progress has been made in understanding the genetics of relatively simply phenotypes (**??**), more complex phenotypes have been challenging (but see **??** for examples of research on the genetics of more complex phenotypes).

For biologists interested complex behavioral phenotypes, the ability to link allelic variation or differences in gene expression to the occurrence of specific behavioral traits promises to create significant new opportunities to explore the proximate and ultimate bases for variation in animal behavior (**?**). Novel high-throughout

sequencing techniques provide a newfound ways in which studies of complex traits, particularly in non-model organisms may proceed. Specifically, the fact that whole transcriptomes can be obtained with relative ease has allowed researchers to begin to think about systems and phenotypes where no obvious set of candidate genes exists.

Genetic bases for social behavior—Social interactions represent a fundamental component of the behavior of numerous animal species. Accordingly, studies of the adaptive bases for variation in social relationships have long been a focus for behavioral biologists (**?????**)). While research into the genetic underpinnings of social behavior has proceeded more slowly, studies of model laboratory organisms (e.g., *Rattus*, *Mus*) have revealed promising relationships between variation in social relationships and specific genes or gene families (**???**). At the same time, studies of less traditional subjects such as prairie voles (*Microtus ochrograster*) are revealing potential links between the formation of pair-bonds and variation at specific loci (**???**). Pair-bonds may represent a fundamentally different type of social bond (e.g. a sexual relationship) than those characteristic of social systems, and therefore the genetic underpinnings of the social behavior of individuals is virtually unstudied.

In addition to research conducted at the individual level, a substantial amount of work has been done, looking at the genetics of the origins of social behavior (**??**). This work, focused mostly on the social Hymenoptera, has provided robust support for the idea that genetic changes may somehow participate in the transition from solitary to social life, though the directionality of causation has not been established firmly. Given the fundamental yet complex role that genes play in shaping phenotypic variation, efforts to understand the genetic underpinnings for social interactions are of considerable general interest to behavioral biologists.

Gene expression and behavior. - Studies of gene expression provide a particularly powerful means of linking patterns of behavioral and genetic variation. By comparing rates of gene transcription in individuals that differ with respect to specific phenotypic traits, such studies serve to delineate the genes and gene pathways that underlie the production of particular behavioral phenotypes (**??**). For example such analyses have been used to link aggressive tendencies in mice to differential expression of G-protein coupled neuropeptide receptors such as GABA (**?**) as well as differential expression of the loci coding for the proteins septin (**?**) and calcineurin (**?**). Relationships between behavioral variation and patterns of gene expression have also been reported for insects (**??**) and birds (**???**), thereby underscoring the general importance of such studies for elucidating the genetic underpinnings of social interactions.

Though studies using a candidate gene approach have been useful in understanding the mechanisms

underlying social behavior (e.g. (**?**)), these studies are fundamentally limited by our current understanding of the genetics of social behavior, which is likely myopic. Alternatively, the use of a microarray to identify loci that are differentially expressed in individuals displaying distinct behavioral phenotypes may be of value. Indeed, microarrays have been employed to examine behaviorally relevant differences in gene expression in model systems like *Drosophila*, uncovering patterns of expression related to mating behavior in post-copulatory females (**??**) and reproductive success in males (**?**). Like the candidate gene approach, however, constructing microarrays requires a priori knowledge of the sequences of either the specific gene under investigation (**?**) or that of a closely related homolog, which limits the suitability of this approach for studies of many non-model organisms.

Transcriptome sequencing and behavior. - The recent development of high throughput sequencing technologies provides a resolution to the challenges posed by the candidate gene and microarray approaches (**??**). For studies of complex behavioral traits – in particular those thought to be mediated by multiple loci or those for which genetic control is completely unknown – high-throughput sequencing of mRNA represents an emerging opportunity to examine either complete or tissue-specific transcriptomes (**??**) that can be used to quantify differences in levels of gene expression (**?**). Importantly, this approach does not require *a priori* knowledge of the genes associated with a given behavioral trait, therefore making it particularly promising for studies of non-model organisms. At the same time, the sensitivity of this approach to differences in gene expression suggests that transcriptome sequencing may be used to detect epistatic interactions of the type expected to underlie many complex behavioral phenotypes.

Given these advantages, we used Illumina sequencing to generate a large mRNA (0.5 billion 100nt and 150nt sequencing reads) dataset from the hippocampi of 20 lab-reared female *Ctenomys sociabilis* that were housed in either social or solitary conditions. Our explicit goals included the identification of genes related to social behavior, and to gain a deeper understanding of the differences in gene expression that may be related to the maintenance of social behavior.

Materials and Methods

Sampling Design. - Whole brains were collected from 20 members of a captive population of colonial tuco-tucos (*Ctenomys sociabilis*) housed on the Berkeley campus. This captive population was founded from 12 free-living individuals captured in Neuquen Province, Argentina, in January, 1996. In captivity, the animals were housed in artificial burrow systems constructed of clear Plexiglas tubes connecting several

Plexiglas boxes that served as nest chambers and latrines (**?**). Typically, the captive population consisted of approximately 45 individuals. Although the social structure and demographic history of *C. sociabilis* (**?**) have resulted in relatively high levels of inbreeding within natural populations of this species, reproductive partners in the captive population were assigned so as to minimize inbreeding within the study subjects. Animal care and use committee approval was sought and obtained prior to the initiation of this work, and is covered under protocol number R224-2011.

Twenty unrelated (i.e., non-littermate) females were used in this study. Analyses focused on yearlings as this is the adult age class that is most abundant in nature and that displays the greatest variation in social relationships, ranging from solitary to social lifestyle (**?**). Captive females used in this study were housed with 0-3 other adult females, thereby imitating naturally occurring intra-population variation in group size for this species (**?**). The social groups in which the test animals were housed were stable for at least one month prior to the collection of brain samples.

Animals were euthanized via overdose with Isoflurane followed immediately by decapitation. The brain was extracted from each individual, after which the hippocampus was dissected out, then placed in a cryotube containing RNAlater (Ambion, Inc), and then flash frozen in liquid nitrogen, and stored at -80C. RNA was later extracted in a dedicated RNAse-free workspace using a TRIzol extraction (Invitrogen). mRNA purification and Illumina sequencing library construction was done using the standard TruSeq RNA kit (Illumina) following the manufacturers recommendation. Each sample was subjected to qPCR using a KAPA kit (Kapa Biosystems, Woburn, MA) for precise quantification. Libraries were then pooled to contain equimolar quantities of each individual library and submitted for Illumina sequencing on a HiSeq 2000 sequencer at The Vincent Coates Genome Sequencing Lab at UC Berkeley.

Assessing sequence quality and pre-assembly procedures. - Accurate assembly of complete transcriptome sequences requires that sequence reads be as error free as possible; random sequencing errors substantially increase the complexity of the *de bruijn* graph, which may result in assembly error (**??**). To address this, we used the open-source software package Trimmomatic (**?**) to identify and to trim nucleotides falling below a given quality threshold (PHRED=5) as well as to remove adapter sequences. To visualize the impact of quality trimming, sequence qualities were assessed pre- and post-quality trimming by the program SolexaQA (**?**). The trimmed quality filtered dataset was then subjected to an error correction pipeline that uses the software package Seecer (**?**), recently shown to significantly improve assembly quality (**?**).

De novo sequence assembly was completed using the program Trinity (**?**), which was run on the Pittsburg Supercomputing Center hardware resource Blacklight (http://www.psc.edu/index.php/computing-resources/

blacklight). This assembly was produced using a solitary individuals (animal ID 406A) whose transcriptome was sequenced with 32M–150nt paired-end Illumina sequencing reads. These data are available under accession SRR488338 at the Short Read Archive. This assembly strategy was chosen, as opposed to concatenating reads from all individuals, based on recommendations from a recent paper suggesting that approximately 30M reads provides a reasonable balance between depth and sequence error (**?**).

The raw assembly was filtered using multiple methods. First, transcript quantitation was accomplished following a re-mapping strategy implemented in the program eXpress (**?**). This procedure employs read mapping produced using the very-sensitive-local and report all alignments settings in Bowtie2 (**?**). Low confidence contigs were defined as having FPKM values less than 1 and were removed from the dataset. Second, after deleting low confidence contigs, we attempted to remove contig redundancy. This redundancy can be a product of sequencing error, polymorphism, or alternative splicing; because alternative splicing may have important phenotypic consequences, only very conservative removal of redundant contigs was attempted. To accomplish this, we used the program cd-hit-est (**?**), allowing clustering only when 99% sequence similarity occurred. Finally, we removed sequences corresponding to ribosomal RNA and mitochondrial DNA from the dataset.

Contig annotation - After *de novo* transcriptome assembly, concatenation, and removal of contaminant and duplicate contigs, we attempted to identify putative transcripts by using a BlastN (**?**) search against the nt database. We then translated putative transcripts from nucleotides to amino acid, and searched for protein coding genes using the TransDecoder software. We then implemented a BlastX search using the Uniprot database to identify protein coding regions, we well as to determine the completeness of the transcript. Matches were considered significant if the e-value for the sequences compared was $>10^{-10}$. Within the group of significant hits, we chose the best Blast hits based on percent sequence similarity. We attempted to identify signal peptides that may have secretory function with the software package SignalP (**?**), and putative transmembrane regions with the program tmhmm (**?**). Putative transcripts were assigned GO terms using the Blast2GO software package (**?**). GO terms were then clustered into 3 groups: biological process, molecular function, and cellular component.

Differential Expression - Testing for differential expression between the two treatment groups was conducted as follows. Sequencing reads were mapped to contigs contained within the minimally-redundant assembly; the number of reads mapping to each contig was summarized using the program eXpress **?**. A data matrix of count data was supplied to the program edgeR (**?**), in which TMM normalization was completed. Differential expression was calculated within edgeR after estimating dispersion. False discovery

rate was set at 0.05 using the Benjamini–Hochberg procedure.

Real-Time Reverse Transcriptase PCR - As a control for assembly and *in silico* transcript abundance estimation, real-time reverse transcriptase PCR was run on TRIzol-extracted RNA further purified with DNase (DNA-free, Ambion). *Rattus* and *Cavia* primers were designed using NCBI Primer BLAST software, which verifies specificity. Tuco-specific primers were designed using partial tuco sequences from Illumina Sequencing results described above. The Ct values were determined using PCR miner (**?**) and normalized to the reference gene, RPLP. For all studies, two-step PCR was used, following the manufacturers instructions for iScript cDNA synthesis kit (BioRad) and then the manufacturers instructions for SsoAdvanced SYBR supermix (BioRad). Samples were run in a BioRad CFX96 real-time PCR system. After the PCR was complete, specificity of each primer pair was confirmed using melt curve analysis, and all samples run on a 2% ethidium bromide agarose gel with a 50bp DNA ladder (Invitrogen) to verify correct product size. Primer sequences are available in Table 1.

Results

In total, 20 individual Illumina sequencing libraries were prepared from the 20 yearling female individuals. Sequencing resulted in over over 500 million 100nt and150nt sequencing reads, distributed over the 20 individuals as per Table 1. Sequence read data are available in the SRA. Adapter removal and quality trimming using Trimmomatic resulted in the removal of 3,410,717 reads. AllPathsLG error correction procedures removed an additional 2,491,765 sequence reads, and corrected more that 100 million putative nucleotide errors. Figure 1 illustrate patters of sequence quality prior to Trimmomatic trimming (Figure 1A) as well as post-trimming (Figure 1B), using the sequence reads used for the *de novo* assembly.

The *de novo* assembly was constructed from 32,541,324–150nt paired-end reads that had been quality trimmed and error corrected. The raw assembly contained 98,239 contigs >200nt in length (N50=2,495). Contigs that were not well-supported by sequence reads, as evidenced by having a FPKM <1, were removed from the dataset, as were contigs corresponding to ribosomal RNA and mitochondrial DNA. After this filtration step, 46,797 contigs remained. CD-HIT-EST was then used to cluster contigs that were >99% similar, which resulted in the removal of 370 contigs. After these steps, 46,418 putative transcripts (N50=2,463) remained. The size of the hippocampal transcriptome is estimated to be 53,960,000 nt in size.

Of the 46,797 contigs, XXX matched a known sequence within the nt database. Approximately 18k putative protein coding regions were discovered, of which YYk were unique. The vast majority of transcripts

were full-length, having both start and stop codons Figure 2. Of these 18k coding regions, 1096 had predicted signal peptides, and 2877 had putative transmembrane regions.

Mean and standard deviation of gene expression were calculated for all each gene, using all 20 samples, we well as after partitioning the samples into social and solitary groupings. The 20 most-highly expressed transcripts, when the mean is calculated using all 20 samples, are listed in Table 2. Five of 20 are predicted to contain a transmembrane region, 4 of 20 have a putative signaling peptide region. Nineteen of the 20 most-highly expressed transcripts appear to be coding sequence– a single sequence comp9226_c0_seq1, does not appear to contain an open-reading frame, and may represent a lincRNA or other non-coding transcript.

Discussion

Elucidating the genetic and genomic mechanisms that underlie complex phenotypes remains a central problem in modern evolutionary biology. In particular, understanding how patterns of allelic variation or gene expression lead to observable differences in phenotypic traits remains a substantial challenge for studies of most organisms. Although substantial progress has been made in understanding the genetics of the evolution of social behavior (**?**), understanding how gene underlie individual-level differences in complex phenotypes– including behavior, remains an outstanding question. This study attempts to identify the genetic correlates of social behavior in a mammalian system characterized by a flexible social system. After careful analysis, we were unable to identify significant differences in gene expression. Whether any of the non-significantly different expression profiles are biologically relevant is an unanswered question.

These results are intriguing– indeed, ours is one of the 1^{st} to attempt to relate differences in gene expression to differences in complex phenotypes like behavior. Indeed, though we predict that differences exist, it is obviously possible that behavior and other complex phenotypes are fine tuned via a different mechanism.

In addition to there being a potential biological explanation for our finding, it is possible that we lack statistical power to detect true differences. This seems unlikely, as we were careful to standardize many aspects of animal care, and handled samples in a consistent fashion. We included 10 biological replicates per group, which is well above the typical number of replicates for mRNAseq studies.

The brain is not a homogeneous organ– we now know that discrete regions of the brain are functionally in-

dependant. Previous work has suggested that the hippocampus is an important center of behavioral control (**???**) and therefore we feel confident that had there been changes, this brain area would have been involved. Future studies looking at different brain regions is in progress.

Lastly, it could be that mammalian social behavior is hardwired in specific individuals, or is formed in response to stimuli experienced earlier during development. These differences may be observable only during specific developmental stages. We are currently in the process of collecting samples from different developmental stages to gain an ontogenetic perspective on gene expression in the hippocampus.

Acknowledgments

Figures & Tables

Figure 1

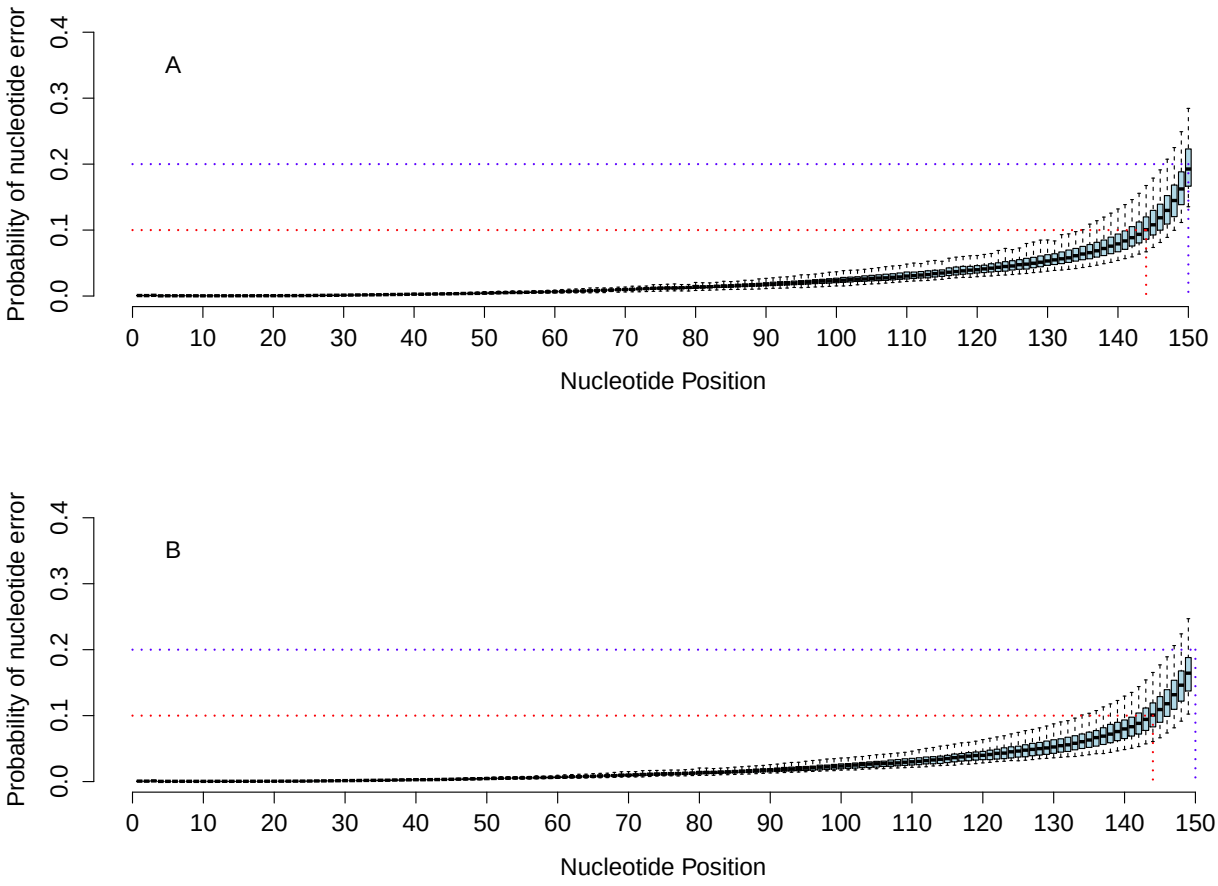

Fig. 1. A. The probability of error increases from the 5' to 3' end of the sequencing read. This error pattern is typical of Illumina sequencing. The red line indicates the 10% error threshold, which is crossed at the 144^{th} nucleotide. The blue line indicates the 20% error threshold, which is crossed at the 149^{th} nucleotide. B. Read error profile after trimming. Note that a very gentle trimming has been conducted.

Figure 2

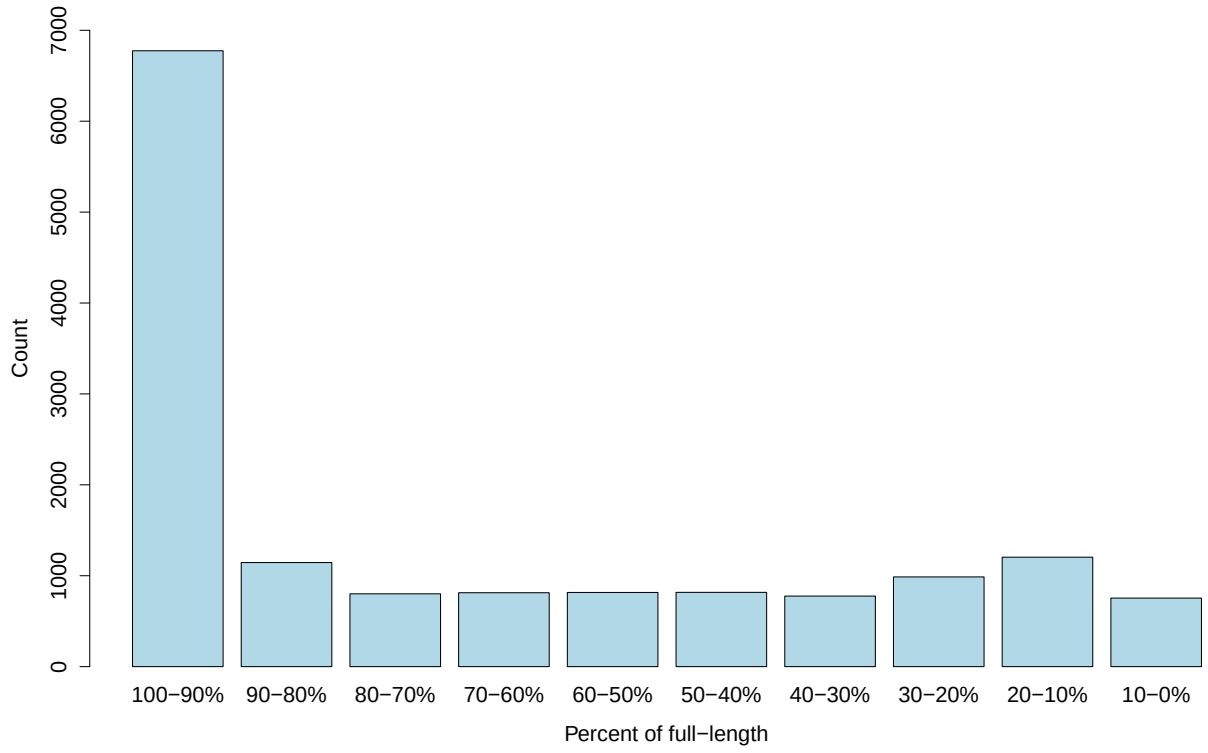

Fig. 2 The vast majority of contigs are full-length transcripts.

Figure 3

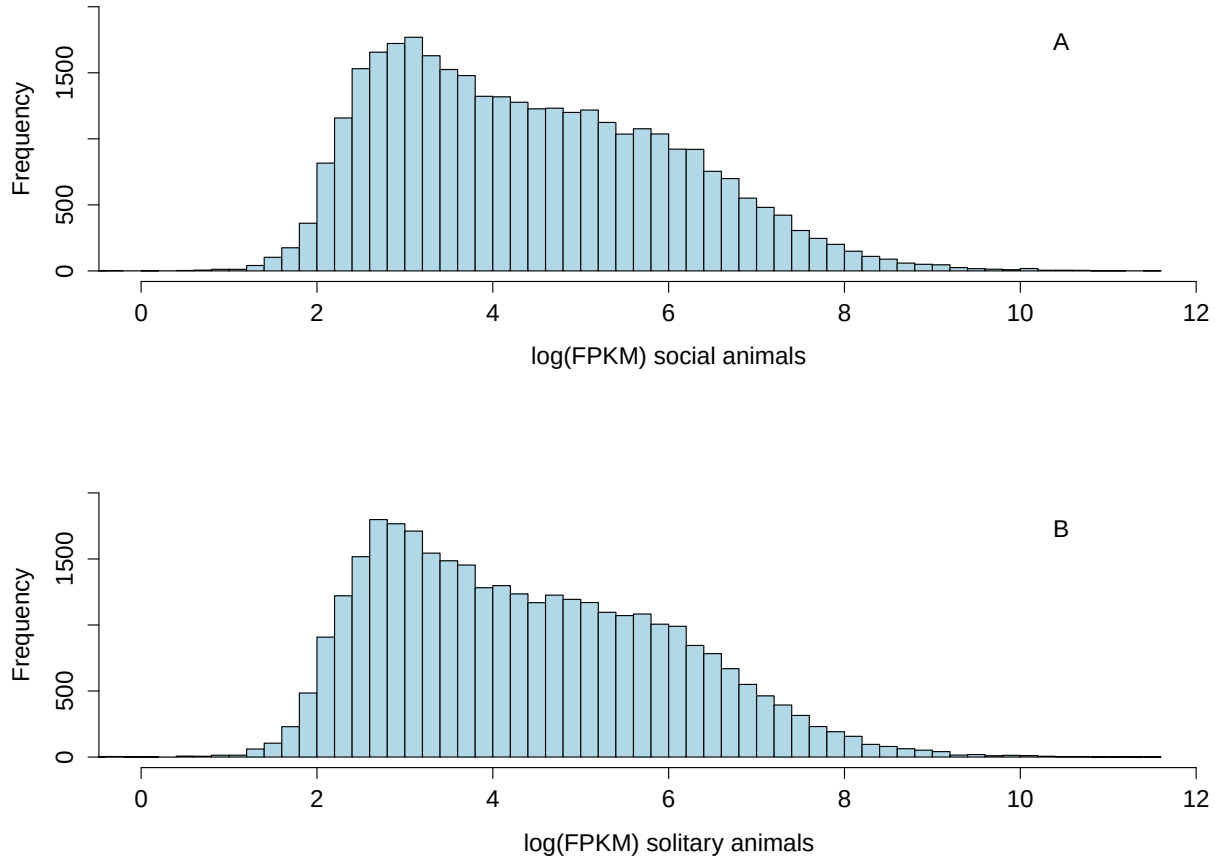

Fig. 3 Patterns of expression are similar between social and solitary species.

Table 1

Gene	Organism	Forward (F) or Reverse (R)	Sequence 5' → 3'
POMC	Rat	F	CTCACCACGGAAAGCAACCT
		R	TTCAGTCAAGGGCTGTTC
CRHBP	Rat	F	GCAGAGGAGCAGCCCTACCG
		R	CCCACCCTGGCAGTCCATGG
CRH	Rat	F	GGAGCCGCCCATCTC TCT
		R	TCCTGTTGCTGTGAGCTTGCT
CRHR1	Rat	F	TGCAGGCCAGAAACATTGC
		R	TCCACCTCCCTTCAGGATCA
CRHR2	Rat	F	CTGGAACCTCATCACCACCT
		R	AGGTAGCAGCCTTCCACAAA
CRHR2	Tuco	F	CGCCTGGGCCATTGGCAAAC
		R	TCCACCAGGTCGCCAGGCTC
MR	Rat	F	GTGCGGCTGCAGCTGACCTT
		R	AGGCCACAGTTCCACGCCAC
GR alpha	Rat	F	ACAGACTTTCGGCTTCTGGA
		R	CTGAAGATGCATCCGAGTGA
GluR1	Rat	F	CAGATCGATATTGTGAACATCA
		R	CCTGAAAGAGCATCTGGTAT
NMDAR1	Rat	F	CTCATCTCTAGCCAGGTCTA
		R	TCGCATCATCTCAAACCAGAC
RPLP1	Rat	F	ATCTACTCCGCCCTCATCCT
		R	GCAGATGAGGCTTCCAATGT

Table 1. Primer sequences for quantitative PCR experiments.

Table 2

Name	DatabaseID	Name	FPKM	TMM	Signal Peptide
comp13664_c0_seq1	P02686	Myelin basic protein	105199.7 ± 65429	0	NO
comp5136_c1_seq1	P02767	Transthyretin	69252.1 ± 114867	0	YES
comp5136_c0_seq1	HQ326642	Calmodulin	63590.6 ± 26236	0	NO
comp9206_c0_seq1	P26779	Saposin-A	49256.7 ± 37465	0	YES
comp9199_c0_seq1	P15087	Carboxypeptidase E	48264.4 ± 16964	0	NO
comp14077_c0_seq1	Q14515	SPARC-like protein 1	47868.9 ± 19865	0	YES
comp14063_c0_seq1	P68105	Elongation factor1-alpha 1	45470.5 ± 24411	0	NO
comp9202_c0_seq1	Q71U34	Heat shock cognate 71 kDa protein	37875.7 ± 16183	0	NO
comp14108_c0_seq1	Q4R4P1	Heat shock protein HSP 90-alpha	37150.7 ± 13513	0	NO
comp14104_c0_seq1	Q8TB61	Adenosine3'-phospho 5'-phosphosulfate transporter 1	33338.4 ± 11724	9	NO
comp13104_c0_seq1	P62972	Polyubiquitin	32834.2 ± 12678	1	NO
comp14076_c0_seq1	P04075	Fructose-bisphosphate aldolase A	31671.2 ± 13672	0	NO
comp5152_c0_seq2	P60203	Myelin proteolipid protein	30604.3 ± 16900	4	NO
comp11953_c0_seq1	P07214	SPARC	30594.7 ± 15972	0	NO
comp9226_c0_seq1		? lincRNA	30082.1 ± 9541	0	NO
comp14070_c0_seq1	Q4KYY3	Glyceraldehyde-3-phosphate dehydrogenase	29708.9 ± 13529	0	NO
comp13888_c0_seq1	P13637	Sodium/potassium-transporting ATPase subunit alpha-3	25332.9 ± 14498	8	NO
comp14068_c0_seq1	P20065	Thymosin-beta-4	25274.4 ± 11164	0	NO
comp12851_c0_seq3		NDRG4	24781.7 ± 8885	0	YES
comp9194_c0_seq2	Q812E9	Neuronal membrane glycoprotein	24707.6 ± 10530	4	NO

Table 2. The twenty most highly expressed genes in the *C. sociabilis* hippocampus. Name refers to the contig ID, while DatabaseID indicates the UniProt transcript identity. FPKM is the estimate of gene expression, averages across the 20 individuals. TMM=transmembrane regions, with the number indicating

the of putative transmembrane domains. Signal peptide denotes whether of not a significant signal peptide is discovered, which is often indicative of secretory function.

Table 3

contig name	**gene name**	**soc (FPKM± SD)**	**sol (FPKM ± SD)**
comp10487_c0_seq1	BDNF	256 ± 76	228 ± 70
comp11013_c0_seq1	Neuroligan-3	242 ± 115	197 ± 109
comp12135_c1_seq2	NEC-1	977 ± 409	857 ± 368
comp12180_c0_seq2	Neuronal PAS3	80 ± 56	72 ± 60
comp12249_c0_seq1	NMDA R1	1175 ± 809	1087 ± 685
comp12444_c0_seq1	5-HT2C	673 ± 925	1312 ± 2144
comp12616_c0_seq2	Histone Acetyltransferase p300	180 ± 95	101 ± 67
comp12659_c0_seq2	PSD-95	6851 ± 4701	5835 ± 2453
comp12745_c0_seq2	Peroxin-5	247 ± 106	278 ± 135
comp12920_c1_seq2	Ataxin-3	104 ± 44	105 ± 46
comp13367_c0_seq1	Guanine Nucleotide Binding Protein	99 ± 38	95 ± 28
comp13702_c0_seq12	Abelson Interactor 2	269 ± 157	285 ± 143
comp14536_c0_seq1	MAPKK1	6096 ± 3119	4948 ± 1532
comp14614_c0_seq1	Shank1	3387 ± 1423	3100 ± 902
comp15294_c0_seq1	MCAD	1044 ± 450	1324 ± 1068
comp16411_c0_seq1	POMC	48 ± 111	54 ± 124
comp18034_c0_seq1	PTEN	2411 ± 861	2454 ± 943
comp19765_c0_seq1	ProSAP2	239 ± 174	226 ± 90
comp20237_c0_seq1	CLN8	599 ± 331	527 ± 218
comp21149_c0_seq1	MeCp-2	177 ± 66	180 ± 73
comp21157_c0_seq1	Caveolin-1	195 ± 96	202 ± 104
comp22526_c0_seq1	GR	483 ± 279	535 ± 325
comp24215_c0_seq1	Retinoic Acid receptor	31 ± 21	33 ± 14
comp25793_c0_seq1	CRH-BP	136 ± 31	139 ± 50
comp28662_c0_seq1	Neuronal Acetylcholine Receptor B2	732 ± 419	581 ± 215
comp32945_c0_seq1	NMDA R2A	181 ± 160	172 ± 161
comp35997_c0_seq1	Tenascin-R	687 ± 539	604 ± 494
comp3934_c0_seq2	Chorein	1217 ± 410	1187 ± 417
comp4697_c1_seq1	Nuclear Receptor TLX	151 ± 44	141 ± 32
comp49052_c0_seq1	Dopamine D2 Receptor	89 ± 52	76 ± 48
comp49278_c0_seq1	NMDA R2B	85 ± 46	70 ± 46
comp51753_c0_seq1	Caspr4	302 ± 113	248 ± 100
comp5378_c0_seq1	HGPRT	149 ± 57	160 ± 72
comp5525_c0_seq2	Dishevelled-1	295 ± 79	284 ± 72
comp5533_c0_seq1	JIP-2	2631 ± 1359	2298 ± 894
comp5545_c0_seq1	Hungtingtin	2935 ± 1854	2472 ± 1191
comp5551_c0_seq1	nPKC-epsilon	7542 ± 3889	6562 ± 1839
comp6018_c1_seq1	MR	1483 ± 648	1429 ± 538
comp6176_c0_seq1	Neuroligan-2	2668 ± 1527	2167 ± 816

Continued on next page

Table 1 – *Continued from previous page*

contig name	**gene name**	**soc (FPKM± SD)**	**sol (FPKM ± SD)**
comp6729_c0_seq1	Dardarin	165 ± 103	125 ± 77
comp7178_c0_seq1	CRF-R1	212 ± 129	180 ± 90
comp72570_c0_seq1	IL-1	49 ± 33	48 ± 30
comp8432_c1_seq2	Annexin A3	213 ± 163	279 ± 267
comp9154_c0_seq2	Junctophillin-3	3684 ± 1533	3139 ± 849
comp9338_c0_seq1	Somatostatin Receptor	94 ± 39	83 ± 37

232

233 Table 3. A comparison of the mean gene expression levels (± stand. dev.) for solitary and social animals in
234 45 transcripts previously implicated in social behavior. *Eileen– I still need to add the qPCR data*