



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Adolescent rat social play: Amygdalar proteomic and transcriptomic data



Navatha Alugubelly^{a, b}, Afzaal N. Mohammed^{a, b},
 Mariola J. Edelmann^c, Bindu Nanduri^b, Mohammed Sayed^d,
 Juw Won Park^{d, e, **}, Russell L. Carr^{a, b, *}

^a Center for Environmental Health Sciences, USA^b Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, MS, USA^c Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA^d Department of Computer Engineering and Computer Science, USA^e KBRIN Bioinformatics Core, University of Louisville, KY, USA

ARTICLE INFO

Article history:

Received 5 September 2019

Accepted 24 September 2019

Available online 30 September 2019

Keywords:

Social play

Proteome

Transcriptome

Behaviour

Amygdala

ABSTRACT

This data article contains the proteomic and transcriptomic data of the amygdala of adolescent rats involved in social play compared to non-behavioural animals. Social play was performed on male Sprague Dawley rats on postnatal day 38 and protein and gene expression in the amygdala was determined following behavioural testing. The protein expression was measured by analysing trypsin digested protein samples using a LTQ Orbitrap Velos mass spectrometer equipped with an Advion nanomate ESI source. The obtained tandem mass spectra were extracted by Thermo Proteome Discoverer 1.3 and the data were displayed with Scaffold v 4.5.1. The transcriptomic data were generated by Illumina HiSeq 4000 system. Cuffdiff (v2.2.1) program was used to calculate RNA-seq based gene expression levels. For further interpretation of data presented in this article, please see the research article 'Proteomic and Transcriptional Profiling of Rat Amygdala Following Social Play' (Alugubelly et al. 2019).

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

DOI of original article: <https://doi.org/10.1016/j.dib.2019.104589>.

* Corresponding author. Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Box 6100, MS 39762-6100, USA.

** Corresponding author. Department of Computer Engineering and Computer Science, KBRIN Bioinformatics Core, 213 Duthie Center, University of Louisville, Louisville, KY 40292, USA.

E-mail addresses: juw.park@louisville.edu (J.W. Park), rlcarr@cvm.msstate.edu (R.L. Carr).

<https://doi.org/10.1016/j.dib.2019.104589>

2352-3409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications Table

Subject	Neuroscience
Specific subject area	Behavioural Neuroscience
Type of data	Tables
How data were acquired	RNA-sequencing using an Illumina HiSeq 4000 system. Proteomic analysis using LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an Advion nanomate ESI source (Advion, Ithaca, NY).
Data format	Analysed
Parameters of data collection	The data were collected from both behavioural and non-behavioural group of rats. Social play was performed by keeping two rats in a cage and allowed them to interact for 10 minutes. In behavioural pair, one rat was sacrificed for transcriptomic analysis at 15 min following social play and the other rat was sacrificed for proteomic analysis at 3 hours following social play. The non-behavioural rats were also sacrificed on the same day. After sacrifice, whole brains were collected from each rat and amygdala was isolated. The collected amygdala tissue was used for protein and gene expression analysis.
Experimental features	The protein analysis was done by using Thermo Proteome Discoverer 1.3 and Scaffold v 4.5.1 software. The RNA-seq based gene expression analysis was performed by using the Illumina HiSeq 4000 system, STAR aligner (v.2.5.2b), and Cuffdiff (v2.2.1) programs.
Data source location	Mississippi State University, Mississippi State, Mississippi, United States of America.
Data accessibility	The RNA-seq data has been deposited into the NCBI Gene Expression Omnibus under the accession number (GSE126023). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012677.
Related research article	Alugubelly, N., Mohammad, A.N., Edelmann, M.J., Nanduri, B., Sayed, M., Park, J.W., and Carr, R.L., 2019. Proteomic and Transcriptional Profiling of Rat Amygdala Following Social Play, Behav. Brain Res.

Value of the Data

- The data provide the insights on proteome and transcriptome of amygdala of adolescent rats participated in social play.
- The data provide a full list of the proteins and genes detected. It also provides additional outcomes of the bioinformatics analysis including (1) all significant GO terms identified for both differentially expressed proteins and genes; (2) all identified canonical pathways, molecular and cellular functions, and physiological functions associated with the expressed differentially expressed proteins and genes; and (3) lists of the differentially expressed genes and proteins associated with each pathway and functions
- The data will be useful for understanding the neurobiology of social play. It also provides a model by which to begin to chart the dynamics of the interactions among the neural circuits involved in play.
- The data can be utilized in future experiments to understand the role of specific proteins and genes involved in social play and to further explore the role of different neurological pathways in social play.

1. Data

In this data article, the transcriptomic and proteomic analysis of social play of adolescent rats has been performed for the first time. The data are directly available in this article and related to our previous publication (Alugubelly et al., 2019). From isolated rat amygdala, the collected trypsin digested peptides were analysed by mass spectrometer. Proteome discoverer was used to identify the differentially expressed proteins. The list of total proteins along with fold change values and p-values were presented in Supplementary File 1. The isolated RNA from amygdala was used for poly A selected RNA-seq library preparation. Cuffdiff (v2.2.1) program was used to calculate RNA-seq based gene expression levels. The list of total genes along with fold change and p-values were listed in Supplementary File 2.

2. Experimental design, materials, and methods

2.1. Animal maintenance

Adult male and female Sprague Dawley rats (CD IGS; Envigo, Indianapolis, IN) were used for breeding. These rats were housed in an Association for Assessment and Accreditation of Laboratory Animal Care–accredited facility under constant temperature (22 °C), on a 12 h light and 12 h dark cycle with lights on between 0700 and 1900. Rats were provided with food and water freely available during experimentation. The procedures used in this project were approved by the Mississippi State University Institutional Animal Care and Use Committee. Females were separated from males once they appeared to be pregnant. The date of birth was designated as PND0. Male rat pups within each litter were used for behavioural testing. Rats were weaned on PND21 and marking was continued until behavioural testing to allow identification.

2.2. Behavioural testing

The behavioural arena was a clear empty plastic cage with bright light (~700 lux). Each test session was recorded using a remotely operated Canon EOS Rebel digital camera. Testing was performed on PND38. Following a 24-h isolation period, two rats of the same, age, and size but from different litters were placed into different corners of the behavioural arena. The rats remained in the arena together for 600s. After each test, the cage was emptied, cleaned with 70% ethanol, dried, and refilled with fresh litter.

For each behavioural pair, one rat was sacrificed for transcriptomic analysis at 15 min following social play and the other rat was sacrificed for proteomic analysis at 3 hours following social play. The non-behavioural rats were also sacrificed on PND 38 following a 24-h isolation period. At sacrifice, whole brains were collected and stored at –80 °C. An RNAase free environment was maintained throughout the tissue collection process. A total of three rat brains from each group were used for each analysis. Frozen brains were sliced using a manual tissue slicer to obtain 500- μ m sections which were stored on microscopic slides until the amygdala was collected using punches (1mm size). The Paxinos and Watson (1998) atlas was used as a reference. The obtained amygdala tissue was processed for proteomic and transcriptomic analysis.

2.3. Proteomic analyses

2.3.1. Sample preparation

Collected amygdala tissue was lysed in NP-40 lysis buffer (150 mM NaCl, 20 mM MgCl₂, 50 mM Tris-HCl pH 7.4, 0.5% NP-40) supplemented with 1mM of the serine protease inhibitor, phenyl-methylsulfonyl fluoride (PMSF) using a Microson™ ultrasonic cell disruptor. The debris was removed by centrifugation at 21,000g at 4 °C for 30 min. The protein concentration was measured using a Pierce™ BCA protein assay kit (Thermo Scientific). From each sample, 100 μ g of protein was precipitated by chloroform/methanol extraction. Briefly, the sample volume was adjusted to 200 μ l using NP-40 lysis buffer. To each sample, 600 μ l of methanol, 150 μ l of chloroform, and 450 μ l of milliQ-H₂O were added, vortexed, and centrifuged at room temp for 1 min, at 21,000g. The upper aqueous phase was discarded and 450 μ l of methanol was added to the lower phase, vortexed, and centrifuged under the same conditions for 2 min. The supernatant was discarded, and protein digestion was performed by suspending the pellet in 33 μ l of 100 mM Tris-HCl (pH 7.8) containing 6 M urea. The samples were reduced with 1.6 μ l of 200 mM dithiothreitol (DTT) for 45min at room temperature and alkylated with 6.6 μ l of 200 mM iodoacetamide (IAA) for 45 min at room temperature. The alkylation reaction was then quenched by adding 20 μ l of 200 mM DTT for 45 min at room temperature. The urea concentration was reduced by adding 258 μ l of milliQ-H₂O. Finally, the proteins were digested with trypsin (sequencing grade modified trypsin, Promega) at 1:50 ratio for 18 hr at 37 °C. Protein digestion was terminated by lowering the pH of each sample to <6 by adding concentrated acetic acid. The samples were desalted using C18 SepPak columns (Waters, USA). The sample was then dried down in speed vac.

All samples were submitted to the University of Arizona Proteomic Consortium for analysis by tandem mass spectrometry coupled to liquid chromatography (LC-MS/MS).

2.3.2. Mass spectrometry

The LC-MS/MS analysis of trypsin digested protein samples was carried out using a LTQOrbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an Advion nanomate ESI source (Advion, Ithaca, NY). Peptides were eluted from a C18 precolumn (100- μ m id \times 2 cm, Thermo Fisher Scientific) onto an analytical column (75- μ m ID \times 10 cm, C18, Thermo Fisher Scientific) using a beginning concentration of 2% solvent B (acetonitrile, 0.1% formic acid) for 5 minutes, then a 2–7% gradient of solvent B over 5 minutes, followed by a 7–15% gradient of solvent B over 50 minutes, a 15–35% gradient of solvent B over 60 minutes, a 35–40% gradient of solvent B over 28 minutes, a 40–85% gradient of solvent B over 5 minutes, held at solvent 85% B for 10 minutes, 85–2% gradient of solvent B for 1 minute then held at 2% solvent B for 16 min. All flow rates were at 400 nl/min. Solvent A consisted of water and 0.1% formic acid. Data dependent scanning was performed by the Xcalibur v 2.1.0 software using a survey mass scan at 60,000 resolutions in the Orbitrap analyser scanning m/z 400–1600, followed by collision-induced dissociation (CID) tandem mass spectrometry (MS/MS) of the fourteen most intense ions in the linear ion trap analyser. Precursor ions were selected by the monoisotopic precursor selection (MIPS) setting with selection or rejection of ions held to a \pm 10 ppm window. Dynamic exclusion was set to place any selected m/z on an exclusion list for 45 seconds after a single MS/MS.

2.3.3. Data processing and quantitation

The tandem mass spectra were extracted by Thermo Proteome Discoverer 1.3 (Thermo Scientific, USA) using the Sequest algorithm (Thermo Fisher Scientific, San Jose, CA, USA; version 1.3.0.339). Sequest was set up to search *Rattus Norvegicus*_UniprotKB assuming the digestion enzyme as trypsin. Fully tryptic peptides with up to 2 missed cleavage sites were selected. While searching with Sequest, fragment ion mass tolerance of 0.8 Da and a parent ion tolerance of 10.0 PPM were used. Oxidation of methionine and carbamidomethyl of cysteine were specified in Sequest as variable modifications. The results were also validated using X!Tandem, another search engine and displayed with Scaffold v 4.5.1 (Proteome Software Inc., Portland OR), a program that relies on various search engine results (i.e.: Sequest, X!Tandem, MASCOT) and uses Bayesian statistics to reliably identify more spectra [1]. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [2]. Proteins that contain similar peptides and cannot be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Label-free protein quantitation using the sum of weighted spectra associated with a protein was performed in Scaffold. The proteins that passed the Fisher's exact test with a p-value of ≤ 0.05 were used for biological interpretation. Differentially expressed proteins (DEP) were identified based on a fold change value, which was calculated by applying normalization in Scaffold. A minimum value of 0.2 was used for the samples in which a protein was not identified. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012677 [3].

2.4. RNA-seq and transcriptomic analysis

Total RNA was isolated from amygdala using a Qiagen miRNA easy micro kit (Germantown, MD) which is specialized for isolation of RNA from lipid rich tissue. The RNA quality was verified by NanoDrop. Each RNA sample containing 1 μ g of RNA was used for poly A selected RNA-seq library preparation using the NEB Ultra-Directional RNA Library Prep Kit for Illumina, E7420. Stranded paired-end sequencing data with read lengths of 100 bp (2 \times 100 bp) were generated by Institute for Genomics, Biocomputing & Biotechnology core using the Illumina HiSeq 4000 system. An average of 52.8 million read pairs per replicate were generated. The RNA-seq data has been deposited into the NCBI Gene Expression Omnibus under the accession number (GSE126023). RNA-seq reads were mapped to

the rat reference genome (rn6) and transcriptome (Ensembl, release 84) by STAR aligner (v.2.5.2b) [4] allowing up to 3 mismatches per read. Cuffdiff (v2.2.1) [5] program was used to calculate RNA-seq based gene expression levels using the FPKM (fragments per kilobase of exon per million fragments mapped) and then differentially expressed genes (DEG) were identified between two conditions at $FDR \leq 5\%$, and greater than two-fold difference in average FPKM.

2.5. Bioinformatics

The differentially expressed genes and proteins were first subjected to gene ontology (GO) enrichment using the Database for Annotation, Visualization and Integrated Discovery (DAVID) to determine the cellular locations and biological processes that were enriched. Enrichment of genes/proteins was designated when GO terms possessed a p-value < 0.05 and $FDR < 0.05$ and statistical significance was determined by the EASE Score Threshold. For proteins, the list of significant GO terms for cellular locations and biological processes are presented in Supplementary Files 3 and 4, respectively. For genes, the list of significant GO terms for cellular locations and biological processes are presented in Supplementary Files 5 and 6, respectively.

Additional analysis was performed using Ingenuity Pathway Analysis (IPA, QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>). Fisher's exact test was utilized in all those analyses to identify the overrepresented proteins or genes with a p-value of less than 0.05. The software was used to identify significant canonical pathways associated with the differentially expressed proteins/genes. Lists of associated canonical pathways are presented in Supplementary File 7 for proteins and in Supplementary File 8 for genes. In addition, the functional annotation tool in the software was used to classify the genes/proteins based on their molecular and cellular functions and physiological functions. Lists of associated molecular and cellular functions are presented in Supplementary File 9 for proteins and in Supplementary File 10 for genes. Lists of associated physiological functions are presented in Supplementary File 11 for proteins and in Supplementary File 12 for genes.

Acknowledgments

Research was supported by the Mississippi Agricultural and Forestry Experiment Station (MAFES), by the College of Veterinary Medicine, Mississippi State University, and by NIH R15ES023162.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2019.104589>.

References

- [1] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, *Anal. Chem.* 74 (20) (2002) 5383–5392.
- [2] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem.* 75 (17) (2003) 4646–4658.
- [3] Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D.J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A.F. Jarnuczak, T. Ternent, A. Brazma, J.A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data, *Nucleic Acids Res.* 47 (D1) (2019) D442–D450.
- [4] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics (Oxford, England)* 29 (1) (2013) 15–21.
- [5] C. Trapnell, D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, L. Pachter, Differential analysis of gene regulation at transcript resolution with RNA-seq, *Nat. Biotechnol.* 31 (1) (2013) 46–53.