Analysis of hippocampal transcriptomic responses to technical and biological perturbations

Rayna M. Harris[1,2], **Hsin-Yi Kao**[2,3], **Juan Marcos Alarcón**[3], **Hans A. Hofmann**[1,2], **André A. Fenton**[2,3,4]

*For correspondence:
afenton@nyu.edu (AAF)

[1]Department of Integrative Biology, Center for Computational Biology and Bioinformatics, Institute for Cell and Molecular Biology, The University of Texas at Austin; [2]Neural Systems & Behavior Course, Marine Biological Laboratory; [3]Center for Neural Science, New York University; [4]Department of Pathology, SUNY Downstate Medical Center; [5]The Robert F. Furchgott Center for Neural and Behavioral Science, SUNY Downstate Medical Center; [6]Neuroscience Institute at the New York University Langone Medical Center, New York University

Abstract Cost-effective next-generation sequencing has made unbiased gene expression analysis possible. Single-neuron gene expression studies may be especially important for understanding nervous system structure and function because of the neuron-specific functionality and plasticity that defines functional neural circuits. Cellular dissociation is a prerequisite technical manipulation for single-cell and single cell-population studies, but the extent to which the cellular dissociation process cells affects neural gene expression has not been determined, nor has it been determined how gene expression is altered by the stress that accompanies many of the behavioral manipulations that are required to study learning and memory and other cognitive functions. Here, we determined to which extent cellular dissociation-induced changes in hippocampal gene expression might confound studies on the behavioral and physiological functions of the hippocampus. We processed tissue punch samples from the dentate gyrus (DG), CA3, and CA1 hippocampus subfields using either a tissue homogenization protocol or a cellular dissociation protocol in preparation for RNA sequencing analysis to evaluate the impact of the tissue preparation. Then, we evaluated the effect of stressful experience and cognitive training on hippocampus subfield specific gene expression and determined to which extent these response overlap with the cellular dissociation response. Finally, we assessed the extent to which the subfield-specific gene expression patterns are consistent with those identified in a recently published hippocampus subfield-specific gene expression database. We report substantial differences in baseline subfield-specific gene expression, that 1% of the hippocampal transcriptome is altered by the process of cellular dissociation, that an even weaker alteration is detected 24 h after stressful experience, and that while these alterations are largely distinct from the subfield specific response of the hippocampus transcriptome to cognitive training, there is nonetheless some important confounding overlap. These findings of the concordant and discordant effects of technical and behavioral manipulations should inform the design of future neural transcriptome studies and thus facilitate a more

41 comprehensive understanding of hippocampal function.

42 _____

Introduction

44 Nervous systems are comprised of diverse cell types that include neurons, glia, and vascular
45 cells, each serving distinct functions and thus expressing different genes. Consider the hip-
46 pocampus, a structure central for spatial navigation and the processing of event memory in
47 the mammalian brain. To date, distinct aspects of navigation and memory processing have
48 been firmly correlated to activity of particular cellular subfields within the hippocampal for-
49 mation. This subfield-specific understanding of hippocampal function, has led to the notion
50 that cells within a given subfield are homogeneous in their molecular blueprint and perform
51 the same function. However, even within the anatomically-defined subfield of CA1, there are
52 identifiable subclasses of pyramidal cells that belong to distinct functional circuits (*Mizuseki*
53 *et al., 2011*; *Danielson et al., 2016*). This diversity is even greater when we consider that spe-
54 cific cells within a functional class can be selectively altered by neural activity in the recent or
55 distant past. For example, only a third of the pyramidal cells of the superficial CA1 sub-layer
56 are expected to be meaningfully active during experience of a particular environment and
57 only a subset of those might have been sufficiently engaged to alter the strength of a synapse
58 which then triggers further gene expression changes within the functional class(*Guzowski*
59 *et al., 1999, 2006*). All this diversity implies distinctive gene expression, very likely at the level
60 of single neurons, and such considerations may strongly curtail interpretations of gene expres-
61 sion studies that use mixtures of cells or microdissected tissue samples.

62 Fortunately, recent advances in tissue harvesting and processing, as well as in sequencing
63 technologies have allowed detailed analyses of genome-scale gene expression profiles at the
64 level of single cell populations, in the context of brain and behavior studies (*Chalancon et al.,*
65 *2012*; *Harris and Hofmann, 2014*; *Mo et al., 2015*). These approaches have led to systems-level
66 insights into the molecular substrates of neural function, along with the discovery or valida-
67 tion of candidate pathways regulating physiology and behavior (*Cembrowski et al., 2016a*).
68 While the complexity of some tissues complicates the interpretation of transcriptome data
69 collected from samples containing hundreds to tens of thousands of cells representing nu-
70 merous cellular subclasses at different levels of diversity, difficulties with interpretation can be
71 minimized by careful experimental design governing both data collection and data analysis.
72 To complement this effort, and optimize experimental designs, it is necessary to understand
73 the extent to which the treatment of tissue samples prior to transcriptome analysis might
74 confound interpretation of the results.

75 We examined the effect of cellular dissociation on the transcriptomes of specific hippocam-
76 pal subfields (CA1, CA3, and DG) by comparing tissue homogenization (as a control) and cel-
77 lular dissociation protocols. We then examined and compared the effect of prior stressful
78 experience that accompanies many protocols to assess learning, memory and innate behav-
79 iors, and cognitive training on hippocampal subfield gene expression. Finally, we compared
80 these results to a public data set of cell type-specific hippocampus gene expression to further
81 validate the patterns of gene expression that we identified. Knowing how technical pertur-
82 bations influence the ability to detect the molecular signature of differences in neural and
83 behavioral variables is an important step in calibrating the ability to mechanistically under-
84 stand hippocampal function. In addition to understanding the impact of cell dissociation
85 and stressful experience on hippocampus gene expression, the present findings allow evalu-
86 ating the extent to which gene expression profiles of heterogeneous tissue samples compare
87 with single neuron population gene expression profiles.

Methods and Materials

Animals

All animal care and use complies with the Public Health Service Policy on Humane Care and Use of Laboratory Animals and were approved by the New York University Animal Welfare Committee and the Marine Biological Laboratory Institutional Animal Care and Use Committee. C57BL/6J mice were used, a generous gift from the Jackson Laboratory (Bar Harbor, ME). All mice were housed on a 12:12 (light:dark) cycle with continuous access to food and water.

Tissue preparation

Each mouse was taken from its cage, anesthetized with 2% (vol/vol) isoflurane for 2 minutes and decapitated. Transverse 300 ㎛m brain slices were cut using a vibratome (model VT1000 S, Leica Biosystems, Buffalo Grove, IL) and incubated at 36°C for 30 min and then at room temperature for 90 min in oxygenated artificial cerebrospinal fluid (aCSF in mM: 125 NaCl, 2.5 KCl, 1 $MgSO_4$, 2 $CaCl_2$, 25 $NaHCO_3$, 1.25 NaH_2PO_4 and 25 Glucose) (Pavlowsky and Alarcon, 2012; Pavlowsky et al., 2017). Tissue samples were collected from CA1, CA3, and DG, respectively in the dorsal hippocampus by punch (0.25 mm, P/N: 57391; Electron Microscopy Sciences, Hatfield, PA). All punches for RNA sequencing came from the slice corresponding to image 74 of the Allen Brain Reference Atlas (RRID:SCR_013286).

Animal and tissue preparation for assessing impact of cellular dissociation

A 1-year-old female C57BL/6J mouse was used for the cellular dissociation experiment. One tissue punch was designated for the control homogenized processing and the other for the cellular dissociation treatment (Fig 1A). Two adjacent tissue samples were collected from each subfield for each mouse. The 'control sample' was processed using the manufacture instructors for the Maxwell 16 LEV RNA Isolation Kit (Promega, Madison, WI). The 'cellular dissocation sample' was incubated for 75 minutes in aCSF containing 1 mg/ml pronase at room temperature, then vortexed and centrifuged. The incubation was terminated by replacing aCSF containing pronase with aCSF. The sample was then vortexed, centrifuged, and gently triturated by 200-㎕l pipette tip twenty times in aCSF containing 1% FBS. The sample was centrifuged and used as input RNA isolation using the Maxwell 16 LEV RNA Isolation Kit (Promega, Madison, WI).

Animals and tissue preparation for assessing impact of stressful experience and cognitive training

Male C57BL/6J mice that were 3-4–months old were used. They were obtained from the Jackson Laboratory (Bar Harbor, ME) and housed at the Marine Biological Laboratory. Gene expression in tissue from mice taken from the home cage was compared to mice that received mild foot shock, to evaluate how gene expression is affected by stressful experience, which is a common confound of behavioral manipulations such as water maze learning, fear conditioning, inhibitory avoidance, active place avoidance and other learning and memory test paradigms. The mice were placed on an elevated circular 40-cm diameter arena that rotated at 1 rpm. The arena wall was transparent and so contained the mice to the arena and allowed it to observe the environment.

Mice in the stressful experience group received a short series of unavoidable mild foot shocks while walking on the arena. Each shock was a constant current 0.2 mA 500 ms 60 Hz shock. The time series of shocks matched the shock time series from training in an active place avoidance task. Each session was 10 minutes and the mice had three sessions a day for 3 days with an inter-trial interval of 2 hours, during which it was returned to the home cage. The shock protocol is initially stressful, assessed by elevated plasma corticosterone, but by the second day, corticosterone levels return to baseline levels (Lesburguères et al. 2016). The mice received an average of 8 shocks per day with a maximum of 29 and a minimum of 1 shock.

136   Comparisons were made to home cage control mice that were not exposed to the arena.

137   Mice in the cognitive training and yoked control groups were placed on the rotating arena
138   and trained in the active place avoidance task. The task conditions the mice to avoid foot
139   shock that can be localized using extramaze visual cues. Acquiring and remembering the
140   avoidance requires intact hippocampus function and long-term synaptic potentiation (Cimadev-
141   illa et al., 2001; Hsieh et al., 2017; Pavlowsky et al., 2017). Yoked control mice were exposed to
142   the arena the same amount of time as the cognitive training and they received the identical
143   time series of foot shocks as an animal in the cognitive training group. Thus the yoked mice
144   experienced the same environment as the trained mice but for the yoked controls, shock was
145   unavoidable. While the trained and yoked animals received the same number of shocks, only
146   the trained animals exhibited an avoidance response and only they yoked animals exhibited
147   a detectable stress response (Lesburgueres et al., 2016).

148   Twenty-four hours after the behavioral manipulations the mice were sacrificed and tissue
149   punches were collected was described above.

RNA sequencing

151   RNA from CA1, CA3, and DG was isolated using the Maxwell 16 LEV RNA Isolation Kit (Promega,
152   Madison, WI). RNA libraries were prepared by the Genomic Sequencing and Analysis Facility
153   at the University of Texas at Austin using the Illumina HiSeq platform.

154   Raw reads were processed and analyzed on the Stampede Cluster at the Texas Advanced
155   Computing Facility (TACC). Quality of the data was checked using the program FASTQC. Low
156   quality reads and adapter sequences were removed using the program Cutadapt (Martin,
157   2011). We used Kallisto for read pseudoalignment to the Gencode MV11 mouse transcriptome
158   and for transcript counting (Mudge and Harrow, 2015; Bray et al., 2016). Transcript counts were
159   converted into gene counts using the reshape2 R package (**Wickham, 2016**).

160   We downloaded the gene counts from the **Cembrowski et al.** (**2016b**) dataset archived
161   (NCBI GEO:GSE74985) (**Cembrowski et al., 2016b**). Briefly, this data set contains hippocam-
162   pal gene expression data for pools of 112 ± 6 cells for each of 5 cell types (CA1, CA2, and CA3
163   pyramidal neurons and DG mossy and granule cells) from behaviorally naive, transgenic mice
164   that express a fluorescent protein label in the specific cell types. The DG, CA1, and CA3 cells
165   were manually sorted from both dorsal and ventral slices and fluorescently-labeled neurons
166   were manually collected, DG granule cells from Rbp4-Cre KL100, CA3 pyramidal cells from
167   Mpp3-Cre KG118, and CA1 pyramidal cells from Vipr2-Cre KE2 mice.

Statistical analyses

169   Gene-level counts were were imported into R for reproducible data management, manipu-
170   lation and analysis using the dplyr, plyr, and knitr packages (**Xie, 2015**, **2014**, **2017**; **Wickham**
171   **and Francois, 2016**).

172   We used DESeq2 for gene expression normalization and quantification of gene level counts
173   (**Love et al., 2014**). We used a threshold of a false discovery corrected (FDR) p value < 0.1. Statis-
174   tics on the principal component analyses (PCA) were conducted in R (**Wickham, 2009**, **2011**).
175   We used the VennDiagram R package (**Schwenk et al., 1984**) for preliminary visualization of
176   differential gene expression, but the final Venn diagrams were drawn with Adobe Illustrator.

177   The hierarchical clustering analysis was conducted and visualized using the R package
178   pheatmap (**Kolde, 2015**) with the RColorBrewer R packages for color modifications (**Neuwirth,**
179   **2014**). The bootstrap probability values for the dendrogram were calculated using the R pack-
180   age pvclust (**Suzuki and Shimodaira, 2006**).

181   PCA was conducted in R using the DESeq2 and genefilter R packages (**Love et al., 2014**;
182   **Gentleman et al., 2017**). The PCA analysis was visualized using the ggplot2, cowplot, and
183   RColorBrewer R packages (**Wickham, 2009**; **Wilke, 2016**).

184   We used GO_MWU for analysis of GO ontology (**Wright et al., 2015**). Figure 2 was generated

185 using –log(p-value) as a continuous measure of significance to identify GO categories that
186 are significantly enriched with either up- or down-regulated genes. No significance cutoff is
187 required for the analysis, but an arbitrary p-value was set to visualize the top 10 most significant
188 GO terms. Figure 5 was generated in two steps. First, a p-value = 0.1 was set for determining
189 significantly expressed genes in each analysis, and these data were converted into a binary (0
190 or 1) for a typical GO enrichment analysis using Fisher's exact test to determine if GO categories
191 are overrepresented among the significantly expressed genes.

Archival of data, code, and figures

193 The raw sequence data and intermediate data files are archived in NCBI's Gene Expression
194 Omnibus Database (accession numbers GSE99765 and GSE100225). The data and code are
195 available on GitHub https://github.com/raynamharris/DissociationTest, with an archived version at
196 the time of publication available at Zenodo (*Harris et al., 2017a*). The schematic images and
197 figure modifications were made using Adobe Illustrator and archived in FigShare under a CC-
198 BY license (*Harris et al., 2017b*).

Results

200 We obtained an average of 5 million reads for each hippocampal tissue sample and quantified
201 the expression representing 22,485 genes in the mouse reference transcriptome MV11.

The effects of cellular dissociation on hippocampal transcriptomes

203 We identified 162 genes that were differentially expressed between the control and dissoci-
204 ated samples, 331 genes that were differentially expressed genes (DEGs) between any of the
205 three hippocampus subfields, and 30 genes were shared between both sets of differentially
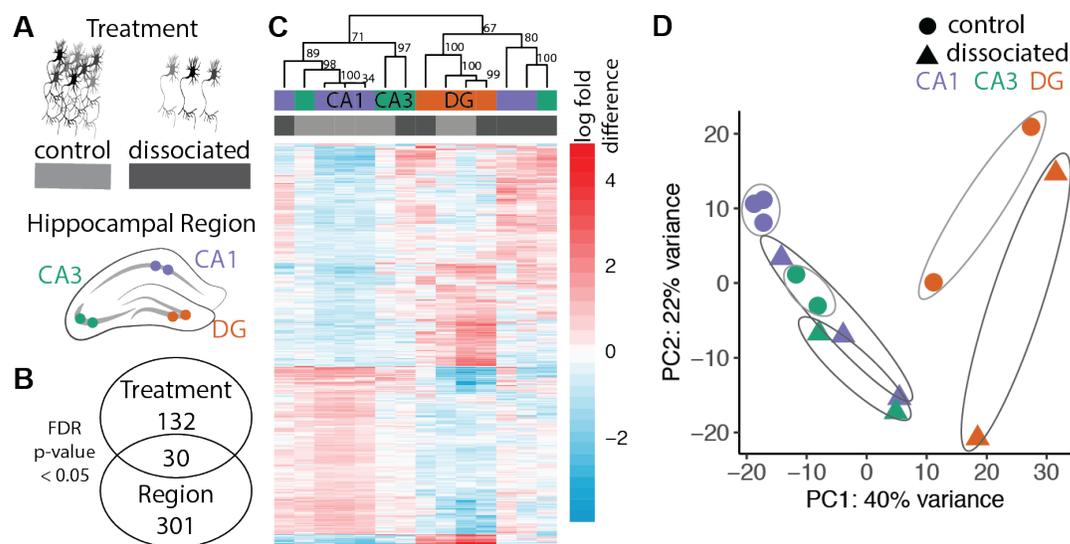206 expressed genes at p-value < 0.05 (Fig. 1A,B).

Figure 1. The effect of cellular dissociation on hippocampal transcriptomes. A) From a single female mouse, we collected 2 CA1, CA3, and DG hippocampal tissue samples. One sample was subjected to a cellular dissociation treatment (dissociated) whereas the control samples (control) were standardly homogenized. **B)** We identified 162 dissociation-induced changes in gene expression, 331 genes with region-specific expression patterns, and 30 genes differentially expressed by both region and treatment (FDR p-value < 0.05). **C)** Hierarchical clustering separates the hippocampal subfields of the homogenized samples (light gray) but not the dissociated samples (dark gray). **D)** PC1 accounts for 40% of all gene expression variation and by inspection, separates the DG samples from the CA1 and CA3 samples. PC2 accounts for 22% of the variation in gene expression and varies significantly with treatment. Ellipses are hand-drawn.

207   A hierarchical clustering analysis of all differentially expressed genes does not give rise
208 to distinct clusters that are separated by subfield or method; however, when examining the
209 control, homogenized samples alone (identified with light grey boxes), the three subfields
210 form distinct clusters, while the dissociated samples do not cluster by subfield (Fig. 1C).
211   Next, we conducted a principal component analysis of all identified genes. PC1 accounts
212 for 40% of the variation and visually separates the DG samples from the CA1 and CA3 sam-
213 ples (Fig. 1D). To confirm statistical significance of this visual pattern, we conducted a two-
214 way treatment x region ANOVA and confirmed a significant effect of region ($F_{2,11} = 17.69$; p
215 = 0.0004). Post hoc Tukey tests confirmed CA1 = CA3 < DG. The effect of treatment and the
216 treatment x region interaction were not significant. PC2 accounts for 22% of the variation in
217 gene expression and varies significantly with treatment ($F_{1,12} = 6.125$; p = 0.03) but not by re-
218 gion or the treatment x region interaction. None of the other PCs showed significant variation
219 according to either region or treatment.

220 ## The effects of stressful experience on hippocampal transcriptomes

221 We examined the effect of stressful experience, which is a common confound of behavioral
222 manipulations because animals in different experimental groups often experience different
223 levels of stress, especially if the experimental procedure is not intentionally stressful. We iden-
224 tified 0 genes that were significantly expressed between samples from the home cage and
225 shocked mouse samples; 1669 genes were significantly differentially expressed between any
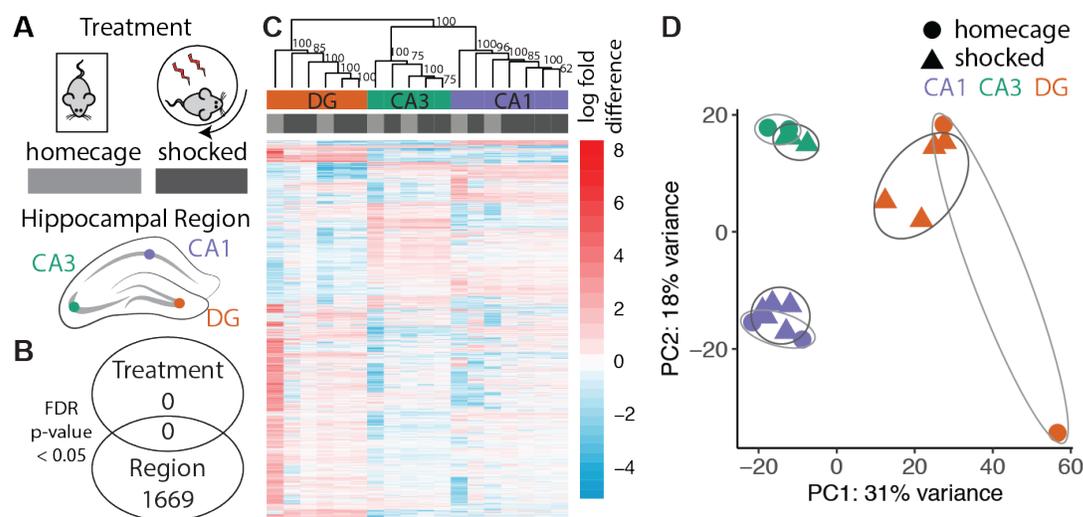226 of the three brain regions at p-value < 0.05 (Fig. 2A, B).

Figure 2. The effects of a stressful experience on hippocampal transcriptomes. A) We compared CA1, CA3, and DG tissue samples from control mice taken directly from their home cage to mice that were subjected to a mild foot shock. **B)** We identified 0 genes that responded to treatment, and 1669 genes that were differentially regulated across regions of the hippocampus (FDR p-value < 0.05). **C)** Hierarchical clusters groups samples by brain region but distinct treatments clusters are not present. **D)** PC1 accounts for 31% of the variation and visually separates the DG samples from the CA1 and CA3 samples. PC2 accounts for 1% of the variation and distinguish the three subfields. Ellipses were hand-drawn.

227   Hierarchical clustering of the differentially expressed genes gives rise to three distinct clus-
228 ters corresponding to the three subfields, with CA1 (purple) and CA3 (green) being more similar
229 to one another than to DG (orange), whereas the effects of the stress manipulation were not
230 distinctive (Fig. 2C).
231   Next, we conducted a principal component analysis of all the genes that were measured.
232 PC1 accounts for 31% of the variation and by inspection, separates the DG samples from the

NOT PEER REVIEWED - Manuscript submitted to Hippocampus.

233 CA1 and CA3 samples (effect of region: $F_{2,15}$= 42.89; p < 0.001; Fig. 2D). Post hoc Tukey tests
234 confirmed CA1 = CA3 ≠ DG. The effects of stress and the stress x region interaction were not
235 significant. PC2 accounts for 18% of the variation and varies significantly between CA1 and
236 CA3 and CA1 and DG (effect of region: $F_{2,15}$= 11.41; p < 0.001; Tukey tests: CA1 ≠ DG = CA3). The
237 effects of stress and the stress x region interaction were not significant. PC3 accounts for 15%
238 of the variation and also explains some brain region specific differences (effect of region: $F_{2,15}$=
239 6.315; p < 0.01; Tukey tests: CA1 = DG ≠ CA3), whereas effects of stress and the stress x region
240 interaction were not significant. PC4 is also influences by region ($F_{2,15}$= 6.315; p = 0.0102; Tukey
241 tests: CA1 ≠ CA3. PC5 did not account for any significant differences according to region or
242 treatment. PC6 significantly accounted for variance associated with the effect of a stressful
243 experience ($F_{1,16}$> 4.774; p's < 0.04).

The effects of cognitive training on hippocampal transcriptomes

245 We identified that 423 genes were differentially expressed between the yoked control and
246 cognitively trained animals, 3485 genes that were differentially expressed across subfields, and
247 324 showed an interaction at FDR p < 0.05 (Fig. 3A, B). We see a large effect of brain region
248 on gene expression, with 20% of detectable genes being differentially expressed between
249 one or more brain-region comparisons (3485 differentially expressed genes /17320 measured
250 genes). This is an order of magnitude greater than the 2% of the transcriptome that changed
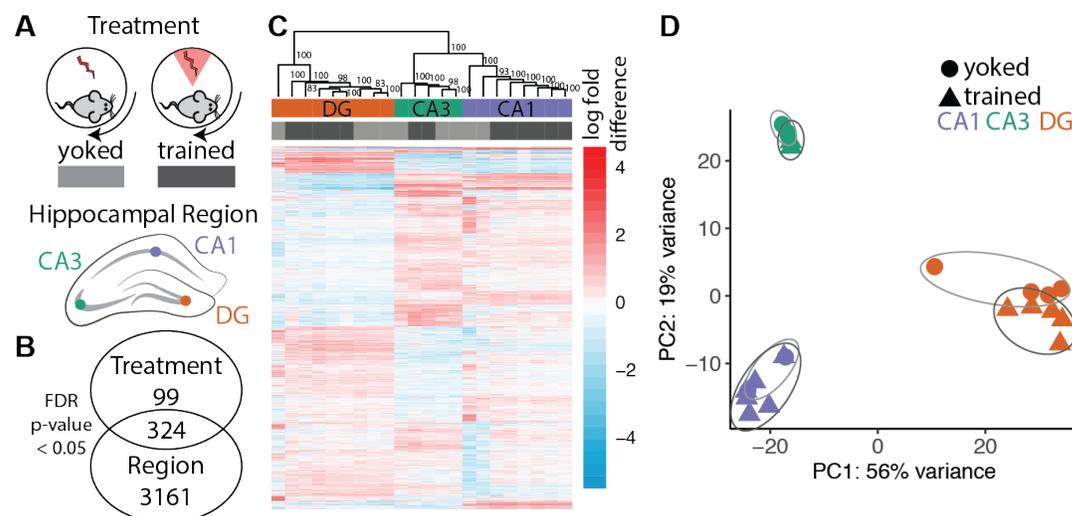251 in response to learning (423 DEGs /17320 genes measured).

Figure 3. Effects of a learned avoidance behavior on hippocampal transcriptomes. A) Mice used in this study were either subjected to random but mild foot shocks (control) or subjected to mild foot shocks conditioned with spatial cues (trained). Tissue samples were collected from CA1, CA3, and DG. **B)** We identified only 285 genes that were significantly expressed according to the behavioral manipulation and identified 3622 genes that were were differentially expressed between any of the three brain regions. **C)** Hierarchical clustering of the differentially expressed genes gives rise to three distinct clusters corresponding to the three brain regions, with CA1 and CA3 being more similar to one another than to DG. **D)** A principle component analysis of all genes in analysis (regardless of level of significance) shows that PC1 accounts for 50% of the variation and distinguishes the DG samples and the CA1 and CA3 samples (Region: $F_{2,19}$= 199.3; p = 1.78e-13). PC2 accounts for 18% of the variation and distinguishes all three subfields ($F_{2,19}$= 220.4; p = 7.15e-14). Ellipses were hand-drawn.

252 Hierarchical clustering of the differentially expressed genes separates samples by both
253 subfield and treatment (Fig. 3C). A principal component analysis of all gene expression data
254 revealed that brain region explains 75% of the variance in the data (Fig. 3D). PC1 accounts for
255 56% of the variance and distinguishes DG from the Ammon's horn samples (effect of region:
256 $F_{2,19}$= 226.1; p < 0.001; Tukey tests: CA1 = CA3 ≠ DG), but the effects of training and the training x

region interaction were not significant. PC2 accounts for 19% of the variance and distinguishes the three subfields (effect of region: $F_{2,19} = 255.3$; $p < 0.001$; Tukey tests: CA1 ≠ CA3 ≠ DG). PC3 and PC4 indicate a significant influence of cognitive training (PC3: $F_{1,20} = 7.451$; $p = 0.01$,) and (PC4: $F_{1,20} = 10.11$; $p = 0.005$), but no significant effects of region and the region x treatment interaction.

Identifying unique and general patterns of hippocampal genomic plasticity

Next, we examined the overlap in genomic response to the technical and biological perturbations. We identified three specific genes that responded to both cellular dissociation and to cognitive training: Grin2a, Epha6, and Ltbp3 (Fig. 4A). There was no overlap in differentially expressed genes compared to the cellular dissociation treatment (Fig. 4A).

A. Gene expression responses to technical and biological perturbations

technical

159

0

0

3

0

0

0

420

biological

biological

Eha6
Grin2a
Itbp3

FDR p-value
< 0.05

B. Dissociation-enriched molecular functions

74/325 structural molecule
42/88 structural constituent of ribosome
15/55 rRNA binding
8/128 helicase
19/245 ligase, forming carbon–nitrogen bonds
32/433 ligase
12/62 oxidoreductase, acting on NAD(P)H
50/596 oxidoreductase
10/36 oxidoreductase, acting on NAD(P)H, quinone or similar
11/66 hydrogen ion transmembrane transporter

UP
Down
$p < 0.00001$
$p < 0.0001$
$p < 0.001$

C. Cognitive training-enriched molecular functions

180/801 poly(A) RNA binding
20/87 structural constituent of ribosome
10/36 oxidoreductase, acting on NAD(P)H, quinone or similar
11/25 glutamate receptor
128/801 signal transducer
105/678 receptor
13/66 hydrogen ion transmembrane transporter
143/735 transmembrane transporter
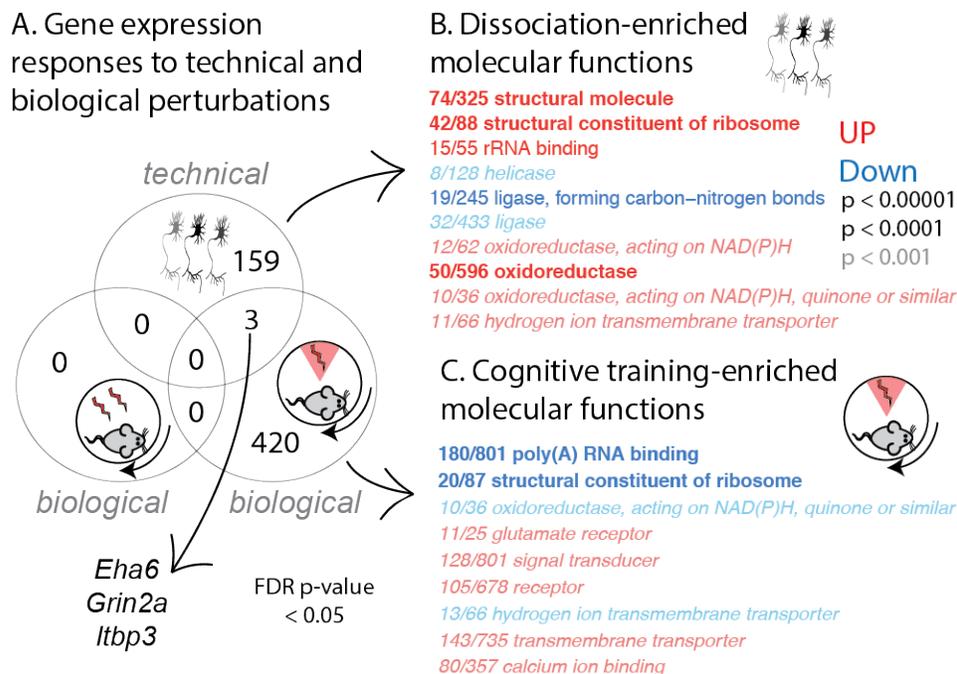*80/357 calcium ion binding*

Figure 4. Unique and shared responses to technical treatments and biological perturbations. A) The number of genes that responded to chemical dissociation (163 genes), a stressful experience (0 genes), and cognitive training (423 genes). The three genes that respond to both technical and biological perturbation are Epha6, Grin2a, and Itbp3. **B, C)** The molecular function of gene ontology (GO) categories that are significantly enriches with either up- or down-regulated genes in response to cellular dissociation **(B)** or cognitive training **(C)**. The top 10 most significant GO terms are visualized, each with a p-value < 0.001. The fraction next to GO category name indicates the fraction of genes in that category that survived a 10% FDR threshold for significance. Zero terms survived a 10% FDR threshold in response to a stressful experience.

We next analyzed gene ontology at 5% FDR significant in each of the data sets to identify the molecular function of genes that changed in response to cellular dissociation (Fig. 4B) or cognitive training (Fig. 4C). The process of cellular dissociation results in a significant up-regulation of molecular processes related to ribosomal activity, rRNA binding, oxidoreductase activity, and proton transport, while it caused a down regulation of ligase and helicase activity (Fig. 4B). The GO analysis detected no Molecular Function GO terms in the significantly overrepresented genes in response to the stressful experience. Cognitive training resulted in a significant upregulation of molecular processes related to glutamate receptors, signal trans-duction, calcium binding, and membrane transport, and it resulted in a significant downregu-lation of ribosomal activity, oxidoreductase activity, mRNA binding, and proton transport (Fig. 4B). The gene ontology analysis identified 91 significant GO terms in response to cognitive

278 training. Among the top 10 are glutamate signaling and membrane transport systems and a
279 downregulation of oxidoreductase and ribosomal activity (Fig. 4C). Notably, learning induces
280 a downregulation of a structural constituent of ribosomes and oxidoreductase, which were
281 both up-regulated in response to cellular dissociation (Fig. 4B,C).

282 ## Recovering robust subfield-specific gene expression patterns

283 Using the public *Cembrowski et al.* (*2016b*) dataset, we identified 10,751 genes that were
284 differentially expressed between hippocampal sub-regions (Fig. 5A). Using meta analyses of
285 the public Cembrowski data with the primary data described herein, we identified 146 genes
286 that showed robust subfield-specific gene expression patterns (Fig. 5A). Those 146 genes are
287 enriched in cellular compartments related to synapses and molecular functions related to
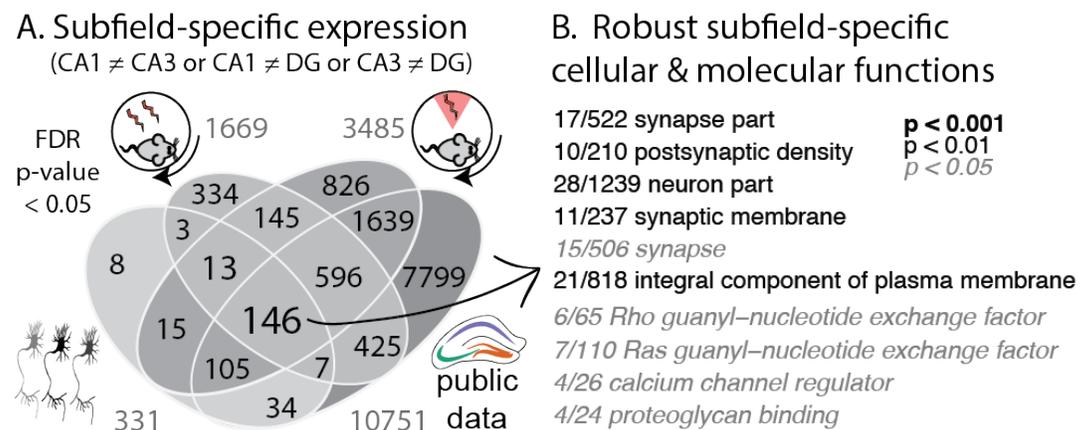288 calcium signaling, GTP exchange, and proteoglycan binding (Fig. 5B).

Figure 5. Meta analysis of primary and public data. A) This Venn diagram shows the overlap in brain-region specific gene expression across all four experiments (cellular dissociation, stressor habitation, cognitive training, and a public dataset examining subfield comparisons). Grey numbers indicate total number of differentially expressed genes between and two-way subfield comparison. Using this approach, we identified 146 genes that were differentially expressed between any two subfields of the hippocampus in all four experiments. **B)** Those 146 provide robust brain-region specific markers of gene expression belong to molecular function and cellular compartment GO. The top 10 most significant GO terms are visualized, each with a p-value < 0.05. The fraction next to GO category name indicates the fraction of genes in that category that survived a 10% FDR threshold for significance.

289 ## Discussion

290 The main purposes of this study were 1) to test whether analysis of gene expression in hip-
291 pocampus subfields is changed by tissue preparation procedures (cellular dissociation versus
292 homogenization) and 2) to evaluate the effects of a stressful experience relative to cognitive
293 training on analysis of gene expression. The work was designed to evaluate the extent to
294 which technical (i.e. cellular dissociation) and biological confounds (i.e. stressful experience)
295 can impact efforts to assess the transcriptomic response to cognitive processes. This is po-
296 tentially important because it is increasingly necessary to dissociate cells in tissue samples for
297 single cell or single population studies.

298 Hippocampal subfield differences are well known (*Lein et al., 2004*; *Hawrylycz et al., 2012*;
299 *Cembrowski et al., 2016a*,b). Across the three experiments with different treatments, the
300 identity of the hippocampal subfield, explained between 40 and 75% of all the variation in
301 gene expression across samples(Fig. 1D, Fig. 2D, Fig. 3D). The samples that were subjected to
302 cellular dissociation show the least amount of region-specific variation, suggesting that this
303 process might alter the genes that normally distinguish the hippocampal subfields from one
304 another. On the other hand, the *Cembrowski et al.* (*2016b*) study identified a larger number

305 of genes with subfield specificity; this is likely due to the cell sorting process that generates
306 a relatively homogenous rather than a heterogeneous population of neurons. These results
307 indicate that cell-type specific differences may be recovered by sorting cells into homogenous
308 populations.

309 Across the data set, many genes consistently or robustly demonstrate hippocampal sub-
310 field specificity in their expression (Fig. 5B). The meta-analysis of the primary data and the
311 public data (*Cembrowski et al., 2016b*) identified 146 genes that could potentially serve as ro-
312 bust subfield specific markers. The molecular functions of these potential marker genes are
313 diverse, related to calcium channel regulation, proteoglycan binding, and guanyl-nucleotide
314 (GTP) exchange, as well as cellular compartment categories related to the synapse and the
315 postsynaptic density. This suggests that the phenotypic and functional differences amongst
316 DG, CA3, and CA1 neurons may be driven or influenced by subfield differences in gene expres-
317 sion.

318 With respect to the effects of cellular dissociation of hippocampal gene expression, we
319 found that 0.9% (162/16,709) of the genes measured changed in response to cellular dissoci-
320 ation (Fig. 1B). This is smaller than the 2.4% (423 /17,320) change we detected in response to
321 cognitive training (Fig. 3B). The stressful experience produces a negligible response (i.e. no
322 significant genes or GO terms were detected), indicating that the mild stress that likely ac-
323 companies most behavioral tasks does not have a lasting influence on hippocampal gene
324 expression (Fig. 1B).

325 The extent to which cellular dissociation and unintended stress impacts the expression of
326 particular genes and signaling pathways, limits the feasibility of investigating how genes con-
327 tribute to behavior and other responses to organismal manipulations. We found that Grin2a
328 responded to both cellular dissociation and cognitive training (Fig. 4A). Grin2a encodes sub-
329 units of N-methyl-D-aspartate (NMDA) type ionotropic glutamate receptors that are crucial for
330 numerous cellular functions throughout the brain, including hippocampus-dependent synap-
331 tic plasticity and learning (*Collingridge et al., 1983*; *Morris, 2013*).Accordingly, care should be
332 taken when studying the role of glutamate and MAPK signaling in combination with cellular
333 dissociation techniques. Epha6 and Ltbp3 also responded to both cellular dissociation and
334 cognitive training (Fig. 4A). Epha6 is involved with the MAPK-Erk signaling pathway. Ltbp3
335 is involved in binding calcium ions and shows altered gene expression in a mouse model of
336 Alzheimer's Disease (*Neuner et al., 2017*).

337 We can look beyond the specific genes and examine which pathway responses are con-
338 cordant or discordant to multiple treatments. In this case, we saw upregulation of ribosomal
339 activity and rRNA biding in response to cellular dissociation, but we saw an opposing downreg-
340 ulation in ribosomal activity and mRNA binding in response to cognitive training (Fig. 5B,C).
341 This suggests that cellular dissociation activates a general transcriptional response whereas
342 cognitive training reduces the transcription of specific protein coding genes. This demon-
343 strates the possibility that such an interaction, in this case a downregulation in response to
344 cognitive training could be overshadowed by technical artifacts if hippocampus tissue is first
345 subjected to the cellular dissociation required for single-cell or single cell population investi-
346 gations.

347 We found no detectable transcriptional response in the CA1, CA3, or DG following the stress-
348 ful experience (Fig. 2B). The shock experience we used causes a large increase in plasma
349 corticosterone levels, comparable to exposure to predator threat, that is observed after the
350 initial shock exposure session but is absent 24-h later after the second training session (Les-
351 burguères et al., 2016b). Our findings support the use of either home cage or shock-yoked
352 animals as controls for active place avoidance training experiments. In the case of the home
353 cage controls that do not experience shock, their stress response is indistinguishable from the
354 trained mice but their sensory experience is very different. In contrast, the shock-yoked mice
355 have identical sensory experience as the experimental mice, but they experience stress that

356 the experimental animals do not (*Lesburguères et al., 2016*). It may be that untrained control
357 mice are optimal, because they would have the identical experience of the environment as
358 experimental mice, except at the brief 500 ms moments of shock that account for roughly
359 3% of the task assuming 20 shocks in 600 s. Depending on the question one control may be
360 preferable over the others, but as demonstrated here, when assessed 24 h after the training
361 experience, they appear to be equivalent in terms of their gene expression profiles (Fig. 2).

Conclusions

363 Many factors contribute to variation in gene expression. We set out to identify the extent to
364 which the process of cellular dissociation – which allows for single cell analysis of neurons
365 – has an appreciable effect on our ability to detect biologically meaningful variation in hip-
366 pocampal gene expression. We conclude that there are specific dissociation-induced and
367 cognitive training-induced changes in gene expression that are largely non-overlapping. It
368 is encouraging that the overlap between cellular dissociation and cognitive training is small,
369 indicating that these technical and biological processes affect different transcriptional pro-
370 cesses. It is also encouraging to know that the stressful experience had no substantial effect
371 on hippocampal gene expression, which if generalizable to other tasks will allow for using
372 behavioral control groups and behavioral manipulations that also induce modest, potentially
373 confounding stress. These findings provide insight into how cellular and biological manipu-
374 lations influence gene expression. Through meta analysis and comparison to the published
375 literature, we have identified a subset of robust sub-region specific markers of gene expres-
376 sion profiling. Further research is clearly needed to uncover the influence of other variables
377 on variation in hippocampal gene expression.

References

395 **Cembrowski MS**, Bachman JL, Wang L, Sugino K, Shields BC, Spruston N. Spatial Gene-Expression Gradi-
396 ents Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons. Neuron. 2016 1; 89(2):351–368. http://
397 www.ncbi.nlm.nih.gov/pubmed/26777276http://linkinghub.elsevier.com/retrieve/pii/S0896627315010867, doi:
398 10.1016/j.neuron.2015.12.013.

399 **Cembrowski MS**, Wang L, Sugino K, Shields BC, Spruston N. Hipposeq: A comprehensive RNA-seq
400 database of gene expression in hippocampal principal neurons. eLife. 2016 4; 5(APRIL2016):e14997.
401 http://elifesciences.org/lookup/doi/10.7554/eLife.14997, doi: 10.7554/eLife.14997.

402 **Chalancon G**, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between
403 gene expression noise and regulatory network architecture. Trends in Genetics. 2012 5; 28(5):221–232.

404 http://www.ncbi.nlm.nih.gov/pubmed/22365642http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=
405 PMC3340541http://linkinghub.elsevier.com/retrieve/pii/S0168952512000157, doi: 10.1016/j.tig.2012.01.006.

406 **Collingridge GL**, Kehl SJ, McLennan H. The antagonism of amino acid-induced excitations of rat hip-
407 pocampal CA1 neurones in vitro. The Journal of physiology. 1983 1; 334:19–31. http://www.ncbi.nlm.nih.
408 gov/pubmed/6134823http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1197297.

409 **Danielson NB**, Zaremba JD, Kaifosh P, Bowler J, Ladow M, Losonczy A. Sublayer-Specific
410 Coding Dynamics during Spatial Navigation and Learning in Hippocampal Area CA1. Neu-
411 ron. 2016 8; 91(3):652–665. http://www.ncbi.nlm.nih.gov/pubmed/27397517http://www.pubmedcentral.nih.
412 gov/articlerender.fcgi?artid=PMC4975984http://linkinghub.elsevier.com/retrieve/pii/S0896627316302987, doi:
413 10.1016/j.neuron.2016.06.020.

414 **Gentleman R**, Carey V, Huber W, Hahne F, genefilter: genefilter: methods for filtering genes from high-
415 throughput experiments; 2017.

416 **Guzowski JF**, McNaughton BL, Barnes CA, Worley PF. Environment-specific expression of the immediate-
417 early gene Arc in hippocampal neuronal ensembles. Nature neuroscience. 1999 12; 2(12):1120–
418 4. http://www.nature.com/doifinder/10.1038/16046http://www.ncbi.nlm.nih.gov/pubmed/10570490, doi:
419 10.1038/16046.

420 **Guzowski JF**, Miyashita T, Chawla MK, Sanderson J, Maes LI, Houston FP, et al. Recent behav-
421 ioral history modifies coupling between cell activity and Arc gene transcription in hippocam-
422 pal CA1 neurons. Proceedings of the National Academy of Sciences of the United States of
423 America. 2006 1; 103(4):1077–82. http://www.pnas.org/cgi/doi/10.1073/pnas.0505519103http://www.ncbi.
424 nlm.nih.gov/pubmed/16415163http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1347968, doi:
425 10.1073/pnas.0505519103.

426 **Harris RM**, Hofmann HA. Neurogenomics of Behavioral Plasticity. Advances in Experimental Medicine
427 and Biology. 2014; 781:149–168.

428 **Harris RM**, Kao HY, Alarcon JM, Hofmann HA, Fenton AA. The Github Repository For Analyses Of Hip-
429 pocampal Transcriptomic Responses To Technical And Biological Perturbations. . 2017 1; https://zenodo.
430 org/record/815081#.WUrDrDOZPdQ, doi: 10.5281/ZENODO.815081.

431 **Harris RM**, Kao HY, Alarcon JM, Hofmann HA, Fenton AA. Figures associated with 'Analysis of hippocampal
432 transcriptomic responses to technical and biological perturbations' manuscript. In: *FigShare*; 2017. doi:
433 10.6084/m9.figshare.5116192.v2.

434 **Hawrylycz MJ**, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehen-
435 sive atlas of the adult human brain transcriptome. Nature. 2012; 489(7416):391–399. http://www.nature.
436 com/doifinder/10.1038/nature11405, doi: 10.1038/nature11405.

437 **Kolde R**, pheatmap: Pretty Heatmaps; 2015.

438 **Lein ES**, Zhao X, Gage FH. Defining a Molecular Atlas of the Hippocampus Using DNA Microarrays and
439 High-Throughput In Situ Hybridization. Journal of Neuroscience. 2004; 24(15). http://www.jneurosci.org/
440 content/24/15/3879.

441 **Lesburguères E**, Sparks FT, O'Reilly KC, Fenton AA. Active place avoidance is no more stressful than
442 unreinforced exploration of a familiar environment. Hippocampus. 2016 12; 26(12):1481–1485. http://doi.
443 wiley.com/10.1002/hipo.22666http://www.ncbi.nlm.nih.gov/pubmed/27701792, doi: 10.1002/hipo.22666.

444 **Love MI**, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data
445 with DESeq2. Genome Biology. 2014 12; 15(12):550. http://genomebiology.com/2014/15/12/550, doi:
446 10.1186/s13059-014-0550-8.

447 **Mizuseki K**, Diba K, Pastalkova E, Buzsáki G. Hippocampal CA1 pyramidal cells form functionally distinct
448 sublayers. Nature Neuroscience. 2011 8; 14(9):1174–1181. http://www.ncbi.nlm.nih.gov/pubmed/21822270, doi:
449 10.1038/nn.2894.

450 **Mo A**, Mukamel E, Davis F, Luo C, Henry G, Picard S, et al. Epigenomic Signatures of Neu-
451 ronal Diversity in the Mammalian Brain. Neuron. 2015 6; 86(6):1369–1384. http://www.ncbi.nlm.
452 nih.gov/pubmed/26087164http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4499463http:
453 //linkinghub.elsevier.com/retrieve/pii/S0896627315004250, doi: 10.1016/j.neuron.2015.05.018.

454 **Morris RGM**. NMDA receptors and memory encoding. Neuropharmacology. 2013 11; 74:32–40. http://
455   www.ncbi.nlm.nih.gov/pubmed/23628345http://linkinghub.elsevier.com/retrieve/pii/S0028390813001512, doi:
456   10.1016/j.neuropharm.2013.04.014.

457 **Neuner SM**, Wilmott LA, Hoffmann BR, Mozhui K, Kaczorowski CC. Hippocampal proteomics defines
458   pathways associated with memory decline and resilience in normal aging and Alzheimer's disease
459   mouse models. Behavioural Brain Research. 2017; 322:288–298. http://www.sciencedirect.com/science/
460   article/pii/S0166432816303564, doi: 10.1016/j.bbr.2016.06.002.

461 **Neuwirth E**, RColorBrewer: ColorBrewer Palettes; 2014.

462 **Schwenk AJ**, Boutros PC, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Venn Diagram for Five Sets.
463   Mathematics Magazine. 1984 11; 57(5):297. http://www.jstor.org/stable/10.2307/2689606?origin=crossref,
464   doi: 10.2307/2689606.

465 **Suzuki R**, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical cluster-
466   ing. Bioinformatics. 2006 6; 22(12):1540–1542. https://academic.oup.com/bioinformatics/article-lookup/doi/
467   10.1093/bioinformatics/btl117, doi: 10.1093/bioinformatics/btl117.

468 **Wickham H**. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2009. http://ggplot2.
469   org, doi: 10.1007/978-0-387-98141-3.

470 **Wickham H**. The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software. 2011;
471   40(1):1–29. http://www.jstatsoft.org/v40/i01/, doi: 10.18637/jss.v040.i01.

472 **Wickham H**, Flexibly Reshape Data: A Reboot of the Reshape Package; 2016.

473 **Wickham H**, Francois R. dplyr: A Grammar of Data Manipulation. R package version 050. 2016; https:
474   //cran.r-project.org/package=dplyr.

475 **Wilke CO**, cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'; 2016.

476 **Wright RM**, Aglyamova GV, Meyer E, Matz MV.   Gene expression associated with white syn-
477   dromes in a reef building coral, Acropora hyacinthus.   BMC Genomics. 2015 12; 16(1):371.
478   http://www.ncbi.nlm.nih.gov/pubmed/25956907http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=
479   PMC4425862http://www.biomedcentral.com/1471-2164/16/371, doi: 10.1186/s12864-015-1540-2.

480 **Xie Y**. Implementing Reproducible Computational Research. Stodden V, Leisch F, Peng RD, editors,
481   Chapman and Hall/CRC; 2014. http://www.crcpress.com/product/isbn/9781466561595.

482 **Xie Y**. Dynamic documents with R and knitr. 2nd ed.; 2015.

483 **Xie Y**, knitr: A General-Purpose Package for Dynamic Report Generation in R; 2017. http://yihui.name/knitr/.