Patisaul Laboratory – Material and Methods

1. BEHAVIORAL TESTING PAPER: Rebuli, M. E. et al. Impact of Low-Dose Oral Exposure to Bisphenol A (BPA) on Juvenile and Adult Rat Exploratory and Anxiety Behavior: A CLARITY-BPA Consortium Study. Toxicol. Sci. 148, 341-354, doi:10.1093/toxsci/kfv163 (2015).

Materials and Methods

This study is a component of the CLARITY-BPA program (Birnbaum, Bucher et al. 2012, Schug, Heindel et al. 2013). Comprehensive study design details are fully described in (Heindel, Newbold et al. 2015) so only applicable portions are summarized here.

Animal Husbandry: Throughout the study, animal rooms were maintained at 23 ± 3°C with a relative humidity of 50 ± 20% and a 12:12 hr light/dark cycle. Food and water were available *ad libitum*. Dams and pre-weaned pups were housed with lights on at 6:00 AM. Animals were then moved at weaning to a different building with a shifted light cycle (off at 11:00, on at 23:00) to accommodate behavioral testing in the dark phase. The diet was soy- and alfalfa-free because these diets contain hormonally active compounds that may be a confounder and obfuscate behavioral sex differences and/or BPA-related effects (Patisaul, Blum et al. 2005, Thigpen, Setchell et al. 2007, Patisaul, Sullivan et al. 2012, Thigpen, Setchell et al. 2013)(5K96 verified casein diet 10 IF, round pellets, γ-irradiated (catalogue # 1810069), Test Diets, Purina Mills, Richmond, IN).

Dose Preparation and Administration: Five dose groups (n = 12 per sex per group) were generated for the studies reported herein: vehicle, BPA 2.5, 25, and 2500 µg/kg bw/day and EE 0.5 µg/kg bw/day (note: The full CLARITY-BPA study has additional groups (see (Heindel, Newbold et al. 2015)). Dosing began on GD 6 and terminated at weaning (PND 21). Direct gavage of the pups began on PND 1 after the litter was culled. For pups younger than PND 5, the gavage needle did not enter the esophagus. Pups were weighed and gavaged daily until PND 21 (weaning).

Weaning and Transfer of Subjects: Offspring were weaned on PND 21, after their last daily gavage, and tattooed on the tail with a unique identification number. Animals used for the present study were then transported to a different building for housing and behavioral testing. The post-weaning housing rooms were held under identical environmental conditions as the pre-weaning housing room described above,

except for the light cycle (23:00 – 11:00), in order to accommodate testing in the dark phase. Only pups from litters with at least 9 live pups on PND 0 and a balanced sex ratio at birth (no litter had more than a 4 pup sex difference except for two litters in load 5, which had a 5 pup sex difference: 9 males and 4 females) were used in this study. Juvenile testing began on PND 25, allowing the animals from PND 21 to PND 25 to habituate to the new building. Juvenile and adult test subjects were siblings; that is, one/sex/litter was assessed as juveniles and another one/sex/litter was assessed in adulthood. At weaning, each subject was housed with one or two conspecifics (same-exposure group, same-sex, same-age, non-siblings). Where needed, treatment-naïve "companion" rats were used to provide cagemates for those study subjects that could not be housed with a conspecific (i.e., those in which only one litter of that exposure group was born on that day). No data were collected from these "companion" rats.

***Behavioral Testing**:* Rats were assessed either as juveniles on PND 25-27 (pre-puberty) or at adulthood using a test battery selected because the tasks have high predictive validity for anxiety and generate corroborative results (Walf and Frye 2007, Chadman, Yang et al. 2009). Juveniles were assessed using the elevated plus maze (EPM) and open field (OF). Adults were first assessed for 7 consecutive days using a Barnes Maze by another CLARITY-BPA consortium team, then on the EPM, OF, and zero maze (ZM). Testing procedures conformed to commonly used standards previously reported and used by us and others (Pellow, Chopin et al. 1985, Shepherd, Grewal et al. 1994, Hogg 1996, Ferguson and Boctor 2010, Ferguson, Law et al. 2012, Patisaul, Sullivan et al. 2012, Cao, Rebuli et al. 2013). Behavioral testing rooms (each containing only one type of maze) contained a white noise generator (producing ~66 dB; Marpac Dohm, Rocky Point, NC), and multiple apparatuses, half designated for males and half designated for females. All apparatuses were cleaned with 70% ethanol after each testing session. Subjects were pre-assigned to an apparatus such that approximately equal numbers of each exposure group were tested in each. When possible, cagemates were tested on the same day, but if not feasible (e.g., when estrous cycle did not match testing protocol), cagemates were tested in sequential sessions (days between testing of cagemates ranged from 1-8). All assessments commenced after housing room lights were off (approximately 11:00) and were completed within four hours. For testing, all subjects were transported to the nearby test room in their home cages on a rolling cart, and remained on the cart until testing. The hallway between the housing room and test rooms was illuminated with red light. The OF was a beam break assay (PAS-Open Field, San Diego Instruments, San Diego, CA). All other tests were video recorded from overhead cameras under dim red lighting and analyzed from the video by TopScan software (Clever Sys Inc., Reston, VA) by the NCSU research team. For all tasks, the number of

defecation events was counted because this behavior is sometimes interpreted to indicate anxiety, but these data were ultimately not used or analyzed because the majority of animals did not defecate during testing. Animals were weighed at the time of testing to look for signs of overt toxicity. As expected, a significant effect of sex on body weight was observed at both ages ($p \leq 0.001$ for both ages) with juvenile and adult males weighing more than same-age females, but exposure had no effect on body weight in either sex at either age.

Juvenile testing spanned PNDs 25-27 to minimize the likelihood that females would be tested after vaginal opening (pubertal onset). For adult behavioral testing, subjects from the two loads (4 and 5) were subdivided into testing intervals for logistical reasons. Both sexes (beginning at PND 77 for half of each load and PND 91 for the remainder) were handled daily to become habituated to human contact. Because behavior varies across the estrous cycle, monitoring and controlling for estrous cycle to the best degree possible is crucial for decreasing biological variability that could result from different estrous cycle phases at the time of the assessment and ensuring testing consistency (Anderson 1940, Archer 1975, Mora, Dussaubat et al. 1996, Diaz-Veliz, Alarcon et al. 1997, Frye, Petralia et al. 2000, Patisaul, Blum et al. 2005). Vaginal lavage began on PND 84 (for half of each load) or 98 (remaining animals) and continued daily until sacrifice. Estrous cycle stage was assessed each morning (between 7:30 and 8:00, or 3-3.5 hours before testing) via vaginal lavage. Slides were categorized by two experienced testers blind to treatment. Rats were tested on the EPM and ZM on the day they were categorized to be in proestrus or any stage of estrus (early to late). OF testing was conducted the day after EPM testing, regardless of estrous stage. Thus, the testing sequence for females was: 1) EPM during proestrus or estrus, 2) OF on the subsequent two days, and 3) ZM during proestrus or estrus. The testing sequence for males was 4 consecutive days (i.e., EPM, two days of OF, and ZM). Adult testing spanned 11 days maximally from PND 97 (for half of the subjects) or PND 111 (for the remainder).

Elevated Plus Maze (EPM)

Juveniles (PND 25) and adults were assessed for anxiety-like behavior during a five min test session using one of four EPMs, as previously described (Ferguson and Berry 2010). Briefly, each apparatus consisted of four connected black Plexiglas arms, each 10 cm wide and 50 cm long, elevated 50 cm above the floor. Two arms were enclosed within 40 cm walls (closed arms) and two arms had a short (8 mm) ledge around the edge (open arms). Each subject was gently placed on the central area facing the closed arm closest to the room wall, and the home cage and rolling cart were moved outside the test room.

Open Field (OF)

Juveniles (PNDs 26-27) and adults were assessed for anxiety and locomotor activity during two 30 min test sessions (over two consecutive days) using one of eight OF apparatuses as previously described (Ferguson, Law et al. 2012). The clear Plexiglas arenas (each 40 x 40 x 40 cm) had a 16 x 16 photo beam detector around the outside floor perimeter for detection of horizontal movements and an elevated photo beam detector to measure vertical activity. Opaque boards between adjacent apparatuses prevented visual contact. Each subject was introduced to the front left corner (same apparatus on both testing days).

For each of the two test days, activity was collected in five min intervals, and summed over the entire 30 min testing session (total activity). Behaviors assessed were total distance traveled (cm), average speed (cm/s), resting time (total time with no activity for > 2s), and time and entries into the center area (defined as the central 20 x 20 cm). An "entry" was defined as consecutive breaking of two beams. PAS-Reporter (San Diego Instruments, San Diego, CA) was used to convert the raw x,y beam break data into the distance, speed, resting, and zone data for statistical analyses.

Zero Maze (ZM)

Adults were assessed for anxiety-like behavior during a five min session using one of two ZM apparatuses, constructed to be consistent with those previously described (Shepherd, Grewal et al. 1994, Patisaul and Bateman 2008, Cao, Rebuli et al. 2013). Each maze consisted of two open arms (9.5 cm wide) and two closed arms (29.5 cm high walls), was 123 cm in diameter, and elevated 75.5 cm above the floor. Each of the two open arm areas had a 10 mm ledge around the edge (so as to be structurally similar to the open arms of the EPMs). The subject was gently placed onto an open arm facing a closed arm and left undisturbed for 5 min.

Summary of Primary Endpoints in the Behavioral Tasks

The strongest indices of anxiety in these tasks are open arm activity in the EPM and ZM (less = heightened anxiety) and center activity in the OF (less = heightened anxiety) (Bailey and Crawley 2009, Gould 2010). The most robust indices of activity are closed arm exploration on the EPM and ZM and total distance traveled in the OF (over the full 30 min task). Habituation was assessed by comparing OF behavior across the two successive testing days (activity declines with experience) (Bailey and Crawley

2009, Gould 2010). Results from all assessed endpoints are presented in the tables and the most commonly reported, salient endpoints for each testing apparatus depicted graphically.

Data Decoding

All behavioral testing was completed and scored blind to exposure group. The blinded raw data were submitted to the NTP Chemical Effects in Biological Systems (CEBS) database. It was then independently verified to account for all expected datasets and data points, and "locked" such that data could not be altered. The NCSU researchers were then provided with the exposure code for data analysis.

Statistical Analysis

The statistical approach was developed to be consistent with prior work and using published guidelines for low dose endocrine disrupting chemical (EDC) studies with sample sizes in this range (Haseman, Bailer et al. 2001). Main effects and their interactions were examined using analysis of variance (ANOVA). A Fisher's protected Least Significant Difference (LSD) was used as the post-hoc test (when main effects or interactions were identified). While the Fisher's protected LSD does not provide the strong family-wise error control of alternative post-hoc procedures, it was selected over a more conservative approach to minimize risk of Type-II error (rejecting a meaningful effect). Because very few BPA effects (versus vehicle control) were identified, controlling for false-positives was not considered of high concern, as doing so would not impact data interpretation. All statistical analyses were implemented in R (Team 2014) and adults and juveniles were analyzed separately. For all endpoints, significance was considered $p \leq 0.05$.

EPM data from two juvenile subjects (one EE male and one BPA 2.5 female) and one adult vehicle control female were excluded from analyses because they fell from the apparatus. One adult EE male was excluded from the ZM analyses because it was an extreme outlier (greater than twice the number of open arm entries as the next highest data point for that sex and exposure group). This exclusion did not affect the statistical identification of any exposure effects. Four adult females (three EE and one BPA 2.5) could not be included in the analysis for the second OF day, because the data collection software was not started. Because of the reduced sample size for the adult female EE group on the second OF test day, data from the second OF day were only used to assess the impact of test day on the outcomes. Data from the first OF day were analyzed in detail, graphed, and included in the figures and tables. For consistency, the juvenile OF data were approached the same way.

For EPM and ZM data sets, ANOVA models assessed effects of sex, exposure, and exposure by sex interactions. Significant main effects were followed up with a Fisher's protected LSD post-hoc test. Because aspects of EPM and ZM behavior are sexually dimorphic, if a main effect of sex was found for any endpoint on that maze, all subsequent analyses for exposure-related effects on that maze were made within sex. Additionally, confirmation of known sex differences in the vehicle controls was considered to be an indication that the test was robust, powered sufficiently to detect a difference in the range of that effect size, and properly conducted. As commonly seen with a sample size of 12/sex/exposure group, achieving normality in all residual distributions within a given endpoint ANOVA model was rare. Because violations of this assumption tend to produce false positives and there were no consistent treatment-associated effects, we did not differentially perform non-parametric tests in cases where deviation from normality may have been present. Rather, we applied a consistent modeling approach to all endpoints across each maze type (Cohen, Cohen et al. 2002).

OF data sets were analyzed in two ways: (1) for each endpoint the data were summed over the entire 30 min session and analyzed and (2) behavior was also assessed in 5 min intervals (i.e., a separate ANOVA was conducted for each 5 min interval). Breaking the 30 min session down into 5 min intervals allows exploratory behavior to be assessed at different points across the session as behavior changes with experience (Goma and Tobena 1978, Bailey and Crawley 2009, Gould 2010). The first 5 min of the test are thought to give the most informative general measures of anxiety (because novelty is highest). As the test progresses, activity declines as the animal becomes familiar with the arena; thus, differences in overall activity or center area behavior during the final intervals could be reflective of anxiety and/or exploratory behavior. Activity towards the end of the 30 min task is thought to reach a steady state so behavior in the final 5 min interval is considered to be the best indicator of general (not driven by novelty stress) locomotion (Gould 2010). For both the 30 min and interval analyses , a three-way ANOVA was conducted to test for main effects of sex, exposure, and test day (across the two days), and their interactions. Day 1 data were then further assessed using two-way ANOVAs with sex and exposure as factors. Significant main effects and interactions were followed up with a Fisher's protected LSD post-hoc test. All tables report *p*-values for the F-test associated with each endpoint across all factors tested.

Because effects were primarily negative, a post-hoc power analysis for a range of treatment effect sizes was performed to evaluate possible risk of a Type-II error (rejecting the null hypothesis when an effect is present). To parameterize these calculations, we used the experimental data (treatment group-wise means and variances) from the adult male EPM measure of time spent in the open arms.

The power calculations were implemented using the G*Power software (Faul, Erdfelder et al. 2009), then plotted using R.

Pairwise correlations between anxiety-related endpoints were conducted to assess data concordance across the OF, EPM, and ZM for the adult testing using methods similar to those described for characterizing inter-maze relationships (Padilla, Barrett et al. 2009).

2. HYPOTHALAMUS/HIPPOCAMPUS PND1 TRANSCRIPTOME PAPER: Arambula, S. E., Belcher, S. M., Planchart, A., Turner, S. D. & Patisaul, H. B. Impact of Low Dose Oral Exposure to Bisphenol A (BPA) on the Neonatal Rat Hypothalamic and Hippocampal Transcriptome: A CLARITY-BPA Consortium Study. Endocrinology 157, 3856-3872, doi:10.1210/en.2016-1339 (2016).

Materials and Methods

PND1 pups were obtained from litters generated for the CLARITY-BPA consortium program (Birnbaum, Bucher et al. 2012, Schug, Heindel et al. 2013). Detailed descriptions of animal husbandry, diet, breeding, and dose preparation and administration have been published elsewhere (Heindel, Newbold et al. 2015); therefore, only relevant methods are summarized here. All elements of the experimental design including doses, timing of exposure, and day of sacrifice were developed and agreed upon by the consortium. The program uses sibling NCTR Sprague Dawley (NCTR-SD) rats from the guideline-compliant chronic two year study, which follows standard protocols and contains classical endpoints typically considered by regulatory agencies in hazard identification and risk assessment (Birnbaum, Bucher et al. 2012, Schug, Heindel et al. 2013). Eight dose groups were included in this study (vehicle, BPA 2.5, 25, 250, 2500, and 25,000 µg/kw bw/day and EE 0.05 and 0.5 µg/kw bw/day). The two EE groups were incorporated into the CLARITY design for the purposes of serving as a reference estrogen and to provide directional guidance to establish if BPA-related effects were consistent with an "estrogenic" effect.

Tissue Collection

On PND1 terminal pup weights were collected and pups (no more than one per sex per litter) were sacrificed by decapitation. Within sex, there were no significant effects of exposure on pup terminal body weight (data not shown). Heads were rapidly frozen in dry ice and shipped to the Patisaul laboratory at North Carolina State University (NCSU) where they were archived and stored at -80°C until processing. All tissues were coded and all testing was done blinded to exposure group.

Each whole head was coronally cryosectioned (Leica CM1900, Nussloch, Germany) from the caudal end until the caudal borders of the hypothalamus and hippocampus were identified. Anatomical landmarks were located with the assistance of a rat and mouse brain atlas (Paxinos 1991, Paxinos 2007). Tissue samples were obtained using a sterilized stainless steel punch that was briefly cooled on dry ice, as previously described (Patisaul, Sullivan et al. 2012). For each animal, two hypothalamic punches and 4 hippocampal punches were obtained. The hypothalamic tissue collected consisted of two sequential punches, one taken anteromedially and one caudomedially, with a micropunch 1.25 mm in diameter and 1.00 mm in depth. These punches were combined and collectively comprised the entire hypothalamus sample. Because the hippocampus is irregular in shape, a series of smaller punches were made to obtain sufficient material but also ensure anatomical specificity. For each animal 4 punches were made, each 0.5 mm in diameter and 1.00 mm in depth: one pair of bilateral anterodorsal punches and one pair of bilateral caudoventral punches. All 4 samples were combined and collectively comprised the entire hippocampus sample. Each punch sample was expelled out of the stainless steel punch directly into a BPA-free Eppendorf tube on dry ice and stored at -80°C.

RNA-sequencing

The experimental design for transcriptome sequencing was developed in consultation with the NCSU Genomic Sciences Laboratory (GSL) and the Bioinformatics Research Core. Transcriptome sequencing was performed by the GSL on 24 hippocampal (n = 4 per sex per group) and 24 hypothalamic (n = 4 per sex per group) samples. Three experimental groups were examined: control, 2.5 BPA and 2500 BPA. These doses were selected in consultation with other consortium members based on effects observed in other subprojects (Rebuli, Camacho et al. 2015, Johnson, Javurek et al. 2016) and prior results in analogous studies using NCTR-SD rats (Cao, Rebuli et al. 2013, Rebuli, Cao et al. 2014). Because the DNA needed to be sent to another CLARITY member for subsequent studies (to be published elsewhere), DNA and RNA were co-extracted from frozen tissue samples using the ZR-duet DNA/RNA miniprep co-purification kit and treated with an on-column DNase I digestion, per the manufacturer's protocol (Zymo, Research). RNA quality was assessed with the Agilent 2100 Bioanalyzer. All hippocampal samples had an RNA integrity number (RIN) ≥9 and all hypothalamic samples had an RIN of 10. To optimize library complexity, all samples used as input material for library preparation had greater than 100 ng of total RNA. Sequencing libraries were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina and the NEBNext Poly (A) mRNA Magnetic Isolation Module (New England Biolabs, Cat. E7420 and Cat. E7490). Per manufacturer's instructions, mRNA was isolated, heat

fragmented, and primed with random primers. First strand cDNA synthesis was performed with actinomycin D and Protoscript II Reverse Transcriptase. Second strand cDNA synthesis was performed with second strand synthesis reaction buffer containing dUTP, which replaces dTTPs and preserves strand orientation information. Following cDNA synthesis, the fragments were purified and size selected using AMPure XP beads (Beckman Coulter Genomics, Cat. A63881). The cDNA library fragments were then treated with an End Repair enzyme mix and ligated onto adaptors specifically designed for the Illumina platform. PCR was used to replace the dUTPs in the adaptor sequence and the second strand of the cDNA fragments, enrich adaptor-ligated cDNA, and add six-nucleotide barcode sequences that allow for pooling of multiple samples for sequencing and sorting of data during analysis. Library clean-up was performed with AMPure XP beads and quality was confirmed on the Agilent 2100 Bioanalyzer. For both experiments (hippocampus and hypothalamus), cDNA libraries were combined into two pools of equal molar amounts. Following a balanced block design (Auer and Doerge 2010), both pools were multiplexed and run across three lanes. Libraries were sequenced using the 125-bp single-end protocol on an Illumina HiSeq2500 sequencer. Approximately 29.9 million reads were generated per hippocampal library and 29.2 million reads per hypothalamic library.

RNA-seq Data Analysis and Bioinformatics

RNA-seq data analysis was performed by the Bioinformatics Core at the University of Virginia. Quality control of read data was assessed with FastQC before and after adaptor trimming and filtering. The STAR alignment tool (Dobin, Davis et al. 2013) was used to align reads to the *Rattus norvegicus* (rn5) reference genomic sequence, downloaded from UCSC's Genome Browser. After aligning data, the number of reads mapping to GENCODE was calculated using the featureCounts software in the Subread package (Liao, Smyth et al. 2014). Count data were normalized for sequencing depth and distortion, and dispersion was estimated using the DESeq2 Bioconductor (Anders and Huber 2010, Love, Huber et al. 2014) package in the R statistical computing environment.

Quantitative real-time PCR (qRT-PCR)

Gene expression analysis by qRT-PCR was performed on eight treatment groups (n = 5 – 6): vehicle, BPA 2.5, 25, 250, 2500, and 25000 µg/kg bw/day and EE 0.05, and 0.5 µg/kg bw/day. A DNA/RNA miniprep co-purification kit (Zymo Research, Cat. D7001) with the addition of an on-column DNase I digestion (Zymo Research, Cat. E1007) was used to extract RNA. Extracted DNA was sent to another CLARITY consortium member for independent analysis. mRNA was reverse transcribed to

single-strand complementary cDNA with the high capacity RNA-to-cDNA kit (Applied Biosystems, Cat. 4387406). Each reverse transcriptase (RT) reaction was incubated for 60 min at 37°C, 5 min at 95 °C, and stored at -20°C until use. Real-time PCR was performed with an ABI StepOnePlus™ Real-Time PCR System. A TaqMan probe-based protocol was used to detect gene expression and primers and probes were included in these predesigned assays. The PCR reactions were incubated in 96-well plates and run using the manufacturer's recommended cycling parameters of 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. All reactions consisted of 1 µl hypothalamic RT product or 2 µl hippocampal RT product, 1 µl of 20X TaqMan gene expression assay mix, 10 µl of TaqMan Universal PCR Master Mix, and nuclease-free H_2O in a quantity sufficient to make a 20 µl total reaction volume. No-template controls were run for each Taqman gene expression assay and each PCR reaction was run in triplicate. An inter-run calibrator and the sample maximization approach was followed to avoid technical and run-to-run variation (Derveaux, Vandesompele et al. 2010). To correct for variation in starting cDNA concentrations, cycle threshold (Ct) values for the gene of interest for each sample were normalized to the Ct for 18s rRNA for each sample. 18s was selected as the normalizing transcript based on preliminary work for these experiments with other candidates showing that 18s was the most reliable and robust. Quantitative mRNA expression data were acquired and analyzed by the Livak ΔΔ cycle threshold (ΔΔ-Ct) method (Livak and Schmittgen 2001).

Data Decoding

All testing and data collection for the RNA-seq was conducted blind to exposure. All individuals had a unique identifier and the samples belonging to each experimental group (grouped by exposure and sex) were designated with a letter (A, B, C etc.). The blinded raw data were submitted to the NTP Chemical Effects in Biological Systems (CEBS) database. It was then independently verified to account for all expected datasets and data points, and "locked" such that data could not be altered. The NCSU researchers were then provided with the exposure code for data analysis. The qRT-PCR experimental design depended, in part, on the RNA-seq results. Thus, the qRT-PCR phase was performed after the RNA-seq data was decoded but blind to the other, remaining, groups.

Genes for the qRT-PCR analyses were selected either *a priori* because they were previously shown to be altered by developmental BPA exposure in brain (*Esr1, Esr2, Oxt*), sexually dimorphic in PND1 hypothalamus (*Ptgds*) (Nugent, Wright et al. 2015), or identified from the RNA-seq results (*Lepr*, *Slc32a1*, and *Slc1a2)*. Additionally, *Slc32a1* and *Slc1a2* were considered gross indicators of altered

inhibitory and excitatory neurotransmission. The blinded raw data were then submitted to CEBS and "locked" as described for the RNA-seq experiments and analyzed once the data were decoded.

Statistical Analysis

The statistical approach was developed to be consistent with previously published transcriptome projects of similar scale (equivalent sample size or smaller) in rat brain (McCarthy, Auger et al. 2009) and guidelines for low dose endocrine disrupting chemical (EDC) studies with sample sizes in this range (Haseman, Bailer et al. 2001). Within each exposure group, no same-sex litter mates were included, so potential litter effects did not need to be statistically accounted for.

RNA-seq

For all data, the hippocampus and the hypothalamus were analyzed separately. The DEseq2 package was used to fit a negative binomial model for each gene using an extended model matrix. To identify sex differences in gene expression, the male and female controls were compared independent of other exposure groups. With that exception, all other data were compared within sex. For each comparison, the p-value was adjusted for multiple testing using the Benjamini-Hochberg False Discovery Rate (padj) (Benjamini and Hochberg 1995). Some genes could not be corrected because their abundance was too low to justify analysis (padj = N/A). Thus, for all comparisons, statistical significance was defined as padj ≤ 0.05. The group for which the greatest number of genes were significantly altered by BPA was the hypothalamic male 2500 BPA group. Thus, to further query the sex-specificity of these effects, an interaction term in the negative binomial model was fit to examine the nonlinear effect of sex-by-exposure compared to unexposed controls. This approach considerably narrowed the prospective list of significantly altered genes.

qRT-PCR

Statistical analyses were performed and graphed using Prism v6 software (GraphPad Software, Inc., La Jolla, CA). Two samples (one for hypothalamic *Oxt* and one for hypothalamic *Esr2)* were excluded due to technical error. Outliers (no more than one per group) were identified using a Grubb's test. Data from the hippocampus and hypothalamus were analyzed independently and, because gene expression was anticipated to be sexually dimorphic in some cases, analyzed within sex. The study had multiple related but independent hypotheses, each of which was tested independently and included only the

relevant groups for addressing that specific hypothesis. Using a two-tailed Mann-Whitney U test, it was first established if expected sex differences in gene expression between the male and female controls were present. Also using a two-tailed Mann-Whitney U test we next addressed the primary goal of the study: establish if BPA impacts gene expression in either region of interest at any of the doses employed, in either sex. When BPA-related effects were found, qualitative comparisons to the EE groups were made to see if directionality was consistent with an "estrogenic" effect. Finally, EE was used as a reference estrogen and thus expected to masculinize sexually dimorphic gene expression in females. Within sex, differences in mean expression values between the controls and each of the two EE groups were identified using a two-tailed Mann-Whitney U test. In all cases, effects were considered significant at $p \leq 0.05$.

3. JUVENILE BRAIN REGION VOLUME PAPER: Arambula, S. E., Fuchs, J., Cao, J. & Patisaul, H. B. Effects of perinatal bisphenol A exposure on the volume of sexually-dimorphic nuclei of juvenile rats: A CLARITY-BPA consortium study. Neurotoxicology 63, 33-42, doi:10.1016/j.neuro.2017.09.002 (2017).

Materials and Methods

The study is a component of the CLARITY-BPA program and used the same animals for which behavioral data are already published (Rebuli, Camacho et al. 2015).

Tissue Collection, Processing, and Nissl Staining

The animals (n = 120) were sacrificed upon completion of the behavior testing, on PND 28 by CO_2 asphyxiation followed by rapid decapitation. Brains were collected, flash frozen on crushed dry ice and shipped from NCTR to North Carolina State University (NCSU) where they were stored at −80 °C. The brain of each animal was cryosectioned (Leica CM1900, Nußloch, Germany) into three serial sets of 20 µm coronal sections, mounted onto Superfrost plus slides (Fisher Scientific, Pittsburgh, PA) and stored at -80 °C. On the day prior to staining, one set of sections was thawed and dried at room temperature overnight. The sections were then defatted in 100% xylene, rehydrated in a series of descending ethanols and Milli-Q water (Merck Group, Darmstadt, Germany), and stained for Nissl substance with thionin (0.2%) to visualize anatomical structures. The slides were then dehydrated in ascending ethanols, cleared in 100% xylene, and cover-slipped with DPX mounting medium (VWR International Inc., Poole, England).

Stereological Quantification

Unbiased stereology was performed using the Stereologer™ software (Stereology Resource Center, Inc., MD) on a Leica DM2500P microscope (Leica Microsystems, Wetzlar, Germany) equipped with a motorized stage (Applied Scientific Instrumentation, Eugene, OR) and a video camera (IMI Technology Co., Seoul, Korea). Procedures for volumetric assessment were similar to those we have used for prior studies (Patisaul, Fortino et al. 2007, McCaffrey, Jones et al. 2013).

Unilateral contours of the AVPV, SDN, and LC and bilateral contours of the MePD were drawn at low magnification (5x) from the live image with the assistance of a rat and mouse brain atlas (Paxinos 1991, Paxinos 2007). Bilateral measurements of the MePD were taken because there are subtle volumetric differences between the left and the right MePD as well as between the sexes (Cooke, Stokas et al. 2007, Johnson, Breedlove et al. 2008). A uniform grid of points with an area per point of 2000 μm^2 was randomly superimposed over each section and all of the points lying within the region of interest were selected. Based on these counts, volume was estimated using the Cavalieri method (Gundersen and Jensen 1987) and coefficient of error for individual volume estimates was less than 10% ($CE < 0.10$).

The volume of the AVPV and SDN were independently defined and measured by two blinded investigators to confirm that the measurement methodology was reproducible. For the analysis, the data from both investigators was averaged. A single investigator, blinded to exposure groups, then quantified the volume of the LC and MePD. Only animals for which every section within regions of interest were perfectly intact were examined. Thus, because of tissue damage or uneven staining, some material could not be analyzed: 2 brains were excluded for the SDN, 3 for the AVPV, 24 for the left MePD, 20 for the right MePD, and 13 for the LC. The number of animals excluded was higher for the MePD because this region is volumetrically larger than the others and thus requires more sections to fully measure. If even a single section was damaged or missing, then the animal was excluded to prevent measurement error. Sample sizes for all endpoints are indicated in the figures.

Data Decoding

To ensure that all investigators remained blinded during data collection, all tissue samples provided by NCTR were given a unique identifier and designated with a letter (A, B, C, etc.) to denote each experimental group (grouped by exposure and sex) but blind the NCSU team to exposure. Coded raw data were submitted to the NTP Chemical Effects in Biological Systems (CEBS) repository after all volumetric measurements were complete. After a CEBS administrator performed a quality control analysis, the raw data was archived, and the NCSU investigators were then unblinded and began the

statistical analysis. This code was not the same as the one used for these animals while they underwent behavioral testing. Thus, all information obtained from these animals was collected under blinded conditions.

Statistical Analysis

Statistical analysis for all of the data was performed and graphed using Prism version 7 (GraphPad Software, Inc., La Jolla, CA). The statistical approach was designed to be consistent with published guidelines for low dose EDC studies (Haseman, Bailer et al. 2001) and previously published stereological studies of similar scale (equivalent or smaller sample sizes) in the rat brain (Adewale, Todd et al. 2011, McCaffrey, Jones et al. 2013). Within each exposure group, no same-sex littermates were included, so potential litter effects did not need to be statistically accounted for. Because it can differ, the volume of the left and right MePDs were first analyzed individually. Significant differences in size were not observed, thus a combined data set was generated by calculating an average (of left and right) MePD size (for animals in which both could be measured) or by using the single available value (from the left or the right). This produced a single, representative MePD volume for each animal.

Prior to all statistical analysis, data were assessed within each region using the Shapiro-Wilk normality test ($\alpha = 0.05$) and violations were only found for residual groups within the MePD and SDN. In some cases this was due to the presence of statistical outliers, which we ultimately chose not to remove in order to ensure a full accounting of all data. Violations of normality may increase the chance of type I error, but are not uncommon with sample sizes of 9-12/sex/group (Cohen, Cohen et al. 2002). Because occurrences of non-normality or outliers did not meaningfully impact the statistical outcome or interpretation of the data, rather than differentially perform non-parametric tests in cases where deviation from normality occurred, we applied a constant modeling approach to all endpoints in each region of interest.

For each region of interest, all data were first analyzed by a two-way analysis of variance (ANOVA) with sex and exposure as factors to identify significant main effects and their interactions. To maximize resolution regarding potential sex-specific effects, the data were then analyzed within sex by a one-way ANOVA and the Dunnett's Multiple Comparison post hoc test was used to compare each exposure group to the same-sex vehicle control group. Lastly, t-tests were used to identify sex differences and, most importantly, to ensure known sex differences were detected in the unexposed control groups. Confirmation of known sex differences in the vehicle controls was considered to be an indication that the measurements were robust, sufficiently powered to detect a difference in the range

of that effect size, and properly conducted. All analyses were two-tailed and the level of significance for all data was set at $p \leq 0.05$.

4. AMYGDALA PND1 TRANSCRIPTOME PAPER (This was an add-on (unplanned; not in the U01 grant proposal) so aspects of the project used deidentified materials): Arambula, S. E., Jima, D. & Patisaul, H. B. Prenatal bisphenol A (BPA) exposure alters the transcriptome of the neonate rat amygdala in a sex-specific manner: a CLARITY-BPA consortium study. Neurotoxicology, 65, 207-220, doi:10.1016/j.neuro.2017.10.005 (2018).

Materials and Methods

PND 1 pups were obtained from litters produced for the CLARITY-BPA program (Birnbaum, Bucher et al. 2012, Schug, Heindel et al. 2013). Methods for animal husbandry, diet, breeding, dose preparation and administration, and necropsy are described in detail elsewhere (Heindel, Newbold et al. 2015); therefore, only relevant methods are reviewed below. To model the exposure route used to establish the NOAEL, dams were gavaged daily with vehicle (0.3% CMC/kg bw/day), BPA (2.5, 25, 250, 2500, or 25000 μg BPA/kg bw/day), or EE_2 (0.05 or 0.5 μg EE_2/kg bw/day) from GD6 until the day of parturition [postnatal day (PND) 0].

Tissue Collection and Preparation:

On PND 1, pups (one per sex per litter) were weighed and euthanized by rapid decapitation. Heads were collected, snap frozen, and shipped coded (blinded) to the Patisaul lab where they were stored at -80°C until processing. Each whole head was coronally cryosectioned (Leica CM1900, Nussloch, Germany) from the caudal end until the caudal borders of the amygdala were identified. Two sequential bilateral punches, each 1.00 mm in diameter and 1.00 mm in depth, were collected caudally to rostrally; this corresponded with plates 69-75 of the Atlas of the Developing Mouse Brain (Paxinos 1991). All four punches, which collectively comprised the entire amygdala, were combined and stored in BPA-free Eppendorf tubes at -80°C. These punches were collected at the same time we collected hypothalamic and hippocampal studies for a prior, published study (Arambula, Belcher et al. 2016) with the intention of performing the amygdalar assessment as a follow-up (secondary analysis) if any significant observations were found in the other brain regions. The outcomes of that prior study informed the selection of the dose groups and primary genes of interest for the present study.

Quantitative real-time PCR:

Analysis was performed on eight exposure groups (n = 5-7 for the predetermined genes, n = 3-7 for the validation genes; sample size based on availability of cDNA): vehicle, BPA 2.5, 25, 250, 2500, and 25000 and EE_2 0.05 and 0.5. Total RNA was extracted with the Qiagen RNEasy Miniprep kit. An Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip was used to determine RNA purity and concentration and each sample had a RIN of 10. Single-stranded cDNA synthesis was performed with 350 ng of RNA input using the high capacity RNA-to-cDNA kit (Applied Biosystems, Cat. 4387406) and samples were stored at -20°C until use. qRT-PCR was performed as previously published (Arambula, Belcher et al. 2016) using a TaqMan probe-based protocol and detected on a StepOnePlus™ Real-Time PCR System (Applied Biosystems, Life Technologies, Grand Island, NY) with the following cycling parameters: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Each sample was run in triplicates the sample maximization approach was followed to avoid technical and run-to-run variation (Derveaux, Vandesompele et al. 2010). Cycle threshold (Ct) values for the gene of interest were normalized to the Ct for 18s rRNA and relative data were determined by the Livak ΔΔ cycle threshold (ΔΔ-Ct) method (Livak and Schmittgen 2001).

RNAseq Data Analysis:

The experimental design for transcriptome sequencing was developed in consultation with the NCSU Genomic Sciences Laboratory (GSL). Transcriptome sequencing was performed by the GSL on 24 amygdala samples (n = 3 per sex per group). Four experimental groups were examined: vehicle, 25 BPA, 250 BPA, and $0.5EE_2$. RNA extraction was performed with the Qiagen RNEasy Miniprep kit according to the manufacturer protocol (Qiagen, Cat. 74134). Total RNA samples were submitted to the North Carolina State Genomic Sciences Laboratory for Illumina RNA library construction and sequencing. Prior to library construction, RNA integrity, purity, and concentration were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, USA). All samples had an RNA integrity number (RIN) of 10. To optimize library complexity, only samples that had greater than 300 ng of total RNA were used as input material for library preparation.

As previously described (Arambula, Belcher et al. 2016), messenger RNA (mRNA) was purified using the oligo-dT beads provided in the NEBNext Poly (A) mRNA Magnetic Isolation Module (New England Biolabs, Cat. E7490). Complementary DNA (cDNA) libraries for Illumina sequencing were prepared with the NEBNext Ultra Directional RNA Library Prep Kit and the NEBNext Mulitplex Oligos (New England Biolabs, Cat. E7420 and E7335). Briefly, mRNA was isolated, heat fragmented, and primed

with random oligos for first strand cDNA synthesis. Second strand cDNA synthesis was performed with dUTPs to preserve strand orientation information. Next, the double-stranded cDNA fragments were purified using AMPure XP beads (Beckman Coulter Genomics, Cat. A63881), end-repaired, and ligated onto adaptors specifically designed for the Illumina platform. Following ligation, the samples were size-selected to a final library size of 400-550 bp (adapters included) using sequential AMPure XP bead isolation. Protocol-specified PCR amplification was performed to enrich adaptor-ligated cDNA and add specific indexes for each sample. The amplified library fragments were purified and quality and final concentration was assessed using an Agilent 2200 Tapestation. The final quantified cDNA libraries were pooled into equimolar amounts for clustering and sequencing on an Illumina HiSeq 2500 DNA sequencer (4 lanes), utilizing a 125 bp single end sequencing reagent kit (Illumina, USA). Approximately 37.5 million reads were generated per sample. The software package Real Time Analysis (RTA), was used to generate raw bcl, or base call files, which were then de-multiplexed by sample into fastq files for data submission.

Data analysis for RNAseq was performed in consultation with the Bioinformatics Core of the NCSU Center for Human Health and Environment. Sequence data was evaluated with FastQC and 12 poor quality bases were trimmed from the 5'-end. The good quality reads were aligned to the *Rattus norvegicus* (rn6) reference genome (downloaded from UCSC) using the STAR software package (Benjamini and Hochberg, 1995). For each replicate, per-gene counts of uniquely mapped reads were calculated using htseq-count script from the HTSeq python package (McCarthy, Auger et al., 2009). Count data were normalized for sequencing depth and distortion, and dispersion was estimated using the DESeq2 Bioconductor [48, 49] package in the R statistical computing environment. We fit a leaner model using treatment levels and differentially expressed genes were identified after applying multiple testing corrections using the Benjamini-Hochberg procedure (padj<0.05) (Benjamini and Hochberg 1995). Lastly, canonical pathway-based functional analyses of transcriptomic datasets, with an adjusted p-value (padj) less than 0.05, were performed with Ingenuity Pathway Analysis (IPA; QIAGEN). To identify relevant pathways, IPA core analysis was initially filtered to only include annotations made in neurons, astrocytes, or the amygdala. Then the analysis was then re-run to evaluate the robustness and relevance of the main findings in the context of a more global analysis by including annotations made in all tissues types and cell-lines. Associated canonical pathways were generated with the negative log probability of a particular network being enriched due to random chance [-log (p-value)] and the p-values were calculated using a right-tailed Fisher's exact tests.

Data Blinding and Statistical Analysis:

All dosing and related work conducted at NCTR was conducted by investigators blinded to exposure as described in detail in Heindel et al. (2015). The brains were then given a unique identifier and grouped by letter (group A, B, C etc.) and sent to NCSU so that analysis could be done blinded to exposure group and sex. All tissue micropunching was conducted blinded and the samples stored at -80°C until the prior, related study was completed and published (Arambula, Belcher et al. 2016) (which necessitated unblinding). Because the present study was considered a secondary, follow-up study, data from the prior study was critically necessary to inform the present one. Thus, the experimental design was conceived and developed under unblinded conditions. Once the groups for RNAseq were selected, the individual samples were selected at random by an investigator blinded to individual and group ID, and all RNAseq work and bioinformatics was conducted blinded. Because qRT-PCR used all remaining samples (not all samples had sufficient cDNA for analysis), that work could not be done fully blinded, but to minimize risk of bias all of the individual samples were randomized across plates.

The statistical approach was developed to be consistent with previously published transcriptome projects of similar scale (equivalent sample size or smaller) in rat brain (McCarthy, Auger et al. 2009, Arambula, Belcher et al. 2016) and guidelines for low dose endocrine disrupting chemical (EDC) studies (Haseman, Bailer et al. 2001). Within each exposure group, no same-sex littermates were included, so potential litter effects did not need to be statistically accounted for.

qRT-PCR: Statistical analysis for all of the data was performed and graphed using Prism version 7 (GraphPad Software, Inc., La Jolla, CA). For each gene of interest, a Grubb's test for outliers ($\alpha = 0.05$) was conducted and up to one outlier per group was removed. In total, only two outliers were removed from all of the qRT-PCR data: a 2.5 BPA male and a vehicle female from the vasopressin receptor (*Avpr1a*) analysis. Next, a two-tailed Mann-Whitney U test was used to determine if any sex differences in gene expression were detected in the unexposed controls. Finally, a two-tailed Mann-Whitney U test was used to compare each exposure group to the same-sex vehicle control. When BPA-related effects were found, qualitative comparisons to the EE_2 groups were made to see if directionality was consistent with an "estrogenic" effect. In all cases, effects were considered significant at $p \leq 0.05$.

RNAseq: The DEseq2 package was used to fit a negative binomial model for each gene using an extended model matrix. To identify baseline sex differences in gene expression, the male and female controls were compared independent of other exposure groups. With that exception, all other data were compared within sex. For each comparison, the p-value was adjusted for multiple testing using the Benjamini-Hochberg method at a false discovery rate of 5% (Benjamini and Hochberg 1995). A cutoff of

padj ≤ 0.05 was used to select differentially expressed transcripts and genes.

References:

Adewale, H. B., K. L. Todd, J. A. Mickens and H. B. Patisaul (2011). "The impact of neonatal bisphenol-A exposure on sexually dimorphic hypothalamic nuclei in the female rat." Neurotoxicology **32**(1): 38-49.

Anders, S. and W. Huber (2010). "Differential expression analysis for sequence count data." Genome Biol **11**(10): R106.

Anderson, E. E. (1940). "The Sex Hormones and Emotional Behavior: I. The Effect of Sexual Receptivity Upon Timidity in the Female Rat." The Pedagogical Seminary and Journal of Genetic Psychology **56**(1): 149-158.

Arambula, S. E., S. M. Belcher, A. Planchart, S. D. Turner and H. B. Patisaul (2016). "Impact of Low Dose Oral Exposure to Bisphenol A (BPA) on the Neonatal Rat Hypothalamic and Hippocampal Transcriptome: A CLARITY-BPA Consortium Study." Endocrinology: en20161339.

Archer, J. (1975). "Rodent sex differences in emotional and related behavior." Behav Biol **14**(4): 451-479.

Auer, P. L. and R. W. Doerge (2010). "Statistical design and analysis of RNA sequencing data." Genetics **185**(2): 405-416.

Bailey, K. R. and J. N. Crawley (2009). Anxiety-Related Behaviors in Mice. Methods of Behavior Analysis in Neuroscience. J. J. Buccafusco. Boca Raton (FL), CRC Press.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society Series B-Methodological **57**(1): 289-300.

Birnbaum, L. S., J. R. Bucher, G. W. Collman, D. C. Zeldin, A. F. Johnson, T. T. Schug and J. J. Heindel (2012). "Consortium-based science: the NIEHS's multipronged, collaborative approach to assessing the health effects of bisphenol A." Environ Health Perspect **120**(12): 1640-1644.

Birnbaum, L. S., J. R. Bucher, G. W. Collman, D. C. Zeldin, A. F. Johnson, T. T. Schug and J. J. Heindel (2012). "Consortium-based science: the NIEHS's multipronged, collaborative approach to assessing the health effects of bisphenol A." Environmental health perspectives **120**(12): 1640-1644.

Cao, J., M. E. Rebuli, J. Rogers, K. L. Todd, S. M. Leyrer, S. A. Ferguson and H. B. Patisaul (2013). "Prenatal bisphenol A exposure alters sex-specific estrogen receptor expression in the neonatal rat hypothalamus and amygdala." Toxicol Sci **133**(1): 157-173.

Cao, J., M. E. Rebuli, J. Rogers, K. L. Todd, S. M. Leyrer, S. A. Ferguson and H. B. Patisaul (2013). "Prenatal bisphenol A exposure alters sex-specific estrogen receptor expression in the neonatal rat hypothalamus and amygdala." Toxicological sciences **133**(1): 157-173.

Chadman, K. K., M. Yang and J. N. Crawley (2009). "Criteria for validating mouse models of psychiatric diseases." Am J Med Genet B Neuropsychiatr Genet **150B**(1): 1-11.

Churchwell, M. I., L. Camacho, M. M. Vanlandingham, N. C. Twaddle, E. Sepehr, K. B. Delclos, J. W. Fisher and D. R. Doerge (2014). "Comparison of life-stage-dependent internal dosimetry for bisphenol A, ethinyl estradiol, a reference estrogen, and endogenous estradiol to test an estrogenic mode of action in Sprague Dawley rats." Toxicol Sci **139**(1): 4-20.

Cohen, J., P. Cohen, G. W. Stephen and L. S. Aiken (2002). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Inc.

Cooke, B. M., M. R. Stokas and C. S. Woolley (2007). "Morphological sex differences and laterality in the prepubertal medial amygdala." J Comp Neurol **501**(6): 904-915.

Delclos, K. B., L. Camacho, S. M. Lewis, M. M. Vanlandingham, J. R. Latendresse, G. R. Olson, K. J. Davis, R. E. Patton, G. Gamboa da Costa, K. A. Woodling, M. S. Bryant, M. Chidambaram, R. Trbojevich, B. E. Juliar, R. P. Felton and B. T. Thorn (2014). "Toxicity evaluation of bisphenol A administered by gavage to Sprague Dawley rats from gestation day 6 through postnatal day 90." Toxicol Sci **139**(1): 174-197.

Derveaux, S., J. Vandesompele and J. Hellemans (2010). "How to do successful gene expression analysis using real-time PCR." Methods **50**(4): 227-230.

Diaz-Veliz, G., T. Alarcon, C. Espinoza, N. Dussaubat and S. Mora (1997). "Ketanserin and anxiety levels: influence of gender, estrous cycle, ovariectomy and ovarian hormones in female rats." Pharmacol Biochem Behav **58**(3): 637-642.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Doerge, D. R., N. C. Twaddle, M. Vanlandingham, R. P. Brown and J. W. Fisher (2011). "Distribution of bisphenol A into tissues of adult, neonatal, and fetal Sprague-Dawley rats." Toxicol Appl Pharmacol **255**(3): 261-270.

Faul, F., E. Erdfelder, A. Buchner and A. G. Lang (2009). "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses." Behav Res Methods **41**(4): 1149-1160.

Ferguson, S. A. and K. J. Berry (2010). "Chronic oral treatment with isotretinoin alters measures of activity but not anxiety in male and female rats." Neurotoxicol Teratol **32**(5): 573-578.

Ferguson, S. A. and S. Y. Boctor (2010). "Cocaine responsiveness or anhedonia in rats treated with methylphenidate during adolescence." Neurotoxicol Teratol **32**(4): 432-442.

Ferguson, S. A., C. D. Law and J. S. Abshire (2012). "Developmental treatment with bisphenol A causes few alterations on measures of postweaning activity and learning." Neurotoxicol Teratol **34**(6): 598-606.

Frye, C. A., S. M. Petralia and M. E. Rhodes (2000). "Estrous cycle and sex differences in performance on anxiety tasks coincide with increases in hippocampal progesterone and 3alpha,5alpha-THP." Pharmacol Biochem Behav **67**(3): 587-596.

Goma, M. and A. Tobena (1978). "Reliability of various measures obtained in open-field test." Psychol Rep **43**(3 Pt 2): 1123-1128.

Gould, T. D. D., D. T.; Kovacsics, C. E. (2010). The Open Field Test. Mood and Anxiety Related Phenotypes in Mice. T. D. Gould, Humana Press.

Gundersen, H. J. and E. B. Jensen (1987). "The efficiency of systematic sampling in stereology and its prediction." J Microsc **147**(Pt 3): 229-263.

Haseman, J. K., A. J. Bailer, R. L. Kodell, R. Morris and K. Portier (2001). "Statistical issues in the analysis of low-dose endocrine disruptor data." Toxicol Sci **61**(2): 201-210.

Heindel, J. J., R. R. Newbold, J. R. Bucher, L. Camacho, K. B. Delclos, S. M. Lewis, M. Vanlandingham, M. I. Churchwell, N. C. Twaddle, M. McLellen, M. Chidambaram, M. Bryant, K. Woodling, G. G. Costa, S. A. Ferguson, J. Flaws, P. C. Howard, N. J. Walker, R. T. Zoeller, J. Fostel, C. Favaro and T. T. Schug (2015). "NIEHS/FDA CLARITY-BPA research program update." Reprod Toxicol **58**: 33-44.

Heindel, J. J., R. R. Newbold, J. R. Bucher, L. Camacho, K. B. Delclos, S. M. Lewis, M. Vanlandingham, M. I. Churchwell, N. C. Twaddle, M. McLellen, M. Chidambaram, M. Bryant, K. Woodling, G. Gamboa da Costa, S. A. Ferguson, J. Flaws, P. C. Howard, N. J. Walker, R. T. Zoeller, J. Fostel, C. Favaro and T. T. Schug (2015). "NIEHS/FDA CLARITY-BPA research program update." Reprod Toxicol **58**: 33-44.

Hogg, S. (1996). "A review of the validity and variability of the elevated plus-maze as an animal model of anxiety." Pharmacol Biochem Behav **54**(1): 21-30.

Johnson, R. T., S. M. Breedlove and C. L. Jordan (2008). "Sex differences and laterality in astrocyte number and complexity in the adult rat medial amygdala." J Comp Neurol **511**(5): 599-609.

Johnson, S. A., A. B. Javurek, M. S. Painter, M. R. Ellersieck, T. H. Welsh, Jr., L. Camacho, S. M. Lewis, M. M. Vanlandingham, S. A. Ferguson and C. S. Rosenfeld (2016). "Effects of developmental exposure to bisphenol A on spatial navigational learning and memory in rats: A CLARITY-BPA study." Horm Behav **80**: 139-148.

Lewis, S. M., F. W. Lee, A. A. Ali, W. T. Allaben, C. C. Weis and J. E. Leakey (2010). "Modifying a displacement pump for oral gavage dosing of solution and suspension preparations to adult and neonatal mice." Lab Anim (NY) **39**(5): 149-154.

Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

McCaffrey, K. A., B. Jones, N. Mabrey, B. Weiss, S. H. Swan and H. B. Patisaul (2013). "Sex specific impact of perinatal bisphenol A (BPA) exposure over a range of orally administered doses on rat hypothalamic sexual differentiation." Neurotoxicology **36**: 55-62.

McCarthy, M. M., A. P. Auger, T. L. Bale, G. J. De Vries, G. A. Dunn, N. G. Forger, E. K. Murray, B. M. Nugent, J. M. Schwarz and M. E. Wilson (2009). "The epigenetics of sex differences in the brain." J Neurosci **29**(41): 12815-12823.

Mora, S., N. Dussaubat and G. Diaz-Veliz (1996). "Effects of the estrous cycle and ovarian hormones on behavioral indices of anxiety in female rats." Psychoneuroendocrinology **21**(7): 609-620.

Nugent, B. M., C. L. Wright, A. C. Shetty, G. E. Hodes, K. M. Lenz, A. Mahurkar, S. J. Russo, S. E. Devine and M. M. McCarthy (2015). "Brain feminization requires active repression of masculinization via DNA methylation." Nat Neurosci **18**(5): 690-697.

Padilla, E., D. Barrett, J. Shumake and F. Gonzalez-Lima (2009). "Strain, sex, and open-field behavior: factors underlying the genetic susceptibility to helplessness." Behav Brain Res **201**(2): 257-264.

Patisaul, H. B. and H. L. Bateman (2008). "Neonatal exposure to endocrine active compounds or an ERbeta agonist increases adult anxiety and aggression in gonadally intact male rats." Horm Behav **53**(4): 580-588.

Patisaul, H. B., A. Blum, J. R. Luskin and M. E. Wilson (2005). "Dietary soy supplements produce opposite effects on anxiety in intact male and female rats in the elevated plus-maze." Behav Neurosci **119**(2): 587-594.

Patisaul, H. B., A. E. Fortino and E. K. Polston (2007). "Differential disruption of nuclear volume and neuronal phenotype in the preoptic area by neonatal exposure to genistein and bisphenol-A." Neurotoxicology **28**(1): 1-12.

Patisaul, H. B., A. W. Sullivan, M. E. Radford, D. M. Walker, H. B. Adewale, B. Winnik, J. L. Coughlin, B. Buckley and A. C. Gore (2012). "Anxiogenic effects of developmental bisphenol a exposure are associated with gene expression changes in the juvenile rat amygdala and mitigated by soy." PLoS One **7**(9): e43890.

Paxinos, G. (1991). Atlas of the developing rat brain. San Diego, Academic Press.

Paxinos, G. (2007). Atlas of the developing mouse brain at E17.5, P0 and P6. Amsterdam ; Boston, Elsevier.

Pellow, S., P. Chopin, S. E. File and M. Briley (1985). "Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat." J Neurosci Methods **14**(3): 149-167.

Rebuli, M. E., L. Camacho, M. E. Adonay, D. M. Reif, D. L. Aylor and H. B. Patisaul (2015). "Impact of Low-Dose Oral Exposure to Bisphenol A (BPA) on Juvenile and Adult Rat Exploratory and Anxiety Behavior: A CLARITY-BPA Consortium Study." Toxicol Sci **148**(2): 341-354.

Rebuli, M. E., L. Camacho, M. E. Adonay, D. M. Reif, D. L. Aylor and H. B. Patisaul (2015). "Impact of Low-Dose Oral Exposure to Bisphenol A (BPA) on Juvenile and Adult Rat Exploratory and Anxiety Behavior: A CLARITY-BPA Consortium Study." Toxicological Sciences **148**(2): 341-354.

Rebuli, M. E., J. Cao, E. Sluzas, K. B. Delclos, L. Camacho, S. M. Lewis, M. M. Vanlandingham and H. B. Patisaul (2014). "Investigation of the effects of subchronic low dose oral exposure to bisphenol A (BPA)

and ethinyl estradiol (EE) on estrogen receptor expression in the juvenile and adult female rat hypothalamus." Toxicol Sci **140**(1): 190-203.

Schug, T. T., J. J. Heindel, L. Camacho, K. B. Delclos, P. Howard, A. F. Johnson, J. Aungst, D. Keefe, R. Newbold, N. J. Walker, R. Thomas Zoeller and J. R. Bucher (2013). "A new approach to synergize academic and guideline-compliant research: the CLARITY-BPA research program." Reproductive toxicology **40**: 35-40.

Schug, T. T., J. J. Heindel, L. Camacho, K. B. Delclos, P. Howard, A. F. Johnson, J. Aungst, D. Keefe, R. Newbold, N. J. Walker, R. Thomas Zoeller and J. R. Bucher (2013). "A new approach to synergize academic and guideline-compliant research: the CLARITY-BPA research program." Reprod Toxicol **40**: 35-40.

Shepherd, J. K., S. S. Grewal, A. Fletcher, D. J. Bill and C. T. Dourish (1994). "Behavioural and pharmacological characterisation of the elevated "zero-maze" as an animal model of anxiety." Psychopharmacology (Berl) **116**(1): 56-64.

Team, R. C. (2014). R: A language and environment for statistical computing. R. F. f. S. Computing. Vienna, Austria.

Thigpen, J. E., K. D. Setchell, G. E. Kissling, J. Locklear, G. F. Caviness, T. Whiteside, S. M. Belcher, N. M. Brown, B. J. Collins, F. B. Lih, K. B. Tomer, E. Padilla-Banks, L. Camacho, F. G. Adsit and M. Grant (2013). "The estrogenic content of rodent diets, bedding, cages, and water bottles and its effect on bisphenol A studies." J Am Assoc Lab Anim Sci **52**(2): 130-141.

Thigpen, J. E., K. D. Setchell, E. Padilla-Banks, J. K. Haseman, H. E. Saunders, G. F. Caviness, G. E. Kissling, M. G. Grant and D. B. Forsythe (2007). "Variations in phytoestrogen content between different mill dates of the same diet produces significant differences in the time of vaginal opening in CD-1 mice and F344 rats but not in CD Sprague-Dawley rats." Environ Health Perspect **115**(12): 1717-1726.

Walf, A. A. and C. A. Frye (2007). "The use of the elevated plus maze as an assay of anxiety-related behavior in rodents." Nat Protoc **2**(2): 322-328.