Supporting Information

For

A combined transcriptomic and proteomic approach to identify and predict toxicity pathways in early-life stages of Japanese medaka (*Oryzias latipes*) exposed to 1,2,5,6-tetrabromocyclooctane (TBCO)

Jianxian Sun[a], Song Tang[b], Hui Peng*[,a], David M.V. Saunders[a], Jon A. Doering[a], Markus Hecker[a,b], Paul D. Jones[a,b], John P. Giesy[a,c,d,e,f], and Steve Wiseman*[,a]

[a] Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5B3, Canada
[b] School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5C8, Canada
[c] Department of Veterinary Biomedical Sciences, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5B4, Canada
[d] Zoology Department, Center for Integrative Toxicology, Michigan State University, East Lansing, Michigan 48824, USA
[e] School of Biological Sciences, University of Hong Kong, Hong Kong Special Administrative Region 999077, People's Republic of China
[f] State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210093, People's Republic of China

***Corresponding author:** Hui Peng, huisci@gmail.com, Tel.: +1 306 966 8733; Fax: +1 306 966 4796. Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5B3, Canada; Steve Wiseman: steve.wiseman@uleth.ca, Tel: (403) 329-2320, Fax: (403) 329-2082. Department of Biology, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada.

Words	1,798
Tables	1
Figure	3

This supporting information provides text of details, tables and figures addressing (1) RNA sequencing; (2) Protein preparation and identification; (3) Identification of proteins by mass spectrometry; (4) Analysis of TBCO; (5) Visual performance assessment; (6) Cardiac function assessment; (7) Concentrations of TBCO in exposure media prior to media-change; (8) Structures of HBCD and TBCO; (9) Heatmap of significantly altered transcripts and proteins in embryos exposed to control and 100 μg/L TBCO; (10) Biplots illustrating diatance-based redundancy analysis (dbRDA) of transcriptome and proteome, and regression between transcripts and proteins from control samples.

RNASeq. To determine responses of the transcriptome to TBCO, 5 embryos hatched at day 8 post-fertilization were selected randomly from each petri dish and were pooled. Total RNA was extracted from each of the seven pools of embryos exposed to the solvent control and medium concentration of TBCO by use of a RNeasy Plus Mini Kit according to the protocol provided by the manufacturer (Qiagen, Mississauga, ON, Canada). Concentrations of RNA were determined by use of a NanoDrop ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). Integrity of RNA was determined by use of an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Three samples of RNA from medaka exposed to the solvent control (n=3) and three samples of RNA from medaka exposed to the medium concentration of TBCO (n=3), each of which had the highest RNA integrity (RIN) value (greater than 8), were selected for sequencing. One RNA-Seq library per sample was prepared by use of the Tru-Seq RNA Sample Prep Kit (*Illumina*, San Diego, CA, USA), according to the protocol provided by the manufacturer. Quality of libraries was confirmed by use of a 2100 Bioanalyzer Agilent Technologies). Each library was loaded onto a separate MiSeq v3 150 cycle cartridge (*Illumina*) and run as 2×75 base-pair (bp) paired-end reads on a MiSeq sequencer (*Illumina*) at the Toxicology Centre (University of Saskatchewan, Saskatoon, SK, Canada).

Sequencing reads were assessed for quality using FastQC (v0.52) on the Galaxy public server (http://galaxy-qld.genome.edu.au).[1] A median of 99.9% (range from 83.3 to 99.9%) of reads had a Phred score of > 30 across all samples. Various FASTQ quality formats were converted by FASTQ Groomer (v1.0.4). Reads were mapped independently for each sample against the reference genome of medaka[2] (Ensembl, v1.78) by use of TopHat2 (v0.6)[3] with

default settings on Galaxy. Across all samples, 80.4%-86.8% of reads were mapped to the reference genome. Alignments from TopHat2 were provided to Cufflinks (v2.02) to assemble transcript for each condition using a reference annotation as a guide.[4] Cuffmerge (v0.0.6) was used to combine all assemblies into a consensus assembly in GTF file format, which provided a uniform basis for calculating changes in abundances of transcripts in each condition. The reads and the merged assembly were fed to Cuffdiff (v0.0.7), and expression levels were calculated and statistical significance of changes in abundances of transcripts was determined.

Protein preparation and identification. Pools of 50 embryos per exposure were homogenized on ice in lysis buffer (20 mM HEPES, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 420 mM NaCl, 20% Glycerol, and protease inhibitor cocktail, pH=7.4) by use of a model-100 sonic dismembrator (Thermo Fisher Scientific, Waltham, MA, USA). After centrifugation at 15,000 × g for 15 min at 4 °C, the supernatant was recovered and total concentration of proteins were quantified by use of the Bradford protein assay with bovine serum albumin (BSA) as a standard.[5] Filter-aided sample preparation (FASP) was used for protein digestion by use of a 30 kDa molecular weight cutoff spin ultrafiltration filter. Briefly, an aliquot of each lysate, containing approximately 100 μg of protein, was transferred to a YM-30 microcon filter unit (Cat No. MRCF0R030, Millipore, Etobicoke, ON, Canada). After centrifugation at 14,000 × g for 30 min, salt and other interferences in the buffer were removed from samples by washing three times with 0.1 M Tris-HCl. Proteins were reduced with 50 μL of 5 mM DTT for 60 min at 37 ^{o}C, and then carboxymethylated for 30 min in the dark, with 15 mM of iodoacetamide. Samples were further digested overnight with 5 μg

trypsin and gentle shaking. Digestion was terminated by adding formic acid to a final concentration of 1% (v/v). The final sample was collected by centrifugation at 14,000 × g for 30 min.

Identification of Proteins by Mass Spectrometry. Each sample was loaded onto a 75 mm inner diameter fused silica microcapillary column (Polymicron Technologies, Phoenix, AZ, USA) packed with 10 cm of Luna 3-mC18 100 Å, reversed phase particles (Phenomenex, Torrance, CA, USA) and placed in-line with a nano-LC-electrospray ion source (Proxeon, Mississauga, ON, USA) interfaced to a LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific) at the University of Toronto (Toronto, ON, Canada). The organic gradient was driven by the EASY-nLC system at 300 nL/min. Mobile phases were (A) 95% acetonitrile with 0.1% formic acid and (B) 5% acetonitrile. Initially, 2% of B was increased to 6% in 2 min, then increased to 24% in 62 min, followed by an increase to 90% in 26 min and held static for 5 min, and then decreased to initial conditions of 2% of B and held for 8 min for equilibration. Positive precursor ions (400 - 2000 *m/z*) were subjected to data-dependent collision-induced dissociation as the instrument cycled through one full mass scan at 60,000 full-width at half maximum followed by 17 successive MS / MS scans targeting the most intense precursors with dynamic exclusion and +2 / +3 charge state selection enabled. The MS proteomics data have been deposited to the ProteomeXchange Consortium[6] via the Proteomics Identifications (PRIDE) partner repository with the dataset identifier PXD003823.

Raw MS files were analyzed by use of MaxQuant software (v1.5.1.2).[7] MS/MS spectra were searched against the database for medaka protein (UniProt) containing forward and

reversed (decoy) sequences, allowing for variable modifications of methionine oxidation and N-terminal acetylation and fixed cysteine carbamidomethylation. Parent mass and fragment ions were matched using a maximal initial mass deviation of 7 p.p.m and 0.5 Th, respectively. The false discovery rate for proteins was set to 0.01. Spectral count was used for label-free quantification of proteins and only proteins with spectral counts greater than 4 were used for data analysis. Proteins were defined as being differentially expressed if the fold-change in abundance was ±1.5 fold, and if the *p*-value was less than 0.05; as conducted previously.[8] Details of significantly changed transcripts and proteome are provided in Supporting Data Sheet.

Analysis of TBCO. Concentrations of TBCO in embryos and in exposure media were quantified. To determine concentrations of TBCO in unhatched embryos/eggs, 3 eggs that were not hatched were selected randomly from each petri dish on days 1, 3, 5, 7 and 9, and eggs from two petri dishes were pooled as one sample so that the number of replicates analyzed per treatment was 3 (n=3). Also on day 9 of the exposure, three embryos that had hatched also were collected and embryos from 2 petri dishes were pooled as one sample so that the number of replicates was 3. F-BDE-47 was spiked to samples of embryos and media as an internal control. Eggs were crushed and 4 mL of nano pure water was added to each sample. Liquid-liquid extraction was conducted twice with a solution of *n*-hexane/methyl tert-butyl ether (1:1 v/v) by shaking for 20 min at 300 × rpm, followed by sonication for 20 min in a 15 mL centrifuge tube and centrifugation for 15 min at 4000 × g. Both supernatants were pooled, evaporated to dryness under a gentle stream of nitrogen, and redissolved in 1

mL of toluene for analysis (approximately 200× dilution). Three laboratory blanks and matrix spikes were extracted for quality assurance purposes. Extracts were analyzed for TBCO by use of an Agilent (Santa Clara, CA, USA) 7890A gas chromatograph (GC) system coupled to an Agilent 5975C mass spectrometer (MS) operating in electron impact ionization mode (EI).

One μL of each sample was injected at an injection port temperature of 280 °C in the splitless mode. Chromatographic separation was achieved with a 15 m × 250 μm i.d. Rtx-1614 fused silica capillary GC column, which had a 0.1 μm film thickness (Restek Corporation, Bellefonte, PA, USA). The carrier gas (helium) was kept at a constant flow rate of 1.5 mL/min. The following GC oven temperature program was used: 100 °C for 1 min, 5 °C/min to 190 °C for 2 min, 20 °C/min to 220 °C for 2 min, and 40 °C/min to 300 °C for 4 min. The GC/MS transfer line was maintained at 280°C. Selected ion monitoring of *m/z* 267/187 and 343/234 was used for quantification/confirmation of TBCO and F-BDE-47, respectively. TBCO was quantified by use of the internal standard method using F-BDE-47. Concentration of TBCO in blanks was less than the limit of detection. Limit of detection of TBCO were 0.12 μg/L and 0.072 μg/g ww in exposure medium and embryos, respectively. The mean and standard error for recovery of TBCO were 95.9 ± 22.0%.

Visual Performance Assessment. To assess adverse effects of TBCO on visual performance, new batches of embryos were exposed to solvent control, 10, 100 and 1,000 μg/L TBCO from 2 hours post-fertilization until 14 days post-hatch. There were four petri dishes for each concentration of TBCO and there with 100 embryos per dish. Visual performance of medaka exposed to TBCO was determined by use of a background color preference assay that was

developed to assess visual performance of *Xenopus laevis*.[14] For the assay, a transparent tank was placed inside another tank that was half black and half white to provide contrasting background colors. For the assay, groups of 20 larvae per replicate were gently moved into the inner tank and allowed to acclimate for 10 s, and then the inner tank was placed into the outer tank. Behaviour of fry was observed for 30 s. Then the inner tank was lifted and the outer tank was rotated 180° to change the position of white and black background so that fry were positioned over the black area of the outer tank. Behaviour of fry were recorded with a camera for another 30 s. Percentages of fry in the black and white areas were determined, as was the time for all fry to move to the white side of the tank. Swimming routs were traced by use of ZooTracer software (Microsoft ResearchCambridge, MA, USA).[9] Suitability of the assay for assessing visual function was first optimized with control fry. Similarly to *Xenopus laevis*, fry of Japanese medaka exposed to the solvent control swam from the side of the tank that overlayed the black backgrouns to the side of the tank that overlayed the white background (96.3 ± 5.2%) within a short period of time (average = 2.7 s).

Cardiac Function Assessment. To assess adverse effects of TBCO on heartbeat, new batches of embryos were exposed to solvent control, 10, 100 and 1000 µg/L TBCO from 2 hours post-fertilization until 5 days post-fertilization. There were four petri dishes for each concentration (n=4) and there were 100 embryos per dish. Effects of TBCO on cardiac function were determined by quantifying heart rate of embryos viewed at 4× magnification by use of an inverted microscope (Zeiss, Toronto, ON, Canada). The number of beats per minute was determined in 3 embryos from each of the four petri dishes per exposure (n=4).

Table S1. Concentrations of TBCO in exposure media prior to media-change. Approximately 5 ml of solution was sampled on days 2, 4 and 6 from each of the 7 petri dishes per exposure group. Data represent mean ± standard deviation.

	Nominal Concentration (μg/L)	Measured Concentration (μg/L)
Control	0	ND
Low	10	2.7±0.4
Medium	100	29.6±7.5
High	1000	151.3±81.9

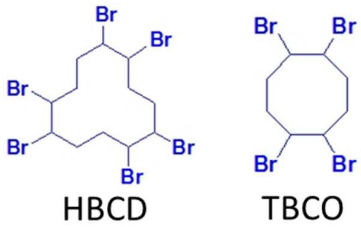

Figure S1. Structures of hexabromocyclododecane (HBCD) and 1,2,5,6-tetrabromocyclooctane (TBCO).

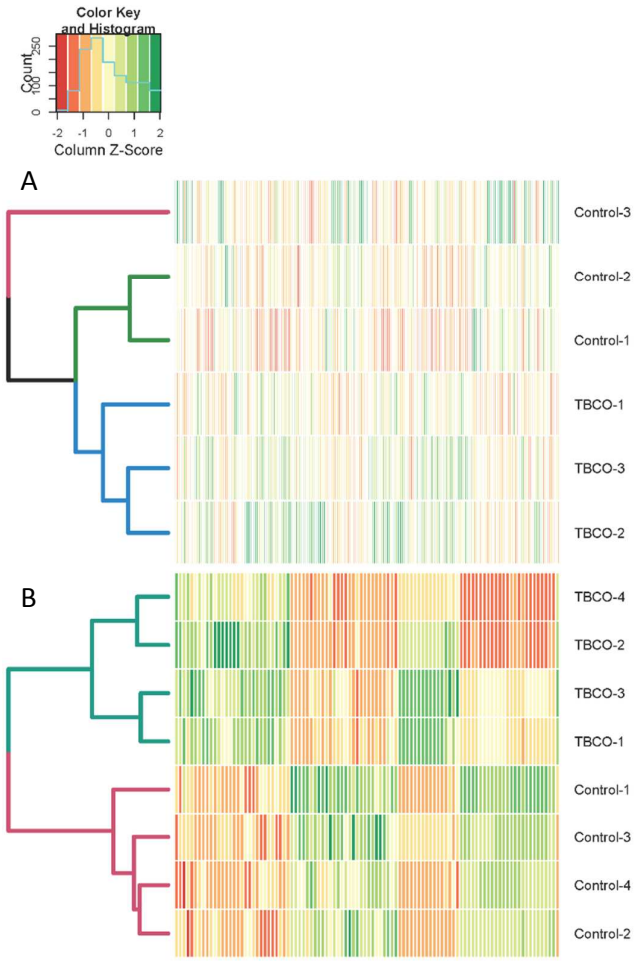

Figure S2. Heatmap of significantly altered transcripts (A) and proteins (B) in embryos exposed to the solvent control and 100 µg/L TBCO.

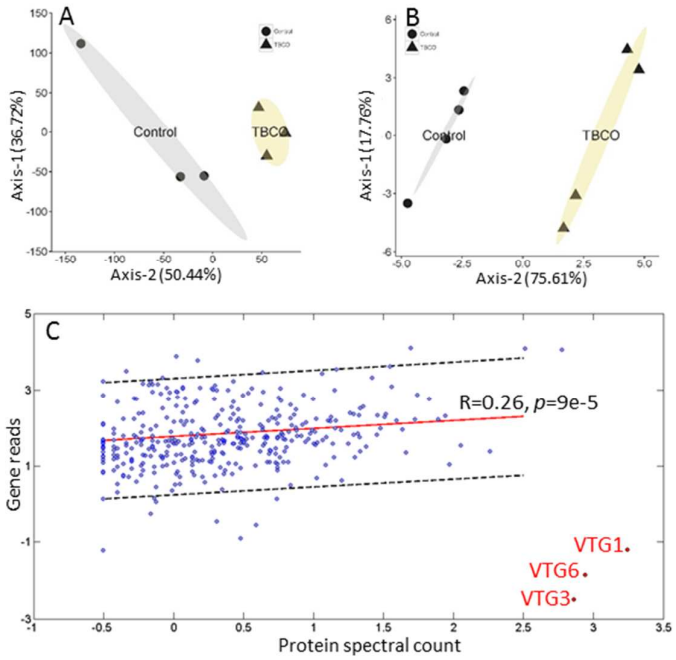

Figure S3. Biplots illustrating diatance-based redundancy analysis (dbRDA) of measured responses of (A) transcriptome (n=3) and (B) proteome (n=4) of embryos of Japanese medaka exposed to the solvent control or 100 µg/L of TBCO. Each data point represents one sample, and proximity of points in the biplots is an approximation of the similarity of samples with respect to their response. Ovals represent the 95% confidence ellipse around the group centroids. In (A), 50.44% of variation in measured responses of the transcriptome is accounted for by the X-axis and 36.72% of variation is accounted for by the Y-axis; In (B), 75.61% of variation in measured responses of the proteome is accounted for by the X-axis and 17.76% of variation is accounted for by the Y-axis. Regression between transcripts and proteins (C) in embryos exposed to the solvent control. Black dashed lines indicate 95% confidence limits.

References

(1) Goecks, J.; Nekrutenko, A.; Taylor, J., Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **2010,** *11*, (8), R86.

(2) Kasahara, M.; Naruse, K.; Sasaki, S.; Nakatani, Y.; Qu, W.; Ahsan, B.; Yamada, T.; Nagayasu, Y.; Doi, K.; Kasai, Y., The medaka draft genome and insights into vertebrate genome evolution. *Nature* **2007,** *447*, (7145), 714-719.

(3) Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S. L., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **2013,** *14*, (4), R36.

(4) Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; Salzberg, S. L.; Rinn, J. L.; Pachter, L., Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. protoc.* **2012,** *7*, (3), 562-578.

(5) Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976,** *72*, (1), 248-254.

(6) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014,** *32*, (3), 223-226.

(7) Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008,** *26*, (12), 1367-1372.

(8) Liu, X.; Hu, Y.; Pai, P.-J.; Chen, D.; Lam, H., Label-free quantitative proteomics analysis of antibiotic response in *Staphylococcus aureus* to oxacillin. *Journal of proteome research* **2014,** *13*, (3), 1223-1233.

(9) A. Buchanan, A. F., Zoo tracer. 〈*http://research.microsoft.com/zootracer*〉 **2014**.