Screening for neurotoxic potential of 15 flame retardants using freshwater planarians

Siqi Zhang[a]*, Danielle Ireland[be]*, Nisha S. Sipes[c], Mamta Behl[c] and Eva-Maria S. Collins[bde]

[a] Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

[b] Division of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92093, USA

[c] Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC, 27709, USA

[d] Department of Physics, University of California San Diego, La Jolla, CA 92093, USA

[e] Department of Biology, Swarthmore College Swarthmore, PA 19081, USA

* These authors contributed equally to this work.

Correspondence should be addressed to Eva-Maria S. Collins; e-mail: ecollin3@swarthmore.edu;

Mailing address: Biology Department, Swarthmore College, 500 College Ave, Swarthmore, PA 19081

Telephone: 610-690-5380; Fax: (610) 328-8663

1   **Abstract**

2   Asexual freshwater planarians are an attractive invertebrate model for high-throughput

3   neurotoxicity screening, because they possess multiple quantifiable behaviors to assess distinct

4   neuronal functions. Planarians uniquely allow direct comparisons between developing and adult

5   animals to distinguish developmentally selective effects from general neurotoxicity. In this

6   study, we used our automated planarian screening platform to compare the neurotoxicity of 15

7   flame retardants (FRs), consisting of representative phased-out brominated (BFRs) and

8   replacement organophosphorus FRs (OPFRs). OPFRs have emerged as a proposed safer

9   alternative to BFRs; however, limited information is available on their health effects. We found

10   11 of the 15 FRs (3/6 BFRs, 7/8 OPFRs, and Firemaster 550) caused adverse effects in both

11   adult and developing planarians with similar nominal lowest-effect-levels for BFRs and OPFRs.

12   This suggests that replacement OPFRs are comparably neurotoxic to the phased-out compounds.

13   BFRs were primarily systemically toxic, whereas OPFRs, except Tris(2-chloroethyl) phosphate ,

14   shared a behavioral phenotype in response to noxious heat at sublethal concentrations, indicating

15   specific neurotoxic effects. By directly comparing effects on adult and developing planarians, we

16   further found that one BFR (3,3',5,5'-Tetrabromobisphenol A) caused a developmental selective

17   defect. Together, these results demonstrate that our planarian screening platform yields high

18   content data resulting from assaying various behavioral and morphological endpoints, allowing

19   us to distinguish selective neurotoxic effects and effects specific to the developing nervous

20   system. Ten of these 11 bioactive FRs were previously found to be bioactive in other models,

21   including cell culture and alternative animal models (nematodes and zebrafish). This level of

22   concordance across different platforms emphasizes the urgent need for further evaluation of

23   OPFRs in mammalian systems.

1

2   **Keywords**: alternative animal models, planarian, flame retardants, OPFRs, polybrominated

3   diphenyl ethers, developmental neurotoxicity

4

5   **Abbreviations**

6   DNT, developmental neurotoxicity; FR, flame retardant; BFR, brominated flame retardant;

7   PBDE, Polybrominated Diphenyl Ether; OPFR, organophosphorus flame retardant; NTP,

8   National Toxicology Program; DMSO, dimethyl sulfoxide; HTS, high-throughput screening;

9   LEL, lowest effect level; HTTK, high-throughput toxicokinetics; POD, point of departure

10

11   **1. Introduction**

12   Traditional hazard assessment using mammalian models cannot keep up with the speed of

13   chemical development. The Toxic Substances Control Act Inventory lists approximately 68,000

14   chemicals currently in use in the United States (https://www.epa.gov/tsca-inventory), but <2%

15   have been evaluated for developmental neurotoxicity (DNT) (Smirnova et al., 2014). *In vitro*

16   testing combined with bioinformatic analyses have emerged as a powerful tool to fill this need

17   for more rapid chemical testing (Fritsche et al., 2018), but this approach faces limitations for

18   predicting adverse effects on neuronal function (Smirnova et al., 2014; Tohyama, 2016). Non-

19   mammalian animal models, amenable to high-throughput, cost-effective screening, are a

20   promising alternative solution for efficient DNT screening since they allow for automated

21   behavioral assays to quantitatively evaluate different neuronal functions (Bal-Price et al., 2012).

22   Flame retardants (FRs), widely added to commercial products for fire safety (Costa and

23   Giordano, 2007; Salimi et al., 2017), are a class of chemicals which only have limited

3

1   mammalian data for safety assessment. Prior to 2005, brominated FRs (BFRs) such as

2   polybrominated diphenyl ether (PBDE) mixtures were the primary FRs used in the United States

3   (Hale et al., 2003). Since then, many were voluntarily phased-out due to growing evidence of

4   their associations with impaired neurodevelopment and decreased fertility, exacerbated by their

5   persistence in the environment and ability to bioaccumulate (Darnerud, 2003; Herbstman et al.,

6   2010; Stapleton et al., 2011; Talsness, 2008). Organophosphorus FRs (OPFRs) have gradually

7   become replacements for PBDEs over the last decade (Stapleton et al., 2012; Stapleton et al.,

8   2014). Concerns exist that OPFRs, similar to PBDEs, may persist in the environment and

9   bioaccumulate, but little is known about their potential toxicity (Hendriks and Westerink, 2015;

10   Meeker and Stapleton, 2010; Stapleton et al., 2009; Van Der Veen and De Boer, 2012).

11   Particularly concerning is the potential DNT of OPFRs, as they share structural similarities to

12   organophosphorus pesticides, which adversely affect neurodevelopment and cause

13   neurobehavioral impairments (González-Alzaga et al., 2014; Muñoz-Quezada et al., 2013;

14   Ricceri et al., 2006; Slotkin et al., 2006). Previous work using a screening battery of *in vitro* and

15   alternative animal models found that several OPFRs had comparable toxicity and potency to

16   BFRs, including phased-out BDE-47 (Behl et al., 2015; Behl et al., 2016; Jarema et al., 2015),

17   indicating that OPFRs may not be the safer alternative to BFRs that they have been marketed to

18   be.

19   Because the freshwater planarian *Dugesia japonica* is a good model to study

20   organophosphorus pesticide neurotoxicity (Hagstrom et al., 2017; Hagstrom et al., 2018a; Zhang

21   et al., 2018) and uniquely allows parallel screening of adult and developing animals to identify

22   developmental-specific effects (Zhang et al., 2018), we herein evaluate the potential

23   neurotoxicity and DNT of 15 FRs using a custom planarian high-throughput screening (HTS)

1 platform. These FRs include 3 phased-out PBDEs (BDE-47, BDE-99, and BDE-153), 3 currently

2 in-use BFRs (TBB, TBBPA, and TBPH), 8 in-use OPFRs (BPDP, EHDP, IDDP, IPP, TCEP,

3 TCPP, TMPP, and TPHP), and one BFR mixture (FM550).

4   Because *D. japonica* is a fairly new model for neurotoxicity studies, we briefly

5 summarize its key features to familiarize the reader and refer to our recent review paper

6 (Hagstrom et al., 2016) for an in-depth review of planarians as an alternative model for DNT. In

7 asexual *D. japonica* planarians, neurodevelopment is solely achieved through neuroregeneration

8 following asexual reproduction by fission, allowing neurodevelopment to be induced by head

9 amputation. In addition, the similar sizes of adult and regenerating/developing planarians allows

10 for both worm types to be screened in parallel with the same assays (Hagstrom et al., 2015;

11 Hagstrom et al., 2016; Zhang et al., 2018). Although the planarian central nervous system is

12 morphologically simple, consisting of a bilobed cephalic ganglion and two ventral nerve cords, it

13 is molecularly complex and highly compartmentalized. It shares the same major

14 neurotransmitters as the vertebrate brain, including dopamine, serotonin, octopamine, GABA,

15 and acetylcholine (Cebrià, 2007; Cebrià et al., 2002a; Cebrià et al., 2002b; Ross et al., 2017).

16 Furthermore, planarians possess a large repertoire of diverse quantifiable behaviors, such as

17 gliding and swimming, phototaxis, thermotaxis, and scrunching. The molecular mediators of

18 some of these behaviors have been characterized, such as serotonergic neurons in gliding or

19 transient receptor potential channels in thermotaxis (Birkholz and Beane, 2017; Inoue et al.,

20 2014; Nishimura et al., 2010) allowing these behaviors to be used to assess specific neuronal

21 functions, and thus provide insight into the cellular effects of neurotoxicity. In summary, by

22 testing diverse behaviors in both adult and regenerating/developing planarians, this platform

23 uniquely allows differentiation of selective neurotoxic effects and effects specific to the

1   developing nervous system.

2   The 15 FRs that are studied here comprise a subset of an 87-compound library from the

3   National Toxicology Program (NTP) (Behl et al., 2018; Hagstrom et al., 2018b; Zhang et al.,

4   2018), which we had previously screened. However, in this previous study we had not evaluated

5   these compounds in detail, because we focused on the comparison of compound classes (Zhang

6   et al., 2018) and concordance with other systems (Hagstrom et al., 2018b). Moreover, we re-

7   screened these 15 FRs independently in a second screen and used the two independent screens to

8   evaluate the robustness of our planarian screening platform. Thus, the present study serves a dual

9   purpose: To more closely examine the potential neurotoxicity of FRs and to quantify screen

10   robustness. To contextualize our results, we evaluated concordance to published data from

11   zebrafish, nematode, and human, mouse, and rat cell-culture models (Behl et al., 2015; Behl et

12   al., 2016; Jarema et al., 2015; Noyes et al., 2015). There was high concordance among the

13   different alternative systems showing comparable DNT of OPFRs to BFRs, warranting further

14   investigation into OPFR toxicity.

15

16

17   **2. Materials and methods**

18   *2.1. Chemical library*
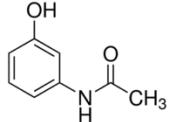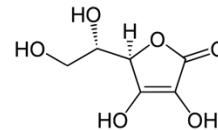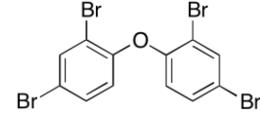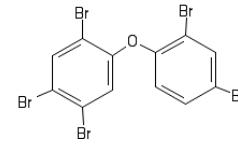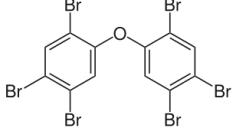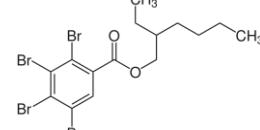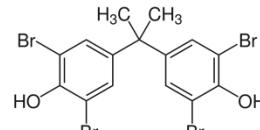
19   18 chemicals consisting of 6 BFRs (including phased-out BDE-47 and BDE-99), 8

20   OPFRs, 1 BFR mixture, 2 negative controls, and 1 duplicate (TPHP) were screened in this study.

21   Table 1 lists the 17 unique chemicals with their name, chemical ID, type, supplier, structure, and

22   purity information provided by the NTP (either as determined by the NTP or, if not available,

23   from the chemical supplier's certificate of analysis). The chemicals were provided by the NTP as
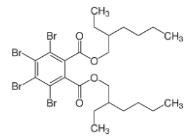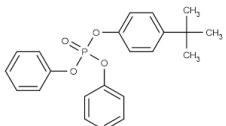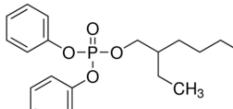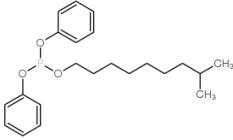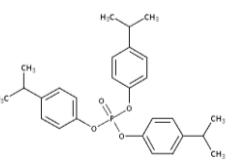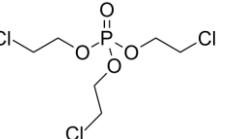
1   ~20mM stocks in dimethyl sulfoxide (DMSO, Gaylord Chemical Company, Slidell, LA), with

2   the exception of 2,2',4,4',5,5'-Hexabromodiphenyl ether (BDE-153), which was provided at

3   10mM.
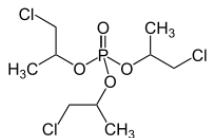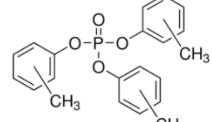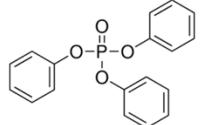
4

5

6   **Table 1**. Summary of the screened chemical library with CAS number, chemical name, ID, type

7   (BFR: brominated flame retardant, OPFR: organophosphorus flame retardant), chemical

8   supplier, structure, and purity. When available, the determined purity (by the NTP) is provided,

9   else the manufacturer supplied purity is provided. NA: not available.

CASRN	Chemical name	ID	Type	Supplier	Structure	Purity
103-90-2	Acetaminophen (4-hydroxyacetanilide)		Negative control	Sigma-Aldrich		100%
50-81-7	L-Ascorbic acid		Negative control	Sigma-Aldrich		100%
5436-43-1	2,2',4,4'-Tetrabromodiphenyl ether	BDE-47	BFR	Cerilliant Corp.		100%
60348-60-9	2,2',4,4',5-Pentabromodiphenyl ether	BDE-99	BFR	Cerilliant Corp.		100%
68631-49-2	2,2',4,4',5,5'-Hexabromodiphenyl ether	BDE-153	BFR	Cerilliant Corp. via Battelle Memorial Institute		100%
183658-27-7	2-ethylhexyl-2,3,4,5-tetrabromobenzoate	TBB	BFR	Toronto Research Chemicals		NA
79-94-7	3,3',5,5'-Tetrabromobisphenol A	TBBPA	BFR	Sigma-Aldrich		99.77%

CAS	Name	Abbreviation	Class	Source	Structure	Purity
26040-51-7	Bis(2-ethylhexyl) 3,4,5,6-tetrabromophthalate	TBPH	BFR	Toronto Research Chemicals		NA
860302-33-6	Firemaster 550	FM550	BFR mixture †	Chemtura Corporation		NA
56803-37-3	tert-Butylphenyl diphenyl phosphate	BPDP	OPFR	Ubichem PLC via MRIGlobal		98.22%
1241-94-7	2- ethylhexyl diphenyl phosphate	EHDP	OPFR	TCI America		98.67%
29761-21-5	Isodecyl diphenyl phosphate	IDDP	OPFR	Bayville Chemical Supply Company Inc.		92.19%
68937-41-7	Phenol, isopropylated, phosphate (3:1)	IPP	OPFR	Ameribrom Inc. (screen 1) Amfinecom Inc. (screen 2)		NA
115-96-8	Tris(2-chloroethyl) phosphate	TCEP	OPFR	Sigma-Aldrich		99.07%

13674-84-5	tris(2-Chloroisopropyl)phosphate	TCPP	OPFR	Albemarle Corporation via MRIGlobal		99.47%
1330-78-5	Tricresyl phosphate	TMPP	OPFR	Sigma-Aldrich (screen 1) Acros Organics (screen 2)		100%
115-86-6	Triphenyl phosphate*	TPHP	OPFR	Sigma-Aldrich		100%

† Firemaster 550 is a chemical mixture, comprised of 2-ethylhexyl-2,3,4,5-tertrabromobenzoate (TBB), bis(2-ethylhexyl) tetrabromophthalate (TBPH), triphenyl phosphate (TPP), and isopropylated triphenyl phosphate (IPTP) (Tung et al., 2017)

*TPHP was provided as a duplicate in FR screen 2. Duplicate vials were from the same supplier with the same lot numbers.

1   The tested FRs were a subset of the NTP 87-compound library, which we previously

2   reported on (FR screen 1; (Zhang et al., 2018)). The NTP later provided us with another set of

3   the library (now including 4 duplicate chemicals for a total of 91 chemicals), in which we only

4   screened the subset of chemicals mentioned in this study (FR screen 2). Chemicals in the new

5   library were provided as individual vials, stored at 4°C and used within 2 months. For all but two

6   chemicals (IPP and TMPP), the same supplier and mostly (14/17) the same chemical lot was

7   used for the chemicals in both FR screen 1 and 2. TPHP was tested in duplicate in FR screen 2,

8   using 2 separate stock vials from the same supplier and lot. The stocks of the negative control L-

9   ascorbic acid were from different lots, but had the same supplier with the same concentration and

10   purities. Five of the chemical stocks differed slightly in concentration between the two libraries,

11   but by at most 0.3mM. Additional information on the NTP library can be found in (Behl et al.,

12   2018).

13   The chemicals were tested at 0.01, 0.1, 1, 10, and 100μM (0.005, 0.05, 0.5, 5, and 50μM

14   for BDE-153), with a final DMSO concentration of 0.5% for all concentrations. Solvent control

15   populations were exposed to 0.5% DMSO. Chemical dilutions for both libraries were done as

16   described in (Zhang et al., 2018).

17

18   ***2.2. FR screen in the planarian system***

19   Asexual *D. japonica*, originally obtained from Shanghai University (Shanghai, China)

20   and cultivated in our lab for > 5 years, were used for all experiments. We screened the library

21   with a custom, fully automated HTS platform using adult (intact) and regenerating, i.e. tail pieces

22   regenerating a new head, planarians in parallel with the same assays, assessing morphological

23   and behavioral endpoints at day 7 and day 12 (Figure 1). Two independent screens, the original

1   NTP 87-compound library screen ("FR screen 1") and the second, FR-only screen ("FR screen

2   2"), were performed with 3 replicates (3 independent runs on different days) for each screen. For

3   each replicate, n=8 regenerating and adult planarians were each screened for every chemical

4   concentration. One adult or regenerating animal was loaded into each well of a 48-well plate

5   (Genesee Scientific, San Diego, CA). Static exposure began on day 1, within 3 hours of

6   amputation of the regenerating planarians, by adding the appropriate chemical concentration to

7   the planarians' aquatic environment. A solvent control population (n=8) exposed to 0.5% DMSO

8   was included in every plate. This DMSO concentration has no effect on planarian morphology

9   or behavior (Hagstrom et al., 2015). The planarians were kept in the dark in the sealed 48-well

10   plates for 12 days, with screening on days 7 and 12 (Figure 1). By day 12, regenerating

11   planarians are considered fully regenerated and behave like adults (Hagstrom et al., 2015; Zhang

12   et al., 2018), thus the two screening days allow us to distinguish regeneration delays from

13   regeneration defects.
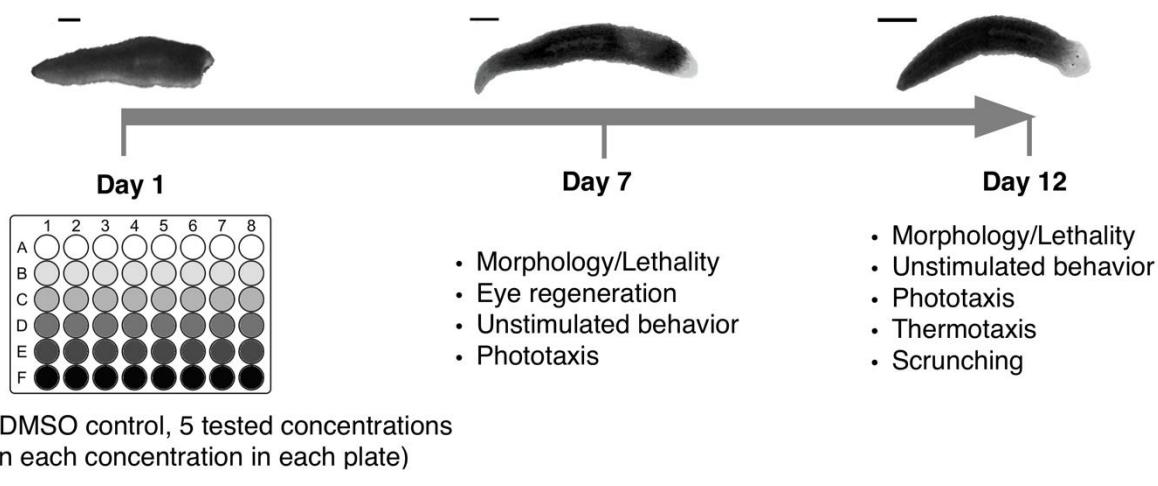
14

15   **Figure 1.** Schematic of overall screen flow in the planarian system. Exposure lasted 12 days.

16   Amputated planarians (amputation at day 1) and adult planarians are loaded into 48-well plates

1   (one worm per well, and one worm type per plate) with the test chemicals at day 1, and assayed

2   for different morphological and behavioral endpoints at day 7 and day 12. Scale bar: 1mm.

3

4   A detailed description of our automated planarian screening platform and statistical

5   workflow can be found in (Zhang et al., 2018). Briefly, adult and regenerating planarians were

6   assessed using morphological and behavioral endpoints at day 7 and day 12 (Figure 1). Since

7   dead worms generally disintegrate, lethality was quantified by automatic determination of the

8   presence or absence of a body in each well. Eye regeneration was quantified by determining the

9   percentage of planarians with 2 eyes regenerated (normal regeneration). For behavioral

10   endpoints, we used computational image analysis to quantify animal behavior to assay

11   unstimulated locomotion, negative phototactic behavior in response to blue light, the ability to

12   react to a heat gradient in thermotaxis, and the ability to scrunch (musculature-driven, oscillatory

13   escape gait (Cochet-Escartin et al., 2015)) in response to noxious heat. For statistical analysis,

14   data from the replicate runs was compiled and each chemical concentration was compared with

15   its own in-plate DMSO controls. Statistical significance was determined using either a one-tailed

16   Fisher's exact test for lethality, eye regeneration, phototaxis, and scrunching endpoints; Mann

17   Whitney U-test for thermotaxis and unstimulated behavior; or two-tailed t-test for unstimulated

18   behavior (depending on normality of the sample) using a significance level of 0.05. To extract

19   biologically significant results from the statistically significant hits, "biological relevancy"

20   cutoffs, based on the distribution of all control populations, were used to exclude hits that fell

21   within the variability of the DMSO controls. Lastly, hits generated from inconsistent data from

22   the different replicates were also excluded. For each endpoint, the lowest concentration which

23   showed a significant effect (statistically and biologically) was defined as the lowest-effect-level

1   (LEL). For each chemical, the overall LEL was determined as the most sensitive LEL when

2   considering all endpoints. All image analyses and statistical analyses were performed in

3   MATLAB.

4   The same experimental procedures and data analysis methods as described in detail in

5   (Zhang et al., 2018) were used, except for the following differences in FR screen 2, building

6   upon our experiences from FR screen 1: Firstly, since we found that food quality affects

7   planarian fitness and sensitivity to chemicals (Zhang et al., 2018), we fed planarians used in FR

8   screen 2 commercial freeze-dried organic chicken liver (Amazon, Seattle, WA) to better control

9   food quality and thus minimize animal fitness variability. Secondly, "biological relevance

10   cutoffs", used to minimize false positives by accounting for variability of the solvent controls in

11   each endpoint (Zhang et al., 2018), were updated by using the combined DMSO control samples

12   from the previous NTP 87-compound screen (n=87 control samples) and the later FR-only screen

13   (n=18 control samples) for a total control population of n=105 samples, with each sample

14   consisting of n=24 planarians. Thirdly, the background noise speed threshold used in the

15   phototaxis assay was recalculated based on the mean speed in the unstimulated behavioral assay

16   of the expanded DMSO controls. Hits due to hyper- (rather than hypo-) activity in unstimulated

17   behavior or hits which were concentration-independent were not included, because they were not

18   conserved when compiling all 6 runs together, suggesting they may be artifacts.

19

20   *2.3. Robustness analysis of planarian screening platform*

21   The robustness of our planarian screening platform was evaluated by comparing the

22   results obtained from six replicate runs (n=8 planarians per condition per replicate) from the two

23   independent FR screens, FR screen 1 and 2. It is reasonable to assume that using all 6 replicates

14

1   provides the most accurate results given the large sample size (n=48). We therefore compared

2   the results of compiling different numbers of replicates (3, 4, or 5) with the results obtained by

3   compiling 6 replicates. As described previously (Zhang et al., 2018), we used a 3-step (statistical

4   test, biological relevance cutoff, and inconsistency check) statistical workflow to determine

5   whether a chemical showed bioactivity, defined as a significant adverse effect, in any endpoint.

6   Herein, we refer to the determination of bioactivity (i.e. bioactive or inactive) in each endpoint as

7   a "readout" to distinguish it from a bioactive "hit", which refers to having an adverse effect at an

8   endpoint. Since the biological relevance cutoffs are meant as a means to distinguish normal wild-

9   type noise levels, biological relevance cutoffs were determined based on the total control

10   population of n=105 control samples (see Section 2.2), and were applied equally to these

11   different sets of compiled data (i.e. the same cutoffs, determined from this total control

12   population, were used for 3, 4, 5, or 6 replicates). For the inconsistency check, instances were

13   excluded where more than half of a set of replicates (i.e. >1 out of 3 replicates, >2 out of 4

14   replicates, >2 out of 5 replicates, or >3 out of 6 replicates) were within the biological relevancy

15   cutoffs. Although the majority of the experimental procedures were consistent between the two

16   screens, some differences existed which could affect the reproducibility of the screens. For

17   example, varying food quality (between different home-made batches and between home-made

18   and commercial sources) affects animal health and thus affects the animals' sensitivity to

19   chemical exposure (Zhang et al., 2018). Different chemical batches can also cause variability due

20   to differences in stock concentrations and potential variability in the serial dilution process.

21   Therefore, we chose replicates (different plates from different days) with the most similar

22   conditions (i.e. in the same screen) to minimize experimental variability and to focus on the

23   robustness of the system itself. To compile 3 replicates, we chose the 3 replicates from FR screen

15

1   2. To compile 4 (or 5) replicates, we chose 1 (or 2) replicate(s) from FR screen 1 in addition to

2   the 3 replicates from FR screen 2. Firstly, overall chemical bioactivity (i.e. whether the chemical

3   was bioactive in any endpoint or inactive in all endpoints) in different sets of replicates (3, 4, 5,

4   or 6) was compared. Secondly, concordance of the bioactivity of the readouts (i.e. whether the

5   chemical was deemed bioactive or inactive in a specific endpoint) between the data using 3, 4, or

6   5 replicates and the data using 6 replicates were compared. Lastly, the concordance of the

7   sensitivity of the concordant bioactive readouts (i.e. LEL of the bioactive readout) was compared

8   between the data using 3, 4, or 5 replicates and the data using 6 replicates.

9

10   *2.4. Comparison with other alternative models*

11   To directly compare our results with published data from other alternative models, we

12   consulted the literature (Behl et al., 2015; Behl et al., 2016; Jarema et al., 2015; Noyes et al.,

13   2015) and focused on studies which had screened the majority of the same FRs. Studies which

14   only tested one or a few of these FRs were not included in the comparison. We extracted the

15   points-of-departures (PODs) and LELs reported in these papers for relevant developmental and

16   developmental neurotoxic endpoints and focused only on the concentration-dependent hits for

17   LELs to facilitate cross-system comparisons. Specifically, we extracted PODs for developmental

18   toxicity and developmental neurotoxicity reported in (Behl et al., 2015), LELs for

19   developmental behavior in (Jarema et al., 2015), LELs for larval development in (Behl et al.,

20   2016), and LELs for all endpoints reported in (Noyes et al., 2015). Only data for regenerating

21   planarians was used for this analysis as the comparative studies used developing models.

22   To quantitatively determine the agreement between the different alternative animal

23   studies, Cohen's kappa (κ) was calculated for each pairwise comparison where κ = (observed

16

1   agreement - probability of chance agreement)/(1- probability of chance agreement) (McHugh,

2   2012). The 95% confidence intervals were quantified as $\kappa \pm$ (1.96 x the standard error). Of note,

3   the Noyes et al. study showed a $\kappa=0$ for all comparisons because it had no inactives in this set of

4   9 FRs, leading to a high probability of agreement due to chance.

5

6   *2.5. Relevance of findings in planarians to human biomonitoring data*

7   To provide toxicologists and regulators with context of how toxicity in planarians may

8   relate to human exposure levels, we compared the concentrations which elicited toxicity in

9   planarians to measured or estimated human plasma concentrations of several of the tested FRs

10   using high throughput toxicokinetic modeling (HTTK) (Rotroff et al., 2010; Sipes et al., 2017;

11   Wetmore et al., 2012; Wetmore et al., 2013). The underlying model assumes that *in vitro* media

12   concentrations are equivalent to human plasma concentrations and is traditionally used for cell-

13   based or high-throughput screens. While recognizing the limitation that the planarian is an

14   integrated system with more extensive kinetics than cell system models, this effort provides a

15   baseline comparison that can be adjusted with increasing information of planarian kinetics. As

16   such, this work directly compares estimated or measured (wherever available) human plasma

17   concentrations with the regenerating planarian LEL media concentrations. Human biomonitoring

18   data were unit-converted using HTTK modeling to estimate internal plasma concentrations in

19   children and consisted of breast milk, handwipes, and house dust, and plasma and cord-blood

20   serum levels of FRs as previously described (Alzualde et al., 2018). Briefly, the 3-compartment

21   model in the HTTK R package (version 1.7) (Pearce et al., 2017; Wambaugh et al., 2015) with

22   input chemical parameters from ADMET Predictor 7.2 (Simulations Plus, Inc., Lancaster, CA,

23   USA) was used. Child peak plasma concentrations (over 1 year) were estimated for TPHP and

17

1   TMPP from breastmilk, child hand wipe, and house dust samples (Kim et al., 2014; Sugeng et

2   al., 2017). Adult plasma values for TBBPA and BDE47, as well as cord blood serum for TBBPA

3   and child plasma for BDE47 were used (Cariou et al., 2008; Stapleton et al., 2012; Wang et al.,

4   2013). Exposure values obtained were the range, as well as the geometric mean for: adult

5   plasma, child plasma, cord serum, and TPHP child handwipe; median for: breastmilk and TMPP

6   child handwipe, and maximum median for house dust.

7

8   *2.6. Cholinesterase activity staining in OPFR treated planarians*

9   To determine whether OPFR toxicity was correlated with inhibition of

10   acetylcholinesterase activity, adult planarians were incubated in either 0.5% DMSO, 10 µM

11   EHDP (an active OPFR in planarians) or 100 µM TCEP (an inactive OPFR in planarians) for 12

12   days and then fixed and stained to visualize acetylthiocholine catalysis as previously described

13   (Hagstrom et al., 2018a; Zheng et al., 2011), with a staining incubation of 3.5 hours. The staining

14   was performed in duplicate on separate days on a total of 6 worms per chemical (3/day) and 12

15   DMSO controls (6/day). Supplementary Figure 1 is representative of all imaged worms.

16

17

18   **3. Results**

19   A library of 15 FRs and 2 negative controls (acetaminophen and L-ascorbic acid) (Table

20   1), was screened in adult and regenerating planarians to assess their effects on mortality, eye

21   regeneration, and behavior. These chemicals were screened twice, once as part of the NTP 87-

22   compound library screen (FR screen 1) (Zhang et al., 2018) and once as a new, independent

23   screen of the FR-specific subset of a newly obtained version of the library (FR screen 2), with 3

1   replicate runs in each screen (total n=24 per screen). Therefore, we compared the results obtained

2   from these two independent screens using the same screening methodology to evaluate the

3   robustness and reproducibility of our planarian screening platform. Moreover, as our previous

4   study of the NTP 87-compound library focused on the screening methodology and major trends

5   seen in the screen, herein we provide more in-depth insight of the toxicological profiles induced

6   by this discrete subset of chemicals. The raw and analyzed data are available in Mendeley Data

7   and Supplementary File 1, respectively.

8

9   *3.1. Robustness of the planarian FR screen*

10   To evaluate the robustness and reproducibility of our planarian screening platform, we

11   compared the results obtained from compiling different numbers of replicates from the two

12   independent screens (FR screen 1 and FR screen 2) (Figure 2). Compiled results for the

13   individual replicates and different sets of replicates can be found in Supplemental File 1. Of the

14   18 screened chemicals, IPP and TMPP were provided by two different suppliers for the two

15   screens (Table 1). Therefore, given potential unknown variability between the different chemical

16   batches, IPP and TMPP were excluded from the robustness evaluation. Notably, a recent screen

17   with developing zebrafish found significant readout differences between IPP stocks from three

18   different suppliers (Noyes et al., 2015), confirming potential variability among suppliers.

Readout with different sensitivity   Loss of bioactive readout   False positive readout

Regenerating planarian   **Adult planarian**

3 replicates

4 replicates

5 replicates

6 replicates

Compounds (rows): TBPH, BDE-153, TBB, BDE-99, BDE-47, TBBPA, FM550, TCEP, IDDP, TCPP, EHDP, BPDP, TPHP, TPHP, Acetaminophen, L-Ascorbic acid, IPP, TMPP

Column labels: Lethality D7, Lethality D12, Unstimulated D7, Unstimulated D12, Phototaxis D7, Phototaxis D12, Thermotaxis, Scrunching, Eye regen, Overall LEL

1

1   **Figure 2.** Comparison of lowest-effect-levels (LELs) identified for each endpoint by compiling

2   data from 3, 4, 5, or 6 replicates in regenerating and adult planarians. Readouts, i.e. whether

3   bioactivity was detected in a particular endpoint, were determined for each endpoint using 3, 4,

4   or 5 replicates from the 2 screens and compared with the readouts determined from 6 replicates.

5   "-" indicates the chemical was inactive in that endpoint. Loss of a bioactive readout (deemed

6   bioactive in 6 replicates but not 3, 4, or 5) are in dark grey. False positive readouts (deemed

7   bioactive in 3, 4, or 5 replicates but not 6) are in yellow. Endpoint readouts which differed in

8   sensitivity/potency (still bioactive in both screens but with different LELs) are in blue. The

9   overall LEL of each chemical is listed in a separate column. IPP and TMPP (in red text) were

10   excluded from the robustness evaluation since the chemical supplier differed between the two

11   screens.

12

13   To determine the robustness of our data, we evaluated both the overall agreement on

14   bioactivity identification and potency (LELs) from any endpoints, and the agreement on

15   bioactivity and LEL per endpoint when comparing across different numbers of compiled

16   replicates. Firstly, we evaluated the key findings, i.e. general determination of toxicity from any

17   endpoint and overall potency. In all sets of compiled data (3, 4, 5, or 6 replicates), the same

18   10/16 chemicals, including the duplicate TPHP, caused adverse effects in at least one assay

19   endpoint in both adult and regenerating planarians. Thus, the overall chemical hit-call was

20   unchanged regardless of the number of replicates used. Furthermore, we also found that the

21   sensitivity of the system generally did not differ among the different numbers of replicates as the

22   overall LELs, i.e. the most sensitive LEL from any assay endpoint for each chemical, were the

23   same for almost all bioactive chemicals in 3/4/5 replicates versus 6 replicates. Only BDE-47 had

21

1 an overall LEL, which differed depending on the number of replicates analyzed. The overall LEL

2 for adult planarians exposed to BDE-47 was lower (i.e. more potent) when using 3 and 4

3 replicates compared to using 6 replicates. Significant effects were found at 10 µM BDE-47 in

4 day 7 and 12 unstimulated behavior when analyzing 3 or 4 replicates but not 6, because in FR

5 screen 1 BDE-47 generally showed decreased sensitivity in this assay, which we attribute to

6 differences in animal fitness caused by differences in food quality (Zhang et al., 2018).

7 Secondly, because one of the strengths of our planarian screening platform is the ability

8 to distinguish effects on different neuronal functions using various behavioral readouts, we

9 compared the bioactivity of each endpoint readout using the compiled data of 3, 4, or 5 replicates

10 with the compiled data of all 6 replicates (Figure 2, Table 2). Of the 144 readouts in regenerating

11 planarians and 128 readouts in adult planarians, the bioactivity (either bioactive or inactive)

12 determined with 6 replicates was highly concordant ($\geq 96\%$) with that of 3, 4, and 5 replicates

13 (Table 2). Differences were mostly due to loss of detection of bioactive readouts in 3, 4, or 5

14 replicates (compared to 6 replicates), although there were a few readouts which were found to be

15 bioactive in 3, 4, or 5 but not 6 replicates. Generally, loss of bioactive readouts in 3, 4, or 5

16 replicates was correlated with fewer replicates, as the smaller sample size limited the ability to

17 detect statistically significant effects. However, this primarily only occurred at concentrations

18 which also yielded significant lethality. Thus, in these cases, there was not enough data from

19 alive animals to yield a statistically significant effect on a morphological or behavioral endpoint.

20 This suggests that some specific defects (e.g. scrunching) may be masked by overt systemic

21 toxicity and lethality, especially in smaller sample sizes. The only exception to this trend was

22 IDDP, where differences in readout bioactivity were not due to low sample size but instead from

23 inconsistency among the different replicates from the two independent screens. Furthermore, we

22

1 evaluated sensitivity concordance by comparing the LELs determined from the different

2 replicate sets for the concordant bioactive readouts (Table 2) to determine whether the same

3 concentration-dependent phenotypic profiles were identified in the different replicates. In

4 regenerating tails, all concordant bioactive readouts for all endpoints had the same LELs across

5 all sets of compiled data. In adult planarians, $\geq 97\%$ of the concordant bioactive readouts had the

6 same LELs using 3/4/5 replicates compared with using 6 replicates. Together, these data show

7 the high reproducibility among the different replicates in both endpoint-specific bioactivity and

8 potency and that our current screening strategy using 3 replicates is sufficiently robust.

9   TPHP was screened as a duplicate in FR screen 2 (FR-specific library screen). The results

10 of the duplicates were consistent in any set of compiled data, being bioactive in the same

11 endpoints at the same LELs (Figure 2), underlining the robustness and reproducibility of our

12 planarian screening platform.

13

14 **Table 2.** Bioactivity concordance of readouts and sensitivity concordance of concordant

15 bioactive readouts of the data using 3, 4, or 5 replicates compared with the data using 6

16 replicates. Data are shown for regenerating (R) and adult (A) planarians. IPP and TMPP were

17 excluded from the analysis. Bioactivity concordance of readouts was determined as the number

18 of readouts showing the same bioactivity (bioactive or inactive) as using 6 replicates, out of the

19 total number of readouts (9 for regenerating and 8 for adult planarians) for all 16 chemicals.

20 Sensitivity concordance of concordant bioactive readouts using 3/4/5 replicates was determined

21 as the number of bioactive readouts with the same LEL as in 6 replicates, out of the total number

22 of concordant bioactive readouts (numbers in first two columns) for all 16 chemicals in this set

23 of compiled data.

23

	Bioactivity concordance of readouts		Sensitivity concordance of bioactive concordant readouts	
	R (144 readouts)	A (128 readouts)	R (concordant bioactive readouts)	A (concordant bioactive readouts)
3 replicates	138 (96%)	123 (96%)	138 (100%)	119 (97%)
4 replicates	139 (97%)	124 (97%)	139 (100%)	123 (99%)
5 replicates	142 (99%)	124 (97%)	142 (100%)	124 (100%)

1

2

3.2. OPFRS are more potent than BFRs in planarians

4   After confirming our data is robust and reproducible, we used the compiled data from all

5 6 replicates to delve into the details of FR toxicity. Eleven of the 15 unique FRs caused adverse

6 effects in both adult and regenerating planarians in at least 1 endpoint (Figure 3). The duplicate

7 chemical, TPHP, was consistent between batches and thus the results are only shown once. The

8 bioactive FRs consisted of 3 of the 6 BFRs (BDE-99, BDE-47, TBBPA), 7 of the 8 OPFRs

9 (IDDP, TCPP, EHDP, IPP, BPDP, IPP, TMPP, TPHP), and the FM550 mixture. TBPH, BDE-

10 153, TBB, TCEP, as well as the two negative control chemicals, were inactive in both worm

11 types at the tested concentrations. All chemicals were screened at 5 test concentrations (see

12 Materials and Methods, Section 2.1); however, hits were only observed at the two highest

13 concentrations, 10 and 100 µM. Due to poor solubility, BDE-153 was tested at a maximum

14 concentration of 50µM, which may explain its inactivity, since the other BDE compounds were

15 only active at 100µM.

24
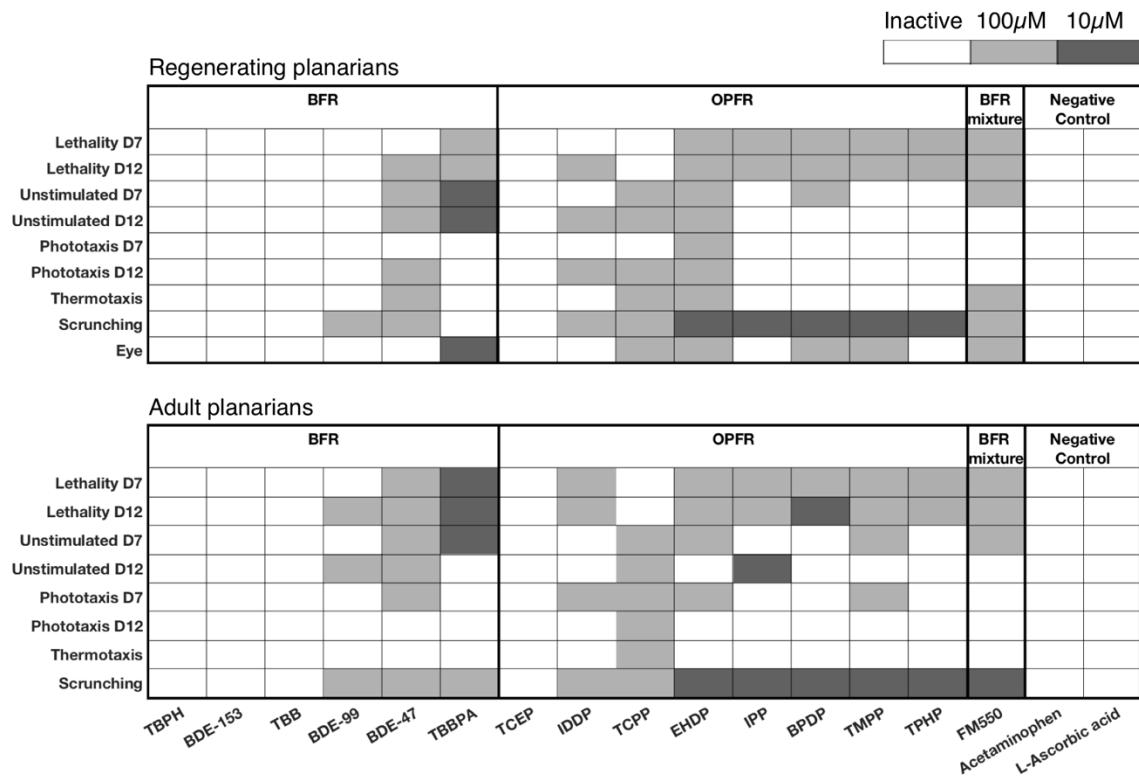
1

Figure 3. Overview of the planarian screening data of 15 unique flame retardants and 2 negative

controls using 6 replicates. Heat maps of effects of brominated flame retardants (BFRs),

organophosphorus flame retardants (OPFRs), a BFR mixture, and negative controls on

regenerating (Top) and adult (Bottom) planarians for all endpoints with lowest-effect-level

shaded. Only the highest two of the five tested concentrations showed effects in planarians.

TPHP was screened as a duplicate in FR screen 2, but only shown once here since the results

were consistent. Note that BDE-153 was screened at a maximum concentration of 50 µM.

Using the high-content data obtained by screening for effects on planarian morphology

and a variety of diverse behaviors, we determined whether any specific toxicities were induced

by the different FRs. Eight FRs in regenerating planarians (BDE-99, TBBPA, TCPP, EHDP,

IPP, BPDP, TMPP, and TPHP) and 6 FRs in adult planarians (TCPP, EHDP, IPP, BPDP, TMPP,

25

1   and TPHP) caused adverse outcomes on morphological and/or behavioral endpoints at

2   concentrations where animal viability was not affected, suggesting specific non-systemic toxicity

3   at these concentrations. Although causing sublethal effects in regenerating planarians, two BFRs,

4   BDE-99 and TBBPA, only showed toxicity concomitant with lethality in adult planarians, due to

5   increased sensitivity to lethality in the adult worms. Similar increased sensitivity of adult

6   planarians to the lethal effects of some chemicals has been observed previously (Hagstrom et al.,

7   2015; Zhang et al., 2018).

8   Next, we further examined which specific endpoints were affected by the sublethal

9   concentrations of the different FRs. Scrunching, a planarian musculature-driven gait used as an

10   escape response to specific adverse stimuli such as noxious heat (Cochet-Escartin et al., 2015),

11   was the most sensitive endpoint with 7 and 6 FRs causing scrunching defects at sublethal

12   concentrations in regenerating and adult planarians, respectively. To a lesser extent, sublethal

13   effects were also seen on unstimulated behavior (regenerating planarians in TBBPA and adult

14   planarians in IPP) and eye regeneration (TBBPA and TCPP). Interestingly, in addition to effects

15   on eye regeneration, TCPP affected almost all behavioral endpoints (with the exception of day 7

16   phototaxis in regenerating planarians) in both worm types in the absence of lethality. Together,

17   these toxicological profiles demonstrate the power of the large number of diverse endpoints to

18   discern specific toxic effects from systemic toxicity and pinpoint effects on specific neuronal

19   functions, such as neuromuscular communication.

20   Our planarian HTS platform has the unique advantage of allowing direct comparisons of

21   chemical effects on adult and developing/regenerating animals. By doing so, we found TBBPA

22   caused a developmental selective defect on day 12 unstimulated behavior, since effects on this

23   endpoint were seen at 10µM in regenerating but not adult planarians (Figure 3). Of note, this

26

1   concentration did induce lethality in 24% of adult planarians at day 12 (see Supplemental File 1).

2   However, this low level of lethality, although statistically significant, may suggest that any overt

3   systemic toxicity at this concentration would not be potent enough to mask effects in other

4   endpoints. Furthermore, TBBPA and TCPP caused eye regeneration defects at a sublethal

5   concentration. Thus, while most compounds adversely affected both adult and developing

6   animals, some compounds displayed additional developmental-specific toxicities.

7   Lastly, to ask whether OPFRs are safer alternatives to BFRs, we compared the

8   toxicological profiles induced by the two types of FRs. The OPFRs generally showed higher

9   potency than the BFRs with LELs of 10µM in 5 out of 7 bioactive OPFRs, whereas the bioactive

10   BFRs, except for TBBPA, had LELs of 100µM. TBBPA had a LEL of 10µM in both adult and

11   regenerating planarians, with the most endpoints affected at this concentration, compared to the

12   other FRs.

13   Taken together, the bioactive OPFRs were at least as potent as BDE-47, which was

14   phased-out due to concerns about its toxicity (Stapleton et al., 2014), emphasizing the

15   importance of evaluating the toxicity of these supposedly safer replacement FRs. Almost all

16   bioactive OPFRs (6/7 and 5/6 for regenerating and adult planarians, respectively) induced

17   specific sublethal effects on scrunching. This was not seen to the same extent in the BFRs, where

18   only 1 bioactive BFR (BDE-99) showed sublethal effects on scrunching and only in regenerating

19   planarians. Thus, by assaying a variety of different neuronal functions in planarians, we were

20   able to differentiate toxicities elicited by different chemical domains, suggesting that our system

21   provides specificity.

22

23   *3.3. Concordance of bioactive hits between planarians and other alternative models*

1   While several of the 15 FRs that we screened have been previously studied in other

2   systems, including rodents (Cope et al., 2015; EFSA, 2011; Moser et al., 2015; Nakajima et al.,

3   2009; National Toxicology Program, 1991; Viberg and Eriksson, 2011) and other alternative

4   models (Alzualde et al., 2018; Bailey and Levin, 2015; Cano-Sancho et al., 2017; Glazer et al.,

5   2018; Oliveri et al., 2015; Slotkin et al., 2017; Usenko et al., 2016), only a handful of studies

6   have performed direct multi-FR comparisons as provided here. Fourteen of the 15 unique FRs

7   tested in this screen have been previously studied in developing zebrafish, nematodes, and *in*

8   *vitro* cell-based (mouse embryonic stem cell, human neural stem cell, and rat neuron) systems

9   with some FRs showing developmental toxicity or DNT. Thus, herein, we compared our

10   planarian data with these published results (Behl et al., 2015; Behl et al., 2016; Jarema et al.,

11   2015; Noyes et al., 2015) to contextualize our findings (Table 3, Figure 4). The FM550 mixture

12   was excluded from this comparison because it was not screened in any of these previous studies.

13   Comparisons were made with the overall regenerating planarian LEL, i.e. most sensitive LEL

14   from any endpoint, determined from the compiled data of 6 replicates from FR screens 1 and 2

15   (see Methods Section 2.4, Table 3 and Figure 4).

16   First, we compared the overall chemical hit calls between the different systems to

17   determine whether the same FRs were identified as bioactive, i.e. caused significant adverse

18   effects, in the different models. Thirteen of the 14 shared FRs were previously found to be

19   bioactive in at least one model; however, not all 14 FRs were tested in all systems or studies. IPP

20   was the only FR found to be bioactive in all tested systems. TBB was the only FR which was

21   inactive in all screens; however, it was only tested in our study and the (Noyes et al., 2015)

22   zebrafish study. Generally, the alternative animal models detected more bioactive FRs than the *in*

23   *vitro* models (Table 3). Of the 9 FRs tested in the *in vitro* models, 8 were bioactive in any model,

28

1   with at most 7 detected in any one model (human stem cell neurite outgrowth). The majority of

2   effects caused by the FRs in the *in vitro* models occurred concomitantly with cytotoxicity,

3   suggesting non-specific effects.

Table 3. Comparison of developmental and/or developmentally neurotoxic endpoints used to evaluate FR toxicity between regenerating planarians and different alternative developing models. Lowest-effect-level (LEL) or point-of-departure (POD) for each model are listed in µM. Dash (-) indicates that there was no toxicity detected at the tested concentrations. An empty cell indicates that this chemical was not tested in this model. The number in brackets represents the LELs when considering concentration-independent and hyper-active effects. Bold text indicates that the effects were detected at sublethal concentrations.

Model	Assay	BFRs						OPFRs							
		TBPH	BDE-153	TBB	BDE-99	BDE-47	TBBPA	TCEP	IDDP	TCPP	EHDP	IPP	BPDP	TMPP	TPHP
Planarian	Morphology and and/or behavior (LEL)	-	-	-	**100**	100	**10**	-	100.5	**100**	**10**	**10.05**	**10.05**	**10**	**9.95**
Zebrafish	Embryonic development (POD)[a]					**28.2**	4.6	-	-		15.3	**4.9**	**9.8**	**15.4**	**2**
	Developmental behavior (LEL)[b]					**4**	-	**12**	-		**12 [2.1]**	**3.8 [1.2]**	-	**4**	**0.4**
	Developmental malformations and/or behavior (LEL)[c]	**0.64 [6.4E-4]**	**6.4 [0.064]**	- [6.4E-4]	**0.64 [0.0064]**	0.0064	**0.0064**	**64 [0.0064]**	**6.4 [0.0064]**	**64 [0.64]**	64 [0.0064]	0.064	**0.64 [0.0064]**	0.0064 [0.0064]	**0.64 [0.0064]**
Nematode	Larval development (POD)[a]					**0.6**	10.1	-	**3.9**		2.3	**3.2**	**3.3**	95.5	**0.9**
	Larval development and/or feeding (LEL)[d]				**0.16**	**0.4**	3.2	-	**2.5**		**1.6**	**1.6**	**2.5**	100	**0.16**
In vitro	Mouse stem cell differentiation (POD)[a]					-	1.1	-	-		-	0.4	-	-	-
	Human stem cell neuroprogenitor proliferation (POD)[a]					-	-	-	2.3		1	1.9	2	**7.4**	-
	Rat neurite outgrowth (POD)[a]					-	-	-	-		-	21	**14.9**	-	-
	Human stem cell neurite outgrowth (POD)[a]				1	1.2	-	1.1		0.9	0.5	1.2	-	1.1

1   a, data from (Behl et al., 2015); b, data from (Jarema et al., 2015); c, data from (Noyes et al., 2015); d, data from (Behl et al., 2016)

	Planarians	Zebrafish[a]	Zebrafish[b]	Zebrafish[c]	Nematode[a]	Nematode[d]
Planarians	1	0.61 ± 0.29	-0.20 ± 0.52	0 ± 0.23	1	1
Zebrafish[a]	0.61 ± 0.29	1	0.18 ± 0.52	0 ± 0.35	0.61 ± 0.29	0.61 ± 0.29
Zebrafish[b]	-0.20 ± 0.52	0.18 ± 0.52	1	0 ± 0.46	-0.20 ± 0.52	-0.20 ± 0.52
Zebrafish[c]	0 ± 0.23	0 ± 0.35	0 ± 0.46	1	0 ± 0.23	0 ± 0.23
Nematode[a]	1	0.61 ± 0.29	-0.20 ± 0.52	0 ± 0.23	1	1
Nematode[d]	1	0.61 ± 0.29	-0.20 ± 0.52	0 ± 0.23	1	1

< 0.20	0.21 - 0.40	0.41 - 0.60	0.61 - 0.80	0.81 - 0.99	1
Chance	Fair	Moderate	Substantial	Near Perfect	Perfect
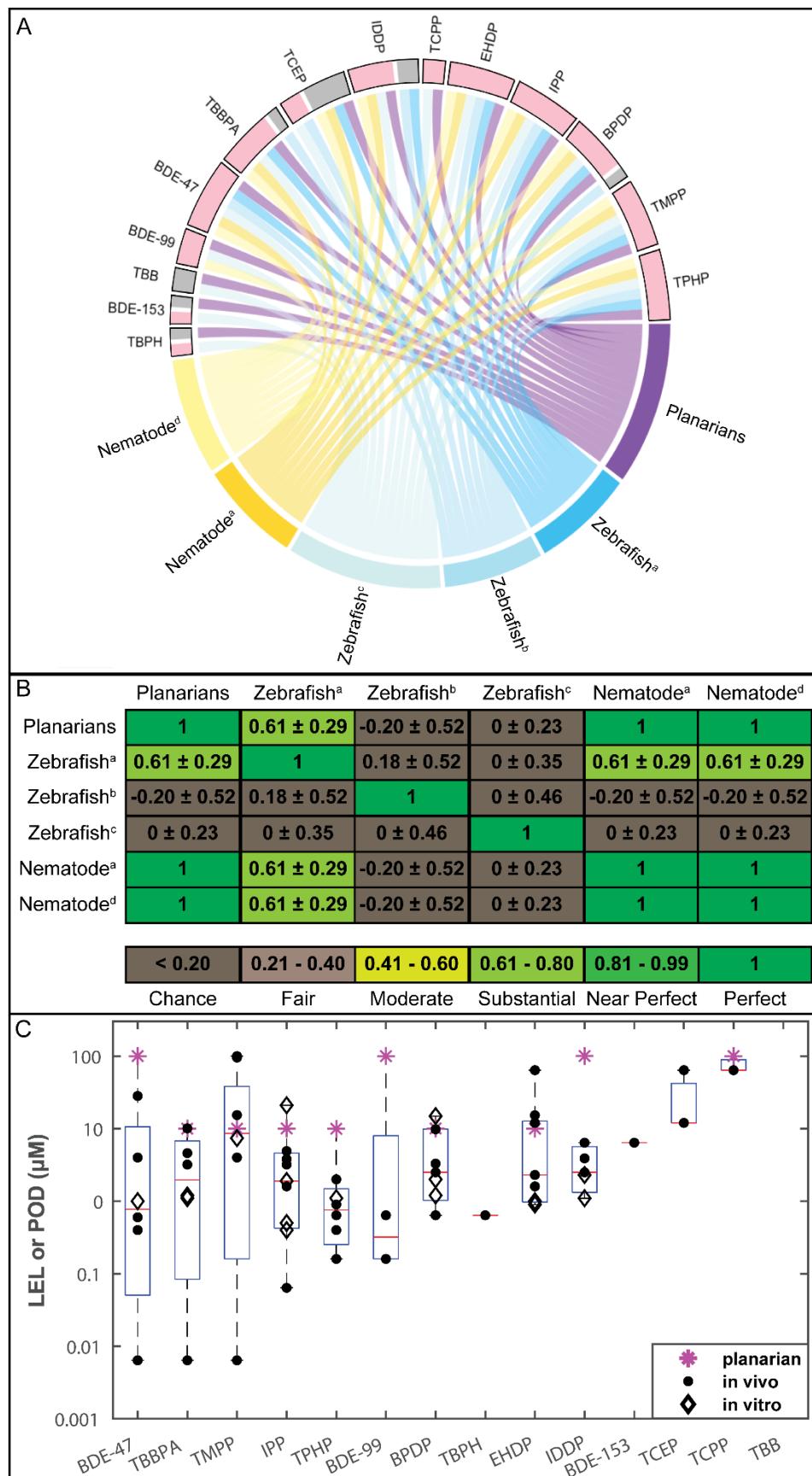
1   **Figure 4.** A. Interaction of bioactivity of the 14 flame retardants (FRs) with the different tested

2   alternative animal models. Each box represents the interactions of this FR with the different

3   models and studies, coded by different colors, where active interactions are in pink and inactive

4   interactions are in grey. FRs with both active and inactive interactions signify that the chemical

5   showed different effects in different studies/models. B. Quantification of concordance between

6   alternative animal models for 9 FRs screened in all alternative animal models using Cohen's

7   kappa (κ), listed as κ ± 95% confidence intervals (McHugh, 2012). Of note, the Noyes et al study

8   showed a kappa of 0 for all comparisons because it had no inactives in this set of 9 FRs. For (A)

9   and (B), studies are listed as: a, data from (Behl et al., 2015); b, data from (Jarema et al., 2015);

10   c, data from (Noyes et al., 2015); d, data from (Behl et al., 2016). C. Comparison of toxicity

11   potency of 14 FRs in regenerating planarians and other alternative developing models. The data

12   for the individual studies are overlaid on a box and whisker plot compiling the lowest-effect-

13   levels (LELs) and points-of-departure (PODs) for the different FRs in all the different models.

14   Stars represent planarian LELs, filled dots represent LELs/PODs in developing zebrafish and

15   nematode models (Behl et al., 2015; Behl et al., 2016; Jarema et al., 2015; Noyes et al., 2015),

16   and diamonds represent PODs in *in vitro* cell-based models (Behl et al., 2015). The x-axis shows

17   the FRs in decreasing order of potency (lowest effect concentration).

18

19   Due to the increased specificity and higher similarity to our planarian screening platform,

20   we focused our concordance analysis on the alternative developing/regenerating animal models

21   (planarians, nematodes, and zebrafish). The chemical hit calls for the 14 FRs tested in these

22   models were highly concordant. The two invertebrate systems, nematodes and planarians, had

23   perfect hit concordance for the 10 FRs tested in both systems. These 10 FRs were also tested in

33

1   at least one zebrafish study and showed high concordance among the three models (considering

2   all studies together) with the same 9 chemicals eliciting adverse effects in at least one study from

3   each system (Table 3, Figure 4A). Of these 10 FRs, only TCEP had discordant results between

4   the different systems, as it was bioactive in two of the zebrafish studies, but inactive in the third

5   zebrafish study and in nematodes and planarians.

6   As this example illustrates, some discrepancies were found across different studies using

7   the same model. Detailed comparisons are shown in Table 3 and Figure 4A. We used Cohen's

8   kappa, which also considers agreement due to chance (McHugh, 2012), to quantify concordance

9   across the different studies for the 9 shared FRs (Figure 4B). As expected, we found perfect

10   agreement between the two nematode studies and our planarian study. However, all other

11   comparisons yielded substantial agreement or only agreement due to chance, even when

12   comparing different studies in the same system, demonstrating that methodological differences

13   between labs studying the same system can be as impactful as interspecies differences. This

14   emphasizes the importance of transparency, so results from different studies can be adequately

15   compared and compiled to leverage the strength of multi-system testing.

16   We also compared the sensitivity (i.e. chemical potency) of our regenerating planarian

17   screening data to those of the comparative alternative model studies (Figure 4C, Table 3).

18   Notably, some of these studies, including ours herein, evaluated toxicity potency using the

19   nominal test concentrations (i.e. LEL) (Behl et al., 2016; Jarema et al., 2015; Noyes et al., 2015)

20   while others used modeling approaches to calculate a point-of-departure (POD) (Behl et al.,

21   2015). Eight of the 10 FRs causing adverse effects in planarians (BDE-47, TBBPA, BPDP,

22   EHDP, IPP, TMPP, TPHP, and TCPP) had overall LELs (from any endpoint) in the regenerating

23   planarian system within an order of magnitude of the effect concentration (LEL or POD) in at

34

1   least one of the other studies using zebrafish, nematode, and *in vitro* cell-based models (Figure

2   4C, Table 3).

3

4

5   **4. Discussion**

6   *4.1. Robustness of planarian screening platform*

7   Robustness of screening data is indispensable for building confidence in a new testing

8   platform. In previous proof-of-concept work (Hagstrom et al., 2015), we used post-hoc statistical

9   power analysis to determine that n=24 specimen per condition were necessary to achieve

10   sufficient power. Here, we experimentally verified the validity of our current test strategy, testing

11   in 3 replicate runs with n=8 each, through direct comparison of the reproducibility of data

12   obtained from using 3 (n=24), 4 (n=32), or 5 replicates (n=40) with 6 replicates (n=48). We

13   found that our planarian platform is robust and that 3 replicates were sufficient to identify all

14   chemical hit-calls, distinguishing the same bioactive chemicals as 6 replicates. Moreover,

15   readout bioactivity and sensitivity (i.e. LELs) of the different endpoints were highly concordant

16   among the different sets of replicates. Loss of bioactive readouts with less replicates was only

17   found for morphological or behavioral endpoints at lethal concentrations. Since all endpoints,

18   except lethality, were quantified from the data from living animals, high lethality led to a small

19   sample size and limited the power to detect a statistically significant effect. Moreover, lethality

20   suggests overt systemic toxicity at these concentrations which may have masked effects at more

21   specific endpoint readouts. Given the overt toxicity at lethal concentrations, many labs only

22   evaluate more sensitive morphological or behavioral effects at sublethal concentrations,

23   disregarding effects at lethal concentrations (Alzualde et al., 2018; Jarema et al., 2015; Truong et

35

1   al., 2014). All together, these data indicate that different number of replicates (from 3-6) yield

2   very similar results, validating that our current screening strategy using 3 replicates is able to

3   robustly identify chemical bioactivity and different toxicities.

4

5   *4.2. The role of in vitro HTS for DNT*

6   As laid out in the introduction, few mammalian guideline studies have been conducted to

7   evaluate the DNT of chemicals, including the FRs tested herein. High-throughput *in vitro* assays,

8   coupled with bioinformatic analysis, have emerged to fill this gap by quickly and efficiently

9   evaluating potential chemical toxicity. However, our comparison of several human and rodent

10   cell culture assays with non-mammalian whole animal models found that the alternative animal

11   models were better at identifying specific developmental or developmental neurotoxic effects at

12   sublethal concentrations (Table 3). When considering all compared *in vitro* assays, 8/9 FRs were

13   found to be bioactive; however, 6 of these bioactive FRs were only identified in 1 or 2 assays

14   (primarily the human stem cell systems). The two rodent cell culture assays only yielded 2 hits

15   each, both of which were also detected by the human stem cell assays. The majority of the *in*

16   *vitro* effects were found concomitantly with cytotoxicity, while selective sublethal effects were

17   more apparent in the alternative animal models. A recent study evaluating a small number of

18   BFRs and OPFRs in rat neural stem cells and neuronotypic PC12 cells was able to distinguish

19   effects on neurodifferentiation from cytotoxicity (Slotkin et al., 2017), suggesting that certain

20   types of cell culture assays may be more suited for DNT studies than others. Thus, while current

21   *in vitro* techniques are extremely powerful for rapid screening of toxicants on other organ

22   systems, they may have limited value for detecting potential developmental toxicity or DNT,

23   because they inherently lack the necessary complexity for detecting functional perturbations that

36

1   are independent of systemic toxicity. Alternative whole animal systems present an opportunity to

2   fill this gap between *in vitro* and mammalian animal models.

3

4   *4.3. Need for transparency in screening methodology and analysis*

5   As shown in Figure 4B, intraspecies differences in chemical hit-calls between different

6   studies were as big as interspecies differences. Additionally, even when a chemical was

7   identified as bioactive in two studies from the same model, sensitivity could vary dramatically.

8   For example, the zebrafish data from (Noyes et al., 2015) shows increased sensitivity (up to 3

9   orders of magnitude difference) to 4 FRs (TBBPA, BPDP, IDDP, and IPP) when compared to

10   the other published zebrafish studies (Table 3).

11   Important methodological differences in experimental design and data analysis may

12   explain these differences. For example, we tested the FRs in log-steps, whereas several of the

13   other studies used smaller concentration steps, thus the large dose gaps in our study limits precise

14   comparisons of potency. In addition, how significant effects are determined can vary

15   dramatically between laboratories. Some use statistics alone, while others, like ourselves

16   incorporate additional stringency tests including biological relevancy cutoffs and inconsistency

17   checks (Zhang et al., 2018). Our statistical pipeline is based on empirical observations to reduce

18   the false positive rate, but may limit sensitivity. For example, we found our biological relevancy

19   cutoffs removed 72% of statistically significant hits, which changed the LELs of 8 chemicals

20   (Supplementary Table 1). However, the majority of the removed hits were concentration-

21   independent. i.e. effects were seen at lower but not the subsequent higher concentrations, and

22   some were hits in negative controls, together suggesting that these excluded hits are likely false

23   positives (Zhang et al., 2018). Variability can also arise due to using LELs vs PODs to determine

37

1   potency. LELs are limited to the tested concentrations, and thus the respective POD could lie

2   somewhere within the range of the LEL and the next lower tested concentration level. Therefore,

3   because the different experimental approaches and statistical criteria used to evaluate bioactivity

4   and potency are as important as differences between species, transparency of methodology is

5   crucial to make system comparisons meaningful and build confidence in these new approaches

6   for complementing mammalian toxicity testing.

7

8   *4.4. Comparisons of different models emphasizes the value of a battery approach*

9   To contextualize the results of our planarian screen with what is known for FR toxicity,

10 we compared the available data from alternative, non-mammalian animal models and found that

11 our results were highly concordant. Eight of 10 common FRs tested in planarians, zebrafish, and

12 nematodes were bioactive in at least one study in all three systems (Table 3 and Figure 4A).

13 Thus, these shared bioactive FRs affected multiple models and multiple endpoints, spanning

14 effects on development, neurodevelopment, and general neuroactivity, indicating multiple

15 mechanisms may be involved in their toxicity. Battery approaches such as these are powerful

16 because they allow parallel interrogation of different species and endpoints, from the cellular to

17 organismal level, to integrate multiple lines of evidence and build confidence in preliminary

18 hazard assessment.

19 In addition to comparing with other toxicological models, comparing with human

20 environmental exposure can contextualize the relevancy of observed toxicity in a given system.

21 Therefore, we overlaid our regenerating planarian screening results to measured or estimated *in*

22 *vivo* plasma concentrations using the HTTK R-package (Figure 5) (Alzualde et al., 2018; Pearce

23 et al., 2017). The nominal water concentrations at which effects were seen in regenerating

38

1   planarians are significantly greater (1-2 orders of magnitude) than the highest estimated

2   concentrations based on reported human exposure levels. These human exposure values have

3   also been recently compared with the nominal and internal LELs in a developing zebrafish

4   screen (Alzualde et al., 2018). While the zebrafish nominal LELs were generally closer to the

5   human exposure levels than planarian nominal LELs, the majority were also outside of the range

6   of human exposure. One possible interpretation of these data is that the toxicity observed in these

7   alternative models are irrelevant because it occurs at higher concentrations than what was

8   estimated from average human exposure data. However, care has to be taken when performing

9   these comparisons, because the modeled data does not take into account differences in uptake

10   routes and levels, metabolism, or possible tissue accumulation over time as it is modeling plasma

11   concentration. Accumulation represents a tangible risk factor particularly for FRs as both BFRs

12   and OPFRs have been suggested to bioaccumulate (Hale et al., 2003; Hendriks and Westerink,

13   2015; Hou et al., 2016; Nakajima et al., 2009). In fact, when comparing internal and nominal

14   concentrations, a zebrafish study found vast differences among the different FRs, with up to

15   ~600X increases in internal FR concentrations in day 4 zebrafish larvae compared to the nominal

16   concentrations (Alzualde et al., 2018).

17   For all systems comparisons, it is imperative to understand the differences between

18   nominal water concentrations and internal doses, particularly at the site of action. For example,

19   although the human exposure data is based on human plasma concentrations, these may not

20   reflect the relevant concentrations at the site of action, i.e. the nervous system. Similarly, the

21   toxicity in the alternative animal models discussed herein are based on nominal water/media

22   concentrations and not doses. Since the effective internal concentrations in the animals are

23   unknown, strong conclusions of comparative sensitivity and their relevancy to human exposure

1   are difficult to make. Thus, understanding the potential for toxicity, even at nominal

2   concentrations which may, at first glance, seem irrelevant to human exposure, are important for

3   risk assessment.

4
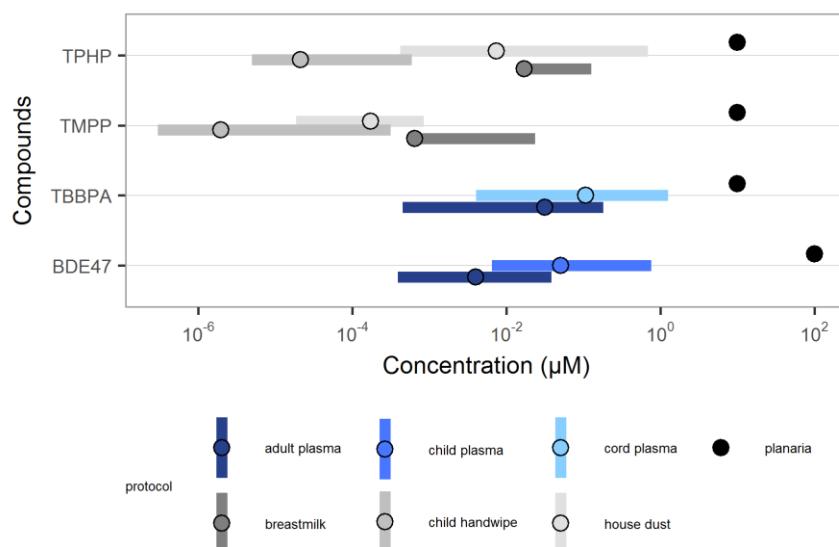
5

6   **Figure 5.** Overlay of the LELs in regenerating planarians with available human exposure data.

7   Unit-converted measured flame retardant concentrations in adult plasma, child plasma, and cord

8   plasma, as well as HTTK R-package estimated child plasma concentrations from breastmilk,

9   child handwipe, and house dust were plotted (colored bars and circles). The bars represent the

10   range, while the respective circles represent the geometric mean, median, or maximum median as

11   noted in the Methods (Section 2.5). The lower and upper range of the unit-converted (plasma)

12   and HTTK-estimated internal concentrations were calculated using the lowest and highest

13   reported values, respectively. The LELs (nominal concentrations) in planarians (black dots) were

14   based on the most potent effect at any endpoint in regenerating planarians (Table 3 and Figure

15   3).

16

4.5. Replacement FRs show comparable neurotoxicity to phased out PBDEs

In this screen, we evaluated the neurotoxicity and DNT of 15 FRs (6 BFRs, 8 OPFRs, and 1 BFR mixture) and found that 11 (3 BFRs, 7 OPFRs, and the BFR mixture) caused adverse effects in both adult and regenerating planarians. It should be noted that the 3 inactive BFRs, BDE-153, TBPH, and TBB have relatively high molecular weights (550-706 g/mol). Moreover, TBPH and BDE-153 had solubility issues at the highest tested concentration (100 and 50 µM, respectively) (Zhang et al., 2018). Thus, these issues could have contributed to their inactivity in our planarian screen. TCEP, the only OPFR inactive in our planarian screen, was also inactive in all studies (*in vivo* and *in vitro*) except for 2 zebrafish studies (Table 3) (Jarema et al., 2015; Noyes et al., 2015). Additionally, inactivity in nematodes was likely not an issue of limited test concentrations as TCEP was tested at up to 1mM without any observed effects. Only the Jarema et al and Noyes et al studies found TCEP to be bioactive in zebrafish developmental behavioral endpoints, albeit at relatively higher concentrations than the other tested FRs, suggesting relatively low potency. In agreement with the results in worms, studies with Long-Evans rats found no significant DNT was induced by oral exposure (gavage) of up to 90 mg/kg/day TCEP from gestational day 10 to weaning (Moser et al., 2015).

Several currently used FRs, including one in-use BFR, were found to be similarly or more toxic than the phased-out PBDEs in planarians. TBBPA, a BFR which is still currently widely in use, was found to cause specific, sublethal effects on unstimulated behavior (day 7 and 12) and eye regeneration in regenerating planarians (Figure 3), and was more potent than the phased-out BDE-47 and -99. In addition, effects on day 12 unstimulated behavior were selective to regenerating planarians. Together, these results suggest TBBPA caused selective adverse effects on planarian development. Although thus far guideline animal studies have not

1   demonstrated any teratogenic effects of TBBPA (EFSA, 2011), developmental defects were

2   observed in several alternative models compared here, including planarians, zebrafish (Noyes et

3   al., 2015), and nematodes (Behl et al., 2015; Behl et al., 2016) at concentrations of 10 µM and

4   below, suggesting conserved mechanisms of toxicity which may need to be re-examined in the

5   context of human risk assessment. In addition, in mice, TBBPA has been reported to cause

6   neurobehavioral defects after acute exposure and act as a thyroid endocrine disruptor (Nakajima

7   et al., 2009; Viberg and Eriksson, 2011).

8   Similarly, almost all of the tested OPFRs, which were marketed as safer alternatives to

9   BFRs, were found to be at least as toxic as the phased-out BDEs in planarians. Our results

10   resonate with previous studies in nematode, zebrafish and human, mouse, and rat cell culture,

11   which similarly found that most OPFRs had comparable activity to BFRs (Alzualde et al., 2018;

12   Behl et al., 2015; Behl et al., 2016; Jarema et al., 2015; Noyes et al., 2015). However, because of

13   the range of behaviors tested in our planarian platform, which assay various neuronal functions,

14   we were able to differentiate the toxicities of OPFRs compared to the BFRs. OPFRs generally

15   caused more sublethal, specific neurotoxic effects whereas BFRs were primarily systemically

16   toxic, as discussed in detail in section 4.6.

17

18   *4.6. Behavioral profiling in planarians adds screening specificity*

19   The majority (5/7) of the bioactive OPFRs in regenerating and adult planarians showed

20   specific, sublethal effects on scrunching, which were not seen to the same extent in the BFRs.

21   Scrunching is an oscillatory, musculature-driven escape gait in planarians that is induced by

22   specific adverse stimuli such as noxious heat (Cochet-Escartin et al., 2015), but is not a generic

23   response to all noxious environments (Cochet-Escartin et al., 2016). Although scrunching shares

42

1   qualitative features, such as muscle-based motion and increased mucus secretion, with other

2   muscle-driven body shape changes, such as peristalsis, it can be quantitatively distinguished

3   using four characteristic parameters: frequency and asymmetry of body length oscillations,

4   speed, and maximum elongation (Cochet-Escartin et al., 2015). Scrunching is also distinctively

5   different from other body undulations described in the literature, including snake-like behaviors

6   and seizures (Ross et al., 2018).

7   Scrunching defects, as tested in our HTS platform, therefore point to specific adverse

8   effects on one or multiple molecular targets controlling the underlying signaling cascade from

9   stimulus sensation to muscle contraction, mucus secretion, and locomotion. Importantly, some of

10   the molecular targets that mediate scrunching have been previously identified. For instance, it

11   has been shown that the Transient Receptor Potential Channel Ankyrin 1 (TRPA1) is required

12   for proper scrunching in response to tail amputation (Arenas et al., 2017). TRPA1 is also an

13   indirect sensor of noxious heat in planarians (Arenas et al., 2017) and we have recently

14   confirmed that TRPA1 activation is required for heat-induced scrunching (unpublished results).

15   Additionally, cholinergic neurotransmission, which is responsible for coordinating planarian

16   motor functions (Nishimura et al., 2012), is likely important for scrunching, as this gait relies on

17   muscle contraction. We have previously characterized two putative genes responsible for

18   cholinesterase activity in *D. japonica* and found them to be sensitive to inhibition by

19   organophosphorus pesticides and linked to behavioral defects (Hagstrom et al., 2017; Hagstrom

20   et al., 2018a). Organophosphorus compounds share the ability to inhibit cholinesterase, though

21   their potency can vary due to structural and pharmacokinetic differences (Taylor, 2018). Because

22   comparable concentrations of organophosphorus pesticides, such as parathion and chlorpyrifos

23   (Zhang et al., 2018), also cause scrunching defects, this phenotype is likely representative of the

1   shared cholinergic toxicity. Consistent with this idea, we found an almost complete loss of

2   cholinesterase activity in day 12 adult planarians exposed to 10 µM EHDP, which causes

3   scrunching defects. In contrast, cholinesterase activity in day 12 adult planarians exposed to 100

4   µM TCEP, which was the only inactive OFPR in our planarian screen, was indistinguishable

5   from controls (Supplementary Figure 1). Similarly, it was previously shown that DNT of 4

6   OPFRs tested in medaka was correlated with changes in acetylcholinesterase activity. For

7   example, TPHP caused developmental and behavioral defects concomitant with significant

8   acetylcholinesterase inhibition whereas TCEP, which had no significant effect on

9   acetylcholinesterase activity, had much milder toxic effects (Sun et al., 2016b). Lack of

10   significant effects of TCEP on acetylcholinesterase activity or transcription were also found in

11   developing zebrafish (Sun et al., 2016a). Together, these studies indicate that the inactivity of

12   TCEP in these different models may result from its inability to significantly inhibit

13   acetylcholinesterase at the tested concentrations.

14   Our behavior-rich planarian system can be a powerful complement to existing developing

15   zebrafish and nematode models, with their respective morphological breadth and genetic

16   tractability. Because we can directly compare toxicities in regenerating and adult planarians, we

17   are able to distinguish toxic from developmental toxic effects, which is not directly possible in

18   these other systems. For example, we found developmentally selective effects on day 12

19   unstimulated behavior following exposure to 10µM TBBPA in regenerating but not adult

20   planarians, suggesting TBBPA may act on specific developmental processes or on processes

21   which are more sensitive during development. The shared effects and potencies of the other

22   bioactive FRs on both regenerating and adult planarians suggest a potential shared mechanism of

23   toxicity, with effects on overall neural function rather than neurodevelopment directly. Together,

44

1   these examples illustrate that, in a single screen, the planarian platform is uniquely able to

2   differentiate DNT from neurotoxicity and distinguish effects on specific neuronal functions.

3

4

5   **5. Conclusions**

6   Here, we showed, using the concrete example of screening 15 FRs, that our planarian

7   HTS platform generates robust data and allows us to distinguish systemic toxicity from

8   neurotoxicity and developmentally selective effects from general neurotoxicity. Additionally, the

9   diverse behaviors tested in this planarian system allowed us to differentiate class-dictated

10   phenotypic signatures, particularly evidenced by the shared effect of OPFRs, but not BFRs, on

11   planarian scrunching. Concordance between this planarian study and previous studies in cell

12   culture, nematodes and developing zebrafish, demonstrating the comparable toxicity of

13   replacement OPFRs with BFRs emphasizes the urgent need for further evaluation of OPFRs in

14   mammalian systems.

15

16

17   **6. Supplementary data description**

18   Compiled data for each endpoint and comparisons between individual replicates can be found in

19   Supplementary File 1. Additional data on the differences between our determined hits and initial

20   statistically significant effects can be found in Supplementary Table 1. Results from the

21   cholinesterase activity staining of EHDP and TCEP can be found in Supplementary Figure 1.

22

23

9. References

Alzualde, A., Behl, M., Sipes, N. S., Hsieh, J.-H., Alday, A., Tice, R. R., Paules, R. S., Muriana,
    A. and Quevedo, C. (2018). Toxicity profiling of flame retardants in zebrafish embryos
    using a battery of assays for teratogenicity, behavior, cardiotoxicity, and hepatotoxicity
    toward human relevance. *Neurotoxicol. Teratol.*

Arenas, O. M., Zaharieva, E. E., Para, A., Vásquez-Doorman, C., Petersen, C. P. and Gallio, M.
    (2017). Activation of planarian TRPA1 by reactive oxygen species reveals a conserved
    mechanism for animal nociception. *Nat. Neurosci.* **20**, 1686–1693.

Bailey, J. M. and Levin, E. D. (2015). Neurotoxicity of FireMaster 550® in zebrafish (Danio
    rerio): Chronic developmental and acute adolescent exposures. *Neurotoxicol. Teratol.* **52**,
    210–219.

1 Bal-Price, A. K., Coecke, S., Costa, L., Crofton, K. M., Fritsche, E., Goldberg, A., Grandjean, P.,

2 Lein, P. J., Li, A., Lucchini, R., et al. (2012). Advancing the science of developmental

3 neurotoxicity (DNT): testing for better safety evaluation. *ALTEX* **29**, 202–15.

4 Behl, M., Hsieh, J. H., Shafer, T. J., Mundy, W. R., Rice, J. R., Boyd, W. A., Freedman, J. H.,

5 Hunter, E. S., Jarema, K. A., Padilla, S., et al. (2015). Use of alternative assays to identify

6 and prioritize organophosphorus flame retardants for potential developmental and

7 neurotoxicity. *Neurotoxicol. Teratol.*

8 Behl, M., Rice, J. R., Smith, M. V, Co, C. A., Bridge, M. F., Hsieh, J.-H., Freedman, J. H. and

9 Boyd, W. A. (2016). Comparative toxicity of organophosphate flame retardants and

10 polybrominated diphenyl ethers to Caenorhabditis elegans. *Toxicol. Sci.* **154**, 241–252.

11 Behl, M., Ryan, K., Hsieh, J.-H., Parham, F., Shapiro, A., Collins, B. J., Birnbaum, L. S.,

12 Bucher, J. R., Walker, N. J., Foster, P. M., et al. (2018). Screening for developmental

13 neurotoxicity (DNT) at the National Toxicology Program: The future is now! *Toxicol. Sci.*

14 Birkholz, T. R. and Beane, W. S. (2017). The planarian TRPA1 homolog mediates extraocular

15 behavioral responses to near-ultraviolet light. *J. Exp. Biol.* **220**, 2616–2625.

16 Cano-Sancho, G., Smith, A. and La Merrill, M. A. (2017). Triphenyl phosphate enhances

17 adipogenic differentiation, glucose uptake and lipolysis via endocrine and noradrenergic

18 mechanisms. *Toxicol. Vitr.* **40**, 280–288.

19 Cariou, R., Antignac, J. P., Zalko, D., Berrebi, A., Cravedi, J. P., Maume, D., Marchand, P.,

20 Monteau, F., Riu, A., Andre, F., et al. (2008). Exposure assessment of French women and

21 their newborns to tetrabromobisphenol-A: Occurrence measurements in maternal adipose

22 tissue, serum, breast milk and cord serum. *Chemosphere* **73**, 1036–1041.

Cebrià, F. (2007). Regenerating the central nervous system: how easy for planarians! *Dev. Genes Evol.* **217**, 733–748.

Cebrià, F., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T. and Agata, K. (2002a). Dissecting planarian central nervous system regeneration by the expression of neural-specific genes. *Dev. Growth Differ.* **44**, 135–46.

Cebrià, F., Kudome, T., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T. and Agata, K. (2002b). The expression of neural-specific genes reveals the structural and molecular complexity of the planarian central nervous system. *Mech. Dev.* **116**, 199–204.

Cochet-Escartin, O., Mickolajczk, K. J. and Collins, E.-M. S. (2015). Scrunching: a novel escape gait in planarians. *Phys. Biol.* **12**, 055001.

Cochet-Escartin, O., Carter, J. A., Chakraverti-Wuerthwein, M., Sinha, J. and Collins, E. M. S. (2016). Slo1 regulates ethanol-induced scrunching in freshwater planarians. *Phys. Biol.* **13**, 1–12.

Cope, R. B., Kacew, S. and Dourson, M. (2015). A reproductive, developmental and neurobehavioral study following oral exposure of tetrabromobisphenol A on Sprague-Dawley rats. *Toxicology* **329**, 49–59.

EFSA (2011). European Food Safety Authority (EFSA) panel on contaminants in the food chain (CONTAM): Scientific opinion on tetrabromobisphenol A (TBBPA) and its derivatives in food. *EFSA J.*

Fritsche, E., Grandjean, P., Crofton, K. M., Aschner, M., Goldberg, A., Heinonen, T., Hessel, E. V. S., Hogberg, H. T., Bennekou, S. H., Lein, P. J., et al. (2018). Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity

(DNT) testing for regulatory purposes. *Toxicol. Appl. Pharmacol.* **354**, 3–6.

Glazer, L., Wells, C. N., Drastal, M., Odamah, K. A., Galat, R. E., Behl, M. and Levin, E. D. (2018). Developmental exposure to low concentrations of two brominated flame retardants, BDE-47 and BDE-99, causes life-long behavioral alterations in zebrafish. *Neurotoxicology* **66**, 221–232.

Hagstrom, D., Cochet-Escartin, O., Zhang, S., Khuu, C. and Collins, E.-M. S. (2015). Freshwater planarians as an alternative animal model for neurotoxicology. *Toxicol. Sci.* **147**, 270–285.

Hagstrom, D., Cochet-Escartin, O. and Collins, E.-M. S. (2016). Planarian brain regeneration as a model system for developmental neurotoxicology. *Regeneration* **3**, 65–77.

Hagstrom, D., Hirokawa, H., Zhang, L., Radić, Z., Taylor, P. and Collins, E.-M. S. (2017). Planarian cholinesterase: in vitro characterization of an evolutionarily ancient enzyme to study organophosphorus pesticide toxicity and reactivation. *Arch. Toxicol.* **91**, 2837–2847.

Hagstrom, D., Zhang, S., Ho, A., Tsai, E. S., Radić, Z., Jahromi, A., Kaj, K. J., He, Y., Taylor, P. and Collins, E. M. S. (2018a). Planarian cholinesterase: molecular and functional characterization of an evolutionarily ancient enzyme to study organophosphorus pesticide toxicity. *Arch. Toxicol.* **92**, 1161–1176.

Hagstrom, D., Truong, L., Zhang, S., Tanguay, R. and Collins, E.-M. S. (2018b). Comparative analysis of zebrafish and planarian model systems for developmental neurotoxicity screens using an 87-compound library. *Toxicol. Sci.* **kfy180**,.

Hale, R. C., Alaee, M., Manchester-Neesvig, J. B., Stapleton, H. M. and Ikonomou, M. G. (2003). Polybrominated diphenyl ether flame retardants in the North American environment. *Environ. Int.* **29**, 771–779.

Hendriks, H. S. and Westerink, R. H. S. (2015). Neurotoxicity and risk assessment of brominated and alternative flame retardants. *Neurotoxicol. Teratol.* **52**, 248–269.

Hou, R., Xu, Y. and Wang, Z. (2016). Review of OPFRs in animals and humans: Absorption, bioaccumulation, metabolism, and internal exposure research. *Chemosphere* **153**, 78–90.

Inoue, T., Yamashita, T. and Agata, K. (2014). Thermosensory signaling by TRPM is processed by brain serotonergic neurons to produce planarian thermotaxis. *J. Neurosci.* **34**, 15701–14.

Jarema, K. A., Hunter, D. L., Shaffer, R. M., Behl, M. and Padilla, S. (2015). Acute and developmental behavioral effects of flame retardants and related chemicals in zebrafish. *Neurotoxicol. Teratol.*

Kim, J. W., Isobe, T., Muto, M., Tue, N. M., Katsura, K., Malarvannan, G., Sudaryanto, A., Chang, K. H., Prudente, M., Viet, P. H., et al. (2014). Organophosphorus flame retardants (PFRs) in human breast milk from several Asian countries. *Chemosphere* **116**, 91–97.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. medica* **22**, 276–82.

Meeker, J. D. and Stapleton, H. M. (2010). House dust concentrations of organophosphate flame retardants in relation to hormone levels and semen quality parameters. *Environ. Health Perspect.* **118**, 318–23.

Moser, V. C., Phillips, P. M., Hedge, J. M. and McDaniel, K. L. (2015). Neurotoxicological and thyroid evaluations of rats developmentally exposed to tris(1,3-dichloro-2-propyl)phosphate (TDCIPP) and tris(2-chloro-2-ethyl)phosphate (TCEP). *Neurotoxicol. Teratol.* **52**, 236–247.

Nakajima, A., Saigusa, D., Tetsu, N., Yamakuni, T., Tomioka, Y. and Hishinuma, T. (2009). Neurobehavioral effects of tetrabromobisphenol A, a brominated flame retardant, in mice. *Toxicol. Lett.* **189**, 78–83.

1   National Toxicology Program (1991). Toxicology and Carcinogenesis Studies of Tris(2-
2   chloroethyl) Phosphate (CAS No. 115-96-8) in F344/N Rats and B6C3F1 Mice (Gavage
3   Studies). *Natl. Toxicol. Program Tech. Rep. Ser.* **391**, 1–233.

4   Nishimura, K., Kitamura, Y., Taniguchi, T. and Agata, K. (2010). Analysis of motor function
5   modulated by cholinergic neurons in planarian Dugesia japonica. *Neuroscience* **168**, 18–30.

6   Nishimura, O., Hirao, Y., Tarui, H. and Agata, K. (2012). Comparative transcriptome analysis
7   between planarian Dugesia japonica and other platyhelminth species. *BMC Genomics* **13**,
8   289.

9   Noyes, P. D., Haggard, D. E., Gonnerman, G. D. and Tanguay, R. L. (2015). Advanced
10   morphological-behavioral test platform reveals neurodevelopmental defects in embryonic
11   zebrafish exposed to comprehensive suite of halogenated and organophosphate flame
12   retardants. *Toxicol. Sci.* **145**, 177–195.

13   Oliveri, A. N., Bailey, J. M. and Levin, E. D. (2015). Developmental exposure to
14   organophosphate flame retardants causes behavioral effects in larval and adult zebrafish.
15   *Neurotoxicol. Teratol.* **52**, 220–227.

16   Pearce, R. G., Setzer, R. W., Strope, C. L., Sipes, N. S. and Wambaugh, J. F. (2017). **httk** : *R*
17   Package for High-Throughput Toxicokinetics. *J. Stat. Softw.* **79**,.

18   Ross, K. G., Currie, K. W., Pearson, B. J. and Zayas, R. M. (2017). Nervous system development
19   and regeneration in freshwater planarians. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e266.

20   Ross, K. G., Molinaro, A. M., Romero, C., Dockter, B., Cable, K. L., Gonzalez, K., Zhang, S.,
21   Collins, E. M. S., Pearson, B. J. and Zayas, R. M. (2018). SoxB1 Activity Regulates
22   Sensory Neuron Regeneration, Maintenance, and Function in Planarians. *Dev. Cell* **47**, 331–

1   347.e5.

2   Rotroff, D. M., Wetmore, B. A., Dix, D. J., Ferguson, S. S., Clewell, H. J., Houck, K. A.,

3   Lecluyse, E. L., Andersen, M. E., Judson, R. S., Smith, C. M., et al. (2010). Incorporating

4   human dosimetry and exposure into high-throughput in vitro toxicity screening. *Toxicol.*

5   *Sci.* **117**, 348–58.

6   Sipes, N. S., Wambaugh, J. F., Pearce, R., Auerbach, S. S., Wetmore, B. A., Hsieh, J.-H.,

7   Shapiro, A. J., Svoboda, D., DeVito, M. J. and Ferguson, S. S. (2017). An Intuitive

8   Approach for Predicting Potential Human Health Risk with the Tox21 10k Library.

9   *Environ. Sci. Technol.* **51**, 10786–10796.

10 Slotkin, T. A., Skavicus, S., Stapleton, H. M. and Seidler, F. J. (2017). Brominated and

11   organophosphate flame retardants target different neurodevelopmental stages, characterized

12   with embryonic neural stem cells and neuronotypic PC12 cells. *Toxicology* **390**, 32–42.

13 Smirnova, L., Hogberg, H. T., Leist, M. and Hartung, T. (2014). Food for Thought …:

14   Developmental Neurotoxicity – Challenges in the 21(st) Century and In Vitro

15   Opportunities. *ALTEX* **31**, 129–156.

16 Stapleton, H. M., Klosterhaus, S., Eagle, S., Fuh, J., Meeker, J. D., Blum, A. and Webster, T. F.

17   (2009). Detection of Organophosphate Flame Retardants in Furniture Foam and U.S. House

18   Dust. *Environ. Sci. Technol.* **43**, 7490–7495.

19 Stapleton, H. M., Eagle, S., Sjödin, A. and Webster, T. F. (2012). Serum PBDEs in a North

20   Carolina toddler cohort: Associations with handwipes, house dust, and socioeconomic

21   variables. *Environ. Health Perspect.* **120**, 1049–1054.

22 Stapleton, H. M., Misenheimer, J., Hoffman, K. and Webster, T. F. (2014). Flame retardant

1 associations between children's handwipes and house dust. *Chemosphere* **116**, 54–60.

2 Sugeng, E. J., Leonards, P. E. G. and van de Bor, M. (2017). Brominated and organophosphorus

3 flame retardants in body wipes and house dust, and an estimation of house dust hand-

4 loadings in Dutch toddlers. *Environ. Res.* **158**, 789–797.

5 Sun, L., Xu, W., Peng, T., Chen, H., Ren, L., Tan, H., Xiao, D., Qian, H. and Fu, Z. (2016a).

6 Developmental exposure of zebrafish larvae to organophosphate flame retardants causes

7 neurotoxicity. *Neurotoxicol. Teratol.* **55**, 16–22.

8 Sun, L., Tan, H., Peng, T., Wang, S., Xu, W., Qian, H., Jin, Y. and Fu, Z. (2016b).

9 Developmental neurotoxicity of organophosphate flame retardants in early life stages of

10 Japanese medaka (*Oryzias latipes*). *Environ. Toxicol. Chem.* **35**, 2931–2940.

11 Taylor, P. (2018). Anticholinesterase agents. In *Goodman and Gilman's The Pharmacological*

12 *Basis of Therapeutics* (ed. Laurence L Brunton), pp. 163–176. San Francisco: McGraw Hill

13 Education.

14 Tohyama, C. (2016). Developmental neurotoxicity test guidelines: problems and perspectives. *J.*

15 *Toxicol. Sci.* **41**, SP69-SP79.

16 Truong, L., Reif, D. M., St Mary, L., Geier, M. C., Truong, H. D. and Tanguay, R. L. (2014).

17 Multidimensional in vivo hazard assessment using zebrafish. *Toxicol. Sci.* **137**, 212–33.

18 Tung, E. W. Y., Ahmed, S., Peshdary, V. and Atlas, E. (2017). Firemaster® 550 and its

19 components isopropylated triphenyl phosphate and triphenyl phosphate enhance

20 adipogenesis and transcriptional activity of peroxisome proliferator activated receptor

21 (Pparγ) on the adipocyte protein 2 (aP2) promoter. *PLoS One* **12**, e0175855.

22 Usenko, C., Abel, E., Hopkins, A., Martinez, G., Tijerina, J., Kudela, M., Norris, N., Joudeh, L.

53

and Bruce, E. (2016). Evaluation of Common Use Brominated Flame Retardant (BFR) Toxicity Using a Zebrafish Embryo Model. *Toxics* **4**, 21.

Van Der Veen, I. and De Boer, J. (2012). Phosphorus flame retardants: Properties, production, environmental occurrence, toxicity and analysis. *Chemosphere* **88**, 1119–1153.

Viberg, H. and Eriksson, P. (2011). Differences in neonatal neurotoxicity of brominated flame retardants, PBDE 99 and TBBPA, in mice. *Toxicology* **289**, 59–65.

Wambaugh, J. F., Wetmore, B. A., Pearce, R., Strope, C., Goldsmith, R., Sluka, J. P., Sedykh, A., Tropsha, A., Bosgra, S., Shah, I., et al. (2015). Toxicokinetic triage for environmental chemicals. *Toxicol. Sci.* **147**, 55–67.

Wang, H. S., Jiang, G. M., Chen, Z. J., Du, J., Man, Y. B., Giesy, J. P., Wong, C. K. C. and Wong, M. H. (2013). Concentrations and congener profiles of polybrominated diphenyl ethers (PBDEs) in blood plasma from Hong Kong: Implications for sources and exposure route. *J. Hazard. Mater.* **261**, 253–259.

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Sochaski, M. A., Rotroff, D. M., Freeman, K., Clewell, H. J., Dix, D. J., Andersen, M. E., Houck, K. A., et al. (2012). Integration of Dosimetry, Exposure, and High-Throughput Screening Data in Chemical Toxicity Assessment. *Toxicol. Sci.* **125**, 157–174.

Wetmore, B. A., Wambaugh, J. F., Ferguson, S. S., Li, L., Clewell, H. J., Judson, R. S., Freeman, K., Bao, W., Sochaski, M. A., Chu, T.-M., et al. (2013). Relative Impact of Incorporating Pharmacokinetics on Predicting In Vivo Hazard and Mode of Action from High-Throughput In Vitro Toxicity Assays. *Toxicol. Sci.* **132**, 327–346.

Zhang, S., Hagstrom, D., Hayes, P., Graham, A. and Collins, E.-M. S. (2018). Multi-behavioral

54

1   endpoint testing of an 87-chemical compound library in freshwater planarians. *Toxicol. Sci.*

2   Zheng, D.-M., Xie, H.-Q., Wang, A.-T. and Wu, C.-C. (2011). The nerve system identificiation

3   by histochemical localization of acetylcholinesterase in planarian Dugesia japonica. *Chinese*

4   *J. Zool.* **45**, 68–75.