Swarthmore College

Works

Biology Faculty Works

Biology

6-8-2018

Multi-Behavioral Endpoint Testing Of An 87-Chemical Compound Library In Freshwater Planarians

S. Zhang

D. Hagstrom

P. Hayes

A. Graham

Eva-Maria S. Collins Swarthmore College, ecollin3@swarthmore.edu

Follow this and additional works at: https://works.swarthmore.edu/fac-biology Part of the Biology Commons, and the Systems Biology Commons Let us know how access to these works benefits you

Recommended Citation

S. Zhang, D. Hagstrom, P. Hayes, A. Graham, and Eva-Maria S. Collins. (2018). "Multi-Behavioral Endpoint Testing Of An 87-Chemical Compound Library In Freshwater Planarians". *Toxicological Sciences*. DOI: 10.1093/toxsci/kfy145 https://works.swarthmore.edu/fac-biology/560

This work is brought to you for free and open access by . It has been accepted for inclusion in Biology Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.

Multi-behavioral endpoint testing of an 87-chemical compound library in freshwater planarians

Siqi Zhang^{*1}, Danielle Hagstrom^{†1}, Patrick Hayes[‡], Aaron Graham[‡], and Eva-Maria S.

Collins^{†§¶}

*Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

†Division of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92093,

USA

‡ Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA92093, USA

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

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

^IThese authors contributed equally to this work.

Correspondence should be addressed to Eva-Maria S. Collins; e-mail: ecollin3@swarthmore.edu; Telephone: 610-690-5380

Running Title: Planarian 87-compound library screen

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society of Toxicology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Abstract

There is an increased recognition in the field of toxicology of the value of medium-tohigh-throughput screening methods using *in vitro* and alternative animal models. We have previously introduced the asexual freshwater planarian *Dugesia japonica* as a new alternative animal model and proposed that it is particularly well-suited for the study of developmental neurotoxicology. In this paper, we discuss how we have expanded and automated our screening methodology to allow for fast screening of multiple behavioral endpoints, developmental toxicity, and mortality. Using an 87-compound library provided by the National Toxicology Program (NTP), consisting of known and suspected neurotoxicants, including drugs, flame retardants, industrial chemicals, polycyclic aromatic hydrocarbons (PAHs), pesticides and presumptive negative controls, we further evaluate the benefits and limitations of the system for medium-throughput screening, focusing on the technical aspects of the system. We show that, in the context of this library, planarians are the most sensitive to pesticides with 16/16 compounds causing toxicity and the least sensitive to PAHs, with only 5/17 causing toxicity. Furthermore, while none of the presumptive negative controls were bioactive in adult planarians, 2/5, acetaminophen and acetylsalicylic acid, were bioactive in regenerating worms. Notably, these compounds were previously reported as developmentally toxic in mammalian studies. Through parallel screening of adults and developing animals, planarians are thus a useful model to detect such developmental-specific effects, which was observed for 13 chemicals in this library. We use the data and experience gained from this screen to propose guidelines for best practices when using planarians for toxicology screens.

Keywords: planarian, developmental neurotoxicity, behavior, medium-throughput screening, Tox21, NTP-library, alternative animal models

Abbreviations

HTS, high throughput screening; MTS, medium throughput screening; NTP, National Toxicology Program; IO, Instant Ocean; PAH, polycyclic aromatic hydrocarbon; DMSO, dimethyl sulfoxide; rpm, rotation per minute; COM, center of mass; fps, frames per second; LOEL, lowest observed effect level; TRP, Transient Receptor Potential

Introduction

It has been nearly a decade since the launch of the "Toxicology Testing in the 21st century" (Tox21; www.tox21.gov) federal initiative to transform toxicology testing in the United States. Its ongoing goal is to dramatically increase the coverage of chemical testing by replacing traditional mammalian models with alternative testing strategies amenable to high-throughput screening (HTS) (Collins et al., 2008). Since its inception, thousands of chemicals have been screened in vitro using HTS robotic systems to identify mechanisms of action and prioritize chemicals for further targeted testing. However, connecting those HTS data to their in vivo relevancy to be predictive of effects on human health remains challenging as important aspects of biology, such as xenobiotic metabolism and interactions between cell types, are inherently missing in these *in vitro* systems. In addition, although these assays often focus on key molecular and cellular targets underlying known toxicity pathways, more knowledge is needed to connect these molecular and cellular effects to functional consequences on organismal health to discern their significance. Realizing this need and the urgency of the matter, the development of medium-throughput screening (MTS)-amenable alternative animal models, such as zebrafish and nematodes, was encouraged as part of the Tox21 initiative. These animal models are attractive MTS toxicology systems due to their ease of breeding and chemical administration, low cost,

small size, short developmental time, and genetic tractability (Boyd et al., 2012; Boyd et al., 2015; Hill et al., 2005; Tejeda-Benitez and Olivero-Verbel, 2016; Truong et al., 2014). Moreover, each system provides unique advantages. For example, the transparency of zebrafish larvae, which develop externally, allows for a breadth of morphological assessments of the development of internal structures in living animals (Kimmel et al., 1995; Truong et al., 2014). However, despite these advantages, the toxicology community remains divided on the added value of these alternative systems, particularly as each has its own drawbacks, species-specific sensitivities and discrepancies with humans, as with any system (Boyd et al., 2015; Scholz, 2013).

A battery approach using multiple complementary testing platforms allows for comparative analyses to find concordance between systems and produce more weight of evidence for reliable and relevant predictions of effects on human health, as demonstrated by a recent battery screen on organophosphorus flame retardants (Behl et al., 2015). These predictions can then be verified by targeted testing in mammalian models, which, although not without caveats, are still considered the gold standard in toxicology, particularly for regulatory decisions (Tsuji and Crofton, 2012).

We have previously introduced the freshwater planarian *Dugesia japonica* as a new alternative animal model for developmental neurotoxicology and shown that it possesses comparable sensitivity to other, more established alternative models (Hagstrom et al., 2015). In addition, the planarian system offers the unique advantage to study adult and regenerating/developing animals in parallel with the same assays, because in this asexual species the sole form of neurodevelopment is neuroregeneration of a head from a tail piece following fission. Finally, planarians have a large behavioral repertoire that can be quantified and assessed in a fully automated fashion, providing multiple distinct endpoints of neuronal function.

Importantly, the planarian nervous system contains most of the same neurotransmitters as the mammalian brain and is considered more structurally similar to the vertebrate brain than other invertebrate brains (Buttarelli et al., 2008; Cebrià, 2007; Mineta et al., 2003; Ross et al., 2017; Umesono et al., 2011). A brief review of the planarian nervous system and of neuroregeneration can be found in Supplementary Information, Section 1. Moreover, we have recently reviewed the history, challenges and benefits of planarians as a model for neurotoxicology (Hagstrom et al., 2016).

While our previous work demonstrated the potential of *D. japonica* for toxicology screens, it was limited in scope (10 compounds, including controls) (Hagstrom et al., 2015). Most of the experiments and analysis were conducted manually, which limited throughput and scalability. Our screening platform has since been greatly expanded and optimized to incorporate more behavioral endpoints that are all assayed in a fully automated fashion.

In this study, we evaluate the capabilities and limitations of this improved planarian MTS platform by testing a library of 87 compounds provided by the National Toxicology Program (NTP), consisting of known and suspected developmental neurotoxicants and negative controls. This compound library, which has also been tested in other alternative systems, including zebrafish and *in vitro* cell culture systems (see other articles in this special issue), gives us a unique opportunity to test the robustness and relevancy of the planarian system as a whole and of the specific endpoints we have developed to assay different neuronal functions. We focus on evaluating the technical aspects of our expanded screening platform and the utility of the planarian model system for toxicology screens, setting clear standards and challenges that need to be addressed for the field going forward. A direct comparison of the results of this planarian

screen with a zebrafish model, and with available mammalian data, are the focus of a companion paper in this Special Issue (Hagstrom et al.).

Material and methods

Test animals: Freshwater planarians of the species *D. japonica*, originally obtained from Shanghai University, China, and cultivated in our lab >5 years, were used for all tests. Planarians were stored in 1x Instant Ocean (IO, Blacksburg, VA) in Tupperware containers at 20°C in a Panasonic refrigerated incubator in the dark. Animals were fed organic chicken or beef liver purchased from a local butcher twice a week. Planarian containers were cleaned 3 times a week per standard protocols (Dunkel et al., 2011). Animals were starved for at least 5 days before being used for experiments and their containers were cleaned immediately prior to worm selection for experiments. Test worms were manually selected to fall within a certain range of sizes and we found full worm length, after automated size measurement, to be 7.3mm +/- 2.3mm (mean +/- SD), and tail worm length to be 7.3mm +/- 2.7mm (mean +/- SD). Slightly larger intact planarians (~1-2 mm larger to account for the size of the head) were chosen for regenerating tail experiments such that the final sizes of the amputated tail pieces were similar to the full/adult test planarians. Some animals were recovered after the screen and reintroduced into the normal population after a minimum of 4 weeks of separate care. As planarians undergo dynamic turnover of all cell types within a few weeks (Rink, 2013) and as we observed no qualitative differences in behavior between recovered and wild-type animals, these recovered worms were considered functionally wild-type. For all experiments, only fully regenerated worms which had not been fed within one week and which were found gliding normally in the container were used. To study regenerating animals, on day 1, intact worms were amputated, by

cutting posterior to the auricles and anterior to the pharynx with an ethanol-sterilized razor blade, no more than 3 hours before the compounds were added. During the course of the screen, some animals underwent fission producing at least 2 pieces (a head and a tail piece) (see below and Supplementary Information, Section 4). To obtain full and tail worms of comparable size, we amputate slightly larger worms to obtain the tail pieces. Since fission probability increases with worm size (Carter et al., 2015; Yang et al., 2017) and decapitation (Bronsted, 1955; Hori and Kishida, 1998; Morita and Best, 1984), fission primarily occurred for tail worms. For these cases, only the head piece was considered in all morphological and behavioral analyses, as this would represent the first regenerated brain.

Test compounds: The 87-compound library (summarized in Supplementary Table 1) was provided by the NTP and included 5 categories: pesticides, flame retardants, drugs, industrial compounds and polycyclic aromatic hydrocarbons (PAHs) (Behl et al., 2018). Five negative controls were also included. The compounds were provided as ~20mM stocks (or lower) in 100% dimethyl sulfoxide (DMSO, Gaylord Chemicals, Slidell, LA) in a 96-well plate. The master library was stored at -80°C.

Chemical preparation and screen setup: The 87-chemical library was separated into 5 "Chemical Sets" of 18 (sets 1-4) or 15 (set 5) chemicals (Supplementary Table 1). Chemicals in the same Chemical Set were tested on the same day, i.e. the same experiment. All chemicals, regardless of provided concentration, were treated the same. 0.5% DMSO was used as solvent control, because we have previously shown that there are no effects on planarian morphology or behavior at this concentration (Hagstrom et al., 2015). To keep the final DMSO concentration constant at 0.5%, the highest concentration tested in the screening process was a 200-fold

dilution of the original provided chemical stock. Subsequent concentrations were a 10-fold dilution of the previous. Thus, each compound was tested at 5 concentrations, generally ranging from 10nM to 100µM (with some exceptions, see Supplementary Table 1). Each 48-well screening plate assayed n=8 planarians in a 0.5% DMSO control, and n=8 worms each per concentration of chemical (5 test concentrations per plate in total) (Figure 1). Experiments were performed in triplicate (independent experiments performed on different days, final n=24) with the concentrations shifted down two rows (one row in run D, see raw data in the Dryad Digital Repository (doi: 10.5061/dryad.mk6m608)) with each replicate to control for edge effects. For each chemical and each experiment, 2 plates, one containing full (intact) planarians and one containing regenerating tails, were assayed. Screening was performed on day 7 and day 12.

Plate setup and storage: 200X stock plates of the tested chemicals were prepared ahead of time by transferring 50µl of the provided chemical stock into one well of a 48-well plate (Genesee, San Diego, CA). 10-fold serial dilutions were performed in DMSO in the same plate using a multi-pipettor to create the remaining stock concentrations. The control well contained DMSO only. These plates were sealed with foil seals (Thermo Scientific, Waltham, MA) and stored at -20°C. On the day of plate set-up, the 200X stock plates were thawed at room temperature for approximately 30 minutes. 10X stocks plates were then made by diluting the 200X stocks 20X in IO water. Dilutions were mixed by rotation on an orbital shaker for approximately 10 minutes before use. The highest concentration of some chemicals, noted in Supplementary Table 1, precipitated out of solution in the 10X stock plates due to low solubility in water.

Screening plates were prepared by transferring individual full planarians or amputated tail pieces into the wells of a 48-well plate with 200µl of IO water using a P1000 pipet with a cut-off tip. A multi-pipettor was used to remove 20µl of IO water from each well and add 20µl of the appropriate 10X stock solution. The plates were sealed with ThermalSeal RTS seals (Excel Scientific, Victorville, CA) to prevent evaporation and gas exchange with the environment. The plates were stored, without their lids, in stacks in the dark at room temperature when not being screened. Prepared plates were only moved to the screening platform when screened at day 7 and day 12.

Screening platform: We have further automated and expanded the custom-built planarian screening platform introduced in (Hagstrom et al., 2015). The new platform consists of a commercial robotic microplate handler (Hudson Robotics, Springfield Township, NJ), two custom-built imaging systems and multiple assay stations (Figure 1). One imaging system is specifically used to image individual planarians at high spatial resolution to allow for quantification of lethality, morphology and eye regeneration. It consists of 4 monochromatic Flea USB3 cameras (FLIR Systems Inc., Wilsonville, OR), each equipped with a fixed-focal (16mm) optical lens (Tamron, Saitama, Japan) and 5mm spacer (Edmund Optics, Santa Monica, CA). Each camera is used to image a single well, thus 4 wells are imaged simultaneously and the entire plate is scanned in the x- and y- directions. The second imaging system consists of one monochromatic Flea USB3 camera, equipped with a fixed-focal (25mm) double-gauss lens (Edmund Optics) and red filter (Roscolux, Stamford, CT), which is used to image the whole plate from above for all behavioral assays. To prevent angular distortion on the edge of the wells, a Fresnel lens (MagniPros, South El Monte, CA) is placed on top of the plate when imaging with the single camera. All cameras are mounted on a custom rail platform (Inventables Inc., Chicago,

IL), which enables x-, y- and linear motion. All assays were imaged at a frame rate of 5 frames per second . Different assay stations were designed specifically for different assays, as explained below. The imaging systems, assay stations and plate handler were controlled by the computer. The stimuli and illuminations in the assays were mainly controlled via Arduino (Arduino, Somerville, MA). Image acquisition was controlled through custom LabVIEW scripts. All assays were performed in the following order, whereby the notation in brackets indicates on which day(s) the assay was performed: phototaxis (d7/d12), unstimulated locomotion (d7/d12), lethality/regeneration (d7/d12), thermotaxis (d12) and scrunching (d12) (see also Figure 1). Any data analysis which had to be cross-checked manually was performed blinded by a single investigator, who was not given the chemical identity of the plates. The raw data are provided in the Dryad Digital Repository (doi:10.5061/dryad.mk6m608).

Lethality assay: To assay planarian lethality and eye regeneration, high-resolution imaging of each individual well was performed. Since planarians tend to rest on the edge of the well, prior to imaging each set of 4 wells, the screening plate was placed on a microplate orbital shaker (Big Bear Automation, Santa Clara, CA) and shaken for 1 second at 800 rotations per minute (rpm) to force the worms to the center of the well. Each well was then imaged for 10 seconds. The plate was illuminated from above by red LED strings (Amazon, Seattle, WA) attached around the camera lens.

Semi-automatic analysis was performed on the image sequence of each single planarian to determine whether the animal was alive or dead. Death was determined by the absence of the worm or the presence of a disintegrating body, using the fact that a dead planarian usually disintegrates (Buchanan, 1935). An alive planarian was marked as '0' and a dead one as '1' (Figure 2A-B). If the worm "suicides" by leaving the water and thus drying out, the respective well would be marked as '10' and discarded in the data analysis.

Lethality was calculated as

$$lethality = \frac{total \ number \ of \ dead \ planarians}{total \ number \ of \ planarians}$$

Where "total number of planarians" excludes any suicides. For compounds which showed significant lethality in the concentration range tested (see Statistical Testing section below), the fraction of dead planarians as a function of concentration at days 7 and 12 was fitted as described in (Hagstrom et al., 2015) using a Hill equation to obtain the LC_{50} (Supplementary Figure S1). Of note, fissioned planarians in a single well were marked as one unit. If any fissioned piece was alive in one well, this well was considered to contain an alive worm and marked as '1'.

Eye regeneration assay: Eye regeneration data was also collected from the highresolution imaging performed in the lethality assay (described above). Image analysis was performed with a custom Python-based machine learning algorithm using a transfer learning neural network (Pan and Yang, 2010). A custom pre-processing program was used in Python to crop 100 x 100 pixel² images of a planarian's head region from the original images. The cropped images were imported into the neural network, which categorized the worms based on a prediction of the number of eyes in the images: normal (2 eyes), abnormal (0, 1 eye or >2 eyes), and invalid (for example, when the worm was on the edge of the well, flipped over, or the head region was not properly cropped) (Figure 2D-G). The neural network was trained using a training set consisting of 2206 images of normal eyes, 1047 images of abnormal eyes and 6703 images with undetectable quality. The training set was labeled semi-manually with a customized computer program. The prediction results of each image for each alive planarian were integrated using a custom MATLAB script to make the final decision of the number of eyes in this regenerating animal. If more than 1 image frame predicted normal eyes, the planarian was determined to have normal eyes. If more than 1 image frame predicted abnormal eyes, but no image frame predicted normal eyes, the worm was determined to have abnormal eyes. In all other cases, the image sequence was an invalid case, due to lack of analyzable images resulting from worm positioning in the well which obscured the eyes, see Figure 2G), and discarded in the following analysis. Since the prediction of the "abnormal" category was often inaccurate because of the small training set and large variability in data, we manually double checked all results predicted to be "abnormal" and invalid. For planarians which underwent fission during the course of the screen, resulting in more than 1 animal in a well, the number of regenerated eyes in the head piece was scored manually. The eye regeneration rate was calculated as

 $eye \ regeneration \ rate = \ \frac{number \ of \ planarians \ with \ 2 \ eyes \ regenerated}{number \ of \ analyzable \ planarians}$

Unstimulated behavioral assay: As planarians tend to rest when stored in the dark, screening plates were firstly shaken for 6 seconds at 900rpm on the microplate shaker used in the lethality assay, to encourage motion before imaging. The screening plate was then moved by the automatic plate handler onto a transparent plate holder. There it was imaged for 3 min by the single camera, with a cold LED panel (B&H Photo Video, New York, NY) equipped with a red filter (Roscolux, Stamford, CT) placed under the transparent plate holder to provide illumination for tracking.

Image analysis was performed using a custom MATLAB script, based on center of mass (COM) tracking. To accurately determine the COM of each planarian, the tracking analysis was specifically optimized for fissioned worms (see Supplementary Information, Section 4). This assay provided 2 readouts: the fraction of time spent resting and the instantaneous speed of locomotion. The instantaneous speeds were calculated for all tracks over 2-second intervals to increase the signal-to-noise ratio (Hagstrom et al., 2015). An empirically determined absolute speed cutoff was used to distinguish the planarians' moving and resting behaviors (see Supplementary Information, Section 5). Instantaneous speeds less than 0.5 mm/s were considered to represent resting and were disregarded in speed calculations. The fraction of time spent resting was calculated as the amount of time resting divided by the total time tracked. Speed values > 0.5 mm/s represent planarian locomotion and were averaged to calculate the mean speed for each planarian. Of note, this speed includes instances of both swimming and gliding behaviors and thus differs from our previously used measure ((Hagstrom et al., 2015), Supplementary Information, Section 5). Planarians with no tracking data (i.e. tracking was lost for worms moving at the edge of the well due to low contrast) were considered non-analyzable and excluded for further analysis. In <4% of day 7 plates and <12% of day 12 plates (full animal and regenerating tails), 1 or 2 animals were non-analyzable. In ~1% of the day 12 plates, 3-5 animals were excluded. For fissioned worms, when the head and tail pieces were distinguishable, analysis was only performed on the head piece. Otherwise, when the head and tail pieces were indistinguishable, analysis was only performed on the fastest piece, as heads generally move faster.

Phototaxis: For this assay, the same transparent plate holder was used as for the unstimulated behavioral assay. Planarians are negative phototactic to blue light and insensitive to

red light (Paskin et al., 2014). To study negative phototactic behavior, blue LED lights (SuperNight, Portland, OR) surrounding the screening plate were used to provide the blue light stimulus. Additionally, red backlighting underneath the plate holder provided light for tracking throughout the assay. Similar to photomotor response studies in zebrafish larvae (Kokel and Peterson, 2011; Truong et al., 2014), we used a combination of dark-light-dark-light cycles. First, the plate was imaged for 30 seconds using red light (dark condition) and then imaged for 30 seconds with both red and blue lights (light condition) (Figure 4A). This sequence was then repeated. The red filter on the single camera blocks the blue light, which optimizes the imaging of this assay. Because it was only found after screening was complete that the second dark cycle was too short for animals to adapt, we compared the planarians' behavior in the first dark cycle with that in cycles 2-4 (1st light cycle, 2nd dark cycle and 2nd light cycle) instead of analyzing each dark/light cycle sequence separately.

Image analysis was automated using a custom MATLAB script. The instantaneous speeds were calculated as in the unstimulated assay. The instantaneous speed was averaged in cycle 1 and cycles 2-4. Any average speed value < 0.01 mm/s (background noise level) was set to 0.01 mm/s. Speed cutoffs were set as the mean speed of the control populations in DMSO measured in the unstimulated behavioral assay, for Day 7/Day 12 full worms and regenerating tails. In the test concentrations, planarians with a mean speed in cycle 1 lower than the speed cutoff were excluded due to their relatively high background activity, which would cause false positives in the phototaxis assay. Otherwise, the mean speed in cycles 2-4 was normalized by the mean speed in cycle 1 (background activity). Planarians with a normalized mean speed in cycles 2-4 higher than 1 were defined as having reacted to the light stimulus, and marked as "1". If the normalized mean speed in cycles 2-4 did not exceed 1, the planarian was considered to have no

reaction, and marked as "0". If the planarian was dead or had high background activity, it was discarded and marked as "NaN". The phototaxis response rate was calculated as

 $phototaxis\ response\ rate = \frac{total\ number\ of\ worms\ reacting\ to\ light}{total\ number\ of\ analyzable\ worms}$

Thermotaxis assay: The plate was placed on a custom setup with 12 peltiers (15mm x 15mm) (Digi-key, Thief River Falls, MN) that are evenly spaced and embedded in an aluminum heat sink. The peltiers are arranged in a matrix of 3 rows x 4 columns (i.e. 4 wells share one peltier) and powered by an AC to DC power supply (Genssi, Las Vegas, NV) (Figure 4B). This setup, which is controlled automatically through an Arduino board, creates an identical heat gradient with a temperature difference of 3-4°C in each well of the screening plate. During the assay, the plate was imaged without the heat gradient (ambient temperature) for 2 minutes, and then imaged with the heat gradient for 4 minutes by the single camera. The plate was illuminated from the top by a custom-made red LED ceiling light which does not obscure the view of the camera.

Image analysis was performed using a custom, automated MATLAB script. The COM of each planarian was tracked over time and used to calculate the fraction of time the animal spent in the cold area in the well when the gradient is on. Since it takes time to establish a stable heat gradient across the well, we only accounted for the fraction of time the worm spent in the cold area during the last two minutes of the assay. The cold area in each well was defined as the area of a sector with central angle of 120° (Figure B-D). Since the image analysis worked poorly on fissioned planarians, since it expects one object per well, we manually calculated the fraction of time the head piece spent in the cold area.

Scrunching assay: Scrunching is a musculature-driven escape gait in planarians, which can be triggered by multiple external stimuli, including amputation, high heat, electric shock and low pH. It is characterized by asymmetric elongation-contraction cycles (with elongation time > contraction time), and a species-specific frequency and amplitude (Cochet-Escartin et al., 2015). To induce scrunching in the screening platform, the screening plate was placed on a peltier plate (TE Technology Inc., Traverse City, MI), which was controlled by the computer through a temperature controller board (TE Technology Inc.), to increase the aquatic temperature in the wells. The temperature of the peltier plate was initially set to 65°C for the first 30 seconds to quickly heat up the plate from room temperature. Then, the temperature was gradually decreased to 43°C to stabilize the aquatic temperature across the plate at around 32°C for 4 minutes (Supplementary Figure S3), which was sufficient to induce wild-type *D. japonica* to scrunch. The plate was imaged by the single camera and illuminated by the same type of custom red LED light used in thermotaxis (see above).

Image analysis was performed using a custom, automated MATLAB script. The COM and length of each planarian were tracked over time. The worm's length over time was plotted and smoothed to detect instances of scrunching. We extracted body length oscillations in the smoothed plot which fulfilled the scrunching criteria mentioned above (asymmetric cycles, characteristic frequency) to determine instances of scrunching (Figure 4C). We defined such oscillations consisting of >3 consecutive peaks in the body length versus time plot as scrunching and marked the planarian as "1". If no such characteristic oscillations were found, the worm was marked as "0" for no scrunching. If the planarian was dead or not properly detected (not enough tracking data), it was discarded and marked as "NaN". The automated image analysis was not possible with fissioned planarians and thus these animals were scored manually. Scrunching rate was calculated as

$$scrunching rate = rac{total number of scrunching worms}{total number of analyzable worms}$$

Statistical testing: All data from the triplicate runs were compiled before performing any statistical test. For lethality, eye regeneration, phototaxis and scrunching endpoints, significant effects were determined using a one-tailed Fisher's exact test to compare the rates determined for each chemical concentration with the rate of its own DMSO controls. For thermotaxis and unstimulated behavioral endpoints, Tukey's interquartile test was first used to remove any outliers, with at most 5% (e.g. 1 out of 24 worms) of the data removed. Since the distribution of the thermotaxis data was highly skewed and variable, a non-parametric one-tailed Mann Whitney U-test was used to compare the distributions of the fraction of time in the cold area for each chemical concentration with the respective distribution of its own control. For speed and fraction of time resting from the unstimulated behavior assay, Lilliefors test was first used to test the normality of the samples. Depending on whether the sample distributions were normal or not, we performed either a parametric two-tailed t-test or a nonparametric two-tailed Mann-Whitney U-test, respectively. For all endpoints, any condition with a p-value less than 0.05 was considered statistically different from the controls. However, we observed that due to low variance in some individual plate control populations (and high variability across plates), some statistically relevant hits were likely not biologically meaningful (see Supplementary Information Section 2 and Supplementary Figure S6). Examples such as this resulted in a large number of dose-independent hits and hits in the negative controls, together suggesting these may be false

positives. Thus, to reduce potential false positives, we disregarded hits that had a smaller effect than determined by a "biological relevance" cutoff based on the variability of the DMSO controls in each assay. These cutoffs were meant to disregard hits that fell within the variability of the DMSO controls across all plates and were thus based on the distribution of the compiled control values for each chemical (n=87) and endpoint (Supplementary Figure S4). High variability within animal behavior endpoints has also been observed in zebrafish (Zhang et al., 2017). For endpoints where the distribution of the compiled control values was normal (unstimulated behavior and phototaxis), cutoffs were based on mean +/-2 or 3 SD (see Supplementary Information), respectively. For endpoints where the distribution of the compiled control values was not normal (day 12 lethality, thermotaxis, and scrunching), cutoffs were set as the 5th and 95th quantiles. These cutoffs were empirically determined to encompass the variability of the DMSO controls and to minimize dose-independent hits (see Supplementary Information Section 2 for more details). Similar approaches to creating assay-specific noise threshold levels has been described previously (Behl et al., 2015). Of note, the distributions of control values in the day7 lethality and eye regeneration endpoints were so narrow (Supplementary Figure S4) that biological relevancy cutoffs were not appropriate. However, because controls exhibited few deaths at day 7, some chemical concentrations were designated as statistically significant hits for day 7 lethality but not day 12. These cases were excluded as artifacts. Moreover, we checked for inconsistency in the data to find instances where a single plate was responsible for designating a "hit". Inconsistent hits were defined as instances with only 1 replicate outside of the biological relevancy cutoff range and two replicates within the control variability. These hits were therefore excluded (see Supplementary Figure S5 for the statistical workflow). Other groups have reportedly dealt with similar issues with plate-to-plate variability by rerunning inconsistent plates

(Zhang et al., 2017), whereas we have decided to keep all data. The lowest observed effect level (LOEL) was determined as the lowest concentration which showed a significant effect (i.e. statistically significant and passed inconsistency and biological relevancy tests, Supplementary Figure S5) in any endpoint. All statistical analyses were performed in MATLAB (see Table 1 for a summary).

To determine the observed power of each of the tested endpoints, we performed post-hoc power analysis using G*power (Faul et al., 2007) (Table 1). For some endpoints our distributions were highly skewed and/or multi-modal (unstimulated behavior and thermotaxis assays) and we were unable to transform them into normal distributions. Thus, in these cases power analysis could not be performed, since G-power expects a normal distribution as input.

Results

To evaluate the strengths and weaknesses of the planarian system for toxicology MTS, we screened an 87-compound library, provided by the NTP, consisting of known and suspected developmental neurotoxicants and five negative controls (Supplementary Table 1). Each chemical was tested at 5 concentrations, generally ranging from 10nM to 100µM, in both full (intact) planarians and regenerating tail pieces (n=8 each) (Figure 1), with a 0.5% DMSO solvent control population (n=8) in each plate. Six chemicals (BDE-153, Chrysene and Dibenz(a,h)anthracene, Bis(tributyltin) oxide, Benzo[g,h,i]perylene, and 2,3,7,8-Tetrachlorodibenzo-p-dioxin) were provided at lower than 20mM due to low solubility in DMSO and were thus tested at lower concentrations (see Supplementary Table 1 for concentrations). On day 7, when regenerating animals start to develop their photosensing system and regain motility (Hagstrom et al., 2015; Inoue et al., 2004), adult and regenerating planarians were assessed for viability, regeneration, locomotion and phototactic behavior. On day 12, all of these endpoints, except for regeneration, were tested again. In addition, on day 12, we evaluated the effects on two more stimulated behaviors: thermotaxis and scrunching. Screening on both days 7 and 12 allows us to evaluate the temporal dynamics of possible subchronic toxic effects and effects on regeneration (Figure 1). Raw data are available from the Dryad Digital Repository (doi: 10.5061/dryad.mk6m608).

Lethality and morphology

To evaluate whether the chemicals have an effect on planarian viability (Figure 2A-B), we performed statistical tests for all chemicals and, when appropriate, calculated the LC₅₀ for chemicals with significant lethality (Supplementary Figure SI and Supplementary Table 2). Over the entire 12 days of screening, 29 of the 87 tested chemicals (33%) were significantly lethal for at least one concentration, with 27 of them already being lethal by Day 7. No significant lethality was found in any of the negative controls at the tested concentrations. While lethality was found in at least one chemical from each chemical class tested, the majority of lethal compounds (18 of 29, 62%) consisted of either flame retardants or pesticides (9 lethal chemicals each). As there are only 15 or 16 chemicals comprising each of these classes in the library, respectively, this also means that the majority of the chemicals in these classes (56-60%) were lethal to planarians. Full worms tended to be more sensitive to the lethal effects of some chemicals, as 6 chemicals caused significant day 12 lethality at lower concentrations in full worms than in regenerating tails. This difference was the most striking with the flame retardant 3,3',5,5'-

Tetrabromobisphenol A as significant lethality was observed in full planarians at 1µM but in

regenerating tails at 100µM. We attribute this difference in sensitivity of full and tail worms, which was also observed in a previous screen (Hagstrom et al., 2015), partially to the generally lower motility and potentially lower level of metabolism in regenerating tail pieces. In contrast, only two chemicals, the drug Diazepam and the industrial chemical Auramine O had lower day 12 lethality LOELs in regenerating tails than in full animals.

Eye regeneration was categorized as normal (2 eyes), abnormal (0 or 1 eye) or invalid (could not be analyzed) (Figure 2 D-G). 21 chemicals (~24%) showed significant defects in eye regeneration. In the majority of these chemicals (12 of 21), regeneration defects may have been a consequence of overt systemic toxicity as effects occurred at day 12 significantly lethal concentrations (Figure 5). However, 9 of these 21 chemicals showed selective effects with the eye regeneration LOEL being less than that of the day 12 tail lethality LOEL. These selective chemicals consisted of 3 pesticides, 3 flame retardants, 1 industrial chemical, 1 PAH, and 1 negative control (Acetylsalicylic acid, Figure 2H-P).

Unstimulated behavior

We evaluated whether the chemicals perturbed planarian unstimulated behavior by quantifying the worms' fraction of time resting and mean speed during the assay (Figure 3). Together, these endpoints demonstrate whether the exposed planarians were moving and if so, whether they were moving normally. Control animals, regenerating tails and full worms, were found to move at a mean speed of approximately 1mm/s, and rest little of the time, in agreement with previous studies on planarian locomotion (Hagstrom et al., 2015). For simplicity and

because these endpoints complemented each other (Supplementary Figure S7), a chemical was classified as a hit if there was a defect in either speed or fraction of time resting.

Considering both endpoints together, 43 chemicals (49%) caused decreased locomotion in at least one worm type (full worms or regenerating tails) and time point. The majority of these chemicals (31 of 43) caused behavioral effects at nonlethal concentrations (Figure 7 and Supplementary Table 3). Overall, pesticides comprised the most hits on unstimulated behavior (11 chemicals each for day 7 full and regenerating planarians, and 8 chemicals each for day 12 full and regenerating planarians) (Figure 3E-H). In fact, considering the entire library, planarian unstimulated behavior was the most sensitive to the effects of the pesticide rotenone with defects as low as 101nM in full worms at day 7 and in regenerating tails at days 7 and 12. Interestingly, rotenone-exposed day 12 full worms did not display defects in unstimulated behavior, suggesting potential transient toxicity or adaptation over time. Loss or gain of hits between day 7 and day 12 were found with several other chemicals (Figure 4A). Moreover, although the majority of chemicals affected both full worms and regenerating tails, some effects were worm type-specific (Figure 4B). Together, these demonstrate the power of assaying toxicity at multiple endpoints and developmental stages to discern the temporal dynamics of toxicity.

In addition to hits which caused decreased activity (due to decreased speed and/or increased time resting), in 8 instances we observed one or two chemical concentrations with induced hyperactivity (due to increased speed and/or decreased time resting compared to controls) (Supplementary Table 4). In fact, the pesticide heptachlor caused hyper-activity in lower concentrations but hypo-activity in higher concentrations in day 12 regenerating tails (Figure 3C).

Stimulated behaviors: phototaxis, thermotaxis and scrunching

Planarians are known to be sensitive to a variety of environmental stimuli, including light and low and high temperatures (Birkholz and Beane, 2017; Cochet-Escartin et al., 2015; Inoue et al., 2004; Inoue et al., 2014; Lambrus et al., 2015; Paskin et al., 2014). For some of these stimuli, it has been shown that different neuronal subpopulations are involved in the animal's characteristic responses to the stimuli (Currie and Pearson, 2013; Inoue et al., 2014; Nishimura et al., 2010). We, therefore, assayed three different stimulated behaviors (phototaxis, thermotaxis and scrunching; Figure 5) to potentially differentiate between specific and general neurotoxicity.

First, we tested the planarians response to light (phototaxis). Planarians demonstrate negative phototaxis to blue light while being insensitive to red light (Paskin et al., 2014). Inspired by zebrafish photomotor response assays (Kokel and Peterson, 2011; Truong et al., 2014), we exposed planarians to bright light and compared behavior before (background activity) and after the light stimulus (Figure 5A). We then scored the number of planarians which demonstrated phototaxis. We found 15 chemicals induced phototaxis defects in at least one worm type (full or regenerating planarian) and one time point (day 7 or 12), making this the least sensitive of the tested endpoints. However, the majority of these chemicals (9) caused effects at nonlethal concentrations (Supplementary Table 5). The most hits were found in day 7 regenerating tails. Day 7 regenerating hits were found to largely overlap with hits in eye regeneration and unstimulated behavior (Figure 6A), suggesting these animals have significant regeneration delays. This is exemplified by the chemical Bis(tributyltin)oxide, which showed the most potent effects on planarian phototaxis, with a LOEL of 0.5µM in both worm types and time

points. At this concentration, regenerating tails also had defects in eye regeneration, unstimulated behavior (day 7 and 12) and scrunching, in the absence of lethality, suggesting a strong defect in regeneration. Similar defects were also found in full animals, but in the presence of lethality. The majority of hits at either day were not shared between full animals and regenerating tails (Supplementary Figure S8B).

We also evaluated how the chemicals affected the planarians' ability to react to a temperature gradient (thermotaxis, Figure 5B). The gradient was established using a custom peltier setup to induce individual temperature gradients in each well, thus incorporating our previous manual screening setup (Hagstrom et al., 2015) into the automated screening of 48-well plates. 16 (\sim 18%) of the tested chemicals demonstrated defects in thermotaxis. These active chemicals were mostly evenly distributed among the chemical classes, consisting of 5 industrial chemicals, 4 drugs, 3 flame retardants, 3 pesticides and 1 PAH. In addition, we observed that adults and regenerating animals were often affected differently, with some chemicals only affecting one worm type and not the other, with regenerating tails generally showing greater sensitivity (Supplementary Figure S8C). Moreover, the majority of these effects (10 of the 16 chemicals, ~63%) showed specific neurotoxic effects at nonlethal concentrations (Figure 7 and Supplementary Table 6) suggesting that this is a sensitive endpoint to discern sublethal neurotoxicity, particularly in developing animals. Planarian thermotaxis was most sensitive to the drug Tetraethylthiuram disulfide and the pesticide Aldicarb with LOELs of $\sim 10 \mu M$ for regenerating tails and full worms, respectively. However, at the same concentration, Aldicarb also caused hypoactivity in the unstimulated behavior assay, suggesting the thermotaxis defect may be a consequence of decreased locomotion. Tetraethylthiuram disulfide, on the other hand,

caused thermotaxis defects in the absence of locomotion defects, suggesting defects in thermoreception.

Lastly, we evaluated the planarians' ability to react to noxious stimuli. Scrunching is a musculature-driven escape gait in planarians, characterized by asymmetric elongationcontraction cycles (Cochet-Escartin et al., 2015) (Figure 5C). This gait can be induced by a variety of noxious stimuli, such as heat, amputation and pH. In our screening platform, scrunching is induced by heating the aquatic temperature of the wells by placing the screening plate on a peltier plate. 38 (~44%) of the tested chemicals caused planarians to be unable to scrunch properly. Similar to lethality, active chemicals in this endpoint were dominated by pesticides (12 chemicals) and flame retardants (10 chemicals). Interestingly, we observed this endpoint to often be affected differentially in the full and regenerating animals, with a slight bias towards regenerating tail pieces, as 14 (37%) chemicals showed increased sensitivity in the regenerating tails and 9 (24%) showed increased sensitivity in the full worms, with 15 toxicants (39%) affecting both worm types at the same concentrations (Supplementary Figure S8D). Among the 38 chemicals that caused scrunching defects, 29 (~76%) showed a scrunching defect with a scrunching LOEL lower than the respective lethality LOEL, for at least one worm type (Figure 7 and Supplementary Table 7), suggesting that scrunching is a sensitive endpoint for sublethal neurotoxicity. For example, the most sensitive scrunching defect was seen with the industrial chemical 1-ethyl-3-methylimidazolium diethylphosphate with a LOEL of 101 nM for regenerating tails. This chemical was not found to be lethal to planarians up to the maximum concentration tested (101 μ M).

Because the tested endpoints are not necessarily independent from each other, we evaluated the extent of agreement between endpoints that may be correlated. For example,

phototaxis and thermotaxis responses rely on animal locomotion to respond to the respective stimuli. Moreover, defects in eye regeneration could be expected to be correlated with defects in phototaxis. We don't, however, expect all hits to be concordant, since the blue light, which was used in the phototaxis assay, can be sensed by photoreceptors in the eyes and pigment in the body epithelium (Birkholz and Beane, 2017). While the majority of phototaxis hits in the regenerating tails were also hits in eye regeneration and/or unstimulated behavior (Figure 6A), 1 hit was found in phototaxis alone, suggesting that this assay does add additional sensitivity beyond the other endpoints. Similarly, in full worms, 2 hits were found which were not hits in the unstimulated behavior assay (Supplementary Figure S8A). Moreover, in both thermotaxis and scrunching (Figure 6B-C), a large proportion of hits were found to overlap with unstimulated behavior hits, though endpoint-specific effects were found in all cases. Together, these comparisons demonstrate the value of the large repertoire of planarian behaviors to be able to discern subtler neurotoxic effects from general systemic toxicity or gross motor defects.

Sensitivity of endpoints and global response

Through the discussion of the individual assays, we have shown that the different endpoints possess different sensitivities to different toxicities of the tested chemical compounds. Figure 7 provides a visual summary of these findings in the case of the regenerating tails (see Supplementary Figure S9 for full worms), allowing for direct comparison of the endpoint sensitivities and selectivity. Furthermore, we applied Ward's method of clustering to summarize the hits of all active compounds (49) for regenerating tails (Figure 8) and full worms (47 chemicals) across all endpoints (Figure 9), similar to (Truong et al., 2014). Endpoints were clustered into 3 major groups: lethality/morphology endpoints, unstimulated behavior/scrunching and phototaxis/thermotaxis, suggesting endpoints in the same cluster might be functionally related. Some of these clusters seem to represent particular toxic signatures for the different chemical classes (Table 4). For example, the majority of pesticides were active in the lethality, unstimulated behavior and scrunching assays. Interestingly, while full worms exposed to pesticides showed more hits (higher class concordance) in lethality, the regenerating tails had more hits in scrunching, suggesting differential effects on the adult and developing nervous system. There was also concordance of endpoints in full worms exposed to flame retardants, with most of the flame retardants being hits in lethality and scrunching. These were also the most concordant endpoints for the regenerating tails exposed to flame retardants, but with slightly less concordance. No obvious signatures were found for any of the other chemical classes, which also generally showed less activity across all planarian endpoints.

When comparing active versus inactive compounds, we found that 41 of the active chemicals are shared hits between full planarians and regenerating tails. When comparing potency, we found 13 chemicals were developmentally selective with lower overall LOELs in regenerating tails than that in full worms (Table 2). Our ability to directly compare the effect of chemicals on the brain of adult (full/intact) and developing (regenerating) animals is a unique strength of the planarian system.

Discussion

Robustness of screen and best practices

Robustness and reliability of screening are major concerns in the evaluation and verification of toxicology models (Judson et al., 2013). One aspect is reproducibility of results

between independent experimental runs (technical replicates). Therefore, in our screen, we assayed each chemical concentration in 3 independent runs and provide the data for direct comparison of the replicates in Supplementary File 1. The majority of hits were reproducible with significant activity in all 3 runs, with on average 73% shared for all runs for all endpoints with full and regenerating planarians (Supplementary Table 8). However, variability among runs was evident in some cases potentially due to technical artifacts and variability among animal populations, as described below.

First, technical issues in the scrunching assay contributed to the observed spread in the data for this endpoint. Specifically, in 3.8% of the screened plates (N=522 plates), the contact between the plate and the peltier used for administering the noxious heat stimulus was inadequate, causing variability in the scrunching response. However, the same dose-dependent trends seen in the replicates with properly functioning peltier contact was still evident in these malfunctioning replicates (Supplementary Figure S11).

Next, to account for possible effects of well position within a single plate, we rotated the position of the different chemical concentrations among runs by shifting each concentration down 2 rows with each replicate. This revealed the existence of an "edge effect", whereby planarians located at the outermost rows of the plate displayed a relatively higher lethality rate when compared to the planarians located in the plate interior at the same concentration (Supplementary Figure S10). We thus conclude, as others have previously (Truong et al., 2014), that alteration of well position for a given chemical concentration between replicates is an important aspect of ensuring reliability of results and thus enhancing screen robustness.

Finally, the planarian's diet turned out to be a significant source of biological variability affecting planarian fitness and behavior. Varying quality of food batches caused a measurable influence on the animals' sensitivity to chemical exposure (see Supplementary Information Section 3 for details) and calls for standardization of food quality to eliminate this source of variability within and between experiments and labs.

To minimize the effects of inter-run variability arising from any of these factors, we excluded hits that were determined through a single run and did not have consistent effects across the triplicates (see Material and Methods and Supplementary Figure S5).

Negative controls

The NTP 87-compound library contained 5 compounds indicated as negative controls (acetaminophen, acetylsalicylic acid, D-glucitol, L-ascorbic acid and saccharin sodium salt hydrate). All negative controls were inactive in full planarians. In contrast, in regenerating tails, while 3 of the 5 negative controls (D-glucitol, L-ascorbic acid and saccharin) showed no effects, at least one endpoint was affected by acetaminophen and acetylsalicylic acid. Acetaminophen caused decreased unstimulated speed in day 12 regenerating tails at the highest concentration tested (103µM). However, this hit was right at the biological relevance cutoff (see Materials and Methods), so it is possible that it is a false positive or potentially mild effect.

Acetylsalicylic acid caused defects in eye regeneration, unstimulated behavior (day 7 and 12) and scrunching in regenerating tails (but not full worms) at the highest tested concentration (99.5μM) suggesting developmental defects. While these chemicals were selected by the NTP to be inactive controls at the tested concentrations, toxicity has been observed with these

compounds previously. Data collected by the NTP from different public databases shows that acetaminophen and acetylsalicylic acid have been reported to have "other" and developmental/other toxicity, respectively (https://sandbox.ntp.niehs.nih.gov/neurotox/). Moreover, these 2 compounds have been associated with toxicity in multi-generation and developmental mammalian guideline studies, respectively, reported on ToxRefDB (Hagstrom et al.) For example, oral exposure of 1% (1.43 mg/kg body weight) acetaminophen to Swiss CD-1 mice for 14 weeks caused multi-generational effects on reproduction and growth (Reel et al., 1992), while single dose oral exposure to 500 mg/kg acetylsalicylic acid caused teratogenesis in rats (DePass and Weaver, 1982). Thus, these findings point toward a potential toxic effect of these compounds on developmental processes in various animal systems.

Comparison of hits with existing planarian toxicology data

For some of the chemicals tested in this screen, previous largely manual toxicology studies on planarians exist. We therefore compared our results with the published literature to evaluate concordance (Table 3). Of note, while we studied chronic exposure in both full and regenerating planarians, most of the previous studies evaluated either only regeneration and/or acute exposure. Direct comparisons between different experiments are difficult to make because of differences in experimental methods (chemical concentrations tested, exposure conditions and duration, worm type (full/regenerating), data and statistical analysis, number of replicates, etc.), and differences in planarian species used, which may have differing sensitivity. Together, this experimental heterogeneity emphasizes the need for uniform testing guidelines going forward.

The zebrafish community faces similar challenges, for example see (Truong et al., 2014), with different labs using different experimental methodologies.

Strengths and current limitations of the planarian as a model for developmental neurotoxicity

The performance of this 87-compound screen revealed both the strengths and weaknesses of the planarian screening platform, as summarized in Table 5. As with any toxicology system, the planarian system has its limitations. However, when appropriately utilized, this system can add value to the existing testing pipeline through its unique strengths, such as the ability to screen adults and developing animals in parallel with the same assays to delineate developmental-specific effects and differentiate between DNT and general neurotoxicity (Table 2). For example, of the 38 known developmental neurotoxicants in this library (Supplementary Table 1, (Ryan et al., 2016)), 10 (1 drug, 5 industrial, and 4 pesticides) had greater effects in regenerating planarians, with lower overall LOELs than full planarians.

Another strength of the planarian system is the large repertoire of quantitative behavioral readouts that allow coverage of a wide spectrum of neuronal functions that are currently not assayed in other medium-throughput animal systems, such as zebrafish larvae. Moreover, the molecular mediators of some of these behaviors have been characterized (Birkholz and Beane, 2017; Inoue et al., 2014; Nishimura et al., 2010), allowing for insight into mechanisms of neurotoxicity. For example, 10 μ M Tetraethylthiuram disulfide was found to selectively disrupt thermotaxis in regenerating tails, but not full planarians, in the absence of other affected endpoints. Planarian thermotaxis has been shown to be mediated by Transient Receptor Potential (TRP) channels (Inoue et al., 2014), and Tetraethylthiuram disulfide has been found to be a

selective agonist for human TRPA1 *in vitro* (Maher et al., 2008). Additionally, regenerating planarians were found to be highly sensitive to rotenone, a pesticide and mitochondrial disruptor. We observed significant defects in full and regenerating tails unstimulated behavior and eye regeneration at concentrations as low as 101nM. In rodent models, the effects of rotenone on retinal neurodegeneration and locomotor activity have been well documented (Alam et al., 2004; Normando et al., 2016; Rojas et al., 2008). The similarity of these affected endpoints in both models suggests that similar molecular pathways are targeted in the same way. Together, these demonstrate the utility of the range of planarian morphological and behavioral endpoints to connect adverse functional outcomes with mechanisms, which are likely conserved in higher organisms, including mammals and humans.

In the NTP 87-compound library, 38 chemicals were denoted as known developmental neurotoxicants (Supplementary Table 1) from previous *in vivo* and *in vitro* studies (Ryan et al., 2016) and 23 (~61%) were active in planarian regenerating tails. Concordance varied by class from most to least: pesticide (13/14), industrial (4/10), and drug (6/14). No PAHs or flame retardants were listed as known developmental neurotoxicants. Moreover, in our companion paper (Hagstrom et al.), we found that of the 28 chemicals in this library with associated quality mammalian guideline studies available on the U.S. EPA Toxicity Reference Database, 20 (71%) were active in regenerating planarians. Some of these false negatives may be due to absence of the relevant biological targets in planarians. For example, the inactivity of thalidomide, an infamous teratogen with suggested effects on angiogenesis (Stephens et al., 2000), in planarians may not be surprising given their lack of a circulatory system.

Other factors need to be taken into account when evaluating concordance, such as the extent of uptake and bioavailability in the animals. The reported concentrations in this study are

nominal water concentrations and the internal concentrations within the planarians are unknown. Thus, it is uncertain whether inactivity is due to loss of chemical to the plastic, lack of absorption into the planarian, insufficient metabolic machinery, or other pharmacokinetic (PK) differences. For example, since chemical uptake in planarians occurs through the skin or pharynx (Balestrini et al., 2014; Kapu and Schaeffer, 1991) and planarians possess a protective mucus coating (Martin, 1978; Pedersen, 2008), certain chemical classes may be unable to effectively penetrate into the animal. Future research will have to determine the PK and pharmacodynamics (PD) of this system, and identify which compounds are bioavailable, to be able to connect activity with the relevant exposure in mammals and humans. While this study focused on the planarian system, a companion study in this special issue (Hagstrom et al.) performs a direct comparison using this NTP 87-compound library between the planarian and zebrafish systems, and available mammalian data. Together, both studies demonstrate the added value of comparative screening in multiple complementary models to assay a larger swath of chemical and biological space.

Supplementary data description

Compiled data for each endpoint and comparisons between individual replicates can be found in Supplementary File 1. Additional data can be found in the Supplementary Tables and Supplementary Figures files. Extended methods and discussion can be found in Supplementary Information.

Funding information

This work was funded by the Hellman Foundation, the Burroughs Wellcome CASI award, and the Alfred P. Sloan Foundation. Danielle Hagstrom was partially supported by the National Institutes of Health Cell and Molecular Genetics Training Grant (5T32GM007240-37) and Marye Anne Fox Endowed Fellowship.

Acknowledgements

We thank Alex Fields for help setting up the rail system used in the screening platform, Jessica Soong for help manufacturing the thermotaxis peltier holder and accessories of Fresnel lens design, Noopur Khachane for help setting up the power control system and the temperature measurement kit, Yingtian He for help with animal care and data analysis and Jared Estrada for help with animal care. We also thank Robert Tanguay and Lisa Truong for discussion.

References

- Alam, M., Mayerhofer, A. and Schmidt, W. J. (2004). The neurobehavioral changes induced by bilateral rotenone lesion in medial forebrain bundle of rats are reversed by L-DOPA. *Behav. Brain Res.* 151, 117–124.
- Balestrini, L., Isolani, M. E., Pietra, D., Borghini, A., Bianucci, A. M., Deri, P. and Batistoni, R.
 (2014). Berberine exposure triggers developmental effects on planarian regeneration. *Sci. Rep.* 4, 4914.
- Behl, M., Hsieh, J.-H., Shafer, T. J., Mundy, W. R., Rice, J. R., Boyd, W. A., Freedman, J. H., Hunter, E. S., Jarema, K. A., Padilla, S., et al. (2015). Use of alternative assays to identify and prioritize organophosphorus flame retardants for potential developmental and neurotoxicity. *Neurotoxicol. Teratol.* 52, 181–193.
- Behl, M., Ryan, K., Hsieh, J.-H., Parham, F., Shapiro, A., Collins, B. J., Birnbaum, L. S.,Bucher, J. R., Walker, N. J., Foster, P. M., et al. (2018). Screening for developmental

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

- Birkholz, T. R. and Beane, W. S. (2017). The planarian TRPA1 homolog mediates extraocular behavioral responses to near-ultraviolet light. *J. Exp. Biol.* **220**, 2616–2625.
- Boyd, W., Smith, M. and Freedman, J. (2012). Caenorhabditis elegans as a model in developmental toxicology. In *Developmental Toxicology* (ed. Harris, C.) and Hansen, J. M.), pp. 15–24. Humana Press.
- Boyd, W. A., Smith, M. V., Co, C. A., Pirone, J. R., Rice, J. R., Shockley, K. R. and Freedman, J. H. (2015). Developmental effects of the ToxCastTM phase I and phase II chemicals in Caenorhabditis elegans and corresponding responses in zebrafish, rats, and rabbits. *Environ. Health Perspect.* 124, 586–93.
- Bronsted, H. V. (1955). Planarian regeneration. Biol. Rev. 30, 65-126.
- Buchanan, J. W. (1935). An analysis of physiological states responsible for antero-posterior disintegration in Planaria dorotocephala. *Protoplasma* 22, 497–512.
- Buttarelli, F. R., Pellicano, C. and Pontieri, F. E. (2008). Neuropharmacology and behavior in planarians: Translations to mammals. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 147, 399–408.
- Carter, J. A., Lind, C. H., Truong, M. P. and Collins, E.-M. S. (2015). To each his own. *J. Stat. Phys.* **161**, 250–272.
- Cebrià, F. (2007). Regenerating the central nervous system: how easy for planarians! *Dev. Genes Evol.* **217**, 733–48.
- Cochet-Escartin, O., Mickolajczk, K. J. and Collins, E.-M. S. (2015). Scrunching: a novel escape gait in planarians. *Phys. Biol.* **12**, 55001.

- Collins, F. S., Gray, G. M. and Bucher, J. R. (2008). Transforming environmental health protection. *Science* **319**, 906–7.
- Currie, K. W. and Pearson, B. J. (2013). Transcription factors lhx1/5-1 and pitx are required for the maintenance and regeneration of serotonergic neurons in planarians. *Development* 140, 3577–88.
- DePass, L. R. and Weaver, E. V. (1982). Comparison of teratogenic effects of aspirin and hydroxyurea in the Fischer 344 and Wistar strains. *J. Toxicol. Environ. Health* **10**, 297–305.
- Dunkel, J., Talbot, J. and Schötz, E.-M. (2011). Memory and obesity affect the population dynamics of asexual freshwater planarians. *Phys. Biol.* **8**, 26003.
- Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191.
- Hagstrom, D., Truong, L., Zhang, S., Tanguay, R. L. and Collins, E.-M. S. Comparative analysis of zebrafish and planarian model systems for developmental neurotoxicity screens using an 87-compound library Title. *Toxicol. Sci.*
- 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.
- Hill, A. J., Teraoka, H., Heideman, W. and Peterson, R. E. (2005). Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicol. Sci.* 86, 6–19.
- Hori, I. and Kishida, Y. (1998). A fine structural study of regeneration after fission in the

planarian Dugesia japonica. Hydrobiologia 383, 131-136.

- Inoue, T., Kumamoto, H., Okamoto, K., Umesono, Y., Sakai, M., Alvarado, A. S. and Agata, K. (2004). Morphological and functional recovery of the planarian photosensing system during head regeneration. *Zoolog. Sci.* 21, 275–283.
- 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.
- Judson, R., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R. R., Whelan, M., Xia, M., et al. (2013). Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX* **30**, 51–56.
- Kapu, M. M. and Schaeffer, D. J. (1991). Planarians in toxicology. Responses of asexual Dugesia dorotocephala to selected metals. *Bull. Environ. Contam. Toxicol.* 47, 302–7.
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203, 253–310.
- Kokel, D. and Peterson, R. T. (2011). Using the zebrafish photomotor response for psychotropic drug screening. *Methods Cell Biol.* **105**, 517–524.
- Lambrus, B. G., Cochet-Escartin, O., Gao, J., Newmark, P. A., Collins, E.-M. S. and Collins, J. J. (2015). Tryptophan hydroxylase is required for eye melanogenesis in the planarian Schmidtea mediterranea. *PLoS One* 10, e0127074.
- Maher, M., Ao, H., Banke, T., Nasser, N., Wu, N.-T., Breitenbucher, J. G., Chaplan, S. R. and Wickenden, A. D. (2008). Activation of TRPA1 by farnesyl thiosalicylic acid. *Mol. Pharmacol.* 73, 1225–1234.

Martin, G. G. (1978). A new function of rhabdites: Mucus production for ciliary gliding.

Zoomorphologie **91**, 235–248.

- Mineta, K., Nakazawa, M., Cebria, F., Ikeo, K., Agata, K. and Gojobori, T. (2003). Origin and evolutionary process of the CNS elucidated by comparative genomics analysis of planarian ESTs. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7666–71.
- Morita, M. and Best, J. B. (1984). Effects of photoperiods and melatonin on planarian asexual reproduction. *J. Exp. Zool.* **231**, 273–282.
- Nishimura, K., Kitamura, Y., Taniguchi, T. and Agata, K. (2010). Analysis of motor function modulated by cholinergic neurons in planarian Dugesia japonica. *Neuroscience* **168**, 18–30.
- Normando, E. M., Davis, B. M., De Groef, L., Nizari, S., Turner, L. A., Ravindran, N.,
 Pahlitzsch, M., Brenton, J., Malaguarnera, G., Guo, L., et al. (2016). The retina as an early biomarker of neurodegeneration in a rotenone-induced model of Parkinson's disease:
 Evidence for a neuroprotective effect of rosiglitazone in the eye and brain. *Acta Neuropathol. Commun.* 4, 86.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359.
- Paskin, T. R., Jellies, J., Bacher, J. and Beane, W. S. (2014). Planarian phototactic assay reveals differential behavioral responses based on wavelength. *PLoS One* **9**, e114708.

Pedersen, K. J. (2008). Sliime-secreting cells of planarians. Ann. N. Y. Acad. Sci. 106, 424-443.

Reel, J. R., Lawton, A. D. and Lamb, J. C. (1992). Reproductive toxicity evaluation of acetaminophen in Swiss CD-1 mice using a continuous breeding protocol. *Fundam. Appl. Toxicol.* 18, 233–239.

Rink, J. C. (2013). Stem cell systems and regeneration in planaria. Dev. Genes Evol. 223, 67-84.

- Rojas, J. C., Saavedra, J. A. and Gonzalez-Lima, F. (2008). Neuroprotective effects of memantine in a mouse model of retinal degeneration induced by rotenone. *Brain Res.* 1215, 208–217.
- Ross, K. G., Currie, K. W., Pearson, B. J. and Zayas, R. M. (2017). Nervous system development and regeneration in freshwater planarians. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e266.
- Ryan, K. R., Sirenko, O., Parham, F., Hsieh, J. H., Cromwell, E. F., Tice, R. R. and Behl, M. (2016). Neurite outgrowth in human induced pluripotent stem cell-derived neurons as a high-throughput screen for developmental neurotoxicity or neurotoxicity. *Neurotoxicology* 53, 271–281.
- Scholz, S. (2013). Zebrafish embryos as an alternative model for screening of drug-induced organ toxicity. *Arch. Toxicol.* 87, 767–769.
- Stephens, T. D., Bunde, C. J. and Fillmore, B. J. (2000). Mechanism of action in thalidomide teratogenesis. *Biochem. Pharmacol.* 59, 1489–99.
- Tejeda-Benitez, L. and Olivero-Verbel, J. (2016). Caenorhabditis elegans, a biological model for research in toxicology. In *Reviews of Environmental Contamination and Toxicology* (ed. de Voogt, W.), pp. 1–35. Springer, Cham.
- Truong, L., Reif, D. M., St Mary, L., Geier, M. C., Truong, H. D. and Tanguay, R. L. (2014).Multidimensional in vivo hazard assessment using zebrafish. *Toxicol. Sci.* 137, 212–33.
- Tsuji, R. and Crofton, K. M. (2012). Developmental neurotoxicity guideline study: Issues with methodology, evaluation and regulation*. *Congenit. Anom. (Kyoto).* **52**, 122–128.
- Umesono, Y., Tasaki, J., Nishimura, K., Inoue, T. and Agata, K. (2011). Regeneration in an evolutionarily primitive brain--the planarian Dugesia japonica model. *Eur. J. Neurosci.* **34**,

- Yang, X., Kaj, K. J., Schwab, D. J. and Collins, E.-M. S. (2017). Coordination of size-control, reproduction and generational memory in freshwater planarians. *Phys. Biol.* 14, 36003.
- Zhang, G., Truong, L., Tanguay, R. L. and Reif, D. M. (2017). A new statistical approach to characterize chemical-elicited behavioral effects in high-throughput studies using Zebrafish. *PLoS One* **12**, e0169408.

Figure legends

Figure 1. Overview of planarian screening platform. (A) Schematic of screening workflow. On day 1, for each chemical, one plate each is filled with either full planarians (F) or regenerating tail pieces (R). 5 test concentrations and 1 control concentration (0.5% DMSO) are placed in each row with n=8 animals per concentration. Plate orientation is altered between replicates. Screening is performed on days 7 and 12. (B) The timeline shows which assays are performed on which screening days.

Figure 2. Lethality and eye regeneration endpoints. High-resolution imaging of each well was used to determine whether a planarian was (A) alive or (B) dead. (C) Distributions of lethal chemicals and their day 12 LOEL by chemical class for full worms (F, top row) and regenerating tails (R, bottom row). Chemicals which were not found to be lethal at the tested concentrations are marked as N/D for "not determined". (D-F) High-resolution imaging of day 7 regenerating tails was used to evaluate whether the eyes had regenerated. A custom neural network was used to automatically detect whether the planarian had (D) 2 eyes (normal), or abnormal eyes, (either

(E) 1 eye or (F) no eyes) as described in Materials and Methods. Insets show cropped and zoomed-in head regions. Arrows point to the eyes. (G) In some cases, it was impossible to correctly determine the number of eyes. Such cases were classified as invalid and discarded in the analysis. Black scale bars: 1mm. White scale bars: 0.2mm. (H-P) Eye regeneration rate (percentage of planarians with 2 regenerated eyes) shown for each replicate (dots) and for all combined data (bars) as a function of concentration for chemicals in which defects were seen in the absence of significant lethality. If no individual replicate data is shown, all animals were dead in this sample. Significant defects in eye regeneration are in black bars. Concentrations corresponding to the day 12 regenerating tail lethality LOELs for each chemical are in red text. No red text signifies no significant lethality was found in the range of concentrations tested. Chemicals shown are flame retardants (H) 3,3',5,5'-Tetrabromobisphenol A, (I) Firemaster 500 and (J) tris(2-Chloroisopropyl)phosphate (TCPP), pesticides (K) Bis(tributyltin)oxide, (L) Heptachlor and (M) Rotenone, (N) industrial Bisphenol A, (O) PAH Pyrene and (P) negative control Acetylsalicylic acid.

Figure 3. Unstimulated behavior: gliding and resting. (A) Representative center of mass (COM) track of one gliding planarian color-coded by time. (B) Representative color-coded COM track of a planarian which started to rest after approximately 1 minute. Scale bars: 2mm. (C-D) Example of dose-response curves of (C) mean speed and (D) mean fraction of time spent resting with standard error as error bars, for same groups of regenerating tails in Heptachlor at Day 12. Stars indicate significant differences from controls (p<0.05), showing either hyper-(black, increased locomotor activity) or hypo-activity (red, decreased locomotor activity). (E-F) Distributions of chemicals with defects in unstimulated behavior and their LOEL by chemical

class for full worms (E-F, top row) and regenerating tails (G-H, bottom row) at day 7 (left) or day 12 (right). Chemicals which were not found to have an effect on unstimulated behaviors at the tested concentrations are marked as N/D for "not determined". Chemicals with nonmonotonic dose-response curves are marked as "indeterminate".

Figure 4. Comparison of time-points and worm types for unstimulated behavior hits. (A) Considering both unstimulated behavioral endpoints together, comparison of hits that were conserved between day 7 and day 12 in either full worms (top) or regenerating tails (bottom). (B) Considering both unstimulated behavioral endpoints together, comparison of hits that were conserved between full worms and regenerating tails at either day 7 (top) or day 12 (bottom). All comparisons are performed per chemical, irrespective of concentration.

Figure 5. Stimulated behaviors. (A) Planarians exhibiting phototaxis respond to alternating light and dark cycles with increasing speed. Examples of 3 full worms in DMSO controls at day 7 were plotted. (B) Schematic of thermotaxis. 12 peltier elements (squares) were evenly distributed to create a heat gradient across each well. The cold area (blue sectors) in each well was defined as the area of a sector of 120° in the analysis. Insets show tracks, color-coded by time, of representative planarian responses to the heat gradient. Both images show the motion of 4 planarians in 4 wells over 2 minutes with either the heat gradient (i) off or (ii) on. Scale bar: 5mm. (C) Representative plot of planarian body length over a short time period (160-240 seconds) in the scrunching assay. The body length oscillations which fulfilled the scrunching criteria in the plot are in a red box. The observed low-frequency oscillations are mostly the worm's turns and head wiggling.

Figure 6. Comparison of shared hits in stimulated vs unstimulated behaviors. (A) Venn diagram of overlap of hits in day 7 eye regeneration, with day 7 (left) or day 12 (right) phototaxis and unstimulated behavior assays in regenerating tails. (B) Venn diagram of hits in thermotaxis and unstimulated behavior at day 12 for full worms (top) and regenerating tails (bottom). (C) Venn diagram of hits in scrunching and unstimulated behavior at day 12 for full worms (top) and regenerating tails (bottom).

Figure 7. Analysis of LOEL by endpoint. Regenerating tail LOELs for each endpoint, separated into 5 concentration classes, listed highest to lowest (1-5). Most chemicals were tested in the range of 0.01-100μM (see legend). However, BDE-153, Chryene and Dibenz(a,h)anthracene were tested at 0.005-50μM, Bis(tributyltin) oxide at 0.5-5000nM, Benzo[g,h,i]perylene at 0.4-4000 nM, and 2,3,7,8-Tetrachlorodibenzo-p-dioxin at 0.04 – 400 nM, due to low solubility in DMSO. Each endpoint LOEL is categorized and counted (y-axis) based on the co-occurrence of lethality at the same or higher concentrations.

Figure 8. Summary of screening results for regenerating tail. Bicluster heat map of chemicals affecting at least one endpoint in regenerating tails with LOEL color-coded. The hits were clustered using Ward's method by calculating Euclidean distance between LOELs.

Figure 9. Summary of screening results in full planarians. Bicluster heat map of chemicals affecting at least one endpoint in full planarians with LOEL color-coded. The hits were clustered using Ward's method by calculating Euclidean distance between LOELs.

Tables

Assay	Endpoints	Statistical test	Median observed power
Lethality	Lethality rate	One-tailed Fisher's exact test	1
Morphology	Eye regeneration rate	One-tailed Fisher's exact test	0.99
		Two-tailed T-test	
	Speed	or Mann Whitney U-test	N/D*
Unstimulated		Two-tailed T-test	
behavior	Fraction of time resting	or Mann Whitney U-test	N/D*
Phototaxis	Phototaxis response rate	One-tailed Fisher's exact test	0.75
Thermotaxis	Fraction of time in cold area	One-tailed Mann Whitney U-test	N/D*
Scrunching	Scrunching rate	One-tailed Fisher's exact test	0.98

Table. 1. Summary of statistical testing

* N/D: not determined

Table 2. Developmentally selective chemicals. Chemicals which had overall lower LOELs in regenerating tails than in full planarians.

Class	Chemical	Selective endpoints
Drug	Colchicine	Unstimulated day 12
Industrial	1-ethyl-3-methylimidazolium diethylphosphate	Scrunching
	2-Methoxyethanol	Thermotaxis
	3,3'-Iminodipropionitrile	Thermotaxis
	Bisphenol A	Unstimulated day 12*
	n-Hexane	Scrunching
РАН	Anthracene	Unstimulated day 7/12, Scrunching
	Phenanthrene	Unstimulated day 7* /12*
Pesticide	Chlorpyrifos (Dursban)	Scrunching
	Lindane	Scrunching
	Permethrin	Unstimulated day 7*
Negative	Acetaminophen	Unstimulated day 12
	Acetylsalicylic acid	Unstimulated day 7/12, Scrunching

* dose was non-monotonic

Table 3. Comparison of results with previous planarian studies. As necessary, concentrations were converted to μ M for ease of comparison.

	This screen		Previously published						
Chemical	LOEL	Endpoints affected	Species		Endpoints affected	Exposure	Reference		
6-Hydroxydopamine hydrochloride	0.1*	Unstimulated behavior (d12F)	Dugesia gonocephala	~120 [†]	Mobility	7 days	(Caronti et al., 1999)		
Acetaminophen	100	Unstimulated behavior (d12R)	D. japonica	1000 [§]	Lethality	4 days	(Li, 2013a)		
Acetylsalicylic acid	100	Eye regeneration, unstimulated behavior (d7R, d12R), scrunching (d12R)	D. japonica 520 [§] Lethal		Lethality	4 days	(Li, 2013a)		
Acrylamide	N/D	none	D. japonica 100		Gliding speed, brain regeneration	8,15 days	(Hagstrom et al., 2015)¶		
Benz(a)anthracene	100	Scrunching (d12F/R)	D. dorotocephala	Not available	Lethality	3 months	(Best and Morita, 1982)		
Berberine chloride	10	Unstimulated behavior (d7F/R, d12F/R)	D. japonica	50	Eye regeneration, mobility	7 days	(Balestrini et al., 2014)		
Bisphenol A	0.01*	Lethality (F/R), eye regeneration, unstimulated behavior (d7R, d12F/R), scrunching (d12R)	D. japonica	~20 [§] 2	Lethality Regeneration	4 days 7 days	(Li, 2013b) (Li, 2014)		
Carbaryl	100	Unstimulated behavior (d7F/R, day12F/R), scrunching (day12F/R)	Dugesia tigrina	1	Mobility	2h	(Feldhaus et al., 1998)		
Chlorpyrifos	10	Lethality (F/R), eye regeneration,	D. japonica; D. dorotocephala	1	Lethality, unstimulated	8,15 days	(Hagstrom et al., 2015)¶		

		unstimulated behavior (d7R), scrunching (d12R)		5	behavior, brain regeneration, thermotaxis; lethality, acute behavior, head regeneration	7 days	(Villar et al., 1993)
Colchicine	100	Unstimulated behavior (d12F)	D. dorotocephala	200	Regeneration	10 days	(McWhinnie, 1955)
Diazepam	10	Lethality (F/R), eye regeneration, unstimulated behavior (d7F/R, d12F/R), scrunching (d12F/R), thermotaxis (d12F/R), phototaxis (d7R, d12F)	D. tigrina	35	Lethality, behavior, morphology	3 hours	(Alves and de Melo, 2013)
Dichlorodiphenyltrichloroethane (DDT)	10	Lethality (F), unstimulated behavior (d7F,d12F/T), scrunching (d12F/T), thermotaxis (d12T)	Polycelis feline, Creobia alpina	1	Regeneration, mobility	14 days	(Kouyoumjian and Villeneuve, 1979)
Diethylstilbestrol	10	Lethality (F/R), unstimulated behavior (d12F)	D. japonica	2§ 2	Lethality Regeneration	4 days 7 days	(Li, 2013b) (Li, 2014)
Estradiol	100	Lethality (F/R), scrunching (d12R)	D. japonica	6§ 9	Lethality Regeneration	4 days 7 days	(Li, 2013b) (Li, 2014)
Methyl mercuric (II) chloride	10	Lethality (F/R), Unstimulatd behavior (d7F)	D. dorotocephala; Polycelis tenuis	0.3	Lethality, morphology, regeneration,	5h-10 days	(Best et al., 1981)

			and Dugesia lugubris		behavior		
Parathion	10	Lethality (F/R), eye regeneration, unstimulated behavior (d12F/T), scrunching (d12T)	Schmidtea mediterranea	50	Lethality, regeneration, mobility	1 – 4 days, 7 days, 12 days	(Poirier et al., 2017)
Permethrin	0.1*	Unstimulated behavior (d7F/R, d12R), scrunching (d12R)	D. japonica	20	Lethality, gliding speed, eye regeneration, brain regeneration	8,15 days	(Hagstrom et al., 2015)¶
Rotenone	0.1	Lethality (F/R), eye regeneration, unstimulated behavior (d7F/R, d12R)	D. japonica	0.01	Lethality, acute behavior	4 days	(Kitamura et al., 2003)
Thalidomide	N/D	none	D. tigrina	232	Eye regeneration	5 day exposure, assayed at day 7	(Best and Morita, 1982)

*dose response was non-monotonic

†only 1 tested concentration

$\$LC_{50}$

¶see Supplementary Information Section 6 for more details

Table 4. Summary of percentage of actives observed in different toxicant classes in all endpoints for either full worms (F) or regenerating tails (R). Percentages are based on the total number of chemicals in the respective class.

Endnointe	Day			Fla retai	ime rdant					Pest	icide		
Enupoints	Day	Drug	g (19)	(1	5)	Indust	rial(15)	PAH	I(17)	(1	6)	Nega	tive(5)
		F	R	F	R	F	R	F	R	F	R	F	R
Lethality	12	37%	37%	60%	53%	20%	20%	6%	6%	56%	44%	0%	0%
Eye	7	NA	16%	NA	40%	NA	20%	NA	12%	NA	38%	NA	20%
Unctimulated	7	11%	21%	40%	40%	7%	20%	12%	24%	69%	69%	0%	20%
Unstimulated	12	26%	26%	33%	27%	13%	13%	24%	24%	50%	50%	0%	0%
Dhototovia	7	0%	11%	7%	13%	0%	7%	0%	0%	25%	13%	0%	0%
Phototaxis	12	11%	0%	13%	13%	7%	0%	0%	0%	25%	13%	0%	0%
Thermotaxis	12	5%	21%	0%	20%	7%	27%	6%	0%	19%	13%	0%	0%
Scrunching	12	21%	21%	67%	47%	13%	27%	24%	29%	44%	75%	0%	20%

Table 5.	Summary	of the s	strengths a	nd weak	nesses of	the pla	narian †	toxicology	system.
									~

Strengths	Weaknesses
 Cost- and time-effective screen within 12 days compared to months in mammalian systems Invertebrate system Amenable to full automation Easy administration of compounds in the water Many different behavioral readouts, some with known cellular/molecular pathways Ability to study adult and developing animals in parallel with the same assays Allows for multi-generational studies 	 Limited morphological endpoints due to simple anatomy May be missing some relevant toxicological targets Potential water solubility issues and loss of toxicants into the environment Unknown PK/PD parameters (e.g. internal concentrations and xenobiotic metabolism); Single route of exposure (absorption) Clonal animals, no genetic diversity



Figure 1. Overview of planarian screening platform. (A) Schematic of screening workflow. On day 1, for each chemical, one plate each is filled with either full planarians (F) or regenerating tail pieces (R). 5 test concentrations and 1 control concentration (0.5% DMSO) are placed in each row with n=8 animals per concentration. Plate orientation is altered between replicates. Screening is performed on days 7 and 12. (B) The timeline shows which assays are performed on which screening days.

186x173mm (300 x 300 DPI)



Figure 2. Lethality and eye regeneration endpoints. High-resolution imaging of each well was used to determine whether a planarian was (A) alive or (B) dead. (C) Distributions of lethal chemicals and their day 12 LOEL by chemical class for full worms (F, top row) and regenerating tails (R, bottom row). Chemicals which were not found to be lethal at the tested concentrations are marked as N/D for "not determined". (D-F) High-resolution imaging of day 7 regenerating tails was used to evaluate whether the eyes had regenerated. A custom neural network was used to automatically detect whether the planarian had (D) 2 eyes (normal), or abnormal eyes, (either (E) 1 eye or (F) no eyes) as described in Materials and Methods. Insets show cropped and zoomed-in head regions. Arrows point to the eyes. (G) In some cases, it was impossible to correctly determine the number of eyes. Such cases were classified as invalid and discarded in the analysis. Black scale bars: 1mm. White scale bars: 0.2mm. (H-P) Eye regeneration rate (percentage of planarians with 2 regenerated eyes) shown for each replicate (dots) and for all combined data (bars) as a function of concentration for chemicals in which defects were seen in the absence of significant lethality. If no individual replicate data is shown, all animals were dead in this sample. Significant defects in eye

regeneration are in black bars. Concentrations corresponding to the day 12 regenerating tail lethality LOELs for each chemical are in red text. No red text signifies no significant lethality was found in the range of concentrations tested. Chemicals shown are flame retardants (H) 3,3',5,5'-Tetrabromobisphenol A, (I) Firemaster 500 and (J) tris(2-Chloroisopropyl)phosphate (TCPP), pesticides (K) Bis(tributyltin)oxide, (L) Heptachlor and (M) Rotenone, (N) industrial Bisphenol A, (O) PAH Pyrene and (P) negative control Acetylsalicylic acid.

190x241mm (300 x 300 DPI)



Figure 3. Unstimulated behavior: gliding and resting. (A) Representative center of mass (COM) track of one gliding planarian color-coded by time. (B) Representative color-coded COM track of a planarian which started to rest after approximately 1 minute. Scale bars: 2mm. (C-D) Example of dose-response curves of (C) mean speed and (D) mean fraction of time spent resting with standard error as error bars, for same groups of regenerating tails in Heptachlor at Day 12. Stars indicate significant differences from controls (p<0.05), showing either hyper-(black, increased locomotor activity) or hypo-activity (red, decreased locomotor activity). (E-F) Distributions of chemicals with defects in unstimulated behavior and their LOEL by chemical class for full worms (E-F, top row) and regenerating tails (G-H, bottom row) at day 7 (left) or day 12 (right). Chemicals which were not found to have an effect on unstimulated behaviors at the tested concentrations are marked as N/D for "not determined". Chemicals with non-monotonic dose-response curves are marked as "indeterminate".

193x165mm (300 x 300 DPI)



Figure 4. Comparison of time-points and worm types for unstimulated behavior hits. (A) Considering both unstimulated behavioral endpoints together, comparison of hits that were conserved between day 7 and day 12 in either full worms (top) or regenerating tails (bottom). (B) Considering both unstimulated behavioral endpoints together, comparison of hits that were conserved between full worms and regenerating tails at either day 7 (top) or day 12 (bottom). All comparisons are performed per chemical, irrespective of concentration.

138x75mm (300 x 300 DPI)



Figure 5. Stimulated behaviors. (A) Planarians exhibiting phototaxis respond to alternating light and dark cycles with increasing speed. Examples of 3 full worms in DMSO controls at day 7 were plotted. (B)
Schematic of thermotaxis. 12 peltier elements (squares) were evenly distributed to create a heat gradient across each well. The cold area (blue sectors) in each well was defined as the area of a sector of 120° in the analysis. Insets show tracks, color-coded by time, of representative planarian responses to the heat gradient. Both images show the motion of 4 planarians in 4 wells over 2 minutes with either the heat gradient (i) off or (ii) on. Scale bar: 5mm. (C) Representative plot of planarian body length over a short time period (160-240 seconds) in the scrunching assay. The body length oscillations which fulfilled the scrunching criteria in the plot are in a red box. The observed low-frequency oscillations are mostly the worm's turns and head wiggling.

195x120mm (300 x 300 DPI)



Figure 6. Comparison of shared hits in stimulated vs unstimulated behaviors. (A) Venn diagram of overlap of hits in day 7 eye regeneration, with day 7 (left) or day 12 (right) phototaxis and unstimulated behavior assays in regenerating tails. (B) Venn diagram of hits in thermotaxis and unstimulated behavior at day 12 for full worms (top) and regenerating tails (bottom). (C) Venn diagram of hits in scrunching and unstimulated behavior at day 12 for full worms (top) and regenerating tails (bottom).

166x161mm (300 x 300 DPI)



Figure 7. Analysis of LOEL by endpoint. Regenerating tail LOELs for each endpoint, separated into 5 concentration classes, listed highest to lowest (1-5). Most chemicals were tested in the range of 0.01-100μM (see legend). However, BDE-153, Chryene and Dibenz(a,h)anthracene were tested at 0.005-50μM, Bis(tributyltin) oxide at 0.5-5000nM, Benzo[g,h,i]perylene at 0.4-4000 nM, and 2,3,7,8-Tetrachlorodibenzo-p-dioxin at 0.04 – 400 nM, due to low solubility in DMSO. Each endpoint LOEL is categorized and counted (y-axis) based on the co-occurrence of lethality at the same or higher concentrations.

95x54mm (300 x 300 DPI)



Figure 8. Summary of screening results for regenerating tail. Bicluster heat map of chemicals affecting at least one endpoint in regenerating tails with LOEL color-coded. The hits were clustered using Ward's method by calculating Euclidean distance between LOELs.

187x209mm (300 x 300 DPI)



Figure 9. Summary of screening results in full planarians. Bicluster heat map of chemicals affecting at least one endpoint in full planarians with LOEL color-coded. The hits were clustered using Ward's method by calculating Euclidean distance between LOELs.

187x203mm (300 x 300 DPI)