PLOS ONE

RESEARCH ARTICLE

Distinguishing classes of neuroactive drugs based on computational physicochemical properties and experimental phenotypic profiling in planarians

Danielle Ireland[1], **Christina Rabeler**[1], **Sagar Rao**[1], **Rudy J. Richardson**[2,3,4,5,6]*, **Eva-Maria S. Collins**[1,7]*

1 Department of Biology, Swarthmore College, Swarthmore, Pennsylvania, United States of America, **2** Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Department of Neurology, University of Michigan, Ann Arbor, Michigan, United States of America, **4** Center of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America, **5** Michigan Institute for Computational Discovery and Engineering, University of Michigan, Ann Arbor, Michigan, United States of America, **6** Michigan Institute for Data and AI in Society, University of Michigan, Ann Arbor, Michigan, United States of America, **7** Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

* rjrich@umich.edu (RJR); ecollin3@swarthmore.edu (EMSC)

Citation: Ireland D, Rabeler C, Rao S, Richardson RJ, Collins E-MS (2025) Distinguishing classes of neuroactive drugs based on computational physicochemical properties and experimental phenotypic profiling in planarians. PLoS ONE 20(1): e0315394. https://doi.org/10.1371/journal.pone.0315394

Editor: Michael Schubert, Laboratoire de Biologie du Développement de Villefranche-sur-Mer, FRANCE

Received: August 28, 2024

Accepted: November 25, 2024

Published: January 30, 2025

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files.

Funding: S.R. was funded through a Swarthmore College summer fellowship. The sponsors played no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Abstract

Mental illnesses put a tremendous burden on afflicted individuals and society. Identification of novel drugs to treat such conditions is intrinsically challenging due to the complexity of neuropsychiatric diseases and the need for a systems-level understanding that goes beyond single molecule-target interactions. Thus far, drug discovery approaches focused on target-based *in silico* or *in vitro* high-throughput screening (HTS) have had limited success because they cannot capture pathway interactions or predict how a compound will affect the whole organism. Organismal behavioral testing is needed to fill the gap, but mammalian studies are too time-consuming and cost-prohibitive for the early stages of drug discovery. Behavioral medium-throughput screening (MTS) in small organisms promises to address this need and complement *in silico* and *in vitro* HTS to improve the discovery of novel neuroactive compounds. Here, we used cheminformatics and MTS in the freshwater planarian *Dugesia japonica*–an invertebrate system used for neurotoxicant testing–to evaluate the extent to which complementary insight could be gained from the two data streams. In this pilot study, our goal was to classify 19 neuroactive compounds into their functional categories: antipsychotics, anxiolytics, and antidepressants. Drug classification was performed with the same computational methods, using either physicochemical descriptors or planarian behavioral profiling. As it was not obvious *a priori* which classification method was most suited to this task, we compared the performance of four classification approaches. We used principal coordinate analysis or uniform manifold approximation and projection, each coupled with linear discriminant analysis, and two types of machine learning models–artificial neural net ensembles and support vector machines. Classification based on physicochemical properties had comparable accuracy to classification based on planarian

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: E-MC is the founder of Inveritek, LLC, which offers planarian HTS commercially. RJR currently serves as a member of the advisory board of NeuroX1, a startup biotechnology company that is developing a software platform for the discovery and development of drugs for neurodegenerative diseases. The remaining authors declare no conflict of interest. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

profiling, especially with the machine learning models that all had accuracies of 90–100%. Planarian behavioral MTS correctly identified drugs with multiple therapeutic uses, thus yielding additional information compared to cheminformatics. Given that planarian behavioral MTS is an inexpensive true 3R (refine, reduce, replace) alternative to vertebrate testing and requires zero *a priori* knowledge about a chemical, it is a promising experimental system to complement *in silico* cheminformatics to identify new drug candidates.

Introduction

Mental illness covers a wide range of conditions that cause significant disturbances in cognition, emotional regulation, or behavior. In 2019, 970 million people or 1 in every 8 people across the world had a mental disorder [1, 2]. Moreover, the number of people experiencing anxiety or depressive disorders worldwide has significantly increased since the COVID-19 pandemic [3]. These statistics are worrisome because psychiatric diseases are intrinsically difficult to treat and pose substantial costs to the affected individuals, families, and society [4, 5].

While there is an urgent need for effective therapeutics, neuropsychiatric drug discovery and validation has been stagnant and lags behind that of other diseases [6, 7]. Computational and *in vitro* methods that try to predict interactions with specific molecular targets are widely used for candidate drug discovery for many illnesses. Although some successful psychiatric drugs, such as fluoxetine, have been identified using this hypothesis-driven approach [8], target-specific methods have been largely insufficient for neuropsychiatric drug discovery due to the complex and still enigmatic pathologies of these disorders [6, 9, 10]. Mental illnesses, such as schizophrenia and major depressive disorder, can stem from polygenic and non-genetic etiologies that likely depend on an interplay between many different molecular targets (reviewed in [11]). Consequently, successful psychiatric drugs tend to be neuroactive compounds with multiple pharmacological targets [12]. Thus, organismal models that provide systems-level insight into neuronal function are needed. However, mammalian tests are prohibitively time and cost intensive to screen the large number of chemicals evaluated during the early stages of lead identification [10, 13].

One possible solution to integrate organismal information into first-tier lead identification is to use behavioral phenotyping with small organisms. Fish larvae, worms, and flies are inexpensive to maintain compared to mammals and lend themselves to medium- to high-throughput screening (MTS/HTS, respectively) (reviewed in [14, 15]). Because phenotypic drug discovery does not rely on mechanistic knowledge and can identify polypharmacological drugs (reviewed in [16, 17]), it is especially well-suited for neuropsychiatric drug discovery.

Behavioral barcodes, which reduce the multidimensional phenotypic information into a string of numeric features, provide a quantitative readout of a phenotype [18]. Using statistical tools such as hierarchical clustering or multidimensional scaling, behavioral profiling can be used to identify patterns characterizing drug classes and have been used to predict the effects of novel chemicals for psychiatric uses [17–19]. For example, behavioral barcoding derived from a series of motor responses of zebrafish larvae to different auditory and visual stimuli was used to identify potential novel antipsychotic compounds by comparing to the phenotypic profile produced by classical antipsychotics, such as haloperidol [20]. Similar methodology has also been employed to identify novel monoamine oxidase and acetylcholinesterase inhibitors [21] and sedatives that induce paradoxical excitation [22]. However, the zebrafish larvae used

for these studies were > 7 days old and thus qualify as animal tests that require Institutional Animal Care & Use Committee approval [23].

Here, we test the hypothesis that MTS using an aquatic invertebrate, the planarian *Dugesia japonica*, allows for unbiased classification of neuroactive drugs and augments the information obtained from the physical and chemical properties of the drugs. Freshwater planarians are small (a few mm long) flatworms, have a long history of use in neurotoxicity and neuropharmacology studies (reviewed e.g., in [24–27]), and show promise as a tool for neuroactive drug discovery. The planarian nervous system shares many of the same neurotransmitters and cell types as the mammalian nervous system, yet remains tractable, consisting of approximately 10,000 neurons [26, 28], making the planarian nervous system of intermediate size and complexity compared to nematodes and zebrafish. Genomes and transcriptomes are publicly available [29, 30], and RNA interference can be used to connect genotypes and phenotypes [31–38]. Planarians display stereotypical behavioral responses to drugs of abuse and prescription drugs [39–41]. The development of computational methods for quantitative assessment of these behaviors [42–44] have reignited interest in planarian behavioral studies [15, 26–28, 45]. We have shown that the asexual planarian *D. japonica* is well-suited for behavioral MTS [43, 44, 46, 47] and that the phenotypic profiles gained from planarian MTS can distinguish between different chemical classes/modes of action for neurotoxic compounds [44, 46, 48]. Planarian MTS has unique strengths that complement screening using zebrafish larvae or roundworms for developmental neurotoxicology [14]. Adult nervous system function can be evaluated in planarians at a fraction of the time and cost required for zebrafish.

We hypothesized that *D. japonica* behavioral phenotyping would be a good model for the identification of prospective neuroactive drugs. To test this, we used our robotic screening platform to determine how well our battery of phenotypic endpoints could distinguish between classes of known neuropsychiatric drugs and add value to a classification of these compounds based on their physicochemical properties. We studied 19 neuroactive compounds from three functional classes: antipsychotics, anxiolytics, and antidepressants. We classified these compounds using either physicochemical descriptors or planarian behavioral profiling using four different classification approaches. We used principal coordinate analysis (PCoA) or uniform manifold approximation and projection (UMAP), each coupled with linear discriminant analysis (LDA), and two types of machine learning models—artificial neural net ensembles (ANNE) and support vector machines (SVMs). We found that the classification based on physicochemical properties had comparable accuracy to planarian profiling. Planarian profiling correctly identified polytherapeutic drugs and thus yielded additional information that would have been missed by cheminformatics. Thus, combining these two approaches may be an economical and useful method for identifying novel drug candidates.

Results and discussion

The purpose of this study was to investigate the extent to which members of neuroactive drug classes can be identified and distinguished using computational classification using either cheminformatics or *in vivo* behavioral phenotyping in planarians. To do this, we used a set of 19 neuroactive drugs that were functionally defined by the supplier as one of three categories: antidepressants (7), antipsychotics (7), or anxiolytics (5), (Table 1, Fig 1).

We assigned each chemical a barcode consisting of a numerical string based on either their computed physicochemical properties or responses observed in planarian phenotypic screening. We then used the same computational pipeline to study how well these compounds could be accurately classified based on either *in silico* cheminformatics or *in vivo* planarian behavioral phenotyping. We compared the results obtained from four different computational

Table 1. Study compounds.

CAS #	ID	Common Name[a]	Class[b]	Source[c]	Purity (Method)[d]	Solvent[e]	Study Type[f]	Tested Concentrations (units)
129722-12-9	ARI	Aripiprazole·HCl	2	A	≥98% (HPLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
10457-90-6	BRO	Bromperidol	2	A	100% (EP reference standard)	DMSO	E,C	0.316, 1, 3.16, 10, 31.6, 100 (μM)
437723-96-1	BUA	(*R*)-isomer of BUP	1	NA	NA	NA	C	NA
324548-43-8	BUB	(*S*)-isomer of BUP	1	NA	NA	NA	C	NA
31677-93-7	BUP	Bupropion·HCl Racemic mixture	1	A	≥98% (HPLC)	water	E	10, 31.6, 100, 316, 1000 (μM)
33386-08-2	BUS	Buspirone·HCl	3	A	≥99% (TLC)	water	E,C	0.316, 1, 3.16, 10, 31.6, 100 (μM)
59729-32-7	CIT	Citalopram·HBr Racemic mixture	1	A	quality level 300; certified reference material (HPLC)	DMSO	E	1, 3.16, 10, 31.6, 100, 316, 1000 (μM)
128196-02-1	CIR	(*R*)-isomer of CIT	1	NA	NA	NA	C	NA
5786-21-0	CLO	Clozapine	2	B	≥99% (HPLC)	DMSO	E,C	0.1, 0.316, 1, 3.16, 10, 31.6, 100 (μM)
439-14-5	DIA	Diazepam	3	C	99.3% (GC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
548-73-2	DRO	Droperidol	2	A	≥98% (TLC)	DMSO	E,C	0.1, 0.316, 1, 3.16, 10, 31.6, 100 (μM)
136434-34-9	DUL	Duloxetine·HCl (*S*)-isomer	1	A	≥98% (HPLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
219861-08-2	ESC	Escitalopram·oxalate (*S*)-isomer of CIT	1	A	≥98% (HPLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100, 316, 1000 (μM)
57653-26-6	FEN	Fenobam	3	A	≥98% (HPLC)	DMSO	E,C	56.2, 100, 178, 316, 562 (μM)
100568-03-4	FLR	(*R*)-isomer of FLU	1	NA	NA	NA	C	NA
100568-02-3	FLS	(*S*)-isomer of FLU	1	NA	NA	NA	C	NA
56296-78-7	FLU	Fluoxetine·HCl Racemic mixture	1	A	≥98% (TLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
52-86-8	HAL	Haloperidol	2	A	≥98% (TLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
113-52-0	IMI	Imipramine·HCl	1	A	≥99% (TLC)	DMSO	E,C	0.316, 1, 3.16, 10, 31.6 (μM)
110-16-7	MAL	Maleic acid	4	A	≥99% (HPLC)	DMSO	E,C	200 (μM)
59467-70-8	MID	Midazolam	3	A	100% (USP reference standard)	DMSO	E,C	10, 17.8, 31.6, 56.2, 100 (μM)
132539-06-1	OLA	Olanzapine	2	A	≥98% (HPLC)	DMSO	E,C	0.001, 0.00316, 0.01, 0.0316. 0.1, 0.316, 1, 3.16, 10, 31.6, 100, 316 (μM)
144-62-7	OXA	Oxalic acid	4	A	Analytical standard (redox titration)	DMSO	E,C	100, 316, 1000 (μM)
84-02-6	PRO	Prochlorperazine·dimaleate	2	A	≥98% (TLC)	DMSO	E,C	0.316, 1, 3.16, 10, 31.6, 100 (μM)

(*Continued*)

Table 1. (Continued)

CAS #	ID	Common Name[a]	Class[b]	Source[c]	Purity (Method)[d]	Solvent[e]	Study Type[f]	Tested Concentrations (units)
79559-97-0	SER	Sertraline·HCl *cis*-(1*S*,4*S*)-isomer	1	A	≥98% (HPLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)
7647-15-6	SOB	Sodium bromide	4	A	≥99.9% (trace metal analysis)	DMSO	E	100, 316, 1000 (μM)
7647-14-5	SOC	Sodium chloride	4	A	≥99.0% (titration with AgNO3)	water	E	3.47 (mM)
1135210-68-2	TRA	Tracazolate·HCl	3	B	≥99.0% (HPLC)	DMSO	E,C	1, 3.16, 10, 31.6, 100 (μM)

[a] All common names are those provided in the National Institutes of Health Global Substance Registration System. Except for maleic acid, oxalic acid, sodium bromide and D-sorbitol, all names also represent the recommended International Non-Proprietary Name.

[b] Primary functional pharmacological class as indicated by the manufacturer and literature: 1 = antidepressant; 2 = antipsychotic; 3 = anxiolytic; 4 = counterion control.

[c] Source of chemicals used in experimental determinations: A = Sigma-Aldrich; B = Tocris; C = Spectrum Chemicals; NA = not applicable (molecular structures used only for *in silico* determinations).

[d] Method used by the supplier to assess purity. HPLC: High-performance liquid chromatography; EP: European Pharmacopeia; GC: Gas chromatography; TLC: Thin layer chromatography; USP: US Pharmacopeia.

[e] Solvent used for stock solutions.

[f] Study type: E = experimental; C = computational. For computational studies, counterions were not present.

https://doi.org/10.1371/journal.pone.0315394.t001

methods, including PCoA followed by LDA (PCoA-LDA), UMAP followed by LDA (UMAP-LDA), and two machine-learning approaches (ANNE and SVMs) (Fig 2), to investigate whether a specific method would work best on either or both types of data. In the Methods section, we provide background on each method and explain why we chose to apply it to our data.

Classification based on physicochemical properties

The neuroactive drugs studied here represent a diverse range of chemical structures, as shown by their differing Tanimoto similarity coefficients, compared to ARI as a reference compound (Fig 3). While some clustering by functional pharmacological class is visible, several compounds (e.g., the antipsychotics OLA and CLO and most anxiolytics) are structurally distinct from the other members of their class. This structural diversity reflects the purpose of this library–to serve as a pilot to determine the extent to which our models could assign the compounds to their functional pharmacological classes despite structural dissimilarities within each class.

As shown in Table 1, some drugs (BUP, CIT, FLU) are mixtures of stereoisomers. Thus, we had to consider whether to treat these chemicals as 2D structures, without considering stereochemistry, or as 3D structures, considering each stereoisomer separately thereby increasing the number of entities considered. Because the number of compounds affects the classification and we wanted to keep the data size comparable to the planarian data, which considers each tested compound as a single entity, we present the classifications based on 2D structural features here without consideration of stereochemistry (Figs 4–6). Results from using 3D chemical features, considering each stereoisomer separately, are shown in S1–S3 Figs. The 3D models show similar, though slightly better, classification accuracies than the models using only 2D features, which may be due to the increased information per compound and/or the increase in the number of compounds (18 versus 21) compared to 2D. Of note, when

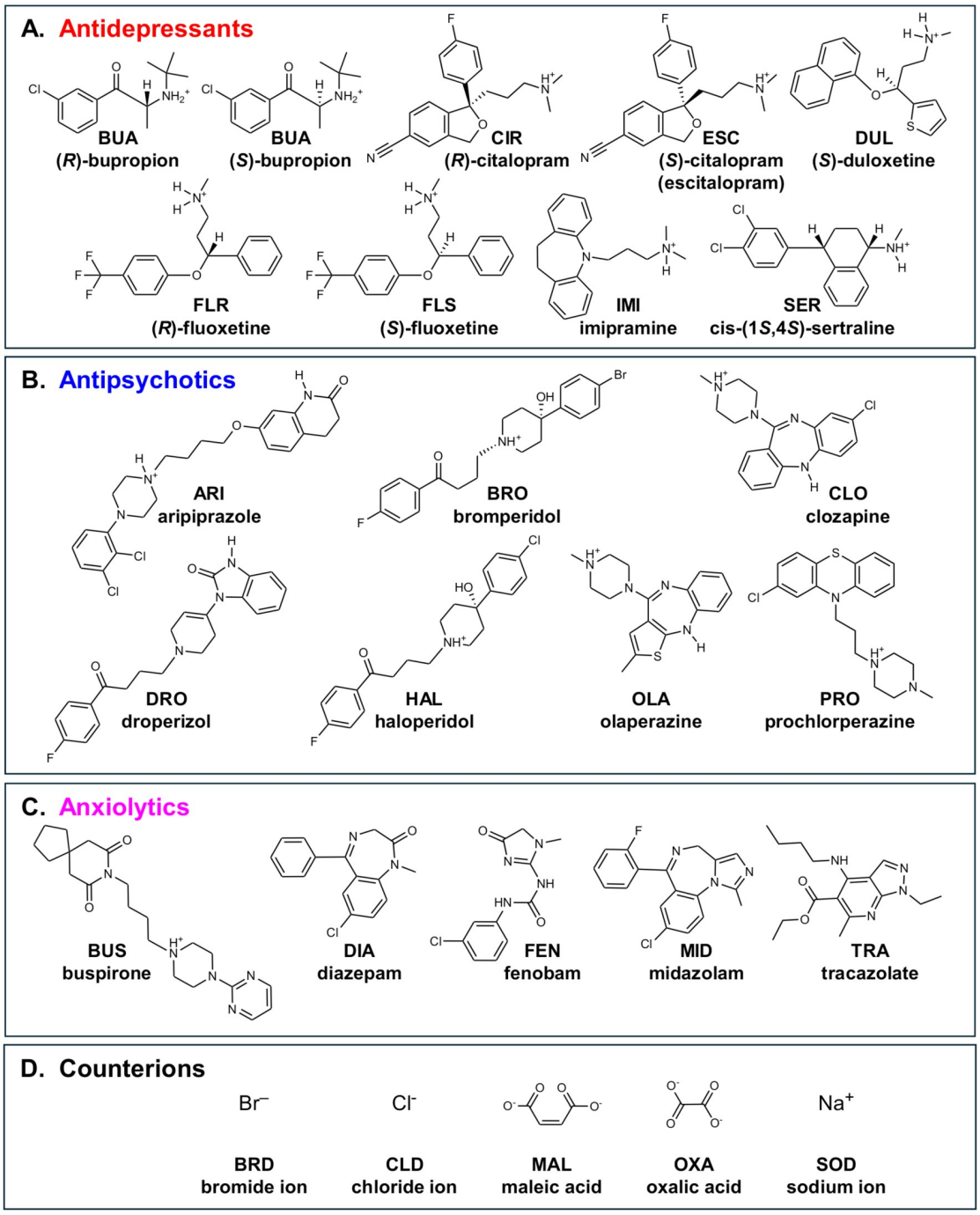

Fig 1. Chemical structures. Molecular structures depicted as 2D with stereochemistry indicated by wedged bonds are provided for the different chemical classes: A) Antidepressants, B) Antipsychotics, C) Anxiolytics, and D) Counterions. Counterions are not shown with their respective drug because they were not included in the computational studies. Table 1 lists the experimental compounds, including counterions, that were tested in planarians. Protonation states were based on the dominant form at pH 7.4 as determined by the protonation module in the ChemAxon Marvin suite (https://www.chemaxon.com).

https://doi.org/10.1371/journal.pone.0315394.g001

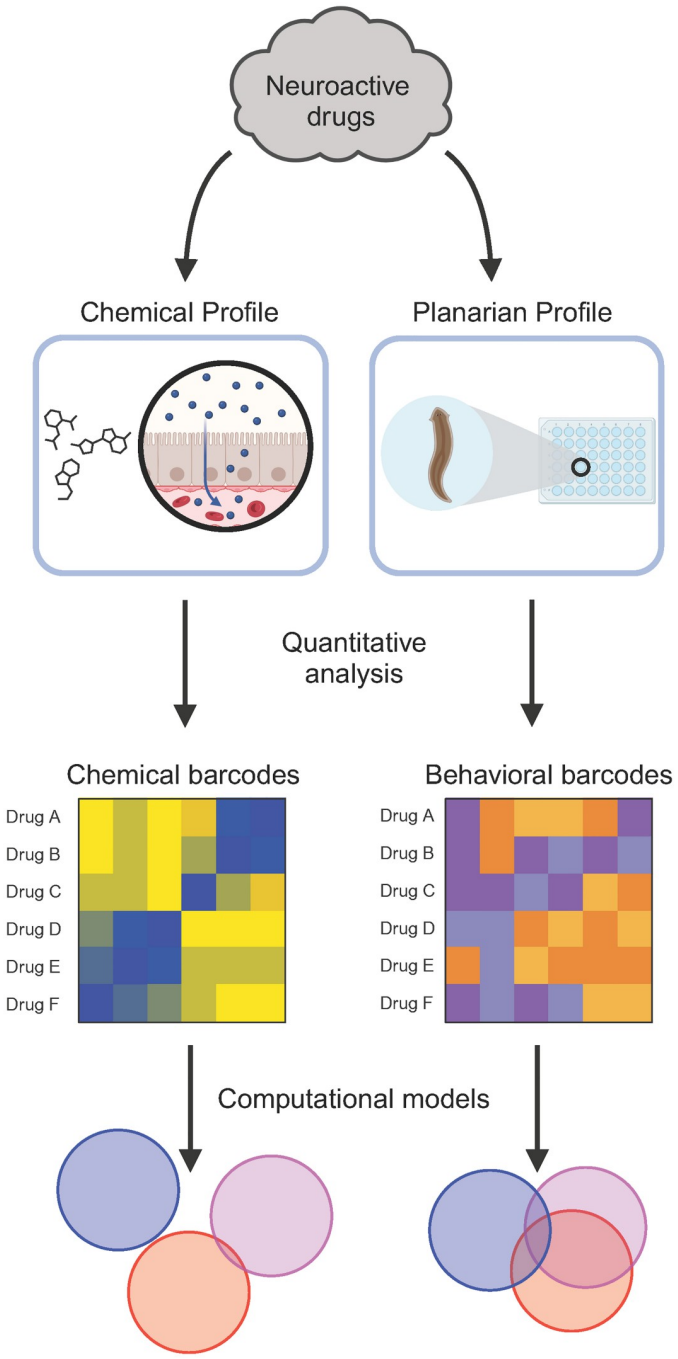

Fig 2. Study overview. Chemical and behavioral profiles were determined for 19 neuroactive drugs, consisting of 7 antidepressants, 7 antipsychotics, and 5 anxiolytics. Using quantitative analyses, we then determined molecular and phenotypic barcodes for each compound. These barcodes were used in the same computational models to determine how well each method performed at classifying the 3 neuroactive drug classes. Created with BioRender.com.

https://doi.org/10.1371/journal.pone.0315394.g002

considering 2D structures, CIR and ESC were treated as one structure, which we labeled CIT, leading to only 18 drugs in these models.

Comparisons were made both with and without the counterions to determine how inclusion of these presumedly null-effect compounds would affect neuroactive drug classification.

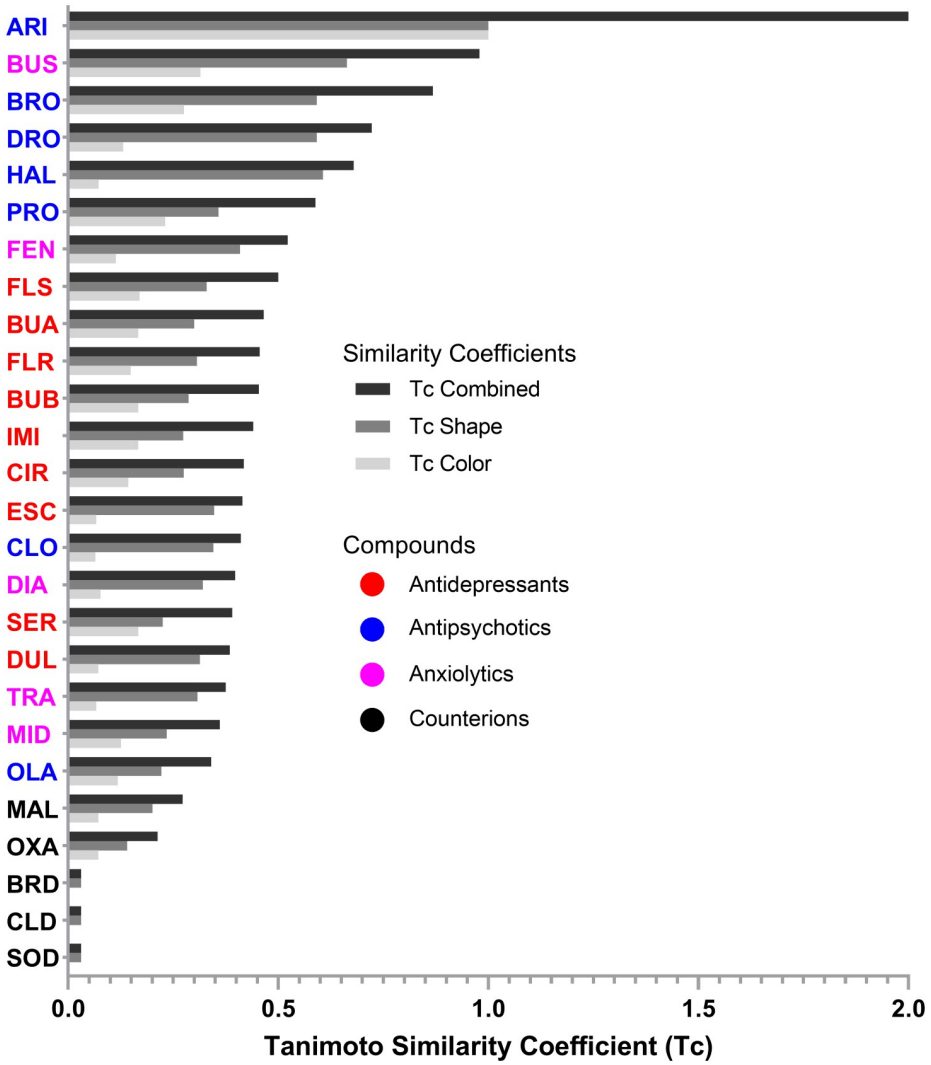

Fig 3. Tanimoto similarity coefficients. Chemical similarity was quantified using the Tanimoto coefficient (Tc). Overlaps were determined after 3D alignment of each structure with the reference compound, ARI. ARI was chosen as the reference compound as it had the largest molecular volume as assessed by the van der Waals radii of the constituent covalently bonded atoms (see Methods). Tc Shape refers to similarity based on overlap of molecular volume. Tc Color refers to similarity based on overlap of 6 pharmacophores (H-bond acceptor, H-bond donor, anion, cation, hydrophobe, or ring). Tc Combined = Tc Shape + Tc Color. Range of Tc Shape and Tc Color = (0,1); range of Tc Combined = (0,2). Chemicals are ordered by their Tc Combined scores and color-coded by functional class.

https://doi.org/10.1371/journal.pone.0315394.g003

ANNE and SVMs were each initially run across 10 different models (S1–S9 Tables) and then the best performing model for each was chosen (see Methods) for comparison across methods.

The accuracy of these models on the small data set used here was high and comparable between PCoA-LDA, ANNE, and SVMs (Figs 5–6). Overall, the UMAP-LDA performed worse than the other three models on all data sets (2D, 3D +/- counterions), which could also be seen in the high overlap between 95% confidence intervals of the chemical clusters in Fig 4. Interestingly, when using the 2D chemical features and excluding the counterions, UMAP-LDA misclassified most anxiolytics (Fig 6B). Only FEN was correctly classified. Generally, classification accuracies were improved by including the counterions, which may be due to the increased number of chemicals.

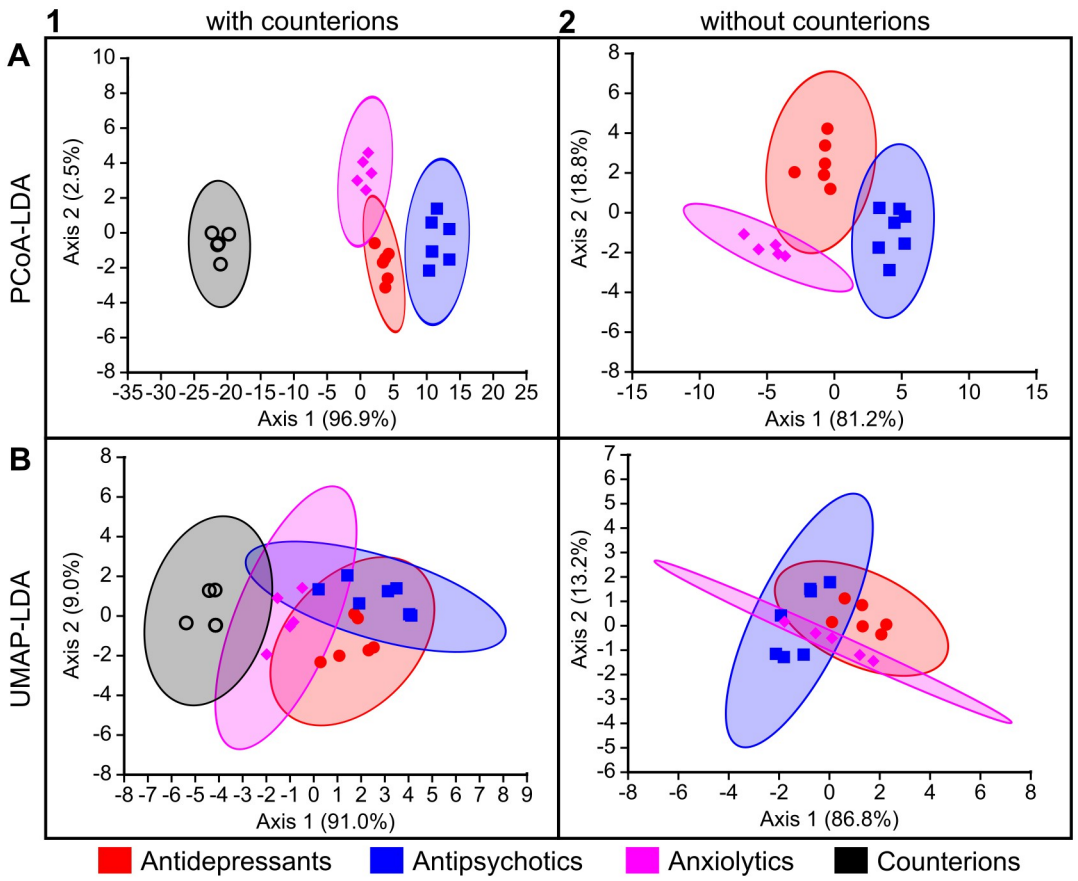

Fig 4. Functional pharmacological classification of the study compounds based on their 2D chemical descriptors using LDA-based approaches. Computational methods are arranged by rows: (A) PCoA-LDA; (B) UMAP-LDA. Counterion inclusion/exclusion is arranged by columns: (1) with counterions; (2) without counterions. Ellipses represent 95% confidence intervals. The axes show the percentage of the total eigenvalues; this does not sum to 100% in (A1) because there was a third axis that accounted for the remaining 0.60%.

https://doi.org/10.1371/journal.pone.0315394.g004

When comparing the misclassifications from the 2D and 3D chemical descriptors, we found that the antidepressant BUP was the most frequently misclassified drug across the different methods when using 2D structural features (Figs 5 and 6). In contrast, the classifications based on 3D chemical structures did not misclassify any antidepressants, including the BUP stereoisomers (BUA and BUB), and most frequently misclassified anxiolytics (S2 and S3 Figs). When considering 3D chemical features, the most misclassified chemical across the different methods was the anxiolytic BUS (S2 and S3 Figs). BUS has a very different structure compared to the other anxiolytics (Fig 1) and looks more similar to some of the antipsychotics (Fig 3), which could explain its frequent misclassification as an antipsychotic. Moreover, while the anxiolytic class had the fewest members (5), its members are also the most diverse as can be seen by the greater spread in the Tanimoto similarity coefficients (Fig 3).

Planarian behavioral phenotyping

Activity and potency determined by benchmark concentration (BMC) modeling. The chemicals were screened acutely (< 3 hour exposure) in adult planarians and assessed for 13 endpoints spanning effects on lethality, body shape, stickiness, locomotion, and reaction to

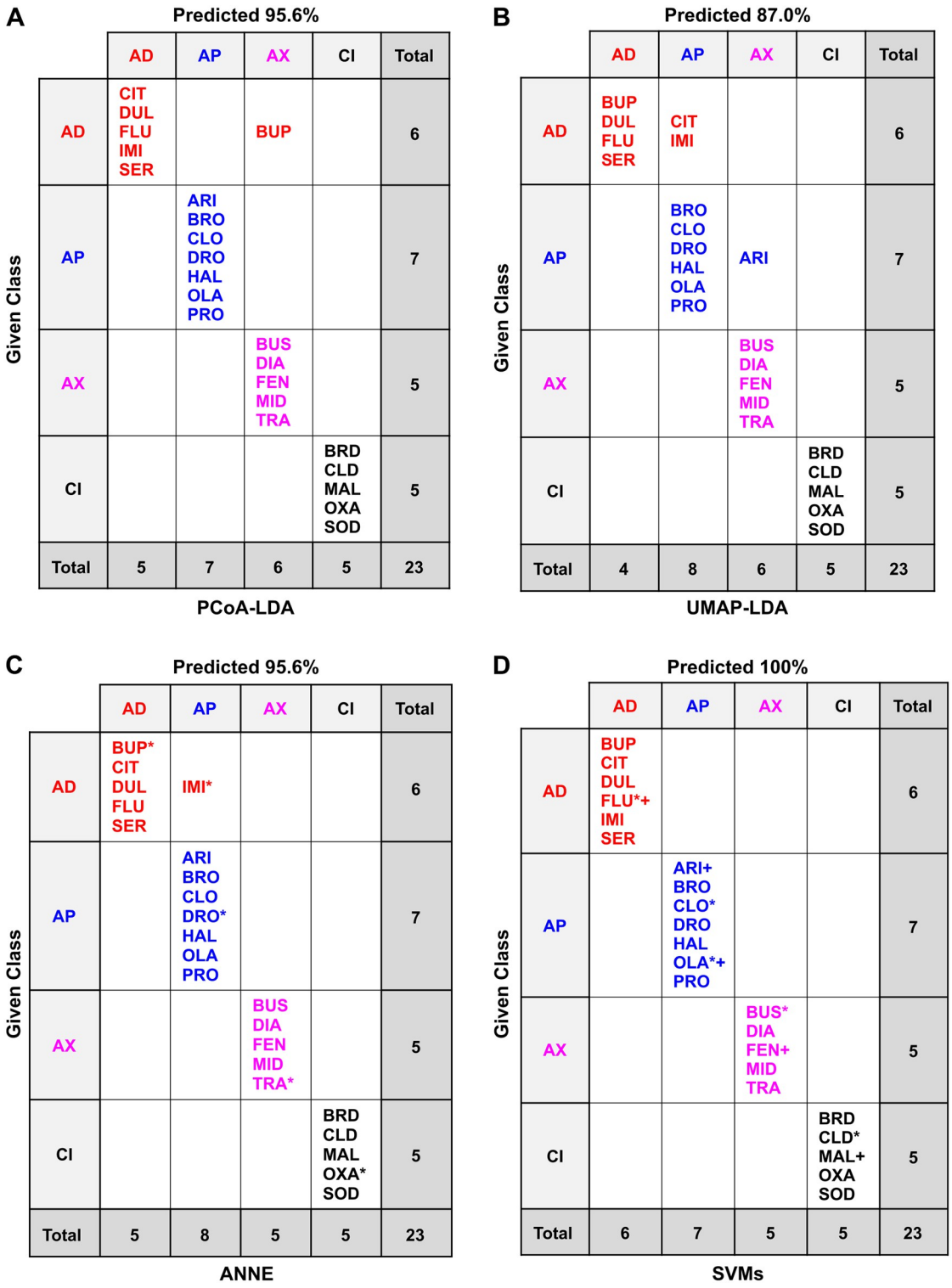

Fig 5. Confusion matrices of the classification methods based on 2D chemical descriptors of the drugs and counterions. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs; AD: antidepressant, AP: antipsychotic, AX: anxiolytic; CI: counterion. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicted accuracy refers to the overall accuracy of the best of 10 models of each type. *indicates randomly chosen members of the test set. In (D) two models were tied for the best so that the test set members are marked with both * (model 07_2i) and + (model 09_2i). See the Methods for model selection. Test set accuracies were 80.0% for ANNE (S2 Table) and 100% for SVMs (S4 Table).

https://doi.org/10.1371/journal.pone.0315394.g005

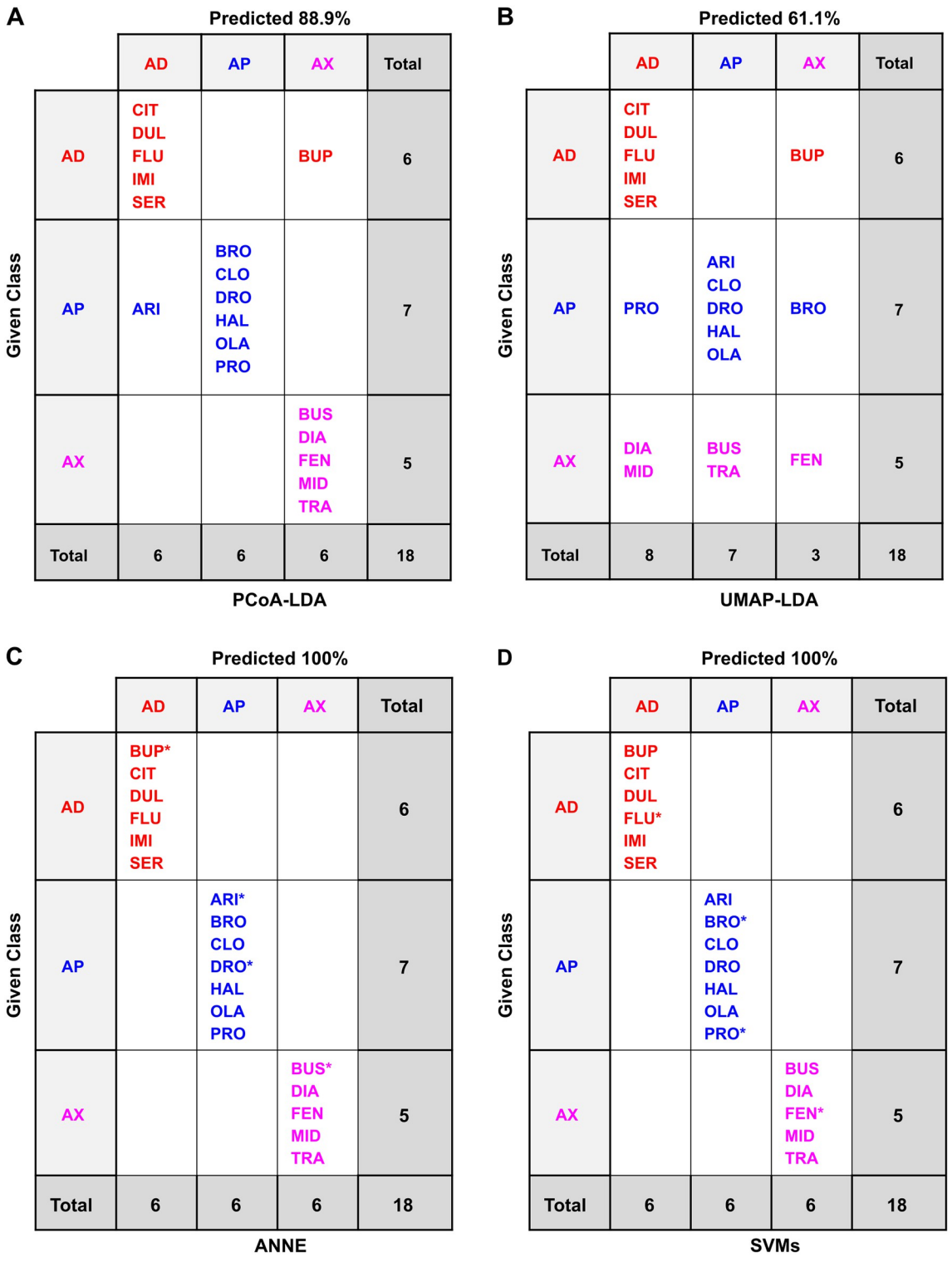

Fig 6. Confusion matrices of the classification methods based on 2D chemical descriptors of the drugs only. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs; AD: antidepressant, AP: antipsychotic, AX: anxiolytic. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy of the best of 10 models of each type. *indicates randomly chosen members of the test set. See the Methods for model selection. Test set accuracies were 100% for ANNE (S3 Table) and 100% for SVMs (S5 Table).

https://doi.org/10.1371/journal.pone.0315394.g006

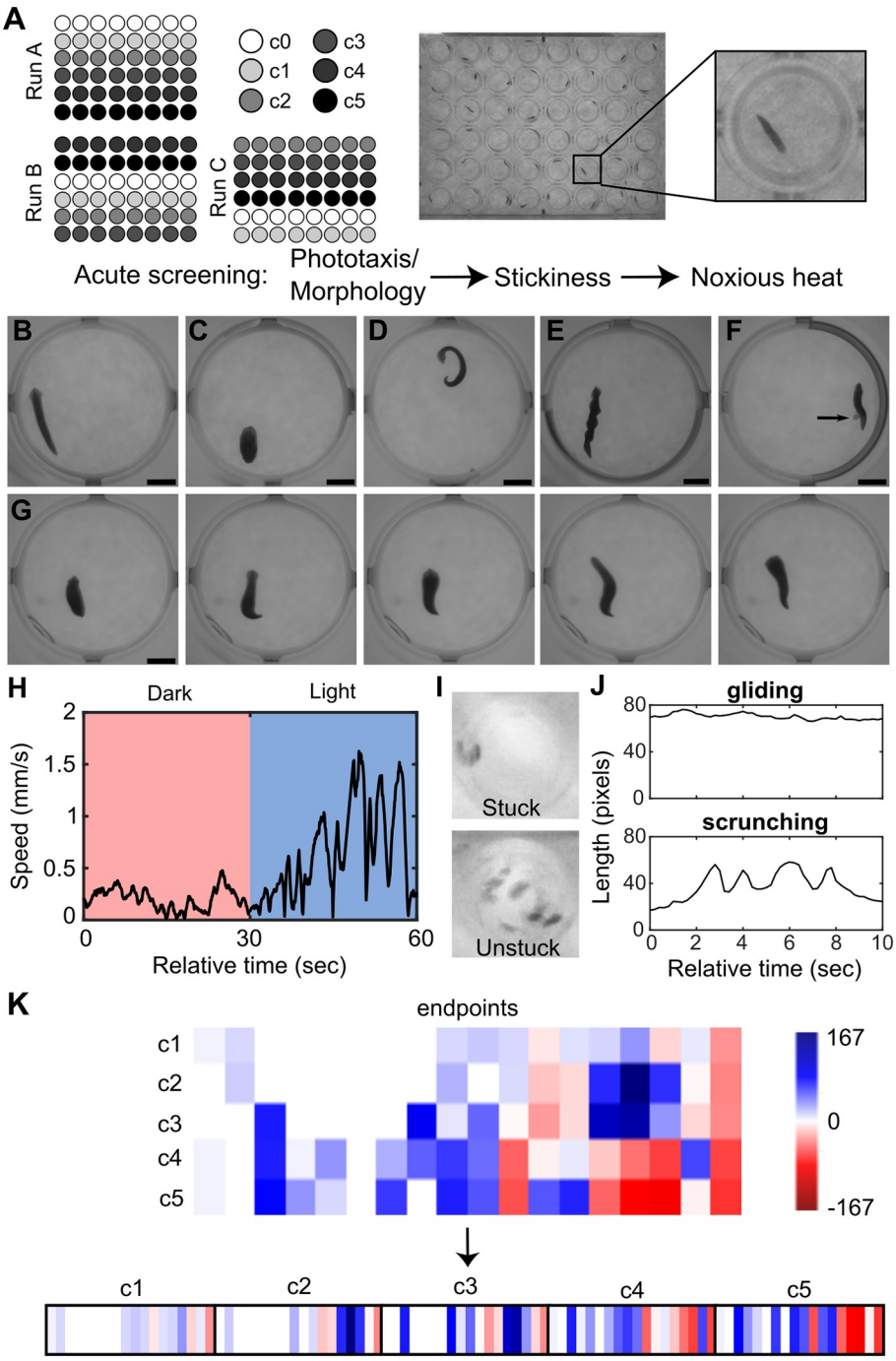

Fig 7. Overview of planarian screening and data analysis. A) Schematic of plate setup, including the orientations used in the triplicate runs, and representative image of a 48-well plate containing one planarian per well. Each plate contained a solvent control (c0) and 5 concentrations of a chemical and was tested acutely in phototaxis, stickiness, and noxious heat assays. B-G) Examples of various planarian body shapes: B) Normal planarian with smooth gliding. C) Contracted with shortened length and often associated with ruffling of the edges of the planarian. D) C-shape/curled and often on its side. E) Corkscrew showing multiple twists across the body axis. F) Pharynx extrusion. Arrow points at the unpigmented pharynx protruding from the underside of the planarian. G) Example image sequence of a planarian undergoing scrunching in response to chemical exposure at room temperature, showing oscillations of body length. This is one type of behavior, among others, which constitute the hyperkinesis category. Each frame is 1 sec apart. Examples shown are exposed to 0.5% (v/v) DMSO (B), 10 μM DRO (C and D), 10 μM BRO (E), 10 μM DUL (F), and 1 μM OLA (G). Scale bars: 2 mm. H) Representative plot of speed over time during the phototaxis assay. During a

normal phototaxis response, planarians increase their speed during the blue light period. I) Minimum intensity projections of the shaking portion of the stickiness assay. A "stuck" planarian adheres to the bottom of the well and is not displaced, whereas an "unstuck" planarian dislodges and is displaced around the well. J) Length over time plots showing normal gliding or the oscillatory scrunching gait [49] that is induced during the noxious heat assay. Given that scrunching is the expected behavior during the noxious heat assay, we scored the number of planarians that do not scrunch. K) Example schematic representation of the behavioral barcodes. First, a barcode is created for each chemical concentration, consisting of a numerical vector of the compiled normalized score for each endpoint (columns). One master barcode is made for each chemical by concatenating the barcodes for the individual concentrations. For details, see Methods.

https://doi.org/10.1371/journal.pone.0315394.g007

Table 2. Binary endpoints.

Endpoint	Description	BMR[a]
Crawl-out	% planarians that crawled out of the water	25[b]
Stickiness	% stuck individuals	25
Shape (any)	% individuals with any abnormal body shape	25[b]
Phototaxis	% planarians that did not phototax [50]	35
Scrunching	% planarians that did not scrunch in response to noxious heat	30

[a]BMR: benchmark response

[b]Variance was already minimized at all tested thresholds, thus the BMR was manually defined. See Methods.

https://doi.org/10.1371/journal.pone.0315394.t002

light and noxious heat (Fig 7, Tables 2 and 3). None of the tested compounds caused lethality at the tested concentrations, though a few compounds caused a substantial number of planarians to crawl out of their wells and dry out ("crawl-out behavior"). The BMC was calculated for each endpoint to determine when phenotypic responses exceeded empirically determined noise levels, defined by the benchmark response (BMR) (Fig 8, Tables 2 and 3, S10–S11 Tables).

Except for FEN, all chemicals showed activity in at least one endpoint in at least one tested concentration. FEN did not show any observable effects up to 562 μM (the highest tested soluble concentration). Moreover, no lethality (0%, $n = 32$) or qualitative defects were observed in 562 μM FEN after a 12-day exposure (S5 Fig), suggesting it may not be bioavailable to planarians. ESC, the (*S*)-stereoisomer of CIT, was only active at ≥316 μM. However, OXA, its counterion control, also caused similar phenotypes at these concentrations (S12 Table). MAL at

Table 3. Continuous endpoints.

Endpoint	Description	normalization	Direction[a]	BMR[b]
Anxiety	Fraction of time spent in outer region of well [48]	$(Response_{chemical}/ Response_{vehicle})*100–100$	-	30
Resting_dark	Fraction of time spent resting in 2nd dark cycle	$(Response_{chemical}-Response_{vehicle})*100$	+	60[c]
Resting_blue	Fraction of time spent resting in 2nd blue cycle	$(Response_{chemical}-Response_{vehicle})*100$	+	35[c]
Speed_dark	Mean speed (mm/s) in 2nd dark cycle	$(Response_{chemical}/ Response_{vehicle})*100–100$	-	60[c]
Speed_blue1	Mean speed (mm/s) in 1st 30 seconds of blue cycle	$(Response_{chemical}/ Response_{vehicle})*100–100$	-	80[c]
Speed_blue2	Mean speed (mm/s) in 2nd 30 seconds of blue cycle	$(Response_{chemical}/ Response_{vehicle})*100–100$	-	70[c]
Locomotor bursts (total)	Sum of locomotor bursts in phototaxis assay	$Response_{chemical}-Response_{vehicle}$	+	20[c]
Noxious stimuli (strength)	Median displacement at end of noxious heat [47]	$(Response_{chemical}/ Response_{vehicle})*100–100$	-	55[c]

[a] The directionality (increasing (+) or decreasing (-) compared to the median control response) of the effect.

[b]BMR: benchmark response

[c]BMR was determined based on the 5th or 95th percentile of the histograms of the normalized vehicle control response (S4 Fig)

https://doi.org/10.1371/journal.pone.0315394.t003

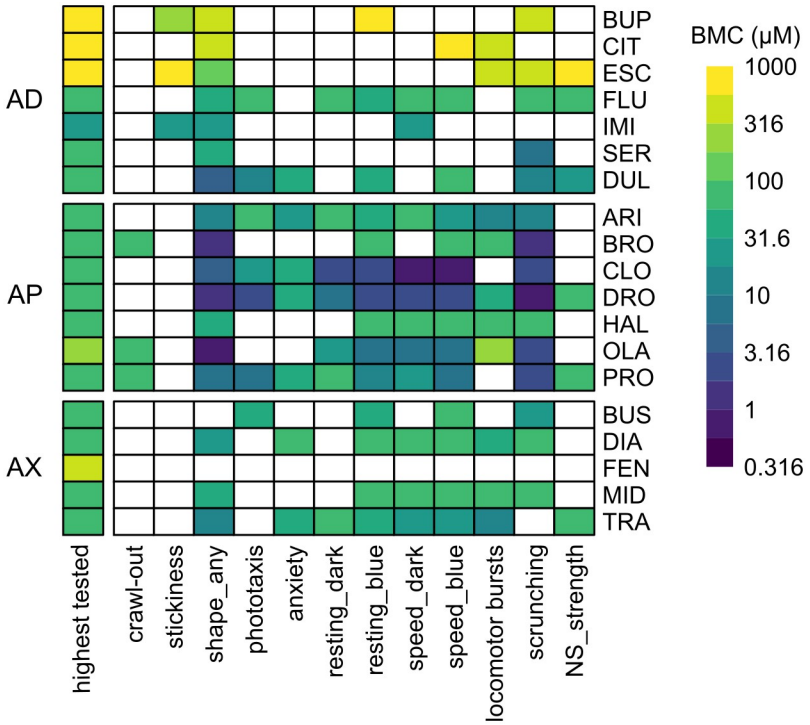

Fig 8. Activity of neuroactive drugs in planarians. Heatmap comparing the benchmark concentrations (BMCs) for all neuroactive compounds in planarians after acute exposure. The first column shows the highest tested concentration. As both "speed_blue" endpoints had similar BMC scores, only "speed_blue2" is shown as this was the more sensitive of the two timepoints. AD: antidepressant, AP: antipsychotic, AX: anxiolytic, NS: noxious stimuli.

https://doi.org/10.1371/journal.pone.0315394.g008

200 μM, the counterion control for PRO, showed a significant effect on scrunching (S12 Table), but as PRO induced scrunching defects starting at much lower concentrations (3.16 μM), this is likely not driven by the effects of the counterion. All other counterions for the salt forms of the drugs were found to be negative. To assess whether any of the observed effects were due to pH effects instead of compound-specific effects, the pH of the highest tested concentration of each chemical was measured (S13 Table). For reference, we measured the pH of Instant Ocean (IO) water as 6.72 ± 0.41 and that of 0.5% (v/v) DMSO as 6.63 ± 0.17 (mean ± SD, $n = 5$), which is consistent with previous reports of acceptable pH ranges for planarian culture conditions [51]. Acute behavioral effects were seen with acidic conditions at pH <4 (S13 Table). Both ESC and OXA had pH values <4 at the two highest tested concentrations (316 and 1000 μM), while at 100 μM, where no behavioral effects were observed, the pH was >4. Together, these data suggest that the effects seen at these concentrations may be due to low pH. Of note, 1000 μM CIT, which did cause behavioral effects, had a pH of 6.42, suggesting that the pH effects of ESC may have been driven by the OXA counterion. All other compounds were within acceptable pH ranges where no adverse effects were observed and thus the observed effects are presumed to be due to the pharmacological activity of the drugs.

Even within the same class, potency differences of several orders of magnitude were seen, which may reflect differences in uptake and metabolism or in affinity to the planarian neuronal targets. Lipophilicity can affect chemical bioavailability; therefore, we plotted BMC versus the calculated distribution coefficient, logD for each chemical. Because logD can vary depending on the pH, we plotted the range of logD values corresponding to the presumed pH range

of the tested concentrations (ranging from the pH of the solvent as a proxy for the pH of the lowest concentration to the measured pH of the highest tested concentration (S13 Table)). We found no correlation between logD and BMC (S6 Fig), suggesting lipophilicity alone cannot explain the differences in potency.

Because of the observed potency differences, we evaluated the phenotypic profiles independent of concentration to identify any class-specific patterns (Fig 9). A significant loss of scrunching behavior, i.e., hits in scrunching, was seen in almost all the tested compounds, agreeing with our previous data showing that absence of heat-induced scrunching is a sensitive readout of disturbed neuronal function across a broad range of chemical types [44, 46, 48]. Generally, the antipsychotics and anxiolytics showed a broader range of effects across multiple endpoints than the antidepressants (Fig 9A). For example, decreased motility in the dark and increased number of locomotor bursts were seen in >70% of the tested antipsychotics and anxiolytics, respectively, but only 29% of antidepressants. Effects in phototaxis and anxiety were more prominent in the antipsychotics than in the other classes. While "any abnormal body shape" was seen with almost all the tested drugs, the specific type of body shape differed across the drug classes (Fig 9B). For example, the antidepressants were largely characterized by the presence of hyperactive shapes/behaviors, without the inclusion of other body shape classes. In contrast, antipsychotics and anxiolytics had more mixed phenotypes with instances of contraction, C-shapes, and hyperactivity or C-shapes and hyperactivity, respectively.

One possible confounding factor in this hit comparison, which aggregates effects at all concentrations, is that some of the highest tested concentrations may reach toxic levels and thus

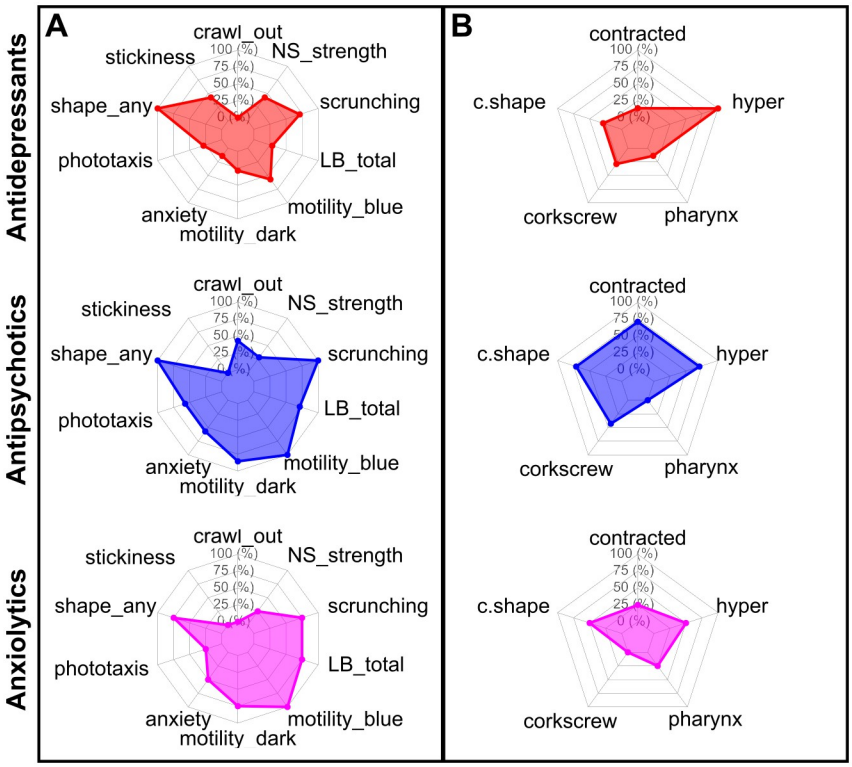

Fig 9. The drug classes showed different planarian phenotypic profiles. Radar plots showing the percentage of chemicals in each class that had a hit at any concentration in (A) any endpoint or (B) specific body shape classes. In (A) speed and resting endpoints in each light period were combined into a "motility" readout. LB: locomotor bursts; NS: noxious stimuli.

https://doi.org/10.1371/journal.pone.0315394.g009

the observed phenotypes may be due to a mix of neuroactivity and toxicity. In fact, a common struggle in studies on neuroactive compounds is to distinguish neuroefficacy from toxicity, which both need to be considered during the drug discovery process. Importantly drug toxicity is the second most common reason (after lack of clinical efficacy) for promising drugs to fail in clinical trials [52]; thus, it is important to identify possible toxicity early in the drug discovery pipeline. While we currently do not have sufficient understanding of the phenotypic readouts to be able to delineate neuroactive vs adverse effects, effects at low concentrations are likely to constitute efficacy and not toxicity. Because nominal concentrations do not necessarily reflect tissue concentrations due to compound-specific differences in uptake and metabolism, it is impossible to define a water concentration threshold that would separate efficacy from toxicity. Future work characterizing planarian drug metabolism will be helpful to gain a better understanding of how physicochemical properties affect chemical bioavailability. Therefore, in the absence of this knowledge, we included all nominal test concentrations for the planarian classifications.

Classification of neuroactive drug classes using planarian behavioral phenotyping. Ultimately, the strength of planarian behavioral screening is in the multidimensional information gained from looking across multiple endpoints testing various neurological functions and not just at each endpoint in isolation. To this end, each tested concentration of a compound was given a phenotypic barcode consisting of the compiled normalized score compared to in-plate solvent controls (see Methods, Fig 7K). For each chemical, one master barcode was created by concatenating the barcodes of all tested concentrations in relative order (S7 Fig). To keep comparisons similar across different test concentrations, the relative order of the concentrations was considered rather than the absolute concentration. For the classification methods, the data were truncated to only the highest 5 concentrations such that negative/missing data were not driving the classifications. These planarian barcodes were then evaluated using the same computational methods as for the chemical features to determine how well the neuroactive drug classes could be separated (Figs 10–12). Compared to Fig 4, which uses the chemical descriptors as input, Fig 10 shows more overlap between the clusters in the LDA-based approaches. The anxiolytics category completely overlaps with the other categories.

As with the chemical descriptors, comparisons were made both with and without the counterions. ANNE and SVMs were each initially run across 10 different models (S15–S18 Tables). The best performing model for each was chosen (see Methods) for comparison across methods. It is important to note that the algorithm used for training the ANNE models incorporated a stopping procedure to guard against overfitting, whereas the SVM algorithm did not provide this feature. However, while our process for selecting the best of ten models of each type included terms to reward accuracy, it also included terms to penalize complexity according to the number of neurons and descriptors for ANNE models or the number of descriptors for SVM models (see Methods for details).

All four classification methods were able to distinguish the neuroactive drug classes based on planarian phenotypic barcodes, with accuracies ranging from 59.1% to 100%. The machine learning methods, particularly SVMs, tended to perform better than PCoA-LDA or UMAP-LDA. Similar to the classification based on physicochemical data, UMAP-LDA performed the worst of the four methods. While this difference was clear when counterions were included (accuracy of 59% versus >72% in all other models (Fig 11)), UMAP-LDA and PCoA-LDA had practically the same adjusted accuracy when counterions were excluded (Fig 12). In contrast to the cheminformatic classifications, inclusion of the counterions decreased accuracy for all methods except SVMs, which had 100% accuracy in all cases. This decreased accuracy is likely because in both PCoA-LDA and UMAP-LDA the counterion OXA was "misclassified" as one of the drug classes due to the observed behavioral effects from the low pH.

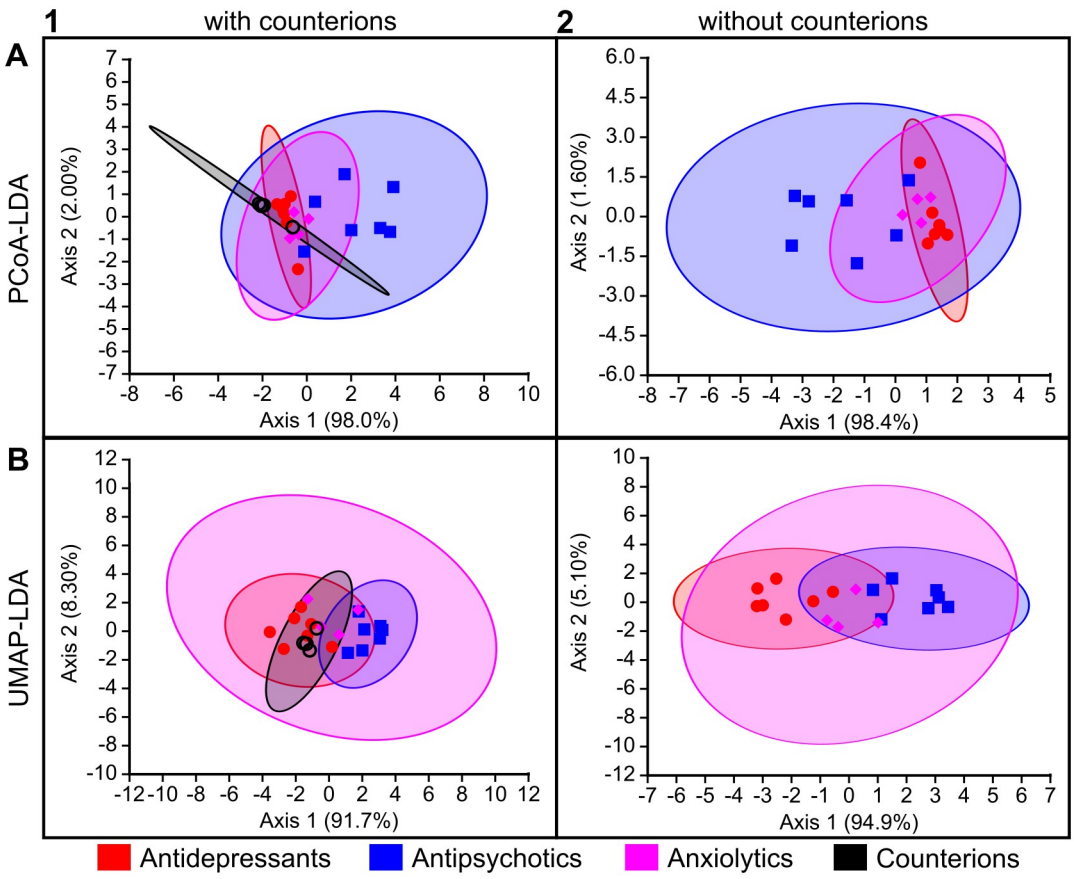

Fig 10. Classification of planarian behavioral phenotyping based on LDA-based approaches. Computational methods are arranged by rows: (A) PCoA-LDA; (B) UMAP-LDA. Counterion inclusion/exclusion is arranged by columns: (1) with counterions; (2) without counterions. Ellipses represent 95% confidence intervals. The axes show the percentage of the total eigenvalues.

https://doi.org/10.1371/journal.pone.0315394.g010

Because FEN was inactive in the planarian screen, it was not included in these classifications. Inclusion of FEN (S8–S10 Figs and S19–S22 Tables) led to similar classification of the remaining drugs but showed overall lower accuracy due to the misclassification of FEN, either as a counterion or as an antidepressant. Notably, the lack of effects with FEN cannot be explained by its size or log D (S6 Fig). FEN is an atypical non-benzodiazepine anxiolytic, which acts as an antagonist to the metabotropic glutamate receptor 5 (mGluR5) [53]. Planarians have glutamatergic neurons [54, 55] and a mGlu5 receptor homolog is present in the *D. japonica* transcriptome [29], suggesting this lack of activity is likely not due to lack of the target. However, FEN has been found to vary in bioavailability in humans [56, 57] and due to various side effects observed in clinical studies [58], is not currently used as an anxiolytic. Thus, the situation with FEN is complex and the lack of both acute and chronic effects with FEN suggest that uptake in planarians may be insufficient to cause effects. Notably, due to the poor water solubility of FEN, we could not test higher concentrations.

Many of the compounds studied here have previously been tested in a behavioral screen in 7-day old zebrafish larvae [20]. Similar to our findings for the classifications based on planarian data, the authors found that they could visually distinguish the three drug classes using multidimensional scaling, with considerable overlap of the clusters, particularly of anxiolytics

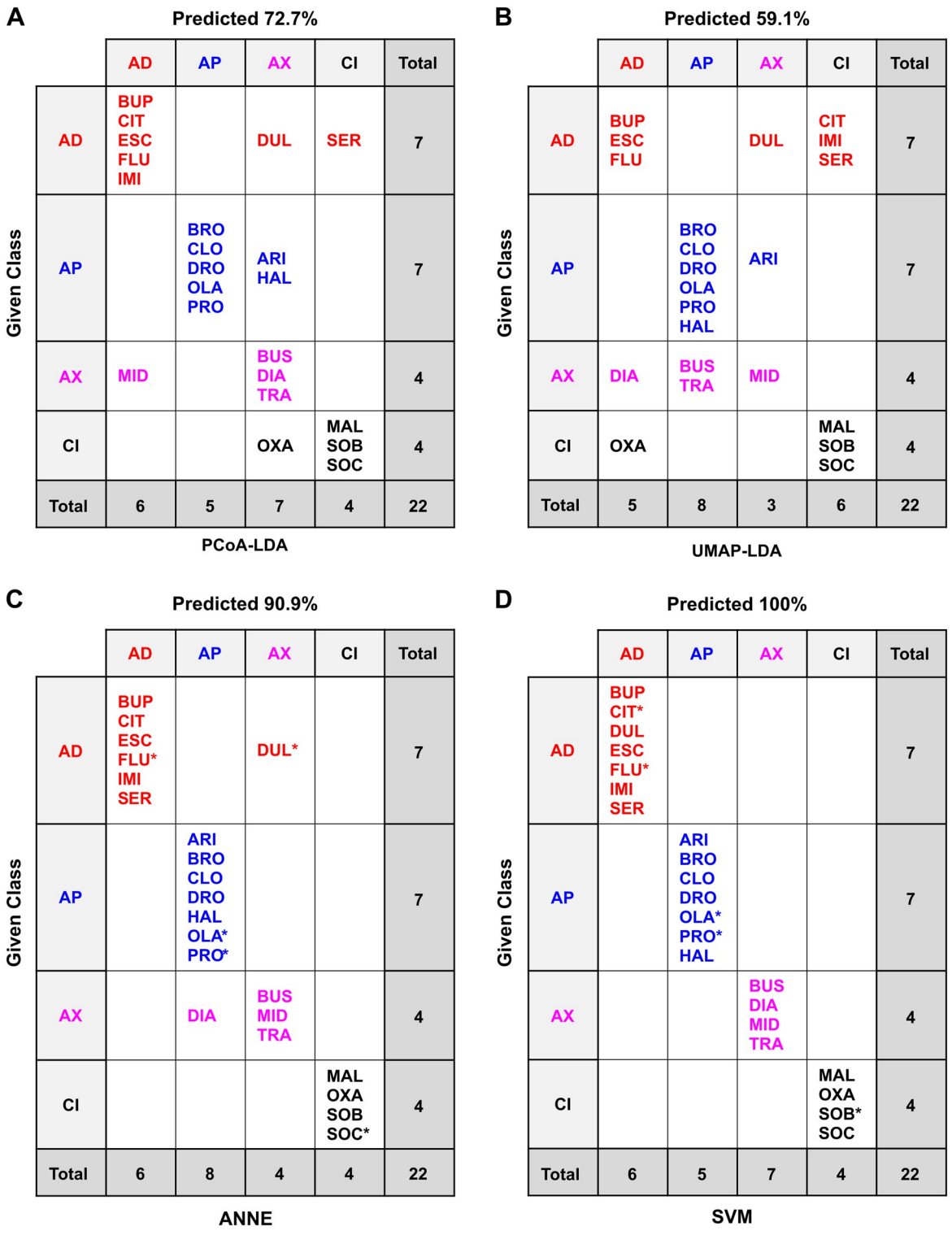

Fig 11. Confusion matrices of the classification methods based on planarian behavioral phenotyping of the drugs and counterions. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs. AD: antidepressant, AP: antipsychotic, AX: anxiolytic; CI: counterion. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. See the Methods for model selection. Test set accuracies were 80.0% for ANNE (S15 Table) and 100% for SVMs (S17 Table).

https://doi.org/10.1371/journal.pone.0315394.g011

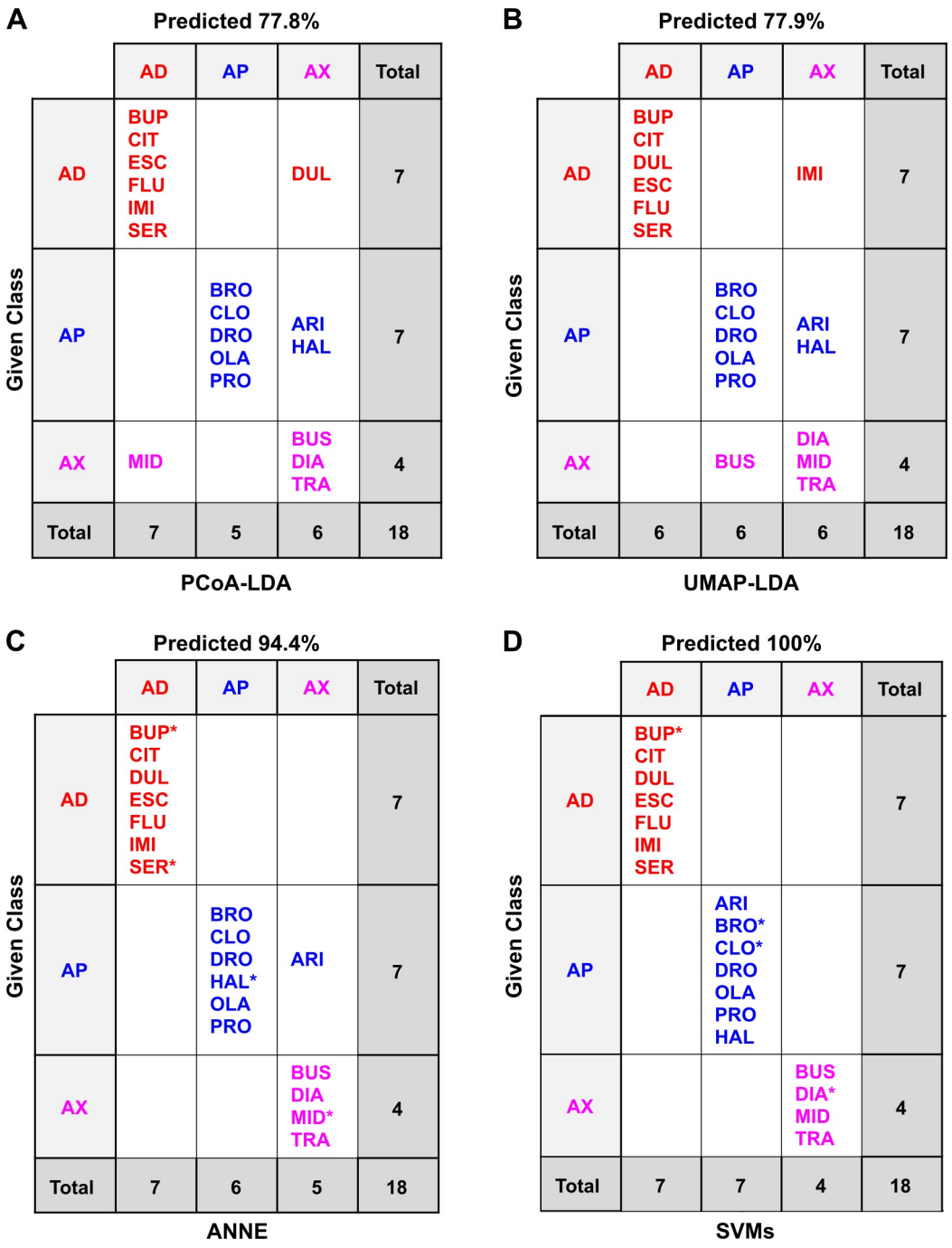

Fig 12. Confusion matrices of the classification methods based on planarian behavioral phenotyping of the drugs only. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs. AD: antidepressant, AP: antipsychotic, AX: anxiolytic; CI: counterion. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. See the Methods for model selection. Test set accuracies were 100% for ANNE (S16 Table) and 100% for SVMs (S18 Table).

https://doi.org/10.1371/journal.pone.0315394.g012

and antidepressants [20]. Given the lower complexity of planarians and their status as an invertebrate model [14], it was encouraging to find that we were able to perform effective behavioral classification schemes in a simple and cost-effective model.

Planarian phenotyping adds information to cheminformatics

The goal of this pilot study was to determine the extent to which planarian behavioral phenotyping could provide additional information for neuroactive drug classification that could not be captured with cheminformatics alone. The value of using a combined cheminformatics and organismal MTS approach has been demonstrated for developmental toxicity and neurotoxicity studies using zebrafish larvae [59]. Here, we show that planarian phenotypic data on neuroactive chemicals augments the cheminformatics data.

The classification of the chemicals based on the planarian behavioral phenotyping did not use any information about the chemicals. Thus, the fact that classification accuracy reached $> 70\%$ for all models (except for UMAP-LDA with counterions included) and $>90\%$ for some models for the planarian data suggests that behavioral barcodes contain meaningful information that reflects the underlying biology. Moreover, while PCoA-LDA and UMAP-LDA classification methods had higher accuracy when using chemical descriptors rather than planarian behavioral barcodes, the machine learning models had similarly high accuracies ($>90\%$) for both data streams. Thus, the choice of classification method can be equally as impactful as the choice of input data. Taken together, our data imply that planarian behavioral screening can identify and differentiate similarly acting compounds without knowledge of the underlying biology or physicochemical properties, making it well suited for unbiased, discovery-based screening of neuropsychiatric drug candidates.

When considering the primary classification indicated on the manufacturer's website, DUL and ARI were among the most misclassified compounds using behavioral responses across the different methods (Figs 11 and 12). However, both compounds have known dual functions. DUL affects both serotonin (5-hyroxytryptamine; 5HT) and noradrenaline reuptake [60], and while it is mainly classified as an antidepressant, it has been shown to have anxiolytic effects upon chronic exposure in mice [60, 61]. Across all classification methods except for SVMs, DUL was often classified as an anxiolytic. Similarly, ARI was originally classified as an atypical antipsychotic but is also used clinically to augment antidepressant efficacy [62, 63]. When using planarian phenotyping, ARI was often classified as either an antidepressant or anxiolytic. Thus, planarian phenotyping was able to detect known "off-label" effects of these drugs that could not be detected by cheminformatics alone. Moreover, planarian behavioral barcoding correctly identified BUP as an antidepressant in all models–in contrast to our findings with the physicochemical data. BUP is an aminoketone and its mode of action as a norepinephrine and dopamine reuptake inhibitor is distinct from that of the other antidepressants which are all selective serotonin reuptake inhibitors [64]. This knowledge of the *in vivo* effects can only be gained from an organismal systems-level approach, where the effects of these pathway interactions can be observed.

Study limitations and considerations for follow-up studies

This pilot study was small in scope (19 compounds) and faced limitations that need to be considered when designing follow-up studies with larger compound libraries to support or refute the findings reported here. First, we tested only 5 anxiolytics because most are controlled substances, and thus more expensive and difficult to obtain and use than other compounds. Furthermore, given that FEN was inactive in planarians up to the solubility limit, this reduced the anxiolytic category to only 4 active compounds for the planarian classifications compared to 7

compounds in each of the other two categories. While the numbers of compounds were still comparable here, class imbalance is something to consider when screening larger libraries [65, 66].

For planarian behavioral phenotyping, we created an aggregated barcode that included effects across all tested concentrations, thus potentially including toxic effects at higher concentrations for some chemicals. By looking at the whole concentration profile for a chemical, we anticipated finding phenotypic signatures for compounds that act similarly, but it is possible that the effects were not aligned by relative concentration or that toxicity of different chemicals can manifest in different ways and confound the classifications. Some zebrafish studies have overcome this issue by selecting single representative concentrations of a chemical to use as a comparator [20]. However, it is not always straightforward to determine which concentration is truly representative and reflective of substantial neuroefficacy without the confounders of toxicity.

We evaluated four computational models and found them to have comparable performance on our data, which made it difficult to decide on a "best" classification method, likely due to the small scope of this study. The machine learning models reached >90% for all cases, thus generally performed better than the LDA methods, but they differed in misclassifications and information content. The LDA methods provide insight into the similarity/distances between chemicals which allows for visualization of the classification (Figs 4 and 10) and are not provided by the ANNE and SVMs, which are more enigmatic in how the models create the classifications.

The next steps to further test these classification methods would be to screen larger libraries of neuroactive compounds that are commercially available, as for example in [20, 22]. Having a larger library would allow for a truly separate and larger test set of chemicals to evaluate model performance instead of the jackknifing approach conducted here. For these initial scaling-up efforts, it will be important to select chemical libraries that contain several members of the same category while also spanning different classes and to include chemicals with both known and unknown modes of action.

One consideration for testing larger libraries of 1000+ chemicals is feasibility. While small organism MTS is substantially faster than vertebrate testing it is still an experimental method and thus will be rate-limiting compared to a purely computational approach. To screen larger libraries, one could select one or two concentrations based on the tests conducted in this pilot study for an initial screen. Further screening for concentration response information could then be performed on promising candidates from the initial screen. This approach has been successfully used in zebrafish larvae to conduct such large library screens [20, 22]. Moreover, the increased incorporation of robotic handling systems will expand the throughput capabilities of small organismal screens. Upon generating larger datasets, the next step would be to integrate the two data streams used here (cheminformatics and planarian phenotyping) to generate a single classification. Other available data streams, such as those generated from *in vitro* HTS [67, 68] or computational modeling of drug-target interactions [69, 70], could also be integrated, similar to the Integrated Approaches to Testing and Assessment (IATA) framework used in toxicology to combine multiple data streams to assess chemical toxicity [71, 72]. Combining multiple data streams allows for a broader understanding of complex biological processes that are not well captured by single assays [73]. Moreover, findings can be validated across models, allowing for weight of evidence-based decision making. This implies that there must be a degree of inter-model information redundancy while simultaneously having each model bring in complementary information, such that the information content from the test battery is more than the sum of its parts.

A few recent studies have implemented such a synthesis based on different types of information, e.g., cheminformatics, and/or HTS data, including *in vitro* cell painting and gene expression profiles, e.g. [74, 75]. To the best of our knowledge, such an integrative approach including MTS behavioral data has not yet been attempted. How heterogeneous data streams are combined and weighed is a non-trivial decision [73] that will itself require validation with known chemical libraries. Thus, there are multiple steps to take and decisions to make that need to be carefully considered and scrutinized to build upon the results from this pilot study. The important contribution of this work is that it has shown the feasibility and promise of invertebrate behavioral MTS as a viable experimental alternative to vertebrate testing for first-tier screening of novel drug candidates.

Materials and methods

Chemicals and chemical structures

Neuroactive compounds that have been functionally classified as antidepressants, antipsychotics or anxiolytics were studied. Functional classifications were assigned based on the primary classification indicated on the manufacturer's website and/or previous literature [20]. Table 1 lists the chemicals studied including the chemical abstracts service (CAS) numbers and information pertaining to the use of the compounds in the experimental studies. Some chemicals were provided in salt form; thus, we also tested the respective counterions alone (Table 1).

For computational models based on two-dimensional (2D) chemical structures of drugs and counterions, 2D coordinates were downloaded as structure-data format (SDF) files from PubChem (https://pubchem.ncbi.nlm.nih.gov/) and concatenated into a single 2D SDF file using OpenBabel 3.1.1 (https://github.com/openbabel/openbabel) [76] for Linux (Linux Mint 21.3 Xfce, https://linuxmint.com/). For computational models based on three-dimensional (3D) chemical structures of drugs and counterions, 3D coordinates were downloaded as SDF files from PubChem, energy-minimized in the AMBER14 force field at pH 7.40 [77, 78] using YASARA-Structure 23.12.24 [79] for Linux and exported as a single concatenated 3D SDF file. Version 2000 was used for both 2D and 3D SDF files. The ionization state of each compound at pH 7.40 was confirmed using the pKa module of Marvin Sketch 23.1.0 for Linux (https://www.chemaxon.com). Structures, names, 3-letter codes, and primary functional pharmacological classes of the drugs and counterions used in the computational studies are shown in Fig 1.

Chemical similarity

Chemical similarity of the 3D SDF structures of drugs and counterions relative to ARI as a reference structure was assessed using vROCS 3.6.1.3 (OpenEye, Cadence Molecular Sciences, Santa Fe, NM, https://www.eyesopen.com; [69]) for Linux. The choice of a reference compound is arbitrary. However, given that the algorithm was based on computing the volume of overlap between each molecule and the reference compound after a 3D alignment of the two structures, we chose to use the reference compound with the largest molecular volume as assessed by the van der Waals radii of the constituent covalently bonded atoms. Similarity was expressed quantitatively in terms of the Tanimoto coefficient (Tc), which is the most widely used similarity metric in cheminformatics [80, 81]. The program calculates three Tc values: Tc Shape, based on a 3D alignment of two molecules that maximizes the volume of overlap; Tc Color, derived from a 3D alignment of two molecules with respect to six molecular features (H-bond donor, H-bond acceptor, anion, cation, hydrophobe, and ring); and Tc Combined, the sum of Tc Shape and Tc Color. The numerical values of Tc Shape or Tc Color range from 0 (no similarity) to 1 (complete similarity); therefore, Tc Combined ranges from 0 (no similarity) to 2 (complete similarity). Note that two identical molecules would have Tc Shape and Tc

Color coefficients of 1. However, if two structures have Tc Shape and Tc Color coefficients of 1, this does not necessarily mean that the molecules are identical because two non-identical compounds could have distinctive features that were not accounted for in the similarity algorithms [82].

Chemical descriptors

Chemical descriptors (comprising computed chemical properties and fingerprints) were calculated for 2D and 3D structures of drugs and counterions using absorption, distribution, metabolism, excretion, and toxicity (ADMET) Predictor® 11.0.3 (Simulations Plus, Lancaster, CA) for Windows (10 Pro 22H2) (S1 and S2 Files). The fingerprint type was extended connectivity fingerprint diameter 6 (ECFP6). Within ADMET Predictor®, fingerprints were converted to descriptor format and combined with the other computed chemical properties descriptors for use in the computational models. For each of the 2D structures (18 drugs and 5 counterions) there were 908 descriptors (514 fingerprints and 394 directly calculated properties), and for each of the 3D structures (21 drugs and 5 counterions) there were 948 descriptors (516 fingerprints and 432 directly calculated properties). Importantly, when considering 2D structures CIT and ESC are treated as one structure which we label as CIT, leading to only 18 drugs. Because 3 of the 18 drugs were administered as racemic mixtures in the behavioral assays, the separate structures of the (*R*) and (*S*) stereoisomers of these compounds were included in the 3D computational models, thereby giving rise to 21 3D drug structures (and 5 counterions). The 2D and 3D descriptors were exported from ADMET Predictor® as two separate data files in CSV format for use in classification models. To keep the data size comparable to the planarian data, which considers each of the 19 tested compounds as one entity, we focused on the 2D structural features in the main text.

Classification models

Linear discriminant analysis (LDA) preceded by principal coordinate analysis (PCoA). LDA is a supervised linear ordination and classification technique that maximizes the discrimination between classes of labeled data [83]. LDA is a preferred classification procedure when there are more than two classes, in which cases the method is also known as canonical variate analysis (CVA) or canonical discriminant analysis (CDA). When the number of predictor variables is greater than the number of samples, LDA is preceded by a dimension reduction technique such as principal component analysis (PCA) or PCoA [84].

We chose PCoA as the dimension reduction step because PCoA can use any desired distance or dissimilarity measure, although Gower distance tends to be preferred [85], whereas PCA is restricted to Euclidean distance. Moreover, PCoA has been combined with LDA to create a successful sequential method for ordination and classification of standardized data called canonical analysis of principal coordinates (CAP) or generalized discriminant analysis (GDA) that preserves the initially chosen distance or dissimilarity measure in the resulting ordination [86].

The procedures for PCoA-LDA were carried out as follows: Either the chemical descriptor files or the barcode files of planarian behavioral responses were imported into the data analysis program, PAST 4.16c for Windows (https://www.nhm.uio.no/english/research/resources/past/; [87, 88]). Classifications based on 2D or 3D chemical descriptors were carried out with or without the counterions. Classifications based on planarian responses were conducted on data with or without FEN and with or without counterions. Columns that were deemed uninformative by the PAST software (non-numeric, all-zero, all-missing, singleton, and constant) were removed, and the remaining data were standardized so that each column had mean = 0

and SD = 1. Following dimension reduction in PAST using PCoA with Gower distances [89–91], LDA was carried out in PAST using the number of PCoA coordinates that yielded the highest classification accuracy (expressed as the percentage of correctly classified chemicals) after applying an exhaustive leave-one-out procedure (jackknifing) [92, 93]. Clusters identified by LDA were displayed as 2D scatterplots with the percentage of the total eigenvalues shown on each of the two axes and the clusters demarcated as 95% confidence ellipses. When the number of PCoA coordinates used for LDA was greater than two, the sum of the percentages of the total eigenvalues on the two axes was less than 100%.

LDA preceded by uniform manifold approximation and projection (UMAP). UMAP was carried out on standardized data files based on chemical descriptors or planarian behavioral responses in PAST 4.16c using Gower distances as described above for PCoA. The UMAP algorithm employed by PAST was derived from the original method of [94]. UMAP is a nonlinear dimension reduction and ordination method that has capabilities for preserving both the local and global structures of the original data, as an alternative dimension reduction method prior to LDA. UMAP provides a complementary alternative to PCoA, which is a linear dimension reduction and ordination technique that emphasizes preservation of the global structure of the original data [95].

The numbers of embedded and UMAP neighbors, as well as the minimum distance between samples, were systematically varied to achieve the maximum separation of clusters after 100 UMAP iterations. Final UMAP coordinates were then used as input for running LDA in PAST as described above for PCoA-LDA. The resulting LDA output included 2D scatterplots with 95% confidence ellipses and percent classification accuracies (unadjusted and jackknifed).

Artificial neural network ensemble (ANNE). An artificial neural network (ANN) is a mathematical construct that mimics a simple array of biological neurons consisting of an input layer, a hidden layer, and an output layer [96]. Inputs with various weights are transmitted to neurons in the hidden layer, where the summed weighted inputs are compared with a threshold value generated by a sigmoid activation function. When the threshold value is exceeded, a signal is conveyed to the output layer. ANNs have the capacity to accept nonlinear and large numbers of inputs from which they can adapt and learn, thereby enabling them to make predictions in the form of continuous outputs (regression models) or discrete assignments to two or more categories (classification models) [97]. Multiple ANNs can be combined into an ensemble (ANNE) so that their outputs are channeled to a voting or averaging algorithm to create an ensembled output, resulting in enhanced performance of ANN-based regression and correlation models [98].

ANNE classification was performed using the Modeler™ 11.0 module of ADMET Predictor® 11.0.3, with CLASS as the dependent variable. The classes consisted of the four primary functional categories (antidepressants, antipsychotics, anxiolytics, and counterions). Ten final models were created for each of the following four sets of chemical descriptors: 2D descriptors with or without counterions; and 3D descriptors with or without counterions. Likewise, ten final models were created for each of the following four sets of planarian behavioral data: including FEN with or without counterions; and excluding FEN with or without counterions.

Before generating the models, the program screened all the descriptors to remove those with the following characteristics: identical or with coefficients of variation $\leq 1\%$; under-represented (non-zero for 1–3 data points); or highly correlated (Pearson $r \geq 0.98$). For chemical descriptors, the culling process reduced the number of candidate 2D descriptors from 908 to 162 and the number of candidate 3D descriptors from 948 to 192. During the running of the ANNE and SVMs models, the algorithms selected the smallest number of descriptors that

optimized performance; for some ANNE models, the final number of descriptors was as low as one, and for some SVMs models, the final number of descriptors was as low as two. S1 Table is a compilation of all of the chemical descriptors that were used in constructing the ANNE and SVM models listed in S2–S9 Tables. For planarian behavioral descriptors, the culling process reduced the number of candidate descriptors from 87 to 80 (FEN with counterions), 87 to 81 (FEN without counterions), 87 to 80 (without FEN with counterions), and 87 to 81 (without FEN and without counterions).

The training:test ratios of chemicals for chemical descriptor files were as follows: 2D with counterions, 18:5; 2D without counterions, 14:4; 3D with counterions, 21:5; and 3D without counterions, 17:4. These ratios for planarian behavioral files were as follows: FEN with counterions, 18:5; FEN without counterions, 15:4; without FEN with counterions, 17:5; without FEN without counterions, 14:4. Test set selection was done via randomized stratified sampling by CLASS. During each run, the members of the ensemble automatically were selected by initially assigning random weights to each submodel and partitioning the training sets into training and verification sets in an approximate 2:1 ratio. After training, scores were assigned to each model according to the summed verification set performance and the absolute difference between verification and training set performance using the Youden index as the criterion. To avoid overtraining of ANN models, training was automatically halted when the verification set score failed to improve or increased beyond a preset number of iterations.

Generation of each of the 10 final models was initiated using a unique random seed number. Final descriptor selection for each model was carried out by the software using a genetic algorithm to explore the effectiveness of different descriptor combinations. Models were then run using default settings for the number of generations, number of neurons, and number of descriptors; these settings were automatically adjusted to be appropriate for the number of chemicals in the training and test sets.

For each run, the following performance metrics generalized to k classes and expressed as percentages were calculated: Youden index (J) [99], Matthews correlation coefficient (MCC or R_k) [100], and Accuracy (Acc) (overall percent correctly classified) [65]. These metrics were calculated by the Modeler™ 11 software for the training set, test set, and all chemicals. The data summaries include the following information for each model: mean and SE (n = 10 models) for each of the 9 performance indicators; number of neurons and descriptors; and the numbers and identities of misclassified chemicals. The rank for each ANNE and SVMs model, respectively, was determined by applying the RANK.AVG function in Microsoft Excel 365 to SUM (training metrics + test metrics + $(100\times N_{min}/N)$ + $(100\times D_{min})/D)$), where N_{min} = minimum number of neurons, N = number of neurons, D_{min} = minimum number of descriptors, and D = number of descriptors. Thus, the ranking rewarded models that produced high performance with the smallest numbers of neurons and/or the smallest numbers of descriptors, in keeping with the parsimony principle [101, 102]

Support vector machines (SVMs). This is a machine-learning technique that seeks to define hyperplanes in n-dimensional space that maximize the separation of data points into classes [103]. The name of the method stems from the fact that the points closest to a given hyperplane are called support vectors [104]. Although SVMs can be used for regression models and outlier detection, they are especially well suited for classification [105]. SVMs are intrinsically a binary classifier, but they have been successfully adapted for classifications involving multiple classes.

SVMs classification was performed using the Modeler™ 11.0 module of ADMET Predictor® 11.0.3, with CLASS as the dependent variable, as described above for ANNE classification. Likewise, the same evaluation statistics and data summaries employed by ANNE were used by SVMs.

The default correlation setting for our machine-learning models was ≥ 0.98. Highly correlated variables were automatically removed for both the physicochemical descriptors (including the molecular fingerprints) and the behavioral endpoints.

Planarian care and culture

D. japonica planarians from an established lab culture were used for all experiments and cultivated according to standard protocols [44, 47]. The planarians were kept in 0.5 g/L Instant Ocean (IO) Salts (Spectrum Brands, Blacksburg, VA, USA) in BPA-free polypropylene plastic containers (approximately 25 cm L x 14 cm L x 8 cm H), with the lid on loosely and stored at 20˚C in a Panasonic refrigerated incubator in the dark when not used for experiments. The planarians were fed 1-2x per week with 100% grass-fed beef liver (purchased from a local farm) or USDA-certified organic chicken liver (Bell and Evans) and cleaned on the feeding day and 2 days later. When not fed, the containers were cleaned weekly. Similarly sized intact planarians that were fasted for 5–7 days were arbitrarily selected to be used in experiments, as in [48].

Chemical exposure

Stock solutions were prepared at 200X of the highest test concentration in either DMSO (Sigma-Aldrich, purity = 99.9%) or Milli-Q water, depending on solubility (Table 1). The 200X stocks were diluted to 10X stocks in IO water just prior to exposure. Planarians were exposed to the chemicals in 48-well tissue culture-treated polystyrene plates (Genesee Scientific, San Diego, CA, USA), with each well containing 1 planarian in 200 μL of chemical solution. When used as a solvent, DMSO was used at a final concentration of 0.5% (v/v) in all test concentrations, which does not cause morphological or behavioral effects in *D. japonica* [43].

For most chemicals, serial half-log dilutions were used to prepare the range of concentrations tested, which were initially guided by previous results in developing zebrafish [20] or our previous studies with *D. japonica* [106]. For FEN and MID, serial quarter-log dilutions were used because preliminary tests had already narrowed an appropriate concentration range to span no effects to strong effects (or maximum solubility). All chemicals were screened over at least 5 concentrations, with each row of the 48-well plate consisting of one concentration ($n = 8$). For some compounds, additional lower concentrations were screened if the original lowest concentration showed effects in the preliminary data analysis. Moreover, some compounds were screened at additional higher concentrations if no statistically significant effects were observed in the original range and the compound was still soluble at higher concentrations. The counterions were screened at the concentration equivalent to what would be found in the highest concentration of the respective salt form of a tested drug. The pH of the highest test concentration for each chemical was first measured with pH strips (VWR, Radnor, PA). As we found the measurements with the pH strips were not precise nor reliable, we remeasured the pH using an Apera PH60-MS pH Tester kit for small volumes (Apera Instruments, Columbus, Ohio), taking measurements after allowing the probe to equilibrate for 20 min.

Plates were sealed with thermal film (Excel Scientific, Victorville, CA, USA) immediately after addition of chemicals. Experiments were run on intact planarians within 15 minutes– 3 hours of chemical introduction. This time window was chosen for throughput, to allow for multiple plates to be set up and tested in one experiment. Which plates were screened first was arbitrarily decided, to avoid any systematic bias in the treatment. Technical triplicates were run for all chemicals (total $n = 24$ per concentration), employing a rotating orientation of the chemical concentrations in the plate rows (Fig 7A) to account for edge effects when screening [44].

One screening plate consisted of negative and positive assay controls, each run at a single concentration, which were used to evaluate proper performance of the assays. Negative controls consisted of D-sorbitol and L-ascorbic acid (both from Sigma-Aldrich) at 100 μM [44]. Ethanol (1% v/v, Greenfield Global, Toronto, Canada), DMSO (3% v/v), and sodium dodecyl sulfate (SDS, 1 mg/L, Life Technologies, Carlsbad, CA) were used as positive controls at concentrations that induce behavioral phenotypes in the absence of lethality in *D. japonica* [43]. The positive controls were prepared fresh at 10x in IO water on the day of the experiment. All assay controls performed as expected (S23 Table).

Planarian screening methodology

Intact planarians were assayed for acute effects on lethality/crawl-out behavior, morphology, stickiness, and various behaviors (locomotion, phototaxis, and noxious heat sensing) using the screening platform and analysis methodology described in [44, 47, 48, 50]. Screening consisted of three assays: Phototaxis, Stickiness and Scrunching (Fig 7A), which have been described in detail in [47]. Image analysis was performed using object tracking [42] in MATLAB (MathWorks, Natick, MA) or Python as described in [44, 47, 48, 50]. Because the liquid volume only takes up 13% of the total well volume, planarians can crawl out of the liquid and dry out. Planarians which crawled out of the water (crawl-out, also previously referred to as "suicide" [44]) were excluded from the analysis of all subsequent endpoints. Worms were included in the analysis for endpoints in which they were still alive and in the exposure solution.

The endpoints used are described in Tables 2 and 3 and consisted of two major classes: unstimulated and stimulated behaviors. Unstimulated behaviors included lethality/crawl-out, presence of abnormal body shapes/behaviors, and general locomotion (speed and resting in the dark, locomotor bursts and anxiety). The "anxiety" endpoint [50, 106] measures the amount of time spent in the outer portion of the well by the planarians [48] and is akin to similar measures in rodents [107] and adult zebrafish [108], wherein spatial exploration is used to describe "anxiety-like" behaviors. It is also referred to as "wall preference" in the literature [48, 109]. Specific distinctive body shape categories (contraction, C-shape, corkscrew, pharynx extrusion, and hyperkinesis) were manually scored blind by a researcher (Fig 7B–7G). Contraction, C-shape, and corkscrew were scored as described in [106]. Hyperkinesis consisted of an array of different types of generally hyperkinetic movements that could not be captured by the other categories, such as muscle waves/scrunching [106], hyperextension, head flailing and convulsive behavior, consisting of uncoordinated muscle twitching (S1 Video). A single planarian could be scored as exhibiting up to three different shape categories. For benchmark concentration (BMC) modeling, described below, the presence of any abnormal body shape was used to determine activity, whereas for phenotypic profiling, both incidence rate of any body shape and the incidence rates in the specific shape classes were used. Stimulated behaviors were measured in response to an environmental stimulus (shaking, light, or noxious heat, Fig 7).

Benchmark concentration modeling

Benchmark concentrations (BMCs) were calculated for every chemical and endpoint to quantify potency using the Rcurvep R package [110] similar to the procedure described in [48]. Briefly, the planarian responses for each endpoint were transformed as needed to allow for determination of directional, concentration-dependent responses, as described in [48]. The transformed data for each individual planarian is provided in S3 File. For the binary endpoints (crawl-out, body shape, stickiness, phototaxis, scrunching), the incidence rates (number of planarians affected and total number of planarians) from the combined data from all replicates

(n = 24) was used. For all binary endpoints except crawl-out, the experimental incidence rates for each concentration were normalized by the incidence number of the respective in-plate vehicle controls. Any negative incidence numbers after normalization were set to 0. For continuous endpoints, the raw response of each individual planarian was normalized either by dividing by or subtracting the median of vehicle control values for that plate (Table 3). Except for locomotor bursts (total), the normalized outcome measures were multiplied by 100 to represent the percent change from the control populations and to provide an appropriate range to perform the BMC analysis. The normalized data were used as input for the Rcurvep package to calculate the benchmark response (BMR) as described in [48]. For some endpoints, the R package could not converge to produce an accurate BMR due to high variance and/or non-monotonic concentration responses. In these cases, the BMR was set to the 5^{th} or 95^{th} percentile of the normalized response of the aggregated vehicle controls (S4 Fig), to set the threshold above which effects would be seen. The determined BMR was then used to calculate the BMC, the concentration that exceeds that BMR in the modeled concentration response curve [110], for each endpoint, using n = 1000 bootstrapped curves. For all endpoints, we report the resulting median BMCs from this bootstrapped analysis. The lower and upper limits (5th and 95th percentiles, respectively) of the BMC for each endpoint are listed in S4 File. Some endpoints can be affected in both directions (e.g., increases or decreases in speed). For these endpoints, BMRs and BMCs were calculated for each direction, but we found that hits were only determined in one direction and thus only report that direction (Table 3).

Phenotypic profiling

To quantitively describe the multidimensional phenotype observed for each test concentration, a "phenotypic barcode" was created for each chemical concentration consisting of either the incidence rate for all binary endpoints, multiplied by 100, or the median normalized response for all continuous endpoints, as were input into the BMC analysis. Next, one master barcode was created for each chemical by concatenating the barcodes of all tested concentrations in relative order (S5 File). To keep comparisons similar across different test concentrations, the relative order of the concentrations was considered rather than the absolute concentration. Because different numbers of concentrations were tested across the different chemicals, we "right-aligned" the barcodes such that the highest test concentrations were aligned across all chemicals. For lower concentrations that were not tested in a specific chemical, the barcodes were filled with 0s (indicative of no effects). When being input into the classification methods, the data were truncated to only the highest 5 concentrations such that negative/missing data were not driving the classifications (S7 Fig).

Supporting information

S1 Fig. Classification based on 3D chemical descriptors for LDA based approaches. Computational methods are arranged by rows: (A) PCoA-LDA; (B) UMAP-LDA. Counterion inclusion/exclusion is arranged by columns: (1) with counterions; (2) without counterions. Misclassifications after jackknifing: (A) none; (B) PRO, BUS; (C) ARI, BUS, TRA; (D) ARI, BUS, TRA. Ellipses refer to 95% confidence intervals. The axes show the percentage of the total eigenvalues; this does not sum to 100% in (A1) because there was a third axis that accounted for the remaining 0.60%.
(TIF)

S2 Fig. Confusion matrices of the classification methods based on 3D chemical descriptors of the drugs and counterions. Confusion matrices for the different classification methods:

(A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVM; AD: antidepressant, AP: antipsychotic, AX: anxiolytic; CI: counterion. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. In (C), model 05_1n7, marked with *, was ranked first but did not include a member of the anxiolytic class in the test set. Thus, we also included the second-place model (model 06_1n8), marked with +, which included at least one member from each of the 4 classes. In (D), four models were tied for the best model but for clarity only the first of the 4 tied models is indicated because there were no misclassifications of any test set members. The members of the 4 test sets for the 4 tied first-place models were as follows: 01_2i: CIR, FLR, DRO, DIA, BRD; 05_21: ESC, FLS, BRO, MID, MAL; 06_2i: BUA, IMI, DRO, PRO, SOD; 08_21: BUA, FLS, BRO, FEN, MAL. Test set accuracies were 100% for ANNE (S6 Table) and 100% for SVMs (S8 Table).
(TIF)

S3 Fig. Confusion matrices of classification methods based on 3D chemical descriptors without counterions. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVM; AD: antidepressant, AP: antipsychotic, AX: anxiolytic. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicted accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. In (C), model 03_1n7, marked with *, was ranked first but did not include a member of the anxiolytic class in the test set. Thus, we also included one of the second-place models (model04_2n2), marked with +, which included at least one member from each of the 4 classes. 04_2n2 was tied for second place with 09_2n2, which had the following test set members: BUA, FLS, OLA, MID. For clarity, we only marked model 04_2n2 in (C). In (D), three models were tied for the best model but for clarity only the first of the 3 tied models is indicated because there were no misclassifications of any test set members. Test set accuracies were 100% for ANNE (S7 Table) and 100% for SVMs (S9 Table).
(TIF)

S4 Fig. Normalized responses of the vehicle controls in the continuous endpoints. Plots show the distribution of normalized responses for each individual vehicle control (n = 768) when normalized by the median of the control population of the respective plate.
(TIF)

S5 Fig. FEN does not cause chronic toxicity. Minimum intensity projections showing the behavior of planarians exposed to either 0.5% DMSO (solvent control) or 562 μM FEN for 12 days over the course of the 5 min phototaxis assay. Images show representative worms out of a total of $n = 32$ per condition. Scale bar: 4 mm.
(TIF)

S6 Fig. LogD cannot explain potency differences. The range of logD values (error bars) for a given chemical were calculated based on the logD at the pH of the solvent and at the measured pH of the highest tested concentration (see S12 Table) to cover the suspected pH for all test concentrations and plotted against the log10 of the most sensitive benchmark concentration (BMC) in μM for that chemical. Markers were added at the median of the logD range to allow for labeling by chemical. N/A indicates that fenobam (FEN) was inactive as this did not have a calculated BMC.
(TIF)

S7 Fig. Heatmap of planarian barcodes used as input for classification methods. Each chemical was assigned a master barcode by concatenating the barcodes of each concentration.

Only the 5 highest concentrations (relative concentrations 8–12) were used for the classification methods so that the inactive/missing data would not dominate the phenotypes. Columns are labeled as "endpoint_relative concentration". CRO: crawl-out, STK: stickiness, SHP: body shape, SCR: scrunching, PTX: phototaxis, ANX: anxiety, RSD: resting_dark, RSB: resting_blue, SPD: speed_dark, SB1: speed_blue1, SB2: speed_blue2, LBT: locomotor bursts_total; NSS: noxious stimuli_strength. Chemicals are colored by class: antidepressants (red), antipsychotics (blue), anxiolytics (magenta), and counterions (black).
(TIF)

S8 Fig. Classification of planarian data including FEN. Computational methods are arranged by rows: (A) PCoA-LDA; (B) UMAP-LDA. Counterion inclusion/exclusion is arranged by columns: (1) with counterions; (2) without counterions. Ellipses show 95% confidence intervals. The axes show the percentage of the total eigenvalues.
(TIF)

S9 Fig. Confusion matrices of the classification methods based on planarian behavioral phenotyping of the drugs, including FEN and counterions. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs. AD: antidepressant, AP: antipsychotic, AX: anxiolytic; CI: counterion. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. Test set accuracies were 100% for ANNE (S19 Table) and 80% for SVMs (S21 Table).
(TIF)

S10 Fig. Confusion matrices of classification of planarian data including FEN without counterions. Confusion matrices for the different classification methods: (A) PCoA-LDA; (B) UMAP-LDA; (C) ANNE; (D) SVMs. AD: antidepressant, AP: antipsychotic, AX: anxiolytic. In A and B, predicted accuracy was calculated following an exhaustive jackknifing. In C and D, predicated accuracy refers to the overall accuracy. *indicates randomly chosen members of the test set. Test set accuracies were 100% for ANNE (S20 Table) and 100% for SVMs (S22 Table).
(TIF)

S1 Table. Chemical descriptors used in the machine learning models (ANNE and SVMs).
(PDF)

S2 Table. ANNE classification models using 2D molecular descriptors of 18 drugs and 5 counterions.
(PDF)

S3 Table. ANNE classification models using 2D molecular descriptors of 18 drugs.
(PDF)

S4 Table. SVMs classification models using 2D molecular descriptors of 18 drugs and 5 counterions.
(PDF)

S5 Table. SVMs classification models using 2D molecular descriptors of 18 drugs.
(PDF)

S6 Table. ANNE classification models using 3D molecular descriptors of 21 drugs and 5 counterions.
(PDF)

S7 Table. ANNE classification models using 3D molecular descriptors of 21 drugs.
(PDF)

S8 Table. SVMs classification models using 3D molecular descriptors of 21 drugs and 5 counterions.
(PDF)

S9 Table. SVMs classification models using 3D molecular descriptors of drugs only.
(PDF)

S10 Table. Calculated benchmark concentration values in logM.
(PDF)

S11 Table. Converted BMC values in μM.
(PDF)

S12 Table. Median normalized responses for each endpoint after exposure to the counterions.
(PDF)

S13 Table. pH at highest tested concentration.
(PDF)

S14 Table. Median normalized responses for extreme pHs.
(PDF)

S15 Table. ANNE classification models using behavioral responses to 18 drugs (-FEN) and 4 counterions.
(PDF)

S16 Table. ANNE classification models using behavioral responses to 18 drugs (-FEN).
(PDF)

S17 Table. SVMs classification models using behavioral responses to 18 drugs (-FEN) and 4 counterions.
(PDF)

S18 Table. SVMs classification models using behavioral responses to 18 drugs (-FEN).
(PDF)

S19 Table. ANNE classification models using behavioral responses to 19 drugs (+FEN) and 4 counterions.
(PDF)

S20 Table. ANNE classification models using behavioral responses to 19 drugs (+FEN).
(PDF)

S21 Table. SVMs classification models using behavioral responses to 19 drugs (+FEN) and 4 counterions.
(PDF)

S22 Table. SVMs classification models using behavioral responses to 19 drugs (+FEN).
(PDF)

S23 Table. Compiled normalized responses for assay controls.
(PDF)

S1 Video. Example planarian behaviors showing the normal gliding behavior compared to two examples of hyperactive behavior.
(AVI)

S1 File. 2D chemical descriptors for drugs and counterions. Raw 2D molecular properties exported from Simulations Plus ADMET Predictor 11 before pre-processing (see Methods).
(CSV)

S2 File. 3D chemical descriptors for drugs and counterions. Raw 3D molecular properties exported from Simulations Plus ADMET Predictor 11 before pre-processing (see Methods).
(CSV)

S3 File. Individual planarian behavioral responses. Responses for continuous endpoints are normalized by the response of in plate controls.
(XLSX)

S4 File. 95th percentile confidence intervals of the benchmark concentration (in logM) for each planarian endpoint.
(XLSX)

S5 File. Concatenated planarian behavioral barcodes. Input data for classification models.
(XLSX)

Acknowledgments

The authors thank Dr. Siqi Zhang for helping with data analysis. The authors thank Dr. Michael Lawless (Simulations Plus) for helpful discussions about the use of ANNE and SVMs models and their statistical evaluation and Prof. Øyvind Hammer (University of Oslo, Norway) for helpful discussions of PCoA-LDA.

Author Contributions

Conceptualization: Danielle Ireland, Rudy J. Richardson, Eva-Maria S. Collins.

Formal analysis: Danielle Ireland, Sagar Rao, Rudy J. Richardson.

Investigation: Danielle Ireland, Christina Rabeler, Rudy J. Richardson, Eva-Maria S. Collins.

Methodology: Danielle Ireland, Rudy J. Richardson, Eva-Maria S. Collins.

Project administration: Eva-Maria S. Collins.

Resources: Eva-Maria S. Collins.

Software: Danielle Ireland, Rudy J. Richardson.

Supervision: Danielle Ireland, Eva-Maria S. Collins.

Visualization: Danielle Ireland, Rudy J. Richardson, Eva-Maria S. Collins.

Writing – original draft: Danielle Ireland, Rudy J. Richardson, Eva-Maria S. Collins.

Writing – review & editing: Danielle Ireland, Christina Rabeler, Sagar Rao, Rudy J. Richardson, Eva-Maria S. Collins.

References

1. World Health Organization. Mental disorders [Internet]. 2022 [cited 2024 Apr 17]. https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

2. Council of Economic Advisers. Reducing the Economic Burden of Unmet Mental Health Needs. 2022 May.
3. World Health Organization. Mental Health and COVID-19: Early evidence of the pandemic ' s impact. Sci Br. 2022; 2(March):1–11.
4. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders. EMBO Rep. 2016 Sep; 17(9):1245–9.
5. Doran CM, Kinchin I. A review of the economic impact of mental illness. Aust Heal Rev. 2019 Nov 13; 43(1):43–8. https://doi.org/10.1071/AH16115 PMID: 29129189
6. Kas MJ, Penninx B, Sommer B, Serretti A, Arango C, Marston H. A quantitative approach to neuropsychiatry: The why and the how. Neurosci Biobehav Rev. 2019 Feb 1; 97:3–9. https://doi.org/10.1016/j.neubiorev.2017.12.008 PMID: 29246661
7. Berk M. Pathways to new drug discovery in neuropsychiatry. BMC Med. 2012 Nov 29; 10(1):1–4. https://doi.org/10.1186/1741-7015-10-151 PMID: 23194414
8. Wong DT, Perry KW, Bymaster FP. The Discovery of Fluoxetine Hydrochloride (Prozac). Nat Rev Drug Discov 2005 49. 2005 Aug 24; 4(9):764–74.
9. Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, et al. Automated design of ligands to polypharmacological profiles. Nature. 2012 Dec; 492(7428):215–20. https://doi.org/10.1038/nature11691 PMID: 23235874
10. Kokel D, Peterson RT. Chemobehavioural phenomics and behaviour-based psychiatric drug discovery in the zebrafish. Briefings Funct Genomics Proteomics. 2008; 7(6):483–90. https://doi.org/10.1093/bfgp/eln040 PMID: 18784194
11. Blokhin IO, Khorkova O, Saveanu R V., Wahlestedt C. Molecular mechanisms of psychiatric diseases. Neurobiol Dis. 2020 Dec 1; 146:105136.
12. Miller AH, Raison CL. Burning down the house: reinventing drug discovery in psychiatry for the development of targeted therapies. Mol Psychiatry 2022 281. 2022 Dec 2; 28(1):68–75. https://doi.org/10.1038/s41380-022-01887-y PMID: 36460725
13. Hartung T. The (misleading) role of animal models in drug development. Front Drug Discov. 2024 Apr 8; 4:1355044.
14. Collins EMS, Hessel EVS, Hughes S. How neurobehavior and brain development in alternative whole-organism models can contribute to prediction of developmental neurotoxicity. Neurotoxicology. 2024; 102:48–57. https://doi.org/10.1016/j.neuro.2024.03.005 PMID: 38552718
15. Henry J, Wlodkowic D. Towards High-Throughput Chemobehavioural Phenomics in Neuropsychiatric Drug Discovery. Mar Drugs. 2019 Jun 6; 17(6):340. https://doi.org/10.3390/md17060340 PMID: 31174272
16. Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: recent successes, lessons learned and new directions. Nat Rev Drug Discov. 2022; 21(12):899–914. https://doi.org/10.1038/s41573-022-00472-w PMID: 35637317
17. McCarroll MN, Gendelev L, Keiser MJ, Kokel D. Leveraging large-scale behavioral profiling in zebrafish to explore neuroactive polypharmacology. ACS Chem Biol. 2016 Apr 4; 11(4):842. https://doi.org/10.1021/acschembio.5b00800 PMID: 26845413
18. Kokel D, Rennekamp AJ, Shah AH, Liebel U, Peterson RT. Behavioral barcoding in the cloud: embracing data-intensive digital phenotyping in neuropharmacology. Trends Biotechnol. 2012/05/29. 2012 Aug; 30(8):421–5. https://doi.org/10.1016/j.tibtech.2012.05.001 PMID: 22652049
19. Bruni G, Lakhani P, Kokel D. Discovering novel neuroactive drugs through high-throughput behavior-based chemical screening in the zebrafish. Front Pharmacol. 2014 Jul 24; 5:153. https://doi.org/10.3389/fphar.2014.00153 PMID: 25104936
20. Bruni G, Rennekamp AJ, Velenich A, McCarroll M, Gendelev L, Fertsch E, et al. Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. Nat Chem Biol. 2016; 12(7):559–66. https://doi.org/10.1038/nchembio.2097 PMID: 27239787
21. Kokel D, Bryan J, Laggner C, White R, Cheung CYJ, Mateus R, et al. Rapid behavior-based identification of neuroactive small molecules in the zebrafish. Nat Chem Biol. 2010; 6(3):231–7. https://doi.org/10.1038/nchembio.307 PMID: 20081854
22. McCarroll MN, Gendelev L, Kinser R, Taylor J, Bruni G, Myers-Turnbull D, et al. Zebrafish behavioural profiling identifies GABA and serotonin receptor ligands related to sedation and paradoxical excitation. Nat Commun 2019 101. 2019 Sep 9; 10(1):1–14.
23. Strähle U, Scholz S, Geisler R, Greiner P, Hollert H, Rastegar S, et al. Zebrafish embryos as an alternative to animal experiments—A commentary on the definition of the onset of protected life stages in animal welfare regulations. Reprod Toxicol. 2012 Apr 1; 33(2):128–32. https://doi.org/10.1016/j.reprotox.2011.06.121 PMID: 21726626

24. Wu JP, Li MH. The use of freshwater planarians in environmental toxicology studies: Advantages and potential. Ecotoxicol Environ Saf. 2018 Oct 15; 161:45–56. https://doi.org/10.1016/j.ecoenv.2018.05.057 PMID: 29859407

25. Raffa RB, Rawls SM, editors. Planaria: A model for drug action and abuse. Austin: Landes Bioscience; 2008.

26. Buttarelli FR, Pellicano C, Pontieri FE. Neuropharmacology and behavior in planarians: Translations to mammals. Comp Biochem Physiol—C Toxicol Pharmacol. 2008 May; 147(4):399–408. https://doi.org/10.1016/j.cbpc.2008.01.009 PMID: 18294919

27. Ireland D, Collins EMS. New Worm on the Block: Planarians in (Neuro)Toxicology. Curr Protoc. 2022 Dec 1; 2(12):e637. https://doi.org/10.1002/cpz1.637 PMID: 36571713

28. Ross KG, Currie KW, Pearson BJ, Zayas RM. Nervous system development and regeneration in freshwater planarians. Wiley Interdiscip Rev Dev Biol. 2017; 6(3):1–26. https://doi.org/10.1002/wdev.266 PMID: 28326682

29. Rozanski A, Moon H, Brandl H, Martin-Duran JM, Grohme MA, Huttner K, et al. PlanMine 3.0-improvements to a mineable resource of flatworm biology and biodiversity. Nucleic Acids Res. 2019 Jan; 47 (D1):D812–20. https://doi.org/10.1093/nar/gky1070 PMID: 30496475

30. Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, et al. The genome of Schmidtea mediterranea and the evolution of core cellular mechanisms. Nature. 2018; 554(7690):56–61. https://doi.org/10.1038/nature25473 PMID: 29364871

31. Inoue T, Yamashita T, Agata K, Inoue T, Yamashita T A K, Inoue T, et al. Thermosensory signaling by TRPM is processed by brain serotonergic neurons to produce planarian thermotaxis. J Neurosci. 2014 Nov; 34(47):5701–14. https://doi.org/10.1523/JNEUROSCI.5379-13.2014 PMID: 25411498

32. Almazan EMP, Ryan JF, Rouhana L. Regeneration of Planarian Auricles and Reestablishment of Chemotactic Ability. Front Cell Dev Biol. 2021 Nov 26; 9:3334. https://doi.org/10.3389/fcell.2021.777951 PMID: 34901022

33. Currie KW, Pearson BJ. Transcription factors lhx1/5-1 and pitx are required for the maintenance and regeneration of serotonergic neurons in planarians. Development. 2013 Sep; 140(17):3577–88. https://doi.org/10.1242/dev.098590 PMID: 23903188

34. Nishimura K, Kitamura Y, Umesono Y, Takeuchi K, Takata K, Taniguchi T, et al. Identification of glutamic acid decarboxylase gene and distribution of GABAergic nervous system in the planarian Dugesia japonica. Neuroscience. 2008 Jun; 153(4):1103–14. https://doi.org/10.1016/j.neuroscience.2008.03.026 PMID: 18440152

35. Arenas OM, Zaharieva EE, Para A, Vásquez-Doorman C, Petersen CP, Gallio M. Activation of planarian TRPA1 by reactive oxygen species reveals a conserved mechanism for animal nociception. Nat Neurosci. 2017 Dec 16; 20(12):1686–93. https://doi.org/10.1038/s41593-017-0005-0 PMID: 29184198

36. Sabry Z, Ho A, Ireland D, Rabeler C, Cochet-Escartin O, Collins EMS. Pharmacological or genetic targeting of Transient Receptor Potential (TRP) channels can disrupt the planarian escape response. PLoS One. 2019; 14(12):e0226104. https://doi.org/10.1371/journal.pone.0226104 PMID: 31805147

37. Shettigar N, Joshi A, Dalmeida R, Gopalkrishna R, Chakravarthy A, Patnaik S, et al. Hierarchies in light sensing and dynamic interactions between ocular and extraocular sensory networks in a flatworm. Sci Adv. 2017 Jul 5; 3(7).

38. Shibata N, Agata K. RNA interference in planarians: Feeding and injection of synthetic dsRNA. Methods Mol Biol. 2018; 1774:455–66. https://doi.org/10.1007/978-1-4939-7802-1_18 PMID: 29916171

39. Ofoegbu PU, Lourenço J, Mendo O, Soares AMVM, Ao J, Pestana LT. Effects of low concentrations of psychiatric drugs (carbamazepine and fluoxetine) on the freshwater planarian, Schmidtea mediterranea. Chemosphere. 2019; 217:542–9. https://doi.org/10.1016/j.chemosphere.2018.10.198 PMID: 30445399

40. Tallarida CS, Bires K, Avershal J, Tallarida RJ, Seo S, Rawls SM. Ethanol and cocaine: Environmental place conditioning, stereotypy, and synergism in planarians. Alcohol. 2014 Sep; 48(6):579–86. https://doi.org/10.1016/j.alcohol.2014.07.006 PMID: 25212751

41. Rawls SM, Patil T, Tallarida CS, Baron S, Kim M, Song K, et al. Nicotine behavioral pharmacology: Clues from planarians. Drug Alcohol Depend. 2011 Nov 1; 118(2–3):274–9. https://doi.org/10.1016/j.drugalcdep.2011.04.001 PMID: 21530106

42. Talbot J, Schötz EM. Quantitative characterization of planarian wild-type behavior as a platform for screening locomotion phenotypes. J Exp Biol. 2011 Apr 1; 214(Pt 7):1063–7. https://doi.org/10.1242/jeb.052290 PMID: 21389189

43. Hagstrom D, Cochet-Escartin O, Zhang S, Khuu C, Collins EMS. Freshwater planarians as an alternative animal model for neurotoxicology. Toxicol Sci. 2015 Jun 26; 147(1):270–85. https://doi.org/10.1093/toxsci/kfv129 PMID: 26116028

44. Zhang S, Hagstrom D, Hayes P, Graham A, Collins EMS. Multi-behavioral endpoint testing of an 87-chemical compound library in freshwater planarians. Toxicol Sci. 2019 Jun 8; 167(1):26–44. https://doi.org/10.1093/toxsci/kfy145 PMID: 29893936

45. Ireland D, Collins E maria S. Planarians as a model to study neurotoxic agents. 1st ed. Alternative Methods in Neurotoxicology. Elsevier Inc.; 2023. 1–32 p.

46. Zhang S, Ireland D, Sipes NS, Behl M, Collins EMS. Screening for neurotoxic potential of 15 flame retardants using freshwater planarians. Neurotoxicol Teratol. 2019; 73:54–66. https://doi.org/10.1016/j.ntt.2019.03.003 PMID: 30943442

47. Ireland D, Bochenek V, Chaiken D, Rabeler C, Onoe S, Soni A, et al. Dugesia japonica is the best suited of three planarian species for high-throughput toxicology screening. Chemosphere. 2020; 253:126718. https://doi.org/10.1016/j.chemosphere.2020.126718 PMID: 32298908

48. Ireland D, Zhang S, Bochenek V, Hsieh JH, Rabeler C, Meyer Z, et al. Differences in neurotoxic outcomes of organophosphorus pesticides revealed via multi-dimensional screening in adult and regenerating planarians. Front Toxicol. 2022; 4:948455. https://doi.org/10.3389/ftox.2022.948455 PMID: 36267428

49. Cochet-Escartin O, Mickolajczk KJ, Collins EMS. Scrunching: a novel escape gait in planarians. Phys Biol. 2015; 12(5):055001.

50. Fuselier SG, Ireland D, Fu N, Rabeler C, Collins EMS. Comparative toxicity assessment of glyphosate and two commercial formulations in the planarian Dugesia japonica. Front Toxicol. 2023; 5(June):1–12. https://doi.org/10.3389/ftox.2023.1200881 PMID: 37435546

51. Dean MRP, Duncan EM. Laboratory Maintenance and Propagation of Freshwater Planarians. Curr Protoc Microbiol. 2020; 59(1):1–20. https://doi.org/10.1002/cpmc.120 PMID: 33058563

52. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022 Jul 1; 12(7):3049–62. https://doi.org/10.1016/j.apsb.2022.02.002 PMID: 35865092

53. Porter RHP, Jaeschke G, Spooren W, Ballard TM, Büttelmann B, Kolczewski S, et al. Fenobam: A clinically validated nonbenzodiazepine anxiolytic is a potent, selective, and noncompetitive mGlu5 receptor antagonist with inverse agonist activity. J Pharmacol Exp Ther. 2005 Nov 1; 315(2):711–21. https://doi.org/10.1124/jpet.105.089839 PMID: 16040814

54. Rawls SM, Gomez T, Stagliano GW, Raffa RB. Measurement of glutamate and aspartate in planaria. J Pharmacol Toxicol Methods. 2006 Jan; 53(3):291–5. https://doi.org/10.1016/j.vascn.2005.10.004 PMID: 16332445

55. Cebrià F, Kudome T, Nakazawa M, Mineta K, Ikeo K, Gojobori T, et al. The expression of neural-specific genes reveals the structural and molecular complexity of the planarian central nervous system. Mech Dev. 2002 Aug; 116(1–2):199–204. https://doi.org/10.1016/s0925-4773(02)00134-x PMID: 12128224

56. Berry-Kravis E, Hessl D, Coffey S, Hervey C, Schneider A, Yuhas J, et al. A pilot open label, single dose trial of fenobam in adults with fragile X syndrome. J Med Genet. 2009 Apr; 46(4):266. https://doi.org/10.1136/jmg.2008.063701 PMID: 19126569

57. Cavallone LF, Montana MC, Frey K, Kallogjeri D, Wages JM, Rodebaugh TL, et al. The metabotropic glutamate receptor 5 negative allosteric modulator fenobam: pharmacokinetics, side effects, and analgesic effects in healthy human subjects. Pain. 2020 Jan 1; 161(1):135. https://doi.org/10.1097/j.pain.0000000000001695 PMID: 31568235

58. Jacob W, Gravius A, Pietraszek M, Nagel J, Belozertseva I, Shekunova E, et al. The anxiolytic and analgesic properties of fenobam, a potent mGlu5 receptor antagonist, in relation to the impairment of learning. Neuropharmacology. 2009 Aug 1; 57(2):97–108. https://doi.org/10.1016/j.neuropharm.2009.04.011 PMID: 19426746

59. Truong L, Rericha Y, Thunga P, Marvel S, Wallis D, Simonich MT, et al. Systematic developmental toxicity assessment of a structurally diverse library of PFAS in zebrafish. J Hazard Mater. 2022 Jun 5; 431. https://doi.org/10.1016/j.jhazmat.2022.128615 PMID: 35263707

60. Mirza NR, Nielsen E, Troelsen KB. Serotonin transporter density and anxiolytic-like effects of antidepressants in mice. Prog Neuropsychopharmacol Biol Psychiatry. 2007 May 9; 31(4):858–66. https://doi.org/10.1016/j.pnpbp.2007.01.020 PMID: 17335951

61. Troelsen KB, Nielsen E, Mirza NR. Chronic treatment with duloxetine is necessary for an anxiolytic-like response in the mouse zero maze: The role of the serotonin transporter. Psychopharmacology (Berl). 2005 Oct 20; 181(4):741–50. https://doi.org/10.1007/s00213-005-0032-5 PMID: 16032412

62. Zajdel P, Marciniec K, Maślankiewicz A, Grychowska K, Satała G, Duszyńska B, et al. Antidepressant and antipsychotic activity of new quinoline- and isoquinoline-sulfonamide analogs of aripiprazole targeting serotonin 5-HT1A/5-HT2A/5-HT7 and dopamine D2/D3 receptors. Eur J Med Chem. 2013 Feb 1; 60:42–50.

63. Conway CR, Chibnall JT, Cumming P, Mintun MA, Gebara MAI, Perantie DC, et al. Antidepressant response to aripiprazole augmentation associated with enhanced FDOPA utilization in striatum: a preliminary PET study. Psychiatry Res. 2014 Mar 3; 221(3):231. https://doi.org/10.1016/j.pscychresns.2014.01.003 PMID: 24468015

64. Stahl SM, Pradko JF, Haight BR, Modell JG, Rockett CB, Learned-Coughlin S. A Review of the Neuropharmacology of Bupropion, a Dual Norepinephrine and Dopamine Reuptake Inhibitor. Prim Care Companion J Clin Psychiatry. 2004; 6(4):159–66. https://doi.org/10.4088/pcc.v06n0403 PMID: 15361919

65. Farhadpour S, Warner TA, Maxwell AE. Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. Remote Sens. 2024; 16(3).

66. Rezvani S, Wang X. A broad review on class imbalance learning techniques. Appl Soft Comput. 2023; 143:110415.

67. Kang J, Hsu CH, Wu Q, Liu S, Coster AD, Posner BA, et al. Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. Nat Biotechnol. 2016; 34(1):70–7. https://doi.org/10.1038/nbt.3419 PMID: 26655497

68. Slack MD, Martinez ED, Wu LF, Altschuler SJ. Characterizing heterogeneous cellular responses to perturbations. Proc Natl Acad Sci U S A. 2008; 105(49):19306–11. https://doi.org/10.1073/pnas.0807038105 PMID: 19052231

69. Hawkins PCD, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools. J Med Chem. 2007 Jan 1; 50(1):74–82. https://doi.org/10.1021/jm0603365 PMID: 17201411

70. Lin X, Xu S, Liu X, Zhang X, Hu J. Detecting Drug–Target Interactions with Feature Similarity Fusion and Molecular Graphs. Biology (Basel). 2022; 11(7). https://doi.org/10.3390/biology11070967 PMID: 36101348

71. Rovida C, Alépée N, Api AM, Basketter DA, Bois FY, Caloni F, et al. Integrated testing strategies (ITS) for safety assessment. ALTEX. 2015; 32(1):25–40. https://doi.org/10.14573/altex.1411011 PMID: 25413849

72. Hogberg HT, Hsieh JH, Chang X, Sipes NS, Shafer T, Behl M. Case study on the use of Integrated Approaches for Testing and Assessment for DNT to prioritize a class of Organophosphorus flame retardants. Series on Testing and Assessment. Paris; 2022.

73. Güvenç Paltun B, Mamitsuka H, Kaski S. Improving drug response prediction by integrating multiple data sources: Matrix factorization, kernel and network-based approaches. Brief Bioinform. 2021; 22(1):346–59. https://doi.org/10.1093/bib/bbz153 PMID: 31838491

74. Moshkov N, Becker T, Yang K, Horvath P, Dancik V, Wagner BK, et al. Predicting compound activity from phenotypic profiles and chemical structures. Nat Commun. 2023; 14(1):1–11.

75. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014; 11(3):333–7. https://doi.org/10.1038/nmeth.2810 PMID: 24464287

76. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011 Oct; 3(10):33. https://doi.org/10.1186/1758-2946-3-33 PMID: 21982300

77. Krieger E, Nielsen JE, Spronk CAEM, Vriend G. Fast empirical pKa prediction by Ewald summation. J Mol Graph Model. 2006 Dec; 25(4):481–6. https://doi.org/10.1016/j.jmgm.2006.02.009 PMID: 16644253

78. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015 Jul 7; 11(8):3696–713. https://doi.org/10.1021/acs.jctc.5b00255 PMID: 26574453

79. Ozvoldik K, Stockner T, Krieger E. YASARA Model-Interactive Molecular Modeling from Two Dimensions to Virtual Realities. J Chem Inf Model. 2023 Oct 23; 63(20):6177–82. https://doi.org/10.1021/acs.jcim.3c01136 PMID: 37782001

80. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015 Dec 8; 7(1). https://doi.org/10.1186/s13321-015-0069-3 PMID: 26052348

81. Todeschini R, Consonni V, Xiang H, Holliday J, Buscema M, Willett P. Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets. J Chem Inf Model. 2012 Nov 26; 52(11):2884–901. https://doi.org/10.1021/ci300261r PMID: 23078167

82. Klei HE, Moriarty NW, Echols N, Terwilliger TC, Baldwin ET, Pokross M, et al. Ligand placement based on prior structures: the guided ligand-replacement method. Acta Crystallogr D Biol Crystallogr. 2014 Jan; 70(Pt 1):134–43. https://doi.org/10.1107/S1399004713030071 PMID: 24419386

83. Zaki MJ, Meira W Jr.. Linear discriminant analysis. In: Data Mining and Machine Learning: Fundamental Concepts and Algorithms. 2nd Editio. Cambridge, UK: Cambridge University Press; 2020. p. 501–16.

84. Xia Y. Correlation and association analyses in microbiome study integrating multiomics in health and disease. Prog Mol Biol Transl Sci. 2020 Jan 1; 171:309–491. https://doi.org/10.1016/bs.pmbts.2020.04.003 PMID: 32475527

85. Laliberte E, Legendre P. A distance-based framework for measuring functional diversity from multiple traits. Ecology. 2010 Jan; 91(1):299–305. https://doi.org/10.1890/08-2244.1 PMID: 20380219

86. Anderson MJ, Willis TJ. Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. Ecology. 2003; 84(2):511–25.

87. Hammer Ø, Harper DAT, editors. Paleontological Data Analysis. Oxford, UK: Blackwell Publishing, Ltd.; 2005. 1–351 p.

88. Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron. 2001; 4(1):1–9.

89. Davis JC. Statistics and Data Analysis in Geology. New York: John Wiley & Sons; 1986. 646 p.

90. Gower JC. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. Biometrika. 1966 Dec; 53(3/4):325.

91. Gower JC. A General Coefficient of Similarity and Some of Its Properties. Biometrics. 1971 Dec; 27 (4):857.

92. Lachenbruch PA, Mickey MR. Estimation of Error Rates in Discriminant Analysis. Technometrics. 1968; 10(1):1–11.

93. White JW, Ruttenberg BI. Discriminant function analysis in marine ecology: some oversights and their solutions. Mar Ecol Prog Ser. 2007 Jan 11; 329:301–5.

94. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arxiv. 2018 Feb 9;(1802.03426).

95. Armstrong G, Martino C, Rahman G, Gonzalez A, Vázquez-Baeza Y, Mishne G, et al. Uniform Manifold Approximation and Projection (UMAP) Reveals Composite Patterns and Resolves Visualization Artifacts in Microbiome Data. mSystems. 2021 Oct 26; 6(5). https://doi.org/10.1128/mSystems.00691-21 PMID: 34609167

96. Han SH, Kim KW, Kim S, Youn YC. Artificial Neural Network: Understanding the Basic Concepts without Mathematics. Dement Neurocognitive Disord. 2018; 17(3):83. https://doi.org/10.12779/dnd.2018.17.3.83 PMID: 30906397

97. Rocha M, Cortez P, Neves J. Evolution of neural networks for classification and regression. Neurocomputing. 2007; 70(16–18):2809–16.

98. Li H, Wang X, Ding S. Research and development of neural network ensembles: a survey. Artif Intell Rev. 2018; 49(4):455–79.

99. Nakas CT, Dalrymple-Alford JC, Anderson TJ, Alonzo TA. Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening. Stat Med. 2013; 32(6):995–1003. https://doi.org/10.1002/sim.5592 PMID: 22949169

100. Gorodkin J. Comparing two K-category assignments by a K-category correlation coefficient. Comput Biol Chem. 2004 Dec; 28(5–6):367–74. https://doi.org/10.1016/j.compbiolchem.2004.09.006 PMID: 15556477

101. Basak SC, Vracko MG. Parsimony Principle and its Proper use/ Application in Computer-assisted Drug Design and QSAR. Curr Comput Aided Drug Des. 2020; 16(1):1–5. https://doi.org/10.2174/1573409916012001061228 54 PMID: 31934837

102. Vandekerckhove J, Matzke D, Wagenmakers EJ. Model comparison and the principle of parsimony. Oxford Handbook of. 2014. 1–29 p.

103. Zaki MJ, Meira W Jr.. Support Vector Machines. In: Data Mining and Machine Learning: Fundamental Concepts and Algorithms. 2nd Editio. Cambridge, UK: Cambridge University Press; 2020. p. 517–45.

104. Noble WS. What is a support vector machine? Nat Biotechnol 2006 2412. 2006 Dec; 24(12):1565–7. https://doi.org/10.1038/nbt1206-1565 PMID: 17160063

105. Brereton RG, Lloyd GR. Support Vector Machines for classification and regression. Analyst. 2010 Jan 25; 135(2):230–67. https://doi.org/10.1039/b918972f PMID: 20098757

106. Bayingana K, Ireland D, Rosenthal E, Rabeler C, Collins EMS. Adult and regenerating planarians respond differentially to chronic drug exposure. Neurotoxicol Teratol. 2023; 96(August 2022):107148. https://doi.org/10.1016/j.ntt.2022.107148 PMID: 36539103

107. Lezak KR, Missig G, Carlezon WA. Behavioral methods to study anxiety in rodents. Dialogues Clin Neurosci. 2017; 19(2):181. https://doi.org/10.31887/DCNS.2017.19.2/wcarlezon PMID: 28867942

108. Stewart A, Gaikwad S, Kyzar E, Green J, Roth A, Kalueff A V. Modeling anxiety using adult zebrafish: A conceptual review. Neuropharmacology. 2012 Jan; 62(1):135. https://doi.org/10.1016/j.neuropharm.2011.07.037 PMID: 21843537

109. Akiyama Y, Agata K, Inoue T. Spontaneous behaviors and wall-curvature lead to apparent wall preference in planarian. Sakakibara M, editor. PLoS One. 2015 Nov 5; 10(11):e0142214. https://doi.org/10.1371/journal.pone.0142214 PMID: 26539715

110. Hsieh JH, Ryan K, Sedykh A, Lin JA, Shapiro AJ, Parham F, et al. Application of benchmark concentration (BMC) analysis on zebrafish data: A new perspective for quantifying toxicity in alternative animal models. Toxicol Sci. 2019; 167(1):282–92.