TITLE: Machine Learning Based Modelling of Human and Insect Olfaction Screens Millions of compounds to Identify Pleasant Smelling Insect Repellents

AUTHORS: Joel Kowalewski[1], Sean M. Boyle[2], Ryan Arvidson[2], Jadrian Ejercito[2], and Anandasankar Ray[1,2]

AFFILIATION:

[1]Interdepartmental Neuroscience Program

[2]Department of Molecular, Cell and Systems Biology

University of California, Riverside, CA 92521, USA

CORRESPONDING AUTHOR: Anandasankar Ray

Department of Molecular Cell and Systems Biology, University of California Riverside, 3401 Watkins Drive, Riverside, CA 92521, USA, Phone:+1(951) 827-5998, Email: anand.ray@ucr.edu

Abstract

The rational discovery of behaviorally active odorants is impeded by a lack of understanding on how the olfactory system generates percept or valence for a volatile chemical. In previous studies we showed that chemical informatics could be used to model prediction of ligands for a large repertoire of odorant receptors in *Drosophila* (Boyle et al., 2013). However, it remained difficult to predict behavioral valence of volatiles since the activities of a large ensembles of odor receptors encode odor information, and little is known of the complex information processing circuitry. This is a systems-level challenge well-suited for Machine-learning approaches which we have used to model olfaction in two organisms with completely unrelated olfactory receptor proteins: humans (~400 GPCRs) and insects (~100 ion-channels). We use chemical structure-based Machine Learning models for prediction of valence in insects and for 146 human odor characters. Using these predictive models, we evaluate a vast chemical space of >10 million compounds *in silico.* Validations of human and insect behaviors yield very high success rates. The discovery of desirable fragrances for humans that are highly repulsive to insects offers a powerful integrated approach to discover new insect repellents.

Introduction

Animals can detect volatile molecules in their environment using specialized protein receptors in their olfactory organs. Mammalian odorant receptors were discovered in 1991 and are G-protein coupled receptors (GPCRs) numbering >400 in humans (Buck L and Axel, 1991). Several years later starting in 1999 three families of distinct insect odorant receptors were discovered from the *Drosophila* genome which are ligand-gated ion channels and number ~100 (Clyne et al., 1999, 2000). Over the subsequent 20-30 years, studies in mice and insects focusing on odor detection, odorant receptors, odorant receptor

41 neurons, and higher-level circuitry, have revealed important principles. A single volatile is usually
42 detected by a combination of receptors, and the activity from the neural circuit processing the
43 information is extremely complex. Previously, we successfully developed a chemical informatics
44 approach to discover ligands for a large number of odorant receptors in *Drosophila* (Boyle et al., 2013).
45 However, the mechanisms connecting olfactory perception (e.g., rose, lemon, pleasant, unpleasant, etc.)
46 to olfactory receptor activity are incompletely understood and it is therefore challenging to predict
47 behavior from odorant receptor activity alone. There is also often a lack of correspondence between
48 simple, 2D representations of odorants and their perceptual descriptions, and the molecular-cellular
49 organization of the olfactory system lacks a simple discernible coding strategy that enables predictions,
50 unlike auditory or visual processing (Buck L and Axel, 1991; Grabe & Sachse, 2018; P Mombaerts, 1999;
51 Peter Mombaerts, 2001; Sachse & Beshel, 2016; Vassar et al., 1993).

52 Yet Machine Learning algorithms have shown promise in accurately predicting human odor
53 perception (Keller et al., 2017; Kowalewski et al., 2021; Kowalewski & Ray, 2020b). Recently, we have
54 shown that the volatile structure-to-percept prediction can be improved by using the activities of a small
55 number of human odorant receptors and applying machine learning algorithms to identify highly
56 predictive combinations (Kowalewski & Ray, 2020b). In insects, we have limited success modeling the
57 behavioral valence of volatiles from olfactory receptor neuron activity in the fruit fly, *Drosophila*
58 *melanogaster*. Since receptor activity depends on chemical structure, this suggests machine learning
59 algorithms can potentially detect a consistent physicochemical basis for simple behavioral or complex
60 perceptual responses (Kowalewski & Ray, 2020b; MacWilliam et al., 2018).

61 To test whether Machine learning based modeling can lead to identification of strong
62 behaviorally active volatiles, we selected a challenge that has a very high potential for easing human
63 suffering and yet had little innovation: pleasant smelling fragrances that are also repulsive to harmful
64 insects that transmit diseases to humans or cause agricultural loss. Annually, mosquitos account for

65 ~500,000-700,000 deaths and a billion infections due to vector-borne illnesses like malaria, Dengue,
66 Yellow Fever. Insects also destroy ~30% of the global agricultural produce causing additional suffering.
67 Insects are currently managed through widespread application of insecticides, but mutations are
68 emerging that confer insecticide resistance and affect insect olfactory responses in both human disease
69 vectors and plant pests (Deletre et al., 2019; Duneau et al., 2018).

70 The most widely used topical repellent for mosquitoes is DEET (N,N-Diethyl-m-toluamide), a
71 chemical with undesirable chemical and cosmetic properties. DEET (N,N-Diethyl-m-toluamide) was
72 discovered >75 years ago through empirical testing. It is not widely used outside the developed world,
73 where vector borne diseases are present globally, and are also use restrictions due to its strong solvent
74 properties, even dissolving plastics and nylons. Recent studies have also raised human health concerns ,
75 as the chemical interferes with inhibitory neurotransmission in mammals via $GABA_A$ and Glycine
76 receptors (Grant & Hall, 2019). Unfortunately, not enough is understood about the odorant receptors
77 for aversion in mosquitoes as a result there has been limited innovation in the search for repellents even
78 a quarter century after discovery of insect odorant receptors. The last repellent registered by the US EPA
79 for commercial use was Picaridin, over 20 years ago.

80 Volatile compounds that have pleasant fragrances are desired by humans in cosmetics and are
81 therefore excellent candidates for novel mosquito repellents. In order to identify fragrances that would
82 be aversive to insects, we assumed machine learning (ML) could be used to identify structural features
83 enriched among repellents and then detect these features within any arbitrary, large chemical space
84 (Maldonado et al., 2006). We trained a support vector machine (SVM) on a comprehensive set of
85 chemicals with known repellency from literature and identified predictive physicochemical features. In
86 parallel, we adapted a collection of recently developed ML models to predict 146 different human odor
87 percepts, helping us organize the candidate repellents by their potential fragrances (Kowalewski et al.,
88 2021; Kowalewski & Ray, 2020b). We then validated the accuracy of behavioral predictions for insects

89 and humans by randomly testing 50 predicted chemicals for repellency in mosquitoes and odor
90 perception in humans. The validation success was >75% in each assay, repellent and perceptual, with
91 ~95% of the chemicals tested for repellency in round 1 showing aversion, which left ~5% as neutral or
92 slightly attractive to mosquitoes. Collectively, this suggests a strong physicochemical basis for odor
93 valence and percept in insects and humans. In addition, we also built *in silico* toxicity models, filtering
94 the compounds across multiple parameters for human use. A small number of the fragrances were
95 tested on evolutionarily distinct insects, finding that they are repellent to mosquitoes and the fruit fly,
96 *Drosophila*.

97 Tens of millions of dollars and several years may be required for identification and adequate
98 human-safety analysis for registration of new repellent chemicals by the EPA (Gupta & Bhattacharjee,
99 2006). Machine learning approaches for modeling toxicity as well as other physical and cosmetic
100 attributes are steadily advancing, suggesting it is possible to accelerate this search and significantly
101 reduce costs. Accordingly, we included additional computational filters to select predicted repellents
102 such as predicted fragrant odor perceptual qualities. Because fragrant molecules that are also insect
103 repellents may be rare to find, we applied this approach to a large library (>10 million) of purchasable
104 chemicals from the ZINC 15 database. These efforts represent a dramatic expansion of the insect
105 repellent chemical space and provides a theory-based high-throughput method to drive rational
106 discovery of novel insect repellents.

107

108 **Results**

109 **Machine learning can learn to predict aversive valence in insects from chemical structure**

110 To predict insect behavior, we developed a novel chemical informatics approach using Machine Learning
111 (ML). We assembled a training set of volatile chemicals known in the literature to have a negative

112 behavioral valence including eucalyptol, linalool, geraniol, citronella, alpha-thujone, beta-thujone (Kline
113 et al., 2003; Klocke et al., 1987; Syed & Leal, 2008), DEET, picaridin, 34 N-acyl piperidines that are
114 structurally related to picaridin (Katritzky et al., 2008) and structurally diverse odorants from fruits and
115 food sources that are not expected to elicit aversion (Carey et al., 2010; Hallem & Carlson, 2006). We
116 generated energy-minimized 3D structures for each and calculated 3,242 physicochemical features for
117 201 training chemicals. From it we identified 18 features that were highly correlated with aversive
118 valence (correlation of 0.912). We then fit a support vector machine (SVM) with these features to learn
119 to predict the repellency of test chemicals (Figure 1).

120 The trained model successfully predicted the valence of the test chemicals (avg AUC = 0.994),
121 indicating that it could distinguish volatiles with aversive valence from non-repellents (Figure 1).
122 Evaluation of the predictive chemical features indicated six membered rings, carbon-nitrogen distances,
123 tertiary amides, and the positioning of oxygen atoms in the molecule are associated with repellency
124 (Figure 1).

125 We subsequently used this repellency model to evaluate a >440,000 small molecules assembled
126 from databases which includes ~3,000 chemicals from natural sources like plants, insects, vertebrates,
127 and compounds that are in food, beverages, and cosmetics. The top 500 predicted compounds (top
128 ~0.1% of hits) represent a structurally diverse group of chemicals, with some showing similarities to
129 volatiles in the training set (Figure 1). Several naturally occurring compounds were also in the list of
130 precited repellents. Since the ML model successfully classified known repellents versus non-repellents,
131 we next experimentally tested the candidate repellents.

132 **Machine learning can prioritize candidate repellents by modeling their odor qualities to humans**

133 Importantly, repellents must meet many physical and cosmetic demands for human use. However,
134 evaluating the scent characteristics of candidate repellents with human subjects is a laborious,

135 expensive, and time-consuming task. In order to create a method to bypass this bottleneck, we adapted
136 our recent perceptual modeling approach involving the prediction of 146 odor perceptual descriptors
137 from chemical structure; example perceptual descriptors include “fruity other than citrus”, “lemon”,
138 “orange”, “cinnamon”, “perfumery” and unpleasant ones such as “sickening”, “rancid”, “animal”, and
139 “dirty linen” to name a few (Kowalewski et al., 2021; Kowalewski & Ray, 2020b). The models were
140 trained with data from prior human behavior studies in which participants evaluated ~150 odorants and
141 mixtures according to the 146 descriptors. Using these models, we predicted odor descriptors for the
142 candidate insect repellents that were discovered using the model described in Figure 1. The compounds
143 had a wide variety of predicted odor characteristics covering both pleasant and unpleasant fragrance
144 profiles. Together with the insect valence predictions, this gives us the foundations for selecting
145 desirable fragrances that can protect humans and plants from pests.

146 **Validation of computational behavioral predictions of human smell and insect aversion**

147 To validate the behavioral predictions of scent or repellency, on humans and insects,
148 respectively, we selected 47 compounds based on their availability and affordability. First, we compared
149 them to reports from human testing of smell properties. While odor qualities remain poorly
150 characterized, 37 of the 47 tested had odor and taste annotation in a sensorial databases at the The
151 Good Scents Company. We could then rigorously evaluate the accuracy of the trained perceptual
152 models by ROC analysis. The perceptual models predicted the observed descriptors for most volatiles
153 with a high success rate (Avg. AUC = 0.77) (Figure 2A,B). Interestingly, many chemicals in the predicted
154 repellents have pleasant odor characteristics, including highly desirable ones such as fruity, floral or
155 perfume-like, such as seen in the word cloud of enriched predicted characters for the predicted insect
156 repellents (Figure 2C).

157 We next tested the compounds against *Aedes aegypti* mosquitoes. We developed a high
158 throughput assay utilizing the innate attraction of the female mosquitoes to a heat block approximating
159 human skin temperature (37 °C) covered with a filter paper (Figure 3A). Each of the test compounds
160 were individually evaluated by applying on the filter paper (3% in acetone) and counting the numbers of
161 mosquitoes landing on the surface with solvent control before and after applying the compound. A
162 repellency index indicated that most of the predicted compounds (~90%) showed >50% repellency to
163 female *A. aegypti* mosquitoes (Figure 3B). These results suggest that the machine learning (ML)
164 algorithm was capable o predicting a chemical's perceptual qualities for humans and its valence for
165 insects. This set of chemicals represents a very large increase in new insect repellents, comprising
166 diverse chemical structures and smell characteristics that are desirable for human use.

167 Since we do not know the molecular receptor targets for these diverse repellents, and also know
168 that different insect species have both highly diverse and conserved olfactory receptors, we cannot tell
169 *a priori* whether the repellency is narrowly effective in mosquitoes or conserved across diverse insects.
170 To determine broad-spectrum aversive valence among insects, we tested another distantly related
171 dipteran, *Drosophila melanogaster*, which is >250 MY apart evolutionarily. All of the predicted
172 compounds displayed concentration-dependent aversive valence to *Drosophila* in a trap-based 2-choice
173 assay (Reeder et al., 2009; Syed et al., 2011). Most of the tested compounds showed strong repellency
174 (Figure 3C). The simplest interpretation is that the machine learning algorithm was able to identify
175 chemical features that predict negative valence across multiple species of dipteran insects.

176 **Repellents show calcium mobilization in mammalian cells**

177 An alternative approach to identifying insect repellents would be to computationally model a known
178 cellular or molecular activity relevant to repellency. Interestingly, one of the aversive compounds used
179 in the repellency training set is the commercial repellent DEET, had previously been seen to cause

180 calcium mobilization in human HEK293 cells at high concentrations (Dennis et al., 2018, Supplement). In
181 order to further evaluate if such Ca^{2+} mobilization is a common feature for repellency, we performed
182 additional experiments.

183 When HEK293 cells were exposed to DEET, a dose-dependent calcium transient was observed,
184 with an EC50 of 56.6 mM (~1% vol/vol), below the concentrations used on skin applied formulations (7-
185 98% vol/vol) of which up to 60% is absorbed into the skin (Schoenig et al., 1996; Selim et al., 1995).
186 (Figure 4A). Additional cell lines from different organisms were tested, including an insect S2 cell line,
187 and all of them were sensitive to DEET at similar EC50s (Figure 4A). The DEET-induced calcium transients
188 were similar in magnitude to the positive control, a 100 µM ATP induced signal (e.g., GPCR mediated)
189 (Figure 4B). To assess the source of DEET induced calcium flux, cells were depleted of intercellular
190 calcium with treatment with thapsigargin (a SERCA antagonist) 10 minutes before the assay. These
191 calcium-depleted cells did not show increased Ca transients when treated with DEET. This indicates that
192 DEET is likely inducing release from intercellular calcium stores, in a dose-dependent manner (Figure
193 4C). Interestingly, two additional commercially used repellents, picaridin and diethyl benzamide (a
194 structural analog of DEET), also evoked a strong intracellular Ca^{2+} signal (Figure 4D). Because these three
195 compounds were part of the training data to build our machine learning insect repellent models, we
196 tested whether the newly identified aversive fragrances also showed similar activity. We tested a
197 sampling of the verified aversive fragrances, and a few traditionally known repellents, at two
198 concentrations (5% and 0.5%) in the HEK293 cell line and found that there were indeed some that
199 showed Ca^{2+} release (Figure 4E). However, there appeared to be no correlation between the repellency
200 index and the Ca^{2+} activation of the fragrant repellents (R^2=0.0093), suggesting that this conserved effect
201 is unlikely to contribute to aversive behavior.

202 Nevertheless, the data provided an opportunity to develop a better understanding of the
203 physicochemical basis of the Ca^{2+} mobilization seen *in vitro* and we trained machine learning models to

204 predict the calcium activity (Figure S1A). Upon analyses, the top chemical features for predicting the
205 calcium activity were different from the features selected by the model to predict repellency. This too
206 suggest that while there is a structural basis for basis for calcium activity which does not overlap with
207 shared features for detection as a repellent. The features fell into 3 broad categories including 2D
208 autocorrelation, RDF, and 3D MoRSE features. Autocorrelations are distance-based metrics, with the 2D
209 autocorrelation referring to the number of bonds between a pair of atoms in the molecule, while radial
210 distribution function (RDF) features encode dependencies between atoms that give rise to molecules'
211 characteristic infrared spectra (Hemmer et al., 1999). The 3D MoRSE features are based on theoretical
212 electron diffraction patterns and are therefore sensitive to differences in 3D geometries (Devinyak et al.,
213 2014). In previous work, these features have been shown to be predictive of ligand bonding and
214 biological activity (Beglari et al., 2020; Cheng et al., 2010; González et al., 2006).

215 Evaluation of models trained with these features on test data showed a high correlation
216 between the predicted and observed Ca^{2+} activities (Figure S1). Using another approach, where we
217 classified the chemicals as broadly falling into one of two categories (high versus low activity), the
218 predictions were also accurate, further emphasizing that the Ca^{2+} signal does indeed have a strong
219 physicochemical basis. Collectively, this suggests the unique 3D shape, rather than generic, possibly
220 indiscriminate information, underlies the observed Ca^{2+} activity, supporting our contention that the
221 activity may be a persistent off target effect among certain chemical repellents and should therefore be
222 further investigated. While the physiological consequences of this effect have not been studied, we
223 incorporated this activity as a parameter into our machine learning pipeline, allowing us to further triage
224 repellents for future use by low estimated Ca^{2+} activity, in addition to other desirable characteristics for
225 human or animal use.

226 **Mining massive commercially available chemical spaces for novel insect repellents**

227 The size and diversity of the chemical library we initially screened (~0.45M) is in the range of high-
228 throughput wet lab screening that pesticide discovery companies can undertake. The Machine learning
229 approach allows us to screen far more without the limitations of cost of synthesis and screening time. To
230 demonstrate the massive throughput capacity of this approach, we increased our search space to >10
231 million commercially available chemicals (from the ZINC 15 database), canvasing far more structural
232 diversity than the ~450,000 chemicals we previously screened. We built a new training set including the
233 chemical repellents we experimentally verified, more than doubling the number of the actives while
234 adding greater structural diversity. The top features fell mostly into 2D and 3D autocorrelations and 3D
235 MoRSE categories, suggesting that certain types of features became increasingly important following the
236 addition of more repellent activities. Notably, some features changed from this vastly improved training
237 set; that is, moving away from simple features that account for amides, or more generally, chemical
238 bonds including nitrogen (N). This possibly reflects an initial bias toward DEET, an amide, that was
239 resolved when including more complex feature selection methods and increasing the size and diversity
240 of the training set.

241 There was a strong correlation between the predicted and observed repellency when validating
242 these updated models on test data (R = 0.73). Then, by applying the filters that we progressively
243 incorporated into the pipeline, and described throughout (e.g., insect valence, odor perceptual qualities,
244 and intracellular calcium release), we identified a set of candidates still numbering in the hundreds. We
245 next predicted the acute oral mammalian toxicity (Rat) using a computational model from our earlier
246 work (Kowalewski & Ray, 2020a) and eliminated all compounds falling into EPA Tox categories I and II.
247 The top predictions contain many aminobenzoates comprised of benzene, amino, and ester groups.
248 Some chemicals notably differ in that they resemble amide repellents like diethyltoluamide (DEET)
249 (Figure 5A). The top hits were predicted to have a variety of different human odor characters, many of

250 which are perceptually pleasant or are useful in agriculture (Figure 5B). This approach accelerates the
251 identification of novel, fragrance-based, broad spectrum insect repellents.

252 **Discussion**

253 Here, we have developed a comprehensive machine learning pipeline to screen small or large
254 chemical spaces for suitable alternatives and verified numerous predicted chemicals for repellency in
255 various behavioral assays and across species, indicating the effectiveness of the pipeline. The rapid
256 identification of fragrances that have excellent behavioral activity, pleasant smell and safety greatly
257 improves the likelihood of regulatory approval and potential for use.

258 Interestingly, we identified a strong intracellular Ca^{2+} signal in response to DEET and picaridin as
259 well as some anthranilates. While the mechanism remains unclear, it has been previously observed in
260 cultured cell lines after DEET exposure (Dennis et al., 2018; Swale et al., 2014). One possibility proposed
261 in the literature is that the elevated intracellular Ca^{2+} is due to activation of octopamine receptors.
262 Octopamine plays a modulatory role, with targets including OctαR and OctαβR in insects and TAAR1 in
263 mammals (Kleinau et al., 2011; Swale et al., 2014), and while this could explain the many complex
264 physiological responses that have reported in the DEET literature (Briassoulis, 2001; Clem et al., 1993;
265 Corbel et al., 2009; Tavares et al., 2019), an unknown but conserved receptor subtype is still possible.
266 Additional research will be needed to determine whether this phenomenon is important for off-target
267 mammalian toxicity. In either case, the model we developed to predict the Ca^{2+} mobilization may help
268 isolate a set of key physicochemical features, enabling identification of compounds that have this effect.
269 However, since the calcium activity does not explain the repellency, additional electrophysiological and
270 genetic experiments are needed to further clarify the mechanism of action.

271 Although many repellents have been studied, DEET remains the most widely used commercially
272 available chemical repellent. Yet research continues to emerge about off-target toxicities in humans

who are repeatedly exposed to DEET; specifically, accounts suggest adverse skin reactions and activities in human or broadly mammalian nervous systems (Briassoulis, 2001; Grant & Hall, 2019; Osimitz et al., 2010). This study demonstrates the successful development and application of a machine learning pipeline that accelerates research into insect repellency and its physicochemical basis. The methods and data presented provide a high-throughput and rational way to identify additional novel repellent chemicals. Here, we emphasized screening candidate chemicals that are most likely to fit multiple requirements (smell and repellency) rather than simply repellency. As these requirements steadily increase, it is obvious that massive, unexplored chemical spaces offer the most promising leads. By predicting and filtering 10+ million purchasable chemicals according to their desirability for human use, we further illustrate an essential role for computational approaches (e.g., machine learning) in future repellency studies, accelerating the discovery of safe, ecologically friendly, and effective repellent chemicals.

Material and Methods

Behavior testing

Repellency Assay in Drosophila

Repellency was tested using a *Drosophila melanogaster* 2-choice trap assay. A chamber was built from DRAM vials, with four air holes made in the plastic lid and the bottom of chamber is coated with 5 ml of 1% agarose to maintain moisture. Traps are made from 1.5 ml Natural Microcentrifuge tubes with bottom 3 mm cut to form an opening and a 1000 ul blue graduated pipette tip catalogue number 1111-2023. To create a plastic funnel, a pipette tip was cut on both ends; cut to 1.0 cm from tip end to expand the opening of the funnel and then cut on the other end of plastic to create a final height of 2.5 cm for funnel. A BioRad PCR 0.2 ml Tube cap was cut from the strip and was inserted into the lid of microcentrifuge tube. Test odor (35 ul) was placed in the PCR tube cap and lure (90 ul) was placed in outer ring between the PCR cap and the microcentrifuge cap. In the dark, 4–7-day old, starved flies (10 males, 10 females) were released per trap and followed for 2 days. Counts were taken manually, and the Preference Index (PI) was computed using the formula below.

$$Preference\ Index\ (PI) = \frac{(\#\ flies\ in\ treated\ trap) - (\#\ flies\ in\ control\ trap)}{(\#\ files\ in\ test\ trap) + (\#\ flies\ in\ control\ trap)}$$

Repellency Assay in Mosquitoes

A non-contact heat attraction assay was performed by simultaneously offering mosquitoes a test chemical on a filter paper coupled with an attractive heat source. Heat for the assay was provided by a Hot hands® Hand warmers HH2 (Heat Max, Dalton, GA) which was activated by shaking in gloved hands. Temperature (~37°C) was verified using IR thermometers. The hand warmer was fitted snugly into a 100 mm x 15 mm petri dish base and covered with 15 cm x 15 cm piece of polyester net secured round the petri dish by a pair or 8-inch plastic cable ties (Gardner Bender, Milwaukee, WI). The petri

306 dishe with hand warmer was covered with a larger 150 mm X 15 mm petri dish with a 10 cm x 7.5 cm
307 rectangular window cut out. This netted window aligned with the hand warmers exposing a heated
308 surface.

309 A polyester net (measuring 9 cm x 8 cm) was treated with 500 µl of either acetone (solvent) or
310 3% (v/v or w/v) test chemical and allowed to air dry for 30 minutes to evaporate solvent. The test
311 compound (~0.2µg/cm^2) treated net was secured by stretching between a frame made using two 10 cm
312 x 7.5 cm flexible magnetic frames with 7.5 cm x 6.2 cm cut out windows. This dose is 40% lower than the
313 WHO, WHOPES "Guidelines for Efficacy Testing of Mosquito Repellents for Human Skin" recommended
314 dose for the positive control stimulus. Three more flexible magnets were added on top as spacers from
315 the heat source. The net-magnet assembly was placed directly over the 10 cm x 7.5 cm cut out openings
316 of the larger 150 mm X 15 mm petri dishes holding the heat pads. The stimulus was placed on the top
317 net of the frames. Temperature was monitored by a laser Infra-red thermometer and assay started at
318 37°C. A puff of 5% CO_2 was blown into the cage to behaviorally activate the mosquitoes and initiate the
319 assay. A webcam timelapse is used to record mosquitoes landing on the net for automated counting
320 offline using ImageJ. A custom Macro is used in ImageJ to count the mosquitoes in the selected area (the
321 net) across the timepoints. The counts from each frame are automatically tabulated.

322 **Macro:**

```
323 run("8-bit");
324 setAutoThreshold("Default");
325 //run("Threshold...");
326 setThreshold(0, 80);
327 run("Analyze Particles...", "size=2-70 show=[Overlay Outlines] clear summarize stack");
```

```
dir = getDirectory("image");
name = getTitle;
index = lastIndexOf(name, ".");
if (index!=-1) name = substring(name, 0, index);
name = name + ".csv";
selectWindow("Summary of "+getTitle);
saveAs("Measurements", dir+name);
```

Since this is a 1-choice assay, the solvent alone controls are done before and after the test material with at least 20 minutes between the trials. Any significant changes in behavior between the control solvent should indicate if mosquito behavior is acceptable.

Repellency index = 100 x [1 - (mean cumulative number of mosquitoes on the window of treatment for 5 seconds at time points 2,3,4,5 min/ mean cumulative number of mosquitoes that remained on window of solvent treatment for 5 seconds at time points 2,3,4,5 min)].

Each time point has N=5 trials, 40 mosquitoes per trial, Except for EA, where N=4.

In vitro assay of Ca^{2+} mobilization

All mammalian cell lines were maintained at 37 C, 5% CO2, and 99 percent relative humidity. S2 Cells were maintained at 26 C and 99% relative humidity. HEK293, CHOk1 and NG108 were grown in DMEM media (Gibco), supplemented with 10% FBS (Gibco), and antibiotic-antimycotic supplement (Gibco). S2 cells were cultured in Shields and Sang M3 Insect Medium (Sigma), supplemented with Insect Medium Supplement (Sigma). Cells were transfected with the luminescent calcium reporter aequorin 24 hours before experiment with X-tremeGENE 9 transfection reagent, in the case of HEK293 and CHOK1, and X-

349 tremeGENE HP, in the case of NG108 and S2 cells. Transfected cells were harvested with enzyme-free
350 cell dissociation buffer (Gibco), centrifuged, and suspended in assay media (DMEM with 0.1% BSA for
351 mammalian cells, unsupplemented SFX-100 insect medium for S2). The aequorin cofactor
352 coelenterazine-F (Nanolight) was added (5 µM final concentration) to the cell suspension, in the dark,
353 with gentle stirring for 2 to 3 hours before assay. DEET emulsion was prepared by adding a percentage
354 of DEET indicated, vol/vol to assay media, vortexed, and bath-sonicated for at least 15 min, but up to
355 one hour. Cells (100 µl, at approx. 50,000 cells) were injected into a 96 well plate containing 100 DEET µl
356 emulsion, and resulting luminescence recorded on a BMG Labtech Lumistar Omega plate reading
357 luminometer. Thapsigargin was added to the cell suspension were indicated to a final concentration of 1
358 µM, 10 minutes before assay.

359 **Chemical Informatics**

360 ***Machine learning predictions of repellency***

361 Physicochemical features for figures 1-2 were selected using Sequential Forward Selection (SFS)
362 (Haddad et al., 2008) from 3,224 using Dragon software (Talete). This algorithm is an iterative search
363 over the chemical feature pool that seeks to maximize the correlation between chemical feature values
364 and, here, the % repellency. These chemical features were subsequently used to fit a support vector
365 machine (SVM) with a radial basis function kernel (RBF). The implementation of the support vector
366 machine algorithm is from the e1071 R package. We performed 20 independent 5-fold cross-validations
367 followed by a receiver operating characteristics (ROC) analysis to analyze the performance of the
368 computational repellency prediction method. The trained SVM then ranked both a library obtained from
369 eMolecules (http://www.emolecules.com) (>400,000) and natural odor libraries (Supplementary
370 Material) (>3,000) by repellency using the optimized descriptor set to identify the best candidate

371 repellent compounds. The EPI Suite (http://www.epa.gov/oppt/exposure/pubs/episuite.htm) was used
372 to calculate predicted LogP and Vapour Pressure values.

373 For Figures 3, 5, S1-S2, we updated the methods and algorithms to further optimize the pipeline.
374 Physicochemical features were selected using AlvaDesc software, which offers > 5000 chemical features.
375 Optimal chemical features were then selected using cross-validation and recursive feature elimination
376 algorithm (CV-RFE). The RFE CV algorithm involves iterative partitioning of the data and the selection of
377 feature subsets (n_1, n_2, …, n_{i+1}; where "n" is the size of the feature subset and "i" is an arbitrary indexer).
378 Given a data partition and n features, an arbitrary ML algorithm is fit, importance is assigned to the
379 feature set and the model of n features predicts a test partition, that is, data excluded from training,
380 differentiating the CV-RFE from RFE alone. In this study, the ML algorithm was a support vector machine
381 (SVM) or a random forest. Over 100s of iterations the algorithm converges on the optimal (n) number of
382 features to maximize predictive success. Additionally, for every round the contribution of each chemical
383 feature is assessed, giving rise to an aggregate rank; namely, the frequency at which a feature has been
384 identified as important to the predictive success. Again, distinguishing the standard RFE algorithm from
385 CV-based adaptation. As for computing importance, the random forest algorithm internally tests the
386 contribution of the features. It accomplished this by fitting multiple decision trees on bootstrap samples
387 of the data. Consequently, a percentage of data has been excluded and serves as a test set. By
388 predicting the test set with and without permuting or shuffling the values of a given feature, the
389 algorithm can then assign an importance according to the % increase in error that is observed after
390 permuting the feature.

391 The support vector machine (SVM) algorithm does not natively perform a comparable
392 bootstrapping procedure as random forest, and therefore does not generate a test set to assess feature
393 importance. In this case, an external metric must be supplied to assign the importance. Depending on
394 the type of fit, classification or regression, this is a pseudo R^2, based on non-linear regression, or the

395 AUC for the feature and the outcome that is being predicted (e.g., % repellency). For additional details,
396 refer to the documentation provided with caret package in R.

397

398 ***Machine learning predictions of Ca^{2+} mobilization***

399 Due to the smaller number of observations (datapoints or chemicals) we opted for an ensemble
400 method that could take advantage of more of the available data. Random forest fits multiple decision
401 trees on bootstrap samples of the data, thereby reducing the tendency to overfit to one or few
402 chemicals in the dataset. As compared to the SVM, it has two simple parameters (mtry and ntrees) and
403 the tuning grid (or number combinations of these parameters that need to be tested to optimize
404 performance tends to be small). Accordingly, it is an ideal approach when little is known about the
405 activity being predicted.

406 To further reduce overfitting, we identified a small set of physicochemical features on different
407 subsets of the data and then randomly sampled from this pool. By doing so, the features are slightly
408 underfit, which is the better approach, again, when less is known about the activity being predicted.

409 Additionally, we used two different methods and metrics to independently verify that
410 physicochemical features were indeed predictive of the activity. First, a regression-based approached
411 predicting the raw assay activities. Second, a classification-based approach where the top ~40% of the
412 activity distribution was converted into an "Active" label, assessing predictive success by ROC analysis.
413 Finally, we fit classification models that were otherwise identical but trained on shuffled labels. This
414 ensured that the chance level-performances significantly differed.

415 ***Machine learning predictions of odor perceptual qualities***

Perceptual training data is from a study published in a reference book: ATLAS of Odor Character Profiles (Dravnieks, 1985). In the study, panellists supplied ratings for ~150 odorants and mixtures across 146 possible descriptors; the ratings are reported as the % usage. The % usage refers to the portion of panellists that rated an odorant (1-5) using a particular descriptor (1-146), indicating its relevance to the odorant. It is on the scale 0-100, with the maximum of 100% suggesting every participant found the descriptor to be relevant.

The top perceptual descriptors were selected by thresholding the predictions based on the mean value in the training data. For each descriptor in the training data, thresholds were computed as 2x the mean. These thresholds were then applied to the predicted repellents to generate a set of exemplar descriptors. Subsequently, these exemplars were converted into a descriptor frequency table, where the frequency is the total number of occurrences of a descriptor. The data and word cloud plot were generated using the R programming language, with additional functionality provided by the Wordcloud package in R.

429 **References**

430 Beglari, M., Goudarzi, N., Shahsavani, D., Arab Chamjangali, M., & Mozafari, Z. (2020). Combination of
431 radial distribution functions as structural descriptors with ligand-receptor interaction information
432 in the QSAR study of some 4-anilinoquinazoline derivatives as potent EGFR inhibitors. *Structural*
433 *Chemistry*, *31*(4), 1481–1491. https://doi.org/10.1007/s11224-020-01505-z

434 Boyle, S. M., McInally, S., & Ray, A. (2013). Expanding the olfactory code by in silico decoding of odor-
435 receptor chemical space. *ELife*, *2013*(2). https://doi.org/10.7554/eLife.01120.001

436 Briassoulis, G. (2001). Toxic encephalopathy associated with use of DEET insect repellents: A case
437 analysis of its toxicity in children. *Human and Experimental Toxicology*.
438 https://doi.org/10.1191/096032701676731093

439 Buck L and Axel, R. (1991). A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis.
440 *Cell*, *65*, 175–187.

441 Carey, A. F., Wang, G., Su, C. Y., Zwiebel, L. J., & Carlson, J. R. (2010). Odorant reception in the malaria
442 mosquito Anopheles gambiae. *Nature*. https://doi.org/10.1038/nature08834

443 Cheng, Z., Zhang, Y., & Fu, W. (2010). QSAR study of carboxylic acid derivatives as HIV-1 Integrase
444 inhibitors. *European Journal of Medicinal Chemistry*. https://doi.org/10.1016/j.ejmech.2010.05.052

445 Clem, J. R., Havemann, D. F., & Raebel, M. A. (1993). Insect Repellent (N,N-Diethyl-m-Toluamide)
446 Cardiovascular Toxicity in an Adult. *Annals of Pharmacotherapy*.
447 https://doi.org/10.1177/106002809302700305

448 Clyne, P. J., Warr, C. G., & Carlson, J. R. (2000). Candidate taste receptors in Drosophila. *Science*.
449 https://doi.org/10.1126/science.287.5459.1830

450 Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J., & Carlson, J. R. (1999). A Novel Family of
451 Divergent Seven-Transmembrane Proteins: Candidate Odorant Receptors in
452 Drosophila. *Neuron*, *22*(2), 327–338. https://doi.org/10.1016/S0896-6273(00)81093-4

453 Corbel, V., Stankiewicz, M., Pennetier, C., Fournier, D., Stojan, J., Girard, E., Dimitrov, M., Molgó, J.,
454 Hougard, J. M., & Lapied, B. (2009). Evidence for inhibition of cholinesterases in insect and
455 mammalian nervous systems by the insect repellent deet. *BMC Biology*.
456 https://doi.org/10.1186/1741-7007-7-47

457 Deletre, E., Martin, T., Duménil, C., & Chandre, F. (2019). Insecticide resistance modifies mosquito
458 response to DEET and natural repellents. *Parasites & Vectors*, *12*(1), 89.
459 https://doi.org/10.1186/s13071-019-3343-9

460 Dennis, E. J., Dobosiewicz, M., Jin, X., Duvall, L. B., Hartman, P. S., Bargmann, C. I., & Vosshall, L. B.
461 (2018). A natural variant and engineered mutation in a GPCR promote DEET resistance in C.
462 elegans. *Nature*. https://doi.org/10.1038/s41586-018-0546-8

463 Devinyak, O., Havrylyuk, D., & Lesyk, R. (2014). 3D-MoRSE descriptors explained. *Journal of Molecular*
464 *Graphics and Modelling*. https://doi.org/10.1016/j.jmgm.2014.10.006

465 Dravnieks, A. (1985). *Atlas of Odor Character Profiles*. ASTM. https://doi.org/10.1520/DS61-EB

466 Duneau, D., Sun, H., Revah, J., San Miguel, K., Kunerth, H. D., Caldas, I. V, Messer, P. W., Scott, J. G., &
467 Buchon, N. (2018). Signatures of Insecticide Selection in the Genome of Drosophila melanogaster.
468 *G3 (Bethesda, Md.)*, *8*(11), 3469–3480. https://doi.org/10.1534/g3.118.200537

469 González, M. P., Terán, C., Teijeira, M., & Helguera, A. M. (2006). QSAR Studies Using Radial Distribution
470 Function for Predicting A1 Adenosine Receptors Agonists. *Bulletin of Mathematical Biology*, *69*(1),
471 347. https://doi.org/10.1007/s11538-006-9127-3

Grabe, V., & Sachse, S. (2018). Fundamental principles of the olfactory code. In *BioSystems*. https://doi.org/10.1016/j.biosystems.2017.10.010

Grant, G. G., & Hall, A. C. (2019). Interactions of N, N- diethyl-meta-toluamide (DEET) and Novel Insect Repellents with Mammalian GABAA and Glycine Receptors. *The FASEB Journal*, *33*(1_supplement), 813.13-813.13. https://doi.org/10.1096/fasebj.2019.33.1_supplement.813.13

Gupta, R. K., & Bhattacharjee, A. (2006). *Discovery and Design of New Arthropod/Insect Repellents by Computer-Aided Molecular Modeling* (pp. 195 – 228). https://doi.org/10.1201/9781420006650.ch10

Haddad, R., Khan, R., Takahashi, Y. K., Mori, K., Harel, D., & Sobel, N. (2008). A metric for odorant comparison. *Nature Methods*. https://doi.org/10.1038/nmeth.1197

Hallem, E. A., & Carlson, J. R. (2006). Coding of Odors by a Receptor Repertoire. *Cell*, *125*(1), 143–160. https://doi.org/http://dx.doi.org/10.1016/j.cell.2006.01.050

Hemmer, M. C., Steinhauer, V., & Gasteiger, J. (1999). Deriving the 3D structure of organic molecules from their infrared spectra. *Vibrational Spectroscopy*. https://doi.org/10.1016/s0924-2031(99)00014-4

Katritzky, A. R., Wang, Z., Slavov, S., Tsikolia, M., Dobchev, D., Akhmedov, N. G., Hall, C. D., Bernier, U. R., Clark, G. G., & Linthicum, K. J. (2008). Synthesis and bioassay of improved mosquito repellents predicted from chemical structure. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0800571105

Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., Mainland, J. D., Ihara, Y., Yu, C. W., Wolfinger, R., Vens, C., Schietgat, L., De Grave, K., Norel, R., Stolovitzky, G., Cecchi, G. A., Vosshall, L. B., & Meyer, P. (2017). Predicting human olfactory perception from chemical features of odor

molecules. *Science*, *355*(6327), 820–826. https://doi.org/10.1126/science.aal2014

Kleinau, G., Pratzka, J., Nürnberg, D., Grüters, A., Führer-Sakel, D., Krude, H., Köhrle, J., Schöneberg, T., & Biebermann, H. (2011). Differential modulation of beta-adrenergic receptor signaling by trace amine-associated receptor 1 agonists. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0027073

Kline, D. L., Bernier, U. R., Posey, K. H., & Barnard, D. R. (2003). Olfactometric Evaluation of Spatial Repellents for Aedes aegypti . *Journal of Medical Entomology*. https://doi.org/10.1603/0022-2585-40.4.463

Klocke, J. A., Darlington, M. V., & Balandrin, M. F. (1987). 1,8-Cineole (Eucalyptol), a mosquito feeding and ovipositional repellent from volatile oil of Hemizonia fitchii (Asteraceae). *Journal of Chemical Ecology*. https://doi.org/10.1007/BF01012562

Kowalewski, J., Huynh, B., & Ray, A. (2021). A systems-wide understanding of the human olfactory percept chemical space. *Chemical Senses*. https://doi.org/10.1093/chemse/bjab007

Kowalewski, J., & Ray, A. (2020a). Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. *Heliyon*. https://doi.org/10.1016/j.heliyon.2020.e04639

Kowalewski, J., & Ray, A. (2020b). Predicting Human Olfactory Perception from Activities of Odorant Receptors. *IScience*, *23*(8), 101361. https://doi.org/10.1016/j.isci.2020.101361

MacWilliam, D., Kowalewski, J., Kumar, A., Pontrello, C., & Ray, A. (2018). Signaling Mode of the Broad-Spectrum Conserved CO2 Receptor Is One of the Important Determinants of Odor Valence in Drosophila. *Neuron*. https://doi.org/10.1016/j.neuron.2018.01.028

Maldonado, A. G., Doucet, J. P., Petitjean, M., & Fan, B. T. (2006). Molecular similarity and diversity in chemoinformatics: From theory to applications. In *Molecular Diversity*. https://doi.org/10.1007/s11030-006-8697-1

516 Mombaerts, P. (1999). Molecular biology of odorant receptors in vertebrates. *Annual Review of*
517 *Neuroscience*, *22*, 487–509. https://doi.org/10.1146/annurev.neuro.22.1.487

518 Mombaerts, Peter. (2001). THE HUMAN REPERTOIRE OF ODORANT RECEPTOR GENES AND
519 PSEUDOGENES. *Annual Review of Genomics and Human Genetics*, *2*(1), 493–510.
520 https://doi.org/10.1146/annurev.genom.2.1.493

521 Osimitz, T. G., Murphy, J. V., Fell, L. A., & Page, B. (2010). Adverse events associated with the use of
522 insect repellents containing N,N-diethyl-m-toluamide (DEET). *Regulatory Toxicology and*
523 *Pharmacology*. https://doi.org/10.1016/j.yrtph.2009.09.004

524 Reeder, N. L., Ganz, P. J., Carlson, J. R., & Saunders, C. W. (2009). Isolation of a Deet-Insensitive Mutant
525 of <I>Drosophila melanogaster</I> (Diptera: Drosophilidae). *Journal of Economic Entomology*.
526 https://doi.org/10.1603/0022-0493-94.6.1584

527 Sachse, S., & Beshel, J. (2016). The good, the bad, and the hungry: How the central brain codes odor
528 valence to facilitate food approach in Drosophila. In *Current Opinion in Neurobiology*.
529 https://doi.org/10.1016/j.conb.2016.06.012

530 Schoenig, G. P., Hartnagel, R. E., Osimitz, T. G., & Llanso, S. (1996). Absorption, distribution, metabolism,
531 and excretion of N,N-diethyl-M-toluamide in the rat. *Drug Metabolism and Disposition*.

532 Selim, S., Hartnagel, R. E., Osimitz, T. G., Gabriel, K. L., & Schoenig, G. P. (1995). Absorption, metabolism,
533 and excretion of N,N-diethyl-m-toluamide following dermal application to human volunteers.
534 *Toxicological Sciences*. https://doi.org/10.1093/toxsci/25.1.95

535 Swale, D. R., Sun, B., Tong, F., & Bloomquist, J. R. (2014). Neurotoxicity and mode of action of N, N-
536 diethyl-Meta-toluamide (DEET). *PLoS ONE*. https://doi.org/10.1371/journal.pone.0103713

537 Syed, Z., & Leal, W. S. (2008). Mosquitoes smell and avoid the insect repellent DEET. *Proceedings of the*

538 *National Academy of Sciences of the United States of America*.
539 https://doi.org/10.1073/pnas.0805312105

540 Syed, Z., Pelletier, J., Flounders, E., Chitolina, R. F., & Leal, W. S. (2011). Generic insect repellent detector
541 from the fruit fly Drosophila melanogaster. *PLoS ONE*.
542 https://doi.org/10.1371/journal.pone.0017705

543 Tavares, E. M., Judge, B. S., & Jones, J. S. (2019). Bug off! Severe toxicity following inhalational exposure
544 to N, N-diethyl-meta-toluamide (DEET). *American Journal of Emergency Medicine*.
545 https://doi.org/10.1016/j.ajem.2019.04.019

546 Vassar, R., Ngai, J., & Axel, R. (1993). Spatial segregation of odorant receptor expression in the
547 mammalian olfactory epithelium. *Cell*, *74*(2), 309–318. https://doi.org/0092-8674(93)90422-M [pii]

548

549

550 **FIGURE LEGENDS**

551 **Figure 1. A chemical informatics pipeline to screen for pleasant smelling insect repellents**

552 **A**, Overview of the cheminformatics pipeline used to identify novel DEET-like ligands from a larger
553 chemical space. Summary table of top 10 physicochemical features that optimally corelate with
554 repellency. Receiver-operating-characteristic curve (ROC), which indicates the success in predicting test
555 repellents. The performance is the average over 100 80/20 splits, where 80% of the data is for training
556 and 20% for testing. The mean area under the curve (AUC) is provided. Representative structures from
557 the top 150 predicted repellent compounds.

558 **Figure 2. Models successfully predict odor perceptual qualities of repellents in humans**

559 **A,** Validation of models predicting the odor perceptual qualities (146 perceptual descriptors) of a set of
560 compounds that were screened for repellency in Figure 3A. The performance metric for the validation is
561 the area under the ROC curve, which assesses the rate of correctly predicted perceptual descriptors
562 relative the rate of incorrectly predicted descriptors. Dashed line is the average Area Under the Curve
563 (AUC). **B,** Sample ROC curves for select compounds as well as the ROC curve across all compounds (e.g.,
564 shown as separate curves in the ROC plot as well as those not shown but whose identities appear in
565 Figure 2A); the respective AUC values are reported inside the plot. **C,** The top 25 perceptual descriptors
566 for the predicted repellents that underwent experimental validation are displayed as a word cloud. The
567 font size of the descriptor is scaled relative to frequency.

568 **Figure 3. Experimental validation of predicted insect repellents show a high success rate several of**
569 **which are pleasant smelling**

570 **A,** Representative still photograph of landing assay of female *Aedes aegypti* mosquitoes on filter paper
571 covered heat block over 5 mins (overlay of all frames), and **B,** mean repellency index for each of the

572 indicated compounds tested at ~0.2μg/cm^2. The repellency coefficient = number mosquitoes attracted
573 to solvent treated control netting, minus number of mosquitoes on treated netting compared to solvent,
574 with 37C heated rectangle as an attractant. Mosquito numbers are counted for the 5 min duration, over
575 multiple frames, as described in the methods. N=2-4 trials, ~20 mosquitoes/trial. For DEET, N=23 trials.

576 **C,** Mean Preference index of *Drosophila* adults to repellents at three different concentrations dissolved
577 in acetone in a Two-choice trap assay baited with 10% Apple Cider Vinegar measured after 48 hrs. N = 7-
578 10 trials each treatment at 48 hrs, 10 flies/trial, error bars = s.e.m.

579 **Figure 4. Some, but not all insect repellents increase Ca^{2+} mobilization in a human cell line**

580 DEET induces a non-specific calcium repose *in vitro*. **A.** Dose response curves for three mammalian cell
581 lines (HEK293. CHOK1, NG108) and one insect cell line (S2). EC50 and Hill coefficients for each cell line
582 are given below. **B.** Representative DEET induced calcium kinetics in HEK293 cells. **C.** HEK293 cells
583 treated with thapsigargin 10 min before assay to deplete intracellular calcium stores no longer respond
584 to DEET, suggesting DEET response is dependent on intracellular calcium. Error bars are standard
585 deviation.**D. F,** Luminescence of the Ca^{2+} indicator at two different concentrations for predicted and
586 previously known repellent chemicals, scaled relative to DEET.

587 **Figure 5. Screen of 10+ million chemicals predicts several insect repellents with different smells**

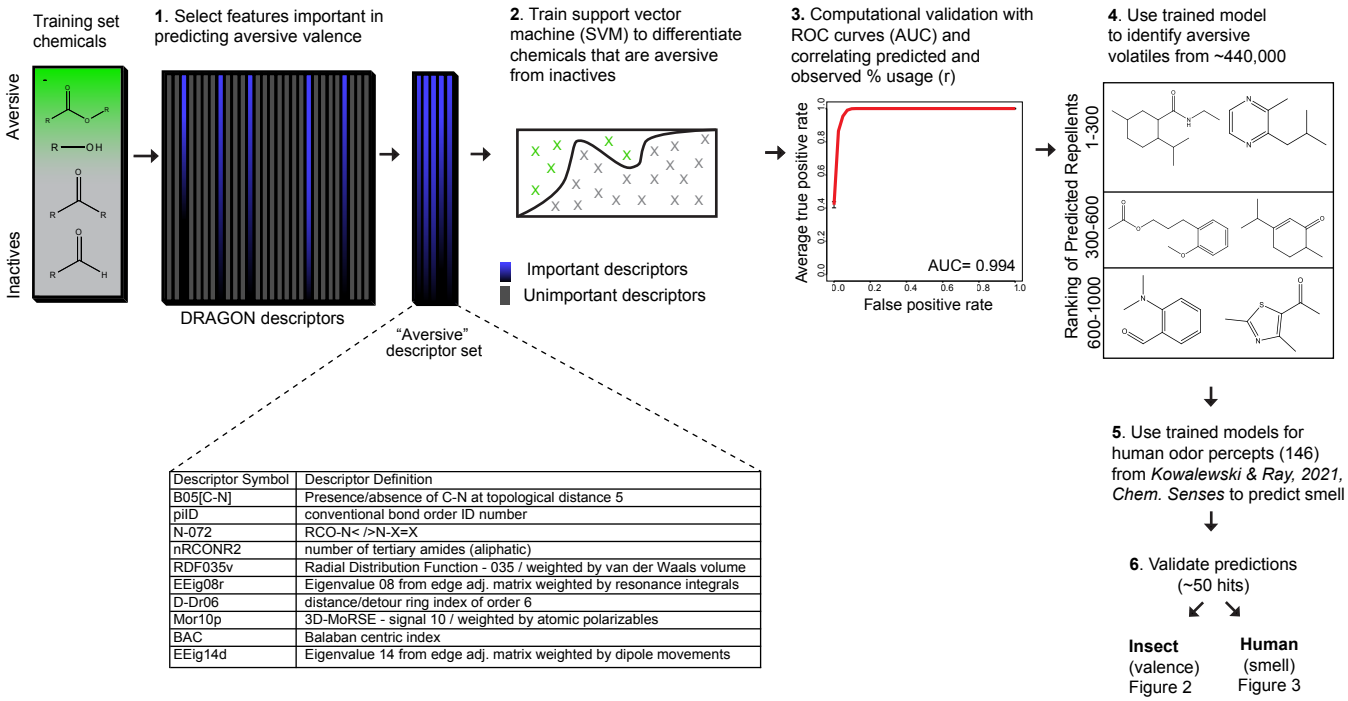
588 **A,** Tabulated predicted % repellency (relative to DEET) from a library of 10+ million chemicals, filtered to
589 the top values. The best matching known repellent is displayed alongside the Euclidean distance, the
590 predicted LD50 and Ca^{2+} mobilization. **B**, heatmap organizing top predictions according to estimated
591 perceptual qualities.

592

593

594 **Supplemental Figure S1**

595 **A**, Overview of the approach to generate and validate the machine learning models. Here, a model-
596 averaged prediction is made, where each model has access to different physicochemical features and is
597 trained on different combinations of training set chemicals. **B,** The R^2 values over the 100 train/test
598 splits, averaged into 10 bins. **C,** The average classification success is reported over 100 train/test splits,
599 assessed by the area under the ROC curve. Active labels (positive cases) were assigned as chemical
600 scoring in the top ~40% of Ca^{2+} values. This is compared to shuffling the labels before training, reported
601 as "Shuffle."



Descriptor Symbol	Descriptor Definition
B05[C-N]	Presence/absence of C-N at topological distance 5
piID	conventional bond order ID number
N-072	RCO-N< />N-X=X
nRCONR2	number of tertiary amides (aliphatic)
RDF035v	Radial Distribution Function - 035 / weighted by van der Waals volume
EEig08r	Eigenvalue 08 from edge adj. matrix weighted by resonance integrals
D-Dr06	distance/detour ring index of order 6
Mor10p	3D-MoRSE - signal 10 / weighted by atomic polarizables
BAC	Balaban centric index
EEig14d	Eigenvalue 14 from edge adj. matrix weighted by dipole movements

Figure 1

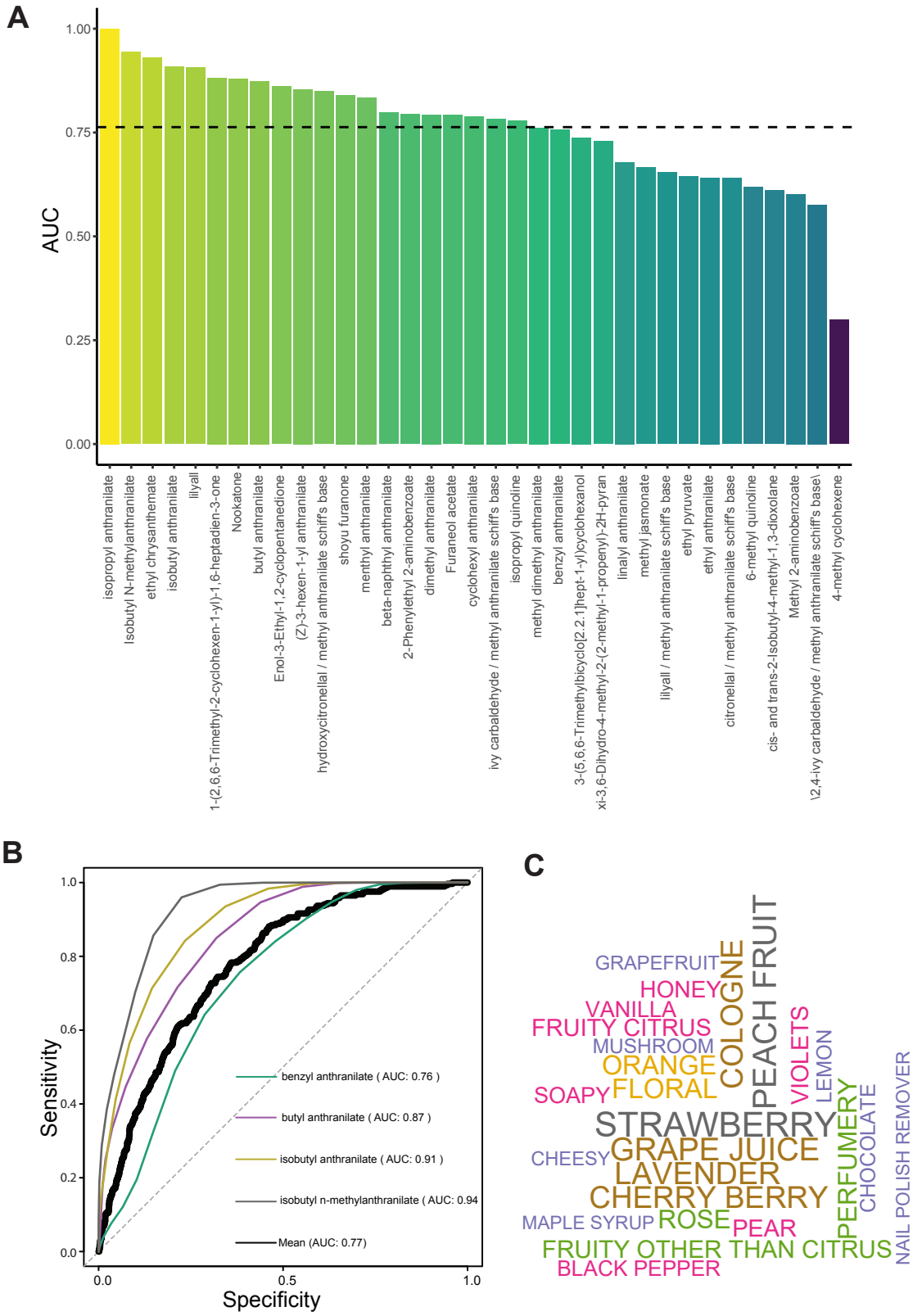

A
AUC
1.00
0.75
0.50
0.25
0.00
isopropyl anthranilate
Isobutyl N-methylanthranilate
ethyl chrysanthemate
isobutyl anthranilate
lilyall
1-(2,6,6-Trimethyl-2-cyclohexen-1-yl)-1,6-heptadien-3-one
Nookatone
butyl anthranilate
Enol-3-Ethyl-1,2-cyclopentanedione
(Z)-3-hexen-1-yl anthranilate
hydroxycitronellal / methyl anthranilate schiff's base
shoyu furanone
menthyl anthranilate
beta-naphthyl anthranilate
2-Phenylethyl 2-aminobenzoate
dimethyl anthranilate
Furaneol acetate
cyclohexyl anthranilate
ivy carbaldehyde / methyl anthranilate schiff's base
isopropyl quinoline
methyl dimethyl anthranilate
benzyl anthranilate
3-(5,6,6-Trimethylbicyclo[2.2.1]hept-1-yl)cyclohexanol
xi-3,6-Dihydro-4-methyl-2-(2-methyl-1-propenyl)-2H-pyran
linalyl anthranilate
methyl jasmonate
lilyall / methyl anthranilate schiff's base
ethyl pyruvate
ethyl anthranilate
citronellal / methyl anthranilate schiff's base
6-methyl quinoline
cis- and trans-2-Isobutyl-4-methyl-1,3-dioxolane
Methyl 2-aminobenzoate
\2,4-ivy carbaldehyde / methyl anthranilate schiff's base\
4-methyl cyclohexene
B
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
benzyl anthranilate (AUC: 0.76)
butyl anthranilate (AUC: 0.87)
isobutyl anthranilate (AUC: 0.91)
isobutyl n-methylanthranilate (AUC: 0.94
Mean (AUC: 0.77)
C
GRAPEFRUIT
HONEY
VANILLA
FRUITY CITRUS
MUSHROOM
ORANGE
SOAPY
FLORAL
COLOGNE
PEACH
FRUIT
VIOLETS
LEMON
STRAWBERRY
CHEESY
GRAPE JUICE
LAVENDER
CHERRY BERRY
MAPLE SYRUP
ROSE
PEAR
PERFUMERY
CHOCOLATE
NAIL POLISH REMOVER
FRUITY OTHER THAN CITRUS
BLACK PEPPER

Figure 2

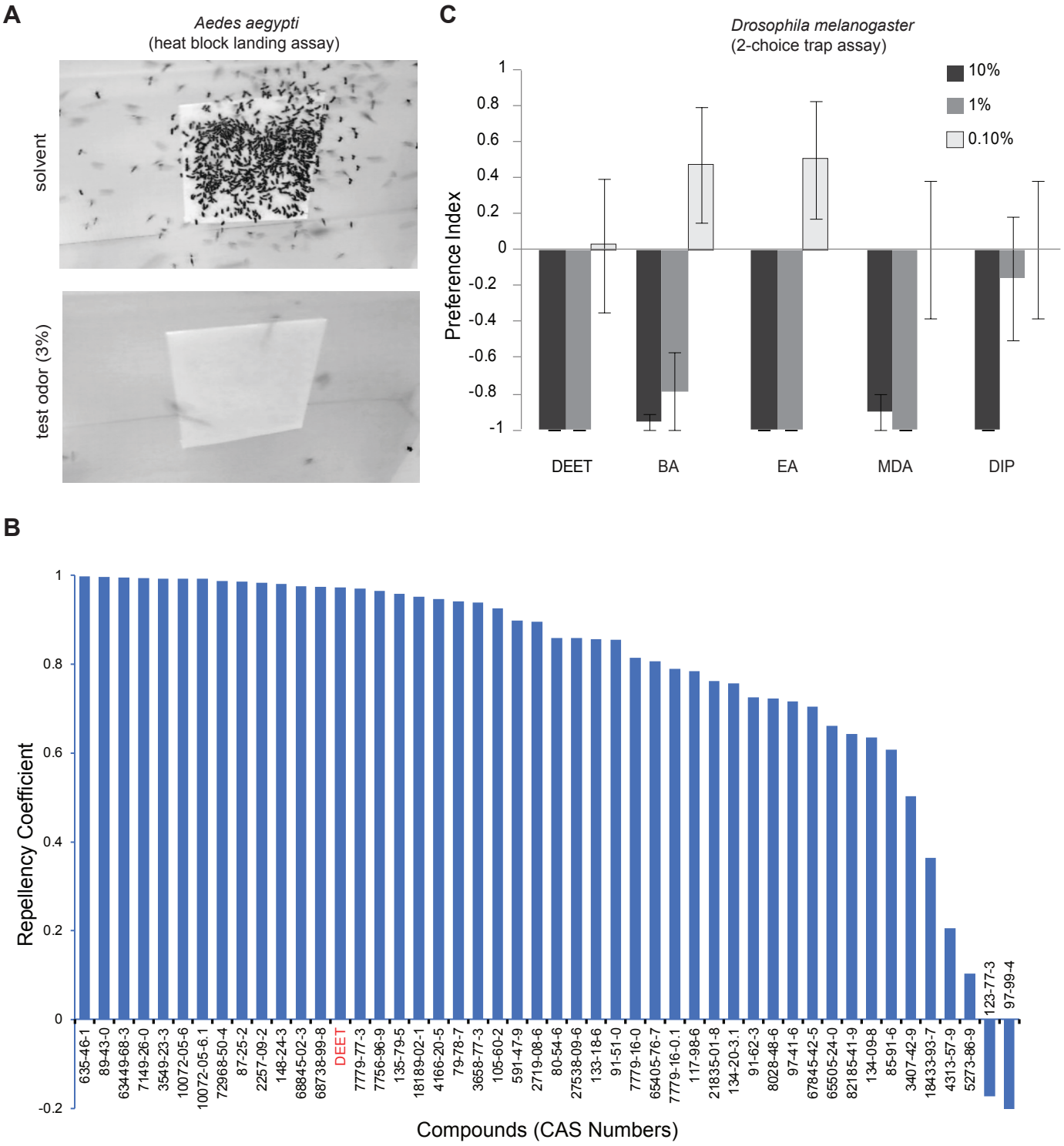
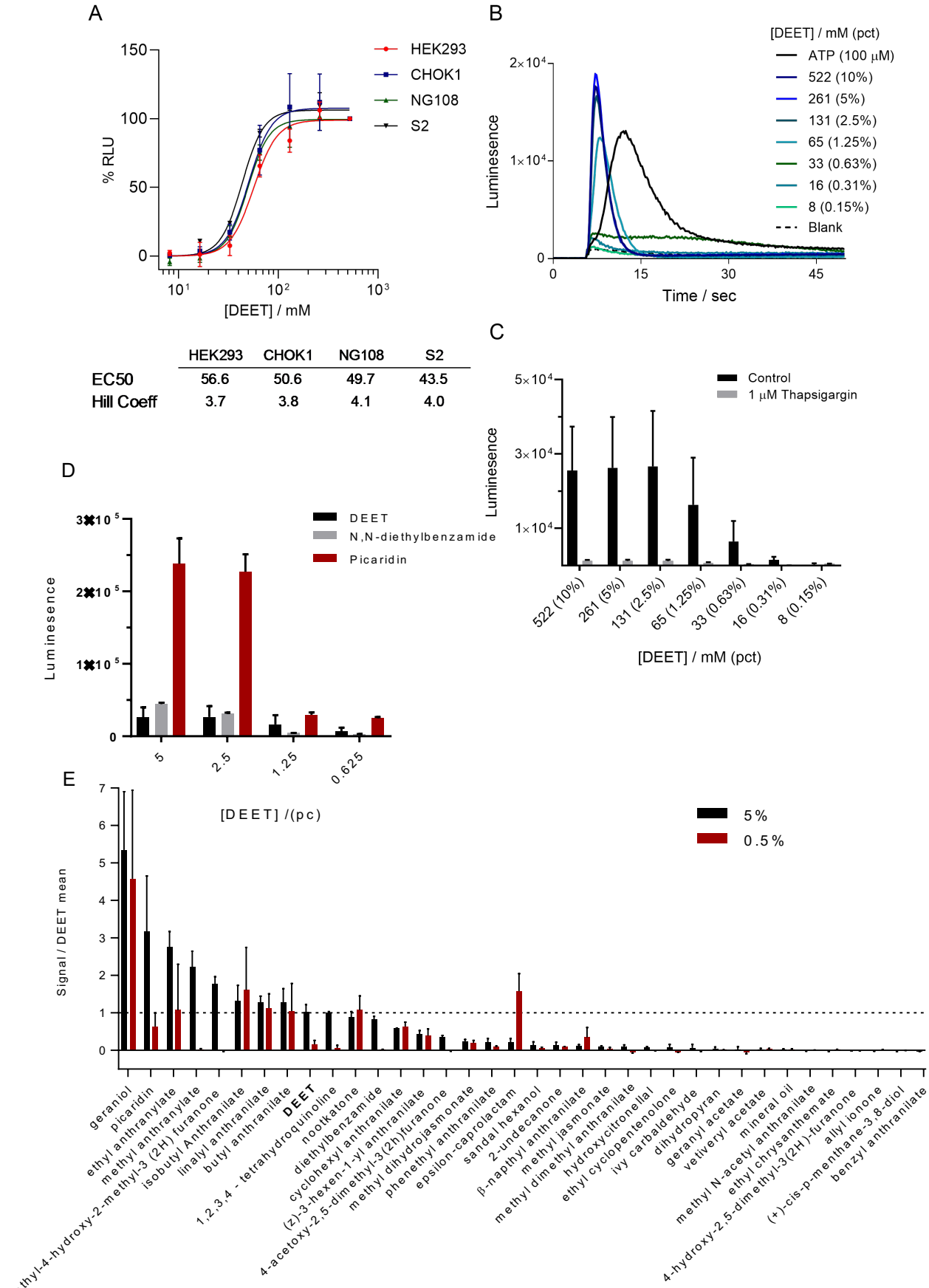

A
Aedes aegypti
(heat block landing assay)
solvent
test odor (3%)
C
Drosophila melanogaster
(2-choice trap assay)
10%
1%
0.10%
Preference Index
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
DEET
BA
EA
MDA
DIP
B
Repellency Coefficient
1
0.8
0.6
0.4
0.2
0
-0.2
635-46-1
89-43-0
63449-68-3
7149-26-0
3549-23-3
10072-05-6
10072-05-6.1
72968-50-4
87-25-2
2257-09-2
148-24-3
68845-02-3
68738-99-8
DEET
7779-77-3
7756-96-9
135-79-5
18189-02-1
4166-20-5
79-78-7
3658-77-3
105-60-2
591-47-9
2719-08-6
80-54-6
27538-09-6
133-18-6
91-51-0
7779-16-0
65405-76-7
7779-16-0.1
117-98-6
21835-01-8
134-20-3.1
91-62-3
8028-48-6
97-41-6
67845-42-5
65505-24-0
82185-41-9
134-09-8
85-91-6
3407-42-9
18433-93-7
4313-57-9
5273-86-9
123-77-3
97-99-4
Compounds (CAS Numbers)

Figure 3



	HEK293	CHOK1	NG108	S2
EC50	56.6	50.6	49.7	43.5
Hill Coeff	3.7	3.8	4.1	4.0

Figure 4

A

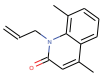
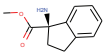
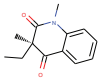
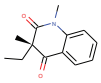
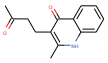
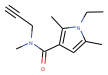
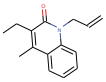
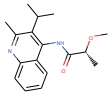
Image	ZINC_ID	Score	LD_{50} mg/kg	Ca^{2+} mobilization
	ZINC000001394030	110.9316	1293.0062	0.9054749
	ZINC000034195873	110.5746	711.4268	0.7932151
	ZINC000005327991	110.4276	883.9981	0.7673523
	ZINC000094992683	110.4188	1067.0473	0.9159867
	ZINC000001735256	110.4169	883.8132	0.7749915

Image	ZINC_ID	Score	LD_{50} mg/kg	Ca^{2+} mobilization
	ZINC000008609700	110.2547	1287.3842	0.5077889
	ZINC000253417706	109.7302	933.5598	0.8336481
	ZINC000034195871	109.6462	684.8090	0.9535496
	ZINC000000506855	109.5113	1299.5812	0.7571195
	ZINC000191321911	109.5082	989.3195	1.1121364

B

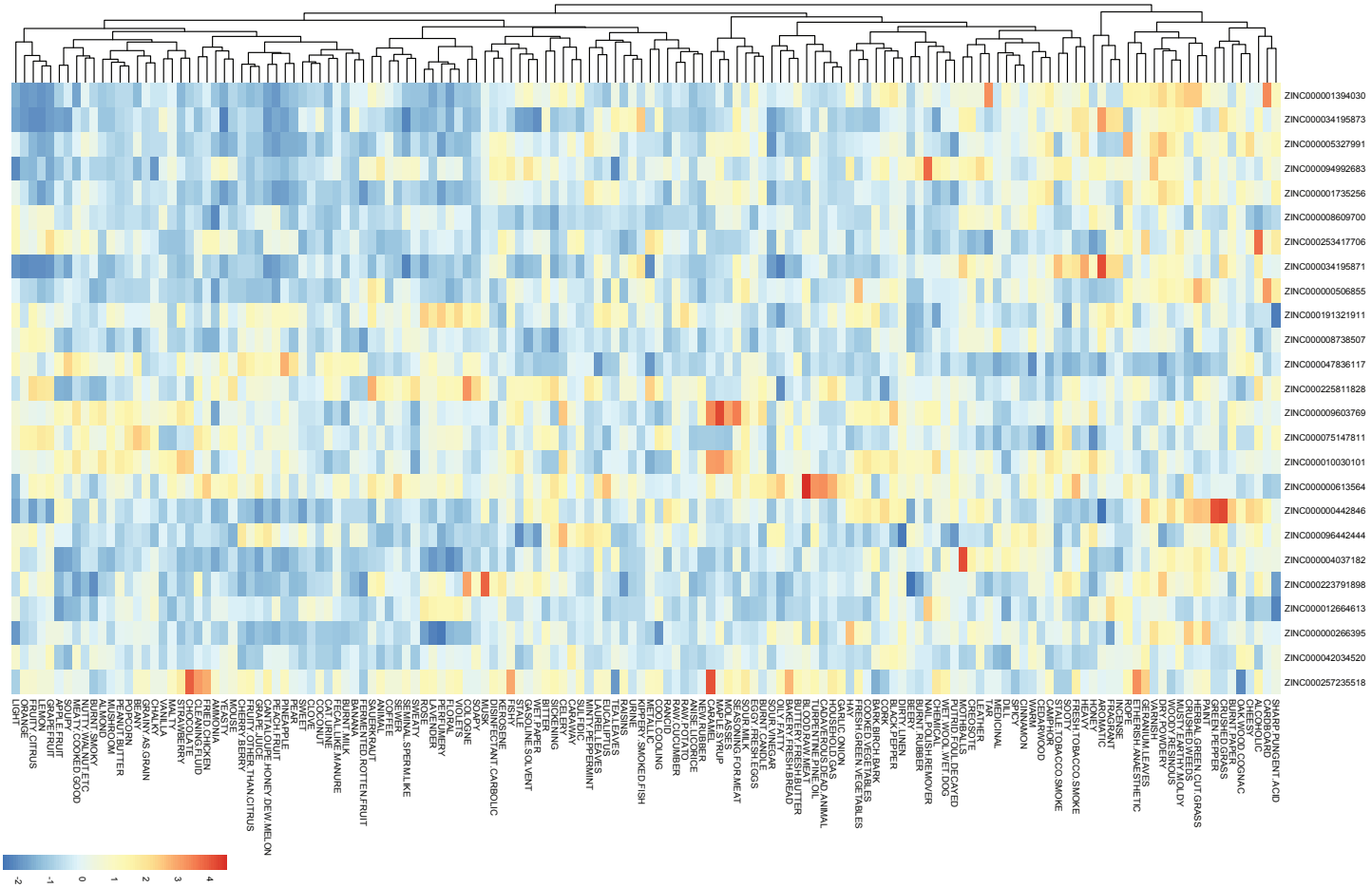


Figure 5