Cell Painting and chemical structure read-across can complement each other for rat acute oral toxicity prediction in chemical early de-risking

Fabrice Camilleri[1,2], Joanna M. Wenda[3], Claire Pecoraro-Mercier[3], Jean-Paul Comet[2], David Rouquié[1,*,]

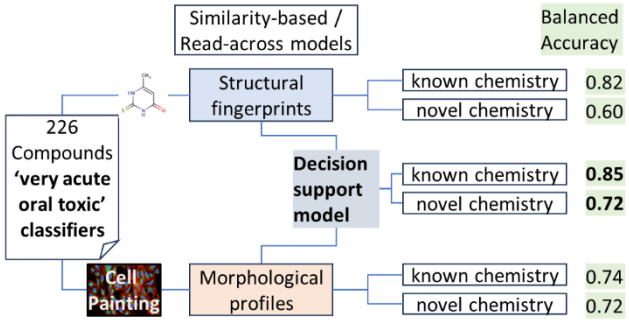
[1] Toxicology Data Science, Bayer SAS Crop Science Division, 355 rue Dostoïevski, CS 90153 Valbonne, 06906 Sophia Antipolis Cedex, France

[2] I3S UMR 7271 du CNRS, Université Côte d'Azur, Bâtiment Algorithme-Euclide-B, 2000 Route des Lucioles, B.P. 121, 06903 Sophia Antipolis, France

[3] Early Toxicology, Bayer SAS Crop Science Division, 355 rue Dostoïevski, CS 90153 Valbonne, 06906 Sophia Antipolis Cedex, France

*: david.rouquie@bayer.com

Keywords: Rat acute oral toxicity, Cell Painting, QSAR, NAMs, Morphological profiles, LD50, KNN, Applicability Domain, compound derisking, Read-Accross.



For Table of Contents Only

ABSTRACT,

Early de-risking decisions in the development of new chemical compounds, enable the identification of novel chemical candidates with improved safety profiles. In vivo studies are traditionally conducted in the early assessment of acute oral toxicity of crop protection products to avoid compounds which are considered "very acutely toxic", with an in vivo Lethal Dose of 50% (LD50) ≤ 60 mg/kg bodyweight. Those studies are lengthy, costly, and raise ethical concerns, catalyzing the use of non-animal alternatives. The objective of our analysis was to assess the predictive efficacy of read-across approaches for acute oral toxicity in rats, comparing the use of chemical structure information, in vitro biological data derived from the Cell Painting profiling assay on U2OS cells or the combination of both. Our findings indicate that the classification of compounds as very acute oral toxic (LD50 ≤ 60 mg/kg) or not is possible using a read-across approach, with chemical structure information, morphological profiles, or a combination of both. When classifying compounds structurally similar to those in the training set, chemical structure was more predictive (balanced accuracy of 0.82). Conversely, when the compounds to be classified were structurally different from those in the training set, the morphological profiles were more predictive (balanced accuracy of 0.72).

Combining the two models allowed for the classification of compounds structurally similar to those in the training set, to slightly improve the predictions (balanced accuracy of 0.85).

44 **Abbreviation table**

AD	Applicability Domain
BA	Balanced Accuracy
BCS	Bayer Crop Science
DMSO	Dimethyl sulfoxide
DS	Decision Support
GHS	Global Harmonized System
KNN	K Nearest Neighbor
LD50	Median Lethal Dose
MAD	Median Absolute Deviation
MCC	Matthews Correlation Coefficient
MIE	Molecular Initiative Event
MoA	Mode of Action
NGRA	Next Generation Risk Assessment
NVAOT	Non Very Acutely Oral Toxic (LD50 > 60 mg/kg)
QSAR	quantitative structure activity relationship
SMILES	Simplified molecular-input line-entry system
SN	Sensitivity
SP	Specificity
TP, TN, FP, FN	True Positives, True Negatives, False Positives, False Negatives

VAOT	Very Acutely Oral Toxic (LD50 ≤ 60 mg/kg)

52 **Introduction**

53 Small molecule discovery involves identifying and developing novel chemical compounds with
54 optimized safety profiles across various target and non-target species, including laboratory
55 animals, humans and environmental species. This complex process requires the integration of
56 chemical and biological exploration. Chemists design diverse compounds to interact with
57 specific biological targets or pathways while biological assays assess efficacy and safety, guiding
58 further chemical optimization. Balancing potency with selectivity and minimizing off-target
59 effects presents a significant challenge. To achieve this, chemists often explore new chemical
60 spaces to discover novel structures with the desired properties[1,2]. Additionally, the complexity
61 of biological systems poses hurdles, as the interplay of various factors influences a molecule's
62 properties. Success in small molecule discovery hinges on a multidisciplinary approach, where
63 chemists, biologists and data scientists collaborate to navigate the intricate landscape of
64 chemical and biological interactions, ultimately advancing the development of new active
65 substances. In agrochemical discovery, early de-risking is crucial involving systematic
66 assessment of compounds for potential safety issues to be addressed as early as possible.
67 Addressing genotoxicity and acute oral toxicity is essential due to established cut-off criteria [3].

68 Traditionally, early genotoxicity evaluations are performed with in vitro test methods (Ames
69 and Micronucleus assays) whereas acute oral toxicity profile screening involve laboratory
70 animal testing in rodent to obtain a first estimate of the LD50 representing the single dose level
71 at which lethality is induced in at least in 50 % of the tested animals. The traditional in vivo
72 studies for acute oral toxicity assessment are time-consuming, expensive, and have low
73 throughput, making it challenging to test a large number of chemicals. Animal studies also raise
74 ethical concerns and must be reduced or if possible, eliminated. Higher throughput alternatives
75 are desired, for the ranking of chemical candidates for their prioritization into the R&D pipeline
76 based on their toxicological profiles. This minimizes resource waste, specifically extended R&D

77 efforts for a chemical which is otherwise only later discovered to not meet required safety
78 standards.

79 Non-animal alternatives are already available to reduce in vivo studies including in vitro
80 approaches and in silico models for early estimation of the LD50. For instance, the in vitro 3T3
81 neutral red uptake assay (NRU) [4] can serve to categorize compounds with in vivo LD50 greater
82 or lower than 2000 mg/kg. However, this LD50 threshold is high and may not discriminate
83 compounds with lower LD50s. In acute toxicity testing, low LD50 value (LD50 ≤ 60mg/kg),
84 means high potential for acute toxicity, which is an unwanted safety profile[3].

85 In silico models are emerging which may close this gap. Quantitative structure activity
86 relationship (QSAR) models rely on machine learning based on structural information of
87 chemical compounds. One notable example is the Collaborative Acute Toxicity Modeling Suite
88 (CATMoS), a public QSAR model, built in a collaboration of several research groups. CATMoS
89 was trained on more than 10,000 compounds and demonstrated high performance in
90 predicting acute oral toxicity in rats [5]. As a regression model, CATMoS can predict the LD50 and
91 classify chemicals into the five categories of Global Harmonized System (GHS). However, it is
92 important to note that QSAR model predictions are only reliable within their applicability
93 domain namely the chemical space covered by the chemistry represented in the training set.
94 This represents a significant limitation when using such models to predict properties of novel
95 chemistry not represented in the training set.

96 In this situation, we hypothesized that using in vitro high biologically dense profiles could be
97 used instead of chemical structure information, especially when exploring new chemical spaces.
98 The Carpenter–Singh Lab at the Broad institute has developed an such in vitro high content
99 biologically profiling assay, Cell Painting, which capture the morphological information of cells
100 perturbed by chemicals [6]. The primary advantage of Cell Painting lies in its untargeted nature,
101 theoretically allowing it to capture any bioactivity inducing a change in cell morphology.
102 Additionally, this assay is more cost-effective than other profiling assays such as

103 transcriptomics[7]. Cell Painting has already been used successfully, in 'hit' discovery and in Mode
104 of Action (MoA) prediction [8–10]. In the field of toxicology, the US Environmental Protection
105 Agency (EPA) has explored its use to screen bioactive compounds for human risk assessment [11].
106 More recently, Cell Painting has also been utilized for the prediction of mitochondrial
107 toxicity[12,13] , and liver toxicity[14].

108 To evaluate if biological information could complement compound structure predictions for
109 acute oral toxicity especially when exploring new chemical spaces, classifiers were employed to
110 predict whether compounds had very high acute oral toxicity (LD50 ≤ 60 mg/kg) or not (LD50 >
111 60 mg/kg). Initially, a well-performing public QSAR model, CATMoS, was utilized for the
112 prediction of a set of Bayer Cros Science compounds. Secondly, a simple chemical structural
113 similarity-based classifier (utilizing a K nearest neighbor [15]) was employed for the prediction of
114 the same set of compounds. This approach, also known as the read-across approach, is
115 commonly used in toxicology specifically in the case of tox data poor chemicals, such as REACH
116 compounds with limited number of in vivo results[16]. To simulate scenarios where predictions
117 using the models are made on novel chemistry, a "novel chemistry" holdout strategy was
118 created to assess the classifier. To verify the efficacy of this holdout strategy, the chemical
119 structural similarity-based classifier was again evaluated. A Cell Painting assay was conducted
120 on a smaller subset of compounds using the U2OS cell line to evaluate read-across using
121 morphological profiles. The results of the two similarity-based classifiers were assessed: one
122 based on chemical structure, and one based on morphological profiles derived from Cell
123 Painting. Both classifiers were also tested in the context of the "novel chemistry" holdout
124 strategy. Finally, a decision support model was constructed to determine which of the two
125 similarity-based classifiers (chemical structure or morphological profile) should be
126 recommended for prediction of acute toxicity.

127 Overall, our results showed that the classification of compounds as very acute oral toxic (LD50 ≤
128 60 mg/kg) is possible using a read-across approach, with chemical structure information,
129 morphological profiles, or a combination of both. When classifying compounds structurally

130 similar to those of the training set, chemical structure was more predictive (balanced accuracy
131 of 0.82). Conversely, when the compounds to be classified were structurally different from
132 those of the training set, the morphological profiles were more predictive (balanced accuracy of
133 0.72). Combining both models allowed for the classification of compounds structurally similar
134 to those used to train the classifiers, to slightly enhance the predictions (balanced accuracy of
135 0.85).

136

Materials and methods

Acute oral toxicity compound classes

The compounds were divided into two classes. The class designated as "Very acutely oral toxic" (abbreviated VAOT) included compounds with a lethal dose (LD50) of 60 mg/kg or less. The class designated as "Not very acutely oral toxic" (abbreviated NVAOT) included compounds with a lethal dose (LD50) greater than 60 mg/kg. (Table 1).

Acute oral toxicity classes	
Very acutely oral toxic – VAOT	Rat oral LD50 ≤ 60 mg/kg
Non very acutely oral toxic - NVAOT	Rat oral LD50 > 60 mg/kg

Table 1

Definition of the two acute oral toxicity classes.

Compound selection

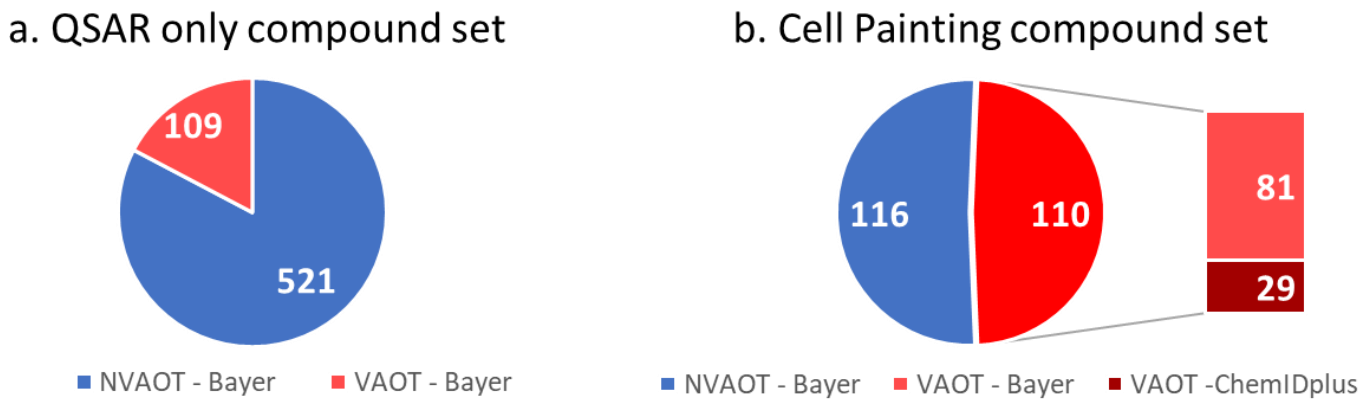
154 Compounds with acute oral toxicity results in rats

155 To select compounds, Bayer internal databases were queried. A total of 765 compounds with in
156 vivo rat acute oral toxicity results were found for two doses: 60 mg/kg and 300 mg/kg. Of the
157 765 compounds, 109 compounds were identified as acute toxic at the dose of 60 mg/kg,
158 meaning compounds belonging to the VAOT class, and 521 compounds were not toxic at 300
159 mg/kg, meaning compounds belonging to the NVAOT class. To create a robust contrast
160 between the morphological profiles of the VAOT and NVAOT classes, we excluded 135
161 compounds that were acute toxic at 300 mg/kg but not at 60 mg/kg. This resulted in the first
162 dataset of 630 compounds, which were used to test the acute toxicity prediction of the
163 collaborative Acute Toxicity Modeling Suite (CATMoS)[5] and to build a chemical structure
164 similarity-based classifier.

165 For the Cell Painting campaign, we checked which of the previous dataset compounds were
166 available in Bayer internal compound repository, 81 VAOT compounds were found. To complete
167 the list of VAOT compounds, we queried the chemIDplus public database [17], and selected 29
168 compounds that were available in Bayer compound logistics, making a total of 110 VAOT
169 compounds. To have a balanced dataset, 116 NVAOT compounds were selected. To have a
170 good chemical structure diversity among them, the Butina algorithm[18] was used to cluster the
171 521 compounds of the previous dataset, based on the Tanimoto similarity of their Morgan
172 fingerprint and a threshold of 0.7: a maximum of clusters was selected and the number of
173 compounds coming from the same cluster were minimized (supplementary information Figure
174 S1). This selection resulted in having a total of 116 NVAOT compounds and 110 VAOT.

175 To summarize, two sets of compounds were defined. The first one, called 'QSAR only compound
176 set', was a set of 630 compounds; 109 were VAOT and 521 were NVAOT. The second one,
177 called 'Cell Painting compound set', was a set of 226 compounds; 110 were VAOT, and 116
178 were NVAOT (Figure 1).

179



180

181 *Figure 1*

182 *Composition of the two datasets by compound class: VAOT (Very Acutely Oral Toxic) and NVAOT (Non Very Acutely Oral Toxic);*
183 *and by source: Bayer and ChemIDplus.*

184 *a: QSAR only compound set. Set used for QSAR classifiers only.*

185 *b: Cell Painting compound set, Set used for chemical structure and morphological profile classifiers.*

186

187 Negative and positive controls

188 For the Cell Painting assay, DMSO (dimethyl sulfoxide) only (0.1%) was used as the negative
189 control.

190 To monitor the Cell Painting assay's performance and assess the quality of experiment's
191 replicates, a set of positive controls was used. These were compounds inducing reproducible
192 and distinct morphological profiles in U2OS cells. The selection was based on published
193 literature and pilot tests in our lab (Supplementary information).

194

CATMoS QSAR model

As a first attempt, the acute oral toxicity class of the compounds was predicted using the collaborative Acute Toxicity Modeling Suite (CATMoS) QSAR model, implemented in the OPERA (version 2.9) [5,19] QSAR suite. To align with the two classes presented in this paper (VAOT and NVAOT), three CATMoS predictions were needed: the EPA classification, the GHS classification and the LD50 range estimation. All compounds being classified by CATMoS as EPA category 1 (LD50 ≤50 mg/kg), or GHS category 1 (LD50 ≤ 5 mg/kg), or GHS category 2 (5 mg/kg < LD50 ≤ 50 mg/kg) were designated as VAOT. For compounds being classified as EPA category 2 (50 mg/kg < LD50 ≤ 500 mg/kg), or GHS category 3 (50 mg/kg < LD50 ≤ 300 mg/kg), the lower limit of the LD50 range predictions was considered: if the inferior limits were smaller than 60, the compounds were classified as VAOT. For all other predictions, the compounds were classified as NVOAT.

The Opera implementation of CATMoS provides three prediction reliability metrics [19] that were used to understand the predictions made by the model. A global Applicability Domain Boolean value is calculated, indicating if a compound falls within the training set chemical space. Additionally, an Applicability Domain index is calculated, ranging from zero to one, revealing proximity of the queried compound to the training set. This index is relative to the similarity of the query chemical to its five nearest neighbors [19]. In particular, a query chemical compound can belong to the CATMoS AD (global AD = 1), but can also be in a 'gap' of the training chemical space (AD index <0.6)[19]. In such cases, the predictions should be considered with caution[19]. Finally, a confidence index is computed, informing on the accuracy of the prediction of the neighbors of the queried compound.

Cell painting campaign

A Cell Painting campaign was conducted in our laboratory to obtain the morphological profiles of our 'Cell Painting compound set' (set of 226 compounds). The Cell Painting Protocol v3 of the

220 Broad Institute was utilized on U2OS human osteosarcoma cells with four biological replicates
221 [20].

222 A previous Cell Painting pilot conducted at the unique dose of 10 µM demonstrated that few
223 agrochemical compounds exhibited morphological changes compared to the negative control
224 (see the 'Morphological change signal measure' section, and the 'Biological response' section).
225 Therefore, in this campaign, to enhance the likelihood of capturing a morphological response,
226 the compounds were screened at three concentrations: 10µM, 31.6µM and 100µM.

227 Cell Culture and Seeding

228 Human osteosarcoma cells U2OS have been purchased from ATCC (ref.: HTB-96, lot: 70025046).
229 The McCoy's 5A Modified Medium with GlutaMAX™ Supplement (Thermo Fisher, ref:
230 36600021) supplemented with 10% Fetal Bovine Serum (Gibco, ref.: 16000044) and
231 penicillin/streptomycin mix (Sigma Aldrich, ref: P4458) was used for culturing cells in T75 or
232 T175 flasks in a standard humidified incubator (37°C, 5% CO2). The passages were performed
233 when the culture achieved about 80% confluency. Trypsin (Thermo Fisher, ref. 25200056) was
234 used to detach the cells during passage and the number of live cells was calculated with an
235 automatic cell counter (Countess II, Thermo Fisher) after staining the cells with trypan blue
236 (Sigma, ref.: T8154). For the creation of a cell bank, the vial with frozen cells received from the
237 supplier was thawed and expanded until internal passage no. 3 (P3). At this stage the cells were
238 cryopreserved in complete media supplemented with 10% DMSO in an ultra-low temperature
239 freezer (-150°C) creating a master bank. One vial of master bank was then thawed, expanded
240 until internal passage no. 6 (P6), and cryopreserved as before to create a working bank. Vials of
241 the working bank were then directly used for seeding the microplates. One vial of cells
242 (containing 4 million cells) was removed from -150°C freezer and thawed in the water bath. The
243 contents of the vial were immediately added to 10 ml of pre-heated complete media and
244 centrifuged (5 min, 120xg). After removing the supernatant, the cell pellet was resuspended in
245 10 ml of complete medium through thorough pipetting. The cell suspension was then added to
246 150 ml of medium in a round bottle with a magnetic stirrer and immediately used for seeding

247 the 384-well microplates (Greiner BioONE CELLSTAR µCLEAR®; ref: 781091). Multidrop (Thermo
248 Fisher) was used to automatically distribute 36 µl of cell suspension per well, resulting in a
249 seeding density of around 900 cells/well. The cells were then incubated at 37°C, in an
250 atmosphere of 5% CO2 in an automatised incubator (Cytomat 2, Thermo Fisher). All
251 experimental replicates were performed on a different day, using a separate cell vial originating
252 from the same working bank (P6).

253 Chemical treatment

254 The test compounds were received in powder form in 96-well deep well plates. They were then
255 dissolved in DMSO (dimethyl sulfoxide) to create 100 mM stock solutions, aliquoted in 96-well
256 V-bottom plates (V96 PP Plate, Thermo Fisher) and frozen at -20°C until the day of the
257 treatment. Every biological replicate of the experiment originates from a separate aliquot of the
258 stock solution plate, so that the compounds undergo only 1 freeze-thaw cycle. On the day of
259 the treatment (24 h post-seeding), the plates containing stock solutions were thawed and the
260 compounds were diluted in DMSO to create dose plates containing 3 concentrations per
261 compound: 100 mM, 31.6 mM, and 10 mM. The dilutions were performed with the use of Viper
262 liquid handler (Synchron). The compound solutions form the dose plates were then
263 administered to the cell plates in a two-step process. Firstly, an intermediate dilution was
264 prepared: 1 µl of the compound solution was diluted in 100 µl of complete cell medium (1:100
265 dilution), next 4 µl of the resulting intermediate solution was administered to the cell plate (4 µl
266 of the diluted compound into 36 µl of cell media, 1:10 dilution). The final concentrations of
267 compounds that the cells were exposed to were therefore: 10 µM, 31.6 µM and 100 µM, the
268 final vehicle (DMSO) concentration was 0.1%. The treated cell plates were subsequently
269 incubated with the compounds for 48 h.

270 Staining

271 The staining and fixation were performed following the published protocol [20] with the use of
272 PhenoVue JUMP kit (Perkin Elmer, ref.: PING23). Briefly, 20 µl/well of the Mitotracker solution

273 were distributed to the cell plates with Multidrop (final concentration: 500 nM). After 30
274 minutes of incubation at 37°C, 20 µl/well of 16% PFA solution (Thermo Fisher, ref.: 28908) were
275 added. The fixation was performed at room temperature (25°C). Two washes with HBSS buffer
276 (Gibco, ref.: 14065-056) were performed with the aid of Mutlifo washer (BioTek). 20 µl/well of
277 the staining solution (HBSS, 1% BSA, 0.1% Triton X-100, 43.7 nM PhenoVue Fluor 555 – WGA; 48
278 nM PhenoVue Fluor 488 - Concanavalin A; 8.25 nM PhenoVue Fluor 568 – Phalloidin; 1.62 µM
279 PhenoVue Hoechst 33342 Nuclear Stain; 6 µM PhenoVue 512 Nucleic Acid Stain) were added
280 and the plates were incubated for 30 min at room temperature before being washed again
281 three times with HBSS. The plates were then sealed with aluminium foil and images were
282 recorded directly.

283

284 Morphological profile generation

285 Image acquisition

286 ImageXpress Micro 4 epifluorescent microscope (Molecular Devices) with 20x air objective was
287 used for recording the fluorescent images (16-bit). The camera binning was set to 2x2. The total
288 imaged area per well spanned 2163 µm x 2163 µm and consisted of 3 by 3 adjacent fields of
289 view placed in the centre of the well. For each field of view images were recorded in 5 channels.
290 The following filter sets were used: DAPI, GFP, Cy3, Texas Red, Cy5. The Z-offset and exposure
291 times were set separately for each channel. A total of 207,360 images were acquired in this
292 campaign.

293 Feature extraction

294 Morphological features were extracted using CellProfiler (version 4.2.1), the cell analysis
295 software developed by the Broad Institute [21]. Two CellProfiler pipelines were used: one pipeline
296 for image illumination correction, and one pipeline for the image analysis. The image
297 illumination correction works at the plate level and averages the intensity of the images of each
298 channel. The image analysis pipeline, segmented objects on each image, labelled them

299 according to the channel they were segmented on, and made thousands of measurements on
300 those objects at the cell level. It also took measurement at the image level. A total of 4,761
301 features were measured and formed the morphological profile of a given cell.

302 Aggregation and normalization

303 After extracting the cell morphological profiles with CellProfiler, features were aggregated at
304 the well level, by taking the means of each feature.

305 The features were then normalized, following the Broad Institute approach[6], using the "mad
306 robustized" method of the Python pycytominer package provided by the Broad Institute [22]. The
307 normalization process involved calculating the median of all wells on a plate for each feature,
308 subtracting this value from the median absolute deviation (MAD) of the wells on the plate, and
309 then multiplying the result by 1.4826 to obtain an unbiased estimator. To avoid a null
310 denominator when the MAD was null, a value of 10^{-18} was added to the MAD.

311 $$x_{norm} = \frac{x_{well} - median\,(x_{Plate})}{mad\,(x_{plate}) + \varepsilon}$$

312 With: x_{norm} the normalized value of a morphological profile feature

313 x_{well} the value of a morphological profile feature

314 x_{Plate} the values of a morphological profile feature of a plate

315 $$mad\,(x_{Plate}) = 1.4826 \times \text{median}\,(|x_{Plate} - \text{median}(x_{Plate})|)$$

316 ε an infinitesimally small positive quantity (10^{-18}) to avoid a null denominator.

317

318 Quality check

319 To assess the quality of the Cell Painting experiment, several metrics were calculated. First, the
320 number of cells of the negative control treatment (DMSO) was monitored, which was an output

321 of the CellProfiler segmentation. The cell numbers should fall within the range of [1800; 3000]
322 cells per well. The coefficient of variation of the number of cells for the negative control
323 treatment should not exceed 15% per plate. All plates met the initial quality control standard.

324 To identify any other potential technical issues with the experiment, the Pearson correlations of
325 positive controls across plates were calculated. The positive controls were selected to elicit very
326 distinct and reproducible morphological profiles and were included in each plate. In a well-
327 executed experiment, the replicates of these treatments should be well correlated. A
328 correlation threshold of 0.8 was set for the Pearson correlation between replicates. Replicates
329 were considered non well correlated if the correlation fell below this threshold. No outlier
330 plates were identified at this stage, and all plates passed this quality control step.

331 There were 10 missing feature values at the well level in the morphological profiles. The
332 majority of these values originated from the "Cells_AreaShape_FormFactor" feature. This
333 feature was removed, along with two wells, in order to remove all missing values.

334 Further outliers were identified based on the number of cells within a group of replicates. Some
335 wells exhibited a significant discrepancy in cell counts, with values exceeding 1800 cells,
336 compared to other replicates of the same treatment. Twenty-two wells were identified as
337 outliers and removed from the analysis.

338

339 Unsupervised feature selection

340 To reduce the dimensionality of the normalized morphological profiles, an unsupervised feature
341 selection was performed with the “feature_select” function of the pycytominer Python package
342 [22].

343 This function performed several steps to select the features. First, highly correlated features
344 were removed. For a pair of features with a Pearson correlation greater or equal to 0.9, the
345 feature with the smallest sum of correlations with other features were removed. Second,

346 features with low variances were removed. For a given feature, if the count of the second most
347 common feature value divided by the count of the most common feature value was lower than
348 0.05, the feature was removed. Furthermore, features with the ratio defined as the number of
349 unique feature values divided by the number of samples, below 0.01, were excluded. Third, a
350 list of features (contained in the package), that are known to be noisy and generally unreliable
351 were removed. Fourth, features with at least one absolute value greater than 500, values
352 considered as outliers, were not retained. Finally, within each treatment group any features
353 with a standard deviation greater than 1.2 were removed. This was done to identify and
354 remove any noisy features. This process resulted in a total of 644 features, which were then
355 used for downstream analysis.

356

357 ## Consensus profiles

358 For a given treatment, in our case a chemical compound at a given concentration, the
359 consensus profiles were obtained by aggregating the replicates. This was done by taking the
360 median values of the replicates for each of the remaining features after the unsupervised
361 feature selection.

362

363 ## Morphological change signal measure

364 We utilized the Grit score[23,24], a metric developed by the Broad Institute, to measure the
365 morphological changes in a treatment replicate relative to the negative control treatment
366 (DMSO). To calculate this metric, different Pearson correlation coefficients were calculated.
367 First, the correlations between the morphological profiles of a given treatment replicate and
368 each of the negative control treatments were calculated. The distribution of those correlations
369 was defined by its mean and its standard deviation. Subsequently, the correlations between the
370 morphological profiles of a given treatment replicate and other replicates of this treatment

371 were calculated. Each of the previous correlation coefficients was z-transformed using the
372 distribution of the correlations with the negative controls. The mean was subtracted, and the
373 results were divided by the standard deviation. The grit scores were then obtained by taking
374 the mean of the transformed values.

375 The grit score indicated how much a given replicate profile deviated from the negative control
376 profiles. A high grit score indicated a high deviation of the profile from the negative control
377 profiles.

378 Median grit scores for a given treatment were also calculated, taking the median of the
379 treatment replicate grit scores. This median grit score value allowed measuring how much a
380 treatment impacts the morphology of U2OS cells, compared to the negative controls (DMSO).

381 A threshold of 1 was set to indicate a treatment that induces a morphological change compared
382 to the negative control. A grit score of 1 means that the correlation of the morphological profile
383 of a treatment to its replicates, is one standard deviation away from the mean of its correlation
384 with the negative control profiles.

385

386 ## Molecular fingerprints

387 The compound structures were extracted from the Bayer database as SMILES format (Simplified
388 Molecular-Input Line-Entry System). To reproduce the case when new chemical structures fall
389 outside the applicability domain of chemical structure based predictive models the Morgan
390 fingerprints were employed to describe the chemical structures of the compounds. To ensure
391 sufficient Butina clusters for our data holdout strategy (see 'Chemical compounds clustering
392 with Butina' and 'Dataset splits section'), we selected a Morgan fingerprint on 1024 bits to
393 provide the most detailed chemical structure description and thus differentiate more structural
394 differences [18,25,26]. To obtain them from the SMILES, we performed the following steps using the
395 RDKIT Python package [27].

396 First, the SMILES were cleaned using the MolStandardize module of RDKIT. This involved
397 removing hydrogens, disconnecting metal atoms, normalizing and ionizing the molecule, and
398 keeping the parent fragments when several fragments of a compound existed. The molecule
399 was then neutralized and, the canonical tautomer was returned. Finally, the cleaned SMILES
400 were used to compute the Morgan fingerprints on 1024 bits, with a radius of three.

401

402 ## Chemical compounds clustering with Butina

403 The Butina clustering algorithm groups molecules based on their structural similarity [18]. The
404 RDKit implementation of the Butina algorithm was used to cluster the chemical compounds [27].
405 The clustering was based on the Tanimoto similarity of the Morgan fingerprints of the
406 molecules, with a cut-off value of 0.7.

407

408 ## Dataset splits

409 To assess the performance of the binary classifiers, the dataset was split several times into
410 training and testing sets. Two types of splits were performed: a random one, which did not
411 consider the chemical similarities of the compounds, and another split that aimed to create sets
412 of structurally different chemicals. This was done to produce cases where the compounds to be
413 classified were novel structures and therefore outside the Applicability Domain (Figure 2).

414 For the random split, called the 'known chemistry case', a stratified 10-fold-cross-validation was
415 performed to split the dataset into 10 different training and testing sets. The scikit-learn Python
416 package [28] was used to perform those splits with the StratifiedKFold function. The dataset was
417 split 10 times with a 10-fold cross-validation with each cross-validation having a different
418 random state, resulting in a total of 100 different splits. Each testing set included 22 or 23
419 compounds.

For the splits based on the chemical structures of the molecules, called the 'novel chemistry case', the compound structures were clustered using the Butina clustering algorithm [18]. A cluster number was then assigned to each compound. The StratifiedGroupK-Fold function of scikit-learn was used to make a 10-fold cross-validation based on the cluster number [28]. Indeed, this function assigned cluster numbers to the testing sets that differed from those in the training set. Additionally, it attempted to maintain a consistent ratio of VAOT and NVAOT classes within each set. The dataset was split in this manner 10 times, with different random states, resulting in a total of 100 unique splits. If any of these splits did not include compounds from both classes in the test sets, they were discarded.

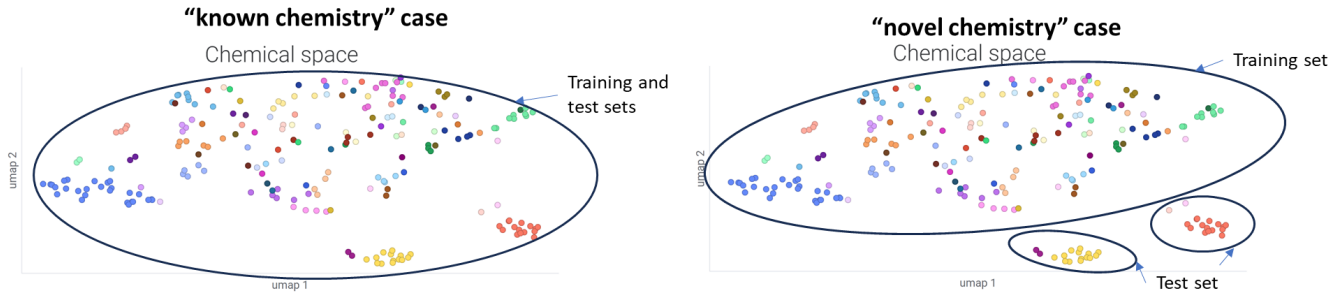


Figure 2

Example on an UMAP of the chemical space of training and test sets from the two holdout strategies. For the 'known chemistry' case, the training and test sets are originated from the same space. For the 'novel chemistry' case, the training and test sets are originated from different Butina clusters. Each color on the UMAP corresponds to a Butina cluster.

Binary classification classifier

To classify compounds as VAOT or NVOAT, several algorithms were tested. For this analysis, we decided to use a K Nearest Neighbors (KNN) algorithm [15], as it showed good performances

440 (supplementary data). The KNN algorithm has also the advantage of being explainable and
441 functions like a read-across, technique commonly used for toxicity prediction [29].

442 We used the scikit-learn [28] implementation of the K Nearest Neighbors (KNN) classification
443 algorithm. Several classifiers were built depending on the data that were used as input. When
444 using the chemical Morgan fingerprints, the Tanimoto (Jaccard) distance was used and when
445 using the morphological profiles, the Pearson correlation-based similarity measure (1 -Pearson
446 correlation) was used. For all classifiers, we set the number of neighbors to one. The choice of
447 the distances and the number of neighbors were the results of benchmarking done on both
448 datasets (supplementary data).

449

450 Decision support model

451 To aid the decision when the two types of classifiers (Morgan fingerprint and Cell Painting
452 morphological profile classifiers) did not predict the same class, a model was built, similar to
453 the Similarity-based merger model [30]. This ensemble model takes as input the predictions of the
454 two KNN classifiers, along with the distances of the nearest neighbors of each prediction (in
455 total four values). A classifier was trained in each training set, for the cases where the two KNN
456 classifiers did not agree on the predicted class. In the test sets, we used this model only when
457 the two KNNs did not predict the same classes, otherwise the consensual predicted classes of
458 the two classifiers were set as the final class (Figure 3).

459 For this decision support model we used a SVM classifier [31] implementation of Scikit learn [28].

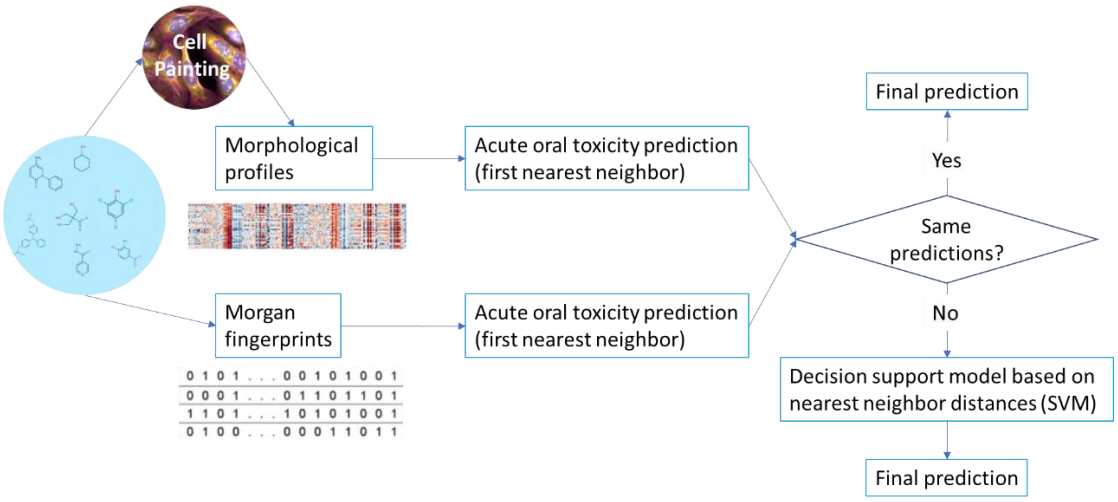


Figure 3

Decision support model

In the 'novel chemistry' case, the training set did not have enough examples with high distances. To remedy this, synthetic examples of distant structures were added. To do this, we subset the cases in each training set where the chemical structure-based predictions did not match the real class, and we updated the nearest neighbor distances with a random number between 0.7 and 0.9 and added these synthetic examples to the dataset used to train the model.

Model performance evaluation

To evaluate the performance of the classifiers, we used different metrics. They were all based on the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN), which were the results of the model classification of a given testing set in a confusion matrix.

Sensitivity: SN = $T\frac{TP}{TP+FN}$

Specificity: SP = $\frac{TN}{TN+FP}$

Balanced accuracy: BA = $\frac{SN+SP}{2}$

Matthews Correlation Coefficient: MCC = $\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

Accuracy: ACC = $\frac{TP+TN}{TP+FN+TN+FP}$

When using cross validation, those metrics were averaged over all testing sets, and their standard deviations calculated.

We performed a corrected t-test [32] to compare the balanced accuracy values over the splits of classification models.

Visualization of the chemical and biological spaces

To visualize on two-dimensional scatter plots the chemical and the biological spaces, UMAP embeddings were generated using the Umap Python package [33]. For the chemical space, the chemical compound structure similarities embeddings were calculated with the Tanimoto distances of the compound Morgan fingerprints. As for the biological space, the morphological profile similarities embeddings were computed using their Pearson correlation-based similarity measures. The plots were visualized in TIBCO Spotfire software.

494 # Results

495 The objective of our analysis was to assess the predictive efficacy of read-across approaches for
496 acute oral toxicity in rats comparing the use of chemical structural information, in vitro
497 biological data derived from the Cell Painting profiling assay on U2OS cells or the combination
498 of both. Two distinct types of inputs were utilized to construct KNN models that classify
499 compounds as Very acutely oral toxic (VAOT) or Non very acutely oral toxic (NVAOT).

500 Initially, we categorized 630 Bayer Crop Science (BCS) agrochemical compounds with known
501 acute rat toxicity results using the public QSAR model CATMoS [5]. Additionally, we used this
502 specific unbalanced set of 630 compounds to build a structure similarity-based classifier.
503 Subsequently, KNN classifiers were used on a reduced but balanced set of 226 compounds
504 using either the chemical structures or their morphological profiles in U2OS cells. A
505 comprehensive analysis of both chemical space (represented by the chemical structures) and
506 biological space (revealed by the U2OS morphological profiles) was conducted to enhance our
507 comprehension of the classifier results. Finally, we investigated whether combining the
508 predictions of the two classifiers could enhance the accuracy of the predictions.

509 The analysis demonstrated that a simple read-across approach, based on chemical structure
510 information and biological data from the Cell Painting profiling assay on U2OS cells, can be used
511 to predict acute oral toxicity, even in the context of new chemical space exploration.

512

513 ## Results of the QSAR classifiers

514 Initially, CATMoS was employed to classify BCS compounds as either VAOT or NVAOT. We used
515 the Opera CATMoS implementation of the model on the 'QSAR only compounds set' (Figure 1
516 1a), with 630 compounds as external test set. The CATMoS predictions were mapped to the two
517 classes, with the majority of compounds being classified as NVAOT. Specifically, 5 of the 109
518 VAOT compounds, and 514 of the 521 NVAOT compounds were correctly predicted, resulting in

519 a low sensitivity of 0.05, a high specificity of 0.99, a balanced accuracy of 0.52 and a MCC of
520 0.09 (Table 2a, supplementary data table S5). This outcome may be attributed to the fact that
521 the CATMoS QSAR model was not trained on Bayer chemistry, but on mostly publicly available
522 industrial chemical compounds. This indicates a possible mismatch in the applicability domain
523 of the model for BCS chemistry.

524 For the predictions of the set of 630 BCS compounds, we could determine that most
525 compounds, 628 (99,6%) were within the CATMoS Applicability Domain (Table 2b). Most of
526 them, 448 (71%), had an Applicability Domain index below 0.6, suggesting that the predictions
527 should be considered with caution (Table 2c). The remaining 180 (29%) predictions having an
528 Applicability Domain index above 0.6, displayed an average confidence level of 0.57, indicating
529 a relatively low level of confidence in the predictions (Table 2b).

530 Subsequently, we developed a classifier, based on our 630 compound set, using a KNN classifier
531 on Morgan fingerprints of chemical compounds. The classifier's performance was evaluated
532 through cross-validation with two data holdout strategies: the 'known chemistry' case, where
533 compounds from the test sets resemble those from the training sets, and the 'novel chemistry'
534 case, where compounds from the test sets differ structurally from those in the training sets.

535 In the 'known chemistry' case, the classifiers achieved an average balanced accuracy of 0.81, a
536 sensitivity of 0.70, a specificity of 0.92 and a MCC of 0.61 (Table 2c).

537 In the 'novel chemistry' case, out of 100 theoretical cross-validation splits, only 44 of the 100
538 theoretical cross-validation splits included the two classes (VAOT and NVAOT) in both the
539 training and testing sets, making them valid. The classifier demonstrated an average balanced
540 accuracy of 0.60, a sensitivity of 0.33, a specificity of 0.87 and a MCC of 0.19 (Table 2c).

541 The 'novel chemistry' case demonstrated that chemical structure-based classifiers perform less
542 well when classifying compounds that are structurally distinct from those in the training set. In
543 summary, the chemical structure similarity-based models demonstrated good performance in
544 handling known chemistry. However, as expected by the design of the 'novel chemistry' case,

545 they exhibited a decrease in performance when confronted with unfamiliar chemical
546 structures. This limitation becomes apparent when exploring new areas of chemical space. To
547 address this limitation, we leveraged the biological effects of chemical compounds for
548 predictions. The subsequent section outlines our approach, employing Cell Painting assay on
549 U2OS cells to capture the biological effects of the chemical compounds.

550 **QSAR only compounds set (set of 630 compounds)**

a. Classification using CATMoS

		Predicted class	
		VAOT	**NVAOT**
True class	**VAOT**	5	104
	NVAOT	7	514

551

Balanced accuracy	**MCC**	**Sensitivity**	**Specificity**
0.52	0.09	0.05	0.99

552

b. Prediction reliabilities

Within Applicability Domain	**Applicability Domain Index**	**Number of compounds (percentage)**	**Average confidence index**
No	all	2 (0.3%)	NA
Yes	< 0.6	448 (71.1%)	0.5
Yes	≥ 0.6	180 (28.6%)	0.57

553

554

c. Performance of the KNN classifiers trained on 630

Holdout strategy	**Balanced accuracy**	**MCC**	**Sensitivity**	**Specificity**
'known chemistry' **case**	0.81 ±0.07	0.60 ±0.12	0.69 ±0.13	0.92 ±0.04
'novel chemistry' **case**	0.61 ±0.16	0.17 ±0.24	0.31 ±0.29	0.92 ±0.21

555 *Table 2*

556 *a. Confusion matrix and metrics for the classification of CATMoS for 630 compounds from Bayer Crop Science.*

557 *b. Reliability of the predictions: Number of compounds outside the CATMoS applicability domain, number of compounds*
558 *and average confidence index for compounds within the CATMoS applicability domain and having an Applicability Domain index*
559 *below or above 0.6.*

560 *c. Mean of 4 metrics assessing the performance of the KNN binary classifiers built out 630 Bayer CropScience*
561 *agrochemical candidates, over the 100 splits of the 'known chemistry' case where training and testing sets are split randomly,*
562 *not taking into account chemical structure similarities, and over the 44 valid splits of the 'novel chemistry' case where training*
563 *and testing sets are split in order to have structurally different compounds over the two sets.*

564

565 ## Comparison of chemical structure and Cell Painting morphological based classifiers

566 The objective of our study was to compare two inputs for predicting acute oral toxicity, utilizing
567 a dataset called the 'Cell Painting set', a subset of the 630 'QSAR only compound set',
568 augmented with additional public chemical compounds. The Cell Painting set included a total of
569 226 compounds (Figure 1b). KNN classifiers were trained using both types of input and
570 employing two data holdout strategies: 'known chemistry' and 'novel chemistry' cases.

571 Similarly to the previous chemical structure similarity-based classifier on the QSAR only
572 compound set (630 compounds), KNN classifiers were trained on the Morgan fingerprint of the
573 molecules.

574 For classifiers based on the morphological profiles obtained from Cell Painting, consensus
575 profiles were utilized after normalization, unsupervised features selection, and replicate profile
576 aggregation at the treatment level. Regarding the chemical structure similarity-based classifiers,
577 we used KNN algorithm. Classifiers were built for each tested concentration (10 µM, 31.6 µM,
578 100 µM).

579

580 ### Results in the 'known chemistry' Case

581 In the 'known chemistry' case, the chemical structure similarity-based classifier demonstrated
582 superior performance compared to other classifiers, achieving a mean balanced accuracy of
583 0.82. This was followed by the 31.6µM morphological profile classifier, with a mean balanced

584 accuracy of 0.74 (Table 3a). The two other morphological profile classifiers at 10 and 100 µM
585 demonstrated lower performance. (Table 3a).

586 The distribution of the balance accuracy values over the 100 splits for each input type of input
587 showed a narrow range (Figure 4a), which was confirmed by low standard deviations ranging
588 from 0.08 to 0.09 (Table 3a).

589 A statistical analysis using the Nadeau and Bengio's corrected t-test to compare the balance
590 accuracy values over the 100 splits of the two top classifiers, indicated that the chemical
591 structure similarity-based classifier significantly outperformed the 31.6 µM morphological
592 profile classifier (p-value= 0.05).

593 Results in the 'novel chemistry' Case

594 In the 'novel chemistry' case, the 31.6 µM morphological profile classifier demonstrated
595 superior performance achieving a mean balanced accuracy of 0.72. This was followed by the
596 chemical structure similarity-based classifier, with a mean balanced accuracy of 0.60 (Table 3b).
597 The remaining two classifiers demonstrated lower performance (Table 3b).

598 The distributions of the balanced accuracies for each classifier showed that in certain splits, the
599 chemical structure similarity-based classifier encountered challenges in making accurate
600 predictions (Figure 4b). This was also the case, to a lesser extent for the 31.6 µM morphological
601 profile-based classifiers (Figure 4b).

602 The Nadeau and Bengio's corrected t-test indicated that the 31.6 µM morphological profile
603 classifier significantly outperformed the chemical structure similarity-based classifier (p-
604 value=0.045).

605 In summary, for the chemical structure similarity-based classifiers, we reproduced the results of
606 the previous chemical structure similarity based classifier, which was trained on approximately
607 three times more compounds (630 compounds) with good performances in the 'known

608 chemistry' case, and a decrease in performance in the 'novel chemistry' case (the balanced
609 accuracy dropped from 0.82 to 0.60).

610 Overall, our findings emphasize the superior performance of chemical structure similarity-based
611 classifier in the 'known chemistry' case. However, the morphological profile-based classifier
612 remains a valuable tool, particularly in the 'novel chemistry' case, where the classifiers based
613 on the 31.6 µM morphological profile demonstrated the highest performance (balanced
614 accuracy of 0.72).

615

616

617

618

619

620

621

622

623

624

625

Performances of the classifiers

	a: known chemistry case				b: novel chemistry case			
	BA	**MCC**	**SN**	**SP**	**BA**	**MCC**	**SN**	**SP**
Morgan FP	0.82 ± 0.08	0.65 ± 0.16	0.82 ± 0.11	0.82 ± 0.12	0.60 ± 0.16	0.20 ± 0.29	0.49 ± 0.28	0.71 ± 0.18
CP 10 µM	0.57 ± 0.08	0.14 ± 0.16	0.55 ± 0.14	0.58 ± 0.14	0.50 ± 0.14	0.04 ± 0.26	0.46 ± 0.24	0.54 ± 0.15
CP 31.6 µM	0.74 ± 0.08	0.49 ± 0.17	0.70 ± 0.13	0.78 ± 0.12	0.72 ± 0.12	0.42 ± 0.23	0.66 ± 0.22	0.78 ± 0.15
CP 100 µM	0.63 ± 0.09	0.27 ± 0.19	0.61 ± 0.13	0.65 ± 0.14	0.53 ± 0.13	0.07 ± 0.25	0.47 ± 0.23	0.60 ± 0.18
DS	0.85 ± 0.07	0.71 ± 0.14	0.86 ± 0.11	0.84 ± 0.11	0.72 ± 0.13	0.42 ± 0.26	0.67 ± 0.23	0.77 ± 0.16

Table 3

Mean and standard deviations of 4 metrics: Balanced Accuracy (BA), Matthew's correlation coefficient (MCC), Sensitivity (SN), Specificity 'SP). Four different input data were used to classify compounds as VAOT and NVAOT: chemical structural data (Morgan FP) and Cell Painting (CP) morphological profiles of U2OS cells exposed to chemicals at 3 concentrations (MP 10 µM, MP 31.6 µM and MP 100 µM). Orange highlight = the best average metric. The decision support (DS) model performance (combining the Morgan Fingerprint and the morphological profile 31.6 µM classifier predictions, supplemented with synthetic examples), are shown in the last row.

Section a. Reports the performance of the binary classifiers, over the 100 splits of the 'known chemistry' case where training and testing sets are split randomly, not considering chemical structure similarities.

Section b. Reports the performance of the KNN classifiers, over the 99 valid splits of the 'novel chemistry' case where training and testing sets are split to have structurally different compounds over the two sets.

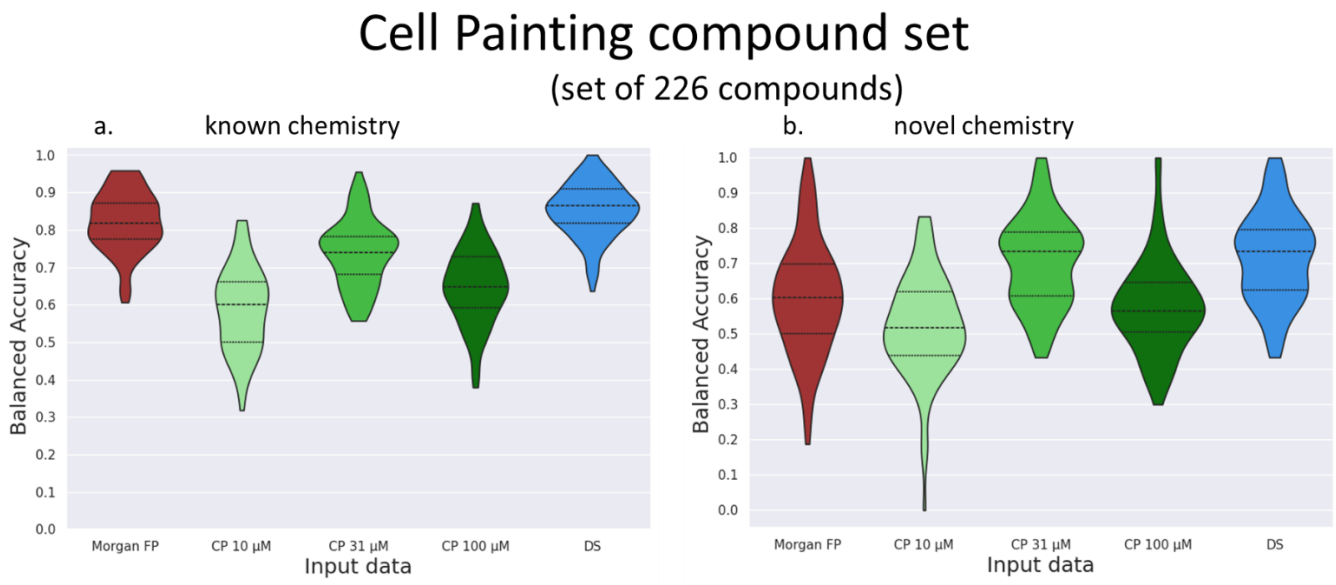


Figure 4

Section a: violin plots represent the balanced accuracies of the binary classifier for the 10 x 10-fold cross validation splits not considering the structure similarities (known chemical case).

Section b: violin plots represent the balanced accuracies of the KNN binary classifier for 99 valid splits of the 10 x 10-fold cross validation that put in the testing set chemical structurally different from the training set (novel chemical case).

Legend: In red, the classifier using the Morgan Fingerprint, in light green the classifier using the morphological profiles at 10 µM, in mid- green the classifier using the morphological profiles at 31.6 µM and in dark green, the classifier using the morphological profiles at 100 µM, in blue, the decision support (DS) model. Inside each violin plot the quartiles are indicated as dash lines.

652

653

654

655

656 ## Comparison of the chemical and biological spaces

657 To gain insight into the performance of the classifiers based on the chemical structures and the
658 morphological profiles, we investigated into the chemical and biological spaces, for the Cell
659 Painting set of compounds.

660 ### Chemical space

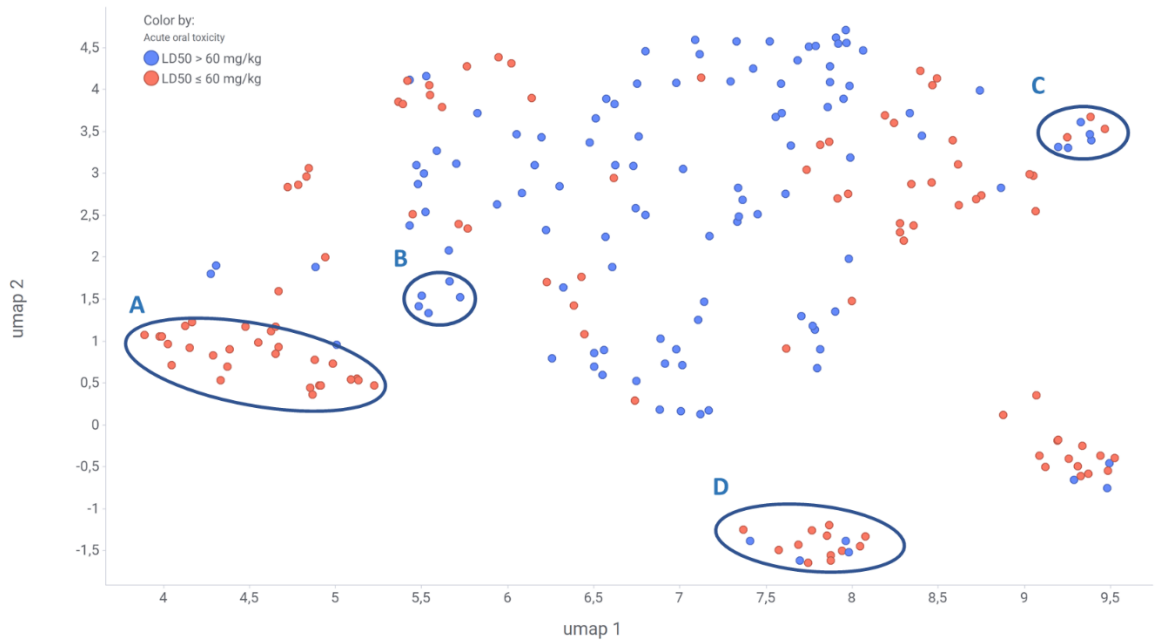
661 Our dataset comprises 226 compounds primarily originating from Bayer Crop Science chemistry
662 and supplemented with 29 public compounds. The Butina algorithm, using the Tanimoto
663 distance and a threshold of 0.7, identified 91 clusters. Of these, 61 were represented by a
664 unique compound indicating structural diversity.

665 By plotting the similarity of the structures on a UMAP (Figure 5), using Morgan fingerprints and
666 Tanimoto distance, it is possible to visually identify distinct clusters. Notably, certain clusters
667 exclusively comprised of VAOT compounds (e.g., cluster A), while others were exclusively
668 comprised of NVAOT compounds (e.g., cluster B). Additionally, several clusters contained a
669 mixture of both (e.g., clusters C and D).

670 In summary, the chemical space exhibited diversity in the form of different clusters. Specific
671 areas demonstrated a prevalence of either VAOT or NVAOT compounds, while others
672 presented a combination of both classes.

673

674 Chemical space



675

676 *Figure 5*

677 *Scatter plot of the 2-dimensional UMAP embedding of the chemical compound morgan fingerprints. In blue, the chemical*
678 *compounds that are NVAOT. In red, chemical compounds that are VAOT. Four clusters of compounds are designated by the*
679 *letter A, B, C and D. Cluster A is an example of a cluster with only VAOT compound. The cluster B is an example of a cluster with*
680 *only NVAOT compounds. B and C are two examples of clusters with a mix of VAOT and NVAOT compounds.*

681

682 Biological space

683 Figure 6 illustrates the degree of similarity in the biological response of compounds, as
684 measured by Pearson correlation-based similarities, on a UMAP plot. In contrast to the
685 chemical space, a limited number of clusters visually emerged with only two notable clusters

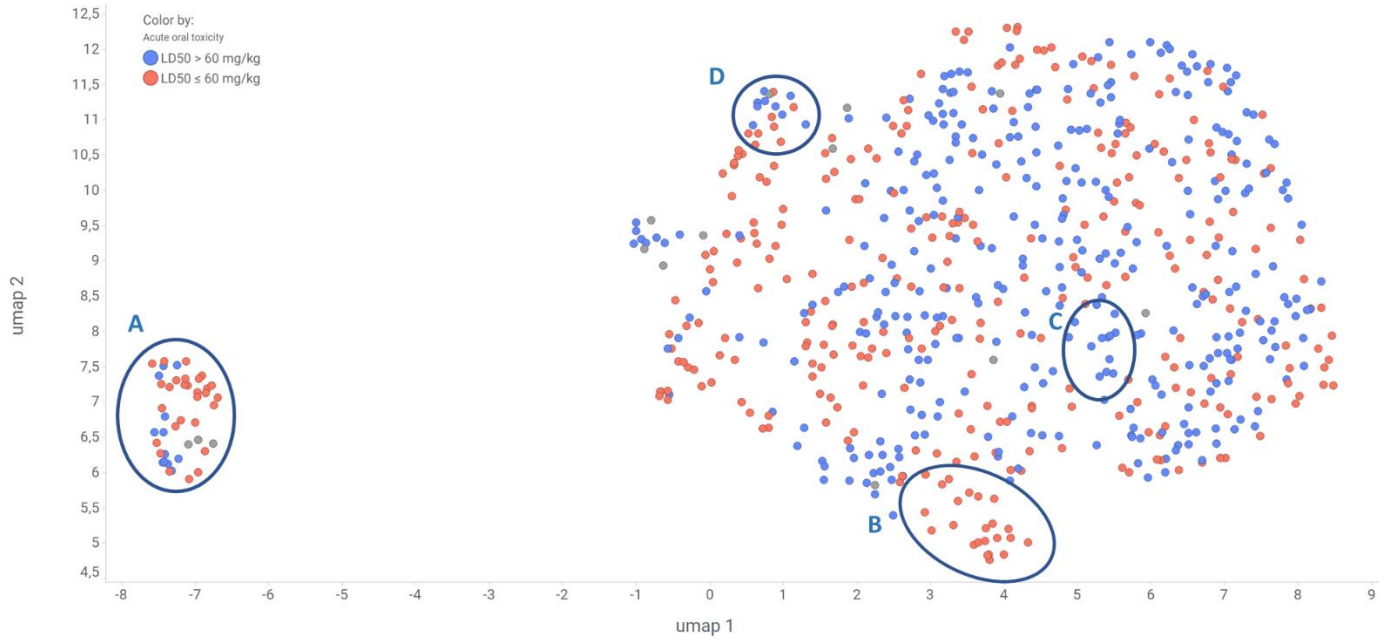
686 observed. An isolated small cluster (cluster A) was clearly separated from the other profiles,
687 and upon inspection these profiles corresponded to instances with notably low cell count.

688 The second cluster exhibited diverse areas: including regions with a high number of VAOT (e.g.,
689 grouping B), NVAOT compounds (e.g., grouping C), and areas with a mix of both classes (e.g.,
690 grouping D).

691 In Figure 7, we focused on the 31.6µM concentration, the concentration, which yielded optimal
692 performance for the classifier using morphological profiles. Similar observations were made,
693 with an isolated cluster corresponding to profiles with a very low number of cells. Additional
694 distinct areas emerged, showcasing regions with high number of VAOT or NVAOT compounds,
695 as well as areas with a mixed representation of both classes.

696

697 Biological space (all concentrations)



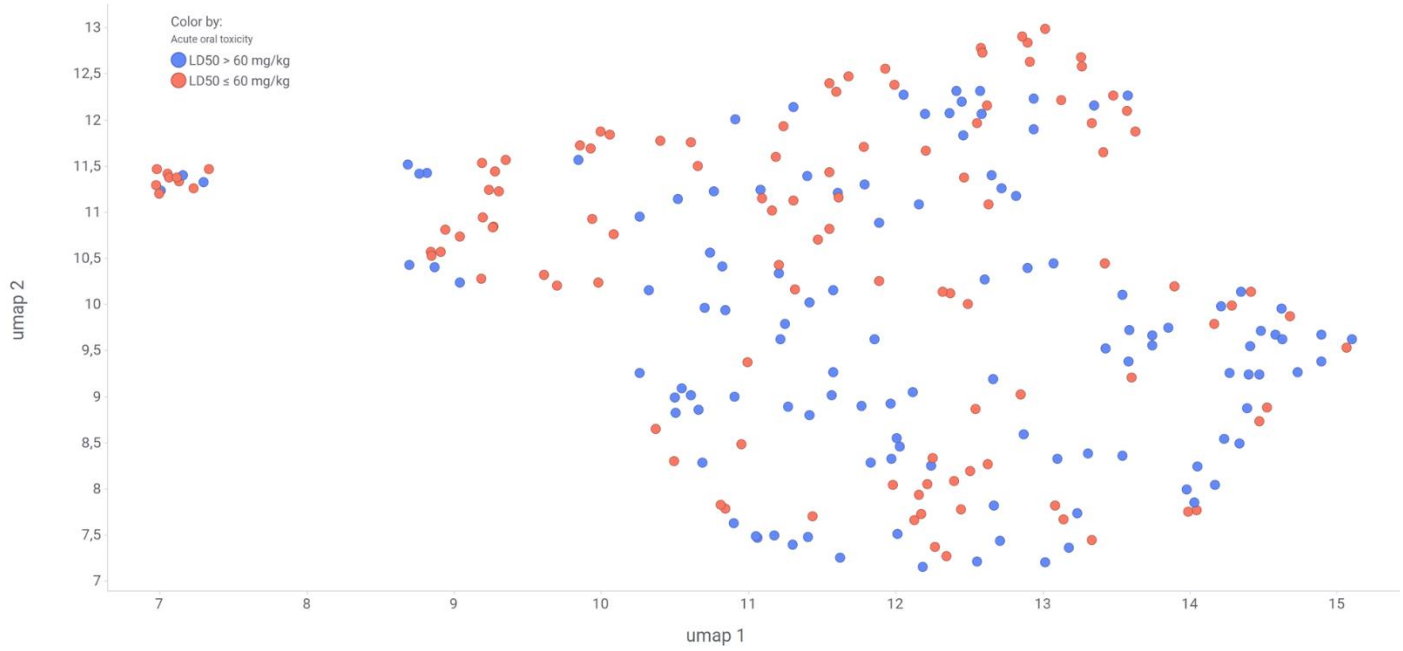
698

699 *Figure 6*

700 *Two-dimensional representation of the consensus morphological profile similarities for all treatments and 3 concentrations,*
701 *using Uniform Manifold Approximation (UMAP) embedding on 2 components with the Pearson correlation-based similarity*
702 *measure. In red, the morphological profiles of U2OS cells perturbated by VAOT compounds. In blue, the morphological profiles of*
703 *U2OS cells perturbated by NVAOT compounds. Four groups of compounds are designated by the letter A, B, C and D. The group*
704 *A of compounds corresponds to treatment with very low cell counts. The group B of compounds is an example of grouping with a*
705 *high number of VAOT compounds. The group C of compounds is an example of grouping with a high number of NVAOT*
706 *compounds. The group D of compounds is an example of grouping with a mix of VAOT and NVAOT compounds.*

707

708 Biological space (31.6 µM)



709

710 *Figure 7*

711 *Two-dimensional representation of the consensus morphological profile similarities for all treatments at 31.6 µM, using Uniform*
712 *Manifold Approximation (UMAP) embedding on 2 components with the Pearson correlation-based similarity measure. In red,*
713 *the morphological profiles of U2OS cells perturbated by VAOT compounds. In blue, the morphological profiles of U2OS cells*
714 *perturbated by NVAOT compounds.*

715

716 Comparison of the chemical and biological spaces

717 We compared different groups of compound structures and groups of morphological profiles to
718 better understand the chemical and biological space interrelationship. Notably, we observed
719 that chemicals clustering together in the chemical space could elicit a diversity of biological
720 responses in the biological space, emphasizing that structurally similar compounds may
721 manifest distinct biological responses (Figure 8, left column). We illustrate a specific case
722 involving a group of chemicals, the carbamates, inducing similar morphologies in U2OS cells
723 (Figure 8, center column). Interestingly, similar morphological profiles induced by structurally
724 different compounds could also be observed (Figure 8, right column). This comparison
725 illustrates that: 1) the biological effects of structurally similar molecules may not necessarily be
726 identical, 2) Structurally similar compounds could trigger different biological effects and 3)
727 conversely, compounds with different structures could result in comparable biological
728 responses.

729

730

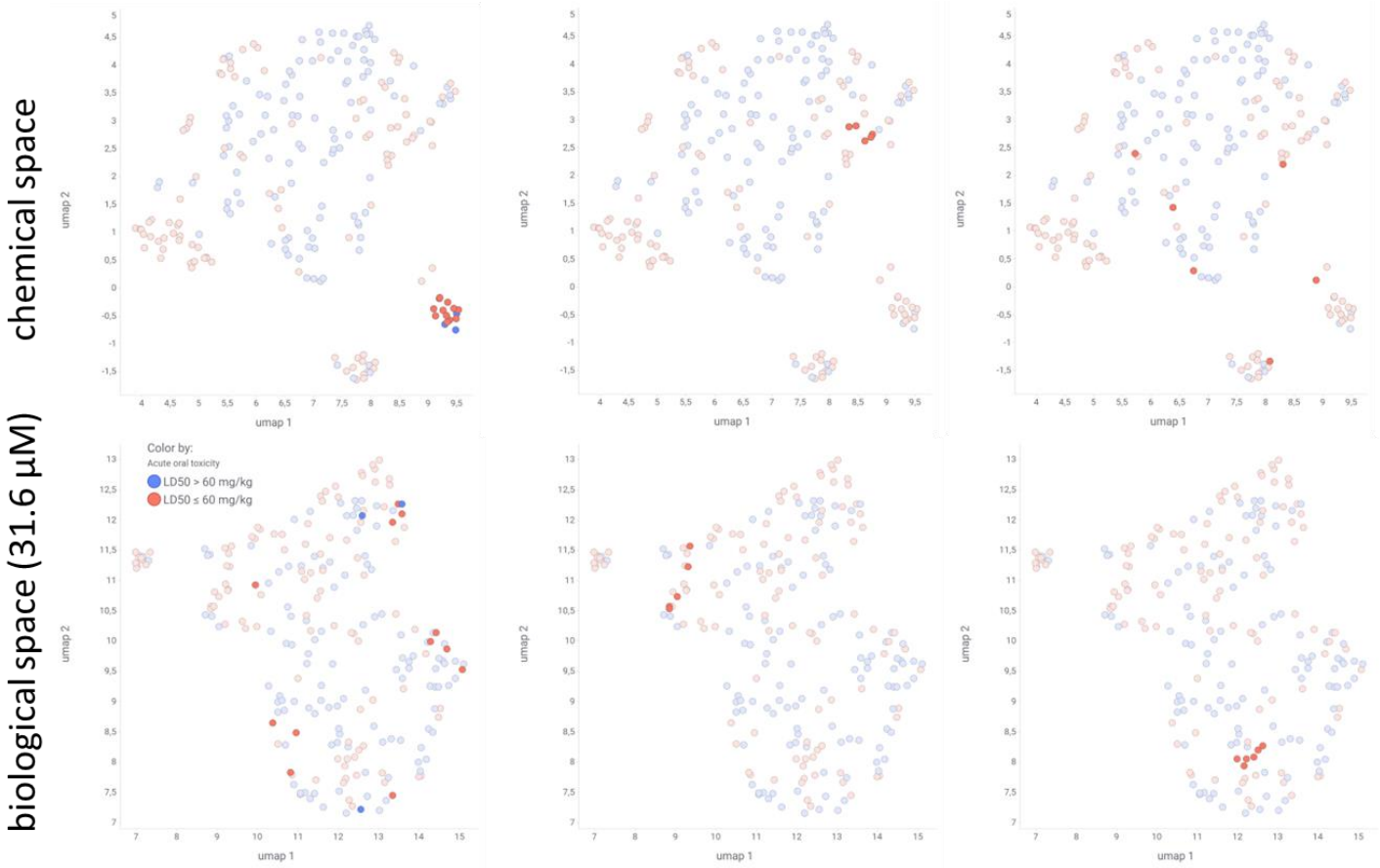


Figure 8

(top) Chemical space. Scatter plot of the 2-dimensional UMAP embedding of the chemical compound morgan fingerprints.

(bottom) Biological space. Scatter plot of the 2-dimensional UMAP embedding of the Cell Painting morphological profiles of the chemical compounds at 31.6 µM.

In red, the morphological profiles of U2OS cells perturbated by VAOT compounds.

In blue, the morphological profiles of U2OS cells perturbated by NVAOT compounds.

On each column, the chemical profiles and morphological profiles of the same compounds are selected.

(left) Example of structurally similar compounds, inducing different morphological profiles.

(center) Example of structurally similar compounds (carbamates), inducing similar morphological profiles.

742 *(right) Example of structurally different compounds, inducing similar morphological profiles.*

743

744

745 Biological response

746 To analyze the biological response of U2OS cells to chemical compound perturbations, we
747 employed morphological profiles, using two metrics: the Grit score [22], and the number of cells.
748 The analysis was conducted to better understand how U2OS cells reacted to our set of
749 compounds, which in turn helped us to understand the results of the classifiers.

750 Grit score

751 The Grit score indicated the extent to which the average morphology of U2OS cells perturbed
752 by a compound deviated from the average negative control morphology of non-perturbed
753 U2OS cells. A high Grit score indicated a more distinct cell morphology from the negative
754 controls. For example, the average Grit score of the positive controls was 4.8.

755 A Mann-Whitney U rank test on the grit scores, for the two compound groups, VAOT and
756 NVAOT, demonstrated that VAOT compounds elicited a marginal though significant stronger
757 biological response compared to NVAOT (Grit score respectively 3 and 2.5, p-value of 0.004).
758 (Table 4).

759 Regarding concentrations, on average, the 10 µM treatments had a grit score of 1.9, the 31.6
760 µM had a grit score of 2.6 and the 100 µM treatment had a grit score of 3.7. This aligns with our
761 assumption that higher concentrations lead to increased biological responses, a consideration
762 made when designing the Cell Painting campaign with 3 concentrations (Table 4).

763 Identifying compounds with no induced morphological changes, we set a Grit score threshold of
764 1. Below this threshold, we considered a treatment not inducing any morphological change. Of
765 the 23 compounds falling below this threshold 6 were VAOT.

766 Number of cells

767 An additional output of the image analysis was the number of cells per well. For this analysis
768 the number of cells was not normalized, and the median number of cells per well for a given
769 treatment were computed. The average number of cells for the negative controls was 2231. We
770 arbitrarily set the number of cells that defines cytotoxicity as a cell count below 50% of the
771 average negative control cell count, meaning a cell count per well of below 1115 defined a
772 cytotoxic treatment.

773 In total 44 compounds exhibited cytotoxicity: 11 compounds, at 10 µM, 22 compounds, at 31.6
774 µM and 44 compounds, at 100 µM (Table 4).

775 Categorizing by class, 28 VAOT compounds (25%) and 16 NVAOT compounds (14%) displayed
776 cytotoxicity for at least one concentration. A chi-square test of independence of variables, with
777 the null hypothesis that the number of cytotoxic compounds is independent of the class (VAOT
778 and NVAOT) gave a p-value of 0.1. We could not conclude that there was a higher percentage
779 of cytotoxicity for VAOT compounds (Table 4).

780

781

782

783

784

785

786

787

788

profiles	Average Grit score	Number of cytotoxic treatments
Negative control	NA	0
VAOT	3	28
NVAOT	2.5	16
10 µM all	1.9	11
10µM - VAOT	2.1	8
10µM - NVAOT	1.64	3
31.6 µM all	2.6	22
31.6µM - VAOT	2.98	14
31.6µM - NVAOT	2.17	8
100 µM all	3.7	44
100µM-VAOT	4.01	27
100µM-NVAOT	3.20	17

789 *Table 4*

790 *Average grit score and number of cytotoxic treatments for different groups of profiles: VAOT (very acutely oral toxic*
791 *compounds), NVAOT (non-very acutely oral toxic compounds), 10 µM treatment profiles, 31.6 µM treatment profiles and 100*
792 *µM treatment profiles.*

793

794 Results of the decision-support model

795 The decision-support model aided the decision when the two KNN classifiers did not predict the
796 same class. The model combined four pieces of information: predictions from the KNN classifier
797 based on the chemical structure information, predictions from the KNN classifier based on the
798 morphological profiles and distances to the nearest neighbor in each classifier.

799 In the 'known chemistry' case, the model demonstrated an average balanced accuracy of 0.84,
800 slightly above the chemical structure similarity-based classifier's average balanced accuracy of
801 0.82 but was not significantly different (Nadeau and Bengio's corrected t-test p-value of 0.47).
802 In the 'novel chemistry' case, the model had on average a balanced accuracy of 0.65, below the
803 31.6µM morphological profile classifier's average balanced accuracy of 0.72.

804 To understand why in the 'novel chemistry' case, this model did not yield better performances,
805 we computed the mean Morgan fingerprint Tanimoto distances between each chemical
806 compound of the training set and its nearest neighbor in the training set, and the mean
807 distances between each chemical compound of the testing set and its nearest neighbor in the
808 training set.

809 In the 'known chemistry' case, on average, in the training set, each compound has a distance to
810 its nearest neighbor of 0.50 and in the testing set 0.48. For the 'novel chemistry' case, on
811 average, in the training set, each compound had a distance to its nearest neighbor of 0.49, and
812 in the testing set 0.73.

In the ‘novel chemistry’ case, the training set did not have enough examples of distant chemical structures. To help the model, we added synthetic examples of distant chemical structures in the training set. To do so, we subset in each training set the cases where the predictions of the chemical structure similarity-based did not match the real class, and we updated the distances of the nearest neighbors with a random number between 0.7 and 0.9 and added those as synthetic examples in the dataset used to train the model.

By applying this approach, in the ‘known chemistry’ case the model achieved an average balanced accuracy of 0.85, slightly above, the chemical structure similarity-based classifier’s average balanced accuracy of 0.82 but was not significantly different (Nadeau and Bengio’s corrected t-test p-value of 0.12). In the ‘novel chemistry’ case, the model achieved an average balanced accuracy of 0.72 (Table 3).

826 ## Discussion

827 Our results showed that the classification of compounds as very acute oral toxic or not, using a
828 similarity-based approach, was possible using chemical structure information, morphological
829 profiles of U2OS cells or the combination of both. When classifying compounds coming from
830 the same chemical space as those in the classifier's training set, the chemical structure
831 information was more predictive. Conversely, when the compounds to be classified came from
832 a different chemical space than those in the classifier's training set, the morphological profiles
833 of the U2OS cells were more predictive.

834 Initial attempts to use the publicly available QSAR model, CATMoS , for the prediction of acute
835 oral toxicity on a set of 630 Bayer compounds did not yield good predictions [5]. The CATMoS
836 performance is hindered as Bayer Crop Science chemistry could be considered to be locally
837 outside its training set (Table 2b). Although almost all compounds were globally within the
838 CATMoS applicability domain, most resided in gaps of the training chemical space. It is well
839 known that QSAR models excel when the compounds to be classified fall within their
840 applicability domain of the models, and can perform poorly when they do not [34]. In summary,
841 CATMoS, which is a QSAR model trained on more than 10,000 compounds, has very good
842 performance for the prediction of acute oral toxicity of those chemicals, but it does not work as
843 effectively with the structurally diverse BCS chemistry.

844 To test our hypothesis, we trained a simple KNN classifier, resembling a similarity-based (or
845 read-across) approach, on this set of 630 BCS compounds, using their chemical structure
846 information. Working with two data-holdout strategies to simulate scenarios within and outside
847 the applicability domain of a QSAR model, we evaluated our classifiers under two conditions:
848 the 'known chemistry' case, simulating scenarios within applicability domain case, and the
849 'novel chemistry' case, attempting to simulate outside applicability domain case. In the 'known
850 chemistry' case, our classifier exhibited strong performance, comparable to CATMoS :CATMoS
851 achieved a balanced accuracy of 0. 84, in classifying compounds as very toxic (VT) (LD50 < 50

852 mg/kg) whereas our classifier had a balanced accuracy of 0.81 (Table 2c) for the classification of
853 compounds as VAOT (LD50 < 60 mg/kg) [5].

854 However, as designed, the performance of the classifier dropped in the 'novel chemistry' case
855 due to the data-holdout strategy, which placed Butina compound clusters, which are not
856 present in the training sets, into the test sets. This effectively simulated scenarios outside of the
857 models applicability domain, although the decrease in balanced accuracy (from 0.82 to 0.60)
858 (Table 3) was not as drastic as that observed with the Bayer CropScience chemistry using
859 CATMoS (from 0.84 to 0.52) (Table 2a).

860 To overcome this chemical applicability domain limitation, we explored whether using the
861 compound biological effects could mitigate this problem. Compound-induced biological effects
862 characterized by transcriptomics have previously been used to predict target activities, in
863 association and comparison with QSAR models[35,36]. Here we utilized Cell Painting to generate
864 morphological profiles at a more reasonable cost compared to transcriptomics.

865 Using a smaller set of 226 compounds, (balanced with 49 % of compounds where LD50<60
866 mg/kg; 51% where LD50>60 mg/kg), we trained KNN classifiers based on either chemical
867 structure information or U2OS morphological profiles at 3 concentrations (10, 31.6 or 100 µM).

868 Morphological profiles at 31.6 µM concentration demonstrated better performance, compared
869 to the other concentrations, in both the 'known chemistry' and 'novel chemistry' cases. With a
870 balanced accuracy of 0.74 (Table 3) in the 'known chemistry' case and 0.72 (Table 3) in the
871 'novel chemistry' case, Cell Painting U2OS profiles demonstrated the ability to predict acute
872 oral toxicity classes, interestingly, independent of the structural similarity of the tested
873 compounds.

874 Cell Painting can indeed identify morphological patterns associated with specific mode of action
875 (MoA) and molecular initiation event (MIE) of compounds [37,38]. Typically, acute toxicity involves
876 a limited number of MIEs [39]. such as narcosis (activity at the lipid bilayer of the membrane),
877 acetylcholinesterase inhibition, ion channel modulators and inhibitors of cellular respiration [40].

878 The Cell Painting experiment revealed morphological profiles (initiated by MIE) associated with
879 acute oral toxicity, as evidenced by the grouping of morphological profiles associated with
880 VAOT compounds. For example, four carbamates (Promecarb, Methiocarb, Propoxur, m-
881 Cumenyl methylcarbamate) known as acetylcholinesterase inhibitors, produced similar
882 morphological profiles in U2OS cells (Figure 8, middle row). This partially explains why the
883 morphological profile-based classifiers were able to correctly classify the compounds. In the
884 'known chemistry' case, the performance of the classifiers based on the 31.6 µM morphological
885 profile did not surpass the classifiers based on the chemical structure information but
886 outperformed them in the 'novel chemistry' case.

887 The comparison of the two chemical structure based KNNs for the two datasets showed similar
888 performances, but this comparison is limited because the two datasets are different. The QSAR
889 dataset consisting of 630 compounds is unbalanced (109 VAOT, 521 NVAOT), while the Cell
890 Painting dataset, made of 226 compounds is balanced, but of smaller size.

891 Capturing the biological effects of compounds had limitations: the limitation of the cell system
892 to reveal the effects causally related to acute toxicity, together with technical limitations of the
893 laboratory experiment itself.

894 For the limitation of the cell system, we observed, by Grit score analysis, that not all the
895 compounds induced a biological response in U2OS cells (10% of the tested compounds),
896 regardless of the concentration used. Six known VAOT compounds did not elicit any
897 morphological changes compared to the negative controls. Among these compounds, five were
898 public compounds and some of them with information on their possible mechanism of action.
899 The warfarin, a vitamin K antagonist, and the methamidophos, a potent acetylcholinesterase
900 inhibitor, did not induce any biological response in U2OS cells. This suggests that U2OS cells
901 have their own biological applicability domain and may not capture all the bioactivities
902 associated with acute oral toxicity observed in a whole organism such as a rat in our case study.

903 Nevertheless, for our set of compounds, Cell Painting on U2OS managed to capture bioactivities
904 for most of the VAOT compounds.

905 On the contrary, when analyzing the number of cells, we could also identify a limitation due to
906 the cytotoxicity of the compounds: 44 compounds showed cytotoxicity at least at one
907 concentration, and 12 exhibited cytotoxicity even at the lowest concentration. The
908 morphological profiles of the cytotoxic treatments were not informative as they consisted
909 mainly of debris and dying cells. It appears that the 31.6 µM concentration represented a good
910 compromise between inducing bioactivity and avoiding cytotoxicity. However, building a model
911 using only this concentration is a limitation, as morphological profiles coming from other
912 concentrations could also have been associated with acute oral toxicity. Different rules for
913 selecting the best concentration per compound, using the Grit score to ensure that the
914 compound was active, or the number of cells to ensure that the compound was not cytotoxic
915 did not produce better models.

916 For the limitation of the experiments, several quality issues can arise when conducting an
917 experiment in a laboratory. Experiments are technically demanding and prone to variability and
918 error. Seeding variability can affect the cell morphologies and thus morphological profiles.
919 There are other common problems that can occur in laboratory experiments, such as treatment
920 errors, compounds with low purity or precipitation at high concentrations. These problems can
921 affect the quality of the morphological profiles and thus the performance of a classifier based
922 on morphological profile similarities.

923 The chemical structural information did not suffer from these limitations, because this
924 information was not subject to quality issues, was not cell system dependent, and was not
925 assay design dependent. This information was intrinsic to the description of a given compound.
926 This may partly explain why chemical structure similarity-based classifiers, in the 'known
927 chemistry' case, performed better than biological based classifiers: the full structural

928 information is available, whereas the biological information is partially available and subject to
929 quality issues, in particular reproducibility.

930 For both types of input data, the optimal number of neighbors for the KNN algorithm
931 (supplementary information, S4) was 1, indicating that few examples of identified profiles
932 leading to high acute oral toxicity or not were present in the dataset. A larger set of
933 compounds, such as the CATMoS training set, would help to identify more examples of cell
934 painting profiles associated with acute oral toxicity.

935 In the 'novel chemistry' case, the morphological profile-based classifiers did not experience as
936 large a performance drop as the chemical structure similarity-based classifiers, suggesting that
937 the biological space did not cluster in the same way as the chemical space. This indicates that
938 similar compounds did not consistently induce the same response in U2OS cells (for example
939 due to activity cliffs), and vice versa. The presence of different Butina clusters in the training
940 and test sets did not necessarily result in different morphological profiles, explaining why the
941 morphological profile-based classifier performance did not drop drastically in the 'novel
942 chemistry' case.

943 The use of biological responses of compounds could be also an advantage with respect to
944 enantiomers. The Morgan fingerprint used in this analysis does not take chirality into account.
945 Enantiomers may have different acute oral toxicity, and the classifier based on chemical
946 structure will not distinguish between them, where morphological profiles may be different.

947 The decision support model combined both predictions along with the nearest neighbor
948 distances to make the final predictions, slightly improving the classification performance in the
949 'known chemistry' case but decreasing in the 'novel chemistry' case. By adding a few synthetic
950 examples in each training set with higher distances in the chemical spaces, it was possible to
951 increase the classification accuracy in the 'known chemistry' case, but not in the 'novel
952 chemistry' case, where the model performed like the 31.6µM morphological-based classifier.
953 Notably, in the 'novel chemistry' case the classifier preferred the predictions of the 31.6µM

954 morphological-based classifier predictions over the predictions of the chemical structure
955 similarity-based classifier.

956 Further results could extend and refine these findings by employing a broader set of
957 compounds covering additional molecular initiating events (MIE) associated with acute oral
958 toxicity. Additionally, a larger set of compounds could facilitate the identification of additional
959 morphological profiles associated with acute oral toxicity. A larger dataset would also allow
960 isolation of a set of compounds as an external dataset to further evaluate the performance of
961 the classifier.

962 The choice of the KNN algorithm in this analysis was deliberate due to its simplicity and
963 similarity to read-across approach commonly used in toxicology. Since the amount of in vivo
964 data is often limited, the read-across approach is often the only analysis that can be performed.
965 For chemical structure-based classifiers, other algorithms yielded similar performances
966 (Supplementary Information).

967 To support the creation of public QSAR models with a wider applicability domain,
968 representations of compound structures and results of acute toxicity studies for early
969 candidates which failed to be placed on the markets could be shared by companies and
970 organizations to expand the chemical space coverage.

971 In addition, in this analysis, the Morgan fingerprint was the only computed chemical fingerprint.
972 The use of additional fingerprint or descriptors could help to achieve better QSAR and chemical
973 structure similarity-based classifier performance.

974 Similarly, hand-crafted morphological features were used in this analysis. To capture a broader
975 representation of morphological profiles, other deep learning based representations could also
976 be tested[41].

977 The decision support model uses the nearest neighbor distance to decide on which prediction
978 to select. Other metrics, such as the 'distance to model', which is used to estimate a prediction
979 uncertainty could also be used to decide on which prediction to use[42].

980 We have also seen the limitation of the U2OS cell line with not capturing all the bioactivities of
981 the compounds. Trying different cell lines could allow capturing more bioactivities linked to MIE
982 leading to acute oral toxicity. Several cell lines have already been used with Cell Painting [43,44]
983 and could help define a set of cell lines capable of capturing a maximum, if not all, MIE leading
984 to acute oral toxicity.

985 Finally, absorption, distribution, metabolism, and excretion (ADME) properties of compounds
986 were not taken into consideration in this study but incorporating such data could enhance
987 predictive models. We tried to use predicted maximum concentration in plasma and AUC from
988 a predictive model [45], but this information did not improve our results (data not shown).

989 In addition, pre-incubation of the compounds with liver S9 fractions (the 9000g supernatant of
990 a liver homogenate), containing phase I and II metabolic enzymes, to generate the possible
991 metabolites of a parent compound, could be helpful when the toxicity is driven by a metabolite,
992 as it is done for example in the Ames assay to test the mutagenic potential of chemical
993 compounds [46,47].

994 In conclusion, a combined approach utilizing chemical structure, and Cell Painting
995 morphological profiles-based classifiers based on chemical and biological spaces distances holds
996 promise for predicting acute oral toxicity. These classifiers could be used in the context of early
997 de-risking and in the future serve in the context of Next Generation Risk Assessment (NGRA),
998 which aims at refining if not replacing laboratory animal testing.

999

1000

Funding Information.

Fabrice Camilleri holds a doctoral fellowship from the Association Nationale de la Recherche Technique (ANRT CIFRE PhD funding).

Acknowledgment.

The authors thank Dr. Nina Hallmark and Dr. Oscar Mendes Lucio for their insightful suggestions to improve the article.

Data Availability

GitHub (https://github.com/Bayer-Group/cellPainting_acuteTox)

Consensus Cell Painting morphological profiles with oral acute toxicity label, and code are provided.

Supporting Information

Lisf of reference compounds, list of public compounds, Umap of the chemical space with Butina clusters, other classifiers results, detailed CATMoS predictions, and python package versions.

1018 # References

1019 (1) Khalak, Y.; Tresadern, G.; Hahn, D. F.; De Groot, B. L.; Gapsys, V. Chemical Space
1020 Exploration with Active Learning and Alchemical Free Energies. *J. Chem. Theory Comput.*
1021 **2022**, *18* (10), 6259–6270. https://doi.org/10.1021/acs.jctc.2c00752.

1022 (2) Scannell, J. W.; Bosley, J.; Hickman, J. A.; Dawson, G. R.; Truebel, H.; Ferreira, G. S.;
1023 Richards, D.; Treherne, J. M. Predictive Validity in Drug Discovery: What It Is, Why It
1024 Matters and How to Improve It. *Nat Rev Drug Discov* **2022**.
1025 https://doi.org/10.1038/s41573-022-00552-x.

1026 (3) Henriquez, J. E.; Badwaik, V. D.; Bianchi, E.; Chen, W.; Corvaro, M.; LaRocca, J.; Lunsman, T.
1027 D.; Zu, C.; Johnson, K. J. From Pipeline to Plant Protection Products: Using New Approach
1028 Methodologies (NAMs) in Agrochemical Safety Assessment. *J. Agric. Food Chem.* **2024**, *72*
1029 (19), 10710–10724. https://doi.org/10.1021/acs.jafc.4c00958.

1030 (4) Erhirhie, E. O.; Ihekwereme, C. P.; Ilodigwe, E. E. Advances in Acute Toxicity Testing:
1031 Strengths, Weaknesses and Regulatory Acceptance. *Interdisciplinary Toxicology* **2018**, *11*
1032 (1), 5–12. https://doi.org/10.2478/intox-2018-0001.

1033 (5) Mansouri, K.; Karmaus, A. L.; Fitzpatrick, J.; Patlewicz, G.; Pradeep, P.; Alberga, D.; Alepee,
1034 N.; Allen, T. E. H.; Allen, D.; Alves, V. M.; Andrade, C. H.; Auernhammer, T. R.; Ballabio, D.;
1035 Bell, S.; Benfenati, E.; Bhattacharya, S.; Bastos, J. V.; Boyd, S.; Brown, J. B.; Capuzzi, S. J.;
1036 Chushak, Y.; Ciallella, H.; Clark, A. M.; Consonni, V.; Daga, P. R.; Ekins, S.; Farag, S.;
1037 Fedorov, M.; Fourches, D.; Gadaleta, D.; Gao, F.; Gearhart, J. M.; Goh, G.; Goodman, J. M.;
1038 Grisoni, F.; Grulke, C. M.; Hartung, T.; Hirn, M.; Karpov, P.; Korotcov, A.; Lavado, G. J.;
1039 Lawless, M.; Li, X.; Luechtefeld, T.; Lunghini, F.; Mangiatordi, G. F.; Marcou, G.; Marsh, D.;
1040 Martin, T.; Mauri, A.; Muratov, E. N.; Myatt, G. J.; Nguyen, D.-T.; Nicolotti, O.; Note, R.;
1041 Pande, P.; Parks, A. K.; Peryea, T.; Polash, A. H.; Rallo, R.; Roncaglioni, A.; Rowlands, C.;
1042 Ruiz, P.; Russo, D. P.; Sayed, A.; Sayre, R.; Sheils, T.; Siegel, C.; Silva, A. C.; Simeonov, A.;
1043 Sosnin, S.; Southall, N.; Strickland, J.; Tang, Y.; Teppen, B.; Tetko, I. V.; Thomas, D.;

1044 Tkachenko, V.; Todeschini, R.; Toma, C.; Tripodi, I.; Trisciuzzi, D.; Tropsha, A.; Varnek, A.;
1045 Vukovic, K.; Wang, Z.; Wang, L.; Waters, K. M.; Wedlake, A. J.; Wijeyesakere, S. J.; Wilson,
1046 D.; Xiao, Z.; Yang, H.; Zahoranszky-Kohalmi, G.; Zakharov, A. V.; Zhang, F. F.; Zhang, Z.;
1047 Zhao, T.; Zhu, H.; Zorn, K. M.; Casey, W.; Kleinstreuer, N. C. CATMoS: Collaborative Acute
1048 Toxicity Modeling Suite. *Environ Health Perspect* **2021**, *129* (4), 047013.
1049 https://doi.org/10.1289/EHP8495.

1050 (6) Bray, M.-A.; Singh, S.; Han, H.; Davis, C. T.; Borgeson, B.; Hartland, C.; Kost-Alimova, M.;
1051 Gustafsdottir, S. M.; Gibson, C. C.; Carpenter, A. E. Cell Painting, a High-Content Image-
1052 Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes. *Nat Protoc*
1053 **2016**, *11* (9), 1757–1774. https://doi.org/10.1038/nprot.2016.105.

1054 (7) Chandrasekaran, S. N.; Ceulemans, H.; Boyd, J. D.; Carpenter, A. E. Image-Based Profiling
1055 for Drug Discovery: Due for a Machine-Learning Upgrade? *Nat Rev Drug Discov* **2021**, *20*
1056 (2), 145–159. https://doi.org/10.1038/s41573-020-00117-w.

1057 (8) Lapins, M.; Spjuth, O. Evaluation of Gene Expression and Phenotypic Profiling Data as
1058 Quantitative Descriptors for Predicting Drug Targets and Mechanisms of Action. March 17,
1059 2019. https://doi.org/10.1101/580654.

1060 (9) Tian, G.; Harrison, P. J.; Sreenivasan, A. P.; Carreras-Puigvert, J.; Spjuth, O. Combining
1061 Molecular and Cell Painting Image Data for Mechanism of Action Prediction. *Artificial*
1062 *Intelligence in the Life Sciences* **2023**, *3*, 100060.
1063 https://doi.org/10.1016/j.ailsci.2023.100060.

1064 (10) Simm, J.; Klambauer, G.; Arany, A.; Steijaert, M.; Wegner, J. K.; Gustin, E.; Chupakhin, V.;
1065 Chong, Y. T.; Vialard, J.; Buijnsters, P.; Velter, I.; Vapirev, A.; Singh, S.; Carpenter, A. E.;
1066 Wuyts, R.; Hochreiter, S.; Moreau, Y.; Ceulemans, H. Repurposing High-Throughput Image
1067 Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chemical Biology*
1068 **2018**, *25* (5), 611-618.e3. https://doi.org/10.1016/j.chembiol.2018.01.015.

1069 (11) Nyffeler, J.; Willis, C.; Lougee, R.; Richard, A.; Paul-Friedman, K.; Harrill, J. A. Bioactivity
1070 Screening of Environmental Chemicals Using Imaging-Based High-Throughput Phenotypic
1071 Profiling. *Toxicology and Applied Pharmacology* **2020**, *389*, 114876.
1072 https://doi.org/10.1016/j.taap.2019.114876.

1073 (12) Garcia De Lomana, M.; Marin Zapata, P. A.; Montanari, F. Predicting the Mitochondrial
1074 Toxicity of Small Molecules: Insights from Mechanistic Assays and Cell Painting Data.
1075 *Chem. Res. Toxicol.* **2023**, *36* (7), 1107–1120.
1076 https://doi.org/10.1021/acs.chemrestox.3c00086.

1077 (13) Seal, S.; Carreras-Puigvert, J.; Trapotsi, M.-A.; Yang, H.; Spjuth, O.; Bender, A. Integrating
1078 Cell Morphology with Gene Expression and Chemical Structure to Aid Mitochondrial
1079 Toxicity Detection. *Commun Biol* **2022**, *5* (1), 858. https://doi.org/10.1038/s42003-022-
1080 03763-5.

1081 (14) Lejal, V.; Cerisier, N.; Rouquié, D.; Taboureau, O. Assessment of Drug-Induced Liver Injury
1082 through Cell Morphology and Gene Expression Analysis. *Chem. Res. Toxicol.* **2023**, *36* (9),
1083 1456–1470. https://doi.org/10.1021/acs.chemrestox.2c00381.

1084 (15) Kowalski, B. R.; Bender, C. F. K-Nearest Neighbor Classification Rule (Pattern Recognition)
1085 Applied to Nuclear Magnetic Resonance Spectral Interpretation. *Anal. Chem.* **1972**, *44* (8),
1086 1405–1411. https://doi.org/10.1021/ac60316a008.

1087 (16) Zhu, H. Supporting Read-across Using Biological Data. *ALTEX* **2016**, 167–182.
1088 https://doi.org/10.14573/altex.1601252.

1089 (17) ChemIDplus, 2023. https://pubchem.ncbi.nlm.nih.gov/source/ChemIDplus.

1090 (18) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and
1091 Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets. *J.*
1092 *Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747–750. https://doi.org/10.1021/ci9803381.

1093 (19) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA Models for Predicting
1094 Physicochemical Properties and Environmental Fate Endpoints. *J Cheminform* **2018**, *10* (1),
1095 10. https://doi.org/10.1186/s13321-018-0263-1.

1096 (20) Cimini, B. A.; Chandrasekaran, S. N.; Kost-Alimova, M.; Miller, L.; Goodale, A.; Fritchman,
1097 B.; Byrne, P.; Garg, S.; Jamali, N.; Logan, D. J.; Concannon, J. B.; Lardeau, C.-H.; Mouchet,
1098 E.; Singh, S.; Shafqat Abbasi, H.; Aspesi, P.; Boyd, J. D.; Gilbert, T.; Gnutt, D.; Hariharan, S.;
1099 Hernandez, D.; Hormel, G.; Juhani, K.; Melanson, M.; Mervin, L. H.; Monteverde, T.; Pilling,
1100 J. E.; Skepner, A.; Swalley, S. E.; Vrcic, A.; Weisbart, E.; Williams, G.; Yu, S.; Zapiec, B.;
1101 Carpenter, A. E. Optimizing the Cell Painting Assay for Image-Based Profiling. *Nat Protoc*
1102 **2023**, *18* (7), 1981–2013. https://doi.org/10.1038/s41596-023-00840-9.

1103 (21) Stirling, D. R.; Swain-Bowden, M. J.; Lucas, A. M.; Carpenter, A. E.; Cimini, B. A.; Goodman,
1104 A. CellProfiler 4: Improvements in Speed, Utility and Usability. *BMC Bioinformatics* **2021**,
1105 *22* (1), 433. https://doi.org/10.1186/s12859-021-04344-9.

1106 (22) Serrano, E.; Chandrasekaran, S. N.; Bunten, D.; Brewer, K. I.; Tomkinson, J.; Kern, R.;
1107 Bornholdt, M.; Fleming, S.; Pei, R.; Arevalo, J.; Tsang, H.; Rubinetti, V.; Tromans-Coia, C.;
1108 Becker, T.; Weisbart, E.; Bunne, C.; Kalinin, A. A.; Senft, R.; Taylor, S. J.; Jamali, N.;
1109 Adeboye, A.; Abbasi, H. S.; Goodman, A.; Caicedo, J. C.; Carpenter, A. E.; Cimini, B. A.;
1110 Singh, S.; Way, G. P. Reproducible Image-Based Profiling with Pycytominer. **2023**.
1111 https://doi.org/10.48550/ARXIV.2311.13417.

1112 (23) *Benchmarking Grit*. Benchmarking Grit. https://github. com/broadinstitute/grit-
1113 benchmark (accessed 2024-05-29).

1114 (24) *Cytominer-eval: Evaluating quality of perturbation profiles*. Cytominer-eval: Evaluating
1115 quality of perturbation profiles. https://github.com/cytomining/cytominereval (accessed
1116 2024-05-29).

(25) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5* (2), 107–113. https://doi.org/10.1021/c160017a018.

(26) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. https://doi.org/10.1021/ci100050t.

(27) Landrum, G.; Tosco, P.; Kelley, B.; Ric; Cosgrove, D.; Sriniker; Gedeck; Vianello, R.; NadineSchneider; Kawashima, E.; N, D.; Jones, G.; Dalke, A.; Cole, B.; Swain, M.; Turk, S.; AlexanderSavelyev; Vaucher, A.; Wójcikowski, M.; Ichiru Take; Probst, D.; Ujihara, K.; Scalfani, V. F.; Godin, G.; Lehtivarjo, J.; Walker, R.; Pahl, A.; Francois Berenger; Jasondbiggs; Strets123. Rdkit/Rdkit: 2023_03_3 (Q1 2023) Release, 2023. https://doi.org/10.5281/ZENODO.591637.

(28) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(29) Escher, S. E.; Bitsch, A. Read-Across Methodology in Toxicological Risk Assessment. In *Regulatory Toxicology*; Reichl, F.-X., Schwenk, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2021; pp 1–14. https://doi.org/10.1007/978-3-642-36206-4_132-1.

(30) Seal, S.; Yang, H.; Trapotsi, M.-A.; Singh, S.; Carreras-Puigvert, J.; Spjuth, O.; Bender, A. Merging Bioactivity Predictions from Cell Morphology and Chemical Fingerprint Models Using Similarity to Training Data. *J Cheminform* **2023**, *15* (1), 56. https://doi.org/10.1186/s13321-023-00723-x.

(31) Cristianini, N.; Ricci, E. Support Vector Machines: 1992; Boser, Guyon, Vapnik. In *Encyclopedia of Algorithms*; Kao, M.-Y., Ed.; Springer US: Boston, MA, 2008; pp 928–932. https://doi.org/10.1007/978-0-387-30162-4_415.

1142 (32) Nadeau, C.; Bengio, Y. Inference for the Generalization Error. *Machine Learning* **2003**, *52*
1143 (3), 239–281. https://doi.org/10.1023/A:1024068626366.

1144 (33) Sainburg, T.; McInnes, L.; Gentner, T. Q. Parametric UMAP Embeddings for Representation
1145 and Semisupervised Learning. *Neural Computation* **2021**, *33* (11), 2881–2907.

1146 (34) Kar, S.; Roy, K.; Leszczynski, J. Applicability Domain: A Step Toward Confident Predictions
1147 and Decidability for QSAR Modeling. In *Computational Toxicology*; Nicolotti, O., Ed.;
1148 Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1800, pp 141–
1149 169. https://doi.org/10.1007/978-1-4939-7899-1_6.

1150 (35) Baillif, B.; Wichard, J.; Méndez-Lucio, O.; Rouquié, D. Exploring the Use of Compound-
1151 Induced Transcriptomic Data Generated From Cell Lines to Predict Compound Activity
1152 Toward Molecular Targets. *Front. Chem.* **2020**, *8*, 296.
1153 https://doi.org/10.3389/fchem.2020.00296.

1154 (36) Moshkov, N.; Becker, T.; Yang, K.; Horvath, P.; Dancik, V.; Wagner, B. K.; Clemons, P. A.;
1155 Singh, S.; Carpenter, A. E.; Caicedo, J. C. Predicting Compound Activity from Phenotypic
1156 Profiles and Chemical Structures. *Nat Commun* **2023**, *14* (1), 1967.
1157 https://doi.org/10.1038/s41467-023-37570-1.

1158 (37) Ljosa, V.; Caie, P. D.; Ter Horst, R.; Sokolnicki, K. L.; Jenkins, E. L.; Daya, S.; Roberts, M. E.;
1159 Jones, T. R.; Singh, S.; Genovesio, A.; Clemons, P. A.; Carragher, N. O.; Carpenter, A. E.
1160 Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to
1161 Small-Molecule Treatment. *J Biomol Screen* **2013**, *18* (10), 1321–1329.
1162 https://doi.org/10.1177/1087057113503553.

1163 (38) Way, G. P.; Natoli, T.; Adeboye, A.; Litichevskiy, L.; Yang, A.; Lu, X.; Caicedo, J. C.; Cimini, B.
1164 A.; Karhohs, K.; Logan, D. J.; Rohban, M. H.; Kost-Alimova, M.; Hartland, K.; Bornholdt, M.;
1165 Chandrasekaran, S. N.; Haghighi, M.; Weisbart, E.; Singh, S.; Subramanian, A.; Carpenter, A.
1166 E. Morphology and Gene Expression Profiling Provide Complementary Information for

1167 Mapping Cell State. *Cell Systems* **2022**, *13* (11), 911-923.e9.
1168 https://doi.org/10.1016/j.cels.2022.10.001.

1169 (39) Prieto, P. Investigating Cell Type Specific Mechanisms Contributing to Acute Oral Toxicity.
1170 *ALTEX* **2019**, *36* (1), 39–64. https://doi.org/10.14573/altex.1805181.

1171 (40) A. Leblanc, G. Mechanisms of Acute Toxicity. In *A Textbook of Modern Toxicology*; 2004.

1172 (41) Caicedo, J. C.; McQuin, C.; Goodman, A.; Singh, S.; Carpenter, A. E. Weakly Supervised
1173 Learning of Single-Cell Feature Embeddings. In *2018 IEEE/CVF Conference on Computer*
1174 *Vision and Pattern Recognition*; IEEE: Salt Lake City, UT, 2018; pp 9309–9318.
1175 https://doi.org/10.1109/CVPR.2018.00970.

1176 (42) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Cherkasov, A.; Li, J.; Gramatica, P.;
1177 Hansen, K.; Schroeter, T.; Müller, K.-R.; Xi, L.; Liu, H.; Yao, X.; Öberg, T.; Hormozdiari, F.;
1178 Dao, P.; Sahinalp, C.; Todeschini, R.; Polishchuk, P.; Artemenko, A.; Kuz'min, V.; Martin, T.
1179 M.; Young, D. M.; Fourches, D.; Muratov, E.; Tropsha, A.; Baskin, I.; Horvath, D.; Marcou,
1180 G.; Muller, C.; Varnek, A.; Prokopenko, V. V.; Tetko, I. V. Applicability Domains for
1181 Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *J.*
1182 *Chem. Inf. Model.* **2010**, *50* (12), 2094–2111. https://doi.org/10.1021/ci100253r.

1183 (43) Cox, M. J.; Jaensch, S.; Van de Waeter, J.; Cougnaud, L.; Seynaeve, D.; Benalla, S.; Koo, S. J.;
1184 Van Den Wyngaert, I.; Neefs, J.-M.; Malkov, D.; Bittremieux, M.; Steemans, M.; Peeters, P.
1185 J.; Wegner, J. K.; Ceulemans, H.; Gustin, E.; Chong, Y. T.; Göhlmann, H. W. H. Tales of 1,008
1186 Small Molecules: Phenomic Profiling through Live-Cell Imaging in a Panel of Reporter Cell
1187 Lines. *Sci Rep* **2020**, *10* (1), 13262. https://doi.org/10.1038/s41598-020-69354-8.

1188 (44) Nyffeler, J. Phenotypic Profiling for High-Throughput Chemical Bioactivity Screening at the
1189 U.S. EPA. **2020**, 11457644 Bytes. https://doi.org/10.23645/EPACOMPTOX.13198346.

1190 (45) Schneckener, S.; Grimbs, S.; Hey, J.; Menz, S.; Osmers, M.; Schaper, S.; Hillisch, A.; Göller,
1191 A. H. Prediction of Oral Bioavailability in Rats: Transferring Insights from in Vitro
1192 Correlations to (Deep) Machine Learning Models Using in Silico Model Outputs and

1193 Chemical Structure Parameters. *J. Chem. Inf. Model.* **2019**, *59* (11), 4893–4905.
1194 https://doi.org/10.1021/acs.jcim.9b00460.

1195 (46) Hakura, A.; Suzuki, S.; Sawada, S.; Motooka, S.; Satoh, T. An Improvement of the Ames
1196 Test Using a Modified Human Liver S9 Preparation. *Journal of Pharmacological and*
1197 *Toxicological Methods* **2001**, *46* (3), 169–172. https://doi.org/10.1016/S1056-
1198 8719(02)00186-7.

1199 (47) Hopperstad, K.; DeGroot, D. E.; Zurlinden, T.; Brinkman, C.; Thomas, R. S.; Deisenroth, C.
1200 Chemical Screening in an Estrogen Receptor Transactivation Assay With Metabolic
1201 Competence. *Toxicological Sciences* **2022**, *187* (1), 112–126.
1202 https://doi.org/10.1093/toxsci/kfac019.

1203

1204