MORE IS DIFFERENT: DRUG PROPERTY ANALYSIS ON CELLULAR HIGH-CONTENT IMAGES USING DEEP LEARNING

Xiangrui Gao[1†], Xueyu Guo[1†], Fan Zhang[1†], Mengcheng Yao[1†], Xiaoxiao Wang[1†], Dong Chen[1], Xiaodong Wang[1]*, Lipeng Lai[1]*

[1] XtalPi Innovation Center, Beijing, China.

[†] Authors contributed equally to this work.

* Correspondence: xiaodong.wang@xtalpi.com, lipeng.lai@xtalpi.com

ABSTRACT

High-content analysis (HCA) holds enormous potential for drug discovery and research, but widely used methods can be cumbersome and yield inaccurate results. Noise and high similarity in cell images impede the accuracy of deep learning-based image analysis. To address these issues, we introduce More Is Different (MID), a novel HCA method that combines cellular experiments, image processing, and deep learning modeling. MID effectively combines the convolutional neural network and Transformer to encode high-content images, effectively filtering out noisy signals and characterizing cell phenotypes with high precision. In comparative tests on drug-induced cardiotoxicity and mitochondrial toxicity classification, as well as compound classification, MID outperformed both DeepProfiler and CellProfiler, which are two highly recognized methods in HCA. We believe that our results demonstrate the utility and versatility of MID and anticipate its widespread adoption in HCA for advancing drug development and disease research.

INTRODUCTION

Developing novel pharmaceutical drugs represents a substantial investment that involves significant amount of time and resources, but with a low success rate. The major obstacles hindering drug development include ineffective drug activity, intractable drug toxicity, as well as marketing difficulties. However, the recent development of cellular phenotyping technology in drug discovery has demonstrated as a valuable tool to overcome the above issues. For example, Recursion's phenotypic drug development system, the Recursion OS, has successfully driven several drugs to the clinical stage, showcasing its effectiveness [1].

31 Image processing and feature extraction are essential steps in HCA. One of the main tools currently in
32 use is CellProfiler [2], developed by the imaging platform of Broad Institute, which relies on traditional
33 image processing techniques. It enables biologists with no training in computer vision or programming to
34 automatically identify and quantify phenotypes from thousands of images. Furthermore, deep learning
35 models are increasingly employed for image feature extraction. Cimini and Carpenter et al. utilized
36 ImageNet pretrained models for feature extraction and compared them with CellProfiler's features. Their
37 results showed some level of improvement, leading them to present DeepProfiler [3]. This tool extracts
38 cell-slice images based on the cellular localization provided by CellProfiler and uses a weakly supervised
39 approach to classify DMSO and drug-treated images.

40 Today, the success of deep learning heavily relies on the availability of vast amounts of data and large
41 advanced models. Massive data helps to train the model effectively without premature overfitting, while
42 the reasonableness and complexity of the network structure endow the model with good memory and
43 information extraction ability. In addition, the diversity of images within the enormous data makes training
44 and prediction of models easier by eliminating complex preprocessing steps, such as segmentation,
45 tracking, tracing, and spatial conversion. As a result, the task can perform end-to-end prediction directly,
46 thereby simplifying the entire process.

47 However, simply migrating advanced computer vision models to HCA is not always successful, and
48 larger models do not necessarily yield better results. For example, DeepProfiler, in practice, is less
49 effective than a machine learning model based on cell phenotype features extracted by CellProfiler when
50 predicting the drug properties of a compound. The problem lies in the fact that DeepProfiler is not
51 optimized for cell images and only classifies individual cell image, making it difficult to be utilized for
52 compound classification or compound property prediction from the cell-slice images [3]. Moreover, a
53 single-cell phenotype may not always represent the property distinction of a compound, as it may not
54 have been influenced by the compound or may have died naturally during the experiment. Determining
55 the phenotypic changes involving many cells provides a more reasonable resource for predicting a
56 compound's properties. Although CellProfiler counts the characteristics of many cells to make judgments,
57 the phenotype of most cells may be undifferentiated, and some important information may be ignored
58 during the counting process.

59 To address the aforementioned issues, we have developed MID - a deep learning-based method for
60 processing cell phenotype images (Figure 1). The concept behind MID is inspired by P.W. Anderson's
61 quote "More is different" [4]. In condensed matter physics, the behavior of large and complex elementary
62 particle aggregates cannot be understood by simply extrapolating the properties of a few elementary
63 particles. This also holds true for the problem of cellular phenotypes. During cellular experiments, the
64 compound effects on cells are random, and the properties of compounds cannot be accurately
65 distinguished from images of one or a few cells. Therefore, a more reasonable approach would be to
66 consider many cell-slice images. By examining more cell-slice images, the model can learn to distinguish
67 between important and unimportant images. Additionally, the information contained in multiple cell-slice
68 images, even some contain noise, may reveal certain indiscernible patterns. Using a combination of
69 convolutional neural networks and Transformers, MID can adaptively filter out irrelevant cell-slice images
70 and extract features that accurately model cell phenotype characteristics. For different cell-slice images
71 exposed to the same compound or compounds with similar drug properties, MID can provide similar
72 features, while features extracted by MID differ more between compounds with distinct drug properties.

73   Compared to DeepProfiler and CellProfiler, MID is a novel HCA process that leverages our

74   understanding of cellular images and deep learning. It excels in screening and utilizing informative and

75   high-quality cell-slice images to extract task-relevant cellular features, resulting in superior performance

76   in three drug property-related tasks, i.e. 1) determination of drug inhibition on hEGR ion channels, 2)

77   prediction of drug-induced mitochondrial toxicity, and 3) classification of compounds. In addition, MID can

78   concatenate different compounds based on similar cell phenotypes, highlighting its potential in the field of

79   drug repurposing and indication expansions.

80   **RESULTS**

81   **MID performance on hERG inhibition classification**

82   hERG, also known as Kv11.1, is a gene (KCNH2) that codes for an alpha subunit protein of the

83   potassium ion channel [5]. This channel, which is essential for the electrical activity of the heart, mediates

84   the repolarizing IKr current in the cardiac action potential and coordinates heartbeats [6]. However, some

85   drugs in the market that inhibit hERG have the potential to prolong the QT interval and cause a

86   dangerous irregularity of the heartbeat called torsades de pointes, making hERG inhibition an important

87   factor to consider during drug development [7,8,9]. Currently, there are several in vitro methods for

88   evaluating hERG inhibition, electrophysiological (membrane clamp, the gold standard) [10],

89   cytofluorimetric-based assays [11], and ligand binding assays [12,13], but they are often complicated, low

90   throughput, and costly. We believe that the MID model can effectively determine whether a drug has the

91   potential to inhibit hERG.

92   To test this hypothesis, we conducted high-content screening (HCS) experiments on 100 compounds,

93   including water and DMSO, 57 of which were toxic and 45 non-toxic, using human induced pluripotent

94   stem cell-derived cardiomyocytes (hiPSC-CMs). From these compounds, we selected 38 (19 non-toxic

95   and 19 toxic) for the test dataset. We then used the images generated from our HCS experiments to

96   evaluate the ability of three models - MID, DeepProfiler, and CellProfiler - to classify the cardiotoxicity of

97   the drugs based on their ability to inhibit hERG channels.

98   Our results show that MID outperformed both DeepProfiler and CellProfiler, achieving an accuracy score

99   of 90.6 in classifying images produced by drugs with or without hERG inhibition. We also calculated

100   precision and recall scores for each model and found that MID had a precision score of 80.0 and a recall

101   score of 95.5, indicating that it was capable of making correct predictions (Fig. 2a). Although

102   CellProfiler-LightGBM had a higher recall score, it had a lower precision score, incorrectly predicting

103   almost all negative cases as positive. In summary, on the classification task of hERG channel inhibition,

104   MID performed best, followed by DeepProfiler, with CellProfiler performing the worst. Moreover, in this

105   task, we try to demonstrate the differential impact of hERG channel inhibitors and non-hERG channel

106   inhibitors on cell phenotype, as well as the effectiveness of MID in capturing this kind of difference. To

107   achieve this, we randomly selected 10 compounds (5 labeled as 1 and 5 labeled as 0) and investigated

108   their effects on cells at varying drug concentrations using MID model prediction scores. As depicted in

109   Fig. 2d, the model evaluation scores exhibited distinct patterns for hERG channel inhibitors and

110   non-hERG channel inhibitors as drug concentration increased. The scores for hERG channel inhibitors

111   were higher, and with increasing concentration, the model evaluation scores increased and gradually

112   reached a plateau or remained at a high level. The scores for non-hERG channel inhibitors were

113   relatively low, and with increasing concentration, the model evaluation scores showed a trend of first

114   increasing and then decreasing (possibly because these drugs have less impact on cells at low

115   concentrations) or remained at a low level.

116   To fully interpret the performance of the MID model, we analyzed samples of its false predictions,

117   including eltrombopag, pitolisant, and sildenafil, in the last section of the Results (MID bad cases contain

118   information that deserves in-depth analysis).

119

120   **MID performance on Mitochondrial toxicity classification**

121   Mitochondrial toxicity caused by certain compounds is a widespread form of organ toxicity, which can

122   result in multiorgan damage in the heart, liver, bone, and brain [14, 15, 16]. Regulators have withdrawn

123   several drugs, including benfluorex, isoprenaline, nifedipine, and rosiglitazone, from the market due to

124   their side effects, including mitochondrial toxicity and other adverse reactions [17]. Furthermore,

125   mitochondrial toxicity may be associated with drug resistance to linezolid [18]. Therefore, mitochondrial

126   toxicity is a critical factor in determining the success of drug development. Multiple mechanisms

127   contribute to mitochondrial toxicity [19], resulting in various changes in cell phenotype. The alterations in

128   cell morphology, texture, and intensity caused by compounds are strongly correlated with mitochondrial

129   toxicity, suggesting that cell phenotype analysis is a reliable method for predicting mitochondrial toxicity

130   [21].

131   In this study, we utilized public HCA images of 285 compounds from previous research [22]. After labeling

132   with PubChem annotation data, we selected a test set of 95 compounds, of which 18 were labeled

133   'Active', 51 were labeled 'Inactive', and 26 were labeled 'Inconclusive'. Similar to the hERG task, we

134   evaluated the classification performance of the three models. The MID model still outperforms the other

135   two models in terms of specific metrics, as shown in Fig. 2b.

136   There may be two reasons for this. Firstly, the label's definition is unclear. Secondly, the cellular portraits

137   used in this task were stained with generic cellular dyes and were not specifically designed for observing

138   drug mitochondrial toxicity. The use of general-purpose dyes to predict mitochondrial toxicity makes it

139   more challenging for the models to capture the exact information or features that genuinely represent the

140   mitochondrial toxicity of drugs, resulting in poor classification performance. In contrast, for the hERG task,

141   we used cell dyes that specifically reflect the cardiotoxicity of drugs in HCA experiments. Customized

142   experiments facilitated the MID model's ability to capture information that is valid for making judgments.

143   This demonstrates the importance of combining dry and wet experiments in areas related to drug

144   discovery.

145

146   **MID performance on Compound classification**

147   The compound prediction task involves classifying cell images based on their response to different

148   compounds.In DeepProfiler, compound classification is a supervised training task that helps the model

149   extract valuable image features. Our results showed that the MID model outperformed DeepProfiler,

150   achieving an accuracy score of 86.7 compared to 10.7, a precision score of 89.2 compared to 11.4, a

151 recall score of 87.1 compared to 9.7, and a F1 score of 88.1 compared to 10.5. Note that the machine

152 learning model that utilizes CellProfiler features is not available here (Fig. 2c).

153

154 **MID can act as an effective cell-slice filter**

155 To demonstrate the ability of MID to extract relevant information, we generated an attention heatmap

156 using MID, which highlighted the correlation between cells in the HCA image and the cardiotoxicity task

157 (Fig. 3a and Fig. 3b). As indicated by the self-attention mechanism, the similarity between the

158 embeddings of single cell-slice image and the embeddings of the CLS token is positively associated with

159 the classification task [23]. In the heatmap, cells that are more relevant to the downstream task are

160 represented by brighter and warmer dots, and the dots in Fig. 3b clearly illustrate the differences between

161 different cells in terms of task correlation.

162 To determine whether the Transformer can isolate cell slices of interest, we analyzed the features of

163 task-relevant single cell-slice image and the irrelevant ones, and based on the attention map values, we

164 can define how relevant the image is to the task. As shown in Fig. 3c, using MID embeddings, the cell

165 slices unrelated to the cardiotoxicity of the compound are tightly clustered regardless of whether the

166 compounds are identical, while the cell slices related to drug-induced cardiotoxicity are separately

167 clustered according to compounds, with most of the compounds widely spaced apart. This demonstrates

168 that the multiple cell slices encoder can function as a cell filter. The clustering results also provide two

169 other insights. Firstly, in cellular experiments using different compounds with different properties, there

170 are always some dead cells or normal cells that are not affected by the compound, which is the common

171 part and does not contribute much to classification. The characteristics of all these cells after the single

172 cell-slice encoder are similar, and their clustering together indirectly confirms that the model has learned

173 the correct information. Secondly, after being exposed to different compounds, the phenotype of the cells

174 changes more significantly and they become very distinct from each other, even when they are dead. We

175 also clustered the single cell-slice image features extracted by CellProfiler (Fig. 3d) and DeepProfiler (Fig.

176 3e), and because CellProfiler and DeepProfiler did not distinguish whether those single cell-slice images

177 were task-related, the features they extracted were not representative of different compounds and had

178 low discriminatory power compared to MID.

179

180 **Task-relevant cell-slice images are highly informative**

181 To further demonstrate the screening capabilities of MID for cell phenotype images, we analyzed and

182 compared cell-slice images that were considered relevant and irrelevant based on the MID attention map

183 (Fig. 3f, Fig. 3g). Each cell-slice image was comprised of three channels: nucleus, mitochondrial reactive

184 oxygen species (ROS), and mitochondrial membrane potential (MMP). Highly correlated images (Fig. 3f)

185 exhibited deeper staining, higher fluorescence intensity, greater contrast between channels, and more

186 accurate staining in the corresponding channels, as compared to the low-correlation images (Fig. 3g).

187 Furthermore, highly correlated images displayed more distinct cell structure in the ROS and MMP

188 channels, resulting in more accurate cell identification, while most cells in the low-correlation images

189 exhibited the opposite characteristics. In terms of image quality, high-correlation images were clear and

190 free of impurities, whereas low-correlation images had a halo on the image surface due to optical

191 structure formation, leading to indistinct images and high noise. Thus, the attention mechanism utilized in
192 MID allowed for the extraction of highly informative, accurate, and structurally intact cell-slice images
193 while filtering out low-informative, erroneous, and poor-quality images, thereby improving the
194 performance of the model after image preprocessing.
195

196 **Multiple cell slices enable accurate grasp of the compound property**

197 To illustrate the core idea of MID, we clustered the features from single cell-slice relevant images and
198 multiple cell slices relevant images (Figure 4). First, using the MID features of single cell-slice images
199 (Fig. 4a), the clusters of different compounds were mixed compared to those of multiple cell-slice images
200 (Fig. 4b). This suggests that incorporating more image slices provides more information, and MID can
201 anchor the useful parts from numerous single cell-slice images and integrate them to predict the specific
202 property of compounds. The feature clusters of compounds (benzethonium chloride and betrixaban)
203 changed from being mixed (Fig. 4a) to forming independent clusters (Fig. 4b), which verifies this idea well.
204 Secondly, when comparing the feature clusters of single cell-slice images obtained using MID with those
205 obtained using CellProfiler (Fig. 4c) and DeepProfiler (Fig. 4e), the distances between the single
206 cell-slice image embeddings of MID, under the influence of different compounds, were still large, and
207 their boundaries were still distinct (Fig. 4a). Moreover, the feature clusters of MID using multiple cell-slice
208 images (Fig. 4b) were more accurate compared to those of CellProfiler (Fig. 4d) and DeepProfiler (Fig.
209 4f). When using multiple cell-slice images, the advantages of MID are evident as it can efficiently
210 differentiate between different compounds. The clustering of belzutifan, carvedilol and daclatasvir, which
211 are closely clustered compared to other compounds (Fig. 4b), may suggest that the cell phenotype
212 images corresponding to these three compounds are similar.
213

214 **The MID embeddings can represent the cell phenotypes**
215 If the embedding of MID can accurately represent the cellular phenotype, this suggests that compounds
216 whose MID embeddings cluster together are likely to have very similar cell phenotypes. To test this
217 hypothesis, we compared images of three drugs (belzutifan, carvedilol, and daclatasvir) as shown in Fig.
218 5a. However, it was challenging for us to discern the differences between the three drugs based solely on
219 their cellular phenotypes.
220 Then, we separately analyzed the changes in MMP and ROS intensity of cells treated with different
221 concentrations of three drugs, each of which may have different effects on certain cell statistical
222 indicators. We also compared the effects of these three drugs with that of betrixaban, a drug that is much
223 farther away in terms of embedding distance. Our results show that the changes in MMP (Fig. 5b) and
224 ROS (Fig. 5c) intensity are similar for the three drugs with similar cell phenotypes, while betrixaban
225 behaves differently.
226 To better understand the results presented in Figure 5, we researched the mechanism of action (MOA) of
227 the three drugs in question. Despite their different indications - Belzutifan for von Hippel-Lindau (VHL)
228 syndrome-associated clear-cell renal cell carcinoma (ccRCC), Carvedilol for hypertension, and
229 Daclatasvir for Chronic hepatitis C genotype 3 (GT-3 HCV) - their MOAs suggest that they may have
230 similar effects on cardiomyocytes.

231 Belzutifan inhibits hypoxia-inducible factor 2α (HIF-2α), and research suggests that HIF-2α inhibitors can

232 reverse pulmonary hypertension and that there are shared pathophysiologic mechanisms between

233 cancer and heart failure [24, 25]. Carvedilol, a nonselective beta-adrenergic antagonist, has anti-free radical

234 and antioxidant effects, and can resist oxidation and reduce ROS production [26]. It also inhibits the

235 Cardiac Mitochondrial Permeability Transition (MPT), which can depolarize mitochondrial membranes

236 and uncouple oxidative phosphorylation (OXPHOS) [27]. Daclatasvir, a pangenotypic NS5A replication

237 complex inhibitor with a dual antiviral effect, inhibits RNA replication and viral assembly. Cellular ROS

238 levels rise during HCV infection, and evidence suggests that anti-RNA viral drugs are associated with

239 intracellular ROS levels [28]. ROS may also play a critical role as a signal molecule in the regulation of

240 viral replication and organelle function [29].

241

242 **MID bad cases contain information that deserves in-depth analysis**

243 During the hERG classification task, our model inaccurately identified three drugs, eltrombopag,

244 pitolisant, and sildenafil, as cardiotoxic based on hERG IC50 values, but being predicted as non-toxic by

245 MID. To understand the reasons behind these false predictions, we analyzed and compared the cell

246 painting staining images of cells treated with toxic and non-toxic compounds.

247 Figure 6 displays the cell painting staining results of the cells after treatment with the compounds, where

248 the nucleus is blue, ROS is green, MMP is yellow, and dead cells are red. The images depicted in Fig. 6b

249 represent the staining results of three significantly non-toxic compounds, namely irbesartan with hEGR

250 IC50 values of 194 μM, riluzole with 25 μM, and belzutifan with 50 μM. The staining channels showed an

251 intact cell structure with uniformly distributed green fluorescence and almost no dead cells. The images

252 shown in Fig. 6c illustrate the staining results of three significantly toxic compounds, namely nintedanib

253 with hEGR IC50 values of 2.6 μM, sunitinib with 0.25-10 μM, and ponatinib with 0.77 μM. In nintedanib,

254 green fluorescence presented a punctate distribution around the nuclei, and dead cells were visible. In

255 sunitinib, the number of dead cells increased, and most of the cells lost their intact cell structure. In

256 ponatinib, almost all the cells were dead. Based on the recognizable cardiotoxicity represented by the

257 images, for the compounds with incorrect predictions shown in Fig. 6a, the images were more similar to

258 those in Fig. 6b. That is, the cells had an intact structure, the green fluorescence was uniformly lamellar,

259 and there were almost no dead cells. Therefore, we suggested that the model tends to identify these

260 three compounds as non-toxic. Additionally, as shown in Supplementary Figure 1, the clusters of MID

261 embeddings of all 100 compounds, plus DMSO and water (N), eltrombopag was located close to

262 irbesartan and surrounded by six other non-toxic compounds (darunavir, lesinurad, omarigliptin,

263 miltefosine, glibenclamide, and rosiglitazone). Pitolisant was close to belzutifan, and sildenafil almost

264 overlapped with riluzole. Furthermore, based on some previous literature, we found that none of the three

265 drugs themselves had significant cardiotoxicity in clinical or noncliİnical studies [30,31,32,33,34,35]. In summary,

266 the comparison of cell painting images and supporting evidence from the literature suggest that the

267 results of MID's wrong predictions are informative and worthy of attention.

268

269 **DISCUSSION**

270   In this study, we developed a highly accurate computational model (MID) using a deep learning
271   framework to extract relevant information from high-content images in HCS experiments for drug
272   property analysis. We thoroughly studied the characteristics of cell phenotype images to construct the
273   method from data processing to model training and validation. To assess the potential value of our
274   method in high-content cellular image processing, we selected three widely used downstream tasks in
275   drug discovery: Task 1 - classification of drug inhibition of the hERG ion channel, Task 2 - classification of
276   drug mitochondrial toxicity, and Task 3 - classification of compounds.
277   To validate our approach, we designed HCS experiments using a batch of compounds with known hERG
278   IC50 values to produce cell images for tasks 1 and 3. We used open-source high-content image data to
279   construct datasets for task 2. After applying MID to the downstream tasks, we analyzed the results in
280   detail. Our findings indicate that MID outperforms CellProfiler and DeepProfiler in tasks 1 and 3.
281   Additionally, MID performs better than CellProfiler and DeepProfiler in task 2, with results consistent with
282   those reported in the literature [21] (Figure 2).
283   In the hERG task, we compared the original cell-slice images input to the MID model and analyzed the
284   valid and invalid information extracted by MID. Our analysis revealed that MID accurately identifies
285   informative and high-quality cell-slice images using a self-attention mechanism (Figure 3). We also found
286   that MID predicts the properties of compounds more accurately when using more cell-slice images
287   (Figure 4). Furthermore, we analyzed biologically significant cell phenotype features extracted by
288   CellProfiler for images judged to be similar by MID but disturbed by different drugs. We found that
289   compounds with different indications but similar cell phenotypes, such as belzutifan, carvedilol, and
290   daclatasvir, were highly consistent in the ROS and MMP intensities of the corresponding images (Figure
291   5). This confirms the accuracy of the judgment from MID and suggests that MID models may be useful in
292   high-throughput applications in the field of drug indication expansion and drug repurposing. Lastly, we
293   analyzed MID bad cases by comparing images and searching the literature. Our analysis suggested that
294   three compounds defined as cardiotoxic by hERG IC50 values were nontoxic according to the image,
295   literature, and MID predictions (Figure 6). Overall, our results demonstrate that MID is a highly effective
296   tool for high-content cell image analysis, with potential applications in drug discovery and development.
297   In conclusion, HCS is an emerging field that is still rapidly evolving in terms of experimental
298   implementation and analytical methods and has the potential to solve diverse biological problems. The
299   key to achieving credible results in downstream tasks lies in the ability of analytical tools or computational
300   models to select useful parts from a large number of cell images of varying quality. Among the tested
301   solutions for downstream tasks, CellProfiler and DeepProfiler performed mediocre, while MID showed
302   promising results. This is because MID, which leverages deep learning to use multiple cell slices
303   simultaneously for model training and verification, can accurately capture key information and eliminate
304   noise interference. Additionally, MID can distinguish between cells with similar phenotypes but perturbed
305   by different compounds, suggesting many possibilities for downstream applications. Overall, the
306   conclusion emphasizes the potential of MID as a valuable tool for high-content cell image analysis, which
307   can contribute to various fields such as drug discovery and development, disease diagnosis, and
308   personalized medicine.

309 In the next phase of our research, we plan to expand the application of MID to more cell phenotyping
310 tasks and high-content images induced by other perturbations, such as RNA interference (RNAi) or
311 CRISPR-Cas9. Our goal is to demonstrate the versatility and effectiveness of our model in analyzing
312 various types of cellular image data. However, we also acknowledge that algorithms based on cell
313 phenotype images have limitations, particularly when the changes in cell phenotype induced by a
314 compound are minimal, which may make it difficult for the model to accurately assess its toxicity or other
315 characteristics. To address this issue, we plan to explore the integration of other high-level information,
316 such as transcriptome data or videos of cells under bright field, using multimodal learning techniques. We
317 believe that incorporating multiple sources of information will enhance the accuracy and robustness of
318 our model and enable it to handle a wider range of biological problems.

319 **METHODS**

320 **Constructing the cardiotoxicity high-content image data**

321 To obtain an appropriate image dataset for the hERG inhibition task, we selected 100 compounds for
322 HCS experiments and generated corresponding high-content images using a high-content imager.
323 Supplementary Table 2 provides detailed information for each compound, including plate name, drug
324 name, hERG IC50 value range, and cardiotoxicity label. We first downloaded all compounds with hERG
325 IC50 values using the python API interface of the chEMBL Database and subsequently selected 100
326 compounds by deduplicating and filtering drug molecules with ambiguous hERG IC50 values. The
327 compounds were then labeled as cardiotoxic and noncardiotoxic based on their hERG IC50 values using
328 a threshold of 10 µM (less than or equal to 10 µM for toxic labeling and greater than 10 µM for nontoxic
329 labeling). This resulted in 47 cardiotoxic and 53 non-cardiotoxic compounds, on which we performed
330 cellular experiments using hiPSC-CMs.
331 The hiPSC-CMs at day 30 after cardiac induction were cryopreserved as Cauliscell hiPSC-CMs
332 (Cauliscell Company, Nanjing, China) and thawed in a 37°C water bath with gentle shaking. After
333 centrifugation and counting, the cells were added to 384-well plates precoated with 10 µg/ml recombinant
334 human vitronectin at 12,000 cells/well (Cauliscell) in cardiomyocyte plating medium (Cauliscell). After 24
335 hours, the volume was replaced with cardiomyocyte maintenance medium, which was changed every
336 other day. Once the cells started to beat rhythmically, we added the compounds, which were selected
337 from the FDA-approved compound library (FDA-approved drug library, MedChemExpress) for
338 cardiotoxicity testing. The working concentrations of each compound were 10 µM, 3.33 µM, 1.11 µM,
339 0.37 µM, 0.12 µM, and 0.04 µM, with 3 replicate instances set for each compound at each concentration.
340 The control group was set with 0.1% DMSO (Sigma), and the blank control was set with water. After
341 incubating the cells with compounds for 72 hours, working concentration dyes of CM-H2DCFDA (Thermo)
342 at 5 µM, TMRM (Thermo) at 20 nM, and Hoechst33342 (Thermo) at 5 µg/ml were prepared with
343 maintenance medium (Cauliscell) and added to the cells for 30 min. The cells were washed with HBSS
344 (Beyotime Biotechnology), and YOYO-3 (Thermo) was added at a final concentration of 2 nM, followed
345 by incubation at room temperature for 10 min and washing with HBSS (Beyotime Biotechnology). Finally,
346 the 384-well plate (PerkinElmer) was placed on a high-content imager (Operetta CLS, PerkinElmer), and
347 images were taken with a 20x water objective in the confocal model with 9 fields of view for one well,
348 corresponding to the filter parameters shown in Supplementary Table 1.

349

350 **Constructing the mitochondrial toxicity high-content image data**

351 For the mitochondrial toxicity task, we utilized open-source high-content images from Bray et al. [20]. A
352 total of fifty-five plate high-content images with numerous compounds were downloaded, and the
353 mitochondrial toxicity labels, obtained from the PubChem Database
354 (https://pubchem.ncbi.nlm.nih.gov/assay/pcget.cgi?query=download&record_type=datatable&actvty=all
355 &response_type=save&aid=720637), were assigned to each compound. After labeling, 285 compounds
356 were collected, of which 47 were classified as active, 141 as inactive, and 97 as inconclusive. Detailed
357 information regarding the drug names and mitochondrial toxicity labels from PubChem can be found in
358 Supplementary Table 3. Supplementary Table 4 shows the datasets and models used in the three tasks.

359

360 **CellProfiler and DeepProfiler data processing**

361 For the cardiotoxicity and mitochondrial toxicity high-content images, we extracted cell phenotype
362 features using standard procedures of CellProfiler (version 4.2.4). Subsequently, we utilized the machine
363 learning model called LGBMClassifier from scikit-learn to perform classification tasks on toxicity data,
364 after dealing with cell phenotype features. Once the locations of single cells were identified in the images
365 by CellProfiler, we inputted the images and locations to DeepProfiler (version 0.3.1). During classification
366 tasks, we employed a pretrained model named EffientNet, which was deployed within DeepProfiler, and
367 calculated embeddings using DeepProfiler. The CellProfiler-LightGBM and DeepProfiler results were
368 evaluated and compared with MID to assess their accuracy and generalizability. In all three models (MID,
369 DeepProfiler and CellProfiler-LightGBM), we partitioned the 285 compounds labeled with mitochondrial
370 toxicity into a training set comprising 190 compounds and a test set containing 95 compounds. For the
371 100 cardiotoxic compounds, the training and testing sets contained 68 and 32 compounds, respectively.

372

373 **ROS and MMP Measurement**

374 To measure ROS and MMP, we utilized cell phenotype features calculated from our high-content images
375 of 100 compounds, which were used for cardiotoxicity assessment. We designed two CellProfiler
376 analysis protocols that can automatically detect and quantify fluorescence intensity, which proved to be
377 useful for analyzing large image datasets. The ROS and MMP measurements were obtained from the
378 suppressed fluorescent channels of the H2DCFDA and TMRM cell-based assay kit. We computed a
379 reduced dataset with the well-mean feature vector per well, followed by normalizing all features by
380 subtracting the mean of each plate layout from each feature.

381

382 **MID model Design**

383 Cell phenotype images are distinct from general images in two key ways. First, they contain a high
384 degree of redundancy, with dark backgrounds and bright cells dominating most of the image. As a result,
385 it can be challenging to distinguish differences between cells using generic training. Second, there is a lot
386 of noise in cell phenotype images due to experimental manipulations and batches, requiring different
387 concentration gradients and experimental replicates to eliminate stochastic factors. Existing models and
388 processing techniques have struggled to address these challenges, resulting in poor results.

389 To address these issues, we developed the MID model, a deep learning model specifically designed for
390 cell phenotype images. MID is a plug-and-play flexible model framework that is not restricted to a
391 particular backbone. The model processes cell-slice images by normalizing and grouping them by
392 compound and concentration. First, single cell locations are extracted from CellProfiler by calculating the
393 center coordinate of the nucleus. Second, images are rescaled with the global fluorescence intensity, and
394 third, DeepProfiler crops cell slices from 3 channels based on the x and y coordinates for the center of a
395 single nucleus. In all experiments, cell slices were cropped from a region of 96 x 96 pixels centered on
396 the nucleus without resizing. The resulting 96 x 288-pixel images were preserved for model construction.
397 Cell-slice images stained with different dyes were integrated as different channels into a single cell-slice
398 multichannel image. This approach reduces noise interference and allows the model to perform the
399 classification task using a few representative cells. Furthermore, the training phase inputs a limited
400 number of cell images at a time to prevent the model from overfitting prematurely, which brought by
401 complex information of too many cell images.

402 The MID model consists of three main parts: a single cell-slice encoder, a multiple cell slices encoder,
403 and a classifier. The single cell-slice encoder uses a uniform CNN network to obtain an unbiased latent
404 representation of single cell-slice image and improve the generalization of the latent representation. The
405 multiple cell slices encoder is built by Transformer, which has strong contextualization capabilities to
406 integrate information from each element in the sequence. A CLS token is added to the top of the
407 sequence to filter out irrelevant cell representations and improve model robustness. The self-attentive
408 mechanism of Transformer is used to fuse the latent representations corresponding to the CLS token,
409 which are then fed into the classifier constructed by a linear layer for classification.

410

411 **MID Training and Inference**

412 During the training process, we used the standard supervised training method with cross-entropy as the
413 loss function, along with learning rate warm-up and cosine decay techniques. In order to account for
414 various experimental and computational factors, as well as potential inaccuracies in the labels
415 themselves, we also employed label smoothing to reduce the model's confidence.

416 In the testing phase, we made a slight modification to the training approach. While in training we
417 randomly selected 12 cell-slice images to form multiple cell-slices sequences, in testing we increased the
418 number of cell slices included in each sequence. To determine the number of cell slices used in the test
419 phase, we tested the performance of the model with different patch numbers selected. As shown in
420 Figure 1, the overall performance of the model increases with the increase of patch number. In order to
421 take into account the performance of the model and the consumption of the calculation at the same time,
422 we randomly sampled 50 sets of multiple cell slices and computed the average classification result from
423 those sequences as the final outcome. This approach offers the advantage of improving the chances of
424 selecting valuable cell-slice images while also reducing computational costs by not evaluating all
425 individual cells.

426

427 **MID attention map**

428 As we utilized the CLS token embeddings generated by Transformer for classification purposes, we
429 obtained the attention map by calculating the dot product between the query representation and key

430 representation among the tokens in Transformer. By applying the Softmax function, each element of the

431 attention map was assigned a value ranging from 0 to 1. Subsequently, we visualized the individual

432 attention map values as solid dots with distinct colors and labeled the corresponding cells in the cell

433 phenotype image (Fig. 3b).

434

435 **Cluster and statistical analysis**

436 To evaluate the performance of feature extraction methods for cell-slice image analysis, we employed

437 the t-SNE algorithm in the python sklearn package to reduce the dimensionality of features or

438 embeddings computed by MID, CellProfiler, and DeepProfiler. We then examined the resulting

439 component distributions to investigate the relationship between MID and the cell phenotype observed in

440 images, as well as the effectiveness of MID compared to CellProfiler or DeepProfiler in capturing useful

441 information. To assess the clustering performance, we used the silhouette_score method in sklearn to

442 calculate the silhouette coefficient. For the statistical analysis presented in Figure 5, we applied the

443 Kruskal⌐Wallis test to test the null hypothesis that the MMP or ROS intensity of images treated with

444 compound at six different concentrations (0.04 µM, 0.12 µM, 0.37 µM, 0.11 µM, 3.33 µM, and 10 µM)

445 were equal, setting significance at $p < 0.05$.

446

447 **DATA AVAILABILITY**

448 The data used in this article will be provided after the paper is accepted 'in principle'.

449 **CODE AVAILABILITY**

450 The MID code was implemented in Python using the deep learning framework of PyTorch. Code, trained

451 models, and scripts reproducing the experiments of this article will be provided after the paper is

452 accepted 'in principle'.

453 **AUTHOR INFORMATION**

454 All authors have given approval to the final version of the manuscript.

455 **Contributions**

456 Xiaodong W and Lipeng L conceived this research. Fan Z, Mengcheng Y and Xueyu G curated the

457 dataset. Xiangrui G, Fan Z and Xueyu G performed data analysis. Xiangrui.G devised deep learning

458 algorithms. Xiaoxiao W ang Dong C conducted the HCS experiments. Xiangrui Gao and Xueyu Guo

459 wrote and modified the paper. Xiaodong W and Lipeng L supervised this work.

460 **EITHICS DECLARATIONS**

461 **Competing interests**

462 The authors declare no competing interests.

463 **ACKNOWLEDGEMENTS**

464 The authors thank Genwei Zhang for helpful comments and proofreading.

465

466 **ABBREVIATIONS**

467 ccRCC
468 *clear-cell renal cell carcinoma*
469 CLS token
470 *Classification* Token
471 CRISPR-cas
472 *Clustered Regularly Interspaced Short Palindromic Repeat- CRISPR-associated （Cas） systems*
473 DMSO
474 *Dimethyl sulfoxide*
475 HCA
476 *High-content analysis*
477 HCS
478 *high-content screening*
479 hEGR
480 *the human Ether-à-go-go-Related* Gene
481 hiPSC-CMs
482 *human induced pluripotent stem cell-derived cardiomyocytes*
483 GT-3 HCV
484 *hepatitis C genotype 3*
485 MID
486 *More Is Different*
487 MMP
488 *mitochondrial membrane potential*
489 MOA
490 *mechanism of action*
491 MPT
492 *Mitochondrial Permeability Transition*
493 OXPHOS

494 *oxidative phosphorylation*

495 RNA

496 *Ribonucleic Acid*

497 ROS

498 *mitochondrial reactive oxygen species*

499 VHL

500 *von Hippel-Lindau*

501

REFERENCES

503   1.   Gustafsdottir SM, Ljosa V, Sokolnicki KL, Anthony Wilson J, Walpita D, Kemp MM, Petri Seiler K, Carrel HA, Golub TR, Schreiber SL,
504   Clemons PA. Multiplex cytological profiling assay to measure diverse cellular states. PloS one. 2013 Dec 2;8(12):e80999.

505   2.   Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. CellProfiler 4: improvements in speed, utility and
506   usability. BMC bioinformatics. 2021 Dec;22(1):1-1.

507   3.   Moshkov N, Bornholdt M, Benoit S, McQuin C, Smith M, Goodman A, Senft R, Han Y, Babadi M, Horvath P, Cimini BA. Learning
508   representations for image-based profiling of perturbations. bioRxiv. 2022 Jan 1.

509   4.   Anderson PW. More is different: broken symmetry and the nature of the hierarchical structure of science. Science. 1972
510   Aug4;177(4047):393-6

511   5.   Gutman GA, Chandy KG, Adelman JP, Aiyar J, Bayliss DA, Clapham DE, Covarriubias M, Desir GV, Furuichi K, Ganetzky B, Garcia ML.
512   International Union of Pharmacology. XLI. Compendium of voltage-gated ion channels: potassium channels. Pharmacological reviews. 2003
513   Dec 1;55(4):583-6.

514   6.   Sanguinetti MC, Tristani-Firouzi M. hERG potassium channels and cardiac arrhythmia. Nature. 2006 Mar;440(7083):463-9.

515   7.   Redfern WS, Carlsson L, Davis AS, Lynch WG, MacKenzie IL, Palethorpe S, Siegl PK, Strang I, Sullivan AT, Wallis R, Camm AJ.
516   Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs:
517   evidence for a provisional safety margin in drug development. Cardiovascular research. 2003 Apr 1;58(1):32-45.

518   8.   Haverkamp W, Breithardt G, Camm AJ, Janse MJ, Rosen MR, Antzelevitch C, Escande D, Franz M, Malik M, Moss A, Shah R. The potential
519   for QT prolongation and pro-arrhythmia by non-anti-arrhythmic drugs: clinical and regulatory implications: report on a Policy Conference of
520   the European Society of Cardiology. Cardiovascular research. 2000 Aug 1;47(2):219-33.

521   9.   Fanoe S, Kristensen D, Fink-Jensen A, Jensen HK, Toft E, Nielsen J, Videbech P, Pehrson S, Bundgaard H. Risk of arrhythmia induced by
522   psychotropic medications: a proposal for clinical management. European heart journal. 2014 May 21;35(20):1306-15.

523   10.   Danker T, Möller C. Early identification of hERG liability in drug discovery programs by automated patch clamp. Frontiers in pharmacology.
524   2014 Sep 2;5:203.

525   11.   Molokanova E, Savchenko A. Bright future of optical assays for ion channel drug discovery. Drug discovery today. 2008 Jan 1;13(1-2):14-22.

526   12.   Deacon M, Singleton D, Szalkai N, Pasieczny R, Peacock C, Price D, Boyd J, Boyd H, Steidl-Nichols JV, Williams C. Early evaluation of
527   compound QT prolongation effects: a predictive 384-well fluorescence polarization binding assay for measuring hERG blockade. Journal of
528   pharmacological and toxicological methods. 2007 May 1;55(3):255-64.

529   13.   Diaz GJ, Daniell K, Leitza ST, Martin RL, Su Z, McDermott JS, Cox BF, Gintant GA. The [3H] dofetilide binding assay is a predictive
530   screening tool for hERG blockade and proarrhythmia: Comparison of intact cell and membrane preparations and effects of altering [K+] o.
531   Journal of pharmacological and toxicological methods. 2004 Nov 1;50(3):187-99.

532   14.   Dykens JA, Will Y. The significance of mitochondrial toxicity testing in drug development. Drug discovery today. 2007 Sep
533   1;12(17-18):777-85.

534   15.   Begriche K, Massart J, Robin MA, Borgne-Sanchez A, Fromenty B. Drug-induced toxicity on mitochondria and lipid metabolism: mechanistic

535   diversity and deleterious consequences for the liver. Journal of hepatology. 2011 Apr 1;54(4):773-94.

536   16.   Hargreaves IP, Al Shahrani M, Wainwright L, Heales SJ. Drug-induced mitochondrial toxicity. Drug safety. 2016 Jul;39(7):661-74.

537   17.   Tang X, Wang Z, Hu S, Zhou B. Assessing Drug-Induced Mitochondrial Toxicity in Cardiomyocytes: Implications for Preclinical Cardiac

538   Safety Evaluation. Pharmaceutics. 2022 Jun 21;14(7):1313.

539   18.   Soriano A, Miró O, Mensa J. Mitochondrial toxicity associated with linezolid. New England Journal of Medicine. 2005 Nov

540   24;353(21):2305-6.

541   19.   Meyer JN, Hartman JH, Mello DF. Mitochondrial toxicity. Toxicological Sciences. 2018 Mar 1;162(1):15-23.

542   20.   Bray MA, Gustafsdottir SM, Rohban MH, Singh S, Ljosa V, Sokolnicki KL, Bittker JA, Bodycombe NE, Dančík V, Hasaka TP, Hon CS. A

543   dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. Gigascience. 2017

544   Dec;6(12):giw014.

545   21.   Seal S, Carreras-Puigvert J, Trapotsi MA, Yang H, Spjuth O, Bender A. Integrating cell morphology with gene expression and chemical

546   structure to aid mitochondrial toxicity detection. Communications Biology. 2022 Aug 23;5(1):858.

547   22.   Haghighi M, Caicedo JC, Cimini BA, Carpenter AE, Singh S. High-dimensional gene expression and morphology profiles of cells across

548   28,000 genetic and chemical perturbations. Nature Methods. 2022 Dec;19(12):1550-7.

549   23.   Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural

550   information processing systems. 2017;30.

551   24.   Ghosh MC, Zhang DL, Ollivierre WH, Noguchi A, Springer DA, Linehan WM, Rouault TA. Therapeutic inhibition of HIF-2α reverses

552   polycythemia and pulmonary hypertension in murine models of human diseases. Blood. 2021 May 6;137(18):2509-19.

553   25.   de Wit S, Glen C, de Boer RA, Lang N. Mechanisms shared between cancer, heart failure and targeted anti-cancer therapies. Cardiovascular

554   Research. 2022 Jul 26.

555   26.   Niemann B, Li L, Simm A, Molenda N, Kockskämper J, Boening A, Rohrbach S. Caloric restriction reduces sympathetic activity similar to

556   beta-blockers but conveys additional mitochondrio-protective effects in aged myocardium. Scientific reports. 2021 Jan 21;11(1):1-5.

557   27.   C Pereira G, M Silva A, V Diogo C, S Carvalho F, Monteiro P, J Oliveira P. Drug-induced cardiac mitochondrial toxicity and protection: from

558   doxorubicin to carvedilol. Current pharmaceutical design. 2011 Jul 1;17(20):2113-29.

559   28.   Lozano-Sepúlveda SA, Rincón-Sanchez AR, Rivas-Estilla AM. Antioxidants benefits in hepatitis C infection in the new DAAs era. Annals of

560   hepatology. 2019 May 1;18(3):410-5.

561   29.   Reshi ML, Su YC, Hong JR. RNA viruses: ROS-mediated cell death. International journal of cell biology. 2014 May 8;2014.

562   30.   Zaninetti C, Gresele P, Bertomoro A, Klersy C, De Candia E, Veneri D, Barozzi S, Fierro T, Alberelli MA, Musella V, Noris P. Eltrombopag for

563   the treatment of inherited thrombocytopenias: a phase II clinical trial. Haematologica. 2020 Mar;105(3):820.

564   31.   Kim TO, Despotovic J, Lambert MP. Eltrombopag for use in children with immune thrombocytopenia. Blood advances. 2018 Feb

565   27;2(4):454-61.

566   32.   Yang R, Li J, Jin J, Huang M, Yu Z, Xu X, Zhang X, Hou M. Multicentre, randomised phase III study of the efficacy and safety of eltrombopag

567   in Chinese patients with chronic immune thrombocytopenia. British Journal of Haematology. 2017 Jan;176(1):101-10.

568   33.   Ligneau X, Shah RR, Berrebi□Bertrand I, Mirams GR, Robert P, Landais L, Maison□Blanche P, Faivre JF, Lecomte JM, Schwartz JC.

569   Nonclinical cardiovascular safety of pitolisant: comparing International Conference on Harmonization S7B and Comprehensive in vitro

570   Pro□arrhythmia Assay initiative studies. British journal of pharmacology. 2017 Dec;174(23):4449-63.

571   34.   Yang X, Ribeiro AJ, Pang L, Strauss DG. Use of human iPSC-CMs in nonclinical regulatory studies for cardiac safety assessment.

572   Toxicological Sciences. 2022 Dec;190(2):117-26.

573   35.   Shinlapawittayatorn K, Chattipakorn S, Chattipakorn N. Effect of sildenafil citrate on the cardiovascular system. Brazilian journal of medical

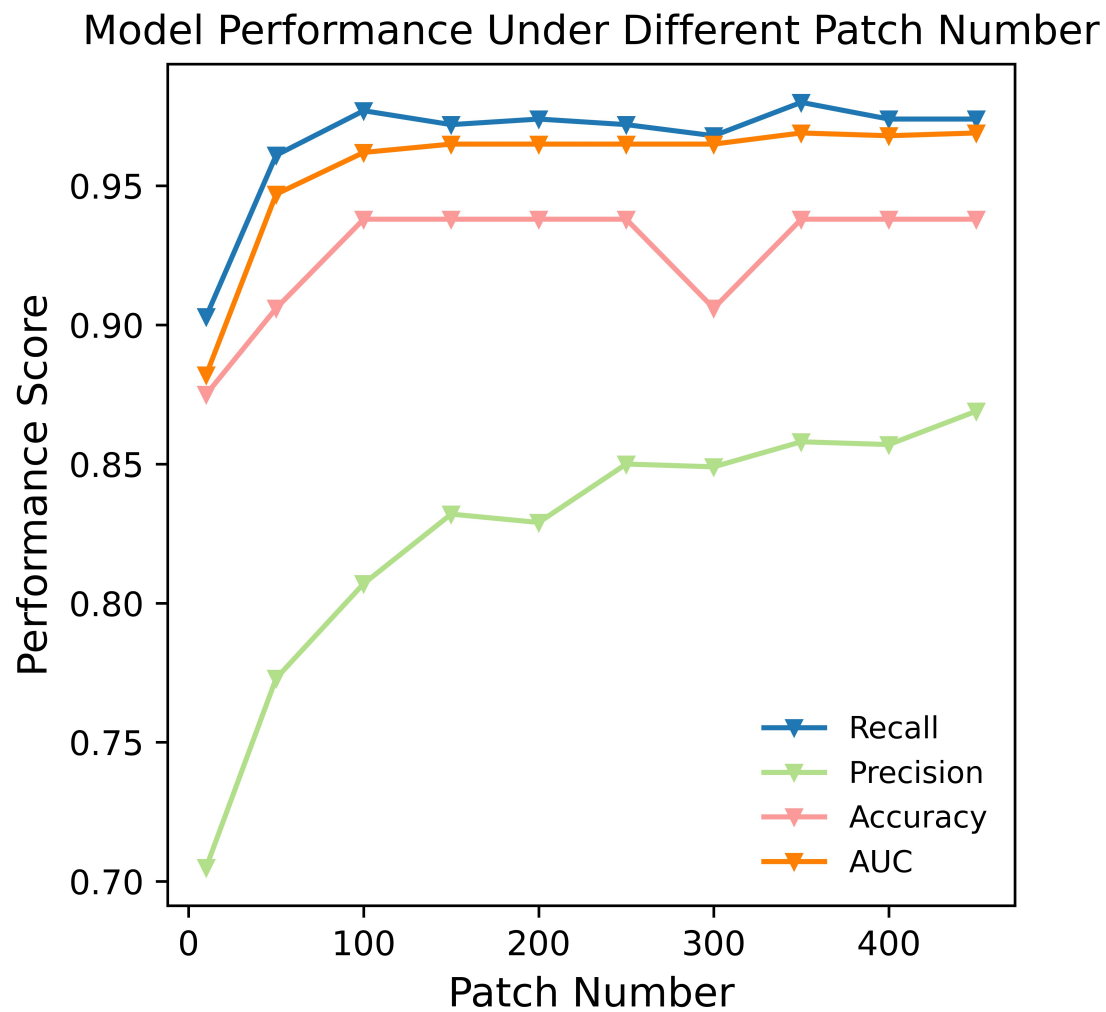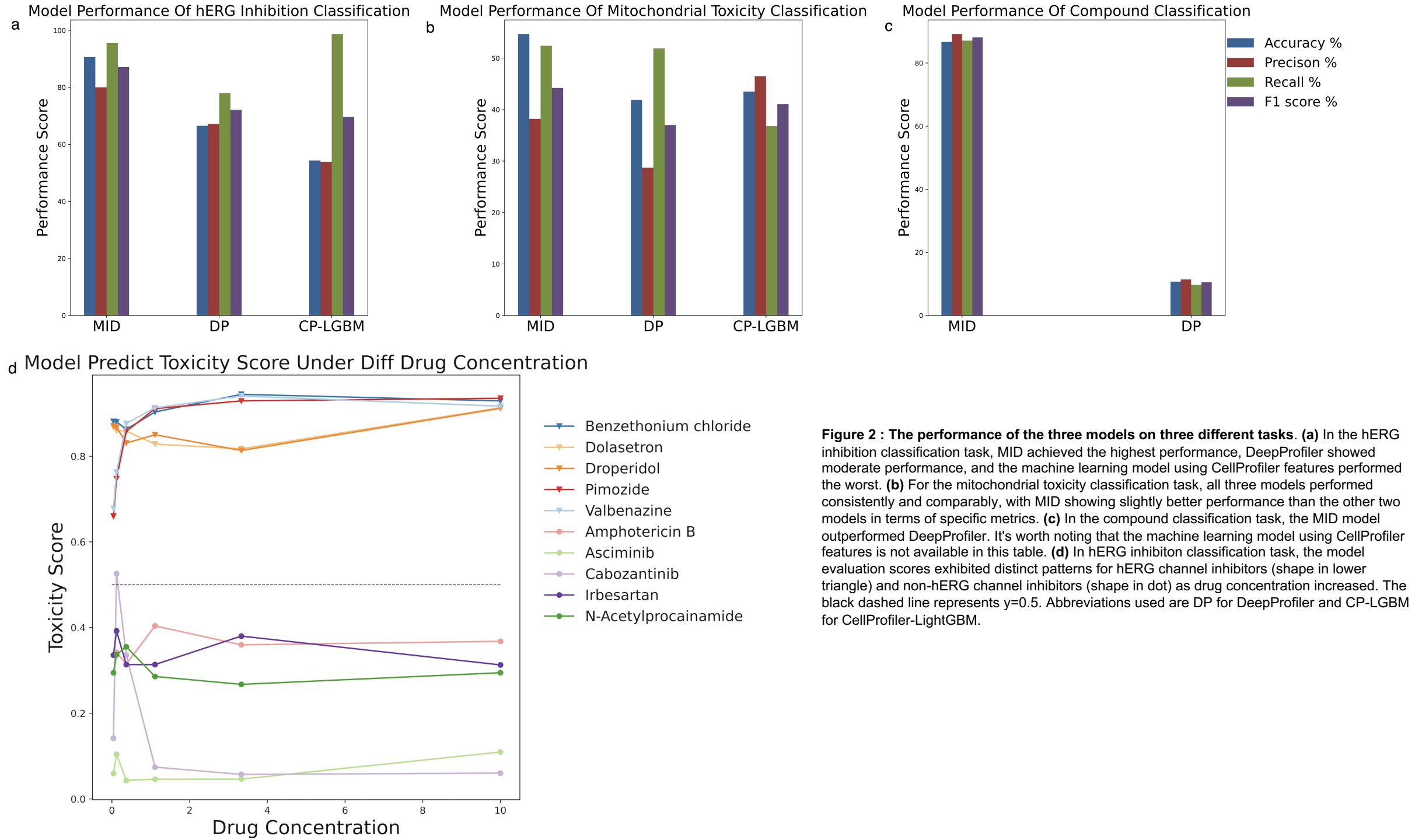574   and biological research. 2005;38:1303-11.

Figure 1: The performance of the MID (More Is Different) model under different patch number. The model prediction performance under different patch number of single cell-slice images in the test phase. The overall performance of the model increases with the increase of patch number.
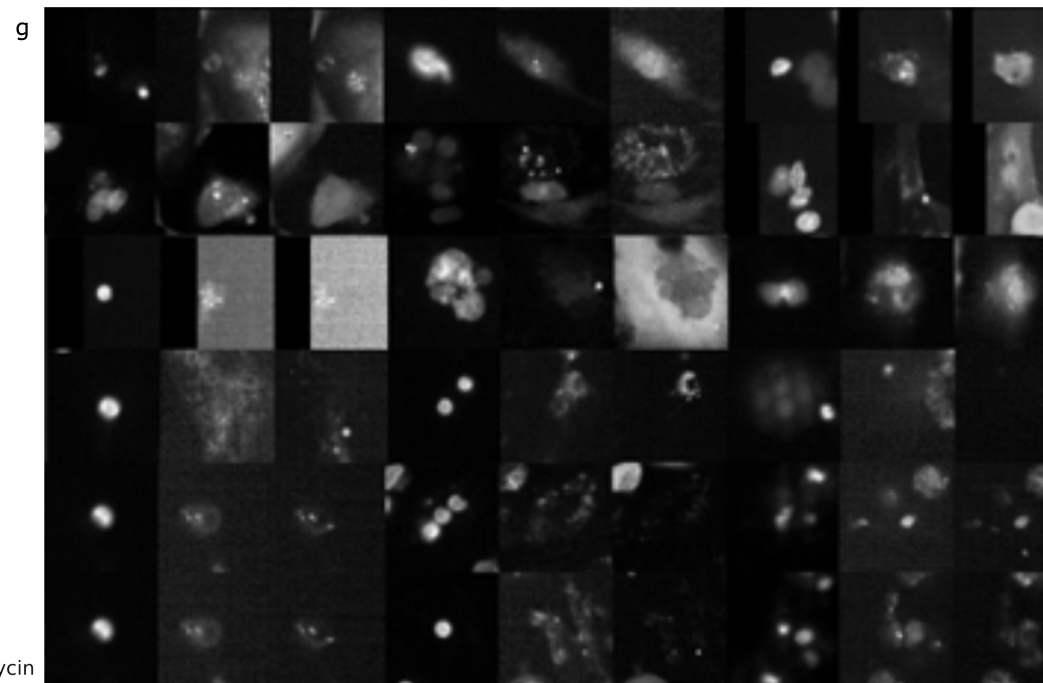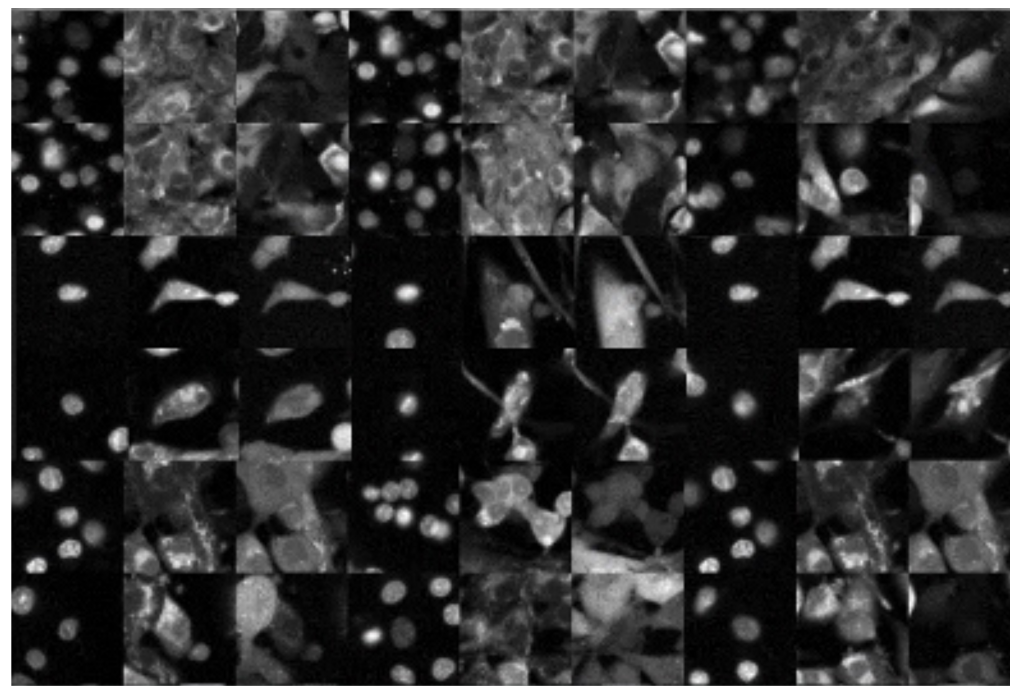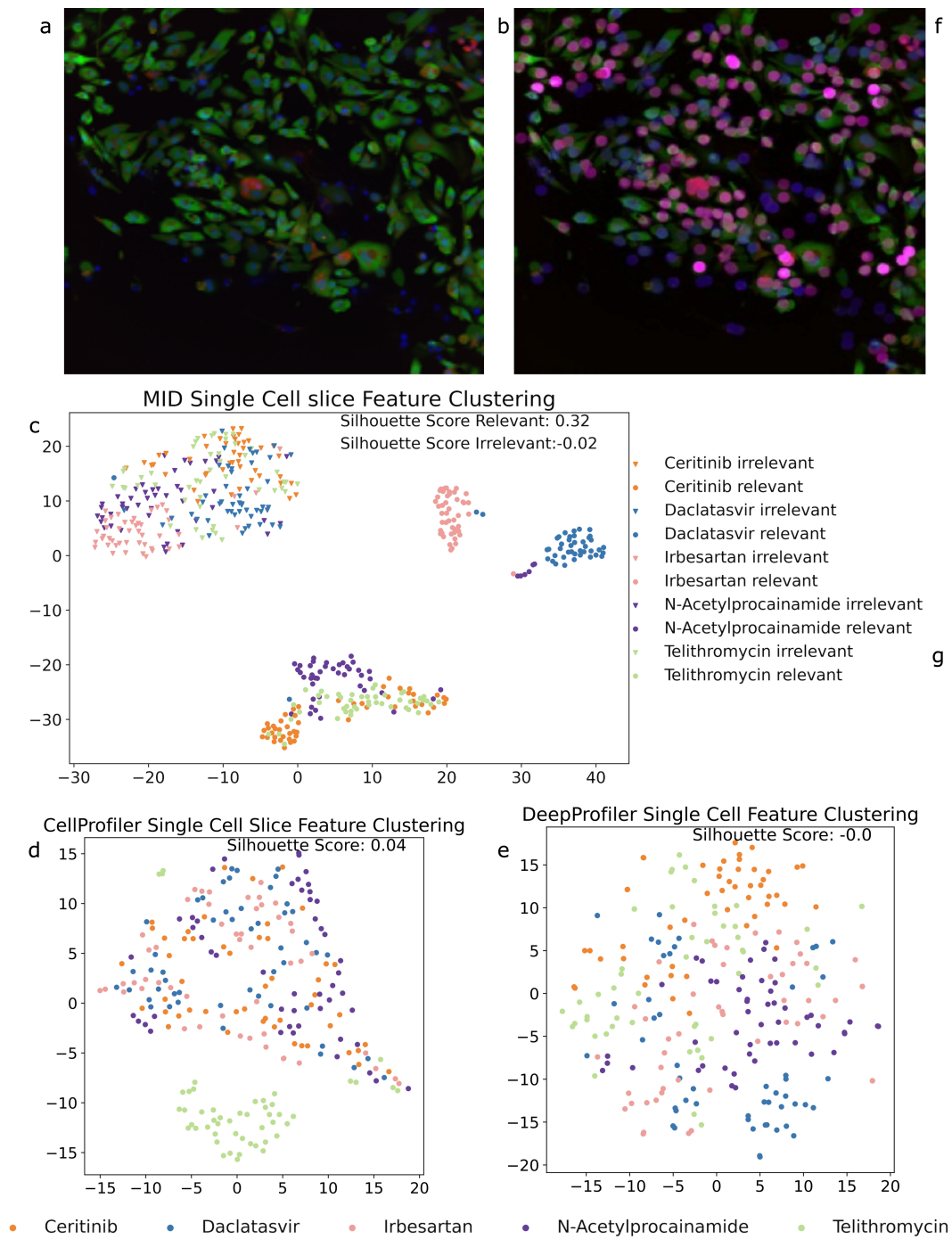
Figure 2 : The performance of the three models on three different tasks. **(a)** In the hERG inhibition classification task, MID achieved the highest performance, DeepProfiler showed moderate performance, and the machine learning model using CellProfiler features performed the worst. **(b)** For the mitochondrial toxicity classification task, all three models performed consistently and comparably, with MID showing slightly better performance than the other two models in terms of specific metrics. **(c)** In the compound classification task, the MID model outperformed DeepProfiler. It's worth noting that the machine learning model using CellProfiler features is not available in this table. **(d)** In hERG inhibiton classification task, the model evaluation scores exhibited distinct patterns for hERG channel inhibitors (shape in lower triangle) and non-hERG channel inhibitors (shape in dot) as drug concentration increased. The black dashed line represents y=0.5. Abbreviations used are DP for DeepProfiler and CP-LGBM for CellProfiler-LightGBM.

Figure 3: MID can act as a filter to identify cell slices that are relevant to a downstream classification task. We used an attention heatmap to visualize the correlation between the CLS embeddings of cell phenotype images and the downstream task. Cells that are more relevant to the task are highlighted in brighter, warmer colors. We added these colors to the Cell Painting image to demonstrate the ability of MID to recognize task-related and nonrelevant cells. By clustering the MID embeddings of relevant and irrelevant cell-slice images from different compounds, we found that MID can effectively filter out nonrelevant cells and identify those that are relevant for classification. The relevant images have a deeper staining level, more intact cell structure, and higher quality than the irrelevant images. **(a)** Cell painting image of cells treated with alfuzosin. Nucleus in blue, ROS in green, MMP in red. **(b)** Attention heatmap of the cell painting image of cells treated with alfuzosin. Some randomly selected cells in the cell painting image were marked with colored solid dots. The color changes gradually from dark blue to bright pink, which means that the marked cells are increasingly relevant to the downstream task. **(c)** We clustered the MID embeddings of relevant (shape in dot) and irrelevant (shape in lower triangle) cell-slice images of 5 compounds and found that the relevant cell-slice images from different compounds clustered separately and were apart from each other. In contrast, all the irrelevant cell-slice images gathered into a large cluster regardless of which compound they belonged to. This phenomenon indicates that the multiple cell slices classifier in our MID model can act as a cell filter. **(d)** Feature clusters of CellProfiler of cell-slice images. For each compound, we randomly selected 50 cell slices **(e)** Feature clusters of DeepProfiler of cell-slice images. For each compound, we randomly selected 50 cell slices. Compared with MID, the cell-slice feature distribution of CellProfiler and DeepProfiler was more dispersed; on the one hand, features of the same compound were not aggregated into a compact cluster, and on the other hand, features of different compounds were mixed. **(f)** Relevant images of belzutifan, carvedilol and daclatasvir. Each three columns (the first column is Nucleus, the second is ROS, the third is MMP) correspond to one compound and each compound has six rows of sampled images, of which the cell staining level is sufficient, the cell structure is clear and easy to read, and the image is of high quality and has less background noise. **(g)** Irrelevant images of belzutifan, carvedilol and daclatasvir. Each three columns (the first column is Nucleus, the second is ROS, the third is MMP) correspond to one compound and each compound has six rows of sampled images, of which the cell staining level is inadequate, the cell structure is ambiguous, and the quality is unqualified and full of noise. ROS, mitochondrial reactive oxygen species. MMP, mitochondrial membrane potential.
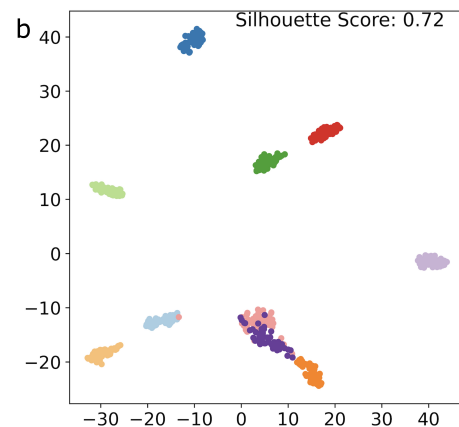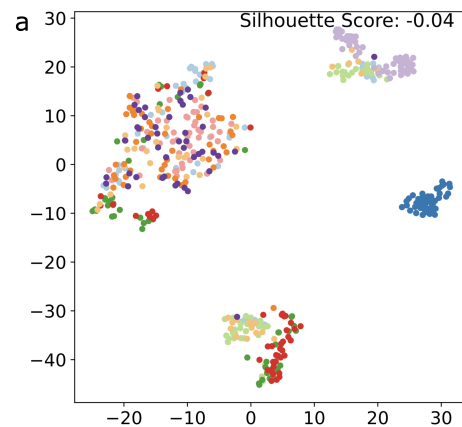
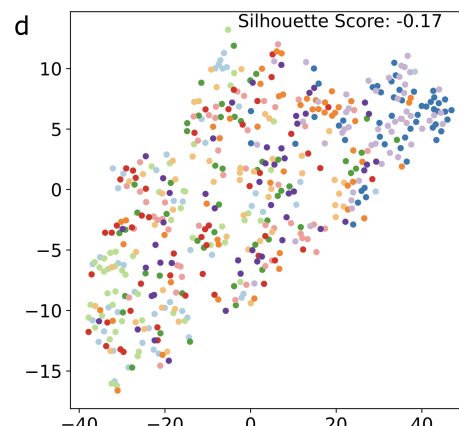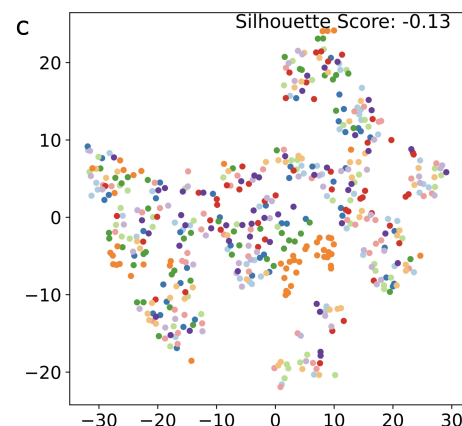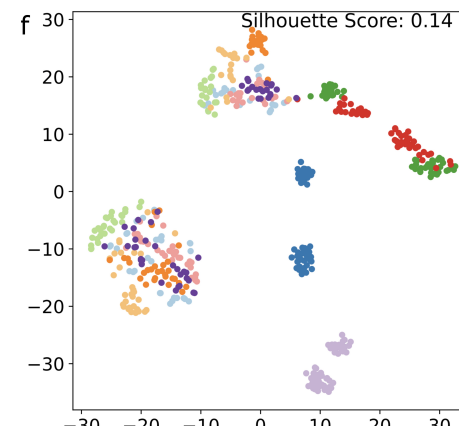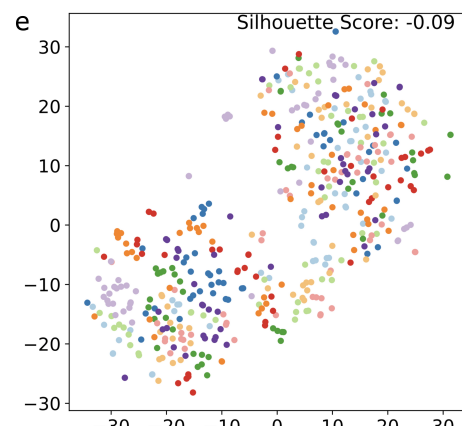**Figure 4: Feature clusters of single cell-slice image and multiple cell-slice images**. **(a)** The feature clusters of single cell-slice images using the MID model. Similar to CellProfile **(c)** and DeepProfiler **(e)**, the feature distribution of the 10 compounds appeared scattered and confounded, making it impossible to distinguish between different compounds. **(b)** The feature clusters of multiple cell-slice images using the MID model. The MID approach captures enough relevant information from multiple cell-slice images to be representative of the sample. Clustering using features of multiple cell-slice images showed that the same compounds were clustered together and clusters of different compounds were dispersed at long distances, indicating that these features have a high degree of discrimination. **(d)** The feature clusters of multiple cell-slice images using CellProfiler, which uses simple averaging to fuse information from multiple cell slices to represent the sample. However, this simple mechanical integration employed by CellProfiler was insufficient to obtain information related to the properties of the compound when compared with MID. **(f)** The feature clusters of multiple cell-slice images using DeepProfiler. Although the effect of DeepProfiler was better than that of CellProfiler, it still lagged behind the performance of MID. Furthermore, when using the features of multiple cell-slice images, the clusters of compounds benzethonium chloride and betrixaban were distributed very separately, and the individual cluster was more compact, indicating the advantage of using multiple cell slices. That is, the MID method can extract valid information from all the useful cell-slice images to represent the tested sample. Abbreviations used are DP for DeepProfiler and CP for CellProfiler.
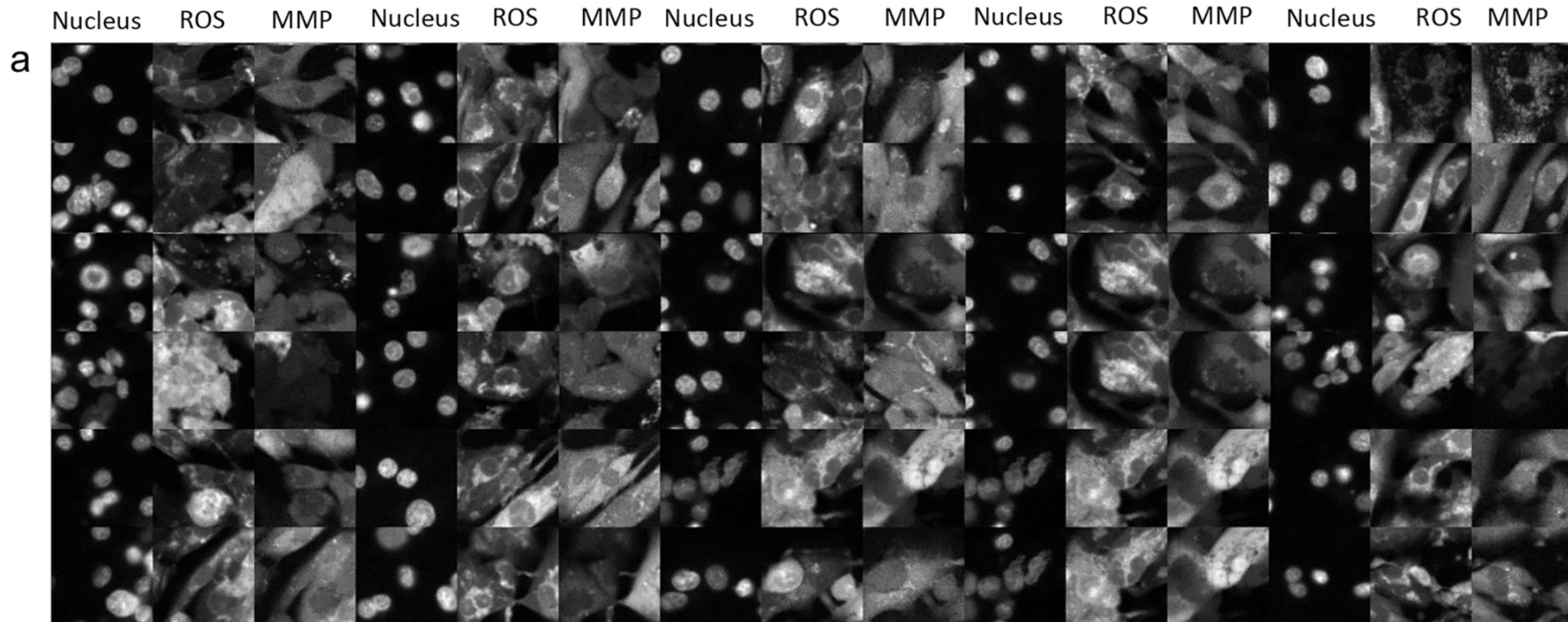
Figure 5: Analysis of cell phenotype similar compounds show that MID embeddings can represent the cell phenotypes. **(a)** cell-slice images of three compounds clustered together by MID embeddings in the hEGR classification task. For the three compounds (belzutifan, carvedilol and daclatasvir), each compound has two rows of images, where the first row is for relevant images and the second row is for irrelevant ones. The three compounds are clustered together in Fig. 4b, and we cannot differentiate any one of them from each other based on the images shown here, regardless of whether they are relevant or irrelevant, which illustrates the consistency of the cell phenotype images and the MID embeddings. ROS, mitochondrial reactive oxygen species. MMP, mitochondrial membrane potential. The intensity of MMP **(b)** and ROS **(c)** calculated from cell images separately perturbated by belzutifan, carvedilol, daclatasvir and betrixaban, of which the first three drugs clustered together and separated from the fourth drug in Fig. 4b. The MMP or ROS intensity of the first three at different concentrations shows a nearly consistent level while significantly differing from the fourth, which interprets the judgment of MID and the similarity of cell phenotype for the first three drugs. ROS, mitochondrial reactive oxygen species. MMP, mitochondrial membrane potential.
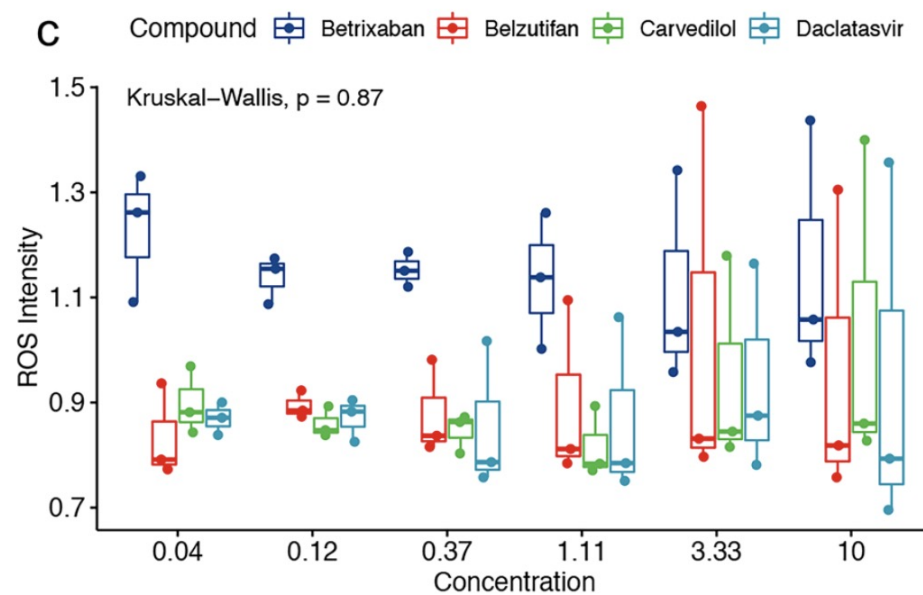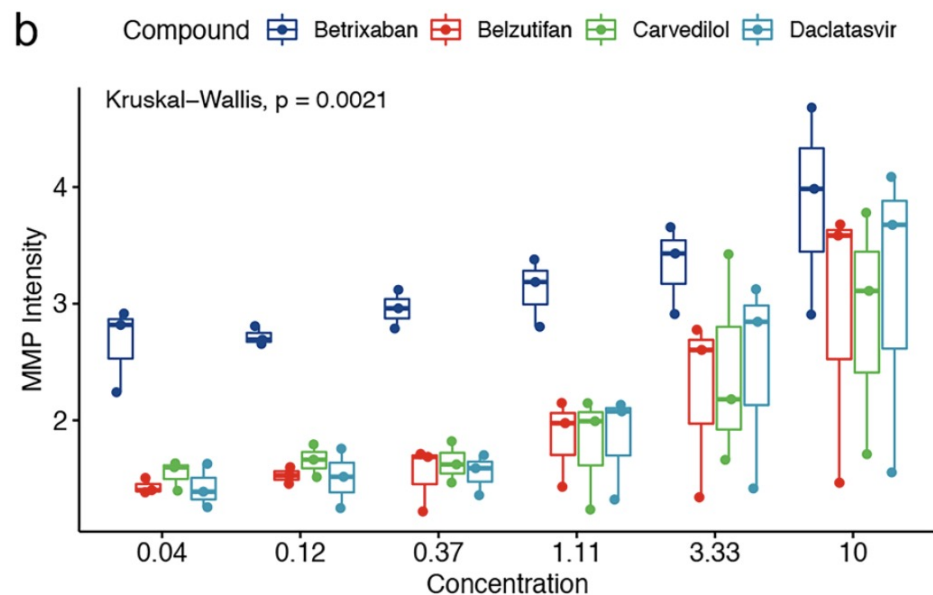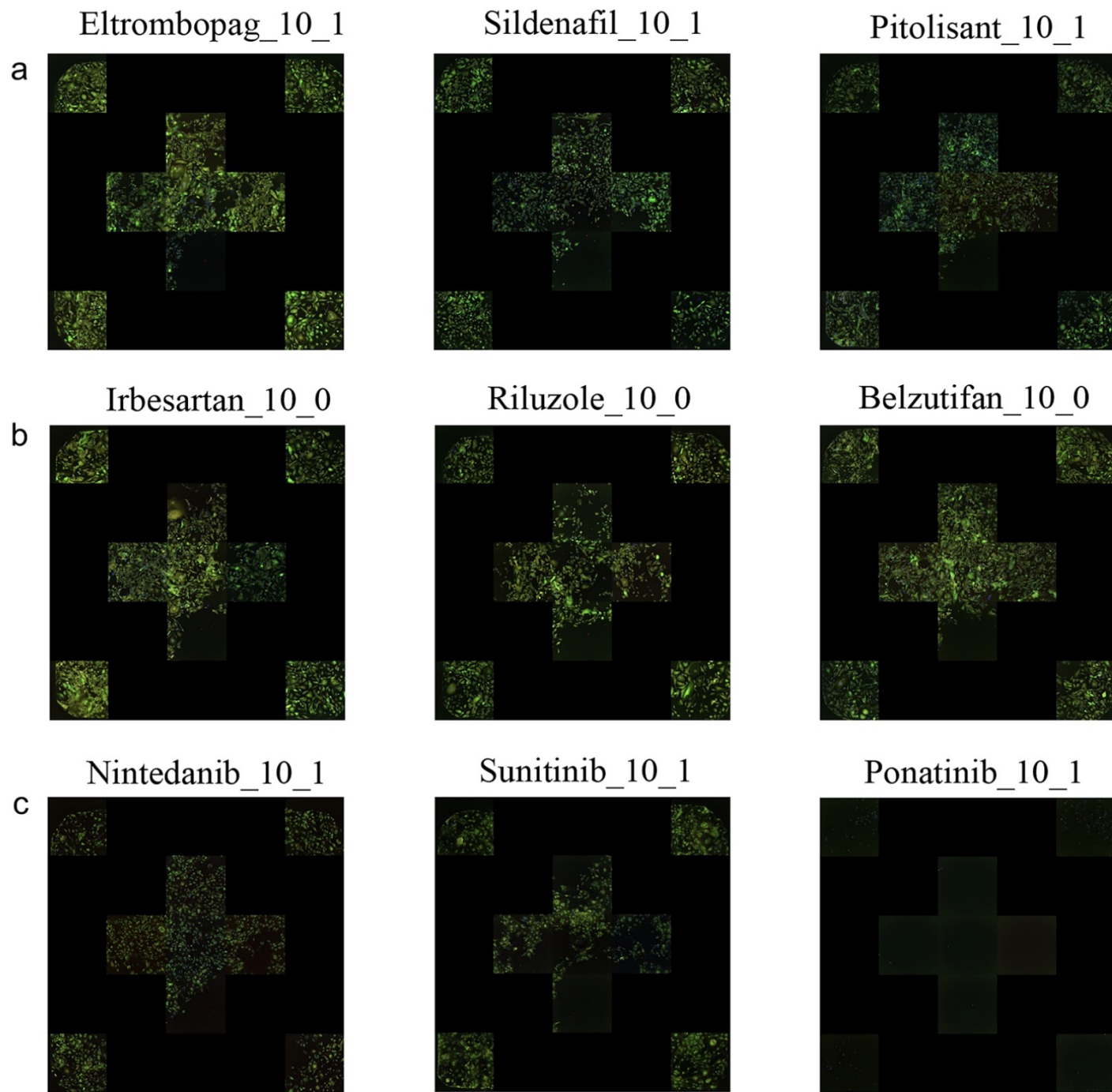
Figure 6: The results of cell painting staining for cells treated with compounds that either do or do not exhibit cardiotoxicity. (a) The staining results for three compounds that MID model incorrectly predicted as nontoxic. **(b)** Cell Paintings staining results of cells treated with three significantly nontoxic compounds, in which the cell structure is intact with uniformly distributed green fluorescence and nearly no dead cells. **(c)** Cell painting staining results of cells treated with three significantly toxic compounds, in which the cell structure is cracked with separately distributed green fluorescence and dead cells are clearly visible. Notably, the staining patterns for the incorrectly predicted compounds in (a) are more similar to those observed for nontoxic compounds.