Cell Chemical Biology

A novel machine learning and live cell imaging tool kit uncovers small molecules induced phospholipidosis

Huabin Hu[1,2,#], Amelie Tjaden[3,4,#], Stefan Knapp[3,4], Albert A. Antolin[1,5,*], Susanne Müller[3,4,6,*]

[1] Centre for Cancer Drug Discovery, Division of Cancer Therapeutics, The Institute of Cancer Research, London, UK

[2] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-751 24, Uppsala, Sweden

[3] Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, Frankfurt am Main, 60438, Germany

[4] Structural Genomics Consortium (SGC), Buchmann Institute for Life Sciences, Johann Wolfgang Goethe University, Frankfurt am Main, 60438, Germany

[5] ProCURE, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona, Catalonia, Spain

[6] Lead contact

[#] H.H. and A.T. contributed equally to this study

* Correspondence: aantolin@idibell.cat (A.A.A.), susanne.mueller-knapp@bmls.de (S.M.).

Tags: (1) Phospholipidosis, (2) live cell imaging, (3) machine learning, (4) chemical probes, (5) chemogenomics, (6) drug-induced phospholipidosis.

~~Summary~~SUMMARY

Drug-induced phospholipidosis (DIPL), characterized by excessive accumulation of phospholipids in lysosomes, can lead to clinical adverse effects. It may also alter phenotypic responses in functional studies using chemical probes. Therefore, robust methods are needed to predict and quantify phospholipidosis (PL) in early in drug discovery and in chemical probe characterization. Here, we present a versatile high-content live-cell imaging approach, which was used to evaluate a chemogenomic and a lysosomal modulation library. We trained and evaluated several machine learning models using the most comprehensive set of publicly available compounds and interpreted the best model using SHapley Additive exPlanations (SHAP). Analysis of high-quality chemical probes extracted from the Chemical Probes Portal using our algorithm, revealed that closely related molecules, such as chemical probes and their matched negative controls, can differ in their ability to induce PL, highlighting the importance of identifying PL for robust target validation in chemical biology.

~~Introduction~~INTRODUCTION

Phospholipidosis (PL) is a pathological condition characterized by an excessive accumulation of phospholipids in lysosomes and a distinct morphological appearance of lamellar bodies.[1] PL is frequently induced by the exposure of cells to exogenous substances, including various pharmacological agents.[2] This drug-induced phospholipidosis (DIPL) can have significant implications for drug safety and side effect profiles, including toxicity in the liver and lungs.[3,4] Currently, over fifty clinically studied drugs that are used for treating a broad spectrum of disorders such as neoplastic, psychiatric, and infectious diseases, are known to induce phospholipidosis (**Figure 1A**).[2,4]

Importantly, DIPL is also a critical risk factor in experiments that utilize chemical probes in cell culture or *in vivo* as the accumulation of phospholipids can have confounding effects in many biological processes.[5] In particular, when using cell-based phenotypic screens, the mechanism of action of the used small molecules needs to be critically assessed.[6] It has been recently reported that several drugs identified in cellular repurposing screens showed apparent antiviral activity against SARS-CoV-2 via DIPL mechanisms and not by acting on their primary targets.[7] It is therefore of utmost importance to comprehensively characterize libraries used in functional cellular assays not only regarding their general toxicity,[8] but also for subtler cellular alterations, such as phospholipidosis induction to facilitate more reliable target validation in chemical biology.

The exact molecular mechanism underlying the occurrence of PL has not yet been completely elucidated.[2] There are divergent views on whether it is a target-specific process, or whether PL is a predominantly due to non-specific effects caused by the chemical properties of the small

molecules. For example, the accumulation of small molecules in the lysosomal membrane can lead to the release of lysosomal hydrolases and sphingolipid activator proteins from the surfaces of intralysosomal luminal vesicles.[2] Several targets have been proposed to mediate DIPL, such as lysosome-specific phospholipase A2 (PLA2G15),[9] or the sphingomyelinase (Asm)-ceramide system, but currently no single target can explain all the observed DIPL cases.[10,11]

DIPL has been associated with cationic amphiphilic drugs (CADs) and the underlying non-specific mechanisms that contribute to phospholipidosis are widely believed to be associated with the physicochemical properties of CADs. Specifically, CADs feature hydrophobic and hydrophilic domains and carry a positive charge at physiological pH which may contribute to their accumulation within lysosomes.[2] This physicochemical property could be attributed to the ability of small molecules to diffuse passively across lipid bilayers in their unionized form, and subsequently become ionized and trapped within lysosomes due to the lower pH values in this cellular compartment.[12,13] For these cases, it is also evident that the binding of CADs to phospholipids may lead to the formation of drug-lipid complexes that cannot be digested.[14,15] Given the impact of PL for drug development, several approaches have been developed to identify molecules that induce phospholipidosis in order to deprioritize them during the early stages of drug discovery.[16] Electron microscopy (EM) has been recognized as the most reliable experimental technique detecting PL.[17] Despite its effectiveness, the widespread utilization of EM has been limited due to the low throughput, high cost, and prolonged timescales of this analysis. Consequently, high-content imaging techniques have gained popularity as a viable alternative. Among the most commonly employed detection techniques are fluorescently-

labelled phospholipids such as (1-acyl-2-(12[(7-nitro-2-1-,3-benzoxadiazol-4-yl)amino]dodecanoyl) (NBD) NBD-PE staining[18] or Nile red lipid stain.[19] In addition, the fluorescence microscopy-based lysosomal perturbation assay that employs a red fluorescent probe, developed by Coleman et al. in 2010,[20] has also gained widespread adoption as a reliable detection method. However, the implementation of an efficient assay using live-cell imaging methods that could also assess kinetic aspects, is still lacking.

Experimental phospholipidosis profiling can be resource intensive, and thus several *in silico* approaches have been developed to reduce the experimental burden.[16,21] Ploemen et al. introduced the first physicochemical property model that utilized cLogP and pKa as chemical properties to identify cationic amphiphilic compounds inducing PL.[22] This method has been widely adopted and it has been later refined incorporating other properties such as the net charge[23] or volume of distribution (V_d).[24] In general, physicochemical models are useful when applied to small and homogeneous datasets but their predictive capacity is considerably diminished with assessing larger datasets.[25] To further enhance the predictability of phospholipidosis, Przybylak et al. introduced the structural alert model comprising 39 structural fragments that closely capture chemical features associated with phospholipidosis, such as peripheral amine groups and hydrophobic moieties. [26,27] Next, scientists from the U.S.A. National Institute of Health (NIH) developed a machine learning (ML) approach based on Support Vector Machine (SVM) model, trained using molecular descriptors of atom types and whole molecule properties of three large compound collections,[28] achieving a high accuracy of 0.90 area under the receiver operator characteristic curve (AUC). NIH scientists also showed that many phospholipidosis inducers and hERG blocking compounds share a similar structural

feature composed by two aromatic rings and an amino group forming a three-center toxicophore.[29,30]

More recently, researchers at AstraZeneca utilized a significant amount of proprietary data (1,537 compounds with 441 inducers) to create a consensus computational model that combines ML and physicochemical property model (Ploemen's criterion).[31] The established model demonstrated strong performance in an external set of 183 compounds (identified 75% of all PL inducers). However, most machine learning approaches studies do not yet benefit from recently developed explainable approaches that allow for model interpretability to better understand their strengths and limitations.[32] Thus, it is important to enhance existing computational approaches and gain deeper insights into the molecular mechanisms underlying phospholipidosis.

In this study, we developed and validated a high-content live-cell imaging approach and screened two diverse libraries comprising ~ 300 compounds. We also trained a set of ML models using the most comprehensive set of publicly available compounds to date and assessed these models using the two internal experimentally screened sets. We interpreted the best model using the SHAP approach and used it to predict whether high-quality selective probes recommended by the Chemical Probes Portal (https://www.chemicalprobes.org/)[33] may induce PL. Experimental validation confirmed many predictions and allowed us to identify chemical probes that induce PL at recommended concentrations. Finally, our analysis of structurally highly similar compounds, such as chemical probes and their inactive analogs used as negative controls, demonstrated that small changes in compound structure may lead to PL. Our findings have important implications for the practice of chemical biology evaluation and

characterization and our understanding of the complex nature of phospholipidosis induction.

~~Results~~ RESULTS

High-content assay establishment and screening two in-house datasets

Several high-content assays have been proposed in the literature to detect phospholipidosis in a cellular context.[1,18,19] However, these assays are endpoint assays and they have not been adapted for imaging in live cells. To this end, we used the LipidTox red reagent, an already implemented high-content imaging method first described by Nioi et al.[34] and later optimized by Shahane et al.[1] In this method, phospholipids are conjugated with a fluorescent dye to increase their visibility when they accumulate during phospholipidosis. In order to adapt this assay for live-cell monitoring, we incorporated this detection method into our previously published live-cell multiplex assay and further optimized this procedure (**Figure 1B**) (see **STAR methods** for details).[8] The categorization of PL dots encompassed their localization within both the cell body and lysosomes, with the latter being identified using LysoTracker as an additional dye (**Figure 1C**). To rule out possible toxicity of the dyes used in live-cells over long periods of time (up to 72h), and to determine the optimal concentration, we performed a viability assessment using alamarblue after adding different dye concentrations. This experiment revealed no significant effects on cell survival at the concentrations used (**Figure S1A**). To validate the assay, we tested six drugs known to induce phospholipidosis (**Figure 1A**)[2,7] including antidepressant drug sertraline, a selective serotonin reuptake inhibitor, antipsychotic drug haloperidol targeting dopamine D_2 receptor, and the antiestrogenic agent tamoxifen using osteosarcoma (U2OS) cells. Drugs were tested at three different

concentrations over 72h and the normalized PL was assessed based on the total intensity of the dye molecules compared to cells treated with DMSO. After 48h, all reference compounds significantly induced PL, which was detected already at the lowest concentration used (500 nM) (**Figure 1D**). For all control compounds, induction of PL correlated with compound concentration and the time of exposure (**Figure 1E**). To enable subsequent machine learning analysis (a binary classification task mirroring the public compound annotation used for model training) of in-house developed data sets, it was necessary to establish a rigorous threshold for classifying the compounds as either "inducer" or "non-inducer" of PL. Utilizing the initial reference dataset, we established a threshold in U2OS cells by defining normalized PL intensity larger than 2 based on the intensity measured for chloroquine (**Figure 1D**). Nonetheless, in cases where compounds were close to the threshold value, we performed a careful visual assessment to determine whether the compound should be categorized as an "inducer" or "non-inducer". Most PL data described in the literature have been recorded in HepG2 cells. Since our assay was performed in U2OS osteosarcoma cells, we tested context dependency of PL in different cell lines by testing the reference compounds also in HEK293T and HepG2 cells after 24h (**Table S1**). In all cell lines, PL was detected for all reference compounds. Given the distinct morphological characteristics of each cell line, a direct comparison of the total intensities was not feasible (**Figure S1B**). We therefore recommend considering each cell line separately and setting the threshold for significant PL induction accordingly. This will be particularly important in drug discovery projects where PL may be evaluated on patient-derived cell models to assess the corresponding intensity threshold appropriately.

After establishing the high content assay and defining the threshold to distinguish between “inducer” and “non-inducer”, we tested two in-house compound sets in order to identify PL inducers. The first compound set included a diverse set of chemogenomic compounds targeting proteins of different families, including GPCRs, kinases, epigenetic modulators, and others (**Table S2A**). In contrast to chemical probes which are highly selective and potent modulators of their targets, frequently accompanied by a matched structurally related negative controls, chemogenomic compounds are not exclusively selective and they often target more than one protein. However, the selectivity of chemogenomic compounds has been comprehensively assessed and the hope is that phenotypic changes in cell culture can be assigned to a target based on hits from chemogenomic compound set with similar on-target activity but diverse off-target profiles. The chemogenomic compound set consisted of 213 compounds, out of which 53 (25%) were identified as "inducers" with a normalized PL intensity larger than 2 (**Figure 2A**; **Table S2B**). To assess compounds, that are known to play a pivotal role in lysosomal modulation, a second set consisting of 99 compounds was used (**Table S3A**). In order to capture cellular phenotypes over long incubation times, this set was subjected to a longer incubation time of 72 hours. In total, 29 compounds (29%) of this second set were defined as “inducers” (**Figure 2B**, **Table S3B**). Among the PL inducing compounds was crizotinib, an approved receptor tyrosine kinase inhibitor targeting ALK and cMET, as well as a number of other protein kinases, which has been reported to boost lysosomal count.[35] In addition, we detected PL upon exposure of cells with daurisoline, that acts as an autophagy inhibitor suppressing lysosomal acidification.[36] Also, the histone deacetylase (HDAC) inhibitor mocetinostat, which previously has been shown to trigger accumulation of lysosomes in cells induced PL.[37] As an

external control, four reference compounds (amiodarone, chloroquine, chlorpromazine and sertraline) were added to both sets and were correctly identified as “inducers”, demonstrating the reproducibility of our live-cell assay. There were 22 compounds that were present in both sets, which acted as internal controls (**Figure 2C**). Interestingly, although for most compounds phospholipidosis was observed after 24h in both sets, some compounds reached the threshold only at 48h, indicating the importance of assessing compounds at longer time points.

PL has been frequently associated with off-target phenotypic effects and it might contribute to toxicity of diverse approved drugs.[5] Nevertheless, in our 2-D cell culture models, we did not observe correlation between PL and cytotoxicity and only few compounds, such as kinase inhibitor adavosertib, showed experimentally PL induction at late time points and affected cell viability (**Figure S2A and 2B**). Other authors,[4] have suggested a connection between autophagy and PL. However, following assessment of the lysosomal modulation set in an LC3 HiBiT reporter assay, we did not observe a direct correlation between phospholipidosis and autophagy (**Table S4**). Therefore, additional evaluations may be required to further elucidate the relationship between these two phenomena.

Curation of a comprehensive phospholipidosis data set from the literature

First, a thorough analysis of PubMed was carried out, resulting in the identification of nine articles that reported large compound datasets assessed for PL. From these articles, we identified reported compounds, downloaded its chemical structure with the aid of the PubChemPy package (https://github.com/mcs07/PubChemPy) and applied a structural standardization protocol (see **STAR methods** for details). This analysis resulted in a set of

2,422 unique compounds that we termed the curated data set. Among these compounds, 114 (4.7%) had inconsistent phospholipidosis annotations in the literature, a fact that has been observed previously.[28] For example, a number compounds have been labeled as inducers as well as non-inducers in different publications. These inconsistencies illustrate the challenges around comprehensively identifying phospholipidosis, and building computational models with the limited data quality, quantity, and availability. To attain data with a high level of confidence, we eliminated compounds that possessed inconsistent annotations. This led to the creation of an initial set consisting of 2,308 literature compounds, of which 1,683 were non-inducers and 625 were inducers.

Next, we assessed the overlap between the curated set and our two in-house compound sets. This analysis identified 38 compounds that were present in both sets. Of these, only five compounds (~ 13%) had inconsistent annotations due to precipitation or high toxicity and were excluded from the analysis. Overall, a total of 213 compounds for the chemogenomic compound set (**Table S2**), 99 compounds for the lysosomal modulation set (**Table S3**), and 2,303 compounds for the curated literature set (**Table S5**), were obtained. The Venn diagram in **Figure 2C** presents an overview of the distribution of compounds among these three sets, while **Figure 2D** illustrates the distribution of phospholipidosis compositions in each set. Specifically, the curated set obtained from the literature contained 622 inducers and 1,681 non-inducers; the chemogenomic compound set comprised 53 inducers and 160 non-inducers; and the lysosomal modulation set consisted of 29 inducers and 70 non-inducers. In general, the percentage of inducers in each set was quite similar with 24.9% to 29.3% PL inducers.

A simple physicochemical model is limited to differentiate inducers from non-inducers

Using the curated data set obtained from the literature, we initially evaluated the effectiveness of a simple physicochemical property model (Ploemen's criteria using pKa and cLogP values)[22] to differentiate phospholipidosis inducers from non-inducers (see **STAR methods** for details). An overview of the analytical workflow is presented in **Figure 3**. Using this simple model, an accuracy value of 0.74 was obtained for the curated set of 2,303 compounds meaning that 74% of PL inducers and non-induces were correctly predicted. However, we only obtained a recall value of 0.35 (**Figure 3**, left), indicating that this simple model was only able to identify 35% of all PL inducers and illustrating the model's limited ability to comprehensively predict phospholipidosis (i.e., inducers were frequently misclassified as non-inducers). The suboptimal performance of the model prompted us to explore whether ML algorithms could be more effective.

A random forest algorithm coupled with concatenated features outperforms other approaches

Firstly, we evaluated four distinct ML algorithms, namely *k*-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and eXtreme Gradient Boosting (XGBoost), using MACCS and ECFP4 fingerprints as molecular representations. In order to better recapitulate the physicochemical properties that are known to be important to predict PL, the molecular representations were concatenated with two additional properties (pKa and cLogP) and termed 'concatenated features' (see **STAR methods** for details). At first, the curated data was split into two distinct datasets: a training set that served the purpose of training the model and fine-tuning

its hyperparameters, and a testing set that was utilized to evaluate the model's performance internally. Moreover, the two physical phospholipidosis sets evaluated in-house were used as external validation to assess the model's predictability on real-world datasets, as depicted in the workflow illustrated in **Figure 3**. The results presented in **Figure 4A-C** and **Figure S3** indicate the performance of the ML models across 10 independent trials (i.e., 10 times independently splitting the curated data to train and test the models). Our analysis revealed that most of the models had robust predictive performance on the test set, with the RF model using MACCS fingerprint exhibiting the highest average recall value of 0.73. Accuracy values around 0.80 and F1 values around 0.6 were similar across all ML methods explored (**Figure 4A**). We next investigated the generalization ability of these models by assessing their performance on the chemogenomic compound set and the lysosomal modulation set (external validation datasets). Notably, the RF algorithm coupled with MACCS fingerprint also outperformed the SVM, KNN and XGBoost models in this setting, as indicated by the highest average recall and F1 scores of 0.57 and 0.47, respectively, on the chemogenomic compound set (**Figure 4B**). In contrast to the RF algorithm, SVM, KNN and XGBoost exhibited average recall values of less than 0.4 with any of the two fingerprints tested. Similarly, when evaluating the lysosomal modulation set (**Figure 4C**), the random forest algorithm using MACCS fingerprint exhibited superior performance over the other three methods, achieving an average recall value of 0.84.

Following our assessment of the three sets, we decided to further investigate the performance of the RF algorithm in combination with different molecular representations, including MOE 2D descriptors (206 physicochemical property features) and PubChem fingerprint (881 chemical substructure patterns). Our analysis demonstrated that there was little variation in the

predictive performance of the RF algorithm when using different molecular representations in both the test set (**Figure S3A**) and our two in-house sets (**Figure S3B** and **S3C**).

Based on these results and considering also computational efficiency, we decided to focus on RF in combination with MACCS fingerprint for further analysis. To further assess this model, we plotted the ROC curves of the three investigated sets based on this best ML model (**Figure 4D**). The AUC values of the test set and lysosomal modulation set were 0.87 and 0.90, respectively – a high accuracy comparable to previously reported methods.[28] These values suggested that the model has a high predictive capability for these two sets. However, the models' predictability for the diverse chemogenomic compound set was lower with an AUC value of approximately 0.7, indicating that it is less able to make accurate predictions for this set.

The statistical results of the RF model with MACCS on the test set, chemogenomic compound, and lysosomal modulation sets based on the best trial are presented in **Figure 4E**. For comparison, the same sets were predicted using Ploemen's simple physicochemical property model. As shown in **Figure 4E**, both the physicochemical and RF models have similar accuracy values (0.71-0.81) whilst the F1 and, importantly, the recall, is much better with the RF (0.64-0.86) than with the physicochemical model (0.33-0.48). These results demonstrate that the newly developed ML model can identify many more compounds that induce PL that would be misclassified if only physicochemical properties were considered.

To benchmark our ML model further, we applied previously published methods that reported the exact code or were straightforward to use to predict new datasets (e.g. are available via webserver) and we identified FP-ADMET,[38] a repository of molecular fingerprint-based RF

built comprehensively exploring molecular descriptors for several ADMET properties, including PL prediction (see **STAR methods** for details). FP-ADMET built a RF model to predict PL using 1,719 compounds and reported an AUC of 0.88. When we used FP-ADMET to predict our chemogenomic and lysosomal modulation sets, its recall was 0.44 and 0.71, respectively **(Figure 4E)**. It is worth highlighting that although FP-ADMET was able to improve the recall of the simple Ploemen's physicochemical property model (0.33, 0.48), our algorithm displayed a significantly better recall (0.64, 0.86) on these two external validation sets – illustrating how our algorithm compares favorably with available methods and can identify more DIPL compounds from two different compound sets.

As a means of an additional validation, we identified a set of 462 new compounds (117 PL inducers and 345 non-inducers, **Table S5**) from public sources,[1,29] which we used as an additional validation set. Using our RF model, we were able to correctly predict 85 of the compounds as PL inducers, which corresponds to a recall value of ~ 0.73 – in line with our previous recall values of 0.64 – 0.86 in the external validation sets (**Figure 4**). The remaining performance metrics were also in line with the results on the two previous external validation sets or slightly higher (**Figure S4**), hence further highlight the strong predictability and robustness of our algorithm when validated prospectively and when we compared it to other published approaches such as FP-ADMET.

Two different real-world prediction scenarios explain the disparity of prediction performances

Intrigued by the marked discrepancy in the predictive ability of the ML method developed

when applied to the chemogenomic compound (recall = 0.64) or lysosomal modulation (recall = 0.86) sets (**Figure 4E**), we decided to explore the underlying factors.

First, we evaluated the chemical similarity between the training set and the three investigated sets (**Figure S5**). As expected, the training set (used for model construction) had the highest chemical similarity to the test set with a median maximum Tanimoto coefficient (mmTc) of 0.51, followed by lysosomal modulation set (mmTC = 0.33). The chemogenomic set was the most dissimilar (mmTc = 0.29), which may explain why all methods perform worse with this external dataset (**Figure 4E**). However, it is worth highlighting that our model still yielded the best performance across all models tested when predicting the chemogenomic set, illustrating its effectiveness in a challenging, real-world scenario.

In addition to the chemical similarity, we also analyzed the physicochemical properties on each set of compounds to gain more insights in the differences between them. As shown in **Figure S6**, these two sets represent different areas of the property space which translate into different levels of difficulty for real-world computational prediction. In the lysosomal modulation set, most PL inducers fall within the cationic amphiphilic area of the chemical space (cLogP $\geq$ 3 and pKa $\geq$7.4). Therefore, they are easy to predict, which explains the higher recall of the ML method (0.86) and that even the simple physicochemical model achieves a significant recall (0.48). In contrast, in the chemogenomic set, there are far more PL inducing compounds (orange dots in **Figure S6**) that fall outside the cationic amphiphilic area of the property space, and thus are more difficult to predict. Therefore, different external validation sets can be easier or harder to predict depending on the area of the chemical space they cover. Despite these limitations, ML still outperforms the simple model by a substantial margin (**Figure 4E**).

Additionally, we also explored whether incorporating pKa and cLogP properties into the MACCS fingerprint was truly improving the prediction performance. This was clearly demonstrated by the increase in recall values from 0.53 (without pKa and cLogP properties) to 0.64 (with these two properties) on the chemogenomic compound set, as presented in **Table S6**. Therefore, we made the decision to continue utilizing the RF model with MACCS and concatenated features for further analysis.

Model interpretation based on SHAP-based feature analysis

Based on the selected computational model, it was important to elucidate the underlying mechanisms by which the model generates its predictions to understand its strengths and limitations. Therefore, we applied the SHAP approach[39,40] to analyze the contributions of individual features (see **STAR methods** for details). The top 15 features that significantly affected the prediction output, along with their SHAP value distributions, were analyzed and presented in **Figure 5A**. These top features were also mapped back to an exemplary chemical structure to facilitate interpretation, as shown in **Figure 5B**. Due to the relatively low SHAP values observed for all feature bits (< 0.1, **Figure 5A**), it was inferred that there were no dominant features that strongly influenced the predictions generated by the ML model. Thus, using a small number of features may not yield accurate predictions. All top features, except for the MACCS feature "O=A >1", had a positive effect on predicting PL inducers. Thus, the existence of these features additively increases the likelihood of molecules being categorized as inducers. The most influential features were the molecular physicochemical properties of pKa and cLogP, as well as the MACCS fingerprint "CH2QCH2". The presence of the pKa

feature (pKa ≥ 7.4) and cLogP feature (cLogP ≥ 3.0) indicated that these compounds can be protonated to carry a positive charge in the acidic endolysosomal compartment and is more amphiphilic.[41] This observation was in line with previous reports that suggesting that compounds that induce phospholipidosis tend to be cationic amphiphilic in nature.[2] Nevertheless, a significant number of compounds inducing PL fall outside the cationic amphiphilic chemical space (**Figure S6**). The "CH2QCH2" substructure often accounts for the presence of a hydrophilic amine group in a molecule, which is a structural representation of the pKa property. Additionally, other MACCS structural features, such as aromatic or six-membered ring systems, indirectly accounted for the lipophilic~~obic~~ nature of molecules. In summary, the ML model outperformed the simple physicochemical model by giving small weights to multiple features that included pKa and cLogP (indicating the characteristics of whole molecule) as some of the top-ranking features, but the model also ranked highly chemical features representing ionizable and aromatic/aliphatic moieties. Therefore, the ML method “learned” that even if pKa is not ≥ 7.4 or cLogP ≥ 3, the presence of other chemical moieties contributed to correct predictions of PL.

Next, we attempted to deeply explore the prediction of a few representatives from our in-house compound sets shown in **Figure 5C**. Afatinib and BI-2536 are kinase inhibitors that target EGFR/ErbB and PLK1, respectively. While Afatinib induces autophagy and thereby increases lysosomal activity,[42] BI-2536 suppresses lysosomal fusion.[43] Nonetheless, both compounds were correctly categorized as inducers by the ML model, with predicted probability values of 0.87 and 0.80, respectively (**Figure 5C)**. The high probability of an inducer prediction can be attributed to the prevalence of several positive features representing by positive SHAP values

which are highlighted in lime on their corresponding chemical structures. Our analysis also included two misclassified chemogenomic compounds, namely (+)-JQ1 and (S)-ZINC-3573. (+)-JQ1 is a BET bromodomain inhibitor[44] that suppresses cell proliferation and promotes lysosomal modulation. The model wrongly predicts that (+)-JQ1 is not a phospholipidosis inducer, which is primarily attributed to the absence of significant features, such as a hydrophilic amine group with a pKa value below 7.4 (**Figure 5C**). We also analyzed the features of the 18 false negative (FN) cases (**Figure 4E** and **Table S7**) in the chemogenomic compound and lysosomal modulation sets. The majority of them (89%) did not contain the "pKa ≥ 7.4" feature, indicating available amines were not predicted to be sufficiently hydrophilic. These results could prompt the development of better pKa prediction approaches or challenge the necessity of compounds to be protonated in order to induce PL. On the other hand, a certain number of false positive (FP) cases were also identified, such as (S)-ZINC-3573. (S)-ZINC-3573 is the inactive enantiomer of chemical probe (R)-ZINC-3573, a selective agonist of MRGPRX2.[45] (S)-ZINC-3573 was misclassified as inducer of phospholipidosis because it presents several features such as aromatic rings, "CH2QCH2", and pKa ≥ 7.4. These two examples illustrated that the presence or absence of specific moieties and physicochemical properties was not sufficient to correctly classify all compounds. Taken together, the SHAP values assigned to each feature contribute to the transparency of the ML prediction, significantly aiding our comprehension of how the model arrived at the specific prediction.

Structurally similar compounds may differ in inducing phospholipidosis

Intrigued by the existence of compounds that produced a phospholipidosis phenotype that was

contradictory with the general pKa, cLogP and features associated with it, we decided to investigate the structural relationship between compounds with varying phospholipidosis annotations through analog analysis. In this study, we used a substructure-based approach known as the matched molecular pair (MMP)[46] to identify structural analogs among the investigated three compound sets. MMPs can be obtained via the application of a compound fragmentation algorithm (see **STAR methods** for details).[47] Through a systematic fragmentation of compounds, a total of 1,368 MMPs were generated, comprising 1,149 unique compounds (1,078 of which were curated and 71 were chemogenomic/lysosomal modulation compounds). This corresponds to approximately 45% of the entire collection of compounds analyzed indicating high structural diversity in the three sets. To visualize the generated MMPs, they were organized into a network (**Figure 6,** upper panel). In this network, the nodes correspond to compounds, and the edges represent the pairwise MMP relationship. Coordinated MMPs formed by subsets of compounds gave rise to the formation of distinct clusters within the network. Upon analyzing the generated MMPs, we were able to identify a total of 343 clusters. In general, cluster sizes are small, with an average of three to four compounds per cluster. Out of the identified clusters, a significant portion of 213 (62.1%) consisted solely of non-inducers, while a smaller subset of 43 (12.5%) was composed exclusively of inducers. This suggests that compounds with similar structural characteristics tend to exhibit similar phospholipidosis profiles. However, we also detected 87 clusters (25.4%) that contained compounds exhibiting both phospholipidosis inducer and non-inducer characteristics indicating that structurally similar compounds may significantly differ in their phospholipidosis induction properties. The analysis of analogs with varying phospholipidosis

annotations yielded a total of 236 MMPs (17.2%). Among the MMPs, we have identified a total of 214 distinct chemical transformations (**Table S8**), representing variations in their chemical composition. The large number of transformations suggests that there is not a single or a few structural modifications that are able to unambiguously explain PL induction. The most frequent transformation is an H ↔ CH3 transformation (7 MMPs), followed by an H ↔ OH transformation (5 MMPs) and an H ↔ F transformation (4 MMPs). All of them are single-atom modification indicating that minor structural modifications with conservative impact on physicochemical properties are enough to alter the PL behaviors of small molecules. This observation underscores the intricate nature of the PL mechanisms at play. These above results indicated that protonation state, pKa, cLogP and substructural features were insufficient to correctly identify all PL inducing compounds, which has important implications for the development of more robust computational methods and highlights the complexity of PL.

Chemical probes and their corresponding negative controls may induce divergent phospholipidosis phenotypes

It is noteworthy that out of the 236 MMPs with different phospholipidosis annotations, 13 MMPs were related to compounds that we screened in-house. Intriguingly, four MMPs were formed by a chemical probe and the corresponding structurally related negative control, while the fifth MMP was formed by two chemical probes (GSK6853 and PFI-4). Three of them were highlighted in the network (cluster I, II, and III) and were summarized in **Figure 6**.

In three cases, chemical probes were experimentally found to induce phospholipidosis, while their corresponding negative controls were non-inducers. Despite variations in activity against

their intended primary targets (controls are inactive or much less active), pairs consisting of chemical probes and negative controls frequently displayed similar physicochemical properties, including comparable pKa and cLogP values. Therefore, very small chemical modifications can transform a compound from a PL inducer to a non-inducer without significantly affecting physicochemical properties. For example, the chemical probe PPTN functions as an antagonist of the GPCR family member P2RY14 (IC_{50} = 0.5 nM).[48] The mere substitution of a hydrogen atom with a methyl group nearly abrogates affinity for P2RY14 (IC_{50} = 4,381 nM, > 8,000-fold difference) whilst pKa and cLogP remain very similar. MS023 (targeting PRMT1, PRMT3, PRMT4, PRMT6, and PRMT8)[49] and TP-472 (targeting BRD7 and BRD9)[50] represent two additional examples of the same behavior.

Another interesting chemical probe/negative control pair was TP-064/TP-064N (**Figure S7**) targeting PRMT4.[51] In this case, despite replacing the methoxy group with a secondary amine, both compounds remained inducers – demonstrating that these small structural changes did not necessarily induce a change in the propensity of small molecules inducing PL. Finally, in the case of SGC3027 (**Figure S7**), which targets PRMT7,[52] we observe the opposite behavior where the probe was experimentally confirmed to be a non-inducer while the negative control was an inducer of PL.

Overall, the identification of this series of chemical probes/negative control pairs revealing differences in phospholipidosis induction, demonstrated that minor chemical modifications with very limited impact on pKa/cLogP values and overall compound structure can transform a compound from a PL inducer to a non-inducer. These small alterations in compound structure poses challenges on computational predictions explaining why overall compound properties

(pKa, cLogP,) were insufficient predicting PL in most cases. These results also caution on the use of negative controls without further validation or characterization since the observed phenotype differences between probes and controls can be mediated via PL and not the known drug target.

Prediction and experimental confirmation of chemical probes inducing phospholipidosis challenge their use at recommended high concentrations

Finally, we decided to apply our ML model to identify chemical probes that might be inadvertently producing PL at high concentrations and thus represent a risk when they are used for target validation. To identify high-quality probes, we used the Chemical Probes Portal (https://www.chemicalprobes.org/),[33] predicted all of them using our ML algorithm. Of these high-quality probes, 160 probes were predicted to be PL inducers. We then prioritized 31 probes with predicted probability value greater than or equal to 0.7 to induce PL for experimental validation (see **STAR methods** for details). From these, we correctly predicted 21 chemical probes as PL inducers, resulting in a precision value of around 0.7. **Table S9** provides a comprehensive list of the 31 probes that were tested and experimental raw data is provided in **Table S10**. **Figure 7A** summarizes the targets of these 21 chemical probes. In total, 31 targets covering epigenetic and kinase groups were identified illustrating that there is no prominent primary target that is shared or increases the likelihood of compounds inducing PL. **Figure 7B** depicts 10 of these chemical probes that we predicted and experimentally confirmed to induce phospholipidosis and are recommended for utilization at high concentrations (up to 10 µM) by the Chemical Probes Portal. The lower panel highlights the time and concentration-dependent

progression of phospholipidosis induction, revealing an augmented induction at higher doses and prolonged exposure periods. These observations underscored the necessity of carefully accounting for such dose and time-dependent effects when employing these probes in experimental settings. Alarmingly, A-770041 and GSK484 induce phospholipidosis at their recommended 10 µM dose. In addition, A-770041, GSK343, adavosertib, and NVP-2 induced phospholipidosis at even lower concentrations (1 µM) (**Figure 7C**) that also fall within the Portal recommended concentrations. Therefore, when utilizing these probes at the suggested concentrations in functional cell assays, a contribution of phospholipidosis to the observed phenotype should be taken into consideration. We have contacted the Portal with a request to annotate the compounds accordingly.

~~Discussion~~DISCUSSION

Given the importance of improving small molecule induced phospholipidosis for de-risking drug discovery and improving the quality of chemical probes,[6,7] we have developed a live-cell, high-content imaging assay allowing time dependent studies and developed robust prediction models.

Time dependent consequences of PL on cellular signaling are poorly understood. Importantly, our experimental results demonstrated the importance of studying time dependency in live cells, to ensure the correct characterization of compounds of interest. This time dependency may also explain some inconsistencies regarding the effect of compounds to induce DIPL found in the literature, although other factors and experimental setups such as cell line of choice could also have contributed. Overall, the imaging assay that we developed has the advantage of tracking

PL development across time without any dye toxicity and allowing for simultaneously assessing the influence of a compound on cell viability thus reducing costs and resources. Generating large-scale, robust datasets with experimental methods that track DIPL across time might hold the key to improve ML model performance further.

Our computational methods confirmed previous observations that ML outperforms methods relying solely on physicochemical properties.[28] RF was the best performing algorithm in our assessment using public data as well in our two in-house data sets. The AUC value of our model on the test set was similar to previously published methods.[28] However, when benchmarked with two external datasets, our approach outperformed FP-ADMET[38] a recently-published and easy-to-implement RF to predict DIPL (**Figure 4E**). Given that both methods use RF, we believe that that our larger compound set (aprox. 600 more compounds) for ML model construction, and the concatenation of features were key to improve performance and illustrate the advantages of our approach. Therefore, our algorithm, together with the comprehensive dataset that we have curated, represent a valuable addition to the available tool-kit to predict DIPL. Using the SHAP approach[39,40] to interpret the best performing model, it became clear that pKa and LogP are still important features. However, many other features can also significantly contribute to the accuracy of the prediction. While pKa and cLogP values represent a property of the entire molecule, the remaining top features represent substructure properties including but not limited to cationic amphiphilic structures, such as peripheral amines and aromatic rings. However, by conducting this feature analysis on misclassified cases we also identified compounds displaying many PL features that did not induce PL.

Divergent phospholipidosis profiles among closely related analogs that share similar

physicochemical properties suggested that the mechanism for inducing phospholipidosis is complex. Comparing data on related analogs also illustrated the challenge training computational models that reliably predict PL using reduced and simplistic physicochemical properties. On the other hand, identifying highly similar analogs that contain both inducer and non-inducer properties presented an opportunity for medicinal chemistry endeavors aimed at fine-tuning the structural features of phospholipidosis inducers to render them not inducing PL. We identified in ~ 20% of cases very similar compounds displaying opposing PL phenotypes including four pairs comprising a chemical probe and its inactive analog. Chemical probes are frequently used in cell assays to investigate the association between a specific gene and a phenotype[53] and matched negative controls are expected to increase the confidence that observed phenotypes are indeed driven by the targeted protein. Our results highlight the need to also comprehensively characterize negative controls for phenotypes such as PL, as observed phenotype might be caused by PL induction rather than the presumed target. This hypothesis is in agreement with recent reports on distinct off-target effects of probes and their negative controls.[54]

Our ML model and experimental data identified 21 chemical probes of high-quality chemical probes in the Chemical Probes Portal (https://www.chemicalprobes.org/) that induce phospholipidosis, some at the recommended cellular concentrations. Thus, the additional annotation of the compounds evaluated in the Chemical Probes Portal will add value for the biomedical community to this highly annotated compound set. With the improved PL prediction algorithm and the live-cell assay, we moreover provide a tool set to identify masked PL inducers among small molecule screening collections to increase robustness and

reproducibility in biomedical research. However, closely related analogues such as chemical probes and their controls which differ in PL inductions also offer interesting tools studying consequences of PL on cellular signaling and potentially, they will lead to PL biomarkers for drug safety studies in the future.

Limitations of the study

While the ML models investigated in this study demonstrated robust predictive capabilities concerning PL property in small molecules, the existence of structurally related compounds capable of inducing varying PL inductions (as exemplified by the MMPs identified here) raises considerations regarding the efficacy of *in silico* methodologies that heavily rely on similarity-property principle. This not only emphasizes the intricate nature of PL mechanisms and underscores the necessity for more advanced modeling techniques and descriptors to comprehensively characterize molecular structures, but also underscores the elucidation of PL phenomena in target-based mechanisms since current rationale has predominantly centered around chemical patterns and physicochemical properties (a generalized PL mechanism). This is particularly relevant for structurally similar molecules with different PL inductions, which often share common chemical features. Disparities in off-target profiles related to lipid metabolism or synthesis among analogs could potentially serve as an explanatory avenue meriting thorough investigation. Additionally, the pKa calculations focused solely on the most basic center of small molecules, overlooking the contributions of other weak basic groups. Consequently, the representation of overall molecular basicity, particularly in the context of multiprotic molecules, might lack accuracy and fail to adequately characterize their protonation

states. Moreover owing the distinct morphologies inherent to diverse cell lines, specific validation and analysis should be performed in each new cell system. A comprehensive profiling within the cell system of interest or in diverse cell lines is recommended to improve predictability. Regrettably, due to cost and time issues, we did not test all compounds of the Chemical Probes Portal across different cell lines.

~~Significance~~SIGNIFICANCE

Phospholipidosis is a pathological condition characterized by an excessive accumulation of phospholipids in lysosomes, which is frequently induced by the exposure of cells to exogenous substances, including various pharmacological agents. This DIPL ~~drug-induced phospholipidosis (DIPL)~~ can have significant implications for drug safety and side effect profiles, potentially causing toxicity and should therefore be identified early on in the screening process for small molecules. We present a comprehensive analysis of diverse compounds for their capability to induce phospholipidosis. Assessing two diverse chemical libraries in a new high-content live cell assay, we annotated DIPL over time. The results were used to identify and implement an improved state-of-the-art machine learning algorithm to annotate phospholipidosis induction of compounds effectively, extending the currently used features to predict phospholipidosis and rationalizing the prediction via using SHAP ~~the SHapley Additive exPlanations (SHAP)~~ approach. By incorporating these elements, we not only validated the effectiveness of our annotation approach but also demonstrate that close analogs are able to induce divergent phospholipidosis phenotypes. We thus contribute to the knowledge and collective understanding of phospholipidosis as well as developed tools for compound

annotation.

Figure legends

Figure 1. Phospholipidosis detection in cells using compounds known to induce phospholipidosis. (**A**) Chemical structures with their calculated pKa and cLogP values of drugs known to induce phospholipidosis (amiodarone, chloroquine, chlorpromazine, sertraline, tamoxifen and haloperidol) and their pharmaceutical application area. Cationic amine group is highlighted in red. (**B**) Schematic representation of live-cell phospholipidosis protocol. (**C**) Fluorescent images of stained (blue: DNA/Nuclei, green: lysosomes, red: phospholipidosis dots) U2OS cells and the detection of the cellbody area and phospholipidosis (PL) dots after Cellpathfinder analysis. (**D**) Normalized PL total intensity of U2OS cells exposed to 10 µM, 1 µM and 0.5 µM of reference compounds (amiodarone, chloroquine, chlorpromazine, sertraline, tamoxifen, and haloperidol) in comparison to cells exposed to 0.1 % DMSO after 48h. Error bars show standard error of mean (SEM) of ~~two~~ four biological ~~duplicates~~replicates. (**E**) Representative fluorescent images of stained (blue: DNA/Nuclei, red: phospholipidosis dots) U2OS cells treated with an PL inducer (haloperidol at 10 µM) in comparison to cells exposed to 0.1 % DMSO.

Figure 2. Live-cell assay to distinguish phospholipidosis inducer and non-inducers and the screening of two in-house datasets. (**A**) Volcano plot of compounds in the chemogenomic compound set tested for phospholipidosis inducers or non-inducers in a live-cell assay using U2OS cells over 48h. Each dot represents a compound tested at 10 µM. The x-axis shows the

log2-fold change in normalized PL total intensity between cells treated over 6h and cells treated over 48h. The y-axis shows the negative log(q value) of the adjusted p-value for each compound. Reference compounds are shown in red. The orange square indicates compounds that are defined as "inducers" for significant phospholipidosis induction (normalized PL intensity > 2). 53 "inducers" compounds are highlighted in a heatmap showing normalized PL intensity after 12h, 24h and 48h of compound treatment, with orange representing high induction and blue representing low induction. Data was normalized against the average of DMSO (0.1%) treated cells. All experiments were performed in biological duplicates. Significance was calculated using an unpaired multiple test analysis using GraphPad Prism 8.4.3. (**B**) Volcano plot of compounds in the lysosomal modulation set tested for phospholipidosis inducers or non-inducers in a live-cell assay using U2OS cells over 48h. ~~Each dot represents a compound tested at 10 µM. The x-axis shows the log2-fold change in normalized PL total intensity between cells treated over 6h and cells treated over 48h. The y-axis shows the negative log(q value) of the adjusted p-value for each compound. Reference compounds are shown in red. The orange square indicates the threshold for significant phospholipidosis induction (normalized PL intensity > 2).~~ 29 "inducers" compounds are highlighted in a heatmap showing normalized PL intensity after 12h, 24h, 48h, and 72h of compound treatment~~, with orange representing high induction and blue representing low induction~~. (**C**) Schematic overview of the different data sets used in this study (~~Curated~~ curated data set, ~~Chemogenomic~~ chemogenomic compound set, ~~Lysosomal~~ lysosomal modulation set) and the number of overlapping compounds. (**D**) Bar diagram of different data sets used in this study (curated data set, chemogenomic compound set, lysosomal modulation set). Bars with a

blue color are showing the number of compounds defined as “non-inducer”, while orange bars show the number of “inducers” of every compound set.

Figure 3. Flowchart of the analysis. The process began with curation of reported phospholipidosis data from scientific literature, followed by the application of a multistep structural standardization protocol to obtain high-confidence data for model construction. Initially, a simple physicochemical property model was applied to the curated data set. However, it was found to have unsatisfactory performance, indicating the need for a more robust model capable of accurately distinguishing between inducers and non-inducers. Therefore, ML models were subsequently developed. To validate the performance of the ML model, two in-house screening datasets were utilized - a chemogenomic compound set and a lysosomal modulation set, in addition to a test set. The evaluation metrics were used to select the best ML model, which was subsequently used for practical applications in predicting chemical probes in the Chemical Probes Portal. The metric of recall assesses the model's ability to correctly identify positive samples, while the F1 score is a measure of the harmonic mean of precision and recall. Accuracy provides an overall estimate of the model's prediction accuracy.

Figure 4. Performance machine leaning models in combination with different molecular representations. The violin plots illustrate the performance of four different ML algorithms, which were based on MACCS and ECFP4 fingerprints concatenated with calculated LogP and pKa. The models were evaluated using the test set (**A**), chemogenomic compound set (**B**) and the lysosomal modulation set (**C**) across 10 independent trials. Based on the best trial of the

RF with MACCS concatenated features, ROC curves for three sets are plotted (**D**) and the statistical performance measures of the computational models, including the simple physicochemical model and FP-ADMET model are presented (**E**). TP refers to "true positive", TN to "true negative", FP to "false positive", and FN to "false negative". The metric of recall assesses the model's ability to correctly identify positive samples, while the F1 score is a measure of the harmonic mean of precision and recall. Accuracy provides an overall estimate of the model's prediction accuracy; BA is the arithmetic mean of sensitivity and specificity when dealing with imbalanced data; MCC a metric used to assess the performance of binary classification models, a score of 0 indicates that the classifier performs no better than random guessing. The Receiver Operating Characteristic (ROC) curve and its corresponding Area Under ROC (AUC) were utilized as a threshold-independent metric to further assess the overall predictive performance. With the exception of the MCC, which ranges from -1 to 1, all other metrics are scaled between 0 and 1, with a score of 1 indicating optimal classification performance.

Figure 5. Top ranked features analysis and rationalization of the prediction on exemplary cases. (**A**) Shown is the distribution of the top 15 features identified by SHAP analysis, which was conducted using the RF algorithm in conjunction with MACCS (concatenated with cLogP and pKa) for phospholipidosis inducer prediction. Each feature's line is represented by a series of dots, where each dot corresponds to a molecule, and the stacking of dots shows the density of SHAP values. A higher positive SHAP value suggests a more significant contribution to the prediction of phospholipidosis inducers. Atom symbols and bond types are provided at the

bottom of the figure for reference. (**B**) To aid in understanding the MACCS substructure patterns, the top 13 features (excluding pKa and cLogP) have been mapped back to a chemical structure, which is highlighted in lime. (**C**) Four exemplary compounds coming from in-house data sets depict how the model predicts the induction of phospholipidosis. Two of these compounds (top) were correctly predicted, while the other two (bottom) were misclassified. The contribution of each feature is presented alongside the SHAP force plot, which highlights the features that positively or negatively contribute to the inducer prediction with red and blue colors, respectively. In the force plot, features with a value of zero or one indicating the absence or presence of the corresponding feature and the length of the bars indicates the degree of contribution of each feature towards the prediction. The base value is the base probability and the number with black bold font in plot represents the final output (predicted probability to be PL inducer) of all features.

Figure 6. Structure-phospholipidosis network and exemplary analog pairs with disparate phospholipidosis annotation. The network representation at the top displays MMP-based relationships between compounds from three sets. Each node represents a compound and the edges represent pairwise MMP relationships. The circular and rectangular nodes correspond to compounds from curated and in-house sets, respectively. The orange and blue hues of the nodes denote compounds that act as phospholipidosis inducers and non-inducers, respectively. At the bottom, three analog pairs consisting of chemical probes and their corresponding negative controls (highlighted in the network) are presented. Each pair is accompanied by the compound name and two calculated physicochemical properties (cLogP and pKa). The primary target gene

names corresponding to each probe are provided in the center. Structural modifications between pairs are highlighted in red. Additionally, fluorescent stained image (blue: DNA/Nuclei, red: phospholipidosis dots) of U2OS cells for each compound is also provided.

Figure 7. Prediction and experimental validation of high-quality chemical probes from ~~The~~ the Chemical Probes Portal inducing phospholipidosis. (**A**) Target family distribution for the 21 probes confirmed to induce phospholipidosis. **(B)** Details of ten chemical probes correctly ~~predicte~~predicted by the ML model and experimentally confirmed to be phospholipidosis inducers recommended for use at high concentrations (1 µM or 10 µM in cell-based assays). Each probe is accompanied by its name, primary target (gene name), recommended in-cell concentration by the Chemical Probes Portal, and predicted probability of inducing PL. (**C**) Heat map showing normalized PL intensity in U2OS cells after 24h, 48h and 72h of selected compounds ~~(amiodarone, UNC064, AZ191, I-CBP112, A-770041, GSK343, SGC-CBP30, adavosertib, NVP-2, borussertib)~~ treatment at two concentrations (1 µM and 10 µM) in comparison to cells treated with 0.l % DMSO. Compound A-770041 is highlighted as a chemical probe that is recommended at higher concentrations (1-10 µM) and that induces PL at longer time points. Orange is representing high PL induction and blue is representing low PL induction. Values show SEM of two biological duplicates.

Figure 1

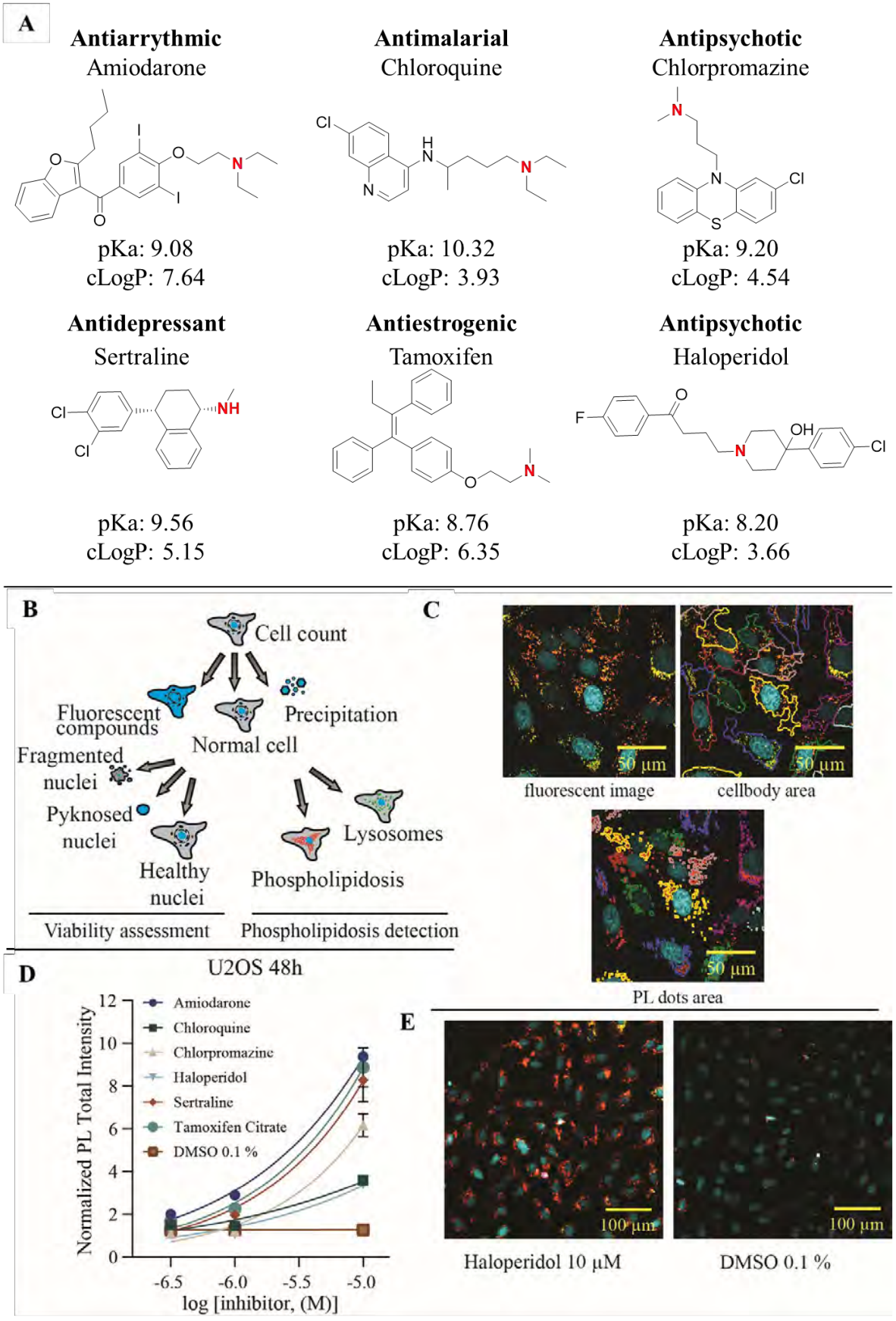

A
Antiarrythmic
Amiodarone
pKa: 9.08
cLogP: 7.64
Antimalarial
Chloroquine
pKa: 10.32
cLogP: 3.93
Antipsychotic
Chlorpromazine
pKa: 9.20
cLogP: 4.54
Antidepressant
Sertraline
pKa: 9.56
cLogP: 5.15
Antiestrogenic
Tamoxifen
pKa: 8.76
cLogP: 6.35
Antipsychotic
Haloperidol
pKa: 8.20
cLogP: 3.66
B
Cell count
Fluorescent compounds
Precipitation
Normal cell
Fragmented nuclei
Pyknosed nuclei
Healthy nuclei
Lysosomes
Phospholipidosis
Viability assessment
Phospholipidosis detection
C
50 µm
fluorescent image
cellbody area
PL dots area
D
U2OS 48h
Normalized PL Total Intensity
log [inhibitor, (M)]
Amiodarone
Chloroquine
Chlorpromazine
Haloperidol
Sertraline
Tamoxifen Citrate
DMSO 0.1 %
E
100 µm
Haloperidol 10 µM
DMSO 0.1 %

Figure 2

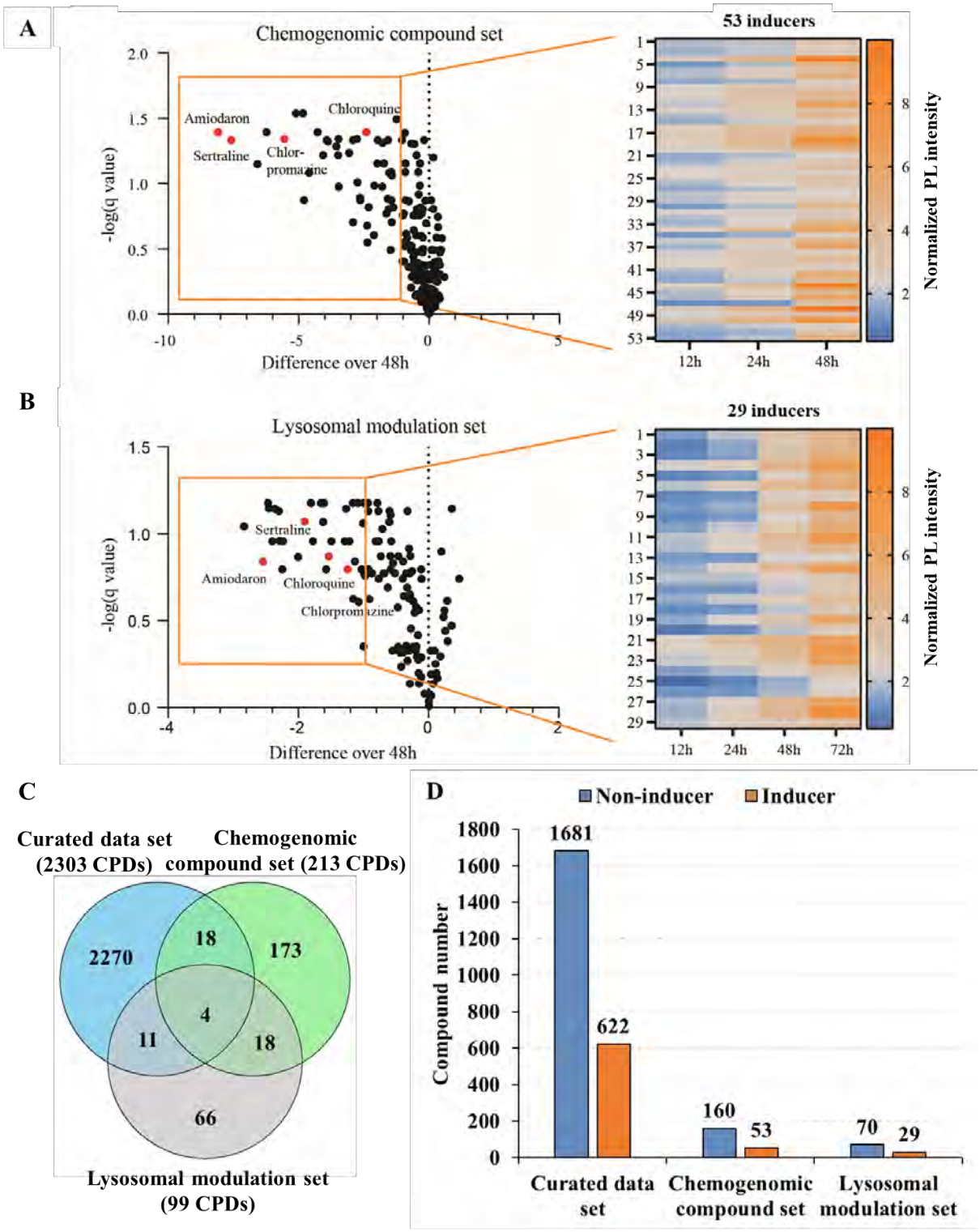

A
Chemogenomic compound set
-log(q value)
Difference over 48h
Amiodaron
Sertraline
Chlor-
promazine
Chloroquine
53 inducers
Normalized PL intensity
12h
24h
48h
B
Lysosomal modulation set
-log(q value)
Difference over 48h
Sertraline
Amiodaron
Chloroquine
Chlorpromazine
29 inducers
Normalized PL intensity
12h
24h
48h
72h
C
Curated data set
(2303 CPDs)
Chemogenomic
compound set (213 CPDs)
2270
18
173
4
11
18
66
Lysosomal modulation set
(99 CPDs)
D
Non-inducer
Inducer
Compound number
1681
622
160
53
70
29
Curated data
set
Chemogenomic
compound set
Lysosomal
modulation set

Figure 3

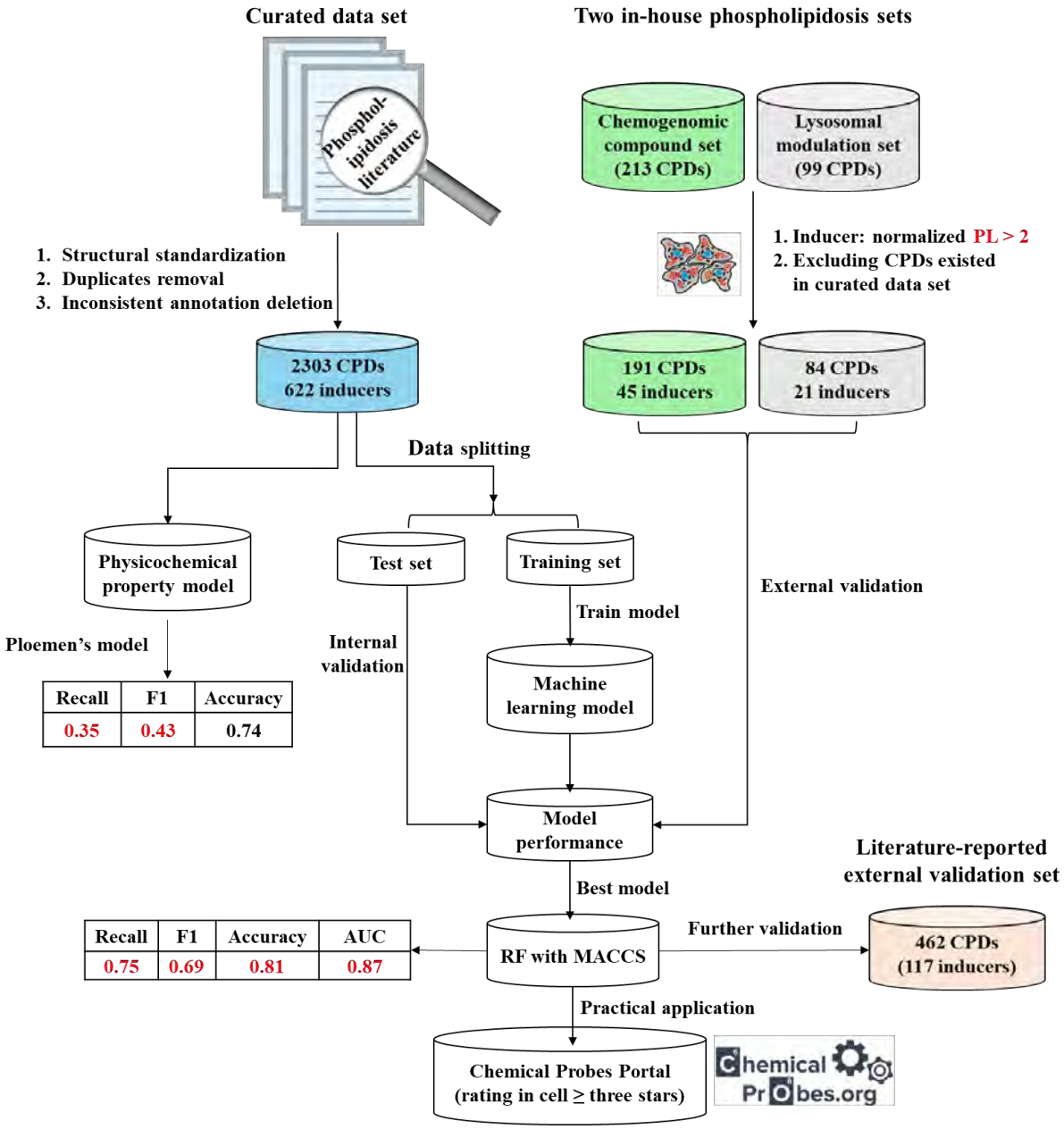
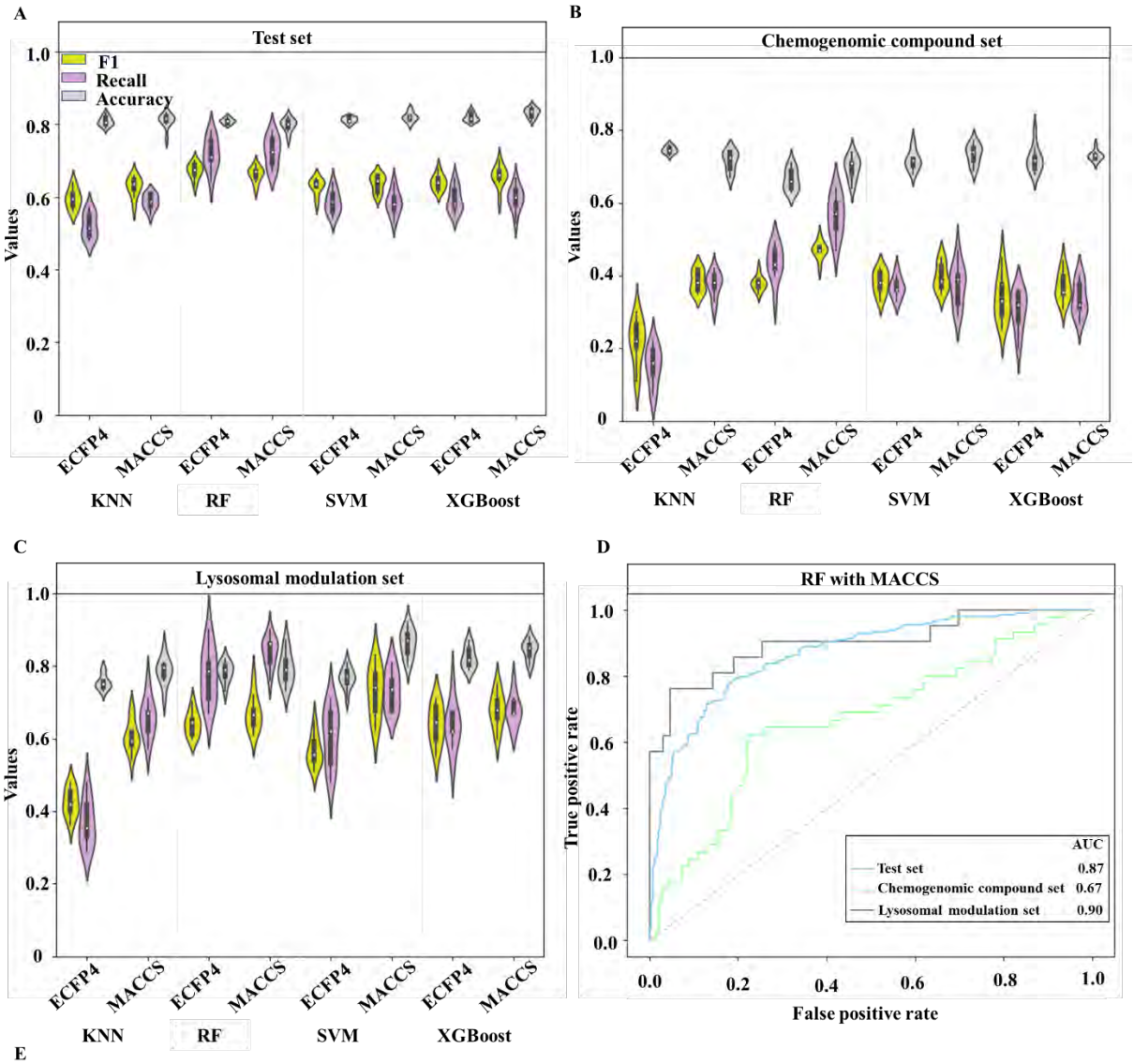

Curated data set
Two in-house phospholipidosis sets
Phospholipidosis literature
Chemogenomic compound set (213 CPDs)
Lysosomal modulation set (99 CPDs)
1. Structural standardization
2. Duplicates removal
3. Inconsistent annotation deletion
1. Inducer: normalized PL > 2
2. Excluding CPDs existed in curated data set
2303 CPDs 622 inducers
191 CPDs 45 inducers
84 CPDs 21 inducers
Data splitting
Physicochemical property model
Test set
Training set
External validation
Ploemen's model
Internal validation
Train model
Machine learning model
Recall F1 Accuracy
0.35 0.43 0.74
Model performance
Literature-reported external validation set
Best model
Recall F1 Accuracy AUC
0.75 0.69 0.81 0.87
RF with MACCS
Further validation
462 CPDs (117 inducers)
Practical application
Chemical Probes Portal (rating in cell ≥ three stars)
Chemical Probes.org

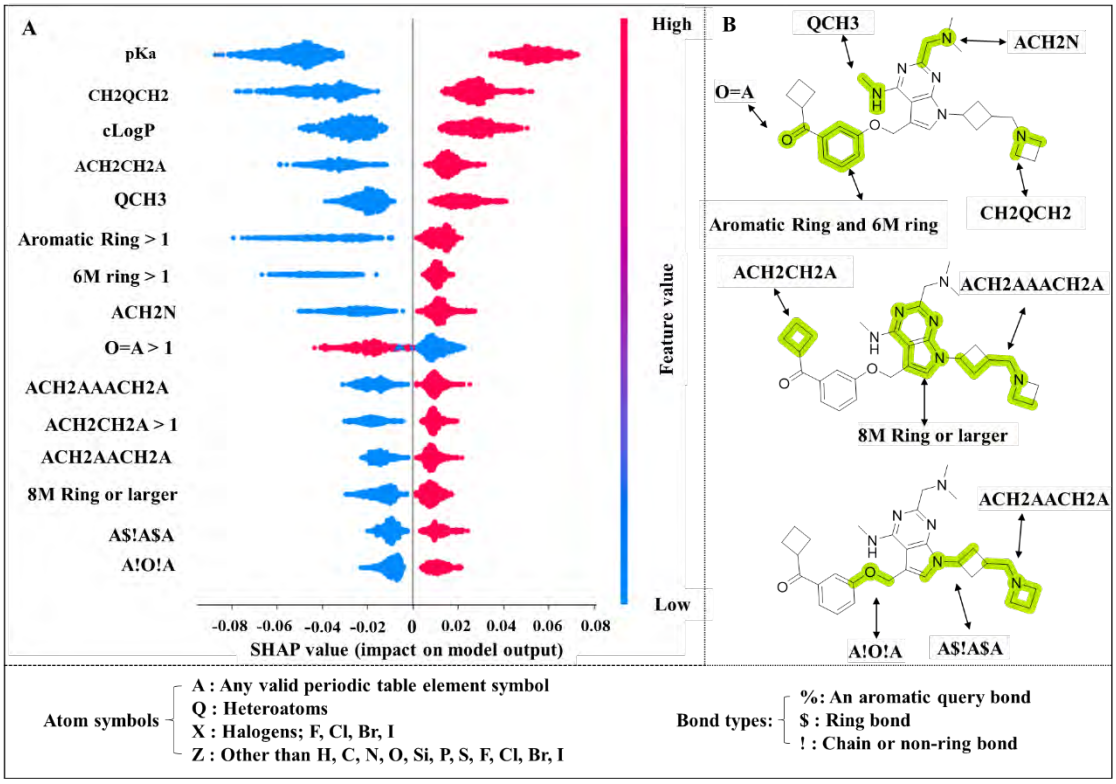
Figure 4



E

Models	Data set (N)	TP	FN	TN	FP	Recall	F1	Accuracy	MCC	BA	AUC
RF with MACCS	Test set (691)	148	49	411	83	0.75	0.69	0.81	0.56	0.79	0.87
	Chemogenomic compound set (191)	29	16	106	40	**0.64**	0.51	0.71	0.33	0.68	0.67
	Lysosomal modulation set (84)	18	3	47	16	**0.86**	0.65	0.77	0.53	0.80	0.90
Ploemen's model	Chemogenomic compound set (191)	15	30	127	19	0.33	0.38	0.74	0.23	0.60	-
	Lysosomal modulation set (84)	10	11	55	8	0.48	0.51	0.77	0.37	0.68	-
FP-ADMET	Chemogenomic compound set (191)	20	25	116	30	0.44	0.42	0.71	0.23	0.62	-
	Lysosomal modulation set (84)	15	6	51	12	0.71	0.62	0.79	0.49	0.76	-

Figure 5


A
pKa
CH2QCH2
cLogP
ACH2CH2A
QCH3
Aromatic Ring > 1
6M ring > 1
ACH2N
O=A > 1
ACH2AAACH2A
ACH2CH2A > 1
ACH2AACH2A
8M Ring or larger
A$!A$A
A!O!A
-0.08 -0.06 -0.04 -0.02 0 0.02 0.04 0.06 0.08
SHAP value (impact on model output)
High
Feature value
Low
B
QCH3
ACH2N
O=A
Aromatic Ring and 6M ring
CH2QCH2
ACH2CH2A
ACH2AAACH2A
8M Ring or larger
ACH2AACH2A
A!O!A
A$!A$A
Atom symbols
A : Any valid periodic table element symbol
Q : Heteroatoms
X : Halogens; F, Cl, Br, I
Z : Other than H, C, N, O, Si, P, S, F, Cl, Br, I
Bond types:
%: An aromatic query bond
$: Ring bond
! : Chain or non-ring bond

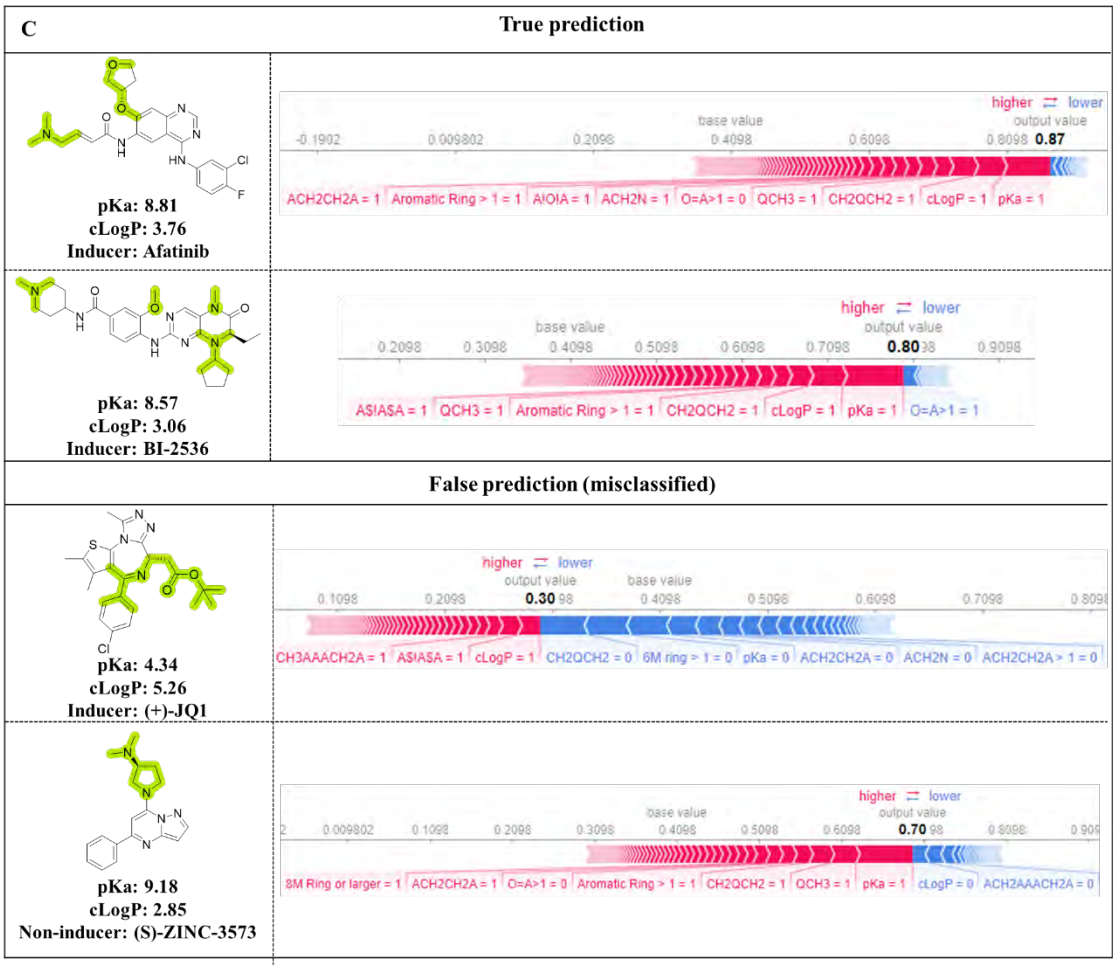

C
True prediction
pKa: 8.81
cLogP: 3.76
Inducer: Afatinib
higher lower
output value
base value
-0.1902 0.009802 0.2098 0.4098 0.6098 0.8098 0.87
ACH2CH2A = 1 Aromatic Ring > 1 = 1 A!O!A = 1 ACH2N = 1 O=A>1 = 0 QCH3 = 1 CH2QCH2 = 1 cLogP = 1 pKa = 1
pKa: 8.57
cLogP: 3.06
Inducer: BI-2536
higher lower
output value
base value
0.2098 0.3098 0.4098 0.5098 0.6098 0.7098 0.80 0.9098
A$!A$A = 1 QCH3 = 1 Aromatic Ring > 1 = 1 CH2QCH2 = 1 cLogP = 1 pKa = 1 O=A>1 = 1
False prediction (misclassified)
pKa: 4.34
cLogP: 5.26
Inducer: (+)-JQ1
higher lower
output value
base value
0.1098 0.2098 0.30 0.4098 0.5098 0.6098 0.7098
CH3AAACH2A = 1 A$!A$A = 1 cLogP = 1 CH2QCH2 = 0 6M ring > 1 = 0 pKa = 0 ACH2CH2A = 0 ACH2N = 0 ACH2CH2A > 1 = 0
pKa: 9.18
cLogP: 2.85
Non-inducer: (S)-ZINC-3573
higher lower
output value
base value
0.009802 0.1098 0.2098 0.3098 0.4098 0.5098 0.6098 0.70 0.8098
8M Ring or larger = 1 ACH2CH2A = 1 O=A>1 = 0 Aromatic Ring > 1 = 1 CH2QCH2 = 1 QCH3 = 1 pKa = 1 cLogP = 0 ACH2AAACH2A = 0

Figure 6

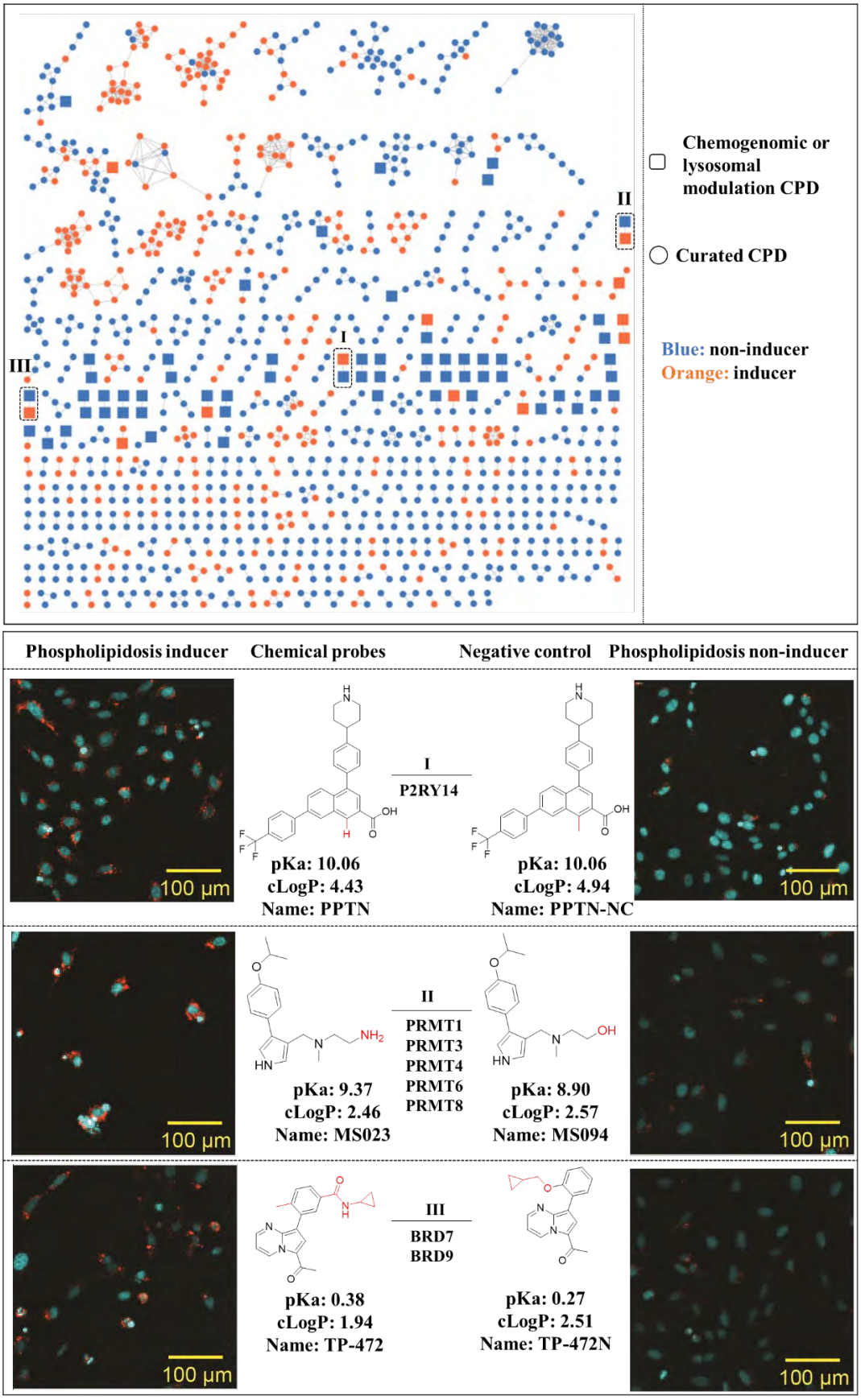
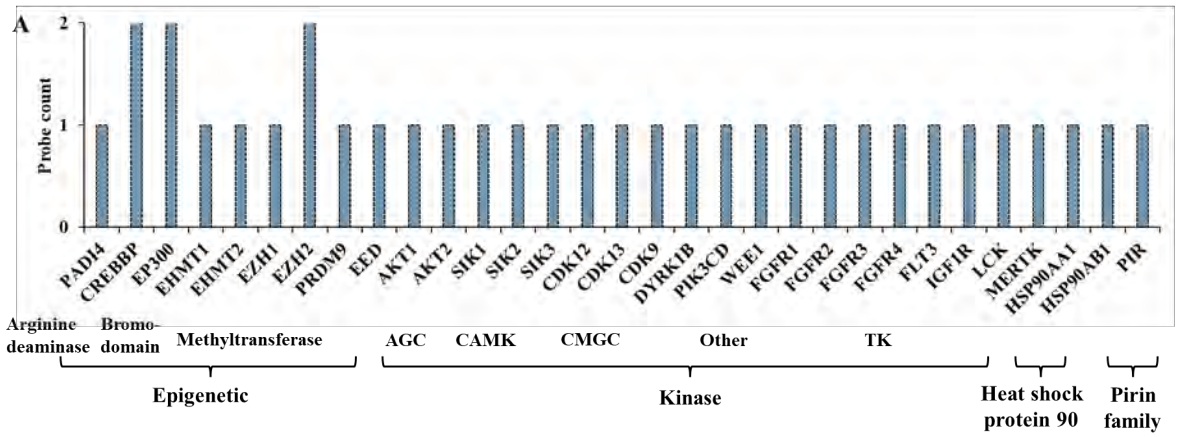
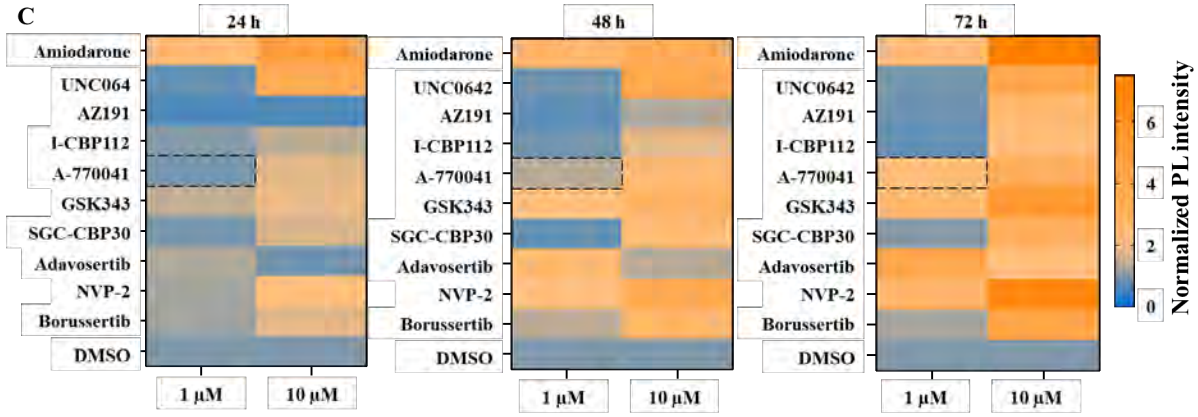

Chemogenomic or lysosomal modulation CPD
Curated CPD
Blue: non-inducer
Orange: inducer
I
II
III
Phospholipidosis inducer
Chemical probes
Negative control
Phospholipidosis non-inducer
I
P2RY14
pKa: 10.06
cLogP: 4.43
Name: PPTN
pKa: 10.06
cLogP: 4.94
Name: PPTN-NC
II
PRMT1
PRMT3
PRMT4
PRMT6
PRMT8
pKa: 9.37
cLogP: 2.46
Name: MS023
pKa: 8.90
cLogP: 2.57
Name: MS094
III
BRD7
BRD9
pKa: 0.38
cLogP: 1.94
Name: TP-472
pKa: 0.27
cLogP: 2.51
Name: TP-472N
100 μm

Figure 7



B

Chemical probe name	Primary target	Recommended in-cell concentration (μM)	Predicted probability (inducers)
UNC0642	EHMT2, EHMT1	[0.1, 1]	0.90
AZ191	DYRK1B	[0.02, 3]	0.86
I-CBP112	EP300, CREBBP	≤ 1	0.83
A-770041	**LCK**	**[1, 10]**	**0.82**
GSK343	EZH2	[0.1, 3]	0.81
SGC-CBP30	EP300, CREBBP	[0.1, 1]	0.79
Adavosertib	WEE1	[0.3, 1]	0.78
NVP-2	CDK9	[0.25, 1]	0.76
Borussertib	AKT2, AKT1	≤ 1	0.74
GSK484	**PADI4**	**≤ 10**	**0.70**



STAR METHODS

KEY RESOURCE TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Hoechst33342	Thermo Scientific	Cat#: 62249 https://www.thermofisher.com/order/catalog/product/62249
HCS LipidTOX Red phospholipidosis detection reagent	Thermo Scientific	Cat#: H34351 https://www.thermofisher.com/order/catalog/product/H34351?SID=srch-srp-H34351
LysoTracker Green DND-26	Thermo Scientific	Cat#: L7526 https://www.thermofisher.com/order/catalog/product/de/en/L7526#:~:text=LysoTracker%20Green%20DND%2D26%20is,excited%20using%20a%20FITC%20filter.
Trypan blue 0.4 %	Thermo Scientific	Cat#: 15250061 https://www.thermofisher.com/order/catalog/product/15250061
Critical commercial assays		
AlamarBlue Cell Viability Reagent	Invitrogen	Cat#: DAL1025
Autophagy LC3 HiBiT Reporter Assay	Promega	Cat#: GA1050
Experimental models: cell lines		
U2OS cells (female, 15-years old)	ATCC	HTB-96™
HEK293T (female, fetus)	ATCC	CRL-1573™
HepG2 (male, 15-years old)	ATCC	HB-8065™
Software and algorithms		
PubChemPy	GitHub	https://github.com/mcs07/PubChemPy
KNIME 4.6.2	KNIME Analytics Platform	https://www.knime.com/

canSARchem workflow	KNIME Hub	https://hub.knime.com/danieladolciami/spaces/Public/latest/canSARchem~hquSdFp3di4kiEv
RDKit version 2022.03.5	Package	https://www.rdkit.org/
PaDELPy	GitHub	https://github.com/ecrl/padelpy
MOE 2020.09	Software	https://www.chemcomp.com/Products.htm
ChemAxon/Instant JChem Version 22.16.0	Software	https://chemaxon.com/instant-jchem
scikit-learn 1.0.2	Package	https://scikit-learn.org/stable/
Python 3.9.10	Programming language	https://www.python.org/
Cytoscape version 3.9.1	Software	https://cytoscape.org/
SHAP 0.41.0	Package	https://shap.readthedocs.io/en/latest/index.html
CQ1 microscope software	Yokogawa	v 1.04.03.01
CellPathfinder Software v3.04.02.02	Yokogawa	N/A
GraphPad Prism v8.4.3	Graphpad Software	https://www.graphpad.com/features
Other		
CQ1 microscope	Yokogawa	N/A
384-well cell culture microplate, PS, f-bottom, µClear®	Greiner	Cat#: 781091
ECHO® 550 Acoustic Liquid Handler	Labcyte	N/A
ECHO® source plate	Labcyte	Cat#: P.05525
Cytomat2C24 incubator	Thermo Scientific	N/A
TC20 Automated Cell Counter	Bio-Rad	https://www.bio-rad.com/de-de/product/tc20-automated-cell-counter?ID=M7FBG34VY
PHERAstar® plate reader	BMG Labtech	N/A
DMEM medium plus L-glutamine (high glucose)	Thermo Scientific	Cat#11965084
McCoy's 5A medium	Gibco	Cat#15410604
Fetal bovine serum	Thermo	Cat#26140079

(FBS)	Scientific	
Penicillin-Streptomycin (10.000 U/ml)	Thermo Scientific	Cat#14190144
DPBS	Thermo Scientific	Cat#14190-094
Geneticin™ (G418 Sulfate)	Thermo Scientific	Cat#10131035

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, susanne.mueller-knapp@bmls.de

Materials availability

This study did not generate any new reagents.

Data and code availability

The published article includes all datasets generated during this study, including the curated literature dataset (**Table S5**). The source code of our best ML model for PL prediction is available on: https://github.com/HuabinHu/ML-for-PL-prediction

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and cell line culture

U2OS (female, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100U/ml penicillin and 100 mg/ml streptomycin) at 37°C and 5% CO_2.

HEK293T (female, fetus) cells were regularly tested for mycoplasma infection. Cells were

grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100U/ml penicillin and 100 mg/ml streptomycin) at 37°C and 5% CO_2.

HepG2 (male, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100U/ml penicillin and 100 mg/ml streptomycin) at 37°C and 5% CO_2.

U2OS Autophagy LC3 HiBiT Reporter (female, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in McCoy's 5A medium supplemented by 10% FBS (Thermo Scientific) and 250 µg/mL G418 (Thermo Scientific) at 37°C and 5% CO_2.

All cell lines were obtained from ATCC (details are described in the key resource table) without further authentication. Influence of sex, gender or both is not expected.

~~**EXPERIMENTAL MODEL AND SUBJECT DETAILS**~~

METHOD DETAILS

~~1.~~ Phospholipidosis detection using confocal microscopy

Human Osteosarcoma cells (U2OS) (ATCC® HTB-96), human hepatocellular carcinoma cells (HepG2) (ATCC® HB-8065™) and human embryonic kidney cells (HEK293T) (ATCC® CRL-1573™) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Thermo Scientific) and Penicillin/Streptomycin (Thermo Scientific). HCS LipidTOX Red phospholipidosis detection reagent (Thermo Scientific) was diluted in cell culture medium 1:1000 and incubated for 5 minutes at 37°C and 5% CO_2. To remove any aggregates, the dye was filtered using 0.2 µm sterile filters. Additionally, the cells were stained with 60 nM Hoechst33342 (Thermo Scientific) and 75 nM LysoTracker Green DND-26 (Thermo Scientific). For every cell line 1500 cells per well in 50 µL/well were seeded in 384- well plates

(cell culture microplate, PS, f-bottom, μClear, 781091, Greiner) in culture medium supplemented with the indicated amount of cell staining dyes. All compounds used in this study, were tested at 10 μM over a 48h time period in U2OS cells. Compounds from the Lysosomal modulation set and hit compounds from the chemical probe portal were additionally tested at 1 μM over a 72h period. Reference compounds (amiodarone, chloroquin, chlorpromazin, sertralin, tamoxifen, haloperidol) were tested at three different concentrations (10 μM, 1 μM and 0.5 μM) in all three cell lines (HEK293T, HepG2 and U2OS) and were used as positive controls for all experiments. Cellular shape and fluorescence were measured at 10x magnification at 6h, 12h, 24h, 48h as well as 72h after compound exposure using a CQ1 high-content confocal microscope (Yokogawa). Reference compounds were also imaged at 60x magnification. The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50% (Hoechst33342); Ex 561 nm/Em 617/73 nm, 100 ms, 40% (HCS LipidTOX); Ex 488/Em 525/50 nm, 50 ms, 40% (LysoTracker Green); bright field, 300 ms, 100% transmission, one centered field per well, 7 z-stacks per well with 55 μm spacing. Images were analyzed using the CellPathfinder software (Yokogawa). Viability assessment using a ML algorithm implemented in the CellPathfinder software was performed as described previously by Tjaden et al.[8,55] In brief, cell body and cell nuclei were detected and gated in different categories based on different features of the cells. Cells showing “Hoechst High Intensity Objects” were excluded. Cells defined as “Normal” were further gated in cells showing healthy, fragmented or pyknosed nuclei.[55] Phospholipidosis dots were defined as being included in lysosomes based on LysoTracker staining. Phospholipidosis was calculated using the total intensity (average) of channel 3. Data was normalized against the average of DMSO (0.1%)

treated cells. All experiments were performed in biological duplicates. Significance was shown which was calculated by using an unpaired multiple t test analysis of timepoints 12h and 48h using GraphPad Prism 8.4.3. ~~Significance was shown using a two-way ANOVA analysis.~~

~~2.~~ AlamarBlueTM ~~Cell~~ cell viability assessment of cell staining dyes

Human Osteosarcoma cells (U2OS) (ATCC® HTB-96~~ATCCHTB-96~~) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Thermo Scientific~~Gibco~~) and Penicillin/Streptomycin (Thermo Scientific~~Gibco~~). 5000 cells per well were seeded in 384-well plates (cell culture microplate, PS, f-bottom, µClear, 781091, Greiner) containing 50 µL/well culture medium one day prior to experimental read out. Cells were incubated over night at 37°C and 5 % CO_2. Six different concentrations of HCS LipidTOX Red phospholipidosis detection reagent (Thermo Scientific) (1:4000, 1:2000, 1:1000, 1:500, 1:250, 1:125) or LysoTracker Green DND-26 (Thermo Scientific) (400 nM, 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM) or both (1:1000 HCD LipidTOX Red phospholipidosis detection reagent and 75 nM LysoTracker Green DND-26) were added in six replicates. Staurosporine 10 µM was added as a positive control. After 24h of compound incubation alamarBlueTM HS was added 1:10 (10 µL/well) to every well according to the manufacturer's instructions. Plates were incubated for 1.5 h protected from light at 37°C and 5 % CO_2. Fluorescence at an excitation wavelength of 560 nm and emission wavelength of 590 nm was measured using a PHERAstar® plate reader (BMG Labtech). Data was normalized against the average of DMSO (0.1%) treated cells. Two biological replicates were tested.

~~3.~~ Autophagy LC3 HiBiT ~~Reporter~~ reporter ~~Assay~~assay

Autophagy flux was detected as described previously.[56] In brief, U2OS Autophagy LC3 HiBiT Reporter Cells (Promega) were cultured in McCoy's 5A medium (Gibco) supplemented by 10% FBS (Thermo Scientific~~Gibco~~) and 250 µg/mL G418 (Thermo Scientific~~Gibco~~). 2000 cells per well were seeded in 384-well plates (cell culture microplate, PS, f-bottom, low volume, 784075, Greiner) containing 10 µL/well culture medium one day prior to experimental read out. Cells were incubated over night at 37°C and 5 % CO_2. Compounds (Lysosomal modulation set) were added at two different concentrations (2.5 µM and 10 µM). Nano-Glo® HiBiT Lytic Reagent was added to the cells 1:100 prior to readout. Luminescence was measured 6h, 24h and 48h after compound exposure with or without autophagy induction of 250 nM Torin using a PHERAstar® plate reader (BMG Labtech).

~~DATA CURATION, COMPUTATIONAL ANALYSIS AND MODEL BUILDING~~

~~1.~~ Chemogenomic compound and lysosomal modulation data set curation

The "chemogenomic compound set" was selected from of our in-house compound library, including compounds covering kinases, epigenetic and GPCR targets. The screen was performed to further annotate the already implemented in-house library. The second set called "lysosomal modulation set" was selected from an in-house library of compounds, containing a collection of small molecules involved in autophagy.

~~2.~~ Literature-reported phospholipidosis data curation and structure standardization.

Following a comprehensive review of phospholipidosis-related literature, we manually compiled compounds with reported *in vivo* or *in vitro* phospholipidosis annotations from nine literature sources.[23,24,27,31,41,57-60] Compounds sourced from diverse origins were subjected to rigorous standardization protocols involving several steps. First, the compound's name was

converted to SMILES-format strings with the aid of the PubChemPy package (https://github.com/mcs07/PubChemPy). Subsequently, the chemical structures of the compounds were standardized using an open-access pipeline, namely the canSARchem registration workflow.[61] This pipeline involved several steps such as salt removal, tautomerization, and charge neutralization. This multistep standardization protocol ensured the identification of duplicates, enabling appropriate subsequent analyses.

~~3.~~ Molecular representations

In cheminformatics, various molecular representations, such as molecular fingerprints and numerical physicochemical descriptors, have been developed and widely adopted to encode chemical or physical information of small molecules in a computer-readable format for multiple applications.[62] To explore the performance of ML prediction, we employed three widely-used two-dimensional (2D) fingerprints and one 2D physicochemical descriptor from Molecular Operating Environment (MOE 2020.09, https://www.chemcomp.com/Products.htm), namely: (1) the Molecular ACCess System (MACCS) fingerprint,[63] which contains 166 predefined structural patterns encoded by a fixed position with a length of 166 bits (i.e., 166 substructures); (2) the extended-connectivity fingerprint (ECFP), which is a circular topological fingerprint capturing local atom environments by considering circular layers centered at each heavy atom with increasing bond diameter up to a maximum of a predefined value.[64] Here, we employed the ECFP fingerprint with a bond diameter of 4 (ECFP4), which was folded into a 2048-bit vector; (3) the PubChem fingerprints, which encodes a total of 881 predefined molecular fragments or patterns to represent the chemical information of a molecule

(https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt); and (4) 206 MOE 2D descriptors, which provide partial charge, pharmacophore feature information, subdivided surface areas, and other physical properties in contrast to the previously mentioned fingerprints. The calculation of MACCS and ECFP4 fingerprints was performed using the RDKit package version 2022.03.5 (https://www.rdkit.org/). PubChem fingerprint were generated the by PaDELPy library,[65] while MOE 2D descriptors were computed by MOE 2020.09 software (https://www.chemcomp.com/Products.htm). In addition, the physicochemical properties of cLogP and pKa (the most basic center) were calculated using ChemAxon/Instant JChem version 22.16.0 (https://chemaxon.com/instant-jchem). These properties were concatenated with the molecular representations mentioned above. To fit with the binary nature of the molecular fingerprints, the continuous cLogP and pKa values were transformed into binary values, with a bit set to one if the cLogP value was greater than or equal to 3.0 (or if the pKa value was greater than or equal to 7.4) and a bit set to zero if the cLogP value was less than 3.0 (or if the pKa value was less than 7.4).[41]

~~4.~~ Building a simple physicochemical property model

Previously calculated pKa and cLogP values were used to calculate Ploemen's rule formulated as: $(pKa)^2 + (cLogP)^2 \geq 90$, in conjunction with $pKa \geq 8$ and $cLogP \geq 1$, as the criteria for classifying a compound as a potential phospholipidosis inducer.[22]

~~5.~~ FP-ADMET model

We utilized a recently developed model (https://gitlab.com/vishsoft/fpadmet) capable of

predicting absorption, distribution, metabolism, excretion, and toxicity (ADMET) for the compounds of interest.[38] This model incorporates 20 distinct binary fingerprints, enabling the modeling of over 50 ADMET-related properties. Specifically, we employed the model to assess its performance on our two external sets on PL prediction, aiming to compare it with our own developed model.

~~6.~~ Chemical similarity calculation

Tanimoto coefficient (Tc), as a popular similarity metric, was used to quantify the structural similarity between two compounds.[66] Tc is generally defined as:

$$\mathrm{Tc}(\mathrm{A},\mathrm{B}) = \frac{\mathrm{c}}{\mathrm{a}+\mathrm{b}-\mathrm{c}}$$

where a and b are the number of features present in compounds A and B, respectively, and c is the number of features shared by A and B. Here, molecules are represented by ECFP fingerprint with radius of 2 (ECFP4) and folded into 2048-bit vector. Tc measurements were carried out in RDKit package version 2022.03.5 (https://www.rdkit.org/).

~~7.~~ Machine learning models and hyperparameter optimization

To discriminate phospholipidosis-inducers from non-inducers, four distinct ML algorithms were employed to generate classification tasks. All ML models were generated using the python-implemented scikit-learn 1.0.2 (https://scikit-learn.org/stable/).[67]

~~7.1~~ *k-nearest neighbor*

A *k*-nearest neighbor (*k*-NN) algorithm is a non-parametric and supervised ML method introduced by Fix and Hodges.[68] It can be used to solve both classification and regression

problems. In classification tasks, the algorithm uses predefined distance metrics to calculate the ranked distances between the compound of interest (unknown sample) and its *k* nearest neighbors in the training dataset.[69] The majority vote (class label) among the *k* neighbors is assigned to the compound of interest. The value of *k* is a critical parameter that needs to be optimized to control the bias and variance of the method.

~~7.2~~ Random ~~f~~Forest

A random forest (RF) classifier is a popular supervised ML algorithm that involves an ensemble of decision trees.[70] During the model training, each individual tree is constructed using a bootstrap sampling method[71] and a random subset of features for node splitting, resulting in a diverse set of trees. The prediction for a given input data is obtained by aggregating the outputs of all individual trees and determining the majority class vote. RF is known for its robustness against overfitting and its ability to handle high-dimensional datasets.

~~7.3~~ Support vector machine

Support Vector Machines (SVMs) are a class of supervised ML algorithms widely used for classification tasks.[72] The main objective of SVM is to find the optimal hyperplane(s) in a high-dimensional feature space that can optimally separate support vectors from each class, thereby maximizing the margin or distance between the classes. However, if the input data points are not linearly separable in the original feature space, a kernel function can be applied to map them into a higher-dimensional feature space where they can be linearly separable.[73] The choice of kernel function, including polynomial, radial basis function (RBF), and sigmoid kernel, is critical and often depends on the nature of the data and the classification task.

~~7.4~~ eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGBoost) is a highly popular supervised ensemble machine learning algorithm used for classification and regression tasks.[74] Unlike RF, which relies on independent decision trees and a bagging-based algorithm, XGBoost follows a gradient boosting approach.[75] It constructs the model in a sequential stage-wise fashion by progressively adding weak learners, typically decision trees that perform slightly better than random guessing. The algorithm iteratively improves the ensemble by leveraging the prediction performance of the preceding learner. It dynamically adjusts the weight distribution, giving greater emphasis to samples that have a significant influence on the construction of the subsequent learner. By aggregating multiple shallow learners and considering the importance or weight assigned to each learner, XGBoost effectively reduces bias and enhances predictive performance. XGBoost is implemented in *xgboost* package (https://xgboost.readthedocs.io/en/stable/).

~~7.5~~ *Hyperparameter optimization*

The ML models were fine-tuned by adjusting their parameters through the use of an internal stratified 5-fold cross-validation technique, which was implemented using the scikit-learn package[67] version 1.0.2 (https://scikit-learn.org/stable/) implemented in python version 3.9.10. For the *k*-NN algorithm, the optimal k values were selected from a range of [1, 3, 5, 7, 10], and the leaf size was adjusted with values of 30, 40, and 50.

For the Random Forest algorithm, the following parameters were optimized: (1) the number of decision trees ("n_estimators": 100, 200, 300, 400); (2) the number of features considered during the search for the best split ("max_features": sqrt, log2, 0.7); (3) the minimum number of samples required at a leaf node ("min_samples_leaf": 1, 3, 5, 10); (4) the maximum depth

of the tree ("max_depth": 7, 10, 12, None). Additionally, the model was built using a balanced class weight.

For the SVM algorithm, the following parameters, which greatly impact the performance of the model, were fine-tuned: (1) the regularization parameter ("C": 0.1, 1, 10, 50, 100, 200, 300, 400, 500, 1000); (2) the selection of the kernel function (polynomial, RBF, and sigmoid kernel); (3) the kernel coefficient gama ("gama": 1, 0.1, 0.01, 0.001).

For XGBoost algorithm, the following parameters were optimized: (1) the step size shrinkage used for update (“learning rate”: 0.05, 0.15, 0.25); (2) maximum depth of a tree (“max_depth”: 3, 5, 7, 9); (3) minimum sum of instance weight (hessian) needed in a child (“min_child_weight”: 1, 3, 5, 7); (4) minimum loss reduction required to make a further partition on a leaf node of the tree ("gamma": 2, 4); (5) the portion of columns to be randomly samples for individual base trees (“colsample_bytree”: 0.3, 0.5); (6) the number of decision trees (n_estimators: 100, 300, 400, 500).

~~8.~~ *Performance evaluation*

For all ML models, 70% of the curated phospholipidosis data were used for model training and hyperparameter optimization. The best performing hyperparameter combinations were then selected to generate the final classifier, which was trained with the complete training set. The remaining 30% of curated data (test set), coupled with two external data sets (chemogenomic compound and lysosomal modulation set), were utilized for evaluating the prediction performance of the model in a realistic scenario using 10 independent trials. To assess the model's performance, the recall, F1-score (F1), accuracy, balanced accuracy (BA), and Matthew’s Correlation Coefficient (MCC)were calculated using the following equations:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

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

$$\text{BA} = \frac{0.5\text{TP}}{\text{TP} + \text{FN}} + \frac{0.5\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

where TP denotes “true positives”, TN “true negatives”, FP “false positives”, and FN “false negatives”. Recall measures the model's ability to detect positive samples; F1 score indicates the harmonic mean of precision and recall; accuracy represents the overall prediction accuracy of the model; BA is the arithmetic mean of sensitivity and specificity when dealing with imbalanced data; MCC a metric used to assess the performance of binary classification models, a score of 0 indicates that the classifier performs no better than random guessing. Moreover, the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under ROC (AUC) are also utilized as a threshold-independent metric to further assess the overall predictive performance. With the exception of the MCC, which ranges from -1 to 1, all other metrics are scaled between 0 to 1 with 1 representing perfect classification.

~~9.~~ Model interpretation and probability calculation

In order to gain better insights into the underlying factors that contribute to the phospholipidosis prediction, we utilized a locally interpretable explanation method called Shapley Additive exPlanations (SHAP).[39,40] This approach, based on game theory, was originally developed to allocate the total gain among players in a cooperative team, allowing for the evaluation of individual player contributions to the outcome of a game. In the context

of our phospholipidosis prediction model, Shapley values were used to estimate the contributions of each feature to the final output. The magnitude and direction (sign) of the feature contribution are manifested by the Shapley values. Positive Shapley values indicate that a feature positively contributes to the prediction of phospholipidosis inducers, whereas for features with negative contributions to inducer prediction (i.e., positive contributions to non-inducer prediction), Shapley values with a negative sign will be used. All features work in a cooperative manner, and the final output prediction reflects the additive Shapley values of all features. To calculate the predicted probability of a test compound using SHAP, the method first calculates the expected probability also referring to as the base probability. The base probability is obtained as the average model output over training set instances and corresponds to the predicted probability of a test compound with unknown feature values. Then, SHAP values are calculated for each feature of the test compound quantifying the contribution of each feature to the predicted probability relative to the expected value. The SHAP values are added to the expected value to modify the base probability. By combining the SHAP values with the base probability, we can obtain the final output probability of the ML model. SHAP analysis is carried out via using SHAP package version 0.41.0 (https://shap.readthedocs.io/en/latest/index.html).

~~10.~~ Matched molecular pair generation

Matched molecular pair (MMP) is defined as a pair of compounds distinguished by a single site.[46] In order to identify structurally similar pairs, a computationally efficient fragmentation algorithm was introduced to systematically generate MMPs in large-scale manner. This

involved systematically fragmenting exocyclic single bonds in the compounds, yielding two fragments at each step.[47] To obtain analogs that typically generated during compound optimization, strict size constraints were implemented during the fragmentation process to derive a core and substituent fragment. In particular, the core fragment was mandated to consist of no fewer than 1.5 times the number of non-hydrogen atoms present in the substituent fragment. Furthermore, the maximal size of fragment substituent was set to 13 non-hydrogen atoms.[76] All MMPs were generated using MMP-related nodes that were integrated into KNIME 4.6.2 (https://www.knime.com/). Following the generation of MMPs, they were organized in a network format using Cytoscape version 3.9.1 (https://cytoscape.org/).

~~11.~~ Chemical Probes Portal

Compounds from the Chemical Probes Portal (https://www.chemicalprobes.org/), a publicly available resource based on expert reviews that aims to enable the assessment, selection, and utilization of chemical probes,[33] were downloaded (accessed in August, 2022). Only the compounds with an in-cell rating of three stars or more were selected, resulting in 321 high-quality chemical probes (from 795 probes in total). The best machine learning model was used to predict if any of them could potentially induce phospholipidosis. After removing those already existing in the curated dataset, we further prioritized the probes with a predicted probability value greater than or equal to 0.7 for experimental validation, yielding 31 probes for experimental confirmation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was performed using GraphPad Prism 8.4.3. Statistical details and definition of parameters are stated in the star methods sections and figure legends. Sample sizes are indicated in the respective figure legend.

Acknowledgements

A.T., S.M., S.K. received financial support for the research, authorship and publication of this article: A.T., S.M., S.K. are supported by SGC, a registered charity (no. 1097737) that receives funds from Bayer AG, Boehringer Ingelheim, the Canada Foundation for Innovation, Eshelman Institute for Innovation, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], EU/EFPIA/OICR/McGill/KTH/Diamond, Innovative Medicines Initiative 2 Joint Undertaking [EUbOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, the Sao Paulo Research Foundation-FAPESP and Takeda as well as support from the German translational cancer network DKTK and the Frankfurt Cancer Institute (FCI). A.T. is supported by the SFB 1177 'Molecular and Functional Characterization of Selective Autophagy'. The CQ1 microscope was funded by FUGG (INST 161/920-1 FUGG). We ~~would~~ thank Martin Schroeder for drawing our attention to this exciting topic. A.A.A. received the support of a fellowship from "la Caixa" Foundation (LCF/BQ/PI22/11910016). A.A.A. and H.H. were formerly supported by a Fellowship from The Institute of Cancer Research.

Author contributions

H.H., A.T., S.K., A.A.A., and S.M. designed the research. A.T.~~, S.K.~~ and S.M performed and managed all the experimental part of the work. H.H. and A.A.A. performed and managed all the computational work. H.H., A.T., S.K., A.A.A., and S.M. conducted data analysis and interpretation. H.H., A.T., A.A.A., and S.M. wrote the manuscript with contributions from S.K.

Declaration of Interests

A.A.A. and H.H. are/were employees of The Institute of Cancer Research (ICR), which has a commercial interest in a range of drug targets, including PARP and protein kinases. The ICR operates a Rewards to Inventors scheme whereby employees of the ICR may receive financial benefit following commercial licensing of a project. A.A.A. has been instrumental in the creation/development of canSAR, the Chemical Probes Portal and Probe Miner. A.A.A. is/was a consultant of DarwinHealth, Inc.

Inclusion and Diversity

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

References

1. Shahane, S.A., Huang, R., Gerhold, D., Baxa, U., Austin, C.P., and Xia, M. (2014). Detection of phospholipidosis induction: a cell-based assay in high-throughput and high-content format. J. Biomol. Screen. *19*, 66-76. https://doi.org/10.1177/1087057113502851.
2. Breiden, B., and Sandhoff, K. (2019). Emerging mechanisms of drug-induced phospholipidosis. Biol. Chem. *401*, 31-46. https://doi.org/10.1515/hsz-2019-0270.
3. Kodavanti, U.P., and Mehendale, H.M. (1990). Cationic amphiphilic drugs and phospholipid storage disorder. Pharmacol. Rev. *42*, 327-354.
4. Bik, E., Orleanska, J., Mateuszuk, L., Baranska, M., Majzner, K., and Chlopicki, S. (2022). Raman and fluorescence imaging of phospholipidosis induced by cationic amphiphilic drugs in endothelial cells. Biochim. Biophys. Acta Mol. Cell Res. *1869*, 119186. https://doi.org/10.1016/j.bbamcr.2021.119186.
5. Donato, M.T., and Gomez-Lechon, M.J. (2012). Drug-induced liver steatosis and phospholipidosis: cell-based assays for early screening of drug candidates. Curr. Drug Metab. *13*, 1160-1173. https://doi.org/10.2174/138920012802850001.
6. Edwards, A., and Hartung, I.V. (2021). No shortcuts to SARS-CoV-2 antivirals. Science *373*, 488-489. https://doi.org/10.1126/science.abj9488.
7. Tummino, T.A., Rezelj, V.V., Fischer, B., Fischer, A., O'Meara, M.J., Monel, B., Vallet, T., White, K.M., Zhang, Z., Alon, A., et al. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. Science *373*, 541-547. https://doi.org/10.1126/science.abi4708.
8. Tjaden, A., Chaikuad, A., Kowarz, E., Marschalek, R., Knapp, S., Schroder, M., and Muller, S. (2022). Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. Molecules *27*. https://doi.org/10.3390/molecules27041439.
9. Hinkovska-Galcheva, V., Treadwell, T., Shillingford, J.M., Lee, A., Abe, A., Tesmer, J.J.G., and Shayman, J.A. (2021). Inhibition of lysosomal phospholipase A2 predicts drug-induced phospholipidosis. J. Lipid Res. *62*, 100089. https://doi.org/10.1016/j.jlr.2021.100089.
10. Gulbins, E., Palmada, M., Reichel, M., Luth, A., Bohmer, C., Amato, D., Muller, C.P., Tischbirek, C.H., Groemer, T.W., Tabatabai, G., et al. (2013). Acid sphingomyelinase-ceramide system mediates effects of antidepressant drugs. Nat. Med. *19*, 934-938. https://doi.org/10.1038/nm.3214.
11. Gulbins, A., Schumacher, F., Becker, K.A., Wilker, B., Soddemann, M., Boldrin, F., Muller, C.P., Edwards, M.J., Goodman, M., Caldwell, C.C., et al. (2018). Antidepressants act by inducing autophagy controlled by sphingomyelin-ceramide. Mol. Psychiatry *23*, 2324-2346. https://doi.org/10.1038/s41380-018-0090-9.
12. Anderson, N., and Borlak, J. (2006). Drug-induced phospholipidosis. FEBS Lett. *580*, 5533-5540. https://doi.org/10.1016/j.febslet.2006.08.061.
13. Goldman, S.D., Funk, R.S., Rajewski, R.A., and Krise, J.P. (2009). Mechanisms of amine accumulation in, and egress from, lysosomes. Bioanal. *1*, 1445-1459. https://doi.org/10.4155/bio.09.128.
14. Reasor, M.J., Hastings, K.L., and Ulrich, R.G. (2006). Drug-induced phospholipidosis: issues and future directions. Expert Opin. Drug Saf. *5*, 567-583. https://doi.org/10.1517/14740338.5.4.567.
15. Joshi, U.M., Rao, P., Kodavanti, S., Lockard, V.G., and Mehendale, H.M. (1989). Fluorescence studies on binding of amphiphilic drugs to isolated lamellar bodies: relevance to phospholipidosis. Biochim. Biophys. Acta *1004*, 309-320. https://doi.org/10.1016/0005-2760(89)90078-7.
16. Chatman, L.A., Morton, D., Johnson, T.O., and Anway, S.D. (2009). A strategy for risk management of drug-induced phospholipidosis. Toxicol. Pathol. *37*, 997-1005.

https://doi.org/10.1177/0192623309352496.

17. Reasor, M.J. (1989). A review of the biology and toxicologic implications of the induction of lysosomal lamellar bodies by drugs. Toxicol Appl Pharmacol *97*, 47-56. https://doi.org/10.1016/0041-008x(89)90054-9.
18. Gum, R.J., Hickman, D., Fagerland, J.A., Heindel, M.A., Gagne, G.D., Schmidt, J.M., Michaelides, M.R., Davidsen, S.K., and Ulrich, R.G. (2001). Analysis of two matrix metalloproteinase inhibitors and their metabolites for induction of phospholipidosis in rat and human hepatocytes(1). Biochem. Pharmacol. *62*, 1661-1673. https://doi.org/10.1016/s0006-2952(01)00823-1.
19. Kasahara, T., Tomita, K., Murano, H., Harada, T., Tsubakimoto, K., Ogihara, T., Ohnishi, S., and Kakinuma, C. (2006). Establishment of an in vitro high-throughput screening assay for detecting phospholipidosis-inducing potential. Toxicol. Sci. *90*, 133-141. https://doi.org/10.1093/toxsci/kfj067.
20. Coleman, J., Xiang, Y., Pande, P., Shen, D., Gatica, D., and Patton, W.F. (2010). A live-cell fluorescence microplate assay suitable for monitoring vacuolation arising from drug or toxic agent treatment. J. Biomol. Screen. *15*, 398-405. https://doi.org/10.1177/1087057110364242.
21. Bauch, C., Bevan, S., Woodhouse, H., Dilworth, C., and Walker, P. (2015). Predicting in vivo phospholipidosis-inducing potential of drugs by a combined high content screening and in silico modelling approach. Toxicol. In Vitro *29*, 621-630. https://doi.org/10.1016/j.tiv.2015.01.014.
22. Ploemen, J.P., Kelder, J., Hafmans, T., van de Sandt, H., van Burgsteden, J.A., Saleminki, P.J., and van Esch, E. (2004). Use of physicochemical calculation of pKa and CLogP to predict phospholipidosis-inducing potential: a case study with structurally related piperazines. Exp. Toxicol. Pathol. *55*, 347-355. https://doi.org/10.1078/0940-2993-00338.
23. Tomizawa, K., Sugano, K., Yamada, H., and Horii, I. (2006). Physicochemical and cell-based approach for early screening of phospholipidosis-inducing potential. J. Toxicol. Sci. *31*, 315-324. https://doi.org/10.2131/jts.31.315.
24. Hanumegowda, U.M., Wenke, G., Regueiro-Ren, A., Yordanova, R., Corradi, J.P., and Adams, S.P. (2010). Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds. Chem. Res. Toxicol. *23*, 749-755. https://doi.org/10.1021/tx9003825.
25. Schieferdecker, S., Eberlein, A., Vock, E., and Beilmann, M. (2022). Development of an in silico consensus model for the prediction of the phospholipigenic potential of small molecules. Comput. Toxicol. *22*, 100226. https://doi.org/10.1016/j.comtox.2022.100226.
26. Przybylak, K.R., and Cronin, M.T. (2011). In Silico Studies of the Relationship Between Chemical Structure and Drug Induced Phospholipidosis. Mol. Inform. *30*, 415-429. https://doi.org/10.1002/minf.201000164.
27. Przybylak, K.R., Alzahrani, A.R., and Cronin, M.T. (2014). How does the quality of phospholipidosis data influence the predictivity of structural alerts? J. Chem. Inf. Model. *54*, 2224-2232. https://doi.org/10.1021/ci500233k.
28. Sun, H., Shahane, S., Xia, M., Austin, C.P., and Huang, R. (2012). Structure based model for the prediction of phospholipidosis induction potential of small molecules. J Chem Inf Model *52*, 1798-1805. https://doi.org/10.1021/ci3001875.
29. Slavov, S., Stoyanova-Slavova, I., Li, S., Zhao, J., Huang, R., Xia, M., and Beger, R. (2017). Why are most phospholipidosis inducers also hERG blockers? Arch Toxicol *91*, 3885-3895. https://doi.org/10.1007/s00204-017-1995-9.
30. Sun, H., Xia, M., Shahane, S.A., Jadhav, A., Austin, C.P., and Huang, R. (2013). Are hERG channel blockers also phospholipidosis inducers? Bioorg Med Chem Lett *23*, 4587-4590.

https://doi.org/10.1016/j.bmcl.2013.06.034.

31. Fusani, L., Brown, M., Chen, H., Ahlberg, E., and Noeske, T. (2017). Predicting the Risk of Phospholipidosis with in Silico Models and an Image-Based in Vitro Screen. Mol. Pharm. *14*, 4346-4352. https://doi.org/10.1021/acs.molpharmaceut.7b00388.
32. Rodriguez-Perez, R., and Bajorath, J. (2021). Explainable Machine Learning for Property Predictions in Compound Optimization. J. Med. Chem. *64*, 17744-17752. https://doi.org/10.1021/acs.jmedchem.1c01789.
33. Antolin, A.A., Sanfelice, D., Crisp, A., Villasclaras Fernandez, E., Mica, I.L., Chen, Y., Collins, I., Edwards, A., Muller, S., Al-Lazikani, B., and Workman, P. (2023). The Chemical Probes Portal: an expert review-based public resource to empower chemical probe assessment, selection and use. Nucleic Acids Res. *51*, D1492-D1502. https://doi.org/10.1093/nar/gkac909.
34. Nioi, P., Perry, B.K., Wang, E.J., Gu, Y.Z., and Snyder, R.D. (2007). In vitro detection of drug-induced phospholipidosis using gene expression and fluorescent phospholipid based methodologies. Toxicol. Sci. *99*, 162-173. https://doi.org/10.1093/toxsci/kfm157.
35. Steen, N.V.D., Honeywell, R.J., Dekker, H., Meerloo, J.V., Kole, J., Musters, R., Ruijtenbeek, R., Rolfo, C., Pauwels, P., Peters, G.J., and Giovannetti, E. (2018). Resistance to crizotinib in a cMET gene amplified tumor cell line is associated with impaired sequestration of crizotinib in lysosomes. J. Mol. Clin. Med. *1*, 99-106. https://doi.org/10.31083/j.jmcm.2018.02.007.
36. Xue, L., and Liu, P. (2021). Daurisoline inhibits hepatocellular carcinoma progression by restraining autophagy and promoting cispaltin-induced cell death. Biochem. Biophys. Res. Commun. *534*, 1083-1090. https://doi.org/10.1016/j.bbrc.2020.09.068.
37. Herp, D., Ridinger, J., Robaa, D., Shinsky, S.A., Schmidtkunz, K., Yesiloglu, T.Z., Bayer, T., Steimbach, R.R., Herbst-Gervasoni, C.J., Merz, A., et al. (2022). First Fluorescent Acetylspermidine Deacetylation Assay for HDAC10 Identifies Selective Inhibitors with Cellular Target Engagement. Chembiochem *23*, e202200180. https://doi.org/10.1002/cbic.202200180.
38. Venkatraman, V. (2021). FP-ADMET: a compendium of fingerprint-based ADMET prediction models. J Cheminform *13*, 75. https://doi.org/10.1186/s13321-021-00557-5.
39. Štrumbelj, E., and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. *41*, 647-665. https://doi.org/10.1007/s10115-013-0679-x.
40. Shapley, L.S. (1953). 17. A Value for n-Person Games. In Contributions to the Theory of Games (AM-28), Volume II, K. Harold William, and T. Albert William, eds. (Princeton University Press), pp. 307-318. https://doi.org/10.1515/9781400881970-018.
41. Muehlbacher, M., Tripal, P., Roas, F., and Kornhuber, J. (2012). Identification of drugs inducing phospholipidosis by novel in vitro data. ChemMedChem *7*, 1925-1934. https://doi.org/10.1002/cmdc.201200306.
42. Liu, X., Suo, H., Zhou, S., Hou, Z., Bu, M., Liu, X., and Xu, W. (2021). Afatinib induces pro-survival autophagy and increases sensitivity to apoptosis in stem-like HNSCC cells. Cell. Death Dis. *12*, 728. https://doi.org/10.1038/s41419-021-04011-0.
43. Li, Z., Yang, C., Li, X., Du, X., Tao, Y., Ren, J., Fang, F., Xie, Y., Li, M., Qian, G., et al. (2020). The dual role of BI 2536, a small-molecule inhibitor that targets PLK1, in induction of apoptosis and attenuation of autophagy in neuroblastoma cells. J. Cancer *11*, 3274-3287. https://doi.org/10.7150/jca.33110.
44. da Motta, L.L., Ledaki, I., Purshouse, K., Haider, S., De Bastiani, M.A., Baban, D., Morotti, M., Steers, G., Wigfield, S., Bridges, E., et al. (2017). The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. Oncogene *36*, 122-

132. https://doi.org/10.1038/onc.2016.184.

45. Lansu, K., Karpiak, J., Liu, J., Huang, X.P., McCorvy, J.D., Kroeze, W.K., Che, T., Nagase, H., Carroll, F.I., Jin, J., et al. (2017). In silico design of novel probes for the atypical opioid receptor MRGPRX2. Nat. Chem. Biol. *13*, 529-536. https://doi.org/10.1038/nchembio.2334.

46. Griffen, E., Leach, A.G., Robb, G.R., and Warner, D.J. (2011). Matched molecular pairs as a medicinal chemistry tool. J. Med. Chem. *54*, 7739-7750. https://doi.org/10.1021/jm200452d.

47. Hussain, J., and Rea, C. (2010). Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. J. Chem. Inf. Model. *50*, 339-348. https://doi.org/10.1021/ci900450m.

48. Robichaud, J., Fournier, J.F., Gagne, S., Gauthier, J.Y., Hamel, M., Han, Y., Henault, M., Kargman, S., Levesque, J.F., Mamane, Y., et al. (2011). Applying the pro-drug approach to afford highly bioavailable antagonists of P2Y(14). Bioorg. Med. Chem. Lett. *21*, 4366-4368. https://doi.org/10.1016/j.bmcl.2010.12.113.

49. Eram, M.S., Shen, Y., Szewczyk, M., Wu, H., Senisterra, G., Li, F., Butler, K.V., Kaniskan, H.U., Speed, B.A., Dela Sena, C., et al. (2016). A Potent, Selective, and Cell-Active Inhibitor of Human Type I Protein Arginine Methyltransferases. ACS Chem. Biol. *11*, 772-781. https://doi.org/10.1021/acschembio.5b00839.

50. Karim, R.M., Chan, A., Zhu, J.Y., and Schonbrunn, E. (2020). Structural Basis of Inhibitor Selectivity in the BRD7/9 Subfamily of Bromodomains. J. Med. Chem. *63*, 3227-3237. https://doi.org/10.1021/acs.jmedchem.9b01980.

51. Nakayama, K., Szewczyk, M.M., Dela Sena, C., Wu, H., Dong, A., Zeng, H., Li, F., de Freitas, R.F., Eram, M.S., Schapira, M., et al. (2018). TP-064, a potent and selective small molecule inhibitor of PRMT4 for multiple myeloma. Oncotarget *9*, 18480-18493. https://doi.org/10.18632/oncotarget.24883.

52. Szewczyk, M.M., Ishikawa, Y., Organ, S., Sakai, N., Li, F., Halabelian, L., Ackloo, S., Couzens, A.L., Eram, M., Dilworth, D., et al. (2020). Pharmacological inhibition of PRMT7 links arginine monomethylation to the cellular stress response. Nat. Commun. *11*, 2396. https://doi.org/10.1038/s41467-020-16271-z.

53. Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. Nat. Chem. Biol. *11*, 536-541. https://doi.org/10.1038/nchembio.1867.

54. Lee, J., and Schapira, M. (2021). The Promise and Peril of Chemical Probe Negative Controls. ACS Chem. Biol. *16*, 579-585. https://doi.org/10.1021/acschembio.1c00036.

55. Tjaden, A., Giessmann, R.T., Knapp, S., Schroder, M., and Muller, S. (2022). High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. STAR Protoc. *3*, 101791. https://doi.org/10.1016/j.xpro.2022.101791.

56. Lazar, D.F., Gillette, A.A., Lewis, S.R., and Butler, B.L. (2021). Chapter 4 - Multimodal assessment of autophagy in mammalian cells with a novel, LC3-based tandem reporter. In Methods Cell Biol., O. Kepp, and L. Galluzzi, eds. (Academic Press), pp. 39-57. https://doi.org/10.1016/bs.mcb.2021.02.004.

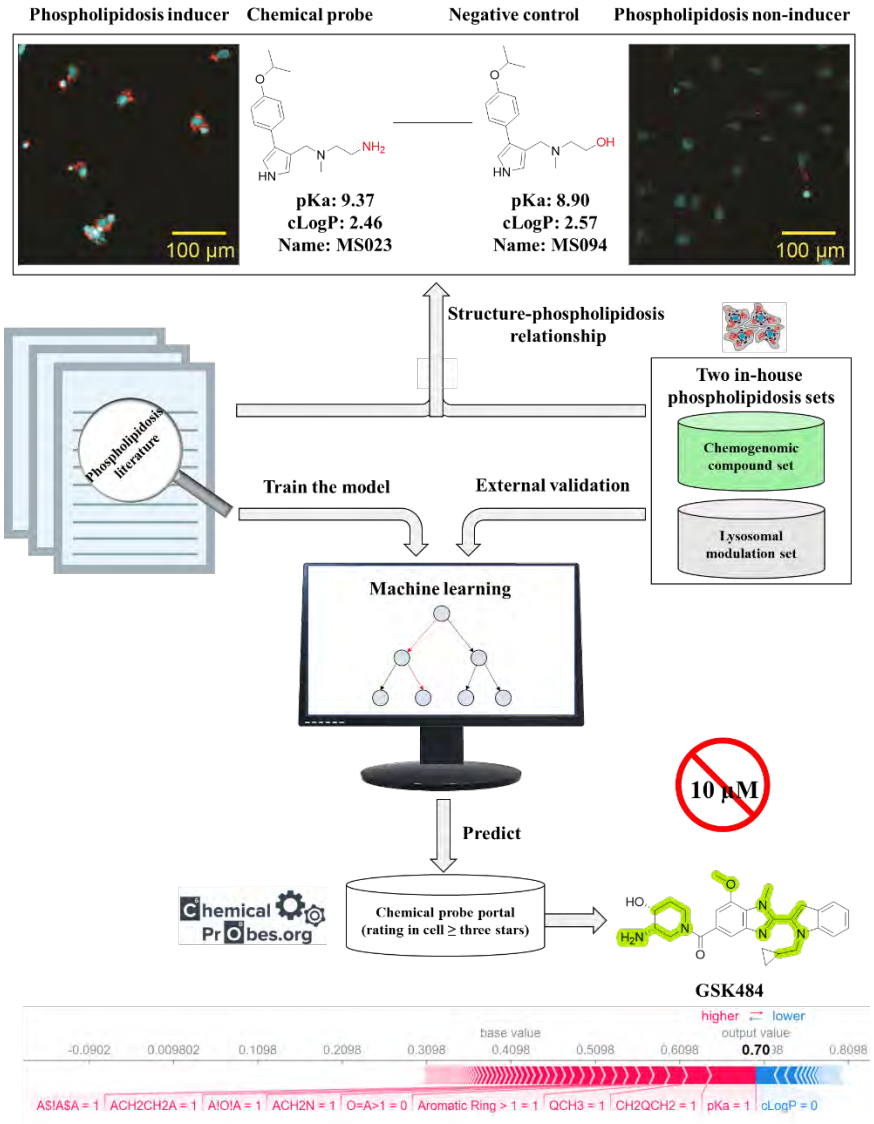
57. Orogo, A.M., Choi, S.S., Minnier, B.L., and Kruhlak, N.L. (2012). Construction and Consensus Performance of (Q)SAR Models for Predicting Phospholipidosis Using a Dataset of 743 Compounds. Mol. Inform. *31*, 725-739. https://doi.org/10.1002/minf.201200048.

58. Goracci, L., Ceccarelli, M., Bonelli, D., and Cruciani, G. (2013). Modeling phospholipidosis induction: reliability and warnings. J. Chem. Inf. Model. *53*, 1436-1446. https://doi.org/10.1021/ci400113t.

59. Pelletier, D.J., Gehlhaar, D., Tilloy-Ellul, A., Johnson, T.O., and Greene, N. (2007). Evaluation of a published in silico model and construction of a novel Bayesian model for predicting phospholipidosis inducing potential. J. Chem. Inf. Model. *47*, 1196-1205. https://doi.org/10.1021/ci6004542.

60. Lowe, R., Glen, R.C., and Mitchell, J.B. (2010). Predicting phospholipidosis using machine learning. Mol. Pharm. *7*, 1708-1714. https://doi.org/10.1021/mp100103e.
61. Dolciami, D., Villasclaras-Fernandez, E., Kannas, C., Meniconi, M., Al-Lazikani, B., and Antolin, A.A. (2022). canSAR chemistry registration and standardization pipeline. J. Cheminform. *14*, 28. https://doi.org/10.1186/s13321-022-00606-7.
62. David, L., Thakkar, A., Mercado, R., and Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. J. Cheminform. *12*, 56. https://doi.org/10.1186/s13321-020-00460-5.
63. Durant, J.L., Leland, B.A., Henry, D.R., and Nourse, J.G. (2002). Reoptimization of MDL keys for use in drug discovery. J. Chem. Inf. Comput. Sci. *42*, 1273-1280. https://doi.org/10.1021/ci010132r.
64. Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. J. Chem. Inf. Model. *50*, 742-754. https://doi.org/10.1021/ci100050t.
65. Yap, C.W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. *32*, 1466-1474. https://doi.org/10.1002/jcc.21707.
66. Maggiora, G., Vogt, M., Stumpfe, D., and Bajorath, J. (2014). Molecular similarity in medicinal chemistry. J Med Chem *57*, 3186-3204. https://doi.org/10.1021/jm401411z.
67. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
68. Fix, E., and Hodges Jr, J.L. (1951). Discriminatory analysis, nonparametric discrimination: consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.
69. Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician *46*, 175-185. https://doi.org/10.1080/00031305.1992.10475879.
70. Breiman, L. (2001). Random Forests. Machine Learning *45*, 5-32. https://doi.org/10.1023/A:1010933404324.
71. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics *7*, 1-26. https://doi.org/10.1214/aos/1176344552.
72. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine Learning *20*, 273-297. https://doi.org/10.1007/BF00994018.
73. Shawe-Taylor, J., and Cristianini, N. (2004). Kernel Methods for Pattern Analysis (Cambridge University Press). https://doi.org/10.1017/CBO9780511809682.
74. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery.
75. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics *29*, 1189-1232, 1144. https://www.jstor.org/stable/2699986.
76. Hu, X., Hu, Y., Vogt, M., Stumpfe, D., and Bajorath, J. (2012). MMP-Cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs. J. Chem. Inf. Model. *52*, 1138-1145. https://doi.org/10.1021/ci3001138.

TOC


Phospholipidosis inducer
Chemical probe
Negative control
Phospholipidosis non-inducer
100 μm
pKa: 9.37
cLogP: 2.46
Name: MS023
pKa: 8.90
cLogP: 2.57
Name: MS094
100 μm
Structure-phospholipidosis relationship
Two in-house phospholipidosis sets
Chemogenomic compound set
Lysosomal modulation set
Phospholipidosis literature
Train the model
External validation
Machine learning
10 μM
Predict
Chemical probe portal (rating in cell ≥ three stars)
GSK484
higher
lower
base value
output value
0.70
-0.0902
0.009802
0.1098
0.2098
0.3098
0.4098
0.5098
0.6098
0.8098
A$!A$A = 1
ACH2CH2A = 1
AlOlA = 1
ACH2N = 1
O=A>1 = 0
Aromatic Ring > 1 = 1
QCH3 = 1
CH2QCH2 = 1
pKa = 1
cLogP = 0