Zhou *et al. Cell Communication and Signaling* (2024) 22:585
https://doi.org/10.1186/s12964-024-01954-7

Cell Communication and Signaling

RESEARCH Open Access

Machine learning-aided discovery of T790M-mutant EGFR inhibitor CDDO-Me effectively suppresses non-small cell lung cancer growth

Rui Zhou[1†], Ziqian Liu[1†], Tongtong Wu[4†], Xianwei Pan[1], Tongtong Li[1], Kaiting Miao[4], Yuru Li[1], Xiaohui Hu[1], Haigang Wu[4†], Andrew M. Hemmings[1,7], Beier Jiang[3*], Zhenzhen Zhang[3*] and Ning Liu[1,2,5,6*†]

Abstract

Background Epidermal growth factor receptor (EGFR) T790M mutation often occurs during long durational erlotinib treatment of non-small cell lung cancer (NSCLC) patients, leading to drug resistance and disease progression. Identification of new selective EGFR-T790M inhibitors has proven challenging through traditional screening platforms. With great advances in computer algorithms, machine learning improved the screening rates of molecules at full chemical spaces, and these molecules will present higher biological activity and targeting efficiency.

Methods An integrated machine learning approach, integrated by Bayesian inference, was employed to screen a commercial dataset of 70,413 molecules, identifying candidates that selectively and efficiently bind with EGFR harboring T790M mutation. In vitro cellular assays and molecular dynamic simulations was used for validation. ***EGFR*** knockout cell line was generated for cross-validation. In vivo xenograft moues model was constructed to investigate the antitumor efficacy of CDDO-Me.

Results Our virtual screening and subsequent in vitro testing successfully identified CDDO-Me, an oleanolic acid derivative with anti-inflammatory activity, as a potent inhibitor of NSCLC cancer cells harboring the EGFR-T790M mutation. Cellular thermal shift assay and molecular dynamic simulation validated the selective binding of CDDO-Me to T790M-mutant EGFR. Further experimental results revealed that CDDO-Me induced cellular apoptosis and caused cell cycle arrest through inhibiting the PI3K-Akt-mTOR axis by directly targeting EGFR protein, cross-validated by sgEGFR silencing in H1975 cells. Additionally, CDDO-Me could dose-depended suppress the tumor growth in a H1975 xenograft mouse model.

†Rui Zhou, Ziqian Liu and Tongtong Wu contributed equally to this work.

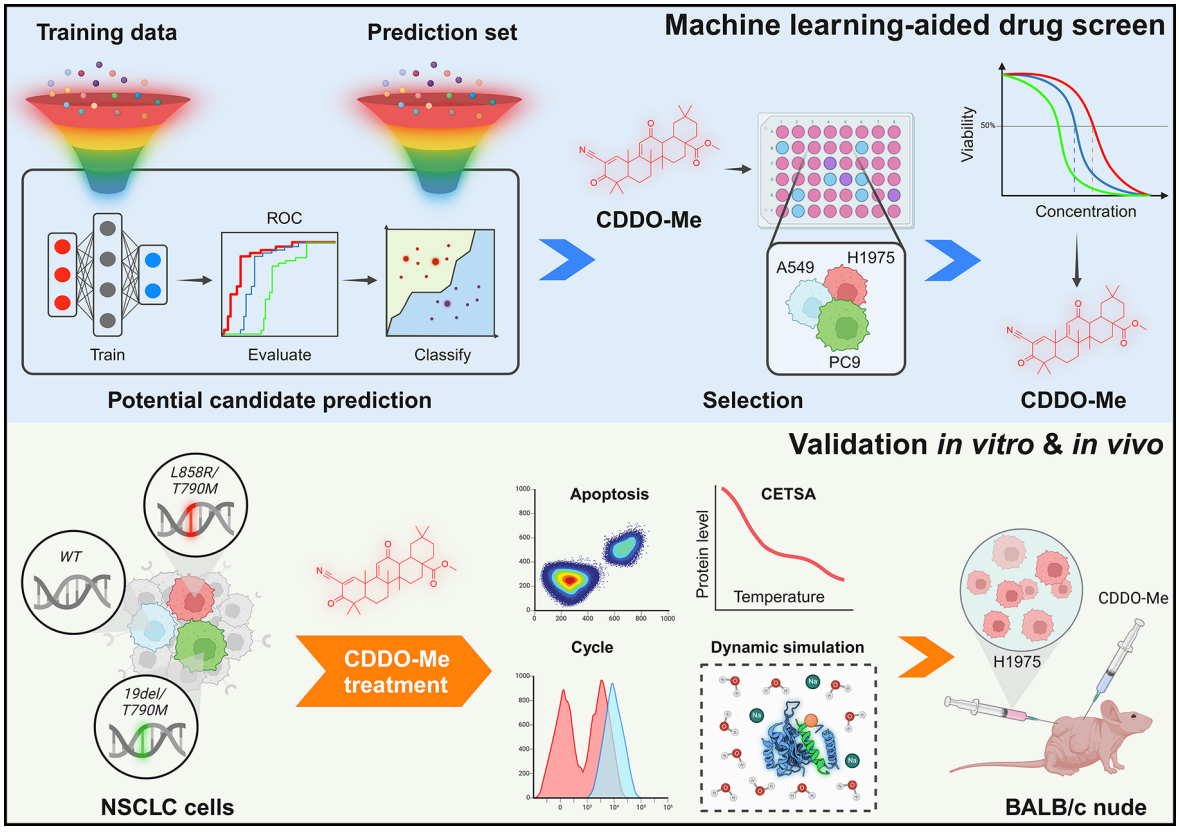
†Haigang Wu and Ning Liu are joint senior author.

*Correspondence:
Beier Jiang
674358923@qq.com
Zhenzhen Zhang
zz_jane@163.com
Ning Liu
nliu@shou.edu.cn

Full list of author information is available at the end of the article

Conclusion CDDO-Me induced apoptosis and caused cell cycle arrest by inhibiting the PI3K-Akt-mTOR pathway, directly targeting the EGFR protein. In vivo studies in a H1975 xenograft mouse model demonstrated dose-dependent suppression of tumor growth. Our work highlights the application of machine learning-aided drug screening and provides a promising lead compound to conquer the drug resistance of NSCLC.

Graphical Abstract



Keywords Non-small cell lung cancer, Epidermal growth factor receptor, T790M mutation, Machine learning-aided drug screening, CDDO-Me

Background

Drug resistance in non-small-cell lung cancer (NSCLC) significantly contributes to serious disease progression and patient mortality [1]. As a vital second-line therapy, Erlotinib, an epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI), is extensively used to manage disease progression in NSCLC patients who have received first-line chemotherapy, thereby improving therapeutic outcomes [2, 3]. The drug interacts with the EGFR catalytic, or ATP-binding, pocket in both its active and inactive states [4]. The ATP-binding pocket of EGFR is a critical site for inhibitor binding, making it a primary target in drug design efforts for NSCLC [5]. Inhibition at this site disrupts EGFR signaling, which is pivotal in the proliferation and survival of cancer cells [6]. This approach forms the basis for using Erlotinib and other TKIs as effective treatments against NSCLC, addressing both active and inactive states of EGFR to prevent escape mechanisms that lead to drug resistance EGFR mutation mediates resistance to EGFR tyrosine kinase inhibitors in NSCLC: From molecular mechanisms to clinical research [7, 8]. However, prolonged use of Erlotinib can lead to reduced binding affinity due to the emergence of the EGFR T790M mutation, resulting in approximately 60% of NSCLC patients developing resistance to both first- and second-line EGFR TKIs [9]. Consequently, the development of new generation EGFR TKIs to counteract T790M mutation-driven resistance has become a critical focus in NSCLC treatment.

To date, discovering novel chemical entities against EGFR TKIs resistance has proven to be greatly challenging through traditional drug screening strategies, including scaffold hopping derived from previous TKIs, and ab initio molecule synthesis [10, 11]. Most EGFR TKIs

used in clinical practice are prepared from existing scaffolds [12], and drug development based on these existing scaffolds may meet clinical needs in a short time. Most of these emerging TKIs with long-time drug administration often faced similar drug-resistant mechanisms [13]. Therefore, new inhibitors of EGFR with T790M mutation should be developed based on new chemical scaffolds, and potential clinical applications rarely limited by the pre-existing drug resistance.

Machine learning has been one of the most promising method that we could utilize biological activity molecule datasets to predict the activities of given molecules, which accelerated the process of identifying potential small molecular inhibitors. In the past few decades, various machine learning algorithms have been widely applied in drug design and screening and presented great potential for conquering drug resistance in precise oncology medicine [14]. For example, Gary Liu et al. [15]. employed a message-passing deep neural network to identify new antibiotics against *A. baumannii* with pan-drug resistance. Authors claimed that the Abaucin can suppress lipoprotein trafficking in *A. baumannii* to alleviate bacterial infection in a wound model. Brent M. Kuenzi et al. [16]. employed an interpretable deep learning model (DrugCell) to predict the drug response of 1235 tumor cell lines to 684 drugs. Henry Gerdes et al. [17]. employed a machine learning algorithm incorporating omics data to predict the anti-proliferative efficacy of more than 400 drugs in cancer cells. However, despite the introduction of numerous emerging strategies aimed at developing more feasible and powerful algorithms for drug discovery, bias screened without biological activity can still be observed in any algorithm [18]. Therefore, it is imperative to address bias in drug screening and enhance the accuracy of identifying target molecules from given datasets.

Beyond screening new chemical scaffolds by exploring the full chemical spaces, molecular scaffolds derived from existing bioactive molecules, for example, ocean-sourced natural products and food-originated bio-active molecules, provide the great potential to accelerate the translational processes, due to the long-term application in food or medicinal applications. Such new scaffolds of kinase inhibitors can benefit from food-resourced chemotypes for the following reasons: (a) Long-time molecular evaluation indicating the higher biological safety compared with ab initio artificial synthesized chemical entities [19]; (b) diversity of natural products from an aforementioned source providing a broader range of options for identifying new scaffolds and mitigating current drug resistance.

An opportunity exists to conduct machine learning-aided in silico screening to discover structurally and functionally emerging inhibitors that selectively target EGFR with L858R and T790M mutation and provide a potential approach to overcome first- or second-generation TKIs resistance. The EGFR L858R mutation, a key driver in NSCLC, confers sensitivity to first-generation EGFR-TKIs but is often followed by the emergence of the T790M resistance mutation, highlighting the need for next-generation inhibitors capable of overcoming both mutations [20]. In this study, we proposed an advanced and integrated platform based on structure-based drug screening to rapidly identify selective inhibitors of $EGFR^{L858R/T790M}$ and explored the selective cytotoxicity among various NSCLC cell lines (H1975, PC9-ER, and A549). Through this method, we successfully identified CDDO-Me, an anti-inflammatory compound against NF-κB. Cellular thermal shift assay (CETSA) and molecular dynamics simulation were employed to determine whether CDDO-Me can specifically and selectively bind with T790M mutant EGFR protein. Further flow cytometry and immunoblot assays were conducted to explore the impacts of CDDO-Me on apoptosis, cell cycle, caspase protein-depended cell apoptosis, and PI3K/Akt/mTOR pathway, a canonical EGFR downstream signaling pathway. *EGFR* knockout in the H1975 cell line was performed to determine whether CDDO-Me activated the apoptosis pathway without directly targeting the T790M mutant EGFR. Finally, CDDO-Me was employed to examine the dose-manner anti-cancer effect in the H1975 xenograft animal model and elicit critical biomarkers in vivo. Our findings highlight the application of machine learning-aided in silico screening in drug discovery and describe a promising lead compound with specific activity against Erlotinib resistance in NSCLC.

Methods

Datasets and chemical descriptors preparation

The first dataset was retrieved from ChEMBL (https://www.ebi.ac.uk/chembl/), DrugBank (https://go.drugbank.com/), and PubChem BioAssay (https://pubchem.ncbi.nlm.nih.gov/) databases. Composed of 11 FDA-approved drugs classified as "active", and 411 compounds demonstrated specific IC_{50} activity targeting $EGFR^{L858R/T790M}$ classified as "active" or "inactive" based on the description of activities in specific scientific literature. The second dataset was obtained from the DUD-E database with 407 decoy compounds, classified as "inactive" [21]. Then, we minimized molecular energy in MM2 forcefield (1,000 iterations with an RMS gradient of 0.01 kcal/mol/Å) for subsequent batch docking screening through Chem3D 14.0.0.17 software. After pretreatments, all 829 compounds were randomly divided into training set and test set with a ratio of 7:3.

A total of 70,413 and 452 compounds were collected from the FooDB and CSNpharm (https://www.csnpharm.cn/, 2023.6) databases respectively. Among them,

452 compounds are pharmaceutical-nutraceutical compounds with both medicinal and dietary functions. Before applying machine learning prediction, we performed preprocessing on the compound library by established protocols. The preprocessing steps included the following: (1) Removal of structurally redundant compounds. (2) Elimination of long-chain alkenes, chloroform, potassium carbonate, potassium hydroxide, potassium nitrate, sodium bicarbonate, potassium permanganate, potassium bicarbonate, and sodium carbonate. (3) Exclusion of lactates, long-chain fatty acids, and vitamins. Finally, the library was utilized for virtual screening to predict their potential binding efficiency with $EGFR^{L858R/T790M}$.

To obtain molecular descriptors of molecules in the selected library for following virtual screening, the Rdkit (https://anaconda.org/rdkit/rdkit, 2020.6) and Deepchem (https://github.com/deepchem, 2020.6) package was used to convert smiles of compounds to their structure and generated the ECFPs fingerprints as the chemical 2D descriptors. Specifically, ECFP4 fingerprints with a 1024-bit length were used, with a radius of 2, which corresponds to the maximum bond distance considered when generating the fingerprint. This radius determines the depth of the atom neighborhood explored around each atom in the molecule, impacting the granularity of the fingerprint and its ability to differentiate between different molecular structures. The RDKit and DeepChem packages were employed with their default settings, ensuring consistency and reproducibility in the generation of the fingerprints.

Ligand-based virtual screening

Quantitative structure-activity relationship (QSAR)-based machine learning classifiers including support vector machine (SVM), Naive Bayes (NB), *k*-nearest neighbor (*k*NN), XGBoost (XGB), and multilayer perceptron (MLP) were established to screen the potential ligands based on the molecular descriptors. Cross-validation is a technique widely used in machine learning for evaluating model performance. It involves partitioning the data into multiple subsets, training the model on different subsets, and evaluating its performance on a separate validation set. In this study, a stratified 10-fold cross-validation method was used on the training set to overcome the model overfitting problem.

SVM

The Scikit-learn python package was employed to implement the SVM algorithm (https://github.com/scikit-learn/scikit-learn, 2023.6), and ECFP4 fingerprints of the training set were used as the SVM input and the active information was used as labels (0 or 1). The setting for the decision_function_shape was 'one-versus-rest' and the probabilistic estimation was put into use. Other parameters were set as default.

MLP

We used the DeepPurpose package to implement the multilayer perceptron algorithm (https://github.com/kexinhuang12345/DeepPurpose, To date Jun 2023). The CompoundPred module of DeepPurpose encoded compounds SMILES with an embedding layer, then fed into a multilayer perceptron decoder to output and predict the probability of activity. The filter length was 4, 6, and 8. In addition, the model with three hidden layers was used to have 1024, 1024, and 512 units, respectively. The other hyperparameters used were as follows: learning rate (0.01), train epoch (100), and batch size (128).

XGBoost

To implement the XGB algorithm, we used the XGBoost package (https://github.com/dmlc/xgboost, 2023.6). The ECFP4 fingerprints of the training set were used as the XGB input and the active information was used as labels (0 or 1). The max_depth of XGB was 4, the learning rate was 0.01, and the objective was 'binary: logistic'. Other parameters were set as default.

Naive bayes

The Scikit-learn python package was employed to implement the Naive Bayes algorithm (https://github.com/scikit-learn/scikit-learn, 2023.6). ECFP4 fingerprints of the training set were used as input and the active information was used as labels (0 or 1). The setting for parameters of the BernoulliNB module was the default.

KNN

The Scikit-learn python package was employed to implement the *k*NN algorithm (https://github.com/scikit-learn/scikit-learn, 2023.6). Molecular descriptors mentioned above of the training set were used as input and the active information was used as labels (0 or 1). The setting for parameters was the default.

Bayesian integration of multiple machine learning models

To enhance the predictive accuracy and robustness of our model, we employed a Bayesian integration approach to combine the outputs of five distinct machine learning algorithms: Support Vector Machine (SVM), Naive Bayes (NB), *k*-Nearest Neighbors (*k*-NN), Extreme Gradient Boosting (XGB), and Multi-layer Perceptron (MLP). This ensemble technique leverages the strengths of each individual model, providing a more comprehensive analysis of the molecular data represented by Simplified Molecular Input Line Entry System (SMILES) strings.

In this method, each of the five machine learning models was independently trained on the same training

dataset and then used to make predictions on the test dataset. The individual predictions from each model were then combined using a Bayesian probabilistic framework.

The Bayesian integration process involves the following steps:

(1) Prediction Generation: Each model M_i where i ranges from 1 to 5, corresponding to SVM, NB, k-NN, XGB, and MLP) produces a probabilistic prediction $P(A|M_i)$ for the activity of a molecule. These predictions represent the likelihood of the molecule being active given the model M_i.

(2) Prior Probabilities: Assign prior probabilities $P(M_i)$ to each model based on their historical performance or expert knowledge. In our case, we assumed equal priors for simplicity, $P(M_i) = 0.2$.

(3) Posterior Probabilities: Calculate the posterior probability for each model using Bayes' theorem:

$$P(M_i|A) = \frac{P(A|M_i)\cdot P(M_i)}{\sum_{j=1}^{5} P(A|M_i)\cdot P(M_j)}$$

This step updates our belief in the likelihood of each model being correct based on its predictions.

(4) Weighted Averaging: Combine the predictions of all models to obtain a final integrated prediction. The final prediction $P(A|D)$ is a weighted average of the individual predictions, where the weights are the posterior probabilities $P(M_i|A)$:

$$P(A|D) = \sum_{i=1}^{5} P(A|M_i)\cdot P(M_i|A)$$

Model evaluation

In this study, five indices were calculated for model evaluation, including the precision, recall, F1-score, MCC, and ROC curve, which were defined as follows:

$$Precision = \frac{TP}{TP+FP}$$

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

$$F1-score = 2\frac{Precision \times Recall}{Precision + Recall}$$

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

TP (true positive) means that the number of active compounds were correctly classified; TN (true negative) means that the number of inactive compounds were correctly classified; FN (false negative) means that the number of active compounds were incorrectly classified; FP (false positive) means that the number of inactive compounds were incorrectly classified. The receiver operating characteristic (ROC) curve is a performance measurement for classification problems at various threshold settings [22]. We calculated the area under the receiver operating characteristic curve (AUC) to evaluate how capable the model is of distinguishing between classes.

Drug-like filter and ADMET prediction

We used SwissADME (http://www.swissadme.ch/, 2023.6) web server to generated five independent parameters of candidates, including molecular weight (MW), MLogP, H-bond acceptors (HBA), H-bond donors (HBD), and LogS. Subsequently, we discarded the unsatisfying molecules according to Lipinski's rule of five for further screening [23]. Then, the remaining compounds were subjected to predict the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties using the ADMETlab2.0 web server tool for further selection [24].

Chemicals

Isoalantolactone (CAS 470-17-7, 98.0% purity, Cat# 12309), and Bardoxolone Methyl (CAS 218600-53-4, 99.4% purity, Cat# 12828) were purchased from CSNpharm (Chicago, USA). Daurisoline (CAS 70553-76-3, 99.7% purity, Cat# HY-N0221) and protease inhibitor cocktail (Cat# HY-K0010) were purchased from MedChemExpress (New Jersey, USA). Erlotinib (Cat# S7786) was purchased from Selleck (Selleck Chemicals, Houston, TX, United States).

Antibodies and reagents

Antibodies against the following proteins were used with source and dilution ratios indicated: EGFR (Cat# 4267, 1:1000); phospho-EGFR (Y1068) (Cat# 2234, 1:1000); PI3K (p110α) (Cat# 4249, 1:1000); phospho-PI3K p85 (Tyr458)/p55 (Tyr199) (Cat# 4228, 1:1000); Akt (Cat# 9272, 1:1000); phospho-Akt (Ser473) (Cat# 9271, 1:1000); mTOR (Cat# 2983, 1:1000); phospho-mTOR (Ser2448) (Cat# 2971, 1:1000); MAPK (p44/42) (Cat# 4695T, 1:1000); Phospho-p44/42 MAPK (Thr202/Tyr204) (Cat# 4370T, 1:1000); Cyclin D1 (Cat# 2978, 1:1000); CDK4 (Cat# 12790, 1:1000); PARP (Cat# 9542, 1:1000); cleaved-PARP (cl-PARP) (Cat# 5625, 1:1000); caspase 3 (Cat# 9662, 1:1000); cleaved-caspase 3 (cl-caspase 3) (Cat# 9661, 1:1000); Bcl-2 (Cat# 4223, 1:1000); β-actin (Cat# 3700, 1:10000), all antibodies for western blotting were purchased from Cell Signaling Technology (Beverly, MA, United States). Peroxidase-conjugated goat anti-rabbit (Cat# 12–348, 1:20000) and mouse (Cat# AP160P, 1:20000) secondary antibodies were purchased from Sigma Aldrich (St. Louis, MO, United States).

Dulbecco's Modified Eagle Medium (DMEM) (Cat# 12491015), Roswell Park Memorial Institute (RPMI)-1640 medium (Cat# 12633012), phosphate-buffered saline (PBS) washing buffer (Cat# 10010023), Fetal bovine serum (FBS) (Cat# A5669701), Trypsin-EDTA solution (Cat# 25200072), and Penicillin-Streptomycin solution (PS) (100×) (Cat# 15140122) were all purchased from Gibco (Carlsbad, CA, United States). Cell Counting Kit-8 (CCK-8) (Cat# C0038), RIPA lysis buffer (Cat# P0013B), Crystal violet (Cat# C0121), and Cell cycle assay kit (Cat# C1052) were purchased from Beyotime Biotechnology (Shanghai, China). Annexin V-Fluorescein isothiocyanate (FITC) apoptosis detection kit (Cat# 556547) was obtained from BD Biosciences (San Jose, CA, United States). Bicinchoninic acid (BCA) Protein assay kit (Cat# PA115) was acquired from TIANGEN Biotech (Beijing, China). Polyvinylidene fluoride (PVDF) membranes (Millipore, IPVH00010) were purchased from Millipore (Billerica, MA, United States).

Cell lines and cell culture

Human non-small cell lung cancer cell lines A549 and H1975 were purchased from the American Type Culture Collection (ATCC, CCL-185, CRL-5908), PC9 cell line was purchased from European Collection of Authenticated Cell Cultures (ECACC, 90071810), HEK 293T cell line was purchased from the American Type Culture Collection (ATCC, CRL-3216). All cell lines were routinely tested to confirm that they were free of Mycoplasma. H1975 and PC9 cells were cultured at 37 °C with 5% CO_2 in RPMI 1640 and A549 cell lines were cultured in DMEM. All culture media were supplemented with 10% heat-inactivated FBS with 100 U/ml penicillin and 100 μg/ml streptomycin. To establish an erlotinib-resistant PC9 subline, the cells were treated with 0.01 μM of erlotinib and the concentration was increased in a stepwise manner. After two months the cells were able to grow in 1 μM of erlotinib, then continuously subcultured with 1 to 2 μM of erlotinib for an additional 6 months and did a single-cell cloning to establish the erlotinib-resistant cell line.

Cell viability assay

A549, H1975, PC9, PC9-ER and HEK-293T cells were cultured in 96-well plates (5×10^3 cells per well). Following the treatments with compounds, cell viability was assessed with CCK-8 Kit at indicated time points. After 2 h of incubation at 37 °C, absorbance was determined with a Microplate Reader (BIO-TEK, Inc., Winooski, VT, United States) at 450 nm. The IC_{50} value was calculated by GraphPad Prism 8.0 (San Diego, CA, United States) software.

Colony formation assay

Colony formation assay was used to examine the long-term effects of CDDO-Me on NSCLC cell growth. A549, H1975, and PC9-ER cells were cultured into 6-well plates (1×10^3 cells per well). Following 24 h of treatment with CDDO-Me (0, 0.2, 0.4, and 0.6 μM), the culture medium was replaced with a fresh medium, and the growth continued for 10–14 days. After culture completion, the colonies were washed with cold PBS, then fixed with 4% paraformaldehyde and stained with 0.5% crystal violet for 15 min at room temperature, Macroscopic colonies of each well were counted by EVOS XL Core Imaging System (Thermo Fisher Scientific Inc., Waltham, MA, United States).

Western blot analysis

Western blot analysis was performed as in our previous reports [25]. For the analysis by sodium salt-polyacrylamide gel electrophoresis (SDS-PAGE), total proteins were obtained from the cells or tumor tissue homogenates with RIPA buffer containing a complete protease inhibitor cocktail. The protein supernatant was collected after centrifugation at 4 °C and measured with the BCA assay kit. Equal amounts of total proteins were resuspended in loading buffer, boiled at 100 °C for 5 min, and separated by 10–15% SDS-PAGE. Proteins were transferred onto PVDF membranes. After blocking the PVDF membrane with 5% non-fat dry milk, incubation with primary antibodies overnight at 4 °C. The membranes were washed three times in TBST followed by 1 h incubation with secondary antibodies, conjugated with horseradish peroxidase (HRP), at room temperature. Next, the membranes were washed before enhanced chemiluminescence and visualized with the Bio-Rad Chemi-Doc XRS system (Bio-Rad, CA, USA) and quantified by ImageJ 1.54 software.

Apoptosis and cell cycle assay

To examine the effect of the compound on apoptosis and cell cycle, NSCLC cells were seeded in 6-well plates at a density of 2×10^5 cells per well and were treated with different doses of CDDO-Me (0, 0.2, 0.3, 0.4, and 0.6 μM) for 48 h. Cell apoptosis cell cycle were detected by FACSCelesta flow cytometer (BD Biosciences, Franklin Lakes, NJ, United States) using Annexin V-FITC Apoptosis Detection and PI Cell Cycle and Apoptosis Analysis Kit respectively. Apoptosis data were analyzed by FlowJo v10.8.1 software, and cell cycle data were analyzed with Modfit LT 4.1 software.

Cellular thermal shift assay (CETSA)

To validate CDDO-Me and target interaction, we expand H1975 cells in 6-well plates with cell culture medium to a cell density of 2×10^5 cells per well, incubated with

DMSO and CDDO-Me (2μM) for 1 h and then extract total protein with RIPA buffer. The protein suspension was separated into six PCR tubes and heated individually at designated temperatures (42–57 °C) for 10 min. Finally analyzed by western blotting. The CETSA curve was visualized using GraphPad Prism 8.0.

Protein expression and purification

The coding sequences for the intracellular kinase domains of human $EGFR^{WT}$ and $EGFR^{L858R/T790M}$ (residues 695–1022) were cloned into pBAD/His A vectors with N-terminal 6xHis tags, synthesized by Tsingke Biotech Co., Ltd (Beijing). Recombinant plasmids were transformed into *E. coli* BL21 (DE3) pLysS cells, and initial cultures were grown in LB medium with 50 μg/ml kanamycin at 37 °C, 180 rpm, overnight. Each culture (4 ml) was scaled up to 400 ml in LB with 50 μg/ml kanamycin and grown to OD600 ≈ 0.6 at 37 °C, then induced with 0.1 mM IPTG for 3 h at 37 °C, 180 rpm. Cells were harvested by centrifugation (4000 rpm, 4 °C, 5 min), resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 1 mM PMSF, 1 mM EDTA, 0.5% Triton X-100, 5% glycerol), and sonicated on ice (50% power, 10 min). Recombinant proteins were confirmed by 12% SDS-PAGE. Lysates were clarified by centrifugation (23,000 rpm, 4 °C, 20 min), and supernatants were filtered (0.22 μm) for purification. Purification was conducted on a HisTrap HP column (Cytiva, Cat# 17524701), equilibrated with buffer A (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM imidazole, 10% glycerol) and eluted with buffer B (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 300 mM imidazole, 10% glycerol) using a gradient (10–30% buffer B). Proteins were loaded at 0.5 ml/min following equilibration with 5 column volumes of buffer A. Concentrations were measured by absorbance at 280 nm.

ADP-Glo biochemical assay

The IC_{50} values of CDDO-Me for $EGFR^{L858R/T790M}$ and $EGFR^{WT}$ were determined using the ADP-Glo™ kinase assay in a 10-dose, serial CDDO-Me dilution format, with concentrations ranging from 0.001 nM to 10,000 nM. The kinase reaction was initiated by incubating proteins with increasing concentrations of CDDO-Me in 1x kinase reaction buffer on ice for 5 min. ATP (final concentration 10 μM) were then added to the reaction mixture and incubated at room temperature for 60 min. Afterward, 5 μl of ADP-Glo™ reagent was added to the reaction mixture, followed by a 40-min incubation at room temperature. Subsequently, 10 μl of kinase detection reagent was added and incubated in the dark at room temperature for 30 min. Luminescence was measured to quantify ADP production. A control reaction was performed with 8% DMSO in the absence or presence of kinase enzymes to serve as negative and positive controls, respectively.

Microscale thermophoresis (MST) analysis

MST analysis was performed using the NanoTemper Monolith NT.115 instrument (NanoTemper Technologies). Monolith Protein Labeling Kit Red-NHS 2nd Generation (Cat# MO-L011) dye was used to label EGFR-WT and EGFR-L858R/T790M. Briefly, 90 μl of protein (10 μM) was mixed with 10 μl of 300 μM Red dye and incubated for 30 min at room temperature. CDDO-Me were diluted in PBS-T buffer to make 16 1:1 serial dilutions from 50 μM. Measuring the affinity of EGFR with drug: 100 nM of labeled EGFR was mixed with drug prepared in 16 different serial concentrations at RT in PBS-T. The mixtures were then loaded into standard glass capillaries (Monolith Capillaries, Cat# MO-K022). After blowing evenly, the machine was tested and the Initial Fluorescence Analysis program (LED 20%, MST power 40%) was used for Analysis through the MO. The K_d values were derived from the concentration-dependent changes in normalized fluorescence (Fnorm). Data were analyzed using MO.Control v2.6.3 software.

Molecular docking and molecular dynamics simulation

Molecular docking was performed by AutoDock v.1.2.4 and AutoDock vina v.1.2.0 [26, 27]. The crystal structure of human $EGFR^{L858R/T790M}$ (PDB ID: 3W2R) and $EGFR^{WT}$ (PDB ID: 7B85) was retrieved from RCSB Protein Data Bank (http://www.pdb.org). The 3D structures of the selected candidates were generated by using Chem3D v.19.0 and utilized MM2 force field for energy minimization. All missing terminal residues of protein structures were repaired by Swiss-PdbViewer v.4.10 software [28], then prepared by removing crystallographic waters, adding polar hydrogen and Kollman charges for docking study. The grid box center was set as coordinates of x, y, z = 1.52, 12.084, 51.751, and the grid size was 80 Å x 98 Å x 76 Å. The other parameters were set as default. The binding interaction of the protein-ligands complex has been observed by using UCSF Chimera v.1.16 and BIOVIA Discovery Studio Visualizer v.21.1.0.20298. After the batch docking-based screening, we selected three commercially available compounds among the top 20% of docking scores for further experimental validation in vitro and molecular dynamics simulation (MD) analysis. The lowest energy conformation was selected for MD simulation analysis. All MD simulations were performed using the GROMACS (2020.6). The topology for the protein-ligand complex was prepared using Amber99 and GAFF force fields respectively [29, 30]. All molecules were solvated in a water environment using the TIP3P water model, with a total of 20,000 water molecules added to the simulation box. To neutralize the system's net charge, 5 Na^+ ions were added, ensuring an electrically neutral environment for the molecular dynamics simulation. The periodic box of simulation

was 50×50×50 $Å^3$ Equilibration was conducted in an NVT ensemble for 100 ps at 300 K, and subsequently in the NPT ensemble for 100 ps at 1.0 bar using the Nose-Hoover thermostat and Parrinello-Rahman barostat. The protein-ligand complex performed the energy minimization to ensure no steric clashes in the system. Before the 200 ns MD simulation, we performed NVT and NPT equilibration. The MD simulation of the protein-ligand complex was performed at 200 ns with a time step of 2 fs. Finally, we analyzed the MD simulation results.

RMSD and RMSF analyses were conducted to evaluate the stability and flexibility of the protein-ligand complex throughout the MD simulation. The RMSD of the Cα atoms of the protein and the ligand was calculated using the GROMACS tool 'gmx rms', which compares the atomic positions at each time step to the initial structure to monitor the overall conformational changes. The RMSF was calculated using 'gmx rmsf' to assess the residue-level flexibility of the protein, focusing on the fluctuations of individual residues around their average positions throughout the simulation.

Hydrogen bonds between the protein and ligand were analyzed using the GROMACS tool 'gmx hbond'. We calculated the number and occupancy of hydrogen bonds formed during the simulation, considering a donor-acceptor distance of 3.5 Å and a hydrogen-donor-acceptor angle of 30° as the criteria for hydrogen bond formation.

Binding free energies was performed using the GROMACS tool gmx_MMPBSA, which integrates molecular mechanics energies with solvation terms to estimate the binding affinity. The binding free energy (ΔG_bind) was decomposed into van der Waals (ΔG_vdW), electrostatic (ΔG_ele), polar solvation (ΔG_pol), and non-polar solvation (ΔG_nonpol) contributions. The MM/PBSA calculation was performed on 1000 snapshots extracted evenly from the 200 ns trajectory to ensure an accurate and representative evaluation of the binding affinity. The dielectric constants used for the solute and solvent were set to 2 and 80, respectively. Then, we extracted the conformations of small molecules bound to $EGFR^{L858R/T790M}$ and $EGFR^{WT}$ at 130 ns. The typical time to calculate a 2D and 3D free energy landscape using MM/PBSA model for 200 ns for protein.

Generation of CRISPR-Cas9 KO cell lines

To knockout the *EGFR* gene in H1975 cell line, we designed two independent guide RNAs (gRNAs) using the CRISPick (https://portals.broadinstitute.org/gppx/crispick/public). The gRNA sequences were cloned into lentiCRISPR v2-Blast plasmid (Addgene #83480). Sequences of the guide RNAs are listed in Table S2.

To produce lentivirus, HEK 293T cells were plated in 100 mm dish one day before transfection at 70% confluency. Two hours before transfection, DMEM media was replaced with 10 ml serum-free OptiMEM media. For each dish, 11 μg of lentiviral transfer vector, 8.25 μg psPAX2 (Addgene Cat# 12260) and 5.5 μg pMD2.G (Addgene Cat# 12259) diluted in 1 ml Opti-MEM were combined with 15 μl Lipofectamine 3000 diluted in 1 ml Opti-MEM. The transfection mixture was left for 20 min and then added dropwise to the cells. 6 h after transfection, the media was replaced by 10 ml fresh complete DMEM media for each dish. Virus-containing media was collected 72 h post-transfection followed by Filtration was then performed with a 0.45 μm (Millipore, #SLGPR33RB). Viral supernatants were concentrated by centrifugation at 4,000 rpm and 4 °C for 35 min. Concentrated viral supernatants were stored in aliquots at -80 °C.

For sgRNA lentivirus infection, target cells were seeded in a 6-well plate 24 h before infection and were grown to 60–80% confluency upon transduction. Culture medium was removed, and cells were incubated with virus supernatant along with 8 mg/ml polybrene (Sigma Cat# TR-1003) overnight. Virus-containing medium was replaced with fresh medium. 10 mg/ml Blasticidin S (Beyotime Cat# ST018) was applied to kill non-infected cells. After 7 days, cells were collected for immunoblotting or other experiments.

Xenograft tumour assay

BALB/c nude mice (male, 5–6 weeks old, 10–14 g), were purchased from Shanghai Model Organisms Center, Inc (Shanghai, China). All mice (n=18) were seeded to adopt the environment under specific pathogen-free (SPF) conditions with 25 °C and 65% relative humidity for 2 weeks, and environment was kept with a 12 h light/dark cycle.

H1975 cells ($5×10^6$ cells) suspended in 200 μl PBS were subcutaneously implanted into the right flank of per nude mice when 6–7 weeks old. Tumor volume (V) was measured with digital calipers and calculated using the following formula: $Width^2 \times Length \times 0.5$. Mice were randomized into three dependent groups (n=6 mice per group): vehicle (2% DMSO in PBS), CDDO-Me (3 mg/kg/d, *i.p.*), and CDDO-Me (6 mg/kg/d, *i.p.*). Mice were treated for 20 days and the tumor volume and body weight were measured after every 2 days. After 20 days, the mice were euthanized, and the xenograft tumors were resected to take bright field images and further histological examinations. The protocols for animal care and euthanasia were approved by the Institutional Animal Care and Use Committee of Shanghai Ocean University (Shanghai, China).

Immunohistochemistry

Immunohistochemical staining was performed by Shanghai RecordBio Co.Ltd. (Shanghai, China). Tumor sections were immuno-stained with specific anti-phospho EGF

Receptor, and anti-Ki-67 antibodies. The images were captured using a Wisleap WS-10 scanner and analyzed by using NDP.view 2.3.

Statistical analyses

All the experimental data were analyzed by GraphPad Prism 8.0 (GraphPad Software Inc., San Diego, CA, United States). Results were presented as mean ± standard deviation (S.D.), and all the biological assays were conducted with $n=3$ independent replicates unless specified mentioned in the legends. Statistics analysis was conducted using one-way ANOVA with Tukey's multiple-comparisons test was applied: NS no significance, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

Results

Machine learning-guided discovery of EGFR mutation inhibitors

Recently, advances in machine learning-aided drug discovery have highlighted the advantages in the discovery of selective disease-associated kinase inhibitors [31]. Here, we applied an integrated machine learning-aided virtual screening approach (Supplementary Fig. S1) to identify new natural products selectively against T790M-mutant EGFR-kinase activities. The applied screening protocol can be divided into different sections: (1) integrated machine learning to identify potential molecules selectively binding to EGFR-T790M; (2) druggability analysis using ADMET and Lipinski's rule; (3) molecular docking to determine interaction between target and small molecules; (4) selected molecules purchased for further biological activity screening.

To obtain the un-bias predictive performance using machine learning, molecules in the testing dataset should be broadly collected from different databases to ensure the chemical diversity. Here, we initially screened a diverse collection of 829 compounds curated from DrugBank, PubChem BioAssay, and ChEMBL databases for classifier training and testing (As given in Fig. 1). Among these, 11 FDA-approved drugs and 235 compounds demonstrated specific IC_{50} activity against T790M-mutant EGFR cells classified as "active". Additionally, 176 compounds exhibited specific activities from public literature and 407 decoy compounds from the DUD-E database were classified as "inactive". All the molecular information was transformed into Extended-Connectivity Fingerprints (ECFPs) format, which were widely acknowledged as highly effective approaches for characterizing features of small molecules in the field of chemical analysis [32].

Next, these datasets were employed to train a binary classifier and predict whether new molecules presented the selective suppression of $EGFR^{T790M}$ kinase activity.

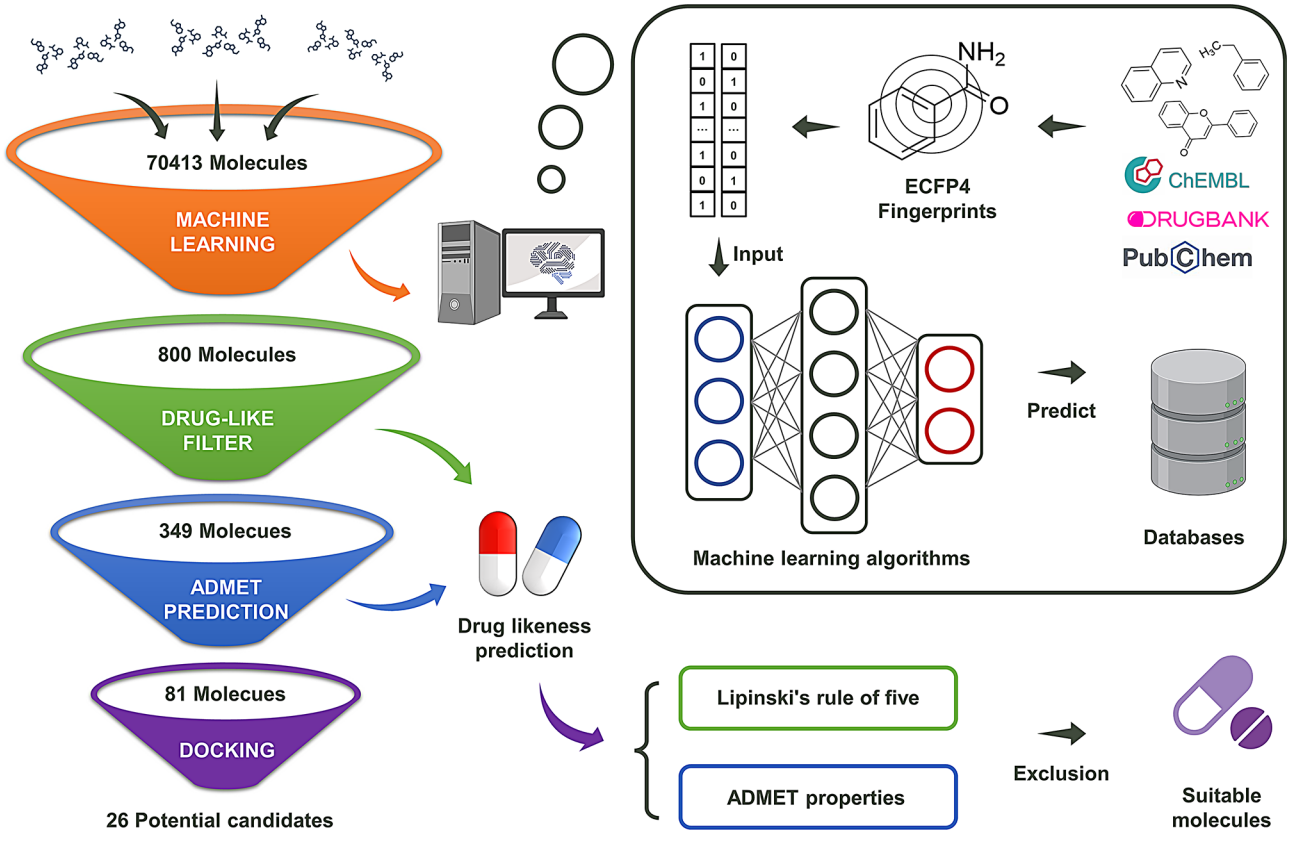


Fig. 1 The workflow of in silico drug screening

The individual algorithm often faces bias in predicting outcome results, which were determined by many factors, including selection of inner probability mass function et al. [33]. In order to exclude the bias originating from selection of individual algorithms, five machine learning algorithms, including Support vector Machine (SVM) [34], Naive Bayes (NB) [35], *k*-Nearest Neighbors (*k*-NN) [36], Extreme gradient boosting (XGB) [37],

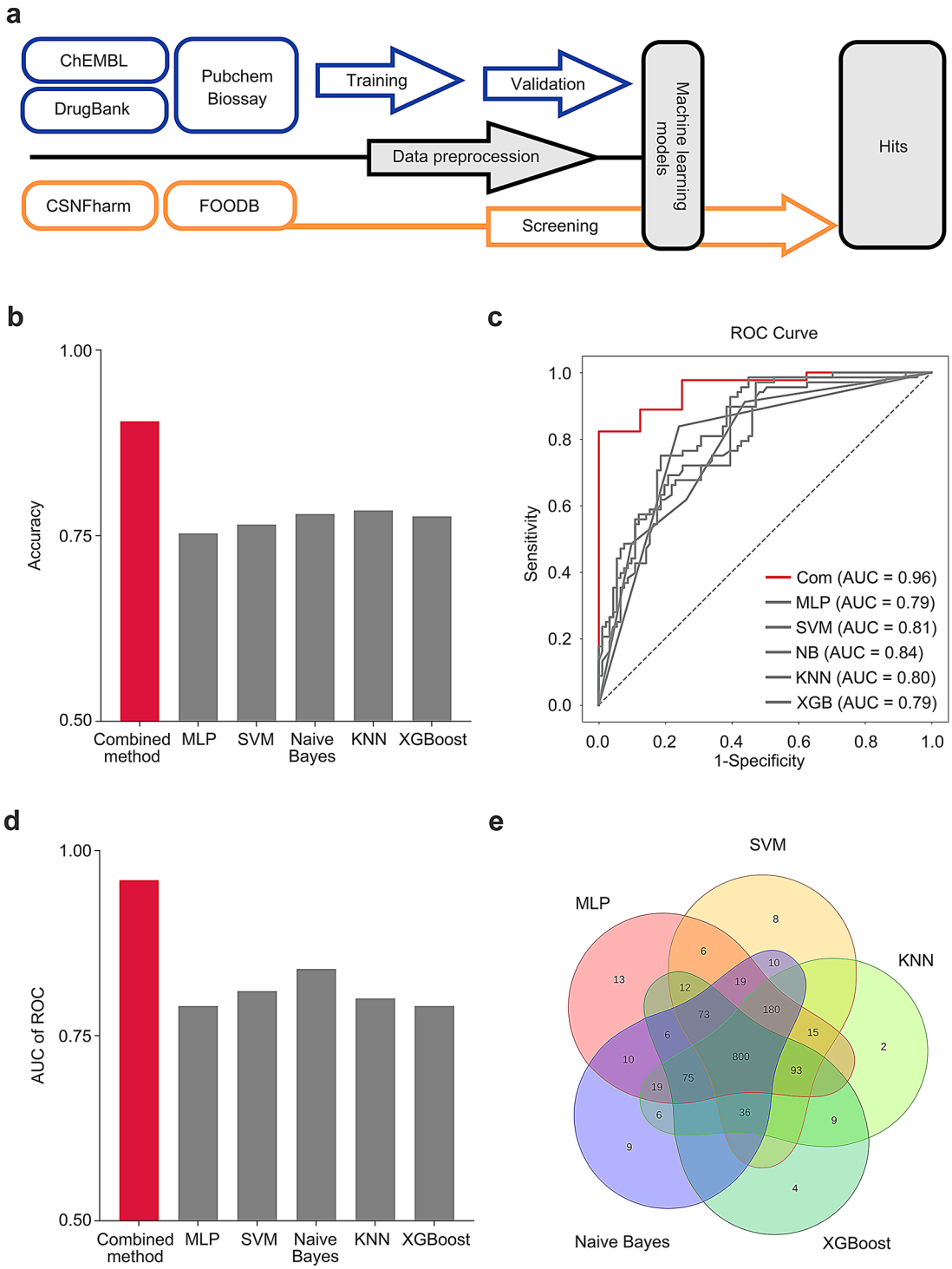


Fig. 2 Integrated machine learning approach demonstrates robust performance. (**a**) Virtual screening workflow. (**b**) The prediction accuracy of ensemble and individual models. (**c**) Receiver Operating Characteristic (ROC) curves for ensemble and individual models. (**d**) AUC values based on ensemble and individual models. (**e**) Venn diagram illustrating the virtual screening results obtained from five integrated machine learning models

and Multi-layer perceptron (MLP) [38] were employed to construct models (as given in Fig. 2a and Supplementary Fig. S1), which were integrated to an emerging and comprehensive algorithm using Bayesian inference and excluded the potential bias. After training unsupervised learning algorithm, this integrated model exhibited superior accuracy compared to the individual models (Table 1; Fig. 2b). Concurrently, the integrated machine learning method attained higher values for the area under the ROC curve (AUC) compared to individual models (Fig. 2c, d).

Post the model training, this model was employed to identify selective inhibitors from the molecular dataset, including 70,865 molecules. These chemical libraries were obtained from the FooDB database and CSN pharm database and assured the diversities of pharmacophores and chemical structures. Compared with screening results conducted by individual algorithms, a total of 800 small molecules were concurrently identified as active by five distinct machine-learning models, and the Venn diagram visually illustrated the overlapping and distinct sets of predicted compounds among five models (Fig. 2e). Therefore, our results suggested that these 800 small molecules presented the potential as selective $EGFR^{T790M}$ inhibitors.

Drug-like properties and ADMET filter

The crucial impact of poor pharmacokinetics and toxicity displayed the critical roles in late-stage of drug development, and it should necessitate prior prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties for meticulously selected drug candidates [39]. Owing to the algorithm, the ADMET features of candidates were not considered during the machine learning-aided screening process. To avoid such issues, an initial screening of selected 800 compounds was conducted to exclude the compounds against Lipinski's rule of five [40]. The ADMET features of left 349 compounds were subsequently explored using the ADMETlab 2.0 web server [24]. Using the crucial criteria based on AMES toxicity, intestinal absorption, water solubility, and BBB penetration, only 81 compounds were successfully obtained for further screening.

Structure-based drug virtual screening using molecular docking

To evaluate the interactions between the candidate and EGFR, 81 compounds were employed to conduct molecular docking using Autodock and predict the binding model between the $EGFR^{L858R/T790M}$ catalytic domain, which contained the phosphate-binding loop (P-loop) and DFG tripeptide motif, and candidate compound. As shown in Table 2, we selected the top-ranked 26 compounds, in which the binding energy was less than −7 kcal/mol. Additionally, considering the off-targeted effects caused by low molecular weight and water solubility [41, 42], (Table 2), 3 commercially available compounds, CDDO-Me (Fig. 3a), Daurisoline, and Isoalantolactone, were purchased for subsequent biological evaluation. In the previous reports, these compounds displayed excellent anti-proliferation activities [43–45] in different cancers, however, few literature reported the association between biological activity and EGFR status.

CDDO-Me treatment selectively suppressed the proliferation and migration of EGFR-mutant NSCLC cells

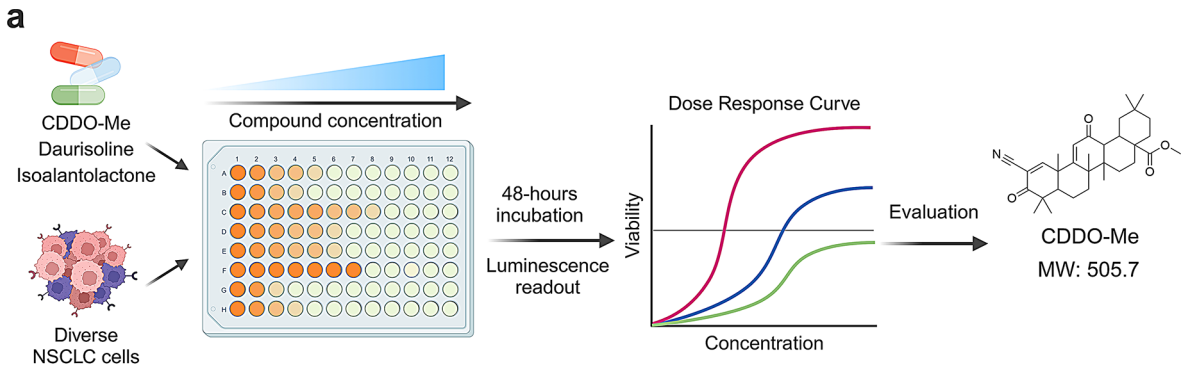
To examine whether these candidates could selectively suppress activities of the EGFR mutant type compared with wild types, A549 ($EGFR^{WT}$), and H1975 ($EGFR^{L858R/T790M}$) were employed to preliminarily assess the dosage-dependent anti-proliferation of these compounds between A549 and H1975. As given in Fig. 3b and Supplementary Fig. S2a-b, CCK-8 results corroborated that CDDO-Me can selectively suppress the anti-proliferation of H1975 cells (0.40±0.03 μM) compared with A549 (2.81±0.30 μM), while Daurisoline, and Isoalantolactone did (Supplementary Fig. S2a-b) not display the selectivity between H1975 and A549. Subsequently,

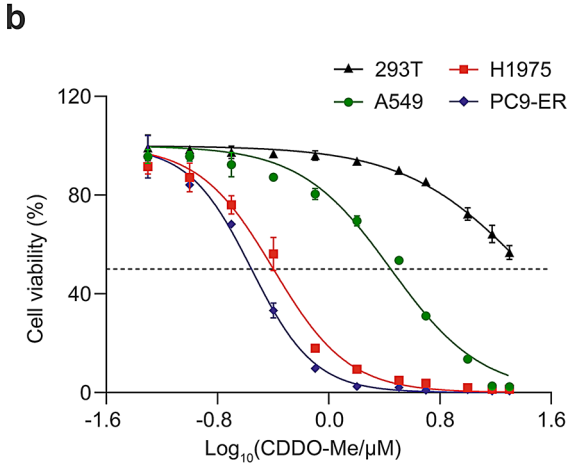
Table 1 Performance of validation results on test set for esemble and individual models

Model		Precision	Recall	F1-score	MSE	MCC	Support
SVM	Inactive	0.63	0.73	0.67	0.20	0.44	180
	Active	0.71	0.43	0.54			69
Naive Bayes	Inactive	0.78	0.64	0.64	0.22	0.49	180
	Active	0.58	0.61	0.64			69
KNN	Inactive	0.72	0.62	0.72	0.22	0.51	180
	Active	0.79	0.68	0.79			69
XGBoost	Inactive	0.73	0.69	0.71	0.22	0.50	180
	Active	0.75	0.73	0.79			69
MLP	Inactive	0.75	0.87	0.74	0.20	0.55	180
	Active	0.81	0.62	0.71			69
Combination	Inactive	0.90	0.97	0.94	0.10	0.75	180
	Active	0.91	0.82	0.81			69

Table 2 Potential candidate predicted by esemble learning classifier

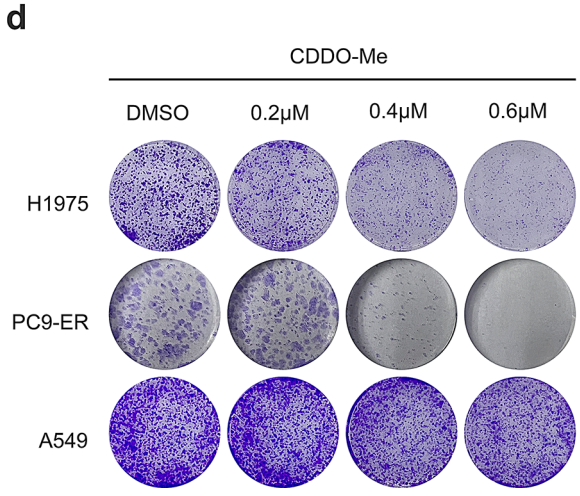
FOODB CSN ID	Name	MW	LogP	RO5	LogS	LogBB	LogKp	LogHERG	LogPPB	Caco-2 Permeability	TPSA	Docking score
CSN15660	Daurisoline	610.70	4.44	0.00	-3.05	-0.23	-5.50 cm/s	-0.06	-0.06	-4.93	83.90	-8.9
CSN12828	Bardoxolone methyl	505.70	5.39	0.00	-6.51	-0.12	-4.61 cm/s	-0.91	-0.01	-5.32	84.20	-8.7
CSN21003	Dehydroandrographolide succinate	532.60	2.70	1.00	-4.23	-0.79	-7.64 cm/s	-2.25	-0.06	-5.51	154.00	-8.5
FDB001822	Lappaol F	714.27	3.80	2.00	-2.96	-0.70	-7.47 cm/s	-0.29	-0.04	-6.12	162.60	-8.4
FDB004435	6-Keto-28-homobrassinolide	478.37	4.29	1.00	-4.44	-0.67	-5.63 cm/s	-0.91	-0.06	-4.73	97.99	-8.2
FDB000488	Ochratoxin A	403.08	4.56	1.00	-3.43	-0.90	-5.40 cm/s	-1.66	-0.02	-5.00	112.93	-8.2
FDB022684	Corticosterone	346.21	2.27	1.00	-3.07	-0.48	-7.04 cm/s	-1.10	-0.07	-4.40	74.60	-7.7
FDB007234	Zonarene	204.19	4.65	1.00	-4.61	0.67	-4.69 cm/s	-1.01	-0.01	-4.48	0.00	-7.4
FDB007717	10-Epizonarene	204.19	4.99	1.00	-4.78	0.67	-4.69 cm/s	-1.01	-0.01	-4.48	0.00	-7.4
FDB007794	cis-Sesquisabinene hydrate	222.20	4.18	1.00	-4.39	0.24	-4.76 cm/s	-0.90	-0.03	-4.31	20.23	-7.3
FDB013845	Sesquisabinene hydrate	222.20	4.55	1.00	-4.38	0.24	-4.76 cm/s	-0.90	-0.03	-4.31	20.23	-7.3
FDB019265	beta-Elemenone	218.17	3.79	2.00	-4.36	0.31	-4.49 cm/s	-1.44	-0.06	-4.20	17.07	-7.3
FDB097411	7E, 9E-Octadecadienoic acid	204.19	5.24	1.00	-5.43	0.46	-4.21 cm/s	-0.78	-0.01	-4.54	0.00	-7.3
FDB001956	Prehumulinic acid	280.17	3.45	2.00	-2.61	-0.34	-5.76 cm/s	-1.39	-0.03	-4.63	68.28	-7.3
FDB022983	N-Phenylacetylphenylalanine	283.12	2.34	1.00	-2.42	-0.65	-6.24 cm/s	-1.36	-0.05	-5.58	66.40	-7.3
FDB014737	Isoalantolactone	232.15	3.11	0.00	-3.97	-0.04	-5.29 cm/s	-1.41	-0.08	-4.57	26.30	-7.2
FDB009193	Citronellyl anthranilate	275.19	5.17	1.00	-5.45	0.18	-3.82 cm/s	-0.88	-0.01	-4.68	52.32	-7.2
FDB005326	alpha-Curcumene	174.14	4.90	1.00	-5.25	0.67	-3.71 cm/s	-0.52	-0.02	-4.50	0.00	-7.1
FDB021371	Isoleptospermone	266.15	2.25	1.00	-2.38	-0.05	-5.76 cm/s	-2.10	-0.12	-4.22	68.28	-7.1
FDB014954	1,3,5,11-Bisabolatetraen-10-one	216.15	4.11	1.00	-4.59	0.40	-4.89 cm/s	-0.66	-0.02	-4.36	17.07	-7.1
FDB016956	Vetiverol	220.18	3.58	1.00	-3.60	0.35	-5.51 cm/s	-1.10	-0.09	-4.45	20.23	-7.1
FDB029178	5′-Carboxy-gamma-chromanol	306.18	4.11	1.00	-3.54	-0.78	-5.29 cm/s	-1.17	-0.01	-5.03	66.76	-7.1
FDB013625	Acolamone	220.18	4.06	1.00	-5.16	0.34	-4.68 cm/s	-1.25	-0.05	-4.20	17.07	-7
FDB021227	Italidipyrone	542.22	4.33	1.00	-3.59	-0.93	-6.60 cm/s	-1.53	-0.01	-4.93	178.64	-7
FDB014444	Isoacolamone	220.18	4.42	1.00	-5.22	0.41	-4.84 cm/s	-1.56	-0.04	-4.27	17.07	-7
FDB016223	Linalyl phenylacetate	272.18	4.59	1.00	-4.88	0.15	-4.04 cm/s	-0.96	-0.03	-4.62	26.30	-7





c

Compounds	Cell line	IC50 for 48h (μM)	EGFR gene status
CDDO-Me	H1975	0.40 ± 0.03	L858R/T790M
	PC9-ER	0.28 ± 0.01	exon 19del
	A549	2.81 ± 0.30	Wild type
	293T	26.83 ± 3.61	Wild type



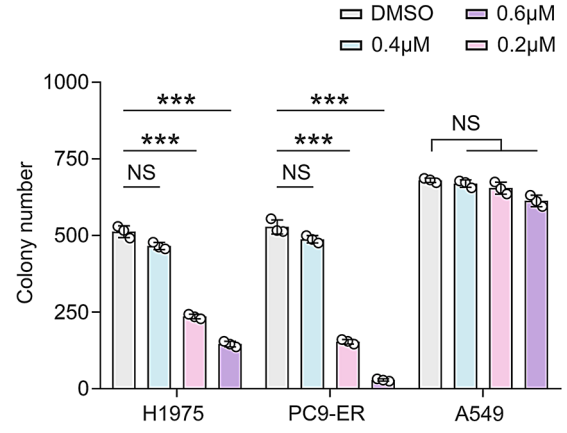


Fig. 3 The cytotoxicity of CDDO-Me on different NSCLC cells. (**a**) Schematic representation of CDDO-Me identification. (**b**) H1975, PC9-ER, A549 and HEK 293T cells were treated with CDDO-Me (0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 5, 10, 15 and 20 μM) for 48 h, respectively. Cell viability was examined by CCK-8 assay and shown as relative viability compared to the untreated control. Each test was performed in triplicate. (**c**) The IC_{50} values of CDDO-Me in H1975, PC9-ER, A549 and HEK 293T cells were presented as the mean ± SD. (**d**) Colony formation of H1975, PC9-ER, and A549 cells were measured after CDDO-Me (0, 0.2, 0.4, and 0.6 μM) treatment for 14 days, and photographs of crystal violet-stained colonies were depicted. (**e**) The statistical result of colony number after CDDO-Me treatment. Data are shown as mean ± S.D., $n = 3$. One-way ANOVA with Tukey's multiple-comparisons test, NS no significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

we further conducted a 200 ns molecular dynamics simulation of Daurisoline-EGFR$^{L858R/T790M}$/EGFR^{WT} and Isoalantolactone-EGFR$^{L858R/T790M}$/EGFR^{WT} complex, and MD results were analyzed by a root-mean-square deviation (RMSD) of the backbone atoms of EGFR and the heavy atoms of Daurisoline and Isoalantolactone. As shown in Fig. S2c-d, the EGFR-ligand complex reached equilibrium at about 130 ns. Analysis of the protein-ligand RMSD profiles from 130 to 200 ns reveals that Isoalantolactone-EGFR^{WT} (Fig. S2c) exhibits greater stability, with a ΔG_{bind} of -10.22, compared to the mutant complex, which has a ΔG_{bind} of -6.13 (Table S1). In contrast, no significant difference in RMSD or ΔG_{bind} was observed for the Daurisoline-protein complexes: ΔG_{bind} values are −15.12 for EGFR$^{L858R/T790M}$ and −14.65 for EGFR^{WT} (Fig. S2d). Additionally, we also analyzed the hydrogen bond formation after 130 ns simulation, and extracted protein-ligand structure (Supplementary Fig. S2e) did not display the hydrogen bond between small molecules and EGFR$^{L858R/T790M}$, implying the weak interaction of these small molecules with EGFR$^{L858R/T790M}$. Overall, these results suggested that the activity difference of CDDO-Me between H1975 and A549 may be associated with EGFR-T790M mutant status.

In order to validate the selective interactions between CDDO-Me and EGFR-T790M mutant, it is necessary to establish an emerging T790M-mutant cells for further examination. As the literature reports, clinical trials clearly showed that T790M mutation of EGFR in lung cancer often caused the failure of Erlotinib treatment and disease progression [46]. Erlotinib was employed to treat PC9 cells at low concentrations and established the Erlotinib-resistant PC9 cell lines (PC9-ER, Fig. S3a), and tested the proliferation of CDDO-Me on PC9-ER cells. Compared with A549 (EGFR^{WT}) and H1975 (EGFR$^{L858R/T790M}$), IC_{50} value of CDDO-Me on PC9-ER is 0.28±0.01 μM, which was about 1/1.4 and 1/10 folds on H1975 (IC_{50}=0.40±0.03 μM) and A549 (IC_{50}=2.81±0.30 μM), respectively. Additionally, we also examined the toxicity of CDDO-Me on the human normal cell 293T (Fig. 3b-c), and IC_{50} value (26.83±3.61 μM) is far higher than the lung cancer cell lines, implying the higher biocompatibility and lower cytotoxicity.

To further assess the anti-proliferative activity of CDDO-Me, concentration-dependent colony formation assays were conducted to examine the difference among A549, H1975, and PC9-ER cells. After exposure to various concentrations of CDDO-Me, the number of colony formations (Fig. 3c-d) in H1975 and PC9-ER was significantly suppressed compared with A549 cells. Additionally, after CDDO-Me treatment, the number of stained migratory H1975 and PC9-ER cells was much greater than A549 (Supplementary Fig. S3b-c). We also detected the phosphorylation level of EGFR in PC9-ER and PC9 cells after incubation with 5μM Erlotinib for 48 h, results indicated that phosphorylation level of EGFR in constructed PC9-ER cells did not altere by erlotinib treatment (Fig. S3d). Collectively, these results suggested that CDDO-Me could significantly suppress the proliferation and the directional migration of EGFR mutant NSCLC cells (H1975 and PC9-ER).

CDDO-Me is a selective inhibitor of EGFR$^{L858R/T790M}$

The above-mentioned assays implied that CDDO-Me molecule displayed the selectivity between EGFR wild type and mutant in NSCLC cells, but it cannot provide enough information about whether CDDO-Me can directly bind with EGFR^{T790M} in vitro. To further validate whether CDDO-Me can specifically bind to EGFR harboring T790M mutation, cellular thermal shift assay (CETSA) and molecular dynamic simulation were conducted to determine whether CDDO-Me directly binds with EGFR^{T790M} and affects its degradation in vitro.

CETSA can precisely examine the thermal stability of target proteins, which were docked with small molecules [47]. We used CETSA to test the cell-level interaction between CDDO-Me and EGFR$^{L858R/T790M}$, temperature-dependent immunoblot analysis of EGFR (Fig. 4a) clearly revealed that the thermal stability of EGFR protein treated with CDDO-Me was higher than control.

Further elucidating the selective binding of CDDO-Me to EGFR$^{L858R/T790M}$ is crucial, and microscale thermophoresis (MST) offers a robust approach to validate this direct interaction. Therefore, we purified the intracellular domains of both EGFR$^{L858R/T790M}$ and EGFR^{WT} followed by the ADP-Glo™ biochemical assay with purified EGFR, following purification via HisTrap HP chromatography, EGFR$^{L858R/T790M}$ and EGFR^{WT} exhibited purities exceeding 90%, as determined by SDS-PAGE analysis (Fig. S3e). As given in Fig .4b, CDDO-Me exhibited an IC_{50} of 17.22±5.26 nM for EGFR$^{L858R/T790M}$ which was about 1/10 than EGFR^{WT} (175.70±17.40 nM). These results demonstrated a significantly higher potency of CDDO-Me against EGFR$^{L858R/T790M}$ compared to the EGFR^{WT}, suggesting a potential therapeutic implication for targeting mutant EGFR. Furthermore, we conducted the MST assay with CDDO-Me and EGFR. MST results indicated that the binding affinity of EGFR$^{L858R/T790M}$ to CDDO-Me (1.10 μM) was notably stronger than that of EGFR^{WT} (79.01 μM) (Fig. 4c-d), demonstrating that CDDO-Me binds directly to EGFR and exhibits significantly higher affinity for the T790M mutant than for the wild type.

Molecular dynamics simulation revealing selectivity of CDDO-Me

Given the selective anti-proliferation activity between wild-type EGFR and EGFR$^{L858R/T790M}$ displayed by CDDO-Me, we conducted a molecular dynamics

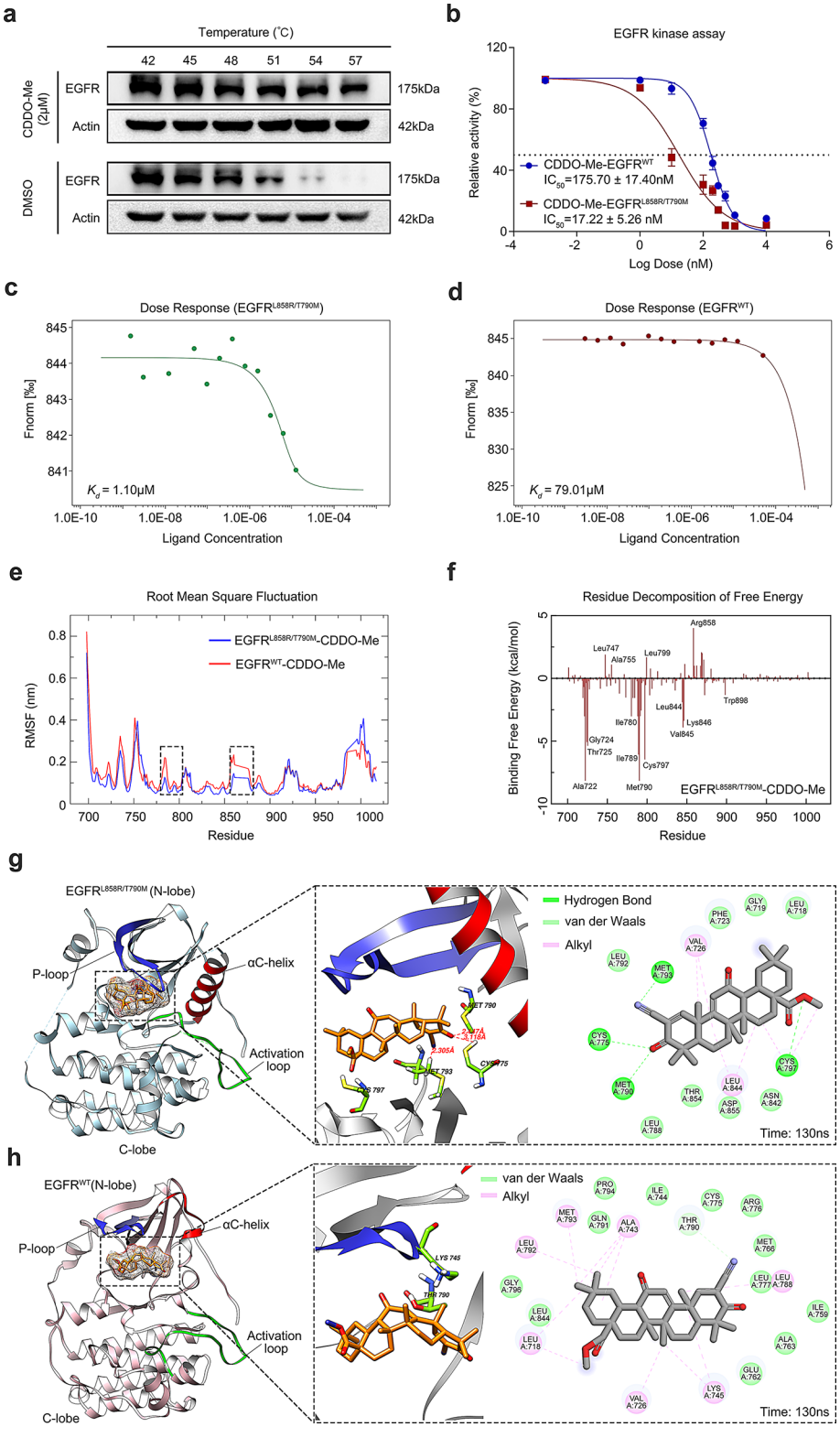


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 CDDO-Me specifically targets the L858R/T790M-mutant EGFR. (**a**) CDDO-Me thermally stabilized $EGFR^{L858R/T790M}$ in cellular level. H1975 cells were treated with CDDO-Me (2 μM) for 1 h. Lysates were divided into eight fractions, followed by heating to indicated temperatures (42, 45, 48, 51, 54, and 57 °C). (**b**) ADP-Glo™ assay determination of the IC_{50} values of CDDO-Me against purified intracellular domain of $EGFR^{WT}$ and $EGFR^{L858R/T790M}$, data are shown as mean ± S.D., $n = 3$. (**c**, **d**) The binding of fluorescently labeled $EGFR^{L858R/T790M}$ and $EGFR^{WT}$ to varying concentrations of CDDO-Me by microscale thermophoresis (MST). The K_d value of the EGFR-CDDO-Me interaction as determined by MST assay. (**e**) Root-mean square fluctuation (RMSF) values of $EGFR^{L858R/T790M}$ in complex with CDDO-Me (blue) and $EGFR^{WT}$ in complex with CDDO-Me (red) over 200 ns. (**f**) Binding free energy decomposition of CDDO-Me-$EGFR^{L858R/T790M}$ complex. (**g**) The 3D binding mode of CDDO-Me (orange) in the catalytic site of $EGFR^{L858R/T790M}$ and diagrammatic illustration of the interaction between $EGFR^{L858R/T790M}$ binding site residues and CDDO-Me by BIOVIA Discovery Studio Visualizer software. The protein and ligand CDDO-Me are shown by cartoon and stick respectively with key residues labeled and demonstrated as green sticks, and the hydrogen bonds are labeled by red dashed lines. Ligand is presented by gray line, green dashed line is conventional hydrogen bonds, light green dashed line is Van der Waals, and light pink dashed line is Alkyl. (**h**) The 3D binding mode of CDDO-Me (orange) in the catalytic site of $EGFR^{WT}$ and diagrammatic illustration of the interaction between $EGFR^{WT}$ binding site residues and CDDO-Me by BIOVIA Discovery Studio Visualizer software. The protein and ligand complex display pattern are same as (**g**)

simulation to elucidate the potential mechanism underlying the interaction between CDDO-Me and EGFR. To this end, we determined the binding stability between CDDO-Me and EGFR using long durational molecular dynamic simulation (200 ns) of CDDO-Me-$EGFR^{WT}$ and CDDO-Me-$EGFR^{L858R/T790M}$.

Root-mean-square deviation

FigureS3f depicts the RMSD trajectories for the $EGFR^{L858R/T790M}$/$EGFR^{WT}$-CDDO-Me complex. The trajectory for $EGFR^{L858R/T790M}$ is illustrated in blue, with RMSD values shown on the left *Y*-axis in nanometers (nm). In contrast, the trajectory for $EGFR^{WT}$ is depicted in red, with RMSD values also presented on the left *Y*-axis in nanometers (nm).

From the original frame, the protein RMSD began to fluctuate from around 0.2 nm at 0 ns. The protein RMSD gradually increased, reaching equilibrium at approximately 130 ns, with an average RMSD of around 0.32 nm during the 130–200 ns simulation period. In contrast, the RMSD trajectory of the $EGFR^{WT}$-CDDO-Me complex continued to ascend without stabilizing, indicating that the $EGFR^{L858R/T790M}$ variant achieved greater conformational stability with the incorporation of the CDDO-Me molecule.

Radius of gyration

Radius of gyration (Rg) was employed to assess the compactness effect of protein during molecular dynamic simulation. If Rg values among these systems presented a significant difference, it implied that ligand may induce compactness or repellency. As given in Supplementary Fig.S3g, The Rg of the $EGFR^{L858R/T790M}$ system was generally lower than that of the $EGFR^{WT}$. Notably, after 130 ns, the Rg of the $EGFR^{L858R/T790M}$ system was significantly reduced compared to $EGFR^{WT}$. These results indicated that the CDDO-Me molecule affect the spatial structure of $EGFR^{L858R/T790M}$ to stabilize the protein-ligand complex.

Root-mean-square fluctuation

To identify the critical residues of protein interacting with ligand, we computed the Root Mean Square Fluctuation (RMSF) of $EGFR^{WT}$-CDDO-Me and $EGFR^{L858R/T790M}$-CDDO-Me. As given in Fig. 4e, there were several critical residues identified as having high association with protein-ligand interaction, i.e. Leu775-Thr800, Ile850-Leu875 amino acid regions (labeled by black box). Among these residues, the Leu775-Thr800 region is situated within the ATP binding pocket, with the steric hindrance induced by Met790 considered a pivotal factor in TKIs resistance [48]. Overall, molecular dynamic simulation results clearly showed that CDDO-Me can stabilize the protein conformation.

Binding free energy analysis

To further elucidate the receptor-ligand interactions, we employed the MM/PBSA approach to quantitatively assess the binding free energy and energy decomposition values (Table. S1). Our analysis unveiled a remarkably strong binding affinity of $EGFR^{L858R/T790M}$ and CDDO-Me, as supported by the calculated values of ΔG_{bind}, ΔG_{vdW}, ΔG_{ele}, ΔG_{nonpol}, and ΔG_{pol}, which were determined to be -31.25, -35.91, 0.61, -6.12, and 10.17 kcal/mol, respectively. Throughout the 200 ns MD simulation, it is noteworthy that several residues, such as Ala722, Met790, Cys797, and Arg858, made significant contributions to the binding energy (Fig. 4f). Notably, Ala722 and Met790 exhibited the lowest binding energy, approximately −8.16 kcal/mol. Additionally, these residues are constituents of the ATP binding pocket, suggesting the interaction between CDDO-Me and the active pocket of $EGFR^{L858R/T790M}$.

Structural analysis after equilibration

To further investigate the interactions between the protein and ligand, we extracted the protein-ligand complex structure at the 130 ns mark for detailed analysis. As given in Fig. 4g-h, conformation of $EGFR^{L858R/T790M}$-CDDO-Me complex revealed that CDDO-Me bound to the active site of $EGFR^{L858R/T790M}$ to form hydrogen bonds with CYS775, MET793, MET790 and CYS797, which

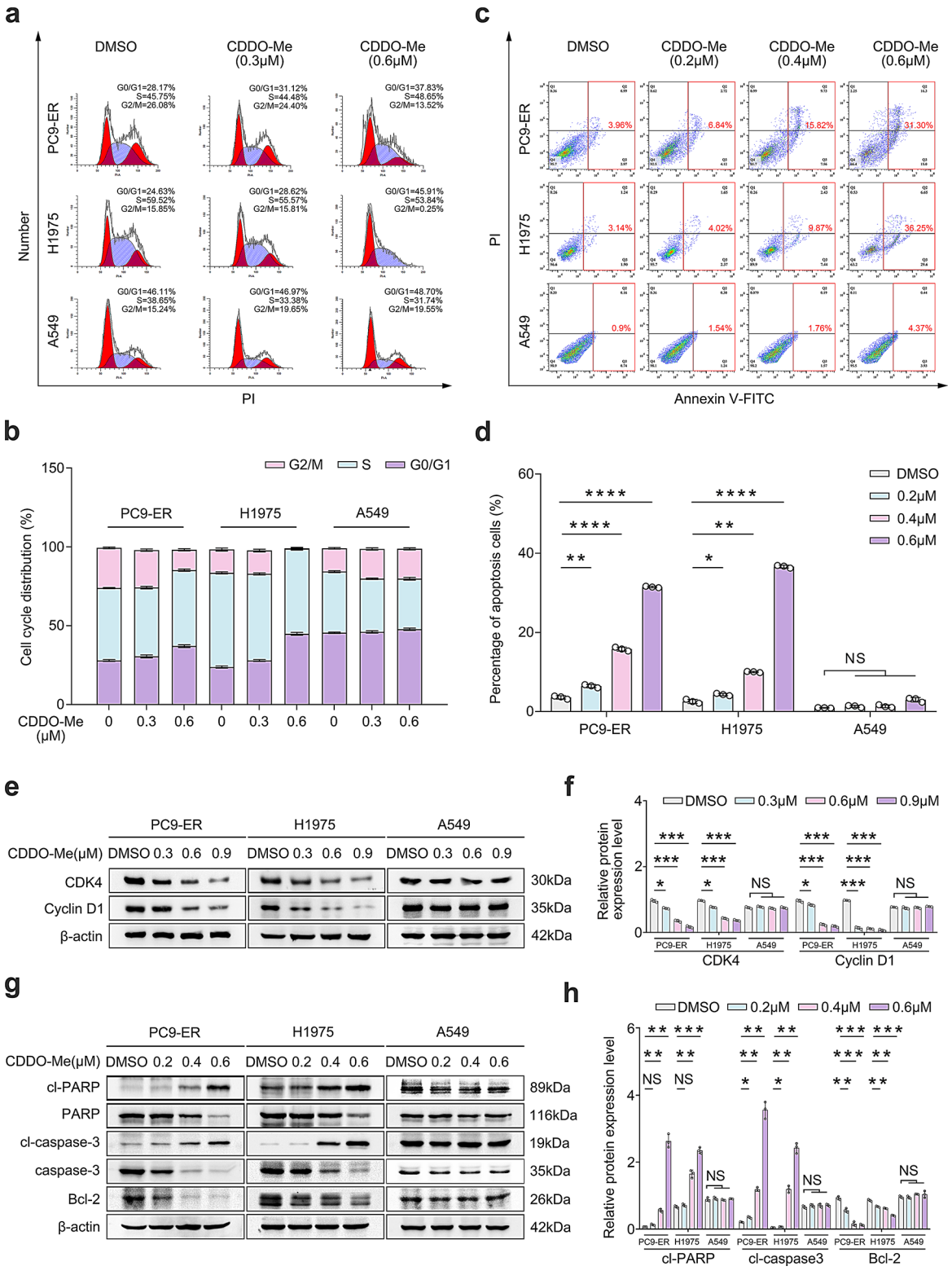


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 CDDO-Me induces cellular apoptosis and modulates cell cycle progression in erlotinib-resistant NSCLC cells. (**a**) Flow cytometry revealed the cell cycle distribution of H1975, PC9-ER, and A549 cells treated with the indicated concentrations of CDDO-Me (0, 0.3, and 0.6 μM) for 24 h. (**b**) The statistical result of cell cycle distribution. (**c**) Apoptosis of H1975, PC9-ER, and A549 cells incubated with the indicated concentrations of CDDO-Me (0, 0.2, 0.4, and 0.6 μM) for 24 h were detected by the Annexin-V FITC/PI double-staining assay. (**d**) Quantitative measurement of cellular apoptosis abundance after CDDO-Me treatment (0, 0.2, 0.4, and 0.6 μM) for 24 h. (**e**) Immunoblot analysis was conducted to reveal the cell cycle distribution of H1975, PC9-ER, and A549 cells after incubation with the indicated concentrations of CDDO-Me (0, 0.3, 0.6, and 0.9 μM) for 24 h. CDK4 and CyclinD1 protein levels were measured by Western blot analysis, and β-actin was detected as the endogenous loading control, respectively. (**f**) Quantitative measurement of CDK4 and CyclinD1 after CDDO-Me treatment (0, 0.3, 0.6, and 0.9 μM) for 24 h. (**g**) Immunoblot analysis was conducted to reveal the expression levels of PARP-mediated apoptosis biomarker expression levels of H1975, PC9-ER, and A549 cells after incubation with the indicated concentrations of CDDO-Me (0, 0.2, 0.4, and 0.6 μM) for 24 h. PARP, cl-PARP, caspase 3, cl-caspase 3, and Bcl-2, protein levels were measured by Western blot analysis, and β-actin was detected as the endogenous loading control, respectively. (**h**) Quantitative measurement of PARP, cl-PARP, caspase 3, cl-caspase 3, and Bcl-2 after CDDO-Me treatment (0, 0.2, 0.4, and 0.6 μM) for 24 h. Data are shown as mean ± S.D., $n = 3$. One-way ANOVA with Tukey's multiple-comparisons test, NS no significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

were all around ATP-binding pocket and contributed to the stronger binding capacity of CDDO-Me (Fig. 4g). However, CDDO-Me did not form any hydrogen bond interactions with $EGFR^{WT}$ (Fig. 4h).

CDDO-Me caused G0/G1 cell-cycle arrest and triggered apoptosis in EGFR-mutant NSCLC cells through caspase-dependent pathway

To dissect whether the anti-proliferation is caused by CDDO-Me-involved cell cycle arrest, we employed flow cytometry to examine the cell cycle distribution of EGFR mutant NSCLC cells and immunoblot analysis was conducted to measure the expression levels of cell cycle regulatory proteins after treatment with CDDO-Me. As given in Fig. 5a-b, CDDO-Me significantly increased the cell percentage of EGFR mutant NSCLC cells (H1975 and PC9-ER) at G0/G1 phase and caused subsequential reduction in G2/M phase compared to the control and EGFR-WT NSCLC cell (A549). Subsequentially, western blot analysis of cell cycle regulatory proteins, including CDK4 and Cyclin D1, was conducted. As shown in Fig. 5e-f, quantitative measurement of CDK4 and Cyclin D1 expression displayed the dose-dependent suppression in EGFR mutant NSCLC cells (H1975 and PC9-ER) compared with controls and EGFR-WT NSCLC cell (A549). Together, these results corroborated that G0/G1 cell cycle arrest may play an essential role in the CDDO-Me-induced antiproliferation against H1975 and PC9-ER cells.

Currently, apoptosis is one of major molecular mechanism in molecular targeted cell death. To investigate whether apoptosis involved into the CDDO-Me-induced cellular inhibition, we examined the apoptotic percentage of NSCLC cells with various concentration of CDDO-Me treatment (0.2, 0.4, and 0.6 μM) for 48 h. As show in Fig. 5c-d, we can observe a significantly increase in late apoptosis in a dose-dependent manner, while the percentage of late apoptosis in PC9-ER and H1975 was increased from 3.96±0.04%, 3.14±0.12% to 31.30±0.05%, 36.25±0.26%, respectively. However, no difference was observed in EGFR wild type NSCLC cell A549 after CDDO-Me treatment.

Furthermore, we examined the expression levels of caspase-related proteins in EGFR downstream to determine whether CDDO-Me activated caspase-dependent pathway. As given in Fig. 5g-h, significantly up-regulation of cleaved PARP and caspase3 was observed after CDDO-Me treatment, while expression levels of PARP, caspase, and Bcl-2 presented the dose-dependent down-regulation compared with negative control (A549). Therefore, these results suggested that apoptosis induced by CDDO-Me may rely on caspase-dependent signaling pathways.

CDDO-Me induced cell apoptosis via suppressing EGFR/PI3K/Akt/mTOR signaling pathway in EGFR-mutant NSCLC cells

As above-mentioned, CDDO-Me directly binds to T790M-mutant EGFR, inducing G0/G1 cell cycle arrest and promoting apoptosis in EGFR-mutant NSCLC cells. However, the molecular targets of small molecules often are not unique, so-called off-target effect, and complex interactions of potential target-CDDO-Me may cause similar phenotypes or molecular mechanisms compared with specific inhibition of given target. Consequently, more comprehensive validations should be conducted to elucidate whether inhibition of T790M-mutant EGFR plays the major role in CDDO-Me-caused cell apoptosis.

Firstly, we primarily determined the main EGFR downstream signaling cascade, PI3K/Akt/mTOR, which plays pivotal roles in NSCLC cellular processes, including proliferation, differentiation, migration, and apoptosis [49, 50]. Western blot analysis revealed that CDDO-Me can dose-dependently (0.2, 0.4, and 0.6 μM) suppress phosphorylation levels of EGFR, PI3K, AKT, and mTOR in EGFR mutant cells (H1975 and PC9-ER) compared with A549 cells at 48 h (Fig. 6a-b). Concomitantly, the expression levels of total EGFR, PI3K, AKT and mTOR did not alter after CDDO-Me treatment (Fig. 6a-b). Additionally, the phosphorylation level of Erk1/2 in H1975 and A549 cells did not obviously change until the concentration of CDDO-Me reached 0.4 μM, indicating that CDDO-Me exerts no significant influence on the MAPK signaling pathway(Fig. S4a-b). These results indicated that CDDO-Me exhibits anti-NSCLC activity by selectively inhibiting

the activation of the EGFR downstream PI3K/Akt/mTOR pathway in EGFR-T790M mutant cells.

To further explore the role of EGFR in the anti-cancer activity of CDDO-Me, we generated *EGFR* KO H1975 stable cell lines using CRISPR/Cas9-mediated knock-out (Fig. 6c). Immunoblot analysis of the sgEGFR group confirmed that EGFR was effectively knocked out(Fig. S4c-d). Subsequently, cells were treated with CDDO-Me, and the number of *EGFR* knockout H1975 cell colonies (Fig. 6d-e) was significantly higher than that of the sgNC group, indicating that *EGFR* KO cells displayed lower sensitivity to the indicated concentration of CDDO-Me treatment. We next sought to investigate whether *EGFR* knockout affected CDDO-Me-induced apoptosis in H1975 cell. As presented in Supplementary Fig. S4e-f, knockout of *EGFR* in H1975 cells significantly reduced the apoptotic rate induced from 21.64±0.02% to 11.09±0.03% by CDDO-Me treatment (0.6 μM). Additionally, compared to the sgNC group, *EGFR* knockout partially reversed the CDDO-Me-induced expression of cleaved PARP and cleaved caspase-3, validating that CDDO-Me induces the H1975 cells apoptosis majorly performing inhibition of EGFR (Fig. 6f-g).

Additionally, we employed ZSTK474, a well-established potent and selective inhibitor of phosphoinositide 3-kinase (PI3K), to determine whether CDDO-Me causes cellular apoptosis through indeed inhibiting the PI3K/Akt/mTOR signaling pathway. H1975 cells were treated either individually or in combination with the specified concentration of CDDO-Me (0.8 μM) and ZSTK474 (1 μM) for a duration of 24 h. The results clarified that co-treatment with CDDO-Me and ZSTK474 significantly suppressed the phosphorylation of EGFR, PI3K, Akt, and mTOR compared to either drug alone (Fig. S4g-h). Therefore, we demonstrated that CDDO-Me induces apoptosis and death in H1975 NSCLC cells through the EGFR/PI3K/Akt/mTOR signaling pathway. In summary, we propose that CDDO-Me directly interacts with mutant EGFR, downregulating its phosphorylation level and thereby suppressing the downstream PI3K/Akt/mTOR pathway to exert its anti-cancer effects (Fig. 6h).

CDDO-Me inhibited tumor growth of the H1975 xenograft model in vivo

We determined the anti-tumor activity of CDDO-Me in a H1975 xenograft model (Fig. 7a). After the tumor model was successfully established, mice were intraperitoneally injected with vehicle or CDDO-Me (3 and 6 $mg{\cdot}kg^{-1}$) every day for a total 20 times. As expected, CDDO-Me dose-dependent suppressed H1975 tumor growth in BALB/c nude mice (Fig. 7b) compared with vehicle group. By collecting the major organs and tumor tissues, CDDO-Me (6 mg/kg) significantly reduced the ~57% volume of H1975 tumor (Fig. 7c, d) compared to vehicle. Additionally, long durational drug administration also showed the low toxicity without observing any significantly difference of body weight between CDDO-Me therapy and vehicle group (Fig. 7e). As growth inhibition was recognized as suppression of EGFR-PI3K/AKT/mTOR axils, we examined the proliferation activity and EGFR phosphorylation in tumor tissue. Illustration in Fig. 7f-g, we observed that the percentage of Ki-67 proliferative cells and p-EGFR intensity were significantly reduced in tumor tissues, suggesting that CDDO-Me dose-dependent reduced tumor cell proliferation via suppression of EGFR phosphorylation levels. Notably, immunoblot analysis was conducted to examine expression levels of EGFR, PI3K, Akt, and mTOR in tumor tissues. Results demonstrated that phosphorylation levels of EGFR, PI3K, Akt, and mTOR were significantly decreased compared with vehicle group (Fig. 7h-i). These findings demonstrated that CDDO-Me yield a significant suppression of tumor in vivo through inhibition of EGFR-PI3K-Akt-mTOR axis without systematic side effects.

Discussion

In summary, aimed to conquer the drug resistance of EGFR mutation-led drug resistance, we have proposed a machine learning-aided drug screening and identify an emerging EGFR-mutant selective inhibitor CDDO-Me. Addressing drug resistance to clinical medications necessitates the identification of novel scaffolds and exploration of chemical space with various diversities, as similar scaffold structures often lead to analogous mechanisms of drug resistance. In comparison to random molecular design across the entire chemical space, marine- or food-sourced natural product libraries offer enhanced opportunities for discovering new scaffold structures capable of suppressing target activities through novel binding modalities. To date, several marine- or food-originated drugs (e.g. Ziconotide, Trabectedin, Eribulin) have been approved by FDA for disease therapy [51], and more candidates with higher bioactivity and safety are under the clinical trials. Moreover, long-standing dietary or edible applications can provide assurance regarding the safety of molecular constituents, mitigating potential adverse effects, for example genetic toxicity. Consequently, in our investigation, we utilized a commercially available molecular library derived from food sources to identify promising lead compounds with selectivity against mutated EGFR.

Our study displays a significant stride in the field of machine learning-aided drug discovery, particularly in the development of selective inhibitors. Indeed, over the past decade, a growing number of researchers have made notable progress in employing machine learning for inhibitor discovery. This has involved the application of diverse neural network methods to identify novel

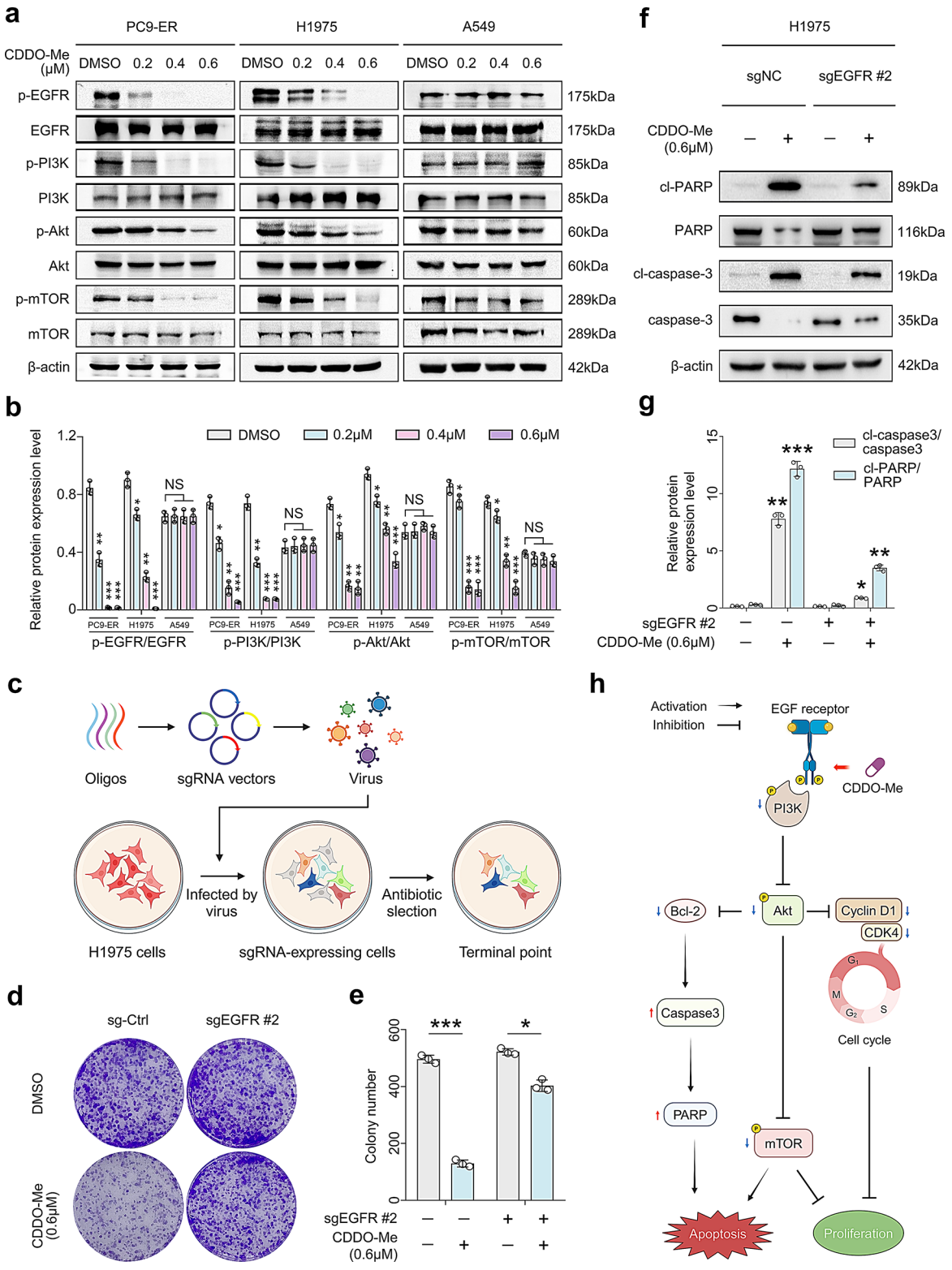


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 CDDO-Me-induced inhibition of NSCLC via the EGFR pathway in Erlotinib-resistant NSCLC cells. (**a**) Immunoblot was conducted to reveal the phosphorylation level of EGFR downstream pathways in H1975, PC9-ER, and A549 cells after incubation with the indicated concentrations of CDDO-Me (0, 0.2, 0.4, and 0.6 μM) for 24 h. p-EGFR, EGFR, p-PI3K, PI3K, p-Akt, Akt, p-mTOR, and mTOR protein levels were measured by Western blot analysis, and β-actin was detected as the endogenous loading control, respectively. (**b**) Quantitative measurement of p-EGFR, EGFR, p-PI3K, PI3K, p-Akt, Akt, p-mTOR, and mTOR band density after CDDO-Me treatment (0, 0.2, 0.4, and 0.6 μM) for 24 h. (**c**) Schematic illustration of experimental flowchart to construct CRISPR knockout H1975 cells. (**d**) Colony formation of *EGFR*-KO H1975 cells were measured after CDDO-Me (0 and 0.6 μM) treatment for 14 days, and photographs of crystal violet-stained colonies were depicted. (**e**) The statistical result of colony number after CDDO-Me treatment. (**f**) Immunoblot analysis was conducted to reveal the expression levels of PARP-mediated apoptosis biomarker in *EGFR*-KO H1975 cells after incubation with the indicated concentrations of CDDO-Me (0 and 0.6 μM) for 24 h. PARP, cl-PARP, caspase 3, and cl-caspase 3, protein levels were measured by Western blot analysis, and β-actin was detected as the endogenous loading control, respectively. (**g**) Quantitative measurement of PARP, cl-PARP, caspase 3, and cl-caspase 3 after CDDO-Me treatment (0 and 0.6 μM) for 24 h. (**h**) Graphical model for molecular mechanism of CDDO-Me-induced cell apoptosis in T790mutatnt NSCLC cell. Data are shown as mean ± S.D., $n = 3$. One-way ANOVA with Tukey's multiple-comparisons test, NS no significance, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

scaffold molecular moieties across various therapeutic domains, including antibiotics [52], antiviral agents [53], and anticancer drugs [54]. Despite advances in machine learning algorithm technologies implemented in the drug discovery, there are still many challenges regarding the implementation and integration of these algorithm into the drug discovery process specifically in general. To our knowledge, the superior points to improve efficacy of machine learning are algorithms, which determined the rate of convergence and precision. Different algorithms focused on different optimizing approaches, and these searching difference may cause the bias of outcomes [55]. As expected, there exist several avenues for enhancing prediction models, such as Boosting Generalization [56], Error Correlation Reduction [57], and Variance Reduction [58]. Compared with Employing individual algorithms within prediction models, introduction of various types of algorithms can enhance predictive efficiency by integrating more intricate algorithms or novel types of neural networks, so-called "hybrid algorithm". Recently, hybrid algorithms have been introduced in machine learning-aided drug discovery, yielding superior outcomes compared to traditional standalone algorithms [59, 60]. Building on these advancements, the hybrid prediction model presented herein strongly supports the enrichment of selective inhibitors derived from commercial datasets. Additionally, due to the discernible correlation between compound structures and bioactivities, we observed optimal training performance through ensemble learning progression. Our findings further suggest that similar machine learning-aided drug discovery pipelines hold promise for future biomedical applications.

To date, structure-based molecular design encounters similar limitations, as molecules with analogous structures often elicit comparable drug resistance mechanisms. Machine learning-assisted drug discovery offers a promising avenue to identify novel scaffold moieties capable of overcoming drug resistance. While EGFR inhibitors such as Gefitinib, Afatinib, and Erlotinib have been extensively utilized in NSCLC therapy, the emergence of drug resistance to these agents underscores the necessity of alternative therapeutic strategies. Although drug resistance to EGFR inhibitors may be mitigated by suppressing alternative signaling pathways, direct targeting of mutant EGFR proteins remains a primary therapeutic approach for NSCLC. For instance, the development of AZD-9291, which covalently binds to the EGFR-T790M mutation, exemplifies this strategy. Remarkably, our findings indicate that CDDO-Me directly binds with T790M-mutant EGFR and shares a similar binding pattern to the catalytic site of EGFR as AZD-9291, located adjacent to the P-loop and directly interacting with the Cys797 residue, thereby activates the EGFR-PI3K-AKT mediated apoptosis and cell cycle pathway. Complete suppression of mTOR catalytic activity demonstrates efficacy in inducing caspase-dependent apoptosis. As anticipated, we also observed a reduction in caspase-3 and phosphorylated mTOR levels induced by CDDO-Me. Additionally, CDDO-Me downregulates G0/G1 cell cycle checkpoint proteins Cyclin D1 and CDK4 to exert anti-mutant NSCLC activity. Specifically, induction of long-term G1 cell cycle arrest in oncogenic-driven cancer growth imparts cytotoxic effects, thereby presenting a potential avenue for reversing drug resistance [61].

Bardoxolone methyl (CDDO-Me), a semi-synthetic triterpenoid derived from the natural product oleanolic acid, exerts its pharmacological effects by activating the Kelch-like ECH-associated protein 1 (Keap1)/nuclear factor erythroid 2-related factor 2 (Nrf2) axis, thereby modulating immunological responses, such as those observed in the treatment of chronic kidney disease [62, 63]. Furthermore, CDDO-Me has demonstrated efficacy in suppressing SARS-CoV-2 replication, mitigating macrophage infiltration to prevent metabolic dysfunction-associated steatohepatitis, and inhibiting osteoclastogenesis to counteract extracellular matrix degradation [64–66]. Furthermore, extensive literature indicates that CDDO-Me inhibits PPAR-γ activity, thereby obstructing differentiation and augmenting the differentiation effects of all-trans-retinoic acid (ATRA) in acute myeloid leukemia (AML) [67, 68]. This compound induces apoptosis in cancer cells through several mechanisms, including the generation of reactive oxygen species (ROS) [69],

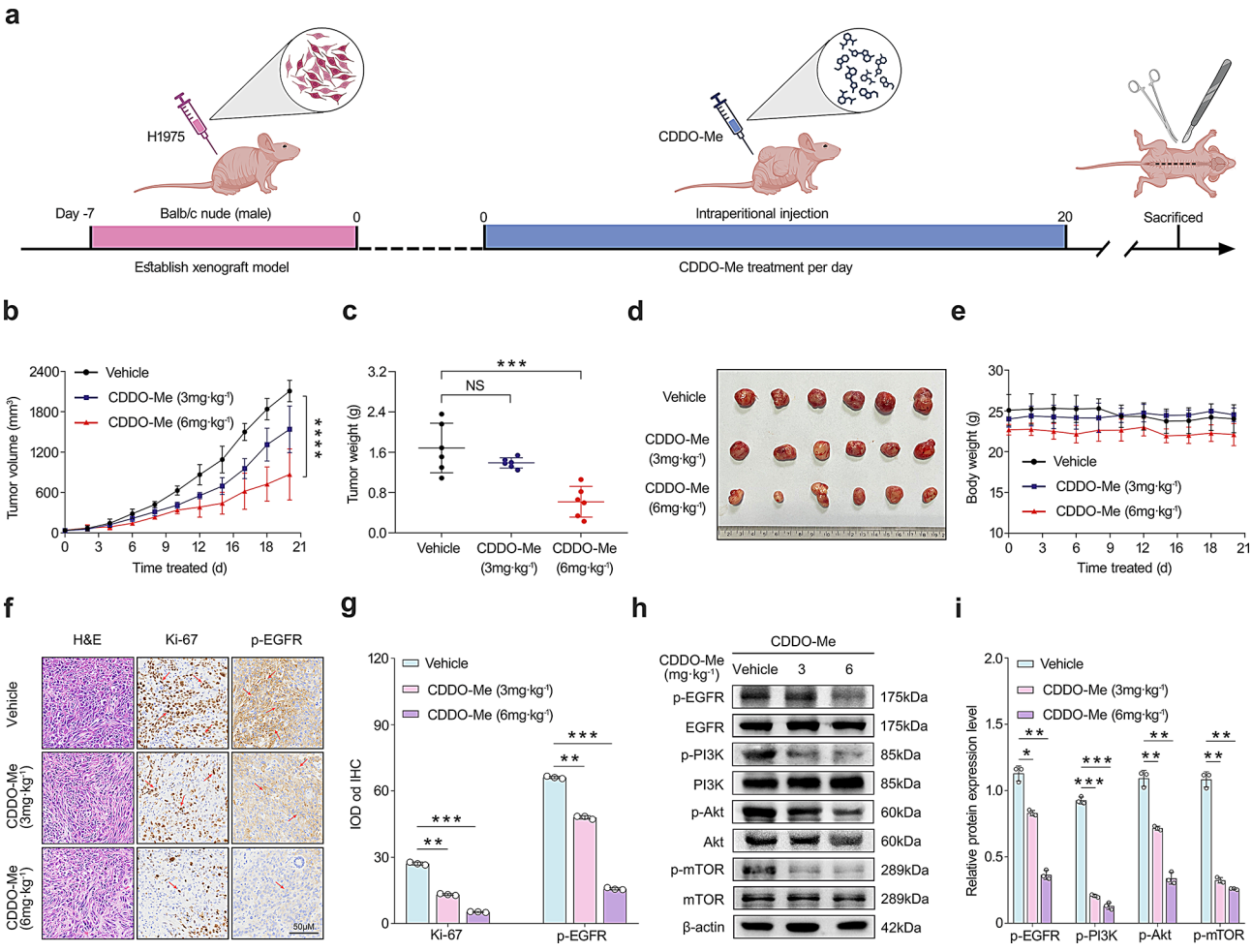


Fig. 7 CDDO-Me suppresses tumor progression in H1975 xenograft models. H1975 cells (5×10^6) were subcutaneously inoculated into BALB/c male nude mice. The mice were allocated to six groups after 7 days of tumor-cell implantation. (**a**) Schematic graph of the CDDO-Me administration protocol in the established H1975 xenograft model. (**b**) Tumor volume (length × width^2 × 0.5 mm^3) was measured every 2 days and treated with vehicle (2% DMSO in PBS, i.p., $n=6$), CDDO-Me (3 mg·kg^{-1}, i.p., $n=6$) and CDDO-Me (6 mg·kg^{-1}, i.p., $n=6$). (**c**) The scatter plot summarized the weight of the tumors. (**d**) The representative stripped images of the tumor entity after being treated with vehicle, 3 mg·kg^{-1} and 6 mg·kg^{-1} of CDDO-Me for 20 days. (**e**) The body weight was quantified in each group. (**f**) Representative hematoxylin and eosin (HE) staining of tumor tissues. And immunocytochemical staining for Ki-67 and p-EGFR expression in tumor tissues from nude mice (Magnification, 400×. Scale bar, 50 μm). (**g**) Quantitative measurement of IHC results were analyzed by Image-Pro Plus 6.0 ($n=5$ fields of view). (**h**) Immunoblot analysis was conducted to reveal the phosphorylation level of EGFR downstream pathways in vivo, mice were treated with vehicle (2% DMSO in PBS, i.p, $n=6$), CDDO-Me (3 mg·kg^{-1}, i.p., $n=6$) and CDDO-Me (6 mg·kg^{-1}, i.p., $n=6$). p-EGFR, EGFR, p-PI3K, PI3K, p-Akt, Akt, p-mTOR, and mTOR protein levels were measured by Western blot analysis, and β-actin was detected as the endogenous loading control, respectively. (**i**) Quantitative measurement of p-EGFR, EGFR, p-PI3K, PI3K, p-Akt, Akt, p-mTOR, and mTOR. Data are shown as mean ± S.D., $n=3$. One-way ANOVA with Tukey's multiple-comparisons test, NS no significance, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$

activation of caspase-3/8 [70], and c-Jun NH2-terminal kinase (JNK) [71], as well as by inhibiting the PI3K/Akt/mTOR signaling pathway to promote autophagy [72]. Additionally, CDDO-Me attenuates Janus kinase (JAK) activity [73, 74], reducing cell proliferation and differentiation, and suppresses telomerase activity, leading to apoptosis. It also inhibits the mitochondrial protease Lonp1, impacting mitochondrial function, and binds to Ubiquitin-Specific-Processing Protease 7 [75], consequently downregulating substrates such as MDM2, MDMX, and UHRF1 [76]. Collectively, these findings suggest that CDDO-Me interacts with a diverse array of molecular targets to exert its anti-cancer effects, demonstrating its potential as a multi-target therapeutic agent.

The established safety profile of CDDO-Me in these clinical applications renders it a promising candidate for cancer therapy, with minimal consideration for adverse effects. Biochemical assays conducted in vivo, coupled with molecular dynamics simulations, have provided insights into the selective binding of CDDO-Me to EGFR mutations, exhibiting a wide margin of selectivity against mutated variants in non-small cell lung cancer (NSCLC). Notably, CDDO-Me demonstrates heightened selectivity against downstream kinases within the EGFR-PI3K

pathway. Moreover, the remarkable dose-dependent antitumor activity of CDDO-Me observed in xenograft H1799 tumor models further underscores its potential as a therapeutic agent targeting T790M mutations, as validated by immunoblot analysis.

Overall, our studies presented the discovery of a selective inhibitor, CDDO-Me, targeting EGFR-T790M for the treatment of drug-resistant NSCLC tumor proliferation, achieved through a machine learning-aided drug discovery approach. Initially, our investigation demonstrated that the hybrid algorithm significantly improves machine learning prediction efficiency of the selective binding ability of T790M-mutant EGFR compared to wild type, thereby facilitating the screening of candidate compounds for subsequent bioactivity assays. Subsequently, we employed various experimental techniques, including CETSA, and molecular dynamics simulations, to cross-validate that CDDO-Me selectively inhibits the degradation rate of EGFR mutants in vivo and forms stable interactions within the EGFR-ATP binding pocket. To determine the molecular mechanisms of CDDO-Me against EGFR-mutated NSCLC, the anti-tumor activity was ascribed to the induction of cellular apoptosis and cell cycle arrest at the G0/G1 phase, achieved through modulation of the EGFR/PI3K/Akt/mTOR signaling cascade. Importantly, our findings underscore the feasibility of employing machine learning-assisted virtual screening methodologies to identify promising EGFR inhibitors. As such, we advocate for a strategy aimed at impeding EGFR kinase-dependent functions, thereby introducing novel and efficacious avenues for the discovery of potential inhibitors and the treatment of NSCLC.

Conclusion

Our study successfully identified CDDO-Me, an oleanolic acid derivative with anti-inflammatory properties, as a selective inhibitor of EGFR-T790M. The efficacy of CDDO-Me was validated through in vitro cellular assays and molecular dynamic simulations, showing significant suppression of NSCLC cell proliferation harboring the EGFR-T790M mutation. Furthermore, CDDO-Me induced apoptosis and caused cell cycle arrest by inhibiting the PI3K-Akt-mTOR pathway, directly targeting the EGFR protein. In vivo studies in a H1975 xenograft mouse model demonstrated dose-dependent suppression of tumor growth, highlighting the therapeutic potential of CDDO-Me in overcoming NSCLC drug resistance.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12964-024-01954-7.

Supplementary Material 1

Supplementary Material 2

Author contributions

N.L., H.W., R.Z., Z.L., B.J. and X.P. designed and conducted the experiments. R.Z., B.J., Z.Z., and N.L. designed and conducted the machine learning-associated assays. R.Z., Z.L., T.W., X.P., T.L., K.M., Y.L., X.H., A.H., H.W. and N.L. participated in data analysis. N.L. and B.J. coordinated and directed the project. R.Z., H.W., and N.L. contributed to the preparation of the manuscript. H.W., B.J., Z.Z. and N.L. revised the manuscript. H.W., B.J., Z.Z. and N.L. obtained the funding. All authors discussed the results and commented on the manuscript. All authors have read and approved the final manuscript.

Funding

This work was financially supported by Natural Science Foundation of Shanghai (No. 21ZR1427300), Shanghai Frontiers Research Center of the Hadal Biosphere, SciTech Funding by CSPFTZ Lingang Special Area Marine Biomedical Innovation Platform, Shanghai Science and Technology Program (No. 21S21902800), the project of Naval Medical Center of PLA (No. 21TPQN0802), and Henan province development breakthrough program (No. 242102310349).

Data availability

The source code is available at https://github.com/Arimakousei1/Hybrid-algorithm-using-parallel-algorithms .

Declarations

Ethics approval and consent to participate

The animal use and experimental procedures were approved by the Institutional Animal Care and Use Committee of Shanghai Ocean University (Shanghai, China).

Competing interests

The authors declare no competing interests.

Author details

[1]International Research Centre for Food and Health, College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China
[2]Marine Biomedical Science and Technology Innovation Platform of Lingang Special Area, Shanghai 201306, China
[3]Naval Medicine Center of PLA, Naval Military University, Shanghai 201306, China
[4]School of Life Sciences, Henan University, Kaifeng, Henan Province 475000, China
[5]Department of Marine Biopharmacology, College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China
[6]Shanghai Engineering Research Center of Aquatic-Product Processing & Preservation, Shanghai 201306, China
[7]School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Received: 2 July 2024 / Accepted: 21 November 2024
Published online: 05 December 2024

References

1. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. Nat Rev Clin Oncol. 2014;11:473–81.
2. Weiss JM, Stinchcombe TE. Second-line therapy for Advanced NSCLC. Oncologist. 2013;18:947–53.
3. Chong CR, Jänne PA. The quest to overcome resistance to EGFR-targeted therapies in cancer. Nat Med. 2013;19:1389–400.
4. Roskoski R. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol Res. 2016;103:26–48.
5. Pao W, Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. Nat Rev Cancer. 2010;10:760–74.
6. Liu Q, Yu S, Zhao W, Qin S, Chu Q, Wu K. EGFR-TKIs resistance via EGFR-independent signaling pathways. Mol Cancer. 2018;17:53.

7. Yun C-H, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong K-K, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci USA. 2008;105:2070–5.
8. Dong R-F, Zhu M-L, Liu M-M, Xu Y-T, Yuan L-L, Bian J, et al. EGFR mutation mediates resistance to EGFR tyrosine kinase inhibitors in NSCLC: from molecular mechanisms to clinical research. Pharmacol Res. 2021;167:105583.
9. Westover D. Mechanisms of acquired resistance to first- and second-generation EGFR tyrosine kinase inhibitors. Ann Oncol. 2018;29.
10. Juchum M, Günther M, Laufer SA. Fighting cancer drug resistance: opportunities and challenges for mutation-specific EGFR inhibitors. Drug Resist Updates. 2015;20:12–28.
11. Shi K, Wang G, Pei J, Zhang J, Wang J, Ouyang L, et al. Emerging strategies to overcome resistance to third-generation EGFR inhibitors. J Hematol Oncol. 2022;15:94.
12. Sabbah DA, Hajjo R, Sweidan K. Review on epidermal growth factor receptor (EGFR) structure, signaling pathways, interactions, and recent updates of EGFR inhibitors. CTMC. 2020;20:815–34.
13. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. Acta Pharm Sinica B. 2015;5:390–401.
14. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18:463–77.
15. Liu G, Catacutan DB, Rathod K, Swanson K, Jin W, Mohammed JC, et al. Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii. Nat Chem Biol. 2023;19:1342–50.
16. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting Drug Response and Synergy using a deep learning model of Human Cancer cells. Cancer Cell. 2020;38:672–e6846.
17. Gerdes H, Casado P, Dokal A, Hijazi M, Akhtar N, Osuntola R, et al. Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. Nat Commun. 2021;12:1850.
18. Seyyed-Kalantari L. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27.
19. Zaky AA, Shim J-H, El-Aty AMA, Bioactivities. Applications, Safety, and Health Benefits of Bioactive Peptides from Food and By-Products: a review. Front Nutr. 2022;8.
20. Ono A. Mutant allele frequency predicts the efficacy of EGFR-TKIs in lung adenocarcinoma harboring the L858R mutation. Ann Oncol. 2014;25.
21. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for Better Benchmarking. J Med Chem. 2012;55:6582–94.
22. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O. Virtual screening Workflow Development guided by the receiver operating characteristic curve Approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem. 2005;48:2534–47.
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25. 1. Advanced Drug Delivery Reviews. 2001;46:3–26.
24. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49:W5–14.
25. Yan S, Zhang B, Feng J, Wu H, Duan N, Zhu Y, et al. FGFC1 selectively inhibits Erlotinib-Resistant Non-small Cell Lung Cancer via Elevation of ROS mediated by the EGFR/PI3K/Akt/mTOR pathway. Front Pharmacol. 2022;12:764699.
26. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2009;30:2785–91.
27. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model. 2021;61:3891–98.
28. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-pdb viewer: an environment for comparative protein modeling. Electrophoresis. 1997;18:2714–23.
29. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004;25:1157–74.
30. Ponder JW, Case DA. Force Fields for Protein Simulations. Advances in Protein Chemistry [Internet]. Elsevier; 2003 [cited 2023 Dec 6]. pp. 27–85. https://linkinghub.elsevier.com/retrieve/pii/S006532330366002X
31. Nurani A, Yamashita Y, Taki Y, Takada Y, Itoh Y, Suzuki T. Identification of a histone deacetylase 8 inhibitor through drug screenings based on machine learning. Chem Pharm Bull. 2024;72:173–8.
32. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50:742–54.
33. Kordzadeh N, Ghasemaghaei M. Algorithmic bias: review, synthesis, and future research directions. Eur J Inform Syst. 2022;31:388–409.
34. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24:1565–7.
35. Yang F-J, An Implementation of Naive Bayes Classifier. 2018 International Conference on Computational Science and Computational Intelligence (CSCI) [Internet]. Las Vegas, NV, USA: IEEE; 2018 [cited 2024 May 10]. pp. 301–6. https://ieeexplore.ieee.org/document/8947658/
36. Cunningham P, Delany SJ. k-Nearest Neighbour classifiers - A Tutorial. ACM Comput Surv. 2022;54:1–25.
37. Chen T, Guestrin C, XGBoost:. A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. 2016 [cited 2024 May 10]. pp. 785–94. http://arxiv.org/abs/1603.02754
38. Arena P, Fortuna L, Re R, Xibilia MG. MULTILAYER PERCEPTRONS TO APPROXIMATE COMPLEX VALUED FUNCTIONS. Int J Neur Syst. 1995;06:435–46.
39. Van De Waterbeemd H, Gifford E. ADMET IN SILICO MODELLING: TOWARDS PREDICTION PARADISE? Nat Rev Drug Discov. 2003;2:192–204.
40. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliver Rev. 1997;23:3–25.
41. Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, Yuan ML, et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. Sci Transl Med. 2019;11:eaaw8412.
42. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45:2615–23.
43. Huang X-H, Yan X, Zhang Q-H, Hong P, Zhang W-X, Liu Y-P, et al. Direct targeting of HSP90 with daurisoline destabilizes β-catenin to suppress lung cancer tumorigenesis. Cancer Lett. 2020;489:66–78.
44. Chen W, Li P, Liu Y, Yang Y, Ye X, Zhang F, et al. Isoalantolactone induces apoptosis through ROS-mediated ER stress and inhibition of STAT3 in prostate cancer cells. J Exp Clin Canc Res. 2018;37:309.
45. Zhou S-F, Wang Y-Y, Yang Y-X, He Z, Zhao R, Hong Z et al. Bardoxolone methyl induces apoptosis and autophagy and inhibits epithelial-to-mesenchymal transition and stemness in esophageal squamous cancer cells. Drug Des Dev Ther. 2015;993.
46. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the Time of Acquired Resistance to EGFR-TKI Therapy in 155 patients with *EGFR* -Mutant lung cancers. Clin Cancer Res. 2013;19:2240–7.
47. Jafari R, Almqvist H, Axelsson H, Ignatushchenko M, Lundbäck T, Nordlund P, et al. The cellular thermal shift assay for evaluating drug target interactions in cells. Nat Protoc. 2014;9:2100–22.
48. Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, et al. Mechanism for activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. Cell. 2009;137:1293–307.
49. Kampa-Schittenhelm K, Heinrich M, Akmut F, Rasp K, Illing B, Döhner H, et al. Cell cycle-dependent activity of the novel dual PI3K-MTORC1/2 inhibitor NVP-BGT226 in acute leukemia. Mol Cancer. 2013;12:46.
50. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141:1117–34.
51. Jiménez C. Marine Natural products in Medicinal Chemistry. ACS Med Chem Lett. 2018;9:959–61.
52. Torres MDT, De La Fuente-Nunez C. Toward computer-made artificial antibiotics. Curr Opin Microbiol. 2019;51:30–8.
53. Rajput A, Thakur A, Mukhopadhyay A, Kamboj S, Rastogi A, Gautam S, et al. Prediction of repurposed drugs for coronaviruses using artificial intelligence and machine learning. Comput Struct Biotechnol J. 2021;19:3133–48.
54. Fan K, Cheng L, Li L. Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. Brief Bioinform. 2021;22:bbab271.
55. Das S, Dey A, Pal A, Roy N. Applications of Artificial Intelligence in Machine Learning: review and Prospect. IJCA. 2015;115:31–41.
56. Suryavanshi A, Kukreja V, Srivastava P, Bhattacherjee A, Rawat RS. Felis catus disease detection in the digital era: Combining CNN and Random Forest. 2023 International Conference on Artificial Intelligence for Innovations in

Healthcare Industries (ICAIIHI) [Internet]. Raipur, India: IEEE; 2023 [cited 2024 May 10]. pp. 1–7. https://ieeexplore.ieee.org/document/10489457/
57. Fujiwara T, Denno S, Hou Y. Out-of-bound Signal Demapping for Lattice reduction-aided iterative Linear receivers in Overloaded MIMO systems. IEICE Trans Commun. 2021;E104B:974–82.
58. Gower RM, Schmidt M, Bach F, Richtarik P. Variance-Reduced Methods for Machine Learning [Internet]. arXiv; 2020 [cited 2024 May 10]. http://arxiv.org/abs/2010.00892
59. Ying Y, Wang L, Ma S, Zhu Y, Ye S, Jiang N, et al. An enhanced machine learning approach for effective prediction of IgA nephropathy patients with severe proteinuria based on clinical data. Comput Biol Med. 2024;173:108341.
60. Godlewski A, Czajkowski M, Mojsak P, Pienkowski T, Gosk W, Lyson T, et al. A comparison of different machine-learning techniques for the selection of a panel of metabolites allowing early detection of brain tumors. Sci Rep. 2023;13:11044.
61. Foy R, Crozier L, Pareri AU, Valverde JM, Park BH, Ly T, et al. Oncogenic signals prime cancer cells for toxic cell overgrowth during a G1 cell cycle arrest. Mol Cell. 2023;83:4047–e40616.
62. Liby KT, Sporn MB. Synthetic Oleanane Triterpenoids: Multifunctional Drugs with a Broad Range of Applications for Prevention and Treatment of Chronic Disease. Esbenshade TA, editor. Pharmacol Rev. 2012;64:972–1003.
63. Nangaku M, Kanda H, Takama H, Ichikawa T, Hase H, Akizawa T. Randomized Clinical Trial on the Effect of Bardoxolone Methyl on GFR in Diabetic kidney disease patients (TSUBAKI Study). Kidney Int Rep. 2020;5:879–90.
64. Sun Q, Ye F, Liang H, Liu H, Li C, Lu R, et al. Bardoxolone and bardoxolone methyl, two Nrf2 activators in clinical trials, inhibit SARS-CoV-2 replication and its 3 C-like protease. Sig Transduct Target Ther. 2021;6:212.
65. Yang R, Guo Y, Zong S, Ma Z, Wang Z, Zhao J, et al. Bardoxolone methyl ameliorates osteoarthritis by inhibiting osteoclastogenesis and protecting the extracellular matrix against degradation. Heliyon. 2023;9:e13080.
66. Onuma K, Watanabe K, Isayama K, Ogi S, Tokunaga Y, Mizukami Y. Bardoxolone methyl prevents metabolic dysfunction-associated steatohepatitis by inhibiting macrophage infiltration. British J Pharmacology. 2024;bph.16374.
67. Wang Y, Porter WW, Suh N, Honda T, Gribble GW, Leesnitzer LM et al. A synthetic triterpenoid, 2-cyano-3,12-dioxooleana-1,9-dien-28-oic Acid (CDDO), Is a ligand for the peroxisome proliferator-activated receptor γ. Mol Endo. 14(10):1550–56.
68. Tsao T, Kornblau S, Safe S, Watt JC, Ruvolo V, Chen W, et al. Role of peroxisome proliferator-activated Receptor-γ and its Coactivator DRIP205 in Cellular responses to CDDO (RTA-401) in Acute Myelogenous Leukemia. Cancer Res. 2010;70:4949–60.
69. Ikeda T, Sporn M, Honda T, Gribble GW, Kufe D. The Novel Triterpenoid CDDO and its derivatives induce Apoptosis by Disruption of Intracellular Redox Balance. Cancer Res 2003;63:5551–58.
70. Hyer ML, Croxton R, Krajewska M, Krajewski S, Kress CL, Lu M, et al. Synthetic triterpenoids cooperate with Tumor Necrosis factor–related apoptosis-inducing ligand to Induce apoptosis of breast Cancer cells. Cancer Res. 2005;65:4799–808.
71. Zou W, Yue P, Khuri FR, Sun S-Y. Coupling of endoplasmic reticulum stress to CDDO-Me–Induced Up-regulation of death receptor 5 via a CHOP–Dependent mechanism involving JNK activation. Cancer Res. 2008;68:7484–92.
72. Wang X-Y, Zhang X-H, Peng L, Liu Z, Yang Y-X, He Z-X et al. Bardoxolone methyl (CDDO-Me or RTA402) induces cell cycle arrest, apoptosis and autophagy via PI3K/Akt/mTOR and p38 MAPK/Erk1/2 signaling pathways in K562 cells. Am J Transl Res 2017;9(10):4652–72.
73. Ryu K, Susa M, Choy E, Yang C, Hornicek FJ, Mankin HJ et al. Oesleearachnaraticnlee triterpenoid CDDO-Me induces apoptosis in multidrug resistant osteosarcoma cells through inhibition of Stat3 pathway. BMC Cancer. 2010;10:187.
74. Deeb D, Gao X, Liu YB, Gautam SC. Inhibition of cell proliferation and induction of apoptosis by CDDO-Me in pancreatic cancer cells is ROS-dependent. J Exp Ther Oncol. 2012;10(1):51–64.
75. Bernstein SH, Venkatesh S, Li M, Lee J, Lu B, Hilchey SP et al. The mitochondrial ATP-dependent Lon protease: a novel target in lymphoma death mediated by the synthetic triterpenoid CDDO and its derivatives. Blood. 2012;119(14):3321–29.
76. Qin D, Wang W, Lei H, Luo H, Cai H, Tang C et al. CDDO-Me reveals USP7 as a novel target in ovarian cancer cells. Oncotarget. 2016;7:77096–109.

Publisher's note