Acta Pharmaceutica Sinica B 2026;16(3):1219−1232

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

TOOLS

PhenoModel: A multimodal phenotypic drug design foundation model for discovering novel potential inhibitors of multiple cancer cells

Shihang Wang[a,†], Qilei Han[a,†], Weichen Qin[b,†], Lin Wang[a,c,†], Junhong Yuan[a,d], Fengyu Cai[b], Yiqun Zhao[e], Pengxuan Ren[a,f], Yunze Zhang[a], Yilin Tang[b], Ruifeng Li[b], Zongquan Li[a,b], Wenchao Zhang[b], Shenghua Gao[e,*], Fang Bai[a,b,g,*]

[a]*Shanghai Institute for Advanced Immunochemical Studies and School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China*
[b]*School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China*
[c]*Institute of Systems Medicine, Chinese Academy of Medical Sciences, Suzhou 215028, China*
[d]*Department of Pediatric Surgery, Children's Hospital of Fudan University, Shanghai 201102, China*
[e]*Department of Computer Science, University of Hong Kong, Hong Kong SAR 999077, China*
[f]*School of Pharmacy, Shanxi Medical University, Taiyuan 030001, China*
[g]*Shanghai Clinical Research and Trial Center, Shanghai 201210, China*

Received 2 April 2025; received in revised form 1 August 2025; accepted 24 August 2025

KEY WORDS

Phenotypic drug discovery;
Cellular morphological profiles;
Cell painting;
Artificial intelligence;
Contrastive learning;
Molecular representation;
Chemical space;

Abstract Phenotypic drug discovery (PDD) focuses on the observable traits or phenotype of cells or organisms in response to drug treatment, rather than relying primarily on specific molecular targets. Drugs discovered through this approach may have better therapeutic relevance, as they are tested in conditions that closely mimic human disease. In this study, we present PhenoModel, a multimodal molecular foundation model developed using our unique dual-space contrastive learning framework. This model effectively connects molecular structures with phenotypic information. PhenoModel is applicable to a range of downstream drug discovery tasks, including molecular property prediction and active molecule screening based on targets, phenotypes, and ligands. Our results demonstrate that PhenoModel outperforms baseline methods in these areas. Building from this model, PhenoScreen is developed to

This article is part of special issue entitled: Machine Learning in Drug Discovery published in Acta Pharmaceutica Sinica B.
*Corresponding authors.
E-mail addresses: gaosh@hku.hk (Shenghua Gao), baifang@shanghaitech.edu.cn (Fang Bai).
†These authors made equal contributions to this work.
Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.09.036

Virtual screening

successfully identify several phenotypically bioactive compounds against osteosarcoma and rhabdomyosarcoma cell lines. These findings highlight the versatility of PhenoModel and its potential to accelerate drug discovery by uncovering novel therapeutic pathways and expanding the diversity of viable drug candidates.

1. Introduction

Phenotypic drug discovery (PDD) is a classic and powerful strategy in drug development[1-3]. Traditionally, drug discovery methods relied heavily on target-based drug discovery (TDD), whose focus was on identifying compounds that interact with a specific biological target[4]. However, phenotype-based approaches do not require prior knowledge about the target. Instead, PDD focuses on identifying compounds that elicit desired phenotypic effects in biological systems, such as cell morphology changes, disease state reversal, or specific signaling pathway modulation[5]. This allows the identification of potential therapeutic compounds that act on novel or multiple targets, making PDD particularly valuable for diseases with unknown molecular mechanisms or limited treatment options. A recent retrospective analysis highlighted that most first-in-class drugs originated from PDD rather than TDD, highlighting the importance of PDD in discovering innovative therapeutics for complex diseases[4,6]. PDD offers several advantages over TDD[4]. First, it provides a broader exploration of biological systems and can identify multi-target drugs (polypharmacology), which often enhance therapeutic efficacy. Furthermore, PDD is well-suited for discovering first-in-class drugs by targeting unknown or novel mechanisms, which is crucial for addressing complex or poorly understood diseases. Liu et al.[7] developed a method for pooling exogenous perturbations, thereby enabling phenotypic screens with information-rich readouts and advancing both drug discovery efforts and fundamental biological research. However, PDD also faces many challenges, including the difficulty in target identification and the high cost of large-scale screening efforts. Advances in high-content imaging, such as the Cell Painting technique, have provided new tools for capturing cellular phenotypes, but real-world screening is still resource-intensive, and efficient computational methods are needed to process and analyze phenotypic data on a large scale[8].

AI, particularly deep learning, has been pivotal in analyzing complex phenotypic data, including high-content screening (HCS) images, transcriptomic profiles, and multi-omics datasets, enabling the discovery of potential drug candidates even without well-defined molecular targets. Employing diverse computational biology methods to aid early drug discovery can help researchers explore a wider chemical space that is currently beyond the reach of experimental methods[9].

AI-driven approaches have the potential to accelerate drug discovery by integrating phenotypic information with molecular representation, enabling faster and more accurate predictions of compound efficacy[10]. For example, Tong et al.[11] proposed the TranSiGen model based on a variational autoencoder to represent phenotypic information related to the cellular environment and compound effects, as well as phenotype-based drug redirection. Ni et al.[12] proposed the PertKGE model based on the knowledge graph, and applied it to the phenotype- and target-based drug discovery process, which led to the discovery of a variety of novel hit compounds. Recent studies have also shown that cellular morphological profiles can be complementary to compound structure data in predicting drug efficacy. Integrating these phenotypic data with machine learning and deep learning techniques has become an emerging field of research in drug discovery[1]. For instance, the Broad Institute's Cell Painting technique, which captures high-content cellular images to analyze cell morphology, has proven effective in revealing drug mechanism of action (MOA), predicting bioactivity, and assessing toxicity[1,13-19]. The Klambauer lab initially used high-throughput microscopy imaging data and convolutional neural networks to predict the results of biological assays for different compounds[20]. Fredin Haslum et al.[16] have developed a deep learning based high-throughput cell image analysis method to predict bioactivities of different compounds. Recently, Dee et al.[19] applied the SwinV2 model to predict the mechanism of action of 10 kinase inhibitor compounds directly from Cell Painting images. When integrated cellular morphological profiles are used with AI models, this approach can enhance molecular representation by incorporating cellular phenotypic responses alongside structural data, offering a more holistic view of a compound's biological effects. In recent years, the Broad Institute has released multiple Cell Painting datasets, completed in collaboration with various pharmaceutical companies and academic institutions. These datasets contain a vast amount of cellular microscopy images and omics analyses, covering hundreds of thousands of compounds as well as perturbations from gene knockouts and overexpressions[21,22], providing an opportunity to explore a powerful deep learning framework to develop phenotypic-based drug screening or even generation methods.

The chemical space is vast, making it impossible to fully explore in search of bioactive molecules for specific bioassays. This remains one of the most challenging issues that needs to be addressed. Some studies are starting to emerge to explore the strategies to combine molecular structure with other characteristics related to phenotypic bioactivities to improve virtual screening[20]. For example, CLOOME and MIGA have employed contrastive learning to combine molecular structures and cellular phenotypes, demonstrating improved performance in predicting biological activity and mechanisms of action[17,18]. Overall, the application of machine learning and deep learning techniques in morphological analysis has greatly enriched phenotypic drug discovery, ranging from identifying the MOA of small molecules, optimizing lead compounds, and predicting toxicological effects. However, these methods often depend heavily on high-quality cell images, which are often scarce, and may struggle to generalize well across different cell types. According to our limited

knowledge, no experimentally validated AI-based computational method for cell morphology-infused phenotypic screening has been reported to date.

Phenotypic screening remains a bottleneck in early drug discovery due to high experimental costs and limited throughput. To address this, we propose a multimodal molecular foundation model, PhenoModel, based on our specifically designed dual-space contrastive learning framework to bridge the gap between molecular structure and phenotypic information. By jointly training a molecular graph encoder and a QFormer-based ViT image encoder, PhenoModel learns unified compound representations that capture both chemical and phenotypic perturbation signals, enhancing its capability to predict compound efficacy across multiple targets and disease models. Through extensive testing, validation, and application studies, PhenoModel demonstrates robust generalization across multiple targets, disease models, and chemical scaffolds. By unifying molecular structure and phenotypic response in a single framework, PhenoModel not only accelerates hit discovery but also establishes a platform for future enhancements, such as integrating multi-omics datasets or primary-cell screening, paving the way for more accessible, mechanism-driven, and cost-efficient drug design.

2. Materials and methods

2.1. Collection of datasets

In this study, we used the cpg0012 dataset released by the Broad Institute for molecule-image contrastive learning[22]. This dataset comprises paired perturbation molecules and microscope images from experiments conducted on 406,384-well plates. For each well, six different viewpoints of cell microscope images were taken, resulting in a total of 30,616 different perturbation molecules and 919,265 five-channel cell microscope images captured from human osteosarcoma cells (U2OS). After conducting data quality control and excluding control group images, the dataset was refined to 30,611 molecules and their corresponding 757,934 microscope images with a resolution of 692 pixels $\times$ 520 pixels. To streamline subsequent processing, all cell images were stored in npz format and divided into training, validation, and test sets with an 8:1:1 split, ensuring that multiple images from the same sample were kept within the same set.

For molecule-molecule contrastive learning, we employed a dataset from our previous study, which includes 8,000,000 pairs of intermolecular conformational space similarity (CSS) metrics used for training the molecule graph encoder[23].

2.2. Target-based virtual screening benchmark

DUD-E and LIT-PCBA are two commonly used compound screening datasets in drug design, utilized to evaluate and test the performance of virtual screening algorithms[24,25]. DUD-E is a dataset extensively used for benchmarking molecular docking and virtual screening algorithms, covering a variety of different biological targets[24]. It includes a large number of active compounds and corresponding inactive compounds (decoy molecules). These decoy molecules are specifically designed to have similar physical properties to the active compounds, but are chemically distinct enough to ensure that the testing algorithms can differentiate between active and inactive compounds. LIT-PCBA is a large compound library containing activity test results from the PubChem BioAssay database[25]. This dataset includes multiple bioactivity tests, providing researchers with extensive resources for cheminformatics and virtual screening studies. Unlike the intentionally designed negative compounds in DUD-E, the inactive molecules in LIT-PCBA come from experimental test results, presenting greater classification challenges. Therefore, testing the capability of screening algorithms in LIT-PCBA is a more challenging task. Due to their scale and diversity, both datasets are frequently used to test and improve the effectiveness of virtual screening technologies in drug discovery.

To ensure fair testing, we follow the approach adopted in previous studies, using ligands provided in the official PDB templates as reference molecules[23,26]. For each target in the DUD-E dataset, there is only one reference molecule, with a positive to negative sample ratio of about 1:50; for LIT-PCBA, there may be multiple reference molecules, with a positive to negative sample ratio of about 1:1000. In the benchmark testing for LIT-PCBA, each target has several lead compounds that can serve as reference molecules. We calculate the *Pearson* similarity between all reference molecules and the query molecules, and select the highest similarity value as the final score for each query molecule.

For virtual screening tasks, we select AUROC (Area Under the Receiver Operating Characteristic curve) and $EF_{1\%}$ (Enrichment Factor at 1%, Eq. (1)) as evaluation metrics.

$$EF_{1\%} = \frac{\text{Number of true positives within the top 1\%}}{\text{Total number of true positives} * 1\%} \tag{1}$$

AUROC is an important index for evaluating the generalization performance of models, focusing on overall screening performance, while $EF_{1\%}$ specifically targets the early enrichment rate, which is crucial for identifying highly effective compounds early in the screening process.

2.3. Phenotype-based virtual screening benchmark

We have collected 73 cell phenotype datasets from the National Cancer Institute Development Therapeutics Program (NCI/DTP) as a multi-target virtual screening benchmark test set. These datasets contain molecular activity test results from 10 different cell lines, including breast cell line (8 datasets: NCI_BT-549, NCI_HS578T, NCI_MCF7, NCI_MDA-MB-231/ATCC, NCI_MDA-MB-435, NCI_MDA-N, NCI_T-47D and NCI-ADR-RES), central nervous system cell line (8 datasets: NCI_SF-268, NCI_SF-295, NCI_SF-539, NCI_SNB-19, NCI_SNB-75, NCI_SNB-78, NCI_U251 and NCI_XF498), colon cell line (9 datasets: NCI_COLO205, NCI_DLD-1, NCI_HCC-2998, NCI_HCT-116, NCI_HCT-15, NCI_HT29, NCI_KM12, NCI_KM20L2 and NCI_SW-620), leukemia cell line (8 datasets: NCI_CCRF-CEM, NCI_HL-60(TB), NCI_K-562, NCI_MOLT-4, NCI_P388, NCI_P388/ADR, NCI_RPMI-8226 and NCI_SR), melanoma cell line (9 datasets: NCI_LOX-IMVI, NCI_M14, NCI_M19-MEL, NCI_MALME-3M, NCI_SK-MEL-2, NCI_SK-MEL-28, NCI_SK-MEL-5, NCI_UACC-257 and NCI_UACC-62), non-small cell lung cell line (11 datasets: NCI_A549/ATCC, NCI_EKVX, NCI_HOP-18, NCI_HOP-62, NCI_HOP-92, NCI_LXFL529, NCI-H226, NCI-H23, NCI-H322M, NCI-H460 and NCI-H522), ovarian cell line (6 datasets: NCI_IGROV1, NCI_OVCAR-3, NCI_OVCAR-4, NCI_OVCAR-5, NCI_OVCAR-8, NCI_SK-OV-3), prostate cell line (2 datasets: NCI_DU-145 and NCI_PC-3), renal cell line (10 datasets: NCI_786-0, NCI_A498, NCI_ACHN, NCI_CAKI-1, NCI_RXF393, NCI_RXF-631, NCI_SN12C, NCI_SN12K1,

NCI_TK-10 and NCI_UO-31), and small cell lung cell line (2 datasets: NCI_DMS114 and NCI_DMS273). The average number of test molecules per dataset is 43 k, and the average number of active molecules is 2.8 k. Given that the results of bioassays may be influenced by synergistic effects across different targets, and multiple active molecule data are often considered in the actual drug development process, we randomly select 50 molecules as reference active molecules for each bioassay. We use the same benchmark testing approach as adopted with LIT-PCBA.

2.4. Molecular property predictions benchmark

Establishing predictive models for molecular properties through molecular representation learning is a crucial component of ligand-based drug design methods. To further evaluate the representation capabilities of PhenoModel, we conducted experimental tests on multiple molecular benchmarks from the Therapeutics Data Commons (TDC)[27]. TDC is a large-scale collection of chemical datasets specifically designed to advance the development of machine learning methods in drug chemistry, materials science, and biology. These datasets cover various types of chemical data, such as the ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) molecular properties datasets. To ensure the generalizability of molecular representation methods, the training, validation, and testing sets for all molecular property modeling tasks are divided based on molecular scaffolds. For all datasets from TDC, we use the default scaffold splitting method provided by TDC, reserving 20% of the data as the test set. Since ADMET properties, compared to molecular activity, represent a higher scale of "phenotype" that often requires consideration of more molecular structural information, this task poses a greater challenge for molecular representation methods that have not utilized extensive molecular pre-training.

For molecular property prediction tasks, AUROC and Spearman's rank correlation coefficient (Spearman's ρ) are chosen as the evaluation metrics for classification and regression tasks, respectively. AUROC assesses the ability of the model to correctly rank compounds based on their likelihood of having a desired property, which is vital for classification. Spearman's ρ measures the strength and direction of a monotonic relationship between predicted and actual values in regression tasks, providing a non-parametric measure of rank correlation.

During the baseline comparison, for Uni-Mol and FP-GNN, we directly employed the officially provided molecular property prediction scripts. For CombineFP, the various molecular fingerprints, and MIGA, we first pre-extracted molecular features (for MIGA, using its supplied pretrained model miga_graphtrans_256.pth), and then applied our custom-designed decoders alongside the suite of models available in AutoGluon to select the optimal architecture and perform hyperparameter tuning. Finally, we retained the best-performing result for each ADMET task and compared it against the other methods.

2.5. The design of the dual-space contrastive learning strategy

The core of PhenoModel is a carefully designed dual-space contrastive learning representation model, which integrates molecules chemical space with their perturbated cellular morphology space, as illustrated in Fig. 1. This representation model consists of four main modules: two feature extraction modules, the Molecule Graph Encoder and the Cell ViT Encoder, which capture information from molecular structures and cell images, respectively; a primary molecule−molecule contrastive learning module to establish relationships within the chemical space of different molecules; and a step-wise contrastive learning module that links a molecule's chemical space with its associated cellular morphology space. This representation model can then be seamlessly integrated with a variety of drug design tasks, including zero-shot tasks such as virtual screening for active compounds, ADMET prediction, and other related applications.

2.5.1. Molecular graph encoder

The Molecule Graph Encoder employs a four-layer Weisfeiler−Lehman Network (WLN) to extract molecular features, building on our previous work with GeminiMol[23,28], a hybrid contrastive learning framework that incorporates conformational space similarities into molecular representation learning for ligand-based drug design tasks.

The molecule is represented as a graph $G = (V, E)$. The initial feature of each node $v \in V$ is denoted $h_v^{(0)} = x_v$ (*e.g.*, atom type, charge), and the feature of each edge $(v, w) \in E$ is denoted e_{vw} (*e.g.*, bond type, aromaticity). The Weisfeiler-Lehman (WL) algorithm enhances graph representations by iteratively refining the molecular graph through the aggregation of information from neighboring nodes (atoms) and edges (bonds). The iteration process can be expressed as Eq. (2).

$$h_v^{(t+1)} = \mathrm{COMBINE}\left(h_v^{(t)}, \mathrm{AGGREGATE}\left(\left\{h_u^{(t)} : u \in N(v)\right\}\right)\right) \quad (2)$$

Here, $h_v^{(t)}$ denotes the feature representation of node v at the t-th iteration, $N(v)$ denotes the set of neighbors of node v, AGGREGATE is the function that aggregates neighbor features, and COMBINE is the function that updates the current node's features. This process distinguishes between non-isomorphic graphs, which is crucial for capturing both local and global structural information and for accurately representing molecules with similar topologies but different chemical properties.

The process begins by converting SMILES (Simplified Molecular Input Line Entry System) notation into a graphical representation, which is then processed by the WLN. The molecular encoder consists of four WLN layers, with the node dimension set to 2048. After processing by the WLN, the 2048-dimensional features undergo further processing by a readout function. This function is designed to aggregate the node features into a single vector that represents the entire molecule. This vector captures the essential molecular features necessary for downstream tasks such as virtual screening and molecular property prediction.

The use of WLN allows for effective learning of the molecular graph's topology and atom-level interactions, making it a robust choice for capturing complex molecular characteristics essential in drug discovery and other chemical informatics applications.

2.5.2. Cell ViT encoder

The Cell ViT Encoder, based on the Vision Transformer (ViT), is designed to extract information from cell images[29]. ViT excels in capturing cell images by utilizing self-attention mechanisms to process entire images, enabling it to grasp both global context and intricate cellular morphology with high flexibility. The computation of the attention layer can be expressed as Eq. (3):

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

where Q, K, and V denote the Query, Key, and Value matrices, respectively. Its adaptability, scalability, and ability to leverage pretraining from large datasets make it highly effective for

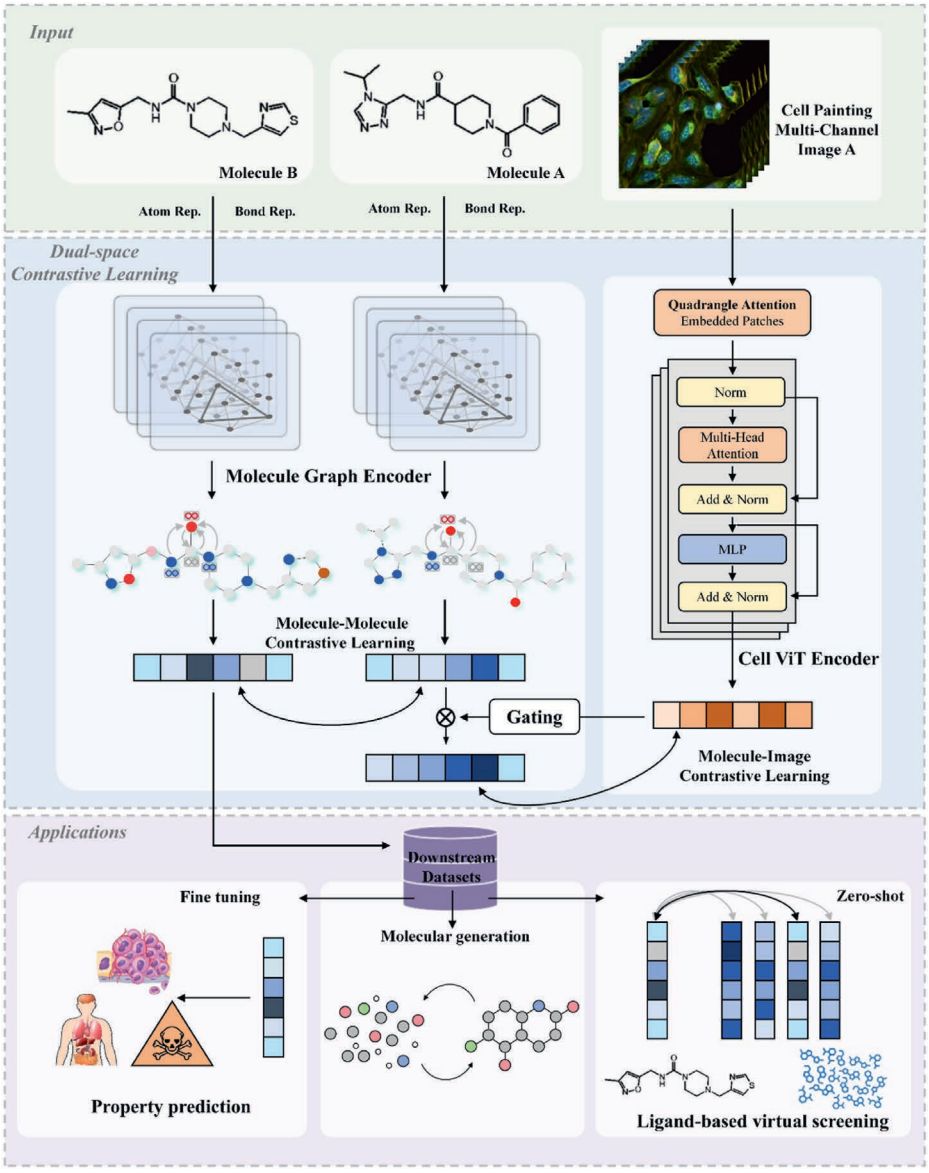

Figure 1 An illustration of the PhenoModel architecture and its potential application in various downstream tasks. PhenoModel consists of a molecule graph encoder, a Cell ViT Encoder, a primary molecule-molecule contrastive learning module; and a step-wise molecule-image contrastive learning module. This representation model can be then seamlessly integrated with a variety of drug design tasks, including zero-shot tasks such as virtual screening for active compounds, ADMET prediction, and other related applications.

biological applications, such as phenotype classification and multimodal learning in drug discovery. The encoder's parameters were initially pretrained on cellular type classification and further optimized using quadrilateral attention.

We randomly cropped each 520 pixels × 696 pixels cell image to 512 pixels × 512 pixels and applied data augmentation techniques such as rotation before inputting them into QFormer for feature extraction. QFormer aims to address the limitations of traditional window-based attention mechanisms in handling objects of varying sizes, shapes, and orientations. Specifically, QFormer employs a novel Quadrangle Attention (QA) mechanism, which uses an end-to-end learnable quadrangle regression module to predict transformation matrices. These matrices convert the default window into a target quadrangle for token sampling and attention computation. By leveraging QFormer, we aim to mitigate the impact of dispersed effective information and noise in cell images on feature extraction.

We first utilized a classification task to train the ViT model based on QFormer to learn the correspondence between different cell images and various drug molecules. This classification task was designed using cell images processed with 30 k different drug molecules collected in the cpg0012 dataset. We designed the network parameters based on QFormer-tiny and set the patch size to 8 to avoid padding zeros in the 512 pixels × 512 pixels images. Finally, we achieved an *ACC*-1 of 26.1% and an *ACC*-5 of 53.0% on the 30 k-class task. After training a robust feature extractor, we froze the main parameters of the ViT model, using it as the image encoder in subsequent contrastive learning tasks.

2.5.3. Molecule-image contrastive learning module

Given the fundamental differences between chemical and cellular spaces, and the large number of parameters in the respective encoders, the molecular encoder requires substantial weight updates to align its embeddings with those of the cellular images. This can lead to network instability and degraded learning performance. To maintain stability, we fixed most of the weights in the Cell ViT image encoder, limiting variations in image embeddings. During the contrastive learning stage between molecules and cell images, each molecule-image pair was transformed into a 2048-dimensional embedding, with InfoNCE loss applied to minimize the distance between the embeddings of positive samples in the feature spaces.

We employed a contrastive learning strategy to align features extracted from molecular structures and cell images. Contrastive learning enhances the model's ability to differentiate by optimizing the similarity between positive sample pairs and maximizing the differences between negative sample pairs. In this study, we utilized the InfoNCE loss function (Eq. (4)):

$$\mathscr{L}_{\text{InfoNCE}} = -\text{Log}\frac{\exp\left(\text{sim}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{N}\exp\left(\text{sim}\left(z_i, z_j\right)/\tau\right)} \tag{4}$$

where sim(z_i, z_j) denotes the similarity measure between features z_i and z_j, typically calculated using cosine similarity (Eq. (5)). The temperature parameter $\tau$is used to adjust the smoothness of the probability distribution, and N represents the number of negative samples.

$$\text{Cosine}_{(z_i, z_j)} = \frac{z_i \cdot z_j}{\|z_i\| \times \|z_j\|} \tag{5}$$

In each batch, the image features and molecular features initially correspond one-to-one. During the training process, these features are cross-paired to generate positive and negative samples. To create training samples, this study randomly selects half of the categories as positive samples, while the remaining categories are considered negative samples. Additionally, we implement a gating mechanism where the 2048-dimensional cell image features, encoded by a ViT, serve as feature weights. After normalization with a sigmoid function, these weights are multiplied by the molecular feature vectors to produce gated molecular feature vectors. These gated features are then used in contrastive training with the cell image features.

2.5.4. Molecule-molecule contrastive learning module

To improve the efficiency of integrating chemical and cellular phenotypic spaces, we introduced molecule-to-molecule contrastive learning. This approach captures molecular interactions by leveraging similarities in chemical space and potential bioactivity, adding a constraint that allows the molecular encoder to update its weights while preserving its ability to learn conformational space similarities. For molecule-to-molecule contrastive learning, we inherited the framework from GeminiMol, which incorporates 3D conformational ensembles and pharmacophore characteristics into the embeddings, effectively linking molecular structure with bioactive information. Ultimately, this strategy achieves balanced and stable weight updates between the two networks.

The molecular graph encoder consisted of 4 WLN graph layers, and the node dimensionality was set to 2048. Molecules are converted into molecular graphs and then subsequently encoded independently. The projection head of the module employed in our study was designed to transform the 2048-dimensional encoding into meaningful CSS predictions. The network consists of a series of fully connected layers with nonlinear activation functions and additional normalization to improve its learning ability. Two encoding vectors of query and reference molecules were input to the projection heads of CSS descriptors.

2.6. Cell culture

The U2OS human osteosarcoma cell line (RRID: CVCL_0042) and RD human Rhabdomyosarcoma cell line (RD-RMS, RRID: CVCL_1649) were cultured in DMEM (Meilunbio, #MA0212, Dalian, China) supplemented with 10% fetal bovine serum (ExCell Bio, #FSP500, Suzhou, China), 1% penicillin/streptomycin (Gibco, #15140122, Grand Island, NY, USA), and 0.1% mycoplasma prevention reagent (TransGen, #FM501, Beijing, China). The K-562 chronic myelogenous leukemia cell line (RRID: CVCL_0004) was cultured in RPMI-1640 medium (Meilunbio, #MA0548, Dalian, China) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin (Gibco), and 0.1% mycoplasma prevention reagent (TransGen). All cells were maintained at 37 °C under 5% CO_2 in a humidified atmosphere.

2.7. Cell viability assay

The U2OS, RD-RMS, and K-562 cells were seeded at a concentration of 5000 cells/100 μL, 4500 cells/100 μL, and 5000 cells/100 μL per well into 96-well plates, respectively. After 24 h incubation, the cells were exposed to DMSO or the compounds with a final DMSO concentration of 0.2%, and cultured for another 48 h. Cell viability was assessed using Cell Counting Kit-8 (CCK8) (TargetMol Chemicals Inc., Boston, MA, USA). The absorbance was determined at 450 nm with the EnSpire Multimode PlateReader (PerkinElmer, Envision, Waltham, MA, USA). The percentage of surviving cells was calculated from the absorbance ratio of compounds treated to untreated cells.

2.8. Cell painting and high-content imaging

The Cell Painting staining was performed according to a modified protocol by Bray et al.[13] using the PhenoVue Cell Painting kit (Revvity, #PING11, Waltham, MA, USA). Briefly, cells were seeded in 384-well microplates (Corning, #3764BC, Kennebunk, ME, USA) at a density of 2000 cells/30 μL per well. The culture medium was replaced with 40 μL of 2% FBS in DMEM with DMSO or compounds 24 h after seeding. The final DMSO concentration was 0.2%. After a 48 h treatment, the PhenoVue 641 Mitochondrial Stain (500 nmol/L) final concentration) was added. The plates were incubated for 30 min in the dark at 37 °C. Cells were then fixed using 16% paraformaldehyde (Alfa Aesar, #043368.9 L; 4% final concentration) at RT for 20 min, followed by four washes with 1 × HBSS (Gibco, #14,065,056). Subsequently, cells were permeabilized with 0.1% Triton X-100 and stained with a staining solution containing the following dyes at their respective final concentrations: PhenoVue Fluor 555-WGA (43.7 nmol/L), PhenoVue Fluor 488-Concanavalin A (48 nmol/L), PhenoVue Fluor 568-Phalloidin (8.25 nmol/L), PhenoVue Hoechst 33,342 Nuclear stain (1.62 μmol/L), PhenoVue 512 Nucleic acid stain (6 μmol/L) in the 1 × PhenoVue Dye Diluent A (HBSS + 1% BSA). After a final incubation period of 30 min at RT, the plates were washed four times with 1 × HBSS prior to imaging.

Cell imaging was conducted using the Operetta High Content Imaging System (PerkinElmer) equipped with a 20 × water-immersion objective. Five channels were employed to visualize the various fluorescent stains: DNA (excitation: 360−400 nm; emission: 410−480 nm), ER (excitation: 460−490 nm; emission: 500−550 nm), RNA (excitation: 460−490 nm; emission: 560−630 nm), AGP (excitation: 560−580 nm; emission: 585−605 nm), and Mito (excitation: 620−640 nm; emission: 650−760 nm). Four fields of view were captured per well.

2.9. Data available

We have released the datasets model weight used in this study, which are publicly available at Zenodo (https://zenodo.org/records/13943032) and Hugging Face (https://huggingface.co/Sean-Wong/PhenoScreen). The cell painting dataset for model training is available at: http://gigadb.org/dataset/100351 [22]. All the source codes are publicly available through GitHub (https://github.com/Shihang-Wang-58/PhenoScreen).

3. Results and discussion

3.1. Establishment of multimodal molecular foundation model by dual-space contrastive learning

In this study, we proposed PhenoModel, which is a dedicatedly designed dual-space contrastive learning model, and integrates molecules' chemical space with their perturbed cellular morphology space, as illustrated in Fig. 1. This foundation model consists of four main modules: two feature extraction modules, the Molecule Graph Encoder and the Cell ViT Encoder; a primary molecule−molecule contrastive learning module; and a step-wise molecule-image contrastive learning module. Detailed information on the parameter settings used for PhenoModel is provided in Supporting Information Table S1. To evaluate PhenoModel, we performed modular ablation on two recognized target-based virtual screening benchmark test sets, DUD-E and LIT-PCBA[24,25], and evaluated the overall active molecule screening ability (AUROC) and early enrichment ability ($EF_{1\%}$) of PhenoModel. In this process, molecular embedding was used to calculate the similarity between the reference molecule and the remaining molecules, thereby identifying the active molecule, which helps to understand the specific contributions of the strategies used in constructing PhenoModel.

We analyzed the contributions of the molecule-image contrastive learning module, the molecule-molecule contrastive learning module, and the gating mechanism. As shown in Table 1, the version of PhenoModel supplemented with cell image information (molecule-image contrastive learning module) outperformed models that lacked this information across all metrics, indicating that cell image information indeed enhances molecular representation. The screening results of active molecules for each target in LIT-PCBA are shown in Tables 2 and 3. Notably, the performance of the model significantly declined when dual-space contrastive training was not utilized (without the molecule-molecule contrastive learning module). This decline could be attributed to changes in the molecular encoder's parameters during contrastive learning, as it aligns the molecule embeddings with image embeddings, potentially losing its original encoding ability. Thus, joint training maintains the molecular encoder's intrinsic capability to effectively encode molecules while supplementing it with molecular activity information from cell images, resulting in improved representation performance of PhenoModel. Additionally, removing the gating mechanism during contrastive training also led to a loss in performance to some extent. These results demonstrate the necessity of each module in PhenoModel.

Additionally, we evaluated the impact of different image encoders on the model's contrastive training, as illustrated in Supporting Information Fig. S1. When using ResNet18 and ViT as image encoders, the different images corresponding to the same molecule were not effectively clustered together after encoding and dimension reduction. However, after incorporating Qformer Attention, multiple cell images corresponding to the same compound exhibited a much tighter distribution in the feature space. This indicates that the Cell ViT Encoder allows the image encoder to more accurately extract cellular features from cell images, enhancing the overall efficacy of the encoding process.

3.2. PhenoModel outperforms the baseline methods on molecular property prediction

The dynamic properties of molecules, such as membrane permeability, are one of the important factors affecting drug efficacy and cellular phenotype. In the process of new drug development, the *in vivo* properties of drug molecules, including absorption, distribution, metabolism, excretion, and toxicity, abbreviated as ADMET properties, are important indicators of druggability in the development of new drugs. By learning the characteristic patterns of different molecules in specific tasks, models can be established that play a crucial role in various drug design processes, such as predicting molecular activity and various ADMET properties.

We compared PhenoModel with GeminiMol[23], Uni-Mol[30], FP-GNN[31], MIGA[17], various combined molecular fingerprints (CombineFP, combined by ECFP4 [32], FCFP6 [32], AtomPairs[33], and Topological torsion[34]), and individual types of molecular fingerprints, testing 22 ADMET property prediction tasks from the TDC dataset that feature public leaderboards, including 13 classification tasks and 9 regression tasks. The results, as shown in Fig. 2A and B, Supporting Information Tables S2 and S3, indicate that PhenoModel's predictive performance is comparable to Uni-Mol,

Table 1 Performance of the PhenoModel variants on DUD-E and LIT-PCBA.

Variant	DUD-E		LIT-PCBA	
	AUROC	$EF_{1\%}$	AUROC	$EF_{1\%}$
PhenoModel	0.757	18.060	0.625	7.099
w/o molecule-image contrastive learning module	0.734	17.718	0.597	6.565
w/o molecule-molecule contrastive learning module	0.695	13.139	0.562	3.795
w/o gating	0.754	15.989	0.578	5.728

Table 2 AUROC of the different models on LIT-PCBA.

Target	PhenoModel	GeminiMol	PhaseShape	ECFP4	Glide SP	PLANET
ADRB2	0.550	0.480	**0.584**	0.552	0.573	0.390
ALDH1	0.551	0.513	0.494	0.512	**0.788**	0.549
ESR1_ago	**0.870**	0.800	0.766	0.790	0.513	0.668
ESR1_ant	0.595	0.546	0.570	0.467	0.450	**0.604**
FEN1	**0.644**	0.596	0.441	0.380	0.408	0.606
GBA	0.553	0.569	0.436	0.441	0.581	**0.689**
IDH1	**0.614**	0.500	0.343	0.393	0.216	0.503
KAT2A	0.601	**0.607**	0.526	0.440	0.490	0.498
MAPK1	**0.653**	0.621	0.558	0.534	0.558	0.612
MTORC1	0.467	0.459	0.427	0.450	**0.490**	0.467
OPRK1	0.585	0.650	0.498	**0.697**	0.528	0.455
PKM2	**0.701**	0.697	0.592	0.605	0.669	0.563
PPARG	0.762	**0.768**	0.747	0.743	0.670	0.705
TP53	0.584	0.608	0.512	0.435	**0.640**	0.587
VDR	**0.643**	0.549	0.420	0.429	0.462	0.441
Average	**0.625**	0.597	0.528	0.525	0.536	0.556

Bold indicates the highest performance among all methods on this target.

which is based on large-scale pre-training using the ChEMBL database. ADMET represents a higher-level "phenotype" data, which in most cases probably cannot be directly linked to molecular properties at the level of cellular changes. Therefore, PhenoModel did not show a significant improvement over the extensively pre-trained Uni-Mol in predicting ADMET properties, but it still displayed competitive performance and exceeded baseline methods in predictions of metabolism and excretion properties. However, PhenoModel isn't pre-trained at the level of hundreds of millions of molecules like Uni-Mol. Meanwhile, compared with MIGA, which also used cell image information for pre-training, PhenoModel significantly outperforms MIGA across all evaluated metrics, reflecting the advantages of the dual-space contrast training approach. We believe that PhenoModel has been able to represent molecules relatively comprehensively and may show greater advantage in tasks that are more closely related to cellular phenotypes.

3.3. PhenoModel surpasses baseline models in phenotype-based virtual screening

The phenotype information of cell images is included in the molecular image contrastive learning module and plays an important role in the PhenoModel model. Considering that cellular phenotypes generally reflect the phenotypic effects of molecules in biological systems or at the level of multi-target modulation. Therefore, we attempted to explore how cellular phenotype information complements the molecular representation capabilities of PhenoModel. To this end, we collected 73 cell phenotype activity experimental datasets from the National Cancer Institute Development Therapeutics Program (NCI/DTP). These datasets come from 73 different cell lines of 10 types of cancer, including breast cell lines, central nervous system cell lines, colon cell lines, leukemia cell lines, melanoma cell lines, non-small cell lung cell lines, ovarian cell lines,

Table 3 $EF_{1\%}$ of the different models on LIT-PCBA.

Target	PhenoModel	GeminiMol	PhaseShape	ECFP4	Glide SP	PLANET
ADRB2	17.643	17.643	**18.750**	11.762	5.555	5.882
ALDH1	2.509	2.438	1.618	2.312	**10.336**	1.381
ESR1_ago	**10.879**	9.791	8.324	0.000	7.106	7.687
ESR1_ant	2.364	2.336	0.980	0.000	0.409	**3.883**
FEN1	**7.583**	3.921	0.815	0.560	0.602	5.149
GBA	3.703	3.124	3.029	**4.907**	4.209	3.011
IDH1	**5.127**	2.563	0.000	5.127	0.140	2.564
KAT2A	1.579	0.000	2.590	0.523	1.507	**3.108**
MAPK1	**2.295**	1.974	2.278	0.990	1.289	1.297
MTORC1	1.030	**5.150**	0.000	1.030	0.842	2.060
OPRK1	0.000	4.165	0.000	4.165	**4.166**	4.165
PKM2	7.873	8.605	4.403	**8.788**	4.020	1.831
PPARG	**31.114**	26.669	22.798	22.224	3.879	3.660
TP53	**7.634**	6.107	5.063	1.527	2.369	2.500
VDR	**5.158**	3.986	2.152	2.817	0.433	1.021
Average	**7.099**	6.565	4.853	4.449	3.124	3.280

Bold indicates the highest performance among all methods on this target.

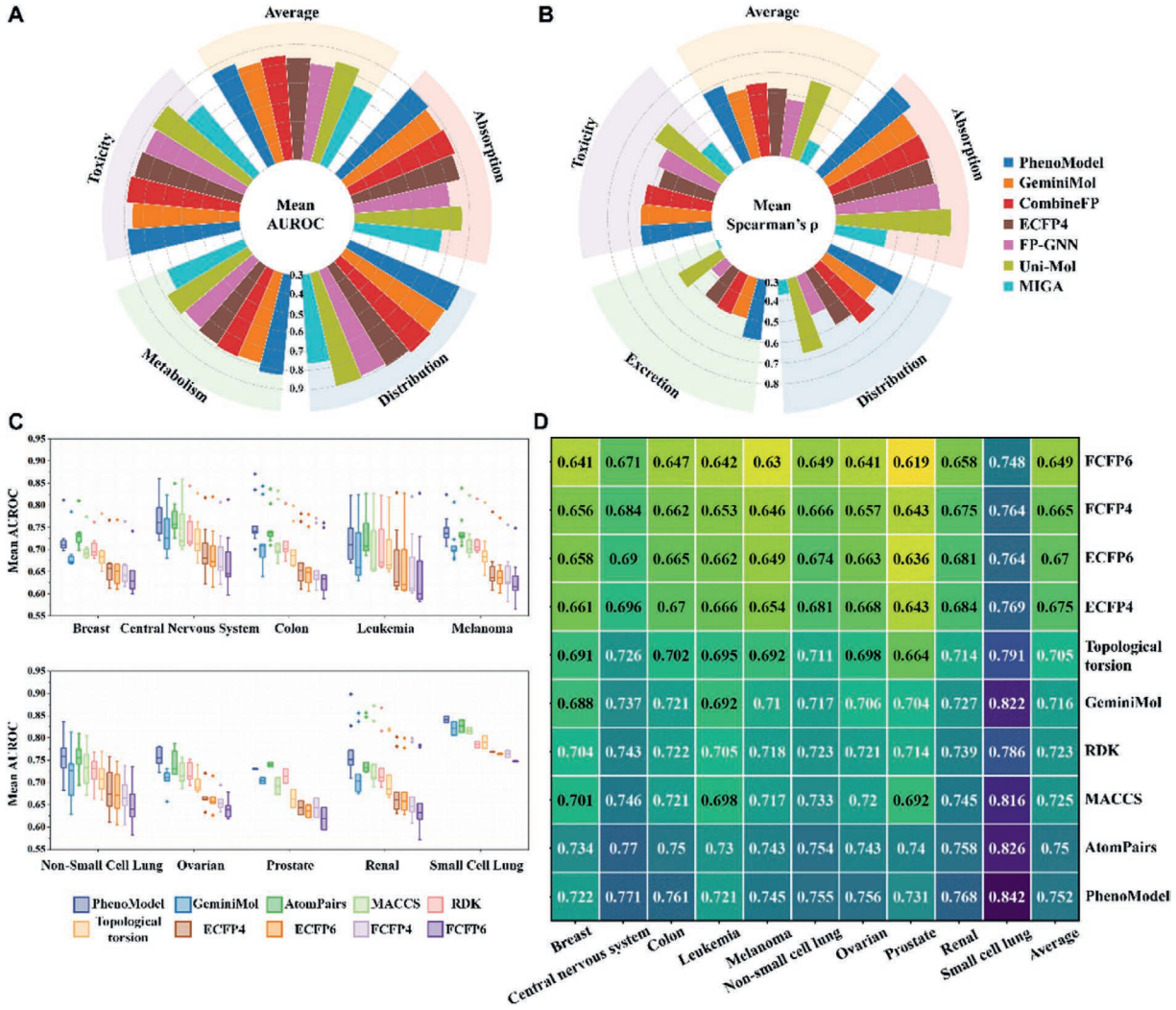

Figure 2 A comparison of PhenoModel with various baseline methods in the tasks of predicting molecular properties and virtual screening. (A) AUROC of thirteen Therapeutics Data Commons (TDC) ADMET classification tasks. (B) Spearman's ρ of nine TDC ADMET regression tasks. (C) AUROC box plots of all cell lines in each cancer type. (D) The horizontal axis represents ten different types of cancer. The vertical axis represents the different types of molecule representation methods. The values in each square of the heatmap represent the average AUROC value of active molecule screening for this method in all cell lines of the denoted type of cancer.

prostate cell lines, renal cell lines, and small cell lung cell lines. The average number of test molecules per dataset is 43 k, and the average number of active molecules is 2.8 k. Given that cell activity may be influenced not just by a single target but also by the synergistic effects of multiple targets, this task presents greater challenges.

We selected eight different molecular fingerprints, and our group previously developed a conformational ensemble-based molecular representation method GeminiMol, and PhenoModel for benchmark testing comparison. During the testing process, all molecules were independently encoded by the corresponding method, and the activity of molecules against various targets was distinguished by calculating the similarity between query molecules and reference molecules. Considering that changes in cell phenotype may be influenced by multiple mechanisms of action, we randomly select multiple reference molecules for each biological assay. Finally, we calculated the average AUROC of these methods across all cell lines included in each type of cancer. As illustrated in Fig. 2C and D, Supporting Information Fig. S2 and Table S4, PhenoModel showed outstanding performance in screening active molecules across all ten types of cancer cell lines, especially compared to the GeminiMol model, which only contains molecule structure information for training, indicating the enhancement of the molecular encoder by cell phenotype information.

3.4. PhenoModel also excels in target-based screening of active compounds

Meanwhile, we tested the performance of the method over the two important virtual screening tasks in drug design: DUD-E and LIT-PCBA[24,25]. DUD-E contained 102 targets and LIT-PCBA contained 15 targets for evaluating the performance of the virtual screening algorithms. All molecules were independently encoded by the PhenoModel model, and the activity of molecules against various targets was distinguished by calculating the similarity

between query molecules and reference molecules. For fair testing, we adopted the approach used in previous studies, selecting ligands provided in the official PDB templates as reference molecules. Notably, in scenarios assessing virtual screening capabilities, LIT-PCBA presents a more challenging task compared to DUD-E, primarily because for each target in DUD-E, the ratio of active to inactive molecules is relatively fixed with considerable structural variation. However, such "standardized" tasks as seen in DUD-E are rare in drug discovery. In contrast, there are typically challenging tasks like LIT-PCBA, especially given the drug activity cliffs, complex mechanisms of action, and significant variations in the proportions of active compounds across different tasks. Therefore, LIT-PCBA is considered the main virtual screening benchmark dataset in our study.

We compared the PhenoModel with multiple baselines, including (I) GeminiMol, which is based on the similarity comparison of molecular conformation ensembles; PhaseShape, a method based on 3D pharmacophore and shape alignment; various molecular fingerprints (including ECFP4[32], ECFP6[32], FCFP4[32], FCFP6[32], MACCS[35], RDK[36], AtomPairs[33], and Topological torsion[34]) represented by ECFP4; (II) the classic molecule docking method Glide SP[37]; and (III) PLANET[26], a drug−target affinity (DTA) prediction-based method. The results indicate that PhenoModel performed well in both the DUD-E and LIT-PCBA tasks (Tables 2 and 3, Supporting Information Tables S5 and S6). For the 102 targets in DUD-E, PhenoModel achieved an AUROC over 0.9 for 32 targets and over 0.8 for 57 targets, with an average AUROC of 0.803, significantly outperforming other baseline methods (Table S4).

Fig. 3 illustrates the visualization of molecular encoding by PhenoModel compared to ECFP4 across various targets, showing that PhenoModel significantly distinguishes between active and inactive molecules in feature space, with most active molecules closely matching the reference molecules in distribution, whereas some active molecules encoded by ECFP4 are further away, indicating that PhenoModel effectively captures the distinctions in features between active and inactive molecules.

For the more challenging LIT-PCBA task (Tables 2 and 3, and S6), PhenoModel's overall screening performance and early enrichment rates surpassed those of various baseline methods in eight out of fifteen targets. Unlike Glide SP and PLANET, PhenoModel does not utilize any target protein information, which further demonstrates its superior performance in molecular representation. Additionally, compared to GeminiMol, PhenoModel showed higher performance in the virtual screening tasks for active molecules over most of the studied targets, suggesting that the active information in cell phenotypes indeed enhances molecular representation capability and the ability to screen active molecules.

3.5. Identification of novel inhibitors aligned with varying drug-induced cellular phenotypes via a PhenoModel-based PDD screening pipeline, PhenoScreen

To assess the ability of PhenoModel for practical drug design applications, we developed an active compound screening pipeline called PhenoScreen to further screen other molecules with similar activities but novel scaffolds according to the known active compounds. We collected seven compounds from the Ardigen PhenAID database (https://phenaid.ardigen.com/) that can affect the phenotype of the U2OS cell line and identified four molecules from literature reports that showed inhibitory effects on osteosarcoma (details of the reference active compounds are provided in Supporting Information Table S7). With the prior information obtained from known phenotypes and activities reported in the literature, we expected to screen more novel molecules with similar or better phenotypes using PhenoScreen. We then conducted a screening of these active reference molecules using PhenoScreen in the Specs compound library (https://www.specs.net/), which comprises 335,000 compounds, following the workflow illustrated in Fig. 4A. This led to the identification of compounds **A1**−**A15**. We measured the cellular activity of the screened compounds using the CCK8 assay (Fig. 4B and C) and found that most of the screened molecules exhibited inhibitory activity at the cellular level. We further observed the effect of the

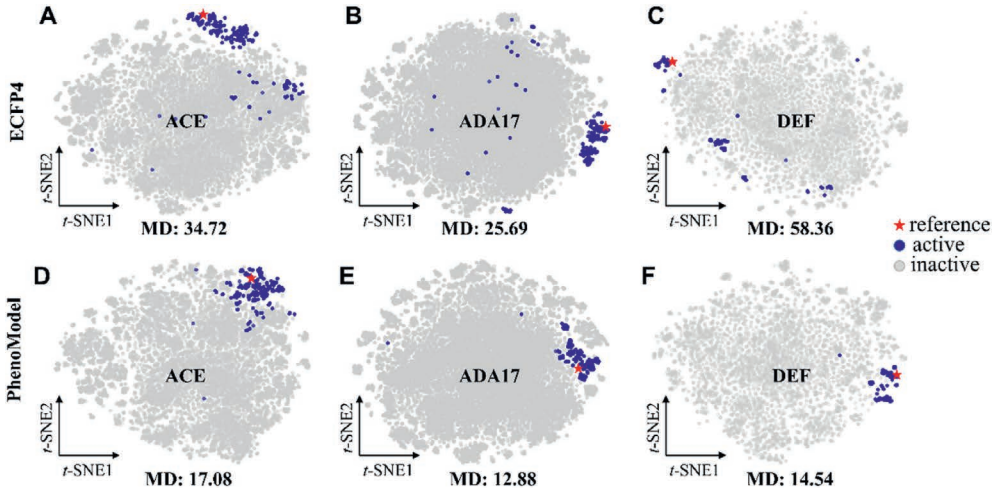

Figure 3 Visualization of molecular representations for targets ACE, ADA17 and DEF *via* t-SNE. MD represents the mean distance between all active molecules and the reference molecule after feature dimensionality reduction. (A−C) Molecules encoded by ECFP4. (D−F) Molecules encoded by PhenoModel. The active molecules encoded by ECFP4 is further away from the reference molecule in feature space compared to PhenoModel. This suggests that our method may be more effective in molecular representation.

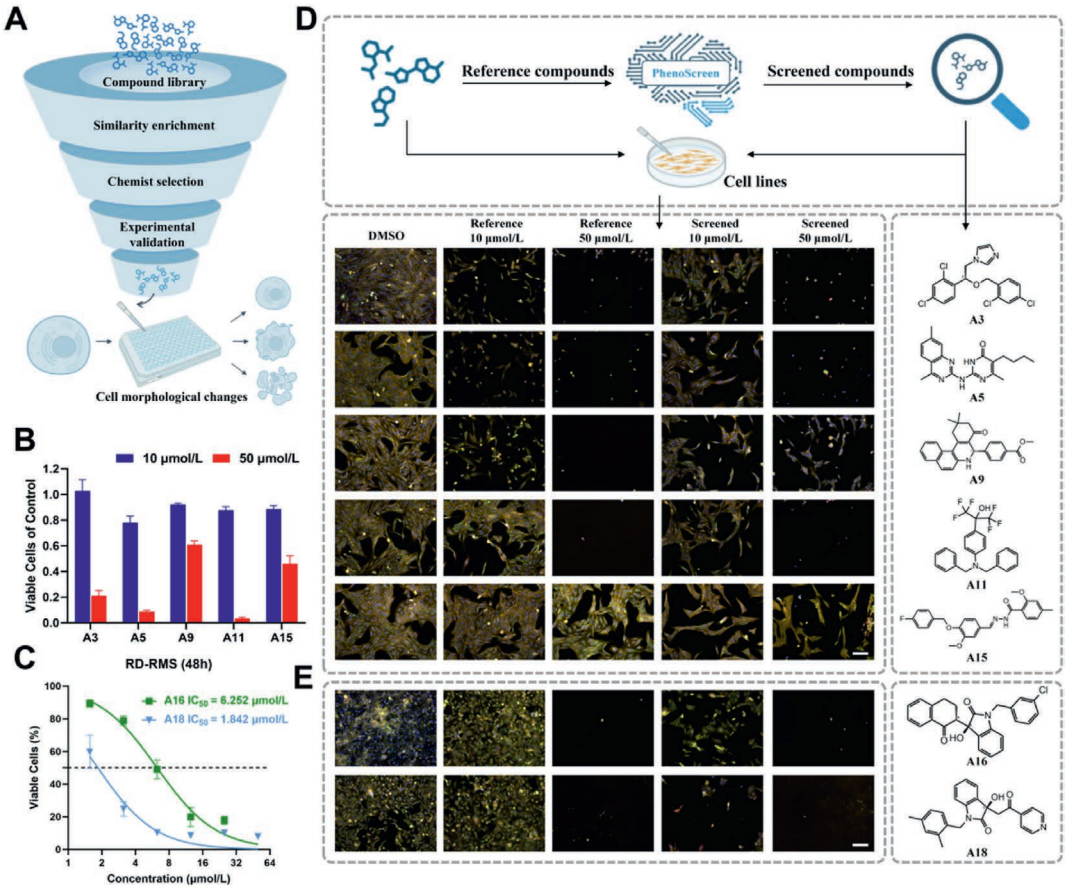

Figure 4 The process and experimental results of using PhenoModel for phenotype-based drug screening. (A) Similarity calculations are performed between compound libraries and predefined reference molecules to enrich highly similar molecules. Chemists then select from the highly similar molecules and conduct experimental verification to observe the inhibitory ability and morphological changes of the compounds at the cellular level. (B) Cell viability assay results of molecules screened for U2OS. (C) The half-maximal inhibitory concentration (IC_{50}) of the identified bioactive compounds **A16** (6.252 μmol/L, CI: 5.590 to 6.994 μmol/L) and **A18** (1.842 μmol/L, CI: 1.470 to 2.148 μmol/L) against rhabdomyosarcoma. (D) Cell phenotype results of cell painting experiments for molecules screened in the U2OS cell line showed that phenotypic changes induced by **A3**, **A5** and **A9** were similar to those of the reference molecules. In contrast, **A11**, and **A15** exhibited more pronounced changes in cellular phenotypes compared to the reference molecules, indicating higher potential inhibitory activities. Scale bar = 100 μm. (E) Cell phenotype results of cell painting experiments for active molecules screened for rhabdomyosarcoma. **A16** and **A18** showed more strong changes in cellular phenotypes compared to the reference molecules, indicating higher potential inhibitory activity. Scale bar = 100 μm.

compounds on the cellular phenotype by cell painting assay, and surprisingly, the results of the cell painting assay showed that the phenotypes of the **A3**, **A5** and **A9** were close to the reference molecules, while **A11** and **A15** showed higher inhibitory activity at the cellular level as compared to the reference molecules (Fig. 4D). These molecules can be further investigated as potential drugs for the treatment of osteosarcoma.

Furthermore, we investigated whether the screening capabilities of PhenoScreen could be generalized to diseases other than osteosarcoma. Based on preliminary studies, we identified QC6352 and ML324 from the literature as the potential therapeutic agents for rhabdomyosarcoma, and indisulam for leukemia. Using PhenoModel, we performed a similarity screen in the commercial compound libraries, which led to the identification of compounds **A16**−**A22**. The Cell painting and CCK-8 assay results showed that these molecules have a stronger inhibitory ability compared to reference compounds (Fig. 4E and Supporting Information Fig. S3). These results suggest that molecules such as **A16**, **A18**, **A21**, and **A22** have potential as therapeutic drugs for rhabdomyosarcoma and leukemia, demonstrating the generalizability of PhenoScreen for active molecule screening.

In summary, PhenoScreen can identify potential molecules with similar phenotypes, and in some cases, even stronger inhibitory activity, by capturing the phenotypic information of molecules and their perturbations. Thus, PhenoScreen serves as an excellent hit discovery method that can be effective across various diseases. In practical applications, since reference molecules often undergo modifications to enhance activity, PhenoScreen leverages both structural and phenotypic information to further improve activity and optimize the pharmacokinetic properties of molecules. Additionally, PhenoScreen can be applied to scaffold hopping of active molecules, thereby exploring more new active molecules.

3.6. A public online web server for efficient virtual screening and target identification

To facilitate the implementation of the proposed PhenoModel model by researchers, an easy-to-use web server has been established at https://bailab.siais.shanghaitech.edu.cn/services/PhenoModel/. The web server supports a compound similarity screening feature. Users can upload the SMILES string of an active compound of interest, choose from supported molecular libraries, and the website will return the top three thousand molecules ranked by similarity and their scores. Considering the diversity of molecular structures in the molecular library and the complexity of synthesis, we provide several molecular libraries such as Specs (more than 335 k compounds), LifeChemicals (more than 616 k compounds), and Chemdiv (more than 1.7 M compounds). The screening time for a single reference molecule in any given library is approximately 1 min.

Moreover, analyzing the target mechanisms of small molecules identified through phenotype-based screening can enrich the strategies for discovering new targets. Therefore, we have constructed a benchmark dataset called DTIDB, based on previous processing of the BindingDB database. This dataset can help researchers identify potential targets of query molecules by searching for known active compounds similar to the query molecules and determining their targets. We believe that these two applications of PhenoModel will assist in advancing more drug discovery research.

4. Conclusions

In this study, we proposed a dual-space-based phenotypic drug discovery foundation model, named PhenoModel, which integrates the information from molecular structure space and bioactive phenotypic space. PhenoModel not only fully characterizes the conformational information of molecules but also incorporates the active information embedded in cell phenotypes. The benchmark results presented in this paper demonstrate that PhenoModel's representation of molecules plays a significant role in virtual screening tasks. It exhibits performance comparable to other specialized models in molecular property prediction tasks. Additionally, PhenoModel also performs exceptionally well in identifying active molecules, both in single-target tasks such as DUD-E and LIT-PCBA, and in various NCI cell line bioassays based on multi-target activity tests. We also developed an active molecules screening pipeline named PhenoScreen based on PhenoModel to pinpoint phenotypically bioactive compounds effective against various cancer cell lines, and successfully found multiple highly active lead compounds, demonstrating its potential in extensive applications.

Of course, we must acknowledge that there is substantial room for improvement in PhenoModel. On one hand, compared to the vast chemical space, PhenoModel's training dataset contains 30,000 individual compound structures, which may not be sufficient for tasks requiring a comprehensive consideration of various molecular structures, such as particular molecular properties prediction. Inevitably, it may not perform much better in related benchmark tests compared to extensively pre-trained molecular representation models like Uni-Mol. While PhenoModel shows robust generalization on small-molecule related tasks, its pre-training dataset contains limited examples of macrocycles, cyclic peptides, and other low-frequency scaffolds. Addressing this gap by incorporating scaffold-diverse datasets will be a key focus of our future efforts to extend PhenoModel's applicability across the full breadth of chemical space. We are able to incorporate additional datasets, such as the cpg0016 dataset, the Tahoe-100 M dataset, and macrocyclic compounds with more unique structures, to expand the coverage of chemical space in future work. On the other hand, the information contained in cell phenotype data may not fully support PhenoModel in tasks that require the representation of higher-dimensional molecular information. The current version of PhenoModel was trained exclusively on cell images from the U2OS cell line. Although we tested its performance on other cancer cell lines, we cannot guarantee consistent results across all cell types or a broader range of cell lines. This is because phenotypic signatures may differ in primary cells or other tissue types, which could limit generalizability. To address this limitation, future work will incorporate a more diverse set of cell backgrounds, including primary cell data, to improve the model's robustness and applicability. Also, it might be possible to enhance PhenoModel's representational capabilities by incorporating data such as the dynamics of molecules in organoid systems. Additionally, cell image data inevitably contains noise due to experimental errors, and how to obtain cleaner and richer phenotypic data is one of the urgent problems to be solved in all related research.

It is noteworthy that changes in transcriptomic information are also crucial in phenotype-based drug design. In future research, interacting molecular, cell image, and transcriptomic information may further enhance the model's ability to represent molecules. Additionally, care must be taken not to compromise the original encoding ability of the molecular Encoder when integrating phenotypic information into the molecular representation model; hence, strategies like joint training may be required to mitigate potential risks, especially for such multimodal models. We plan to further enhance PhenoModel's representational capabilities in our subsequent studies. We believe that with the addition of more task-related training data, PhenoModel could exhibit superior performance in common drug design tasks such as virtual screening and property prediction, and play a significant role in areas like genetic-chemical perturbation association modeling and phenotype-based molecule generation. This would provide researchers with valuable tools to accelerate the drug discovery process.

Acknowledgments

This work was supported by Shanghai Science and Technology Development Funds (Grant IDs: 24JS2850200, 24JS2850100, and 24YF2729200), ShanghaiTech AI4S Initiative SHTAI4S202404, National Key R&D Program of China (Grant IDs: 2022YFC3400501 and 2022YFC3400500), start-up package from ShanghaiTech University, and Shanghai Frontiers Science Center for Biomacromolecules and Precision Medicine at ShanghaiTech University. The authors appreciate the technical support provided by the high-performance computing cluster of ShanghaiTech University. The authors thank the Discovery Technology Platform at Shanghai Institute for Advanced Immunochemical

Studies, ShanghaiTech University for instruments and technical support regarding High-content imaging experiments.

Author contributions

Shihang Wang: Conceptualization, Model development, Benchmark test, Data curation, original draft writing. Qilei Han: Experiment test. Weichen Qin: Model development. Lin Wang: Conceptualization, benchmark test. Junhong Yuan: Experiment test. Fengyu Cai: Model development. Yiqun Zhao: Model development. Pengxuan Ren: Data curation. Yunze Zhang: Data curation. Yilin Tang: Web server development. Ruifeng Li: Benchmark test. Zongquan Li: Data curation. Wenchao Zhang: Data curation. Shenghua Gao: Image processing algorithm design, original draft writing. Fang Bai: Conceptualization, algorithm design, manuscript drafting. All authors contributed to reviewing and editing the manuscript.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2025.09.036.

References

1. Tang QS, Ratnayake R, Seabra G, Jiang Z, Fang RG, Cui LA, et al. Morphological profiling for drug discovery in the era of deep learning. *Briefings Bioinf* 2024;**25**:bbae284.
2. Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: recent successes, lessons learned and new directions. *Nat Rev Drug Discov* 2022;**21**:899−914.
3. Malandraki-Miller S, Riley PR. Use of artificial intelligence to enhance phenotypic drug discovery. *Drug Discov Today* 2021;**26**:887−901.
4. Sadri A. Is target-based drug discovery efficient? Discovery and "Off-Target" mechanisms of all drugs. *J Med Chem* 2023;**66**:12651−77.
5. Baig MH, Ahmad K, Roy S, Ashraf JM, Adil M, Siddiqui MH, et al. Computer aided drug design: success and limitations. *Curr Pharm Des* 2016;**22**:572−81.
6. Calado CRC. Bridging the gap between target-based and phenotypic-based drug discovery. *Expet Opin Drug Discov* 2024;**19**:789−98.
7. Liu N, Kattan WE, Mead BE, Kummerlowe C, Cheng T, Ingabire S, et al. Scalable, compressed phenotypic screening using pooled perturbations. *Nat Biotechnol* 2024;**43**:1324−36.
8. Bai F, Li SL, Li HL. AI enhances drug discovery and development. *Natl Sci Rev* 2024;**11**:nwad303.
9. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell* 2020;**180**:688−702.
10. Moshkov N, Bornholdt M, Benoit S, Smith M, McQuin C, Goodman A, et al. Learning representations for image-based profiling of perturbations. *Nat Commun* 2024;**15**:1594.
11. Tong XC, Qu N, Kong XT, Ni SK, Zhou JY, Wang K, et al. Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery. *Nat Commun* 2024;**15**:5378.
12. Ni S, Kong X, Zhang Y, Chen Z, Wang Z, Fu Z, et al. Identifying compound-protein interactions with knowledge graph embedding of perturbation transcriptomics. *Cell Genomics* 2024;**4**:100655.
13. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* 2016;**11**:1757−74.
14. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, et al. Data-analysis strategies for image-based cell profiling. *Nat Methods* 2017;**14**:849−63.
15. Moshkov N, Becker T, Yang K, Horvath P, Dancik V, Wagner BK, et al. Predicting compound activity from phenotypic profiles and chemical structures. *Nat Commun* 2023;**14**:1967.
16. Fredin Haslum J, Lardeau C-H, Karlsson J, Turkki R, Leuchowius K-J, Smith K, et al. Cell Painting-based bioactivity prediction boosts high-throughput screening hit-rates and compound diversity. *Nat Commun* 2024;**15**:3470.
17. Zheng SJ, Rao JH, Zhang JX, Zhou LY, Xie JC, Cohen E, et al. Cross-modal graph contrastive learning with cellular images. *Adv Sci* 2024;**11**:e2404845.
18. Sanchez-Fernandez A, Rumetshofer E, Hochreiter S, Klambauer G. CLOOME: contrastive learning unlocks bioimaging databases for queries with chemical structures. *Nat Commun* 2023;**14**:7339.
19. Dee W, Sequeira I, Lobley A, Slabaugh G. Cell-vision fusion: a swin transformer-based approach for predicting kinase inhibitor mechanism of action from cell painting data. *iScience* 2024;**27**:110511.
20. Hofmarcher M, Rumetshofer E, Clevert DA, Hochreiter S, Klambauer G. Accurate prediction of biological assays with high-throughput microscopy images and convolutional networks. *J Chem Inf Model* 2019;**59**:1163−71.
21. Chandrasekaran SN, Ackerman J, Alix E, Ando DM, Arevalo J, Bennion M, et al. JUMP cell painting dataset: morphological impact of 136,000 chemical and genetic perturbations. *bioRxiv* 2023. Available from: https://doi.org/10.1101/2023.03.23.534023.
22. Bray MA, Gustafsdottir SM, Rohban MH, Singh S, Ljosa V, Sokolnicki KL, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the cell painting assay. *GigaScience* 2017;**6**:1−5.
23. Wang L, Wang S, Yang H, Li S, Wang X, Zhou Y, et al. Conformational space profiling enhances generic molecular representation for AI-Powered ligand-based drug discovery. *Adv Sci* 2024;**11**:e2403998.
24. 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.
25. Tran-Nguyen VK, Jacquemard C, Rognan D. LIT-PCBA: an unbiased data set for machine learning and virtual screening. *J Chem Inf Model* 2020;**60**:4263−73.
26. Zhang XY, Gao HT, Wang HJ, Chen ZH, Zhang Z, Chen XC, et al. PLANET: a multi-objective graph neural network model for protein−ligand binding affinity prediction. *J Chem Inf Model* 2023;**64**:2205−20.
27. Huang KX, Fu TF, Gao WH, Zhao Y, Roohani Y, Leskovec J, et al. Artificial intelligence foundation for therapeutic science. *Nat Chem Biol* 2022;**18**:1033−6.
28. Coley CW, Barzilay R, Jaakkola T, Green WH, Jensen KF. Prediction of organic reaction outcomes using machine learning. *ACS Cent Sci* 2017;**3**:434−43.
29. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. *ArXiv* 2020. Available from: https://doi.org/10.48550/arXiv.2010.11929.
30. Zhou G, Gao Z, Ding Q, Zheng H, Xu H, Wei Z, et al. Uni-Mol: a universal 3D molecular representation learning framework. In: *International conference on learning representations*; 2023.
31. Cai HX, Zhang HM, Zhao DC, Wu JX, Wang L. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. *Briefings Bioinf* 2022;**23**:bbac408.
32. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;**50**:742−54.
33. Carhart RE, Smith DH, Venkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J Chem Inf Comput Sci* 1985;**25**:64−73.

34. Nilakantan R, Bauman N, Dixon JS, Venkataraghavan R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J Chem Inf Comput Sci* 1987;**27**:82−5.
35. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* 2002;**42**: 1273−80.
36. Landrum G. "RDKit: open-source cheminformatics.". Available from: https://www.rdkit.org.
37. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;**47**: 1739−49.