Methods

Gene dependency analysis using DepMap

The gene dependency analytics was performed from the DepMap release 25Q2. To assess the off-target landscape, gene dependency probabilities for each gene were averaged over all cell models and the mean values were plotted

Post-marketing drug severity score

A drug severity (DS) score was defined to estimate the overall post-marketing severity associated with each drug, integrating both the intensity and the fatality of its safety signals while penalizing for concomitant drug use. We used the Individual Case Safety Reports (ICSRs) in the SafetyVista platform (SafetyVista, Chemotargets SL, 2025, https://safety-vista.com/). It includes data from four spontaneous reporting systems (SRS), namely, FAERS, VAERS, VigiBase and JADER. All SRS sources were standardised, de-duplicated and corrected for masking effects, such as those arising from the massive reporting during the COVID-19 pandemic (Montes-Grajales et al. 2023)

For each drug, we first identified all signals of disproportionate reporting (SDRs). An adverse event was considered an SDR if it had 5 or more reports and a value higher than 1 for the lower limit of the 95% confidence interval of proportional reporting ratio (PRR05).

Then for each drug-SDR pair we calculated the reporting frequency (RF), as well as the proportion of death-related reports (D) related to each event. The mean product of these two measures was then divided by the concomitance factor (C), defined as the mean number of drugs reported together with the drug in individual case safety reports. The resulting DS values were normalized between 0 and 1 to allow comparability across drugs.

$$DS = log\left(\frac{mean(RF * D)}{C}\right)$$

This definition increases the DS score for drugs with frequent and severe (death-associated) safety signals, while decreasing it for drugs whose reports are heavily confounded by concomitant medications.

Systematic drug-target mapping from multiple databases

Bioactivity data were downloaded as csv files from ChEMBL (version 35) (Zdrazil et al. 2024), BindingDB (1.3.2025 update) (Liu et al. 2025), and Drug Target Commons (accessed March 2025) (Tanoli et al. 2018). A molecule with bioactivity of 1μM or more is considered as inactive towards a target. The following filters were applied on ChEMBL data before the download: Target Organism: Homo sapiens; Molecule Max Phase: Early Phase 1, Phase 1, Phase 2, Phase 3, Approved; Protein Classification L3: Protein Kinase; Target Type: Single protein. After accessing the dataset, we obtained the relevant, CDK-related activities by applying a string search on the Target Name column for “cyclin”, followed by manual filtering of target names. Rows with standard relation of “<” or “>” were removed as the exact activity value is not available. However, if the activity was equal to or greater than 1μM, rows with standard relation “>” were kept due to the molecule being inactive regardless. Rows labeled with “Outside typical range”, “Uncertain”, and “Undetermined”, or with a potential duplicate

flag were removed. Nan activity values were filled with 1μM in the presence of a specific comment stating inactivity; otherwise they were removed. The assay confidence scores of the resulting ChEMBL data were 8 or 9 out of 9, indicating high confidence target mapping. For all three databases, only human proteins with activity type IC_{50}, EC_{50}, K_i or K_d were considered. For BindingDB and DTC, we required each target based on "UniProt (SwissProt) Recommended Name of Target Chain" and "target_pref_name" to be a CDK, respectively. We mapped BindingDB Monomer IDs to ChEMBL IDs and removed all instances with no matching ChEMBL ID. To avoid redundant data, we mapped all compounds to their parent form based on ChEMBL compound information. Finally, we concatenated all three data frames, computed the geometric mean of duplicate compound-target rows and applied a negative log10 transformation to yield 2,727 unique, CDK-related bioactivity data points. For each CDK, we separated the available bioactivities based on trial phase status; approved drugs in one group and investigational compounds (phases I-III) in another. The groups were further divided by their activity with a threshold of 1μM to form a two-by-two contingency table for testing. The trial phase information of each compound was extracted from ChEMBL (version 35). It should be noted that the trial status of a compound may have changed since.

ML-based drug-target affinity prediction modeling

Data collection and preparation

To construct benchmark datasets for building regression models that predict quantitative bioactivities of compounds against Casein kinase 2, we collected the drug structures and a wide variety of compound-target bioactivities, including IC_{50}, K_d, K_i values from the ChEBML, BindingDB and DrugTargetCommon. Then, we extracted their canonical SMILES using "rdkit.Chem package" in Python and removed the duplicates in each bioactivity type. Finally, 1282 compounds with IC_{50} values, 361 compounds K_i and 119 compounds K_d values were left.

Generation of compounds features

To construct an effective regression model, we generated and compared different chemical fingerprints for the compounds. Here, five commonly used fingerprints, including standard fingerprints (1024 bits), estate fingerprints (79 bits), graph fingerprints (1024 bits), MACCS fingerprints (166 bits), and substructure fingerprints (307 bits) were generated using the "rcdk" package in R. The impact of various fingerprints on model prediction accuracy were investigated to get the best input features for the regression model.

Chemical diversity of the datasets

It is widely recognized that the chemical diversity of a target activity dataset is a crucial factor for building a robust and comprehensive model for predicting compound-target binding affinities. The molecular descriptors including molecular weight (MW), octanol-water partition coefficient (ALogP), topological surface area (TPSA) were calculated using "rdkit.chem.descriptors" package in Python.

Construction of bioactivity prediction models (BPM): To predict the bioactivity value of various compounds against CK2, we used the compound fingerprints (standard fingerprint) as the input features in five machine learning algorithms, including random forest (RF), artificial neural network (ANN), support vector machine (SVM), extreme gradient boosting (XGBoost) and gradient boosted decision tree (GBDT), accessible via the scikit python

implementation (http://www.scikit-learn.org/). Then, we used the following metrics to score the accuracy of quantitative bioactivity value predictions:

(1) Pearson correlation coefficient between the predicted and actual values, which quantifies the linear relationship between the activity predictions.
(2) Spearman's rank correlation coefficient between the predicted and actual values, which quantifies the ability to rank drug pairs in correct order.
(3) Normalized Root Mean Square Error (NRMSE): a standardized version of root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_p - Y_a)^2}$$

$$\mathrm{NRMSE} = \frac{RMSE}{Y_a\,\max - Y_a\,min} \times 100\%$$

Here, Y_p is the predicted bioactivity value and Y_a is the measured bioactivity value.

The detailed information on how we select the optimal bioactivity value, and input fingerprint feature was shown in **Supplementary Material XX**.

Construction of error prediction models (EPM): Models application domain refers to the chemical space covered by the training data. To ensure reliable bioactivity predictions, only the compounds located in the bioactivity prediction model's (BPM) application domain should be used. To evaluate the model application domain, we used an error prediction model (EPM). In this process, the test set was further divided into a calibration set and a new test set using 1:1 ratio. The EPM was trained using the calibration set, and the features of EPM were the same type fingerprints (1024-bit standard fingerprint) as the BPM. The labels of EPM were shown as Equation 3, where Y_p is predicted bioactivity value from the BPM, and Y_a is actual bioactivity value of these compounds in the calibration set. The machine learning method and parameters of EPM were the same as the BPM.

$$Labels\,(EPM) = |Yp - Ya| \quad (3)$$

Then, the labels of EPM in the calibration set were defined as the set L, where the values were sorted in ascending order.

$$L = \{l1, l2, l3 \ldots ln\},\ l1 \leq l2 \leq l3 \leq \ldots \leq ln \quad (4)$$

When doing prediction in the new test set, the BPM and EPM were firstly used to get all predicted bioactivity values and error values. Then we could define a confidence level, α. The α-quantile of the L set (Q(α),Equation 5 and 6) in the calibration set would be used as a cutoff to decide which compounds would be left by EPM. The compounds whose predicted error values were less than Q(α) in the new test set would be kept and regarded as reliable predictions.

$$i = \alpha \times n \quad (5)$$

$$Q(\alpha) = \begin{cases} \frac{l_i + l_{i+1}}{2} & if\ i\ is\ an\ integer \\ l_{\lceil i \rceil} & otherwise \end{cases} \quad (6)$$

where n is the total number of compounds in the calibration set, i=α×n is the index corresponding to the α, li and li+1 are the i-th and (i+1)-th elements in the sorted label set L, and the second line in equation 6 uses the ceiling function ⌈i⌉ to take the smallest integer greater than or equal to i.

For instance, assuming five calibration samples with ascending error values L={0.1,0.2,0.4,0.6,0.8} and a confidence level α=0.6, the index is i=α×n=3.0. Since i is an integer, the quantile is calculated as Q(0.6)=(l3+l4)/2=(0.4+0.6)/2=0.5. In the subsequent prediction step, if the new test set contains five compounds with predicted error values {0.15,0.25,0.38,0.52,0.70}, those with predicted errors less than Q(α)(e.g., < 0.5) are considered reliable predictions and retained by the EPM, whereas those exceeding this cutoff are excluded.

Combined BPM and EPM for virtual screening: To virtually screen for compounds with strong affinity towards CK2,we chose a commercial compound library from TargetMol, which contains up to 12 million compounds, including different small molecular types such as FDA approved drugs, novel chemicals, natural products and endogenous metabolites. We generated the standard fingerprint of these compounds as the input features for the IC_{50} bioactivity and error predictions using the BPM and EPM models. After that, a specific Q(α) value was set to select the most accurate predicted compounds and an IC_{50} bioactivity value was used as the threshold to find the compounds with promising affinity to act as the inhibitors of CK2. Finally, Lipinski's Rule of 5 was applied to select the drug-like compounds.

Training data size and structural similarity

To explore the effects of training data size and molecular space diversity that contribute to the regression model accuracy, we calculated the Tanimoto similarity using the optimal fingerprint selected above. Structurally similar compounds were removed based on various thresholds (0.4-1.0). The datasets with different data size and compound similarity were generated and the selected regression model was trained using these datasets.

The analysis of key active substructures

To find the key chemical substructures closely related to the binding affinity, we used the "bioalerts" (https://github.com/isidroc/bioalerts) and “RDkit” packages. Firstly, in the training set, we obtained substructures using the radius of the chemical bonds as 2,3,4,5 and 6 (topological distance, number of bonds). The substructures which existed in more than 20% of the training set samples were used in the following analysis. Then in the test set, as for every obtained substructure, the compounds were divided into two groups: the group with this substructure and the group without this substructure. We used the Shapiro-Wilk test to assess whether the pIC_{50} values of these two groups fit normal distributions. If both groups follow normal distributions, we evaluated the difference of pIC_{50} between two groups by student's t test; if not, the difference was evaluated by Kolmogorov Smirnov test.

High-throughput site-specific molecular docking

Two catalytic isoforms of CK2, designated CK2α (encoded by the CSNK2A1 gene) and CK2α' (encoded by the CSNK2A2 gene) conjugated with their studied ligands were downloaded from PDB database with high resolutions (CK2α: 5CU4; CK2α': 8QBU). Using these protein structures, the solvent and other organics were removed from the downloaded proteins, followed by adding the missing hydrogens using AutoDock Tools 1.5.6 software (Eberhardt et al. 2021), and the results were exported in PDBQT format. All small molecule ligands to be docked were converted to PDBQT format and docked using the Vina software. In this process, the ligands were set as flexible structures while the proteins were considered rigid.

The binding energy of each molecular ligand was calculated using the Lamarckian genetic algorithm (Morris et al. 1998), and compounds exhibiting affinity values lower than those of the original inhibitors and effectively binding to the α-D site were screened as potential CK2 inhibitors based on the docking model. Subsequently, the optimal conformations of the selected ligand-protein complexes with the highest binding affinities were visualized using PyMOL.

In vitro CK2 kinase assay

In vitro kinase assays were conducted to evaluate the effects of compounds on CK2 activity. Reactions were performed in 384-well plates with a final volume of 10 μL, containing 1.5 ng/μL recombinant CK2 (New England Biolabs, P6010), 50 μM CK2 substrate peptide (Millipore, 12-330), and CK2 assay buffer [40 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 0.5 mM DTT, and 150 mM NaCl]. DMSO dissolved compounds were added to achieve a final DMSO concentration of 1% (500 nL per well), and the reactions were initiated by adding ATP to a final concentration of 1 μM. Plates were incubated at 37°C for 1.5 hours, after which 10 μL of Kinase-Glo Luminescent Kinase Assay reagent (Promega) was added to quantify residual ATP by luminescence. All tested compounds showed no interference with luciferase activity. IC_{50} values were calculated by fitting dose-response curves using GraphPad Prism.

Inhibition test of BRCA cell lines

Eight breast cancer cell lines with reported high CK2 expression, including CAL-120, MCF-7, BT-549, CAL-51, MFM-223, SK-BR-3, MDA-MB-468, and MDA-MB-231, were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). CAL-120, BT-549, MDA-MB-231, CAL-51, SK-BR-3 and MFM-223 cells were cultured in RPMI-1640 medium, while MCF-7 and MDA-MB-468 were maintained in DMEM. All media were supplemented with 10% fetal bovine serum (FBS), 5 mM L-glutamine and 1% penicillin-streptomycin, and cells were incubated at 37 °C in a humidified atmosphere containing 5% CO_2. For cytotoxicity assays, cells were seeded in 96-well plates at 5×10^3 cells per well, allowed to adhere overnight, and then treated with serial dilutions of Compound 2 or Compound 3 (0–400 μM) for 48 hours. Cell viability was determined by MTT assay which is reliable also for chromogenic small molecules (Wang et al. 2024), adding 10 μL of 5 mg/mL MTT solution per well for 4 hours, followed by dissolution in 100 μL DMSO and measurement of absorbance at 570 nm. IC_{50} values were calculated by fitting dose–response curves using GraphPad Prism 10.1.2.

217 kinase activity profiling

Bioactivity profiling of 217 human kinases, encompassing members of the TK, TKL, STE, CMGC, AGC, CAMK, and CK1 families, was performed using the KinaseProfiler platform (ICE Bioscience).

For this, compounds were serially diluted from 10 mM DMSO stocks, and transferred into 384-well assay plates using an Echo 655 acoustic dispenser to achieve the desired test concentrations. Kinase selectivity profiling was conducted using both HTRF and ADP-Glo functional assays.

For the HTRF assays, pre-mixed 2× ATP & substrate and 2× kinase/metal ion solutions were sequentially added (2.5 µL each) to the wells containing compounds, followed by incubation at 25 °C for 60 minutes. Subsequently, 5 µL of detection reagent containing XL665 and antibody was added and incubated for another 60 minutes, after which fluorescence signals at 620 nm and 665 nm were recorded using a BMG Labtech microplate reader.
In the ADP-Glo assays, compounds were combined with 2 µL of 2× kinase solution and incubated at 25 °C for 10 minutes, followed by addition of 2 µL of 2× ATP & substrate solution for 60 minutes. ADP-Glo reagent (4 µL) was then added and incubated for 40 minutes, followed by 8 µL of kinase detection reagent for an additional 40-minute incubation prior to luminescence measurement.

Percent inhibition was calculated according to the following equation:
% Inhibition = 100% - [(compound - positive control) / (negative control - positive control)] × 100%
where the negative control (1% DMSO) represents 0% inhibition and the positive control (10 µM reference inhibitor) corresponds to 100% inhibition.

Drug combination testing
Drugs diluted in dimethyl sulfoxide (DMSO) were dispensed at a 25-nL volume into 384-well black plates (Corning, #3864) using an Echo 550 acoustic liquid handler (Labcyte). CK2 inhibitors (Compound 2 and Compound 3) were tested in combination with PARP inhibitors (rucaparib, olaparib, and niraparib) at seven different concentrations in threefold serial dilutions, covering a concentration range appropriate for each compound. Benzethonium chloride (BzCl, 100 µM) and DMSO (0.1%) were used as positive and negative controls, respectively. SK-BR-3, MDA-MB-231, and MDA-MB-468 breast cancer cells were seeded at a density of 1,000 cells per well and incubated with drug combinations for 72 hours. Cell viability was assessed using the CellTiter-Glo Luminescent Cell Viability Assay (Promega), and luminescence was measured with a Pheras microplate reader. Bliss synergy scores were calculated using SynergyFinder 2.0, with values higher than 8 considered synergistic and negative values interpreted as antagonistic.

Analysis of structural similarity and scaffold novelty
To evaluate the structural novelty of the newly identified CK2 inhibitors (Comp2 and Comp3) relative to known CK2 inhibitors, a cheminformatics workflow was implemented in Python using RDKit (v2022.09). We collected ten commonly-reported CK2 inhibitors (Supplementary Table S2) as references and all the compounds including our two new hits (Comp2 and Comp3) were compiled and standardized. For each compound, extended-connectivity fingerprints (ECFPs) were calculated using the Morgan algorithm at two different radii (ECFP4, radius = 2; and ECFP6, radius = 3; 2048-bit). Pairwise Tanimoto similarity

coefficients were computed between each new hit and all known CK2 inhibitors, and both the maximum and mean similarity scores were recorded. Then, principal component analysis (PCA) was performed on the ECFP4 fingerprints using scikit-learn (v1.4) to visualize the overall distribution of chemical space. The first two principal components (PC1 and PC2) were plotted to generate a two-dimensional map, with known inhibitors shown as blue circles and new hits as orange stars. To assess scaffold novelty, Bemis-Murcko scaffolds were extracted for all compounds, and the scaffolds of Comp2 and Comp3 were compared with those of the reference inhibitors. Compounds that did not share any scaffold with known inhibitors were considered to possess novel structural frameworks.

For each hit, the most similar known CK2 inhibitor (based on the highest ECFP4 similarity) was selected for maximum common substructure (MCS) analysis. MCS identification was performed using the RDKit FindMCS function with “ringMatchesRingOnly” and “completeRingsOnly” constraints. The number of shared atoms and molecular coverage for both compounds were calculated, and the matched atoms were visualized as highlighted fragments on 2D molecular diagrams.

All analyses and visualizations were performed using RDKit, pandas (v2.2.2), numpy (v1.26.4), matplotlib (v3.9.2), and scikit-learn (v1.5.1).