ACS OMEGA

Open Access



http://pubs.acs.org/journal/acsodf

Article

# Support Vector Machine-Based Prediction Models for Drug Repurposing and Designing Novel Drugs for Colorectal Cancer

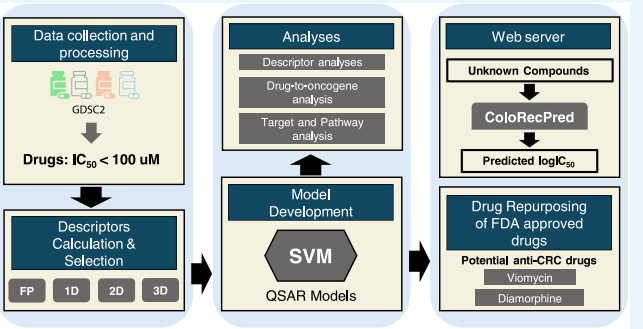
Avik Sengupta, Saurabh Kumar Singh, and Rahul Kumar*

Cite This: *ACS Omega* 2024, 9, 18584−18592

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Colorectal cancer (CRC) has witnessed a concerning increase in incidence and poses a significant therapeutic challenge due to its poor prognosis. There is a pressing demand to identify novel drug therapies to combat CRC. In this study, we addressed this need by utilizing the pharmacological profiles of anticancer drugs from the Genomics of Drug Sensitivity in Cancer (GDSC) database and developed QSAR models using the Support Vector Machine (SVM) algorithm for prediction of alternative and promiscuous anticancer compounds for CRC treatment. Our QSAR models demonstrated their robustness by achieving a high correlation of determination ($R^2$) after 10-fold cross-validation. For 12 CRC cell lines, $R^2$ ranged from 0.609 to 0.827. The highest performance was achieved for SW1417 and GP5d cell lines with $R^2$ values of 0.827 and 0.786, respectively. Further, we listed the most common chemical descriptors in the drug profiles of the CRC cell lines and we also further reported the correlation of these descriptors with drug activity. The KRFP314 fingerprint was the predominantly occurring descriptor, with the KRFPC314 fingerprint following closely in prevalence within the drug profiles of the CRC cell lines. Beyond predictive modeling, we also confirmed the applicability of our developed QSAR models via *in silico* methods by conducting descriptor-drug analyses and recapitulating drug-to-oncogene relationships. We also identified two potential anti-CRC FDA-approved drugs, viomycin and diamorphine, using QSAR models. To ensure the easy accessibility and utility of our research findings, we have incorporated these models into a user-friendly prediction Web server named "ColoRecPred", available at https://project.iith.ac.in/cgntlab/colorecpred. We anticipate that this Web server can be used for screening of chemical libraries to identify potential anti-CRC drugs.




## INTRODUCTION

Colorectal cancer (CRC) holds the third-highest position in terms of worldwide incidence and the second-highest rank in mortality rate with 10% and 9.4%, respectively, across the globe, as per GLOBOCAN 2020,[1] notably affecting both males and females.[1] Some widely used drugs approved for treating CRC are cetuximab, oxaliplatin, 5-FU, and tucatinib.[2] To combat the increasing specter of drug resistance, clinicians and researchers have banked on combinational drug therapies, such as CAPOX, FOLFIRI, FU-LV, and XELOX, among others.[2] Even though this strategic shift toward combination therapy has been useful in many CRC treatments, other strategies need to be introduced continuously to combat CRC.[3] The mechanism of drug resistance is being studied widely and many mechanisms have been recently identified.[4] The primary mechanisms include reduced drug activation, an aberration in downstream signaling processes, drug transport aberration, and changes in drug targets.[4] This emphasizes the clinical importance of augmenting the currently available drug arsenal to enhance the therapeutic methodologies for CRC. Implementing machine learning-based *in silico* methods, e.g., Quantitative Structure Activity Relationship (QSAR) models, is an attractive approach to bypass the time- and cost-exhaustive traditional drug discovery process.[5] The *in silico* methods can be used to screen large chemical libraries to predict novel drugs for CRC and boost drug discovery and development.[5] There is a continuous effort to improve the drug arsenal for CRC worldwide. Recently, many new targets have been used for drug development using 3D-QSAR models. The recently published drug targets are interleukin-6 and DNA topoisomerase II, where 3D-QSAR models have been developed for CRC to find drugs that show anticancer activity.[6,7] To address QSAR model specificity limitations, recent years have seen the evolution of PTML models, combining perturbation theory with machine learning, as detailed by Planche and Cordeiro et al. These models effectively handle complex data sets in drug discovery, proteomics, and nanotechnology, for anticancer studies across

Received: February 6, 2024
Revised: March 28, 2024
Accepted: March 29, 2024
Published: April 9, 2024

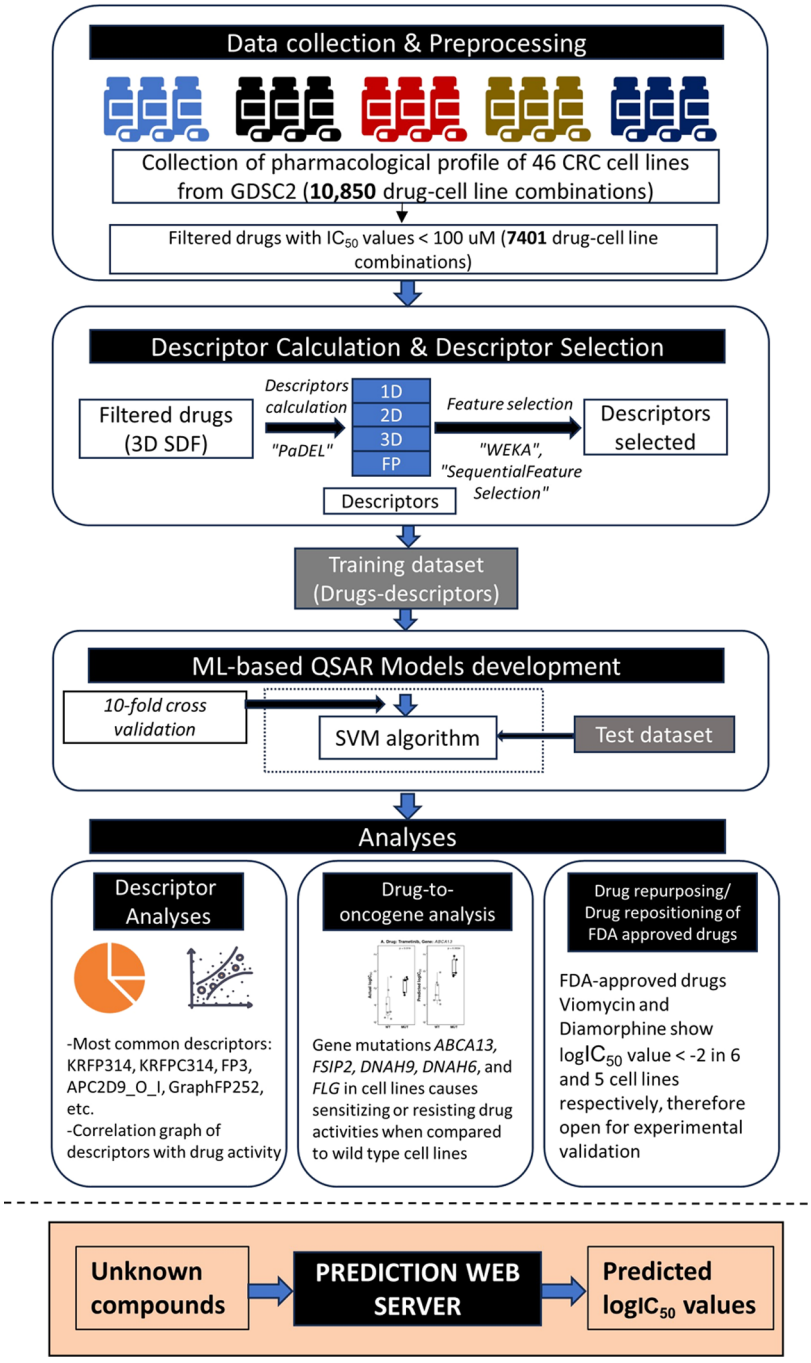



**Figure 1.** Overall study design for the development of QSAR models and prediction Web server.

various parameters, cell lines, and organisms, aiding multitarget anticancer drug development in diverse cancer types through multiple assays.[8−20]

In this study, we leverage the traditional QSAR approach for developing QSAR models for 12 CRC cell lines to identify putative drugs for CRC. The QSAR models' development is done based on the high-throughput pharmacological data available from Genomics of Drug Sensitivity in Cancer (GDSC). Further, we showed the applicability of these models in recapitulating the drug-to-oncogenes relation and repurposing the FDA-approved drugs. Moreover, they can also hasten screening of large chemical libraries to identify novel CRC drugs.[21] Our model will be useful for the research fraternity to complement the ongoing research to identify novel drug candidates for CRC, which can be taken further for experimental validations. For the community-wide utilization of the QSAR models developed in this study, we have integrated these models in a Web server called "ColoRecPred", which is freely available at https://project.iith.ac.in/cgntlab/colorecpred. Figure 1 describes the overall study design adopted in this study.

## METHODS

**Pharmacological Data.** For this study, we have downloaded the pharmacological screens against CRC cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org). A dataset of 297 anticancer drugs and their respective natural logarithmic $IC_{50}$


https://doi.org/10.1021/acsomega.4c01195
ACS Omega 2024, 9, 18584−18592

values were obtained across 46 CRC cell lines. We had a total of 10,850 drug-cell line combinations, each with an $IC_{50}$ value, and then applied a cutoff of 100 $\mu$M to remove the inactive drugs, reducing the total number of drug-cell line combinations to 7401. The total number of drugs for all 46 cell lines is shown in Supporting Information Figure 1. We extracted the individual cell line screened data with their respective $logIC_{50}$ values. PubChem compound IDs (CIDs) of the drugs were also retained to obtain the chemical structures of the drugs.

**Chemical Structure of the Drugs.** The chemical structures of all of the above drugs were downloaded in Spatial Data File (SDF) format from PubChem using their CIDs (https://pubchem.ncbi.nlm.nih.gov). These structures were in 2D format; therefore, they were subjected to 3D conversion using the RDKit toolkit in Python[22] followed by energy minimization using the Merck Molecular Force Field 1994 (MMFF94)[23−27] in OpenBabel software (version 3.1.1).[28]

**Descriptors' Calculation and Selection.** We used PaDEL software[29] to calculate the 1875 chemical descriptors (1D, 2D, and 3D) across 75 descriptor types like number of atoms count, topological, bond count, atom count, 3D autocorrelation, moment of inertia, RDF, WHIM, etc. and 12 different types of binary fingerprints like FP, ExtFP, GraphFP, SubFP, SubFPC, etc. Figure 2 shows the distribution

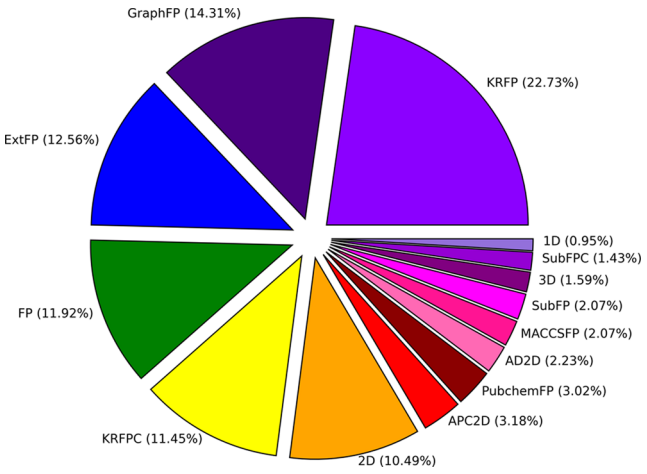



**Figure 2.** Graphical representation of the overview of the distribution of the descriptor types across the 12 selected cell lines. Pie chart representing the types of descriptors (1D, 2D, 3D, and binary fingerprints) across the 12 cell lines.

of the descriptor types. The total descriptor count was equal to 18066 across 75 descriptor types. For descriptor selection, we implemented the "RemoveUseless" function (removes descriptors with no variation or very high variation)[30] to preprocess the dataset, and then implemented "CfssubsetEval" Attribute Evaluator (evaluates the predictive ability of a descriptor and intercorrelation among other descriptors)[30,31] and "BestFirst" Ranker (evaluates features based on Greedy Hillclimbing and Backtracking mechanism)[31] in WEKA to select the descriptors. Further, we calculated the Shapley Additive Explanations (SHAP) values[32,33] for each of the 12 cell line models to understand the individual average impact of the descriptors to the model, which gives the structural and physicochemical interpretation for the developed models globally.

**QSAR Model Development.** To develop QSAR models, we used the Support Vector Machine (SVM) algorithm[34] using "scikit-learn" (version 1.2)[35] library in Python. We implemented a 10-fold cross-validation to avoid overfitting and assessed the model performances using various statistical indices, i.e., Pearson's correlation coefficient ($R$), coefficient of determination ($R^2$), mean squared error (MSE), mean average error (MAE), and root-mean-square error (RMSE). To identify the robust QSAR models, we used a cutoff of $R^2 > 0.6$. Selected models were further subjected to "F-stepping" to reduce the number of descriptors using the "SequentialFeatureSelection" function in "Mlxtend" library in Python environment.[36] During model development, we maintained the drugs to descriptor ratio for each of the selected cell lines close to 2:1 or greater to reduce the chances of overfitting.[37]

**Drug-to-Oncogene Relation.** We used QSAR models to recapitulate the drug-to-oncogene relationships in CRC. We downloaded the mutation data of CRC cell lines from the COSMIC Cell Line Project database (v97, https://cancer.sanger.ac.uk/cell_lines).[38] From the COSMIC mutation data, we removed mutations defined as "Unknown" and "Substitution - coding silent". We selected five genes, i.e., *ABCA13, DNAH6, DNAH9, FSIP2*, and *FLG*, which were mutated in at least five CRC cell lines (Supporting Information Table 1). For these five genes, we identified drugs with significant differences in their respective $logIC_{50}$ between wild-type and mutant cell lines ($p < 0.05$) and predicted their $logIC_{50}$ using QSAR models of respective wild-type and mutant cell lines. Then, we compared the predicted $logIC_{50}$ of the respective drugs in wild-type and mutated cell lines to recapitulate the drug-to-oncogene relation obtained from the experimentally known $logIC_{50}$.

**Drug Repurposing/Drug Repositioning.** We obtained the 1627 FDA-approved drugs from DrugBank (https://go.drugbank.com/) in the 2D SDF format. Their chemical descriptors and fingerprints were calculated using PaDEL (as described above), and we predicted their $logIC_{50}$ values using the QSAR models developed in this study. To select FDA-approved drugs with putative anticancer activity, we applied a cutoff of $logIC_{50}$ value $\leq -2$.

**Target and Pathway Analyses.** The targets of the FDA-approved drugs viomycin and diamorphine were identified using "Super-PRED" (Web site: https://prediction.charite.de/) tool.[39] In "Super-PRED", the criterion "Model accuracy" cutoff of >95% was applied to select the target proteins. The UniProt IDs of the proteins were supplied to the REACTOME Pathway Browser (Web site: https://reactome.org/PathwayBrowser)[40] to find the pathways associated with the target proteins. Further, the pathways were filtered based on FDR value <0.05. Later, we proceeded with the literature survey to find the occurrence of these pathways in CRC. Also, we performed STRING[41] analysis to find the protein−protein Interaction (PPI) networks.

## RESULTS

**QSAR Models' Performance.** To evaluate the performance of the QSAR models, we adopted four statistical indices: (a) Pearson's correlation coefficient ($R$), (b) coefficient of determination ($R^2$), (c) root-mean-square error (RMSE), and (d) mean absolute error (MAE). We developed individual QSAR models for 46 CRC cell lines by splitting the data sets into train (80%) and test (20%). To reduce overfitting, we applied 10-fold cross-validation within the training dataset


https://doi.org/10.1021/acsomega.4c01195
ACS Omega 2024, 9, 18584−18592

Table 1. Performance Measures Calculated for the 12 Cell Lines on the Training Dataset

| | | | performance before F-stepping | | | | | performance after F-stepping | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S. no. | cell line | no. of drugs | no. of descriptors | $R^2$ | $R$ | RMSE | MAE | no. of descriptors | $R^2$ | $R$ | RMSE | MAE |
| 1. | COLO-678 | 56 | 50 | 0.654 | 0.827 | 0.961 | 0.735 | 29 | 0.687 | 0.845 | 0.915 | 0.695 |
| 2. | HT-115 | 138 | 95 | 0.601 | 0.777 | 1.154 | 0.887 | 68 | 0.64 | 0.804 | 1.097 | 0.834 |
| 3. | SW620 | 122 | 86 | 0.572 | 0.793 | 1.719 | 1.263 | 44 | 0.639 | 0.827 | 1.578 | 1.096 |
| 4. | SW1463 | 122 | 79 | 0.621 | 0.792 | 1.224 | 0.873 | 47 | 0.662 | 0.835 | 1.155 | 0.779 |
| 5. | COLO-205 | 142 | 67 | 0.707 | 0.85 | 1.454 | 1.161 | 46 | 0.732 | 0.868 | 1.392 | 1.052 |
| 6. | GP5d | 151 | 102 | 0.713 | 0.845 | 1.306 | 0.958 | 55 | 0.786 | 0.887 | 1.126 | 0.848 |
| 7. | HT-29 | 146 | 134 | 0.601 | 0.782 | 1.491 | 1.097 | 43 | 0.707 | 0.844 | 1.277 | 0.908 |
| 8. | KM12 | 131 | 87 | 0.689 | 0.846 | 1.484 | 1.147 | 46 | 0.749 | 0.871 | 1.334 | 0.992 |
| 9. | SW1417 | 76 | 208 | 0.693 | 0.852 | 1.087 | 0.837 | 48 | 0.827 | 0.927 | 0.814 | 0.52 |
| 10. | MDST8 | 134 | 92 | 0.668 | 0.829 | 1.634 | 1.421 | 64 | 0.71 | 0.869 | 1.527 | 1.103 |
| 11. | SK-CO-1 | 142 | 75 | 0.693 | 0.841 | 1.455 | 1.159 | 50 | 0.747 | 0.88 | 1.322 | 1.051 |
| 12. | CCK-81 | 154 | 138 | 0.56 | 0.764 | 1.311 | 0.961 | 89 | 0.609 | 0.789 | 1.236 | 0.899 |

Table 2. Performance Measures Calculated for the 12 Cell Lines on the Test Dataset[a]

| | | | performance before F-stepping | | | | | performance after F-stepping | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S. no. | cell line | no. of drugs | descriptors | $R^2$ | $R$ | RMSE | MAE | descriptors | $R^2$ | $R$ | RMSE | MAE |
| 1. | COLO-678 | 14 | 50 | 0.903 | 0.952 | 0.555 | 0.414 | 29 | 0.882 | 0.939 | 0.611 | 0.48 |
| 2. | HT-115 | 34 | 95 | 0.704 | 0.84 | 1.225 | 0.994 | 68 | 0.74 | 0.862 | 1.149 | 0.858 |
| 3. | SW620 | 31 | 86 | 0.748 | 0.898 | 1.266 | 1.099 | 44 | 0.693 | 0.861 | 1.396 | 1.184 |
| 4. | SW1463 | 31 | 79 | 0.767 | 0.886 | 1.008 | 0.757 | 47 | 0.681 | 0.846 | 1.18 | 0.859 |
| 5. | COLO-205 | 35 | 67 | 0.688 | 0.835 | 1.301 | 1.076 | 46 | 0.677 | 0.827 | 1.207 | 0.987 |
| 6. | GP5d | 38 | 102 | 0.745 | 0.864 | 1.306 | 0.882 | 55 | 0.676 | 0.833 | 1.383 | 1.02 |
| 7. | HT-29 | 37 | 134 | 0.744 | 0.884 | 1.465 | 1.152 | 43 | 0.675 | 0.834 | 1.651 | 1.264 |
| 8. | KM12 | 33 | 87 | 0.781 | 0.904 | 1.051 | 0.857 | 46 | 0.66 | 0.813 | 1.31 | 1.036 |
| 9. | SW1417 | 19 | 208 | 0.806 | 0.933 | 0.852 | 0.637 | 48 | 0.633 | 0.803 | 1.171 | 0.937 |
| 10. | MDST8 | 34 | 92 | 0.678 | 0.834 | 1.421 | 1.01 | 64 | 0.63 | 0.809 | 1.524 | 1.108 |
| 11. | SK-CO-1 | 36 | 75 | 0.665 | 0.833 | 1.417 | 1.013 | 50 | 0.617 | 0.803 | 1.514 | 1.071 |
| 12. | CCK-81 | 39 | 138 | 0.73 | 0.872 | 1.236 | 0.924 | 89 | 0.603 | 0.785 | 1.5 | 1.179 |

[a]**Key:** $R^2$: coefficient of determination, $R$: Pearson's correlation coefficient, RMSE: root-mean-square error, MAE: mean absolute error

across all of the cell lines. After the cross-validation step, we applied a cutoff of $R^2 > 0.6$ to select the best QSAR models. With these selection criteria, we obtained 12 cell lines (Table 1) and proceeded with them for downstream analysis. For each QSAR model, we ensured the drugs to descriptors ratio of 2:1 or greater. To achieve this, we further reduced the number of descriptors using "F-stepping" as mentioned in the Methods section. The descriptor and drug numbers for the 12 selected cell lines are shown in Table 1. The performances were measured at two different descriptor counts, one after the "CfsSubsetEval" module in WEKA and the other after F-stepping. It was observed that in most of the cell lines, the performance of the models after F-stepping improved (Table 1). The highest performance was achieved for SW1417 cell lines ($R^2$ = 0.827) and the lowest performance was achieved for CCK-81 cell lines ($R^2$ = 0.609) on the training dataset (Table 1). Further, we tested QSAR models on the test dataset and obtained a performance ranging from 0.603 to 0.882 (Table 2). Supporting Information Figure 2 shows the scatter plots (with linear fit) between actual and predicted $\log IC_{50}$ values for the 12 CRC cell lines.

**Descriptor Analysis.** We listed the most occurring descriptors and fingerprints and found that fingerprints KRFP314, KRFPC314 and FP3 were the three most occurring descriptors across 12 selected cell lines (occurring in nine, seven, and six cell lines, respectively), as shown in Supporting Information Table 2. We also analyzed the changes in the drug activity in the presence or absence of the descriptors. The descriptors KRFP314, KRFPC314, FP3, APC2D9_O_I, GraphFP252, KRFP3683, KRFP803, and nC are categorical values, whereas JGI10 is in numerical values. Figure 3 shows significantly associated descriptors in specific cell lines, which shows the increasing drug activity in the presence of the descriptors GraphFP252 and FP3. Supporting Information Figure 3 (A–L) shows extended explorations of drug-descriptor relationships. Supporting Information Figure 4 shows the correlation plot of the JGI10 descriptor among the five cell lines. Supporting Information Table 3 shows the calculated mean absolute SHAP values of the descriptors across the 12 developed QSAR models in the decreasing order of the values. The descriptors with the highest contribution to the model have higher mean absolute SHAP values.

**Drug-to-Oncogene Relation Validation.** We used QSAR models to rehash the drug-to-oncogene relationships obtained from the experimental data. We identified the association between *ABCA13* and Trametinib, where *ABCA13* mutated cell lines were less sensitive for Trametinib as compared to wild-type cell lines ($P$ = 0.019). Using QSAR models developed in this study, we predicted the $\log IC_{50}$ of Trametinib for these cell lines and observed a similar trend with predicted $\log IC_{50}$ ($P$ = 0.0034) (Figure 4A). We found another association between *FSIP2* and SGC0946, where *FSIP2* mutated cell lines were more sensitive to SGC0946 ($P$ = 0.0059). We predicted the $\log IC_{50}$ using QSAR models and recapitulated this association ($P$ = 0.0057) (Figure 4B). We found more such drug-to-oncogene relations and recapitulated



https://doi.org/10.1021/acsomega.4c01195
ACS Omega 2024, 9, 18584−18592

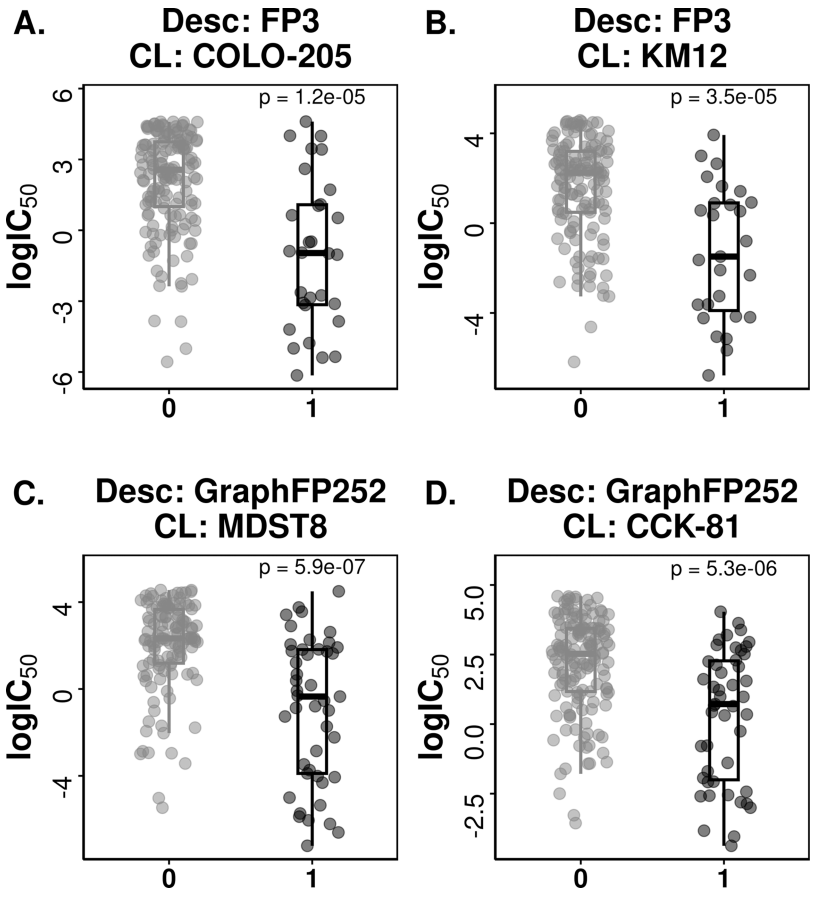



**Figure 3.** Drug-descriptor analysis: (A) Descriptor FP3 increases the drug activity in cell line COLO-205; (B) Descriptor FP3 increases the drug activity in cell line KM12; (C) Descriptor GraphFP252 increases the drug activity in cell line MDST8; (D) Descriptor GraphFP252 increases the drug activity in cell line CCK-81 (0: descriptor absent and 1: descriptor present; Desc: descriptor and CL: cell line).

them using the QSAR models developed in this study (Supporting Information Figure 5), which highlight the predictive power of these models.

**FDA-Approved Drug Analysis.** We further extended our QSAR models to repurpose the FDA-approved drugs for CRC. We predicted the $\log IC_{50}$ of 1627 FDA-approved drugs across the 12 CRC cell lines using our QSAR models. Of these, we filtered the drugs with predicted $\log IC_{50} = < -2\ \mu M$ for each of the 12 cell lines to select the drugs with prominent activity. We found 11 drugs with $\log IC_{50} = < -2\ \mu M$ in at least five cell lines out of 12 (Table 3). Of these 11 drugs, six were known anticancer drugs, three were antimicrobial, which have been shown to have an antineoplastic effect,[42−60] and two drugs, i.e., viomycin (antimicrobial agent) and diamorphine (analgesic), have yet not been experimentally validated as anticancer drugs.[61,62]

**Target and Pathway Analyses.** In a drug target analysis of FDA-approved drugs, i.e., viomycin and diamorphine, we identified 70 and 43 targets, respectively, using Super-PRED with >95% "model accuracy" (Supporting Information Table 4). We performed REACTOME pathway analysis on these targets and identified 137 and 13 pathways for viomycin and diamorphine, respectively (FDR < 0.05). Our analysis of viomycin targets revealed enrichment of pathways associated with CRC, including *WNT5A*-dependent internalization of *FZD4*, receptor tyrosine kinase (RTK) signaling, PI3K/Akt signaling, and G protein-coupled receptor (GPCR) signaling.[63−70] Diamorphine target analysis similarly identified the pathways associated with CRC, i.e., condensation of prometaphase chromosomes, and regulation of *NFE2L2* gene expression and *RHO* GTPase effectors.[67−71] Notably, pathways common across the drug targets of both viomycin and diamorphine highlighted the prominence of regulation of *NFE2L2*, *RHO GTPase*-mediated events, and intermediate filaments-mediated events as potential convergence points, aligning with existing literature[63−65,69,70,72−74] (Supporting Information Table 5). Also, we found 27 targets common between the drugs viomycin and diamorphine (Supporting Information Table 4 and Figure 6).

**Web Server.** We have incorporated these QSAR models into a web server called 'ColoRecPred.' This online platform is designed for the prediction of the anticancer activity, specifically $\log IC_{50}$ values, of unfamiliar compounds against CRC cell lines. We implemented a two-tier Web server architecture, featuring a user-friendly interface built using HTML, CSS, and JavaScript, supported by the Java OpenJDK v11.0.21 framework. The back-end leveraged a combination of programming languages and tools to handle various functionalities. PHP facilitated user interaction and data flow, ensuring a smooth user experience. Bash scripts were employed to orchestrate the Web server's internal processes, ensuring efficient operation. R programming language (v4.1.2) was utilized to manipulate and analyze user-provided data. Python (v3.12.1) played a crucial role in developing and visualizing machine learning models, offering valuable insights directly on the results page. To extract 3D structures from user-uploaded


https://doi.org/10.1021/acsomega.4c01195
ACS Omega 2024, 9, 18584−18592

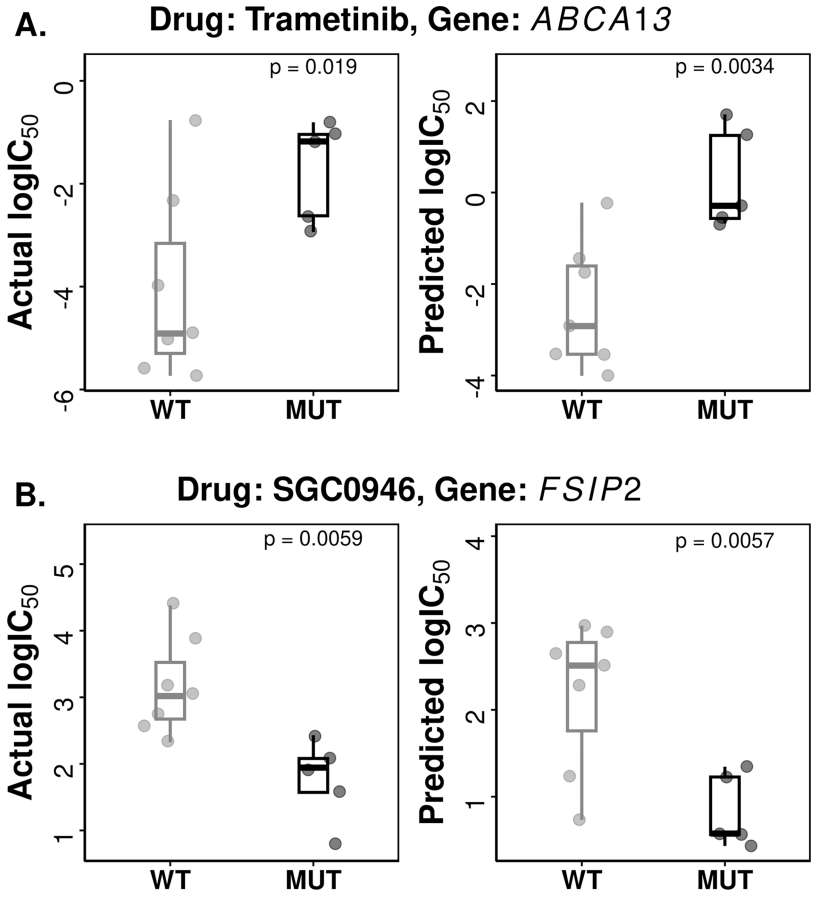



**Figure 4.** Validation of drug-to-oncogene relationship using QSAR models. (A) Mutations in *ABCA13* reduced the sensitivity of CRC cell lines for Trametinib in both actual (experimental) and predicted $logIC_{50}$. (B) Mutations in *FSIP2* increased the sensitivity of CRC cell lines for SGC0946 in both actual (experimental) and predicted $logIC_{50}$.

**Table 3. Most Potent Drugs after Analyzing FDA-Approved Drugs by our QSAR Models**

| S. no. | pubchem ID | drugbank accession # | name | frequency in 12 cell lines | cell lines present in | mode of action | reference(s) (where drug is reported as anticancer) |
|---|---|---|---|---|---|---|---|
| 1. | 31101 | DB01200 | Bromocriptine | 6 | HT-29, SW1463, SK-CO-1, KM12, COLO-205, GP5d | ergot alkaloid | 57−59 |
| 2. | 42890 | DB01177 | Idarubicin | 6 | MDST8, KM12, SW1463, GP5d, COLO-205, SK-CO-1 | antitumor | 55,56 |
| **3.** | **3037981** | **DB06827** | **Viomycin** | **6** | **SK-CO-1, COLO-205, COLO-678, SW1463, KM12, MDST8** | **antimicrobial** | **not yet reported as antitumor** |
| 4. | 5746 | DB00305 | Mitomycin | 5 | CCK-81, COLO-205, SK-CO-1, KM12, GP5d | antimicrobial | 60 |
| 5. | 8223 | DB00696 | Ergotamine | 5 | HT-29, SW1463, SK-CO-1, COLO-205, KM12 | ergot alkaloid | 53,54 |
| 6. | 10531 | DB00320 | Dihydroergotamine | 5 | SK-CO-1, SW1463, HT-29, COLO-205, KM12 | ergot alkaloid | 51,52 |
| 7. | 31703 | DB00997 | Doxorubicin | 5 | KM12, GP5d, COLO-205, MDST8, SK-CO-1 | antibiotic | 49,50 |
| 8. | 41867 | DB00445 | Epirubicin | 5 | KM12, GP5d, COLO-205, MDST8, SK-CO-1 | antitumor | 47,48 |
| 9. | 285033 | DB04865 | Omacetaxine mepesuccinate | 5 | SW620, GP5d, SK-CO-1, HT-29, KM12 | antitumor | 45,46 |
| **10.** | **5462328** | **DB01452** | **Diamorphine** | **5** | **HT-115, COLO-205, KM12, SK-CO-1, GP5d** | **analgesic** | **not yet reported as antitumor** |
| 11. | 11707110 | DB08911 | Trametinib | 5 | SK-CO-1, HT-29, MDST8, COLO-205, SW1463 | antitumor | 42−44 |

2D files, the "RDKit"[22] library provided efficient functionalities. In the web server, the "Predict" page in the Web server allows users to draw a compound or paste a 2D structure (in SDF format) of an unknown compound and select the CRC cell lines for which the user wants to predict $logIC_{50}$ values. On submitting the queries, the Web server shall return the $logIC_{50}$ values of the compound for each of the selected cell lines in a tabular format. "ColoRecPred" is freely available at https://project.iith.ac.in/cgntlab/colorecpred.



https://doi.org/10.1021/acsomega.4c01195
ACS Omega 2024, 9, 18584−18592

## DISCUSSION

In the context of CRC, there is an urgency to swiftly develop alternative therapies and medications to address the challenge of high incidence and mortality.[75] Further, there is the problem of a prolonged timeline of discovery of new drugs (~10 to 15 years), which ends up being very expensive and time-consuming.[76] Computational strategies have been developed and adapted in both pharmaceutical industries and academia to boost the process of drug discovery and development.[77] Keeping this in mind, we have developed robust QSAR models based on SVM. These models will facilitate the screening of potential drugs for CRC treatment, whether as standalone therapies or in combination with conventional drugs. In our study, we developed QSAR models for 12 CRC cell lines to design novel drug compounds against CRC. Drugs have symbolic codes or structures that are quantifiable, known as chemical descriptors that can confer the potent drug activity of the molecule. We utilized the correlation of these chemical descriptors with the drug activity to develop QSAR models. In descriptor analyses, we found the presence of various 1D, 2D, and 3D descriptors and molecular fingerprints across the drugs. With the help of descriptors selection, we could effectively design models with performance ($R^2$) ranging from 0.609 to 0.827. To show the robustness of our QSAR models, we implemented them to recapitulate drug-to-oncogene relations identified using experimental data. We successfully recapitulated the associations of the genes *ABCA13, FLG, FSIP2, DNAH6,* and *DNAH9* with drugs Trametinib, SGC0946, Dinaciclib, Sabutoclax, Vincristine, and SCH772984, respectively. These genetic associations, if studied deeper, shall open dimensions to discover novel biomarkers or drug targets for CRC. To be further affirmative of the results, experimental validation of these drug-to-oncogene relationships is necessary. Also, drug-descriptor analyses showed that the presence of the chemical descriptors KRFP314, KRFPC314, FP3, APC2-D9_O_I, GraphFP252, JGI10, KRFP3683, KRFP803, and nC increased the drug activity. Therefore, it can be assumed that the presence of these chemical descriptors in drugs confers anti-CRC activity. The structural and physicochemical model interpretations using mean absolute SHAP quantifies the contribution of each descriptor to the model's predictions, providing insights into how individual descriptors influence the model's output.

Furthermore, our *in silico* drug repurposing analysis identified two potential FDA-approved drugs, i.e., viomycin and diamorphine, for the CRC treatment. This analysis emphasizes the applicability of QSAR models developed in this study. We further predicted the pathways that viomycin and diamorphine might target in CRC. From the predicted drug targets and pathway analyses of these two drugs, we identified targets that are involved in the pathways associated with CRC, e.g., *WNT*-signaling pathways, *GPCR*-related pathways, and intermediate filaments-related pathways. This warrants that viomycin and diamorphine be explored as effective anti-CRC drugs after further experimental validations. We anticipate that the QSAR models developed in this study will be critical for drug repurposing and designing of novel drugs against CRC.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c01195.

Figure 1: Overview of the drugs count across all 46 cell lines. Figure 2: Scatter plots describing the relationship between the Actual log$IC_{50}$ values and the predicted log$IC_{50}$ values by our developed QSAR models across the 12 cell lines Figure 3: Graphical representation of drug-descriptor relationships of various drug-descriptor combinations. Figure 4: Barplot of JGI10 showing the Pearson's correlation coefficient with drug activity (log$IC_{50}$). Figure 5: Graphical representation of drug-oncogene relationships of various drug-gene combinations. Figure 6: STRING analysis PPI network for A. viomycin and B. diamorphine (PDF)

Table 1: Mutation counts of genes across 12 cell lines. Table 2: Most common descriptors across 12 cell lines. Table 3: Mean absolute SHAP values of F-step selected descriptors for all the 12 developed QSAR models. Table 4: Predicted Targets of the drugs viomycin and diamorphine. Table 5: Predicted Pathways for Viomycin and Diamorphine from REACTOME (XLSX)

## AUTHOR INFORMATION


### Corresponding Author

**Rahul Kumar** − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India;* orcid.org/0000-0002-6927-5390; Email: rahulk@bt.iith.ac.in

### Authors

**Avik Sengupta** − *Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India*

**Saurabh Kumar Singh** − *Department of Chemistry, Indian Institute of Technology Hyderabad, Kandi, Telangana 502284, India;* orcid.org/0000-0001-9488-8036

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c01195


### Author Contributions

Conceptualization and design, data acquisition, analyses, and interpretation: A.S. and R.K.; interpretation of chemical descriptors: A.S., R.K., and S.K.S.; writing of the manuscript draft: A.S. and R.K. All authors contributed to the article and approved the submitted version.


### Funding

Research seed grant from the Indian Institute of Technology Hyderabad.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

AS is thankful to the Council of Scientific & Industrial Research (CSIR) for providing the research fellowship. The authors are thankful to the Indian Institute of Technology Hyderabad (IITH) for providing the necessary infrastructure to help us successfully conduct our research work. The authors acknowledge Ms. Kavita Kundal for the critical reading of the manuscript.

## REFERENCES


(1) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca-Cancer J. Clin.* **2021**, *71* (3), 209−249.

(2) Drugs Approved for Colon and Rectal Cancer - NCI. https://www.cancer.gov/about-cancer/treatment/drugs/colorectal (accessed April 23, 2023).

(3) Jeught, K. V. d.; Xu, H. C.; Li, Y. J.; Lu, X. Bin.; Ji, G. Drug resistance and new therapies in colorectal cancer. *World J. Gastroenterol.* **2018**, *24* (34), 3834.

(4) Wang, Q.; Shen, X.; Chen, G.; Du, J. Drug Resistance in Colorectal Cancer: From Mechanism to Clinic. *Cancers* **2022**, *14* (12), No. 2928, DOI: 10.3390/cancers14122928.

(5) Moshawih, S.; Lim, A. F.; Ardianto, C.; Goh, K. W.; Kifli, N.; Goh, H. P.; et al. Target-Based Small Molecule Drug Discovery for Colorectal Cancer: A Review of Molecular Pathways and In Silico Studies. *Biomolecules* **2022**, *12* (7), No. 878, DOI: 10.3390/biom12070878.

(6) Rai, A.; Raj, V.; Aboumanei, M. H.; Singh, A. K.; Keshari, A. K.; Verma, S. P.; Saha, S. Pharmacophore, 3D-QSAR Models and Dynamic Simulation of 1,4-Benzothiazines for Colorectal Cancer Treatment. *Comb. Chem. High Throughput Screening* **2017**, *20* (8), 658−674, DOI: 10.2174/1386207320666170509153137.

(7) Khaled, D. M.; Elshakre, M. E.; Noamaan, M. A.; Butt, H.; Abdel Fattah, M. M.; Gaber, D. A. A Computational QSAR, Molecular Docking and In Vitro Cytotoxicity Study of Novel Thiouracil-Based Drugs with Anticancer Activity against Human-DNA Topoisomerase II. *Int. J. Mol. Sci.* **2022**, *23* (19), No. 11799, DOI: 10.3390/ijms231911799.

(8) Speck-Planche, A. Multicellular Target QSAR Model for Simultaneous Prediction and Design of Anti-Pancreatic Cancer Agents. *ACS Omega* **2019**, *4* (2), 3122−3132.

(9) Kleandrova, V. V.; Scotti, M. T.; Scotti, L.; Nayarisseri, A.; Speck-Planche, A. Cell-based multi-target QSAR model for design of virtual versatile inhibitors of liver cancer cell lines. *SAR QSAR Environ. Res.* **2020**, *31* (11), 815−836, DOI: 10.1080/1062936X.2020.1818617.

(10) Kleandrova, V. V.; Speck-Planche, A. PTML Modeling for Pancreatic Cancer Research: In Silico Design of Simultaneous Multi-Protein and Multi-Cell Inhibitors. *Biomedicines* **2022**, *10* (2), 491.

(11) Bediaga, H.; Arrasate, S.; González-Díaz, H. PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Comb. Sci.* **2018**, *20* (11), 621−632.

(12) Speck-Planche, A.; Scotti, M. T. BET bromodomain inhibitors: fragment-based in silico design using multi-target QSAR models. *Mol. Diversity* **2019**, *23* (3), 555−572.

(13) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Rational drug design for anti-cancer chemotherapy: Multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.* **2012**, *20* (15), 4848−4855.

(14) Marzaro, G.; Chilin, A.; Guiotto, A.; Uriarte, E.; Brun, P.; Castagliuolo, I.; et al. Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors. *Eur. J. Med. Chem.* **2011**, *46* (6), 2185−2192.

(15) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Fragment-based QSAR model toward the selection of versatile anti-sarcoma leads. *Eur. J. Med. Chem.* **2011**, *46* (12), 5910−5916.

(16) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Multi-target drug discovery in anti-cancer therapy: Fragment-based approach toward the design of potent and versatile anti-prostate cancer agents. *Bioorg. Med. Chem.* **2011**, *19* (21), 6239−6244.

(17) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Chemoinformatics in anti-cancer chemotherapy: Multi-target QSAR model for the in silico discovery of anti-breast cancer agents. *Eur. J. Pharm. Sci.* **2012**, *47* (1), 273−279.

(18) Speck-Planche, A.; Kleandrova, V.; Luan, F.; Cordeiro, M. N. Chemoinformatics in Multi-target Drug Discovery for Anti-cancer Therapy: In Silico Design of Potent and Versatile Anti-brain Tumor Agents. *Anti-Cancer Agents Med. Chem.* **2012**, *12* (6), 678−685, DOI: 10.2174/187152012800617722.

(19) Planche, A. S.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Unified Multi-target Approach for the Rational in silico Design of Anti-bladder Cancer Agents. *Anti-Cancer Agents Med. Chem.* **2013**, *13* (5), 791−800.

(20) Speck-Planche, A.; Cordeiro, M. N. D. S. Fragment-based in silico modeling of multi-target inhibitors against breast cancer-related proteins. *Mol. Diversity* **2017**, *21* (3), 511−523.

(21) Peter, S. C.; Dhanjal, J. K.; Malik, V.; Radhakrishnan, N.; Jayakanthan, M.; Sundar, D.et al.Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2019; Vol. *1*−3, pp 661−676 DOI: 10.1016/B978-0-12-809633-8.20197-0.

(22) RDKit: Open-source cheminformatics. https://www.rdkit.org.

(23) Halgren, T. A. *Performance of MMFF94**; John Wiley & Sons, Inc, 1996; Vol. *17*http://journals.wiley.com/jcc.

(24) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17* (5−6), 520−552.

(25) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 553−586.

(26) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 587−615.

(27) Halgren, T. A. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **1996**, *17* (5−6), 616−641.

(28) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open chemical toolbox. *J. Cheminf.* **2011**, *3* (10), 1−14.

(29) Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466−1474.

(30) Hall, M.; Smith, L.Feature subset selection: a correlation based filter approach. In *Proceedings of International Conference on Neural Information Processing and Intelligent Information Systems*, 1998; pp 855−858.

(31) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA data mining software. *ACM SIGKDD Explor. Newsl.* **2009**, *11* (1), 10−18.

(32) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Guyon, I.; Luxburg, U.; Von Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.et al., Eds.; Curran Associates, Inc., 2017; Vol. *30*.

(33) Jaganathan, K.; Tayara, H.; Chong, K. T. An Explainable Supervised Machine Learning Model for Predicting Respiratory Toxicity of Chemicals Using Optimal Molecular Descriptors. *Pharmaceutics* **2022**, *14* (4), 832.

(34) Mei, H.; Zhou, Y.; Liang, G.; Li, Z. Support vector machine applied in QSAR modelling. *Chin. Sci. Bull.* **2005**, *50* (20), 2291−2296.

(35) Pedregosa FABIANPEDREGOSA, F.; Michel, V.; Grisel OLIVIERGRISEL, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825−2830.

(36) Raschka, S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J. Open Source Software* **2018**, *3* (24), 638.

(37) Kumar, R.; Chaudhary, K.; Singla, D.; Gautam, A.; Raghava, G. P. S. Designing of promiscuous inhibitors against pancreatic cancer cell lines. *Sci. Rep.* **2014**, *4* (1), No. 4668.

(38) Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47* (D1), D941−D947.




https://doi.org/10.1021/acsomega.4c01195
*ACS Omega* 2024, 9, 18584−18592

(39) Gallo, K.; Goede, A.; Preissner, R.; Gohlke, B.-O. SuperPred 3.0: drug classification and target prediction—a machine learning approach. *Nucleic Acids Res.* **2022**, *50* (W1), W726−W731.
(40) Fabregat, A.; Sidiropoulos, K.; Viteri, G.; Forner, O.; Marin-Garcia, P.; Arnau, V.; et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinf.* **2017**, *18* (1), No. 142.
(41) Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47* (D1), D607−D613.
(42) Liu, L.; Shi, H.; Bleam, M. R.; Zhang, V.; Zou, J.; Jing, J.; et al. Antitumor effects of dabrafenib, trametinib, and panitumumab as single agents and in combination in *BRAF* -mutant colorectal carcinoma (CRC) models. *J. Clin. Oncol.* **2014**, *32* (15_suppl), 3513−3513.
(43) Bolf, E. L.; Beadnell, T. C.; Rose, M. M.; D'Alessandro, A.; Nemkov, T.; Hansen, K. C.; Schweppe, R. E. Dasatinib and Trametinib Promote Anti-Tumor Metabolic Activity. *Cells* **2023**, *12* (10), No. 1374, DOI: 10.3390/cells12101374.
(44) Ho, M. Y. K.; Morris, M. J.; Pirhalla, J. L.; Bauman, J. W.; Pendry, C. B.; Orford, K. W.; et al. Trametinib, a first-in-class oral MEK inhibitor mass balance study with limited enrollment of two male subjects with advanced cancers. *Xenobiotica* **2014**, *44* (4), 352−368.
(45) Wetzler, M.; Segal, D. Omacetaxine as an Anticancer Therapeutic: What is Old is New Again. *Curr. Pharm. Des.* **2011**, *17* (1), 59−64.
(46) Nazha, A.; Kantarjian, H.; Cortes, J.; Quintás-Cardama, A. Omacetaxine mepesuccinate (synribo) − newly launched in chronic myeloid leukemia. *Expert Opin. Pharmacother.* **2013**, *14* (14), 1977−1986.
(47) Tsukagoshi, S. [Epirubicin (4′-epi-adriamycin]. *Gan To Kagaku Ryoho.* **1990**, *17* (1), 151−159.
(48) Cersosimo, R. J.; Hong, W. K. Epirubicin: a review of the pharmacology, clinical activity, and adverse effects of an adriamycin analogue. *J. Clin. Oncol.* **1986**, *4* (3), 425−439.
(49) Peter, S.; Alven, S.; Maseko, R. B.; Aderibigbe, B. A. Doxorubicin-Based Hybrid Compounds as Potential Anticancer Agents: A Review. *Molecules* **2022**, *27* (14), 4478.
(50) Kciuk, M.; Gielecińska, A.; Mujwar, S.; Kołat, D.; Kałuzińska-Kołat, Ż.; Celik, I.; Kontek, R. Doxorubicin—An Agent with Multiple Mechanisms of Anticancer Activity. *Cells* **2023**, *12* (4), No. 659.
(51) Chang, S. H.; Lee, A. Y.; Yu, K. N.; Park, J.; Kim, K. P.; Cho, M. H. Dihydroergotamine Tartrate Induces Lung Cancer Cell Death through Apoptosis and Mitophagy. *Chemotherapy* **2016**, *61* (6), 304−312.
(52) He, M.; Liao, Q.; Liu, D.; Dai, X.; Shan, M.; Yang, M.; et al. Dihydroergotamine mesylate enhances the anti-tumor effect of sorafenib in liver cancer cells. *Biochem. Pharmacol.* **2023**, *211*, No. 115538, DOI: 10.1016/j.bcp.2023.115538.
(53) Crider, A. M.; Lu, C. K. L.; Floss, H. G.; Cassady, J. M.; Clemens, J. A. Ergot alkaloids. Synthesis of nitrosourea derivatives of ergolines as potential anticancer agents. *J. Med. Chem.* **1979**, *22* (1), 32−35.
(54) Mrusek, M.; Seo, E.-J.; Greten, H. J.; Simon, M.; Efferth, T. Identification of cellular and molecular factors determining the response of cancer cells to six ergot alkaloids. *Invest. New Drugs* **2015**, *33* (1), 32−44.
(55) Tsuruo, T.; Oh-Hara, T.; Sudo, Y.; Naito, M. Antitumor activity of idarubicin, a derivative of daunorubicin, against drug sensitive and resistant P388 leukemia. *Anticancer Res.* **1993**, *13* (2), 357−361.
(56) Cersosimo, R. J. Idarubicin: an anthracycline antineoplastic agent. *Clin. Pharm.* **1992**, *11* (2), 152−167.
(57) Kamazani, F. M.; Sotoodehnejad nematalahi, F.; Siadat, S. D.; Pornour, M.; Sheikhpour, M. A success targeted nano delivery to lung cancer cells with multi-walled carbon nanotubes conjugated to bromocriptine. *Sci. Rep.* **2021**, *11* (1), No. 24419.
(58) Bai, L.; Li, X.; Yang, Y.; Zhao, R.; White, E. Z.; Danaher, A.; et al. Bromocriptine monotherapy overcomes prostate cancer chemoresistance in preclinical models. *Transl. Oncol.* **2023**, *34*, No. 101707.
(59) Seo, E. J.; Sugimoto, Y.; Greten, H. J.; Efferth, T. Repurposing of bromocriptine for cancer therapy. *Front. Pharmacol.* **2018**, *9*, No. 1030, DOI: 10.3389/fphar.2018.01030.
(60) Tomasz, M. Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* **1995**, *2* (9), 575−579.
(61) Holm, M.; Borg, A.; Ehrenberg, M.; Sanyal, S. Molecular mechanism of viomycin inhibition of peptide elongation in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (4), 978−983.
(62) Rana, S. P. S.; Ahmed, A.; Kumar, V.; Chaudhary, P. K.; Khurana, D.; Mishra, S. Successful Management of a Difficult Cancer Pain Patient by Appropriate Adjuvant and Morphine Titration. *Indian J. Palliat. Care* **2011**, *17* (2), 162.
(63) Kumawat, K.; Gosens, R. WNT-5A: signaling and functions in health and disease. *Cell. Mol. Life Sci.* **2016**, *73* (3), 567−587.
(64) Chen, Y.; Chen, Z.; Tang, Y.; Xiao, Q. The involvement of noncanonical Wnt signaling in cancers. *Biomed. Pharmacother.* **2021**, *133*, No. 110946.
(65) Azbazdar, Y.; Karabicici, M.; Erdal, E.; Ozhan, G. Regulation of Wnt Signaling Pathways at the Plasma Membrane and Their Misregulation in Cancer. *Front. Cell Dev. Biol.* **2021**, *9*, No. 631623, DOI: 10.3389/fcell.2021.631623.
(66) McManus, S.; Chababi, W.; Arsenault, D.; Dubois, C. M.; Saucier, C. Dissecting Oncogenic RTK Pathways in Colorectal Cancer Initiation and Progression. *Methods Mol. Biol.* **2018**, *1765*, 27−42.
(67) Danielsen, S. A.; Eide, P. W.; Nesbakken, A.; Guren, T.; Leithe, E.; Lothe, R. A. Portrait of the PI3K/AKT pathway in colorectal cancer. *Biochim. Biophys. Acta* **2015**, *1855* (1), 104−121.
(68) Maharati, A.; Moghbeli, M. PI3K/AKT signaling pathway as a critical regulator of epithelial-mesenchymal transition in colorectal tumor cells. *Cell Commun. Signaling* **2023**, *21* (1), No. 201.
(69) Chaudhary, P. K.; Kim, S. An Insight into GPCR and G-Proteins as Cancer Drivers. *Cells* **2021**, *10* (12), 3288.
(70) Zeng, Z.; Ma, C.; Chen, K.; Jiang, M.; Vasu, R.; Liu, R.; et al. Roles of G Protein-Coupled Receptors (GPCRs) in Gastrointestinal Cancers: Focus on Sphingosine 1-Shosphate Receptors, Angiotensin II Receptors, and Estrogen-Related GPCRs. *Cells* **2021**, *10* (11), 2988.
(71) Gonzalez-Donquiles, C.; Alonso-Molero, J.; Fernandez-Villa, T.; Vilorio-Marqués, L.; Molina, A. J.; Martín, V. The NRF2 transcription factor plays a dual role in colorectal cancer: A systematic review. *PLoS One* **2017**, *12* (5), No. e0177549.
(72) Kim, M.-S.; Ha, S.-E.; Wu, M.; Zogg, H.; Ronkon, C. F.; Lee, M.-Y.; Ro, S. Extracellular Matrix Biomarkers in Colorectal Cancer. *Int. J. Mol. Sci.* **2021**, *22* (17), 9185.
(73) Li, Z.-L.; Wang, Z.-J.; Wei, G.-H.; Yong, Y.; Wang, X.-W. Changes in extracellular matrix in different stages of colorectal cancer and their effects on proliferation of cancer cells. *World J. Gastrointest. Oncol.* **2020**, *12* (3), 267−275.
(74) Despotović, S. Z.; Milićević, ĐN.; Krmpot, A. J.; Pavlović, A. M.; Živanović, V. D.; Krivokapić, Z.; et al. Altered organization of collagen fibers in the uninvolved human colon mucosa 10 and 20 cm away from the malignant tumor. *Sci. Rep.* **2020**, *10* (1), No. 6359.
(75) Brenner, H.; Kloor, M.; Pox, C. P. Colorectal cancer. *Lancet* **2014**, *383* (9927), 1490−1502.
(76) Mohs, R. C.; Greig, N. H. Drug discovery and development: Role of basic biological research. *Alzheimer's Dementia* **2017**, *3* (4), 651.
(77) Sadybekov, A. V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616* (7958), 673−685.