RESEARCH ARTICLE

Novel target identification towards drug repurposing based on biological activity profiles

Binghan Xue[1], **Yanji Xu**[1], **Ruili Huang**[2], **Qian Zhu**[2]*

1 Division of Rare Disease Research Innovation, National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), Rockville, Maryland, United States of America, **2** Division of Preclinical Innovation, National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), Rockville, Maryland, United States of America

* qian.zhu@nih.gov

Abstract

Rare diseases affect more than 30 million individuals, with the majority facing limited treatment options, elevating the urgency to innovative therapeutic solutions. Addressing these medical challenges necessitates an exploration of novel treatment modalities. Among these, drug repurposing emerges as a promising avenue, offering both potential and risk mitigation. To achieve this goal, we primarily focused on developing predictive models that harness cutting-edge computational techniques to uncover latent relationships between gene targets and chemical compounds towards drug repurposing. Building upon our previous investigation, where we successfully identified gene targets for compounds from the Tox21 in vitro assays, our endeavor expanded to a systematic prediction of potential targets for drug repurposing employing machine learning models built on diverse algorithms such as Support Vector Classifier, K-Nearest Neighbors, Random Forest, and Extreme Gradient Boosting. These models were trained on comprehensive biological activity profile data to predict the relationship between 143 gene targets and over 6000 compounds. Our models demonstrated high accuracy (>0.75), with predictions further validated by using public experimental datasets. Furthermore, several findings were evaluated via case studies. By elucidating these connections, we aim to streamline the drug repurposing process, ultimately catalyzing the discovery of more effective therapeutic interventions for rare diseases.

Citation: Xue B, Xu Y, Huang R, Zhu Q (2025) Novel target identification towards drug repurposing based on biological activity profiles. PLoS One 20(5): e0319865. https://doi.org/10.1371/journal.pone.0319865

Editor: Sheikh Arslan Sehgal, Cholistan University of Veterinary and Animal Sciences, PAKISTAN

Received: October 18, 2024

Accepted: February 9, 2025

Published: May 6, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0319865

Introduction

The concept of drug repurposing has emerged as a beacon of hope, offering a cost-effective and time-efficient approach to addressing the complexities of drug discovery [1–4]. At its core, drug repurposing involves the exploration of alternative applications for already approved or investigational drugs, tapping into their potential

Data availability statement: All relevant data are within the manuscript and its Supporting Information files.

beyond their originally intended uses. This strategy not only capitalizes on existing pharmacological knowledge but also expedites the drug development process by bypassing certain stages of preclinical and clinical trials [5]. A compelling rationale for promoting drug repurposing lies in the interconnected nature of disease mechanisms. Scientific evidence suggests that a single molecular target implicated in a specific disease often exerts influence on various genetic pathways associated with other rare diseases [6,7]. Therefore, unraveling the intricate relationship between chemical compounds and gene targets assumes paramount importance in the quest for breakthroughs in rare disease treatment.

Researchers employ a multifaceted approach to discover the complicated relationship between chemical compounds and gene targets, leveraging a combination of experimental techniques and computational methodologies. Experimental studies often involve high-throughput screening assays, where thousands of compounds are tested against various biological targets to identify potential interactions [8–10]. Concurrently, advanced computational algorithms, such as molecular docking and network analysis, play a crucial role in predicting the binding affinity and specificity of compounds to target proteins [11–15]. These computational models rely on structural and functional data of both compounds and target proteins, allowing researchers to elucidate the molecular mechanisms underlying drug-target interactions.

However, traditional methods like molecular docking and network analysis, while valuable, have certain limitations in the exploration of chemical compounds and gene targets. Molecular docking, for instance, relies heavily on structural data of target proteins and ligands, often overlooking the dynamic nature of protein-ligand interactions and the influence of solvent effects. This can lead to inaccuracies in predicting binding affinities and may not capture the full spectrum of interactions within complex biological systems [15–17]. Similarly, network analysis, while useful for identifying functional relationships between genes and proteins, may struggle to integrate diverse types of data and capture the dynamic nature of biological networks [18]. Additionally, traditional methods often require manual curation and parameter tuning, which is time-consuming and potentially biased [15,18–20]. Moreover, these methods may struggle with the scalability required to analyze large-scale datasets, limiting their applicability in the era of big data [21,22]. These inherent limitations underscore the need for complementary approaches, such as machine learning and artificial intelligence, to overcome these challenges and unlock new insights in drug discovery and molecular biology.

In recent years, machine learning (ML) and artificial intelligence (AI) tools have revolutionized the exploration of the intricate relationship between chemical compounds and gene targets. These tools enable researchers to analyze vast datasets comprising chemical structures, biological activities, and genetic information to uncover novel associations and predict potential interactions [23–30]. ML algorithms, such as Support Vector Classifier (SVC), Random Forests, and deep learning models, are trained on high-dimensional data to classify compounds based on their biological activities or to predict the binding affinity of compounds to specific target proteins [31–34]. Moreover, different AI and ML tools have been widely used for drug

safety evaluation offering beneficial information for drug repurposing [35–37]. Application of advanced machine learning approaches such as k-nearest neighbors (KNN) [38], SVC [39], and extreme gradient boosting (XGB) [40], in drug repurposing exhibited high predictive performance [41,42]. Integrating these computational techniques with experimental validation can accelerate the drug discovery process and offer insights into the mechanisms of action underlying therapeutic effects.

The Tox21 dataset is a pivotal resource in the domains of predictive toxicology and drug discovery [41,43,44]. The Toxicology in the 21st Century program, Tox21, is a collaborative effort between the National Institutes of Health (NIH), the U.S. Environmental Protection Agency (EPA), and the U.S. Food and Drug Administration (FDA) [45–48]. The Tox21 dataset encompasses a plethora of biological activity information derived from screening a collection of ~10,000 drugs and environmental chemicals (Tox21 10K compound library) against a panel of *in vitro* cell-based and biochemical assays, addressing a wide spectrum of biological targets and pathways pertinent to toxicology [48–50]. The inherent advantages of the Tox21 dataset in predictive modeling for drug repurposing are rooted in its extensive scope and diversity. These characteristics facilitate the development of robust machine learning models capable of forecasting the toxicity and potential adverse effects of various chemicals and pharmaceuticals. Harnessing the vast repository of data within Tox21 offers the prospect of expediting the identification of safe and efficacious drug candidates, thereby streamlining the drug discovery process and mitigating reliance on resource-intensive experimental assays.

In this investigation, we compiled gene targets found to be significantly associated with Tox21 chemicals from our previous study [51] and developed predictive models employing four distinct ML algorithms aimed at uncovering novel gene targets associated with specific drugs. Our analysis revealed previously unrecognized gene-drug pairs, which presents opportunities for further exploration in clinical settings, thus facilitating drug repurposing endeavors across a diverse range of medical conditions.

Materials and methods

Data preparation

Tox21 data preparation. The Tox21 10K compound library contains around 10,000 substances, of which 8,971 are distinct entities, covering a broad spectrum of categories including drugs, pesticides, consumer products, food additives, industrial chemicals, and cosmetics [52]. In this study, we employed quantitative high-throughput screening (qHTS) data obtained from screening the Tox21 10K library against 78 in vitro assays. Detailed assay data are accessible from the public Tox21 website (https://tripod.nih.gov/pubdata/). Compound activity was measured by the curve rank metric, which ranges from -9–9 and is determined by various attributes of the primary concentration-response curve, including potency, efficacy, and quality. A notably positive curve rank indicates robust activation, whereas a large negative curve rank signifies potent inhibition of the assay target. Examples of compound activity scores shown in terms of curve rank are shown in Table 1. Structure ID represents CAS Registry Number of each compound. The rest column name shows Tox21 assay name. Among 8,971 substances in the original dataset, 7,170 possessed curve rank data across all Tox21 in vitro bioassays, and only compounds with available activity data were included in subsequent analyses.

Enriched gene target selection. From the previous study, 7,170 compounds in the Tox21 10K library were clustered based on similarity in their activity profiles across the Tox21 *in vitro* assays resulting in 129 clusters [51]. Gene enrichment analysis was performed on each cluster yielding a total of 737 enriched gene targets. For our models to discern patterns effectively and make accurate predictions, it's important that each target has a sufficient number of compounds known to be associated with them. Hence, we tallied the number of associated drugs for each enriched gene target, selecting only those linked with at least 10 different compounds for our models. This selection process enables us to enhance the predictive capacity and significance of our subsequent analyses. The gene targets selected for this study were prioritized based on their significant enrichment in compound activity profiles derived from the Tox21 dataset. This enrichment aligns with their known involvement in key disease pathways, particularly those implicated in rare diseases. For instance,

Table 1. Example of compound activity scores in Tox 21 dataset.

Structure ID	tox21-ache-p1_ ratio	tox21-ache-p3_ ratio	tox21-ahr-p1_ ratio	tox21-ahr-p1_ viability	tox21-ap1-agonist-p1_ch1	tox21-ap1-agonist-p1_ch2	tox21-ap1-agonist-p1_ratio
97612-24-3	0	0	0.667	-0.667	0	2	1.333
207801-27-2	0	0	3.667	0.667	0	0	0
7287-19-6	0	0	0.667	0	0	0.333	0.333
16323-43-6	0	0	0	0	0	0	0
1444-64-0	0	0	0	0	0	0	0
183321-74-6	0	0	7.667	-5	0	-2	-0.667
2404-44-6	0	0	0	0	0	0	0
439-14-5	0	0	0	0	0	0	0
1031-07-8	-5	0	0	0	-4.333	4.667	5.333
78-38-6	0	0	0	0	0	0	0
127-31-1	0	0	0.667	0	0	0	0
91-16-7	0	0	0	0	0	0	0

https://doi.org/10.1371/journal.pone.0319865.t001

the NR3C1 gene—which codes for the glucocorticoid receptor—has well-documented associations with metabolic and inflammatory pathways, making it a compelling target for drug repurposing. Similarly, the compounds included in this analysis were chosen for their robust activity scores, reflecting their potential to modulate these targets effectively. This strategic selection ensures that our predictive models focus on biologically relevant relationships, maximizing their translational potential. In summary, out of the 737 enriched genes, 143 genes associated with 6,925 compounds were included in the training set for our model. For each gene target, the number of associated drugs (represented by a value of 1 in the data matrix) ranged from 10 to 223. Conversely, all unassociated drugs with gene targets were marked with a value of 0 in the data matrix. Selected genes are detailed in Table S1.

Novel gene target prediction

We employed four modeling algorithms by using the Python packages (3.10) of SVC, KNeighborsClassifier, RandomForestClassifier, and XGBClassifier, for the task of gene target prediction. Utilizing compound activity scores as features, our objective was to predict the active or inactive relationship between each gene target and compounds. We developed the k-nearest neighbors (KNN) algorithm, valued for its interpretability. Subsequently, we introduced more sophisticated algorithms, commencing with SVCs, where we explored two different kernels: Radial Basis Function (RBF) and least square. To further augment model performance, tree-based models, namely XGB and Random Forest (RF) were investigated. The selection of four models ensured a comprehensive representation of modeling complexity while embracing popular methodologies in the field. All four modeling algorithms underwent execution on an AWS EC2 instance.

Fine-tuning and assessment of predictive models.

We performed four different modeling algorithms on all 143 gene targets: 1) KNN; 2) SVC; 3) RF; and 4) XGB. To ensure the robustness and accuracy of the models, we systematically explored various parameter configurations for each modeling algorithm, as detailed in Table S2-S5. If the values of the specific parameters were not specified in our table, default settings were employed. Subsequently, we engaged in hyperparameter fine tuning for all models, utilizing grid-search with 5-fold cross-validation (CV). It operates by exhaustively searching through a specified grid of hyperparameters, systematically evaluating the performance of the model for each combination. The dataset was partitioned into five subsets, with four subsets used for training the model and the remaining subset for validation. This process was repeated five times, with each subset serving as the validation set once. Performance metrics, such as accuracy or mean squared error,

were calculated for each parameter combination across all folds. The optimal parameters were then selected based on the average performance across all folds, providing a robust estimation of model performance while mitigating the risk of overfitting. With the fine-tuned parameters, we assessed model performance with the area under the receiver operating characteristic (ROC) curve (ROC_AUC) score and Area Under the Precision-Recall Curve (AUPRC) in our ML models. The ROC_AUC, ranging from 0 to 1, assesses a model's ability to differentiate between two classes by analyzing the true positive rate (sensitivity) versus the false positive rate (1-specificity) at different decision thresholds. A higher ROC_AUC score reflects better discrimination ability. This score offers a holistic view of model performance across thresholds, making it useful for assessing classification models in diverse fields. In our study, we also calculated the Area Under the Precision-Recall Curve (AUPRC) to address the inherent data imbalance in rare diseases, where positive drug-gene pairs are scarce. While ROC-AUC is a widely used metric, it can yield high values even when predictions favor the majority class, making it less reliable for imbalanced datasets. AUPRC, on the other hand, focuses on the precision and recall of positive cases, excluding true negatives from the calculation. This makes AUPRC a more informative and complementary metric to evaluate machine learning performance in scenarios with significant class imbalance.

Predictability of Gene Target

The predictability of a given gene target may exhibit variability across distinct modeling frameworks, while the overall predictivity of different genes may vary across these models. To elucidate this phenomenon, we computed the average testing ROC_AUC score for each gene across all models employed in our study. Genes demonstrating consistently higher mean ROC_AUC scores are deemed to possess heightened predictability, whereas those with lower scores are considered less predictable within the scope of the four machine learning models utilized in our analysis

In seeking to explicate the divergent predictability levels among genes, we inquired into whether predictivity correlates with the informational content provided to the model, specifically the number of associated compounds for each gene. Consequently, we calculated Pearson's correlation coefficient between the count of gene-associated compounds and the average ROC_AUC score. Pearson's correlation coefficient, ranging between -1 and 1, quantifies both the strength and directionality of the relationship between two variables. A negative value signifies an inverse correlation, indicating that as one variable changes, the other changes in the opposite direction. Conversely, a positive value denotes a direct correlation, wherein both variables change in tandem. The magnitude of the correlation coefficient reflects the strength of the relationship: higher absolute values indicate a stronger correlation. In our investigation, Pearson's correlation analysis serves to elucidate whether the number of associated compounds is associated with the predictability of genes.

Parameter Influence on Model Performance

In our pursuit of comprehending the impact of hyper-parameter tuning on the predictivity of the four algorithms, we systematically configured various parameter settings and juxtaposed their average performance across all gene targets. The breadth of parameter settings varies across models, contingent upon the complexity of each algorithm and the extent to which parameters can be feasibly altered. Typically, more intricate models entail a greater number of parameters with a wider range of configurational adjustments. In our exploration, we exhaustively explored parameter spaces, resulting in a myriad of settings across the models under scrutiny. Specifically, we identified 18 configurations for KNN, 54 for SVC, 1621 for RF, and 1154 for XGB. For every model, we categorized these parameter sets into three tiers: those yielding the highest ROC_AUC scores, those with moderate performance, and those resulting in suboptimal outcomes. This approach provided us with a nuanced understanding of the relationship between parameter settings and algorithmic predictivity. Subsequently, we chose 5 parameter sets in each parameter category and visualized the distribution of ROC_AUC scores across all 15 parameter sets for each model using boxplots. By scrutinizing the performance differentials between the best and worst parameter configurations, we aimed to elucidate the extent of parameter influence on algorithmic predictivity. A significant variance in performance between these configurations suggests a robust sensitivity to parameter adjustments,

whereas minimal discrepancies indicate a comparatively stable performance across genes, independent of parameter selection.

Validation of novel gene-drug pairs with chemical assay

Through fine-tuning the parameter configurations of four distinct algorithms, we identified the top three parameter sets for each algorithm, yielding a total of 12 different models. The selected parameters, along with their corresponding ROC_AUC scores for the training dataset, were documented in Table 2. Leveraging these models, we predicted the novel gene targets for the Tox 21 chemicals. To accomplish this objective, we computed the disparity between our prediction results and the records from Pharos [53] and the Board Drug Repurposing Hub (BDRH) [54]. If our model predicts a drug-gene relationship with a probability >0.5 but this relationship is not documented in Pharos or BDRH, we consider it as a potential novel gene-drug pair that can be prioritized for experimental validation. Utilizing the top three best parameter configurations for each of the four algorithms, we cross-referenced these 12 distinct prediction results. This analysis yielded a list of candidate gene-drug pairs along with the number of models supporting each prediction. Subsequently, we compared the predictive outcomes from these models with the gene annotations for the Tox21 chemicals obtained from Pharos and the BDRH. Novel gene-drug pairs were identified where the predictive models indicated a connection between a drug and gene target, yet no such association was found in Pharos or BDRH. To validate these novel relationships, we checked the experimental results of compounds with assay data available for their respective targets. For each predicted gene-drug pair, if the compound acted as an active agonist or antagonist for the gene target, our prediction was deemed correct. Conversely, if the compound was inactive against the gene target according to the assay results, our prediction was considered inaccurate. This validation process underscores the robustness of our predictive models.

Gene-rare disease association identification for drug repurposing application

To achieve the goal of drug repurposing, we aimed to identify possible rare diseases associated with the input chemicals via those newly identified gene targets. We manually searched the OMIM and Orphanet for potential associations

Table 2. Parameters configuration with top 3 average performance.

Models	Parameter Configuration 1	Parameter Configuration 2	Parameter Configuration 3
KNN	n_neighbours:33, p:3	n_neighbours:33, p:2	n_neighbours:31, p:3
SVM	C:20; gamma:0.5; kernel: rbf	C:15; gamma:0.5; kernel: rbf	C:10; gamma:0.5; kernel: rbf
RF	bootstrap: false; max_depth: 6, max_features:auto, min_samples_leaf: 2, min_samples_split:2, n_estimators: 30	bootstrap: false; max_depth: 6, max_features:auto, min_samples_leaf: 2, min_samples_split:3, n_estimators: 30	bootstrap: false; max_depth: 6, max_features:sqrt, min_samples_leaf: 2, min_samples_split:2, n_estimators: 30
XGB	Colsample_bytree: 0.6, max_depth: 5, min_child_weight: 3, reg_alpha: 1, subsample: 0.8	Colsample_bytree: 0.6, max_depth: 3, min_child_weight: 3, reg_alpha: 1, subsample: 0.8	Colsample_bytree: 0.6, max_depth: 3, min_child_weight: 1, reg_alpha: 1, subsample: 0.6

In KNN model: "n_neighbors" represents the number of neighbors considered when making predictions; "p" represents the power parameter for the Minkowski distance metric. In SVC model: The "C" parameter trades off between achieving a low training error and a low complexity model that generalizes well to unseen data. The "gamma" parameter defines how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close'. In RF model: "Max_depth" specifies the maximum depth of each decision tree in the forest. "Max_features" determines the maximum number of features considered for splitting at each node of a decision tree. "Min_samples_leaf" sets the minimum number of samples required to be at a leaf node. "Min_samples_split" sets the minimum number of samples required to split an internal node. "N_estimators" specifies the number of decision trees to be used in the random forest. In XGB model: "Colsample_bytree" determines the fraction of features (columns) to be randomly sampled for each tree during training. Similar to the parameter in Random Forest, "Max_depth" specifies the maximum depth of each decision tree in the ensemble. "Min_child_weight" determines the minimum sum of instance weight (hessian) needed in a child node. "Reg_alpha" is also known as L1 regularization term. "Reg_alpha" adds penalty to the model for large coefficient values, encouraging sparsity in the feature space. "Subsample" determines the fraction of training data to be randomly sampled for each tree.

https://doi.org/10.1371/journal.pone.0319865.t002

between the gene targets and rare diseases. These results establish connections between drugs, gene targets and rare diseases, thereby advancing the process of drug repurposing for rare conditions.

To further demonstrate our predictive results are useful for supporting drug repurposing, we conducted several case studies. We were able to make connections from chemicals to rare diseases via the predicted gene targets with evidence identified from the NCATS Biomedical Data Translator (Translator) [55]. The linkage found between the compound and disease sheds light on the drug discovery and repurposing.

Results

In this study, of 7,170 chemical compounds with curve rank data from the Tox21 10K compound library, we collected 6,925 compounds associated with 143 gene targets for novel repurposing target prediction.

Results on model development

We employed four distinct ML algorithms (SVC, KNN, RF, XGB) with the parameter set that yielded the best average performance. The optimized parameters are listed in Table 3. All four machine learning algorithms demonstrated strong performance, with ROC-AUC values exceeding 0.75. Our results show that the SVC, XGB, and RF models achieved high AUPRC values, indicating their robustness in handling imbalanced data. In contrast, the KNN model exhibited a tendency to favor true negatives, leading to comparatively lower AUPRC scores. This behavior is likely due to the simplicity of the KNN algorithm, which relies on identifying the closest data points in the training set and involves minimal model building and hyperparameter tuning. The overall performance of KNN, SVC, RF, and XGB on the training and test sets is depicted respectively in Fig 1, ROC_AUC scores were utilized to assess model performance as the y-axis in Fig 1. Notably, KNN exhibited the lowest performance compared with the other three algorithms with lowest AUPRC and ROC_AUC, while SVC emerged as the top performer. It's worth mentioning that the predictivity variation across all gene targets was minimal in SVC as well, whereas RF and XGB displayed larger varying degrees of performance.

We evaluated the model performance regarding overfitting and underfitting, as showcased in Fig 1. Impressively, all four algorithms demonstrated performance exceeding 0.7 on the test dataset. Moreover, RF and XGB exhibited enhanced performance on the test data, surpassing 0.8, while SVC and KNN displayed similar performance trends as the training dataset. These results underscore the better predictivity of tree-based models for gene targets compared to other algorithms.

In summary, our prediction models trained on the Tox21 dataset displayed robust performance across all four algorithms with ROC_AUC scores exceeding 0.7. Notably, XGB emerges as the standout performer, showing the best performance on both the training and test datasets, solidifying its status as the premier prediction algorithm among the four algorithms utilized in this study.

2. Predictability of Gene Targets

Apparently, the predictability of gene targets exhibits variations among genes based on the predictive results. It is acknowledged that the ability to accurately forecast the interactions between chemical compounds and specific genes is

Table 3. Selected Best Parameters.

Models	Parameters	Mean ROC_AUC *	Mean AUPRC *
KNN	n_neighbours:33, p:3	0.71268	0.50421
SVC	c:20, gamma:0.5, kernel: rbf	0.77428	0.80494
RF	bootstrap: false; max_depth: 6, max_features:auto, min_samples_leaf: 2, min_samples_split:2, n_estimators: 30	0.75529	0.73537
XGB	Colsample_bytree: 0.6, max_depth: 5, min_child_weight: 3, reg_alpha: 1, subsample: 0.8	0.77504	0.73602

**The mean ROC_AUC/AUPRC was calculated by averaging the ROC_AUC/AUPRC scores across all 143 selected gene targets.*

https://doi.org/10.1371/journal.pone.0319865.t003

A Table_Selected Best Parameters

Model	Parameter	Mean ROC_AUC Score
KNN	n_neighbors: 33, p: 3	0.7126818046805555
SVC	C: 20, gamma: 0.5, kernel: rbf	0.7742757875524475
RF	bootstrap: False, max_depth: 6, max_features: auto, min_samples_leaf: 2, min_samples_split: 2, n_estimators: 30	0.7552852895998671
XGB	colsample_bytree: 0.6, max_depth: 5, min_child_weight: 3, reg_alpha: 1, subsample: 0.8	0.7750385602440503

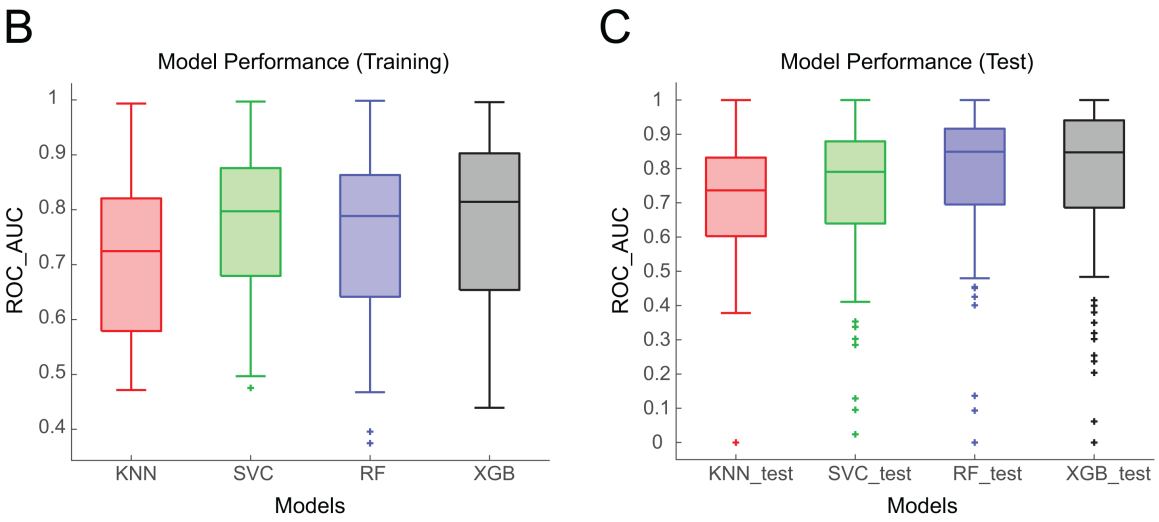

Fig 1. General model performance for different algorithms. A) Model performance on the training set. **B)** Model performance on the test set. For all plots, "+ " denotes the outliers that fall significantly outside the range of the other data points in the dataset.

https://doi.org/10.1371/journal.pone.0319865.g001

influenced by a multitude of factors, such as spanning data quality [56, 57] computational methodologies [27], and Biological Context [58]. Thus, we next aimed to delve deeper into the predictability of the 143 gene targets central to our study.

To assess the predictability of gene targets, we generated a heat map (Fig 2A) with the ROC_AUC calculated for all 143 gene targets across four different ML algorithms. As shown in Fig 2A, each cell in this heat map is the average ROC_AUC score based on the cross-validation results. The map contains 143 rows corresponding to 143 gene targets, and the 4 columns for KNN, RF, SVC, and XGB algorithms, respectively. Notably, certain genes, exemplified by NR3C1, SERPINA6, and PGR, consistently demonstrated high ROC_AUC values regardless of the modeling algorithm employed. Conversely, several genes are consistently associated with low ROC_AUC values across all models. Obviously XGB and SVC showed better performance among most of the gene targets, which is shown in red in Fig 2A.

Further elucidating the predictability landscape, Fig 2B shows the top 20 genes with high predictability across all four predictions. Particularly, the TUBB gene emerged with the highest predictability in both models of SVC and RF. Genes such as NR3C1, NR3C2, HTR2A, HTR2C, and KCNJ6 remain high predictability across all models. Meanwhile, we observed that genes illustrate different performance variability across 4 models. For example, although NR3C1 performs the third top ranked gene with the highest predictability across all models, KNN, and SVC predict the genes with less variability across different parameters. However, the predictability of NR3C1 depends more on hyperparameter tuning with XGB models.

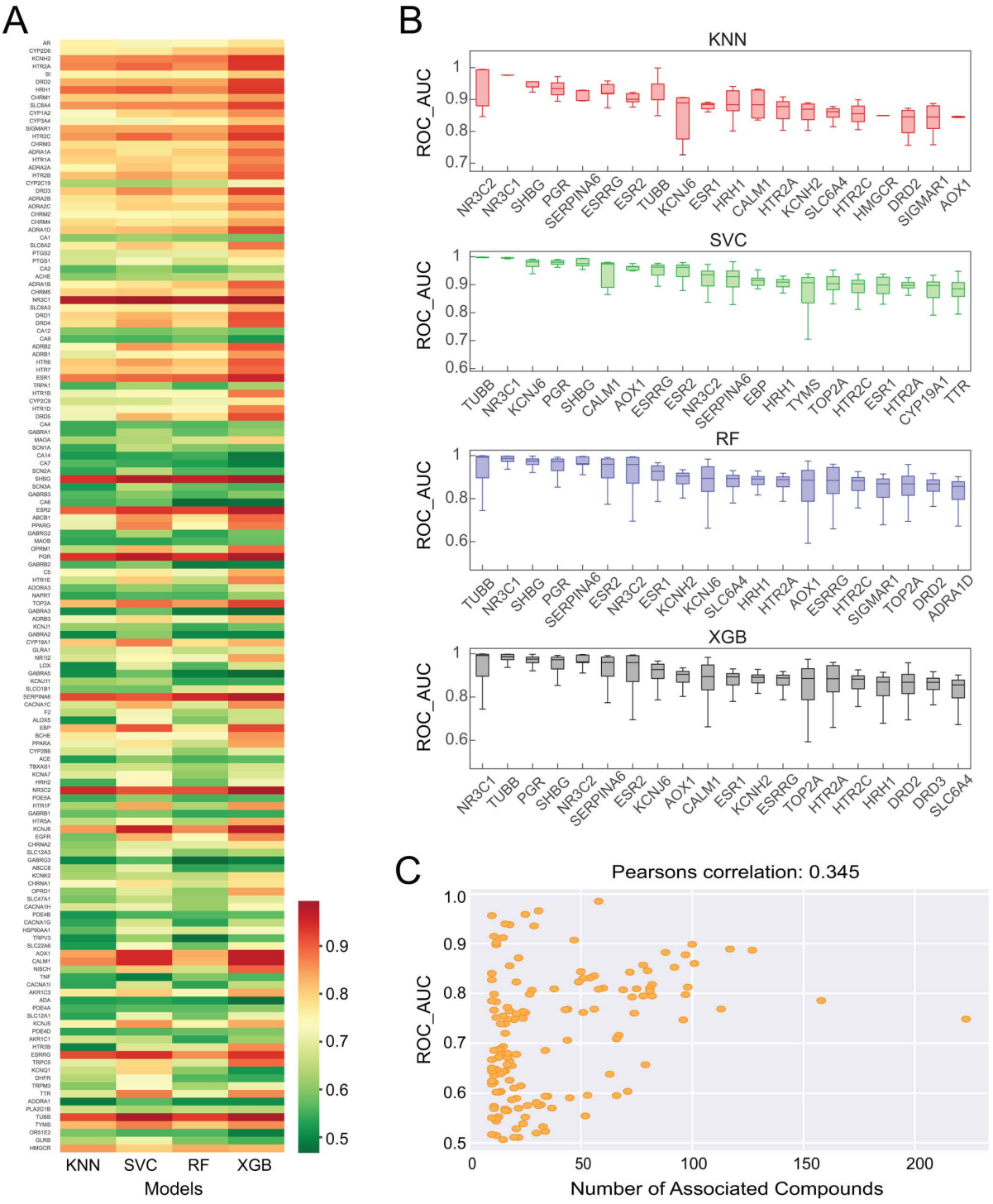

Fig 2. Gene targets prediction assessment. A) Heatmap of ROC_AUC scores across different modeling algorithms and gene targets. High ROC_AUC scores are highlighted in red, whereas low ROC_AUC scores in green. **B)** The boxplot illustrates the top 20 gene targets with the highest ROC_AUC scores for KNN, SVC, RF, and XGB. Within each boxplot, the gene targets are ordered based on their median ROC_AUC scores, from highest to lowest. **C)** The correlation between ROC_AUC scores and the number of associated compounds that each gene targets. Each yellow dot denotes one gene target.

https://doi.org/10.1371/journal.pone.0319865.g002

In our final analysis, we identified 16 gene targets consistently exhibiting low ROC_AUC values (<0.6) across all four algorithms, as outlined in Table S6. To probe the underlying factors contributing to the divergent predictability among these genes, we hypothesized that the number of associated compounds might significantly impact the predictive results. By employing Pearson's correlation coefficient, we found weak positive correlations between the number of associated compounds and ROC_AUC scores, with Pearson's correlation coefficient registering at 0.345 (Fig 2C). Despite this correlation, the weak magnitude prompts consideration of alternative factors causing the variability in gene predictivity, such as the diversity of compounds associated with a gene target.

In summary, our models effectively predicted the remaining 127 gene targets, encompassing nearly 90% of the total gene targets examined. This noteworthy achievement underscores the reliability and robustness of our predictive framework, thereby validating its potential to expedite accurate gene target prediction within the realms of drug discovery and molecular biology research.

Parameter influence on model performance

In our quest to comprehend the extent to which hyper-parameter tuning affects the prediction, we configured different parameter settings and compared their average performance across all gene targets. Fifteen sets of parameters for each model were selected based on their performance, including the top five performers, the bottom five performers, and those in the middle tier of performance (Listed in Table S7). Illustrated in Fig 3, the median performance of various configurations for KNN, RF, SVC, and XGB unveils intriguing insights.

Our analysis reveals that parameter fine-tuning is highly impacting on the RF model, while the performance of the KNN algorithm exhibits relatively minimal fluctuations with varying hyper-parameters. Notably, the RF model consistently yields ROC_AUC values exceeding 0.8 with a particular set of hyper parameters. However, the XGB model demonstrated persistent performance with an average ROC_AUC of around 0.75 regardless of the hyper-parameters. Collectively, our findings disclosed the profound impact of hyper-parameter selection on certain modeling algorithms, such as RF, while exerting less influence on others, such as KNN, and XGB has higher stability across all configurations.

Results on validation of candidate gene-drug pairs

The primary objective of constructing prediction models is to identify potential new gene-drug connections for drug repurposing. To validate our predictions, we explored available in vitro assay data to see if any of the predicted gene-drug pairs have experimental support.

The process of identifying potentially new gene-drug connections is described in Methods. We compared our predictions with records from Pharos and the Board Drug Repurposing Hub (BDRH). If our model predicts a connection with probability >0.5 that was not documented in Pharos or BDRH, we identify it as a potential novel gene-drug pair for further experimental validation. We utilized the top three parameter configurations for each of the four algorithms, resulting in 12 distinct predictions. Cross-referencing these predictions produced a list of candidate gene-drug pairs along with the count of models supporting each prediction. In total, we uncovered 220 gene-drug pairs supported by at least one ML model but not documented in Pharos or the BDRH (Examples are shown in Table 4 and the full list can be found in Table S8). We manually identified 60 gene-drug pairs that have in vitro assay data available, 52 of which were supported by experimental results, that is, the compounds acted either as active agonists or active antagonists of their respective targets in these assays, implying a confirmation rate (86.7%) exceeding 85% (Examples are shown in Table 5 and full list in Table S9).

Results on identifying gene-rare disease associations for drug repurposing

We conducted a manual search of the OMIM and Orphanet database to ascertain whether the gene targets identified in our novel gene-drug pairs are associated with any rare diseases. The findings are summarized in Table 6, revealing a total

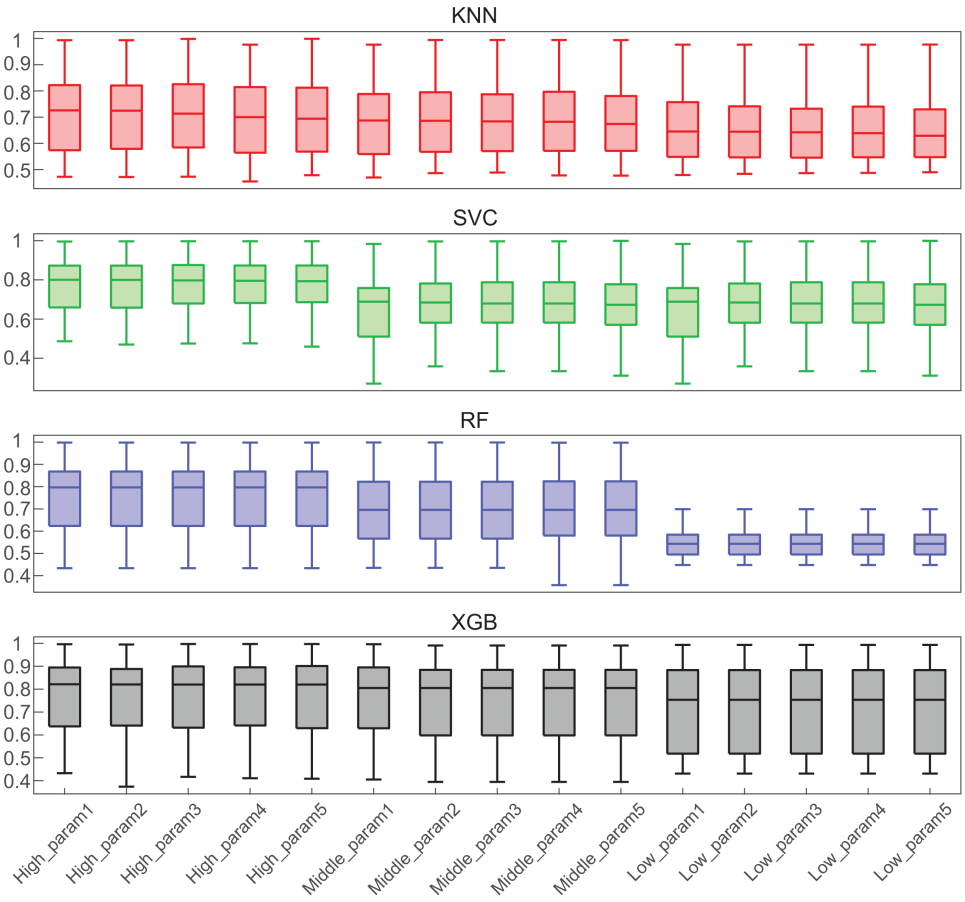

Fig 3. Impact of hyper-parameter fine-tuning on model performance. The boxplot showcases the 15 ROC_AUC scores, comprising the 5 highest, 5 intermediate, and 5 lowest, for KNN, SVC, RF, and XGB. For each boxplot, the parameter configurations are ordered based on their median ROC_AUC scores, from highest to lowest.

https://doi.org/10.1371/journal.pone.0319865.g003

Table 4. Examples of gene-drug pairs predicted by different models.

Models	Interest Gene Targets	Structure ID
[KNN_1, KNN_2, KNN_3, RF_1, RF_2, RF_3, SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'NR3C1'	127-31-1
[KNN_1, KNN_2, KNN_3, RF_1, RF_2, RF_3, SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'NR3C1'	1310709-74-0
[KNN_1, KNN_2, KNN_3, RF_1, RF_2, RF_3, SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'NR3C1'	59198-70-8
[KNN_2, RF_1, RF_2, RF_3, SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'SHBG'	10161-33-8
[SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'ACHE'	102518-79-6
[RF_1, RF_3, SVC_3, XGB_1, XGB_2, XGB_3]	'ADRB3'	32266-10-7
[SVC_1, SVC_2, SVC_3, XGB_1, XGB_2, XGB_3]	'DRD4'	75444-65-4
[KNN_1, KNN_3, RF_1, RF_3, XGB_1, XGB_2]	'HTR2A'	75859-03-9
[KNN_1, KNN_3, RF_2, XGB_1, XGB_2, XGB_3]	'HTR2C'	75859-03-9

https://doi.org/10.1371/journal.pone.0319865.t004

of 35 distinct rare diseases linked to various gene targets for which potential drugs have been identified. Further exploration of these disease-drug relationships promises to provide insights into drug repurposing strategies tailored for rare diseases.

Table 5. Examples of gene-drug pairs validated by Tox21 in vitro assay data.

Gene Targets	Structure ID	Assay Names	Assay Outcome
'CYP1A2'	81-14-1	tox21-p450-1a2-p1	active antagonist
'CYP1A2'	15930-66-2	tox21-p450-1a2-p1	active antagonist
'CYP2D6'	303-49-1	tox21-p450-2d6-p1	active antagonist
'CYP2D6'	1104-22-9	tox21-p450-2d6-p1	active antagonist
'CYP2D6'	113-92-8	tox21-p450-2d6-p1	active antagonist
'DRD2'	959-24-0	tox21-drd2-agonist-p1/tox21-drd2-antagonist-p1	inactive/inactive
'ESR1'	26538-44-3	tox21-er-luc-bg1-4e2-antagonist-p1	active antagonist
'KCNH2'	75859-03-9	tox21-herg-u2os-p1	active antagonist
'KCNH2'	553-08-2	tox21-herg-u2os-p1	active antagonist
'NR3C1'	127-31-1	tox21-gr-hela-bla-antagonist-p1/tox21-gr-hela-bla-agonist-p1	active agonist

https://doi.org/10.1371/journal.pone.0319865.t005

Application of novel gene targets for drug repurposing

The principal objective of this study is to advance the frontier of drug repurposing. Consequently, the elucidation of a cohesive relationship between pharmaceutical compounds and pathological conditions assumes paramount importance. We are now embarking on a quest to identify rare diseases intricately linked to specific gene targets by utilizing the OMIM and Orphanet databases. Our aim in this phase is to outline various pathways that elucidate how drugs implicated in the modulation of these genes may offer therapeutic avenues for the treatment of the rare diseases. Through the following examples, we demonstrate how our research findings facilitate the establishment of robust connections between rare diseases and pharmaceutical agents via the identified novel gene targets, thereby presenting promising avenues for therapeutic exploration and intervention.

Case study 1. Candidate Drugs for Generalized Glucocorticoid Resistance (GCCR). The NR3C1 gene has been predicted as a novel gene for eight distinct compounds, as listed in Table 7. GCCR (GARD:0002499) is a rare adrenogenital syndrome characterized by generalized, partial tissue insensitivity to glucocorticoids, and it is caused by heterozygous mutation in the glucocorticoid receptor gene (NR3C1) on chromosome 5q31 [59]. With the connections between eight chemical compounds and the gene NR3C1, and NR3C1 and GCCR, we aimed to explore the potential use of these drugs in treating GCCR.

Through our analysis of scientific evidence mined from the Translator ecosystem, we found that Fludrocortisone, Rimexolone, and Fluoxymesterone can modulate the NR3C1 gene mainly by binding to it and activating its signaling pathway [60–62] and subsequently influencing GCCR by restoring the function of the glucocorticoid receptor [63–66]. These three drugs as synthetic corticosteroids are commonly used to treat adrenal insufficiency diseases like Addison's disease [62,67–69]. Additionally, research indicates that GCCR patients typically exhibit deficiencies in adrenal corticosteroids, including cortisol and aldosterone [70–73]. Therefore, Fludrocortisone, Rimexolone, and Fluoxymesterone may potentially treat GCCR by providing adrenal corticosteroids. Rimexolone is shown as an example in Fig 4, other examples can be found under the column 'Scientific evidence from the Translator' in Table 7.

Moreover, Rimexolone and Flunisolide, two of the aforementioned drugs, are known to affect the glucocorticoid receptor (GR) by binding to it with high affinity [74]. As cortisol action mediated by the GR is diminished in GCCR patients [75, 76], Rimexolone and Flunisolide may also be effective in treating GCCR by enhancing GR binding. Hormone replacement therapy is another pivotal approach to maintaining hormonal balance in patients [77,78], which could aid in treating GCCR. Among the eight drugs, Diflucortolone valerate and Melengestrol acetate are connected to GCCR through hormone pathways, suggesting their potential use in hormone replacement therapy. Additionally, Deoxycorticosterone acetate (DOCA) is documented as a drug that induces one of the GCCR phenotypes, hypertensive disorder. Research shows that DOCA is often used to induce hypertension in animal models [79–81], thus reducing DOCA levels may partially alleviate GCCR symptoms.

Table 6. Target Gene and Related Rare Diseases based on OMIM and Orphanet.

Gene Targets	Rare Disease Records	OMIM Number
ACHE	YT BLOOD GROUP ANTIGEN	OMIM: 112100
ADRA2A	LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 8; FPLD8	OMIM: 620679
ADRB1	RESTING HEART RATE, VARIATION IN	OMIM: 607276
ADRB1	SHORT SLEEP, FAMILIAL NATURAL, 2; FNSS2	OMIM: 618591
AR	Spinal and bulbar muscular atrophy, X-linked 1	OMIM: 313700
AR	Hypospadias 1, X-linked	OMIM: 30633
AR	ANDROGEN INSENSITIVITY, PARTIAL; PAIS	OMIM: 312300
AR	ANDROGEN INSENSITIVITY SYNDROME; AIS	OMIM: 300068
BCHE	BUTYRYLCHOLINESTERASE DEFICIENCY; BCHED	OMIM: 617936
C5	ECULIZUMAB, POOR RESPONSE TO	OMIM: 615749
C5	COMPLEMENT COMPONENT 5 DEFICIENCY; C5D	OMIM: 609536
CA12	Hyperchlorhidrosis, isolated	OMIM: 143860
CA2	Osteopetrosis, autosomal recessive 3, with renal tubular acidosis	OMIM: 259730
CALM1	Long QT syndrome 14	OMIM: 616247
CALM1	Ventricular tachycardia, catecholaminergic polymorphic, 4	OMIM: 614916
CHRM3	PRUNE BELLY SYNDROME; PBS	OMIM:100100
CYP19A1	AROMATASE DEFICIENCY	OMIM: 613546
CYP19A1	Aromatase excess syndrome	OMIM: 139300
CYP2C9	COUMARIN RESISTANCE	OMIM: 12270
DRD3	TREMOR, HEREDITARY ESSENTIAL, 1; ETM1	OMIM: 190300
ESR1	BREAST CANCER, FAMILIAL MALE, INCLUDED	OMIM: 114480
GABRB2	Developmental and epileptic encephalopathy 92	OMIM: 617829
GABRG2	Generalized epilepsy with febrile seizures plus, type 3	OMIM: 607681
GABRG2	FEBRILE SEIZURES, FAMILIAL, 8; FEB8	OMIM: 607681
GABRG2	DEVELOPMENTAL AND EPILEPTIC ENCEPHALOPATHY 74; DEE74	OMIM: 618396
KCNH2	LONG QT SYNDROME 2; LQT2	OMIM: 613688
KCNH2	SHORT QT SYNDROME 1; SQT1	OMIM: 609620
NR3C1	GLUCOCORTICOID RESISTANCE, GENERALIZED; GCCR	OMIM: 615962
PPARG	LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 3; FPLD3	OMIM: 604367
SI	SUCRASE-ISOMALTASE DEFICIENCY, CONGENITAL; CSID	OMIM: 609845
SIGMAR1	AMYOTROPHIC LATERAL SCLEROSIS 16, JUVENILE; ALS16	OMIM: 614373
SIGMAR1	NEURONOPATHY, DISTAL HEREDITARY MOTOR, AUTOSOMAL RECESSIVE 2; HMNR2	OMIM: 605726
SLC6A2	ORTHOSTATIC INTOLERANCE	OMIM: 604715
SLC6A3	PARKINSONISM-DYSTONIA 1, INFANTILE-ONSET; PKDYS1	OMIM: 613135
SLCO1B1	HYPERBILIRUBINEMIA, ROTOR TYPE; HBLRR	OMIM: 237450

https://doi.org/10.1371/journal.pone.0319865.t006

In summary, our study identifies several drugs with the potential to treat GCCR through different pathways, including increasing adrenal corticosteroids, enhancing GR binding efficiency, hormone replacement therapy and alleviating GCCR symptoms. Meanwhile, there are other compounds, e.g., Cortodoxone, that can be further investigated and shows weak relationship to GCCR according to current research recorded in Translator.

Case study 2. Candidate Drugs for short QT syndrome. In our results, KCNH2 was predicted as a novel gene target of eleven drugs listed in Table 8. KCNH2 provides instructions for making channels that transport positively charged atoms (ions) of potassium out of cells, and mutations in the KCNH2 gene can cause short QT syndrome (OMIM: 609620) [82]. Thus, we speculated that drugs linked to KCNH2 might present promising candidates for short QT syndrome treatment.

Table 7. Candidate compounds for GCCR.

Compounds	Original indication	Use for GCCR	Scientific evidence from the Translator
Rimexolone	Employed in ophthalmology to treat eye inflammation and allergic eye diseases like keratitis and conjunctivitis	Providing adrenal corticosteroids; Enhancing GR binding	https://arax.ncats.io/?r=236272
Fluoxymesterone	Treatment of hypogonadism	Providing adrenal corticosteroids; Enhancing GR binding	https://arax.ncats.io/?r=241052
Fludrocortisone	Adrenal insufficiency diseases like Addison's disease	Providing adrenal corticosteroids	https://arax.ncats.io/?r=236165
Melengestrol acetate	Growth-promoting agent in livestock	Hormone replacement therapy	https://arax.ncats.io/?r=241051
Deoxycorticosterone acetate	Adrenal insufficiency diseases	Alleviate one GCCR symptom, hypertension	https://arax.ncats.io/?r=236273
Diflucortolone valerate	Combat skin inflammation and allergic reactions such as eczema and dermatitis	Hormone replacement therapy	https://arax.ncats.io/?r=241047
Cortodoxone	Adrenal insufficiency diseases like Addison's disease	Weak relationship through ethanol.	https://arax.ncats.io/?r=236270
Halometasone hydrate	Combat skin inflammation and allergic reactions such as eczema and dermatitis	N/A	https://arax.ncats.io/?r=241048

https://doi.org/10.1371/journal.pone.0319865.t007

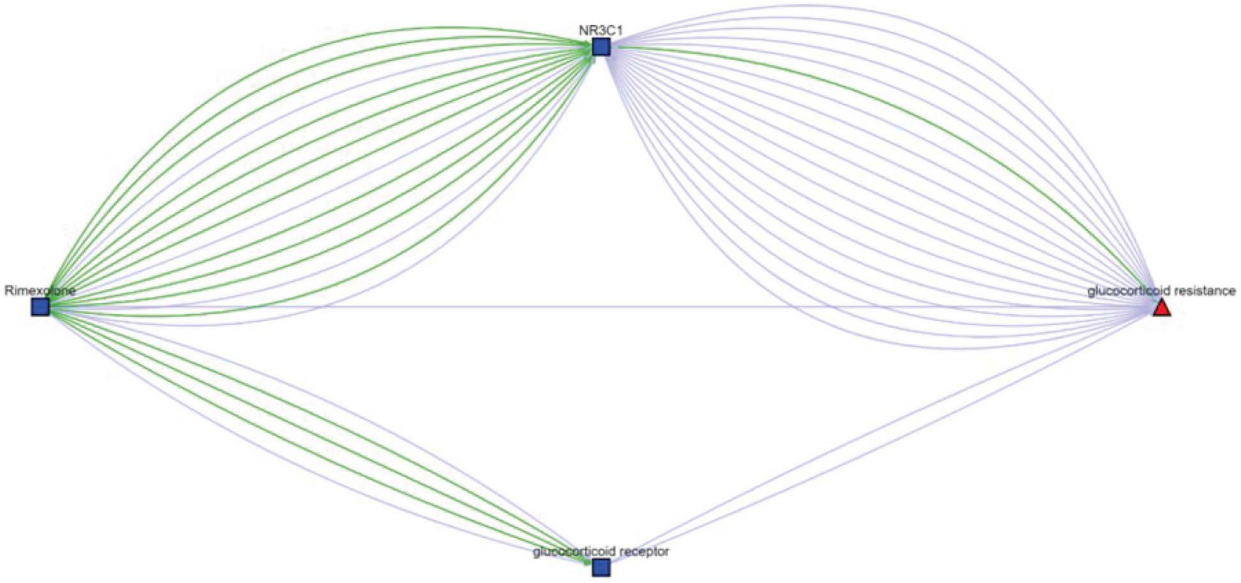

Fig 4. Associations between Rimexolone and Generalized Glucocorticoid Resistance from the Translator. (the original graph can be accessed via https://arax.ncats.io/?r=236272).

https://doi.org/10.1371/journal.pone.0319865.g004

To test our hypothesis, we examined the scientific evidence identified from the Translator, and among these eleven drugs, we found that Bucindolol, triprolidine, Cyproheptadine Hydrochloride, and Thonzonium bromide are connected to short QT syndrome via the KCNH2 gene. Thonzonium bromide is shown as an example in Fig 5, The KCNH2 gene, also referred to as the hERG gene, encodes a protein forming channels in cardiac muscle cell membranes [83]. These channels regulate potassium ion flow, crucial for cardiac rhythm maintenance [84,85]. Cyproheptadine has been demonstrated to interfere with the hERG channel function, while Bucindolol and triprolidine reduce the KCNH2 protein activity [86]. In

Table 8. Candidate compounds for short QT syndrome.

Compounds	Original indication	Use for short QT syndrome	Scientific evidence from the Translator
Bucindolol	Management of heart failure and hypertension	KCNH2 deactivation; β-blocker of calcium channel	https://arax.ncats.io/?r=241069
Thonzonium bromide	Used in topical formulations for its ability to inhibit the growth of bacteria, fungi, and other microorganisms	KCNH2 deactivation	https://arax.ncats.io/?r=241079
triprolidine	Symptomatic relief of allergic conditions such as hay fever	KCNH2 deactivation	https://arax.ncats.io/?r=241074
Cyproheptadine hydrochloride	Allergic conditions; Serotonin Syndrome; Migraine Prophylaxis	KCNH2 deactivation	https://arax.ncats.io/?r=241077
iloperidone	Treatment of schizophrenia	Induced tachycardia	https://arax.ncats.io/?r=241071
Clomipramine hydrochloride	Treatment of various mental health conditions, including: Obsessive-Compulsive Disorder (OCD), panic disorder.	Weak relationship through Chloride ion	https://arax.ncats.io/?r=241073
Trimipramine maleate	Treatment of depression; anxiety disorders; insomnia	Weak relationship connected by prelamin-A/C	https://arax.ncats.io/?r=241078
Mebeverine hydrochloride	Alleviate symptoms associated with irritable bowel syndrome (IBS) and related gastrointestinal disorders	N/A	https://arax.ncats.io/?r=241070
Rimcazole dihydrochloride	Preclinical studies only	N/A	https://arax.ncats.io/?r=241072
Promethazine	Allergic conditions; nausea and vomiting; sedation and anxiolysis; insomnia	N/A	https://arax.ncats.io/?r=241075
AVE8923	potential therapeutic agent for cardiovascular conditions	N/A	Cannot find this drug in translator

https://doi.org/10.1371/journal.pone.0319865.t008

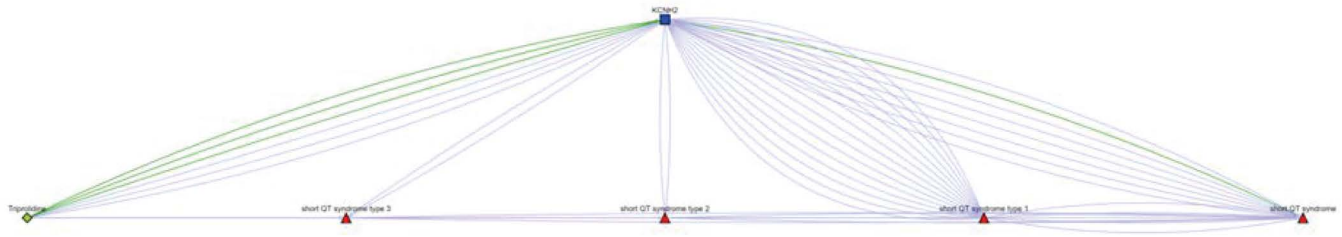

Fig 5. Associations between triprolidine and short QT syndrome from the Translator. (the original graph can be accessed via https://arax.ncats.io/?r=241074).

https://doi.org/10.1371/journal.pone.0319865.g005

the meanwhile, research shows that accelerated KCNH2 deactivation is linked to arrhythmogenesis, a key symptom of short QT syndromes [87]. Thus, these drugs hold potential in short QT syndrome treatment through KCNH2 deactivation. Notably, Bucindolol, a β-blocker formerly employed in heart-related conditions, has been phased out due to its ambiguous efficacy [88,89].

Although no published studies on iloperidone specifically address its relationship with KCNH2 gene targets, we identified the potential indication of iloperidone for short QT syndrome via Translator. Notably, iloperidone-induced tachycardia, a short QT syndrome phenotype, suggests its relevance (table reference column). Hence, iloperidone warrants investigation as a potential treatment for cardiac diseases with similar phenotypes.

In conclusion, our findings suggest that drugs identified in our study can modulate KCNH2 gene target function, holding promise for short QT syndrome treatment.

Discussion

In our study, we introduced a systematic approach to identify novel relationships between chemical compounds and gene targets and rare diseases based on biological activity profiles generated by using Tox21 bioassay screening data towards drug repurposing. With the developed ML models achieving an overall prediction accuracy reflected by ROC-AUC scores exceeding 0.7 and showing comparable values in APRUC metrics, we were able to identify 220 potential pairs of gene targets and compounds. Of these, 60 pairs had public assay records, with 52 of them validated by experimental outcomes. In addition, the two case studies further underscore the robustness of our approach in predicting novel gene targets for drug repurposing and disease treatment initiatives.

Research indicates that algorithms with more intricate architectures, such as SVC or tree-based ML models, often exhibit better statistical performance. Our study corroborates this trend, as our results consistently demonstrate that RF, SVC, and XGB consistently outperform KNN models across all 143 genes analyzed. Furthermore, our investigation into hyper-parameter tuning revealed an interesting pattern: while fine-tuning had a significant impact on RF algorithms, its effect was less pronounced in KNN and XGB models. This phenomenon can be attributed to the intricate nature of the models and the multitude of hyper-parameters that can be fine-tuned to optimize performance. Thus, the complexity of the algorithm architecture and the flexibility in parameter adjustments likely contribute to the varying degrees of sensitivity to fine-tuning observed across different ML models.

In our study, we have identified 16 gene targets that exhibit a lack of predictability. The predictability of gene targets, defined as the capacity of ML models to accurately predict interactions between chemical compounds and specific genes, is influenced not only by computational algorithms but also by various other factors such as data quality, data quantity, and biological complexities. To delve deeper into this issue, we have explored the correlation between the quantity of associated compounds and the predictability of gene targets. Our analysis reveals a slightly positive correlation between the number of known compounds of gene targets and their predictability. It appears that gene targets with a limited number of associated compounds tend to exhibit poorer model performance. This could be the first reason that hinders the predictability of these 16 genes (maximum number of related compounds: 52; minimal number of related compounds: 10; Mean: 22). Conversely, genes involved in well-characterized cellular pathways or disease processes may demonstrate higher predictability due to their well-understood functions and interactions. Additionally, the lack of associated compounds further indicates a poor understanding of these gene targets, contributing to their low predictability across different models. Thus, it becomes evident that the absence of sufficient biological context significantly impacts the prediction accuracy of these gene targets.

To accomplish our primary objective of drug repurposing using our ML models, we have successfully identified 220 gene-target pairs for drugs that were not previous reported. We checked 60 pairs that have in vitro assay data and found that 52 pairs were confirmed by experimental evidence. We also found that NR3C1 genes are associated with multiple compounds that are supported by nearly all different predictive models. The NR3C1 gene codes for the glucocorticoid receptor (GR). Changes in NR3C1 gene targets lead to not only common diseases like polycystic ovarian syndrome but also result in rare diseases like GCCR. Besides that, we also found a great number of compounds related to the family of HTR2, PTGS, and DRD gene targets. All these gene targets are highly related to common and rare diseases such as schizophrenia, gastric ulcer, urticaria, and endogenous depression. Thus, predicting the potential drugs associated with these gene targets will provide meaningful information for drug repurposing and future clinical studies.

Besides those gene-drug pairs that have been validated with Tox21 assay data, the newly predicted drug candidates from this study provide additional opportunities for future investigation. For example, our model predicts the novel association between metoclopramide and the gene ADRA2A. ADRA2A is related to Familial Partial Lipodystrophy (FPLD) [90], and metoclopramide is connected to different kinds of FPLD through multiple pathways including hypertensive disorder [91]. Thus, we hypothesized this might be a new potential solution for treating FPLD with metoclopramide via ADRA2A.

In general, the findings of this study hold significant translational potential for addressing the unmet medical needs in rare diseases. By utilizing machine learning models, we identified novel gene-drug associations that provide a foundation for targeted therapeutic development. For instance, the high predictability of gene targets such as NR3C1 and KCNH2, coupled with their relevance to rare diseases like Generalized Glucocorticoid Resistance and Short QT Syndrome, demonstrates the potential to prioritize drug candidates for preclinical and clinical evaluations. Furthermore, the diversity of associated compounds for these targets enables the exploration of multiple therapeutic pathways. This approach is particularly valuable for rare diseases, where limited research funding often constrains the development of targeted treatments. By focusing on well-characterized gene-disease relationships, such as those supported by existing evidence, this study makes it feasible to design clinical studies with higher confidence in the underlying biology. Our methodology also has broader implications for drug repurposing in rare diseases. The ability to predict novel gene-drug relationships and validate them experimentally highlights a systematic pipeline that can be applied to other underexplored diseases. This not only accelerates the identification of promising therapeutic candidates but also provides a scalable framework for addressing the broader challenges of drug discovery in rare disease contexts.

Despite the promising results from our models, our study acknowledges several limitations. The imbalance of the dataset, characterized by a disproportionately low number of positive drug-gene pairs, can lead to skewed predictions favoring the major class. While we employed metrics like AUPRC to mitigate this, future studies would benefit from augmented datasets with more balanced distributions. Additionally, the varying performance of machine learning models highlights the need to balance model complexity and interpretability. Simpler models like KNN underperform due to their inability to capture intricate relationships, while more complex models like XGB, although more accurate, present challenges in biological interpretation due to their computational complexity. Finally, the predictability of certain gene targets is constrained by limited biological context, emphasizing the importance of integrating additional functional data into future analyses.

Supporting information

Table S1. Selected gene list.
(CSV)

Table S2. Knn parameterSets.
(XLSX)

Table S3. SVC parameterSets.
(XLSX)

Table S4. RF parameterSets.
(XLSX)

Table S5. XGB parameterSets.
(XLSX)

Table S6. Gene with low predictability.
(CSV)

Table S7. High middle low performance parameters.
(XLSX)

Table S8. All gene compound pairs.
(XLSX)

Table S9. Gene compound pairs validated byTox21.
(XLSX)

Acknowledgements

We would like to thank Dr. Shixue Sun for his helpful discussion on study design. The analyses described in this publication were conducted with data and/or tools accessed through the NCATS Biomedical Data Translator. (https://ncats.nih.gov/translator).

Author contributions

Conceptualization: Qian Zhu.

Data curation: Binghan Xue.

Formal analysis: Binghan Xue.

Investigation: Binghan Xue.

Methodology: Binghan Xue.

Project administration: Qian Zhu.

Resources: Yanji Xu, Ruili Huang.

Supervision: Qian Zhu.

Validation: Binghan Xue.

Visualization: Binghan Xue.

Writing – original draft: Binghan Xue.

Writing – review & editing: Binghan Xue, Ruili Huang, Qian Zhu.

Funding

References

1. Oprea TI, Bauman JE, Bologa CG, Buranda T, Chigaev A, Edwards BS, et al. Drug repurposing from an academic perspective. Drug Discov Today Ther Strateg. 2011;8(3-4):61-9.
2. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58. https://doi.org/10.1038/nrd.2018.168 PMID: 30310233
3. De Rosa MC, Purohit R, García-Sosa AT. Drug repurposing: a nexus of innovation, science, and potential. Sci Rep. 2023;13(1):17887. https://doi.org/10.1038/s41598-023-44264-7 PMID: 37857641
4. Kulkarni VS, Alagarsamy V, Solomon VR, Jose PA, Murugesan S. Drug repurposing: an effective tool in modern drug discovery. Russ J Bioorg Chem. 2023;49(2):157–66. https://doi.org/10.1134/S1068162023020139 PMID: 36852389
5. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239–49. https://doi.org/10.1111/j.1476-5381.2010.01127.x PMID: 21091654
6. Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov. 2018;17(3):183–96. https://doi.org/10.1038/nrd.2017.226 PMID: 29217837
7. Wang G, Xu Y, Wang Q, Chai Y, Sun X, Yang F, et al. Rare and undiagnosed diseases: From disease-causing gene identification to mechanism elucidation. Fundam Res. 2022;2(6):918–28. https://doi.org/10.1016/j.fmre.2022.09.002 PMID: 38933382
8. Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov. 2011;10(3):188–95. https://doi.org/10.1038/nrd3368 PMID: 21358738
9. Soheilmoghaddam F, Rumble M, Cooper-White J. High-Throughput Routes to Biomaterials Discovery. Chem Rev. 2021;121(18):10792–864. https://doi.org/10.1021/acs.chemrev.0c01026 PMID: 34213880
10. Yang L, Pijuan-Galito S, Rho HS, Vasilevich AS, Eren AD, Ge L, et al. High-throughput methods in the discovery and study of biomaterials and materiobiology. Chem Rev. 2021;121(8):4561–677. https://doi.org/10.1021/acs.chemrev.0c00752 PMID: 33705116

11. Nagamine N, Sakakibara Y. Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. Bioinformatics. 2007;23(15):2004–12. https://doi.org/10.1093/bioinformatics/btm266 PMID: 17510168
12. Morris GM, Lim-Wilby M. Molecular docking. Methods Mol Biol. 2008;443:365–82.
13. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics. 2008;24(13):i232–40. https://doi.org/10.1093/bioinformatics/btn162 PMID: 18586719
14. Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. 2011;7(2):146–57. https://doi.org/10.2174/157340911795677602 PMID: 21534921
15. Lin W. Major challenges of molecular docking. 2023.
16. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins. 2006;65(1):15–26. https://doi.org/10.1002/prot.21082 PMID: 16862531
17. Huang S-Y, Zou X. Advances and challenges in protein-ligand docking. Int J Mol Sci. 2010;11(8):3016–34. https://doi.org/10.3390/ijms11083016 PMID: 21152288
18. Milano M, Agapito G, Cannataro M. Challenges and limitations of biological network analysis. BioTech (Basel). 2022;11(3):24. https://doi.org/10.3390/biotech11030024 PMID: 35892929
19. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015;20(7):13384–421. https://doi.org/10.3390/molecules200713384 PMID: 26205061
20. Zhao J, Lin F, Liang G, Han Y, Xu N, Pan J. Exploration of the molecular mechanism of Polygonati rhizoma in the treatment of osteoporosis based on network pharmacology and molecular docking. Frontiers in Endocrinology. 2022;12.
21. Korb O, Olsson TSG, Bowden SJ, Hall RJ, Verdonk ML, Liebeschuetz JW, et al. Potential and limitations of ensemble docking. J Chem Inf Model. 2012;52(5):1262–74. https://doi.org/10.1021/ci2005934 PMID: 22482774
22. Zhou Y, Jiang Y, Chen S-J. RNA–ligand molecular docking: advances and challenges. WIREs Computational Molecular Science. 2022;12(3):e1571.
23. Agarwal S, Dugar D, Sengupta S. Ranking chemical structures for drug discovery: a new machine learning approach. J Chem Inf Model. 2010;50(5):716–31. https://doi.org/10.1021/ci9003865 PMID: 20387860
24. O'Boyle N, Dalke A. DeepSMILES: an adaptation of SMILES for use in machine-learning of chemical structures. Journal Name. 2018
25. Xu T, Ngan DK, Ye L, Xia M, Xie HQ, Zhao B, et al. Predictive models for human organ toxicity based on in vitro bioactivity data and chemical structure. Chem Res Toxicol. 2020;33(3):731–41. https://doi.org/10.1021/acs.chemrestox.9b00305 PMID: 32077278
26. Meuwly M. Machine Learning for Chemical Reactions. Chem Rev. 2021;121(16):10218–39. https://doi.org/10.1021/acs.chemrev.1c00033 PMID: 34097378
27. Wu L, Huang R, Tetko IV, Xia Z, Xu J, Tong W. Trade-off predictivity and explainability for machine-learning powered predictive toxicology: an in-depth investigation with tox21 data sets. Chem Res Toxicol. 2021;34(2):541–9. https://doi.org/10.1021/acs.chemrestox.0c00373 PMID: 33513003
28. Staszak M, Staszak K, Wieszczycka K, Bajek A, Roszkowski K, Tylkowski B. Machine learning in drug design: Use of artificial intelligence to explore the chemical structure–biological activity relationship. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2022;12(2):e1568. https://doi.org/10.1002/wcms.1568
29. Tullius Scotti M, Herrera-Acevedo C, Barros de Menezes RP, Martin H-J, Muratov EN, Ítalo de Souza Silva Á, et al. MolPredictX: online biological activity predictions by machine learning models. Mol Inform. 2022;41(12):e2200133. https://doi.org/10.1002/minf.202200133 PMID: 35961924
30. Xu T, Xia M, Huang R. Modeling tox21 data for toxicity prediction and mechanism deconvolution. machine learning and deep learning in computational toxicology: Springer; 2023. p. 463–77.
31. Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;26(9):1169–75. https://doi.org/10.1093/bioinformatics/btq112 PMID: 20236947
32. Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdiscip Rev Comput Mol Sci. 2015;5(6):405–24. https://doi.org/10.1002/wcms.1225 PMID: 27110292
33. Heck GS, Pintro VO, Pereira RR, de Ávila MB, Levin NMB, de Azevedo WF. Supervised machine learning methods applied to predict ligand- binding affinity. Curr Med Chem. 2017;24(23):2459–70. https://doi.org/10.2174/0929867324666170623092503 PMID: 28641555
34. Ji B, He X, Zhai J, Zhang Y, Man VH, Wang J. Machine learning on ligand-residue interaction profiles to significantly improve binding affinity prediction. Brief Bioinform. 2021;22(5):bbab054. https://doi.org/10.1093/bib/bbab054 PMID: 33758923
35. Rácz A, Bajusz D, Miranda-Quintana RA, Héberger K. Machine learning models for classification tasks related to drug safety. Mol Divers. 2021;25(3):1409–24. https://doi.org/10.1007/s11030-021-10239-x PMID: 34110577
36. Al-Worafi YM. Artificial intelligence and machine learning for drug safety. technology for drug safety: current status and future developments: Springer; 2023. p. 69–80.
37. Zhang D, Song J, Dharmarajan S, Jung T, Lee H, Ma Y. The Use of Machine Learning in Regulatory Drug Safety Evaluation. Statistics in Biopharmaceutical Research. 2023;15(3):519–23.

38. Pham T-H, Qiu Y, Zeng J, Xie L, Zhang P. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. Nat Mach Intell. 2021;3(3):247–57. https://doi.org/10.1038/s42256-020-00285-9 PMID: 33796820
39. Zhao K, So H-C. Using drug expression profiles and machine learning approach for drug repurposing. Computational methods for drug repurposing. 2019:219–37.
40. Srisongkram T, Weerapreeyakul N. Drug Repurposing against KRAS Mutant G12C: A Machine Learning, Molecular Docking, and Molecular Dynamics Study. Int J Mol Sci. 2022;24(1):669. https://doi.org/10.3390/ijms24010669 PMID: 36614109
41. Huang R, Xia M, Nguyen D-T, Zhao T, Sakamuru S, Zhao J, et al. Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. Front Environ Sci. 2016;3. https://doi.org/10.3389/fenvs.2015.00085
42. Hsieh JH, Smith-Roe SL, Huang R, Sedykh A, Shockley KR, Auerbach SS, et al. Identifying compounds with genotoxicity potential using Tox21 high-throughput screening assays. Chem Res Toxicol. 2019;32(7):1384–401. https://doi.org/10.1021/acs.chemrestox.9b00053 PMID: 31243984
43. Huang R. Predictive Modeling of Tox21 Data. Advances in Computational Toxicology: Methodologies and Applications in Regulatory Science. 2019:279–97.
44. Idakwo G, Thangapandian S, Luttrell J 4th, Zhou Z, Zhang C, Gong P. Deep learning-based structure-activity relationship modeling for multi-category toxicity classification: a case study of 10K Tox21 Chemicals With High-throughput cell-based androgen receptor bioassay data. Front Physiol. 2019;10:1044. https://doi.org/10.3389/fphys.2019.01044 PMID: 31456700
45. Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. The Tox21 robotic platform for the assessment of environmental chemicals--from vision to reality. Drug Discov Today. 2013;18(15–16):716–23. https://doi.org/10.1016/j.drudis.2013.05.015 PMID: 23732176
46. Tice RR, Austin CP, Kavlock RJ, Bucher JR. Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect. 2013;121(7):756–65. https://doi.org/10.1289/ehp.1205784 PMID: 23603828
47. Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, et al. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chem Res Toxicol. 2016;29(8):1225–51. https://doi.org/10.1021/acs.chemrestox.6b00135 PMID: 27367298
48. Richard AM, Huang R, Waidyanatha S, Shinn P, Collins BJ, Thillainadarajah I, et al. The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chem Res Toxicol. 2021;34(2):189–216. https://doi.org/10.1021/acs.chemrestox.0c00264 PMID: 33140634
49. Schmidt CW. TOX 21: new dimensions of toxicity testing. National Institute of Environmental Health Sciences; 2009.
50. Moukheiber L, Mangione W, Moukheiber M, Maleki S, Falls Z, Gao M, et al. Identifying Protein Features and Pathways Responsible for Toxicity Using Machine Learning and Tox21: Implications for Predictive Toxicology. Molecules. 2022;27(9):3021. https://doi.org/10.3390/molecules27093021 PMID: 35566372
51. Liu F, Patt A, Chen C, Huang R, Xu Y, Mathé EA, et al. Exploring NCATS in-house biomedical data for evidence-based drug repurposing. PLoS One. 2024;19(1):e0289518. https://doi.org/10.1371/journal.pone.0289518 PMID: 38271343
52. Richard AM, Huang R, Waidyanatha S, Shinn P, Collins BJ, Thillainadarajah I, et al. The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chem Res Toxicol. 2021;34(2):189–216. https://doi.org/10.1021/acs.chemrestox.0c00264 PMID: 33140634
53. researchers H. Pharos Website [Available from: https://pharos.habitablefuture.org/.
54. Gould S. Drug Repurposing Hub [Available from: https://www.broadinstitute.org/drug-repurposing-hub.
55. Fecho K, Thessen AE, Baranzini SE, Bizon C, Hadlock JJ, Huang S, et al. Progress toward a universal biomedical data translator. Clin Transl Sci. 2022;15(8):1838–47. https://doi.org/10.1111/cts.13301 PMID: 35611543
56. Wildenhain J, Spitzer M, Dolma S, Jarvik N, White R, Roy M, et al. Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. Cell Syst. 2015;1(6):383–95. https://doi.org/10.1016/j.cels.2015.12.003 PMID: 27136353
57. Kumar SA, Ananda Kumar TD, Beeraka NM, Pujar GV, Singh M, Narayana Akshatha HS, et al. Machine learning and deep learning in data-driven decision making of drug discovery and challenges in high-quality data acquisition in the pharmaceutical industry. Future Medicinal Chemistry. 2022;14(4):245–70.
58. Park K, Ko Y-J, Durai P, Pan C-H. Machine learning-based chemical binding similarity using evolutionary relationships of target genes. Nucleic Acids Res. 2019;47(20):e128. https://doi.org/10.1093/nar/gkz743 PMID: 31504818
59. (NCATS) NCfATS. Generalized glucocorticoid resistance syndrome from Genetic and Rare Diseases (GARD) Information Center 2002. Available from: https://rarediseases.info.nih.gov/diseases/2499/generalized-glucocorticoid-resistance-syndrome%20 https
60. Mayer M, Rosen F. Interaction of anabolic steroids with glucocorticoid receptor sites in rat muscle cytosol. Am J Physiol. 1975;229(5):1381–6. https://doi.org/10.1152/ajplegacy.1975.229.5.1381 PMID: 173192
61. Porcu G, Serone E, De Nardis V, Di Giandomenico D, Lucisano G, Scardapane M, et al. Clobetasol and halcinonide act as smoothened agonists to promote myelin gene expression and RxRγ receptor activation. PLOS ONE. 2015;10(12):e0144550.
62. Schultebraucks K, Wingenfeld K, Otte C, Quinkler M. The role of fludrocortisone in cognition and mood in patients with primary adrenal insufficiency (Addison's Disease). Neuroendocrinology. 2016;103(3–4):315–20. https://doi.org/10.1159/000438791 PMID: 26227663
63. Feng J, Zheng J, Bennett W, Heston L, Jones I, Craddock N, et al. Five missense variants in the amino-terminal domain of the glucocorticoid receptor: no association with puerperal psychosis or schizophrenia. Am J Med Genet. 2000;96(3):412–7.
64. Velayos T, Grau G, Rica I, Pérez-Nanclares G, Gaztambide S. Glucocorticoid resistance syndrome caused by two novel mutations in the NR3C1 gene. Endocrinol Nutr. 2016;63(7):369–71.

65. Al Argan R, Saskin A, Yang JW, D'Agostino MD, Rivera J. Glucocorticoid resistance syndrome caused by a novel NR3C1 point mutation. Endocr J. 2018;65(11):1139–46. https://doi.org/10.1507/endocrj.EJ18-0135 PMID: 30158362

66. Tatsi C, Xekouki P, Nioti O, Bachrach B, Belyavskaya E, Lyssikatos C, et al. A novel mutation in the glucocorticoid receptor gene as a cause of severe glucocorticoid resistance complicated by hypertensive encephalopathy. J Hypertens. 2019;37(7):1475–81. https://doi.org/10.1097/HJH.0000000000002048 PMID: 31145715

67. Smith SJ, MacGregor GA, Markandu ND, Bayliss J, Banks RA, Prentice MG, et al. Evidence that patients with Addison's disease are undertreated with fludrocortisone. Lancet. 1984;1(8367):11–4. https://doi.org/10.1016/s0140-6736(84)90181-8 PMID: 6140341

68. Strickland AL. Long-term results of treatment with low-dose fluoxymesterone in constitutional delay of growth and puberty and in genetic short stature. Pediatrics. 1993;91(4):716–20. https://doi.org/10.1542/peds.91.4.716 PMID: 8464656

69. Shoughy SS, Tabbara KF. Topical tacrolimus solution in autoimmune polyglandular syndrome-1-associated keratitis. Br J Ophthalmol. 2017;101(9):1230–3. https://doi.org/10.1136/bjophthalmol-2016-309808 PMID: 28137823

70. Vingerhoeds AC, Thijssen JH, Schwarz F. Spontaneous hypercortisolism without Cushing's syndrome. J Clin Endocrinol Metab. 1976;43(5):1128–33. https://doi.org/10.1210/jcem-43-5-1128 PMID: 186477

71. Chrousos GP, Vingerhoeds A, Brandon D, Eil C, Pugeat M, DeVroede M, et al. Primary cortisol resistance in man. A glucocorticoid receptor-mediated disease. J Clin Invest. 1982;69(6):1261–9. https://doi.org/10.1172/jci110565 PMID: 6282933

72. Brönnegård M, Werner S, Gustafsson JA. Primary cortisol resistance associated with a thermolabile glucocorticoid receptor in a patient with fatigue as the only symptom. J Clin Invest. 1986;78(5):1270–8. https://doi.org/10.1172/JCI112711 PMID: 3771797

73. Charmandari E, Kino T, Ichijo T, Chrousos GP. Generalized glucocorticoid resistance: clinical aspects, molecular mechanisms, and implications of a rare genetic disorder. J Clin Endocrinol Metab. 2008;93(5):1563–72. https://doi.org/10.1210/jc.2008-0040 PMID: 18319312

74. Hochhaus G, Moellmann HW. Binding affinities of rimexolone (ORG 6216), flunisolide and their putative metabolites for the glucocorticoid receptor of human synovial tissue. Agents Actions. 1990;30(3–4):377–80. https://doi.org/10.1007/BF01966302 PMID: 2386110

75. Aguilera CN, Rodríguez SG. Glucocorticoid resistance syndrome caused by two novel mutations in the NR3C1 gene.

76. Gold PW, Drevets WC, Charney DS. New insights into the role of cortisol and the glucocorticoid receptor in severe depression. Biol Psychiatry. 2002;52(5):381–5. https://doi.org/10.1016/s0006-3223(02)01480-4 PMID: 12242053

77. Chrousos GP, Renquist D, Brandon D, Eil C, Pugeat M, Vigersky R, et al. Glucocorticoid hormone resistance during primate evolution: receptor-mediated mechanisms. Proc Natl Acad Sci U S A. 1982;79(6):2036–40. https://doi.org/10.1073/pnas.79.6.2036 PMID: 6952251

78. Vitellius G, Trabado S, Bouligand J, Delemer B, Lombès M. Pathophysiology of Glucocorticoid Signaling. Ann Endocrinol (Paris). 2018;79(3):98–106. https://doi.org/10.1016/j.ando.2018.03.001 PMID: 29685454

79. Vial JH, Yong AC, Boyd GW. Onset and offset of structural arteriolar changes in DOCA/salt hypertension in the rat. Clin Exp Pharmacol Physiol. 1982;9(3):309–13. https://doi.org/10.1111/j.1440-1681.1982.tb00812.x PMID: 7140010

80. Wang Q, Chen Z, Fan XP, Xu DH, Zhou GH, Liang ZX. A simplified method for preparation of DOCA-salt hypertension model in rats by subcutaneous implantation of DOCA silastic tube. Sheng Li Xue Bao. 1994;46(2):205–8. PMID: 7973805

81. Jin X, Kim WB, Kim M-N, Jung WW, Kang HK, Hong E-H, et al. Oestrogen inhibits salt-dependent hypertension by suppressing GABAergic excitation in magnocellular AVP neurons. Cardiovasc Res. 2021;117(10):2263–74. https://doi.org/10.1093/cvr/cvaa271 PMID: 32960965

82. University JH. Short QT Syndrome from OMIM website 1966 [updated 05/28/2024. Available from: https://www.omim.org/entry/609620?search=%22short%20qt%20syndrome%22&highlight=%22short%20qt%20%28syndromic%7Csyndrome%29%22

83. Farrelly A, Ro S, Callaghan B, Khoyi M, Fleming N, Horowitz B, et al. Expression and function of KCNH2 (HERG) in the human jejunum. American Journal of Physiology-Gastrointestinal and Liver Physiology. 2003;284(6):G883-G95.

84. Pelleg A, Mitamura H, Price R, Kaplinsky E, Menduke H, Dreifus LS, et al. Extracellular potassium ion dynamics and ventricular arrhythmias in the canine heart. J Am Coll Cardiol. 1989;13(4):941–50. https://doi.org/10.1016/0735-1097(89)90240-4 PMID: 2926046

85. Schmitt N, Grunnet M, Olesen S-P. Cardiac potassium channel subtypes: new roles in repolarization and arrhythmia. Physiol Rev. 2014;94(2):609–53. https://doi.org/10.1152/physrev.00022.2013 PMID: 24692356

86. Slavov S, Stoyanova-Slavova I, Li S, Zhao J, Huang R, Xia M, et al. Why are most phospholipidosis inducers also hERG blockers?. Arch Toxicol. 2017;91(12):3885–95. https://doi.org/10.1007/s00204-017-1995-9 PMID: 28551711

87. Itoh H, Horie M, Ito M, Imoto K. Arrhythmogenesis in the short-QT syndrome associated with combined HERG channel gating defects: a simulation study. Circ J. 2006;70(4):502–8. https://doi.org/10.1253/circj.70.502 PMID: 16565572

88. Anderson JL, Krause-Steinrauf H, Goldman S, Clemson BS, Domanski MJ, Hager WD, et al. Failure of benefit and early hazard of bucindolol for Class IV heart failure. J Card Fail. 2003;9(4):266–77. https://doi.org/10.1054/jcaf.2003.42 PMID: 13680547

89. Smart NA, Kwok N, Holland DJ, Jayasighe R, Giallauria F. Bucindolol: a pharmacogenomic perspective on its use in chronic heart failure. Clin Med Insights Cardiol. 2011;5:55–66. https://doi.org/10.4137/CMC.S4309 PMID: 21792345

90. University JH. Online Mendelian Inheritance in Man 1966 [updated 05/28/2024]. Available from: https://www.omim.org/.

91. Sciences NCfAT. Linkage between metoclopramide and FPLD on ARAX translator. Available from: https://arax.ncats.io/?r=241114