HHS Public Access

Author manuscript

Pain. Author manuscript; available in PMC 2024 April 16.

Published in final edited form as:

Pain. 2024 April 01; 165(4): 908–921. doi:10.1097/j.pain.0000000000003089.

Machine learning study of the extended drug–target interaction network informed by pain related voltage-gated sodium channels

Long Chen[a], **Jian Jiang**[a,b], **Bozheng Dou**[a], **Hongsong Feng**[b], **Jie Liu**[a], **Yueying Zhu**[a], **Bengong Zhang**[a], **Tianshou Zhou**[c], **Guo-Wei Wei**[b,d,e,*]

[a]Research Center of Nonlinear Science, School of Mathematical and Physical Sciences, Wuhan Textile University, Wuhan, P R. China

[b]Department of Mathematics, Michigan State University, East Lansing, MI, United States

[c]Key Laboratory of Computational Mathematics, Guangdong Province, and School of Mathematics, Sun Yat-sen University, Guangzhou, P R. China

[d]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, United States

[e]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, United States

Abstract

Pain is a significant global health issue, and the current treatment options for pain management have limitations in terms of effectiveness, side effects, and potential for addiction. There is a pressing need for improved pain treatments and the development of new drugs. Voltage-gated sodium channels, particularly Nav1.3, Nav1.7, Nav1.8, and Nav1.9, play a crucial role in neuronal excitability and are predominantly expressed in the peripheral nervous system. Targeting these channels may provide a means to treat pain while minimizing central and cardiac adverse effects. In this study, we construct protein–protein interaction (PPI) networks based on pain-related sodium channels and develop a corresponding drug–target interaction network to identify potential lead compounds for pain management. To ensure reliable machine learning predictions, we carefully select 111 inhibitor data sets from a pool of more than 1000 targets in the PPI network. We employ 3 distinct machine learning algorithms combined with advanced natural language processing (NLP)–based embeddings, specifically pretrained transformer and autoencoder representations. Through a systematic screening process, we evaluate the side effects and repurposing potential of more than 150,000 drug candidates targeting Nav1.7 and Nav1.8 sodium channels. In addition, we assess the ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties of these candidates to identify leads with near-optimal

*Corresponding author. Address: Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States. weig@msu.edu (G. W. Wei).

Conflict of interest statement

The authors have no conflicts of interest to declare.

Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/B940.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

characteristics. Our strategy provides an innovative platform for the pharmacological development of pain treatments, offering the potential for improved efficacy and reduced side effects.

Keywords

Pain management; Voltage-gated sodium channels; Protein–protein interaction; Drug–target interaction; Machine learning; Virtual drug screen; Repurposing; ADMET

1. Introduction

Pain is a complex phenomenon with various categories, including acute and chronic pain, nociceptive and neuropathic pain, among others. It affects approximately 35% of the US population, surpassing the morbidity rates of cancer and heart disease.[49] There is an urgent demand for new pain medications.

Voltage-gated sodium channels (Nav channels or VGSCs) are vital membrane proteins essential for generating and transmitting action potentials in neurons and excitable cells. They facilitate the rapid entry of sodium ions, leading to cell depolarization and the initiation of action potentials. These channels regulate sodium ion permeability and contribute to various intercellular functions linked to diseases such as chronic pain and cardiac arrhythmia. Notably, specific Nav channel subtypes (Nav1.3, Nav1.7, Nav1.8, and Nav1.9) encoded by genes *SCN3A*, *SCN9A*, *SCN10A*, and *SCN11A*, respectively, are highly expressed in the peripheral nervous system, sympathetic ganglia, olfactory epithelium, and dorsal root ganglion sensory neurons, making them promising targets for pain therapeutics.[13] Much attention has been paid to Nav1.3,[44,46] Nav1.7,[4,12,18] Nav1.8,[5,48] and Nav1.9,[20,21,34,45] in their connection to pain management.[42] However, the specific roles of these pain-related Nav channels in generating and transmitting pain signals remain unclear.

Protein–protein interactions (PPIs) are crucial for various biological processes, including DNA replication, signaling, and metabolism. The String Database v11 (https://string-db.org/) provides a comprehensive collection of protein–protein interactions and can be used to construct PPI networks. By focusing on major sodium channels involved in pain (Nav1.3, Nav1.7, Nav1.8, and Nav1.9), medication treatments and side effects can be analyzed. However, traditional testing methods are time consuming and resource intensive. To address this, artificial intelligence (AI), including machine learning (ML) techniques, can be employed for large-scale predictions in this area.[2]

Recently, many advanced ML methods have been applied to pain treatment and analysis.[36,41,47,53] Currently, numerous in silico methods have been developed for virtual screening of sodium channel inhibitors.[1,2,6,17,22,29,30,35,57] From 2018, more studies on VGSCs can be found in review articles.[24,37,40,43] However, these studies lack consideration of drug–target interaction networks, as well as comprehensive ADMET (absorption, distribution, metabolism, excretion, and toxicity) analysis.

Pain management is not limited to sodium channels and related inhibitors. Opioids, also known as narcotics, have been used for centuries in the treatment of pain. To deal with opioid use disorder (OUD), advanced ML predictors were used to screen and repurpose thousands of DrugBank compounds and evaluate their ADMET properties.[16] More sophisticated AI models were also developed for drug addiction.[58]

In this study, we construct an extended drug–target interaction (DTI) network induced by pain-related sodium channels, which are analyzed by advanced ML models using natural language processing (NLP) tools. We build PPI networks with more than 1000 targets using the String Database v11 and an associated DTI network with 111 targets and more than 150,000 compounds from the Chembl database (https://www.ebi.ac.uk/chembl/). We employ transformer and autoencoder to develop 111 ML models for the screening and repurposing of these compounds and FDA-approved drugs and existing medications. Furthermore, we study ADMET and synthesizability to identify lead compounds as shown in Figure 1. This investigation of the extended DTI network offers an innovative approach to pain therapeutic development.

2. Methods

2.1. Data sets

All inhibitor data sets were collected from the Chembl database for all proteins in the present DTI network, which was informed by 4 investigated sodium channels or treatment targets (Nav1.3, Nav1.7, Nav1.8, and Nav1.9, corresponding to encoded genes *SCN3A*, *SCN9A*, *SCN10A*, and *SCN11A*, respectively). Because the predictive results of machine learning–based models depend on high quality and quantity of data, we set the minimal size of the collected inhibitor data sets to be 250 samples and obtained a total of 111 data sets, including SCN9A and SCN10A. The data sets for SCN3A and SCN11A were not included due to their small data size. The labels for these data sets are binding affinities (BAs) obtained using the following formulas: $BA = 1.3633 \times \log_{10} K_i$[9] and $K_i = IC_{50}/2$.[28] K_i refers to inhibition constant and also represents the dissociation constant describing the binding affinity between the inhibitor and the enzyme, whereas IC_{50} stands for inhibitory concentration 50%, that is, the concentration of inhibitor required to reduce the biological activity of interest to half of the uninhibited value. The drug–target binding affinity is indicated by the dissociation constant $K_d = [L][P]/[LP]$, where $[L]$, $[P]$, and $[LP]$ are the molar concentration of drug, target, and drug–target complex, respectively. Particularly, the Gibbs free energy (kcal/mol) can be derived by $\Delta G = RT\ln K_d$, where R and T are the gas constant and temperature, respectively. $\Delta G = -1.3633 pK_d$ can be obtained with room temperature $T = 298.15K$.[9] Here, pK_d represents $-\log_{10} K_d$ with K_d in the unit of mol. Following the way that PDBbind database mixes K_d and K_i in their refined data sets,[54] in present work, we calculate the binding energy with the above BA calculation formula. In addition, because hERG is a key target for side effects in virtual screening of drug design, an inhibitor data set was also collected from the Chembl database. All details of the data sets are provided in Table S3 in the Supporting Information (available at http://links.lww.com/PAIN/B940).

2.2. Molecular fingerprints

Molecular fingerprints represent the property profiles of a molecule, typically in the form of vectors where each element represents the presence, degree, or frequency of a specific structural characteristic. These fingerprints can be used as features in machine learning (ML) models. The original molecular fingerprints for the inhibitors in the collected 111 data sets are 2D simplified molecular-input line-entry system (SMILES) strings. In this study, we used 2 types of latent-vector molecular fingerprints in the ML models: bidirectional encoder transformer fingerprint (BET-FP) and autoencoder fingerprint (AE-FP). These fingerprints were generated from pretrained models based on natural language processing (NLP) algorithms such as transformers and sequence-to-sequence autoencoders.[10,55] They are latent embedding vectors with a length of 512, obtained by encoding the 2D SMILES strings of the inhibitor compounds using the pretrained models.

2.2.1. Sequence-to-sequence autoencoder fingerprint—Recently, Winter et al.[55] proposed a data-driven unsupervised learning model for extracting molecular information embedded in the SMILES representation. Their approach involved using a sequence-to-sequence autoencoder to translate one form of molecular embedding to another by capturing the chemical structure's complete description in the latent space between the encoder and decoder. This translation model was capable of extracting physical and chemical information during the embedding process, enabling the translation to a distinct molecular representation with the same semantics but different syntax. Notably, the translation model was trained on a large data set of chemical structures and could be used to extract molecular fingerprints for query compounds without the need for retraining or labels.

Typically, the translation model consists of encoder and decoder networks. The encoder network compresses the essential information from the input SMILES, which is then fed as input to the decoder network. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) were employed in the decoder, with fully connected layers mapping the output of the CNN or concatenated cell states of the RNN to intermediate vector embeddings between the encoder and decoder networks. Consequently, the decoder incorporates RNN networks with latent vectors as input. To extract more physical and chemical information from the latent vectors, the translation model was extended based on a classification model that predicts molecular properties using these vectors. The output of the RNN in the decoder network represents the probability distributions of various characters in the translated molecular embeddings. During the training of the autoencoder model, the loss function consists of the sum of cross-entropies between the predicted probability distributions and the correct characters encoded in a one-hot format, as well as the mean-squared errors of the molecular property predictions made by the classification model.

In this study, the translation model was trained on approximately 72 million molecular compounds obtained from the ZINC (https://zinc15.docking.org/) and PubChem databases (https://pubchem.ncbi.nlm.nih.gov/). The compounds underwent preprocessing, including filtering based on criteria such as molecular weight, number of heavy atoms, partition coefficient, and more. By training the translation model on this processed data set, the resulting model generated embedding vectors that served as molecular fingerprints.

2.2.2. Bidirectional transformer—Recently, Chen et al.[10] developed a deep learning network that was pretrained on millions of unlabeled molecules using a self-supervised learning (SSL) platform to extract predictive molecular fingerprints. The SSL approach employed the bidirectional encoder transformer (BET) model, which relies on the attention mechanism. Unlike constructing a complete encoder–decoder framework, SSL used the decoder network solely for encoding the molecular SMILES.

In the SSL pretraining platform, the input consisted of molecular SMILES strings. Pairs of real SMILES and masked SMILES were created by hiding a certain number of meaningful symbols within the strings. The model was then trained using these data-mask pairs in a supervised manner with the SSL method. During the pretraining process, the masked symbols were learned by studying the unprocessed symbols in the SMILES, enhancing the understanding of the SMILES language. Data masking was performed as a preprocessing step before training the model with SSL. A total of 51 symbols were considered as elements in the SMILES strings. The SMILES were used as input to train the model, with a maximum length set to 256. Two special symbols, "<s>" and "</s>," were added to the beginning and end of the SMILES strings. If a string's length was less than 256, the "<pad>" symbol was used to complete the SMILES string. For the data masking process, 15% of the symbols in the SMILES were manipulated, with 80% being masked, 10% remaining unchanged, and the remaining 10% randomly changed.

The BET module plays a crucial role in achieving SSL from a substantial number of SMILES strings. It uses the attention mechanism in the transformer module to extract the importance of each symbol in the SMILES sequence. The BET module consists of 8 bidirectional encoder layers, where each layer includes a multihead self-attention layer and a subsequent fully connected feed-forward neural network. Each self-attention layer has 8 heads, and the embedding size of the fully connected feed-forward layers is 1024. During training, the Adam optimizer with a weight decay of 0.1 is employed, and the loss function chosen is cross-entropy. The input SMILES have a maximum length of 256, including the special symbols added at the 2 ends, and each symbol is embedded in a dimension of 512. Consequently, the resulting molecular embedding matrix consists of 256 embedding vectors, each with a dimension of 512.

The transformer module offers high parallelism capability and training efficiency, allowing for the use of a large amount of SMILES to train deep learning models. In this study, SMILES strings from the Chembl, PubChem, and ZINC databases, either individually or fused together, were used to train 3 separate pretrained models. The resulting transformer-based molecular embeddings generated from the pretrained models using the Chembl database were used as molecular fingerprints.

2.3. Machine learning models

Three classic machine learning algorithms, namely, gradient boosting decision tree (GBDT), support vector machine (SVM), and random forest (RF), are employed to construct our ML models. The GBDT algorithm, an ensemble approach, possesses several advantages such as resistance to overfitting, insensitivity to hyperparameters, and ease of implementation. Consequently, it is competitive when training with small data sets and can yield better

prediction performance compared with deep neural networks (DNNs) and other common ML algorithms. However, it is important to note that one of the challenges of GBDT is to strike a balance between accuracy and efficiency for large data sets. The algorithm assembles multiple weak learners (individual trees) into an iterative prediction model. Although weak learners may produce suboptimal predictions individually, the combination of all weak learners through the ensemble approach helps reduce overall errors. The primary procedure of GBDT involves learning decision trees, where most of the time is consumed in finding the best split points. Gradient boosting decision tree has already demonstrated good performance in various quantitative structure–activity relationship (QSAR) prediction tasks.[25,26] In this study, the GBDT algorithm provided by the Scikit-learn library (version 0.24.1) was used.

Support Vector Machine, introduced by Cortes and Vapnik, is a nonprobabilistic kernel-based supervised learning method that maps input vectors into a high-dimensional feature space.[11] The core concept behind SVM is to identify the optimal decision boundary that separates different classes in the feature space. This decision boundary is defined by a hyperplane that maximizes the margin between the support vectors and the data points closest to the decision boundary. Support vector machine offers advantages such as high efficiency in high-dimensional spaces, robustness against overfitting, and versatility. However, SVM also has some limitations, including computational complexity and sensitivity to parameter tuning.

Random forest, developed by Breiman, is an ensemble of decision trees where the predictions of individual trees are averaged to obtain an ensemble performance.[7] It employs a bootstrap sampling technique, and each decision tree uses only a subset of randomly chosen samples and features, starting with a trunk that splits into multiple branches before reaching the leaves. The leaf nodes represent the final prediction, whereas all other nodes are assigned with molecular features. Random forest is widely used in solving QSAR prediction problems and often does not require a complex feature selection procedure. Moreover, it is robust to redundant features and exhibits insensitivity to parameter variations.

We collected a total of 111 inhibitor data sets in our DTI network. The 3 aforementioned ML algorithms were used to build ML models for these data sets. The details of the hyperparameters for these 3 ML algorithms are provided in Table S5 in the Supporting Information (available at http://links.lww.com/PAIN/B940). In the ML models, we used 2 types of molecular fingerprints, namely, BET and AE fingerprints, to embed the inhibitor compounds. Our ML models were created by pairing these molecular fingerprints with the GBDT, SVM, or RF algorithm. Consequently, we built a total of 111 ML models, each corresponding to one inhibitor data set.

For each data set, 6 individual models were constructed by combining BET and AE fingerprints with the 3 ML algorithms. The average of the predictions from these 6 individual models was considered as our final binding affinity prediction, which we refer to as the consensus method for prediction. The consensus results typically outperform those obtained from individual models. We compared the prediction results using the 3 different algorithms and found that the SVM algorithm with the consensus method performed the

best among the other algorithms using individual fingerprints. This was validated using a set of provided samples, as shown in Table S6 in the Supporting Information (available at http://links.lww.com/PAIN/B940). Hence, the prediction results in the main text are from the SVM algorithm with the consensus method. To reduce the impact of randomness, each individual ML model was trained 10 times using different random seeds, and the average of the 10 predictions was considered as the final result for each individual model. In addition, the Pearson correlation coefficients (R) and root-mean-square deviation (RMSD) of 10-fold cross-validations for the 111 data sets are presented in Table S7 of the Supporting Information (available at http://links.lww.com/PAIN/B940).

3. Results

3.1. Pain-related voltage-gated sodium channel informed drug–target interaction networks

Voltage-gated sodium channels, which consist of a family of 9 distinct proteins or genes (Nav1.1–1.9), exhibit different pharmacological properties. Specifically, the proteins Nav1.3, Nav1.7, Nav1.8, and Nav1.9 are involved in neuropathic pain and are associated with both human Mendelian pain disorders and common pain disorders such as small fiber neuropathy.[3] These 4 VGSC proteins play a role in modulating different types of pain, offering potential for the development of specific sodium channel inhibiting agents for chronic pain treatment. Functionally, Nav1.7 is classified as tetrodotoxin sensitive (TTX-S), whereas Nav1.8 and Nav1.9 are considered tetrodotoxin resistant (TTX-R). Anatomically, these proteins exhibit broad and distinct expression patterns across neuronal and smooth muscle cells throughout the body, as well as in cells of the immune system where they participate in migration and phagocytosis.[14] Traditionally, Nav1.3 is primarily expressed in the brain and spinal cord, whereas Nav1.7, Nav1.8, and Nav1.9 tend to be expressed in the peripheral nervous system. Furthermore, these channels are regulated by a variety of enzymes and structural proteins, such as kinases, auxiliary β-subunits, and ubiquitin-protein ligases, which collectively influence sodium channel biophysical properties and expression.[32,52]

Pain-related VGSCs are widely distributed throughout the body, and their interactions with various upstream and downstream proteins play a crucial role in specific biological functions. To analyze these interactions, we constructed protein–protein interaction (PPI) networks centered around each of the 4 pain-related VGSCs or treatment targets, namely, SCN3A, SCN9A, SCN10A, and SCN11A. These 4 targets were used as inputs to the String database to extract the corresponding PPI networks. The resulting networks, shown in Figure 1A, represent direct and indirect interactions between proteins and each pain-related VGSC. Each PPI network contains 401 proteins, focusing on critical interactions rather than considering a larger number of proteins. It is important to note that there is some overlap between the networks, indicating interdependencies among the VGSCs, and there are 1032 unduplicated proteins in 4 PPI networks. In addition, blocking these proteins may result in other severe off-target effects. Hence, these proteins could be critical sources of side effects, and thus, 4 proteomic PPI networks provide a pool of 4 potential treatment targets and critical side effect targets. It is necessary to systematically explore potential compounds that

inhibit distinct pain targets and the putative side effects from compounds blocking these targets.

Considering that compounds that act as agonists or antagonists on pain-related VGSCs can influence their pharmacological behavior in pain treatment, we aimed to identify additional compounds that bind to these VGSCs. To evaluate the binding effects of inhibitors on VGSCs and other proteins in the PPI networks, we searched and collected inhibitor compounds from the Chembl database for each protein. This process resulted in an extended DTI network, encompassing 111 targets or related data sets and a total of 150,147 inhibitor compounds, which is illustrated in Figure 1B. The protein names of these 111 data sets are listed in Table S2 in the Supporting Information, and additional details about the collected data sets can be found in Table S3 in the Supporting Information (available at http://links.lww.com/PAIN/B940).

The framework of present work is illustrated in Figure 1. Essentially, for 4 pain-related VGSCs, ie, 4 treatment targets, namely, SCN3A, SCN9A, SCN10A, and SCN11A, we take a proteome-informed approach through protein–protein interaction (PPI) networks to identify potential side effect targets. As such, we found more than 1000 targets within the 4 PPI networks centered around SCN3A, SCN9A, SCN10A, and SCN11A, as shown in Figure 1A. We hope to set up machine learning (ML) models for all these targets, in principle. In practice, when we checked databases, we could only build 111 models because of insufficient inhibitors in these DTI networks, as displayed in Figure 1B. Among the 111 models, 2 models are designated for the treatment targets SCN9A and SCN10A, and the remaining 109 models are allocated to side effect targets (one of the side effect targets is hERG). We found a total of 150,147 inhibitors for these 111 targets. As shown in Figure 1C, all inhibitors associated with side effect targets are screened for their repurposing potential using the ML models associated with the treatment targets SCN9A and SCN10A. Moreover, all SCN9A and SCN10A inhibitors, including the repurposed ones, are screened for their potential side effects with respect to the 109 side effect ML models and ADMET models, leading to nearly optimal leads for the treatment targets.

3.2. Binding affinity predictions for the extended drug–target interaction network

Using autoencoder and transformer embeddings, we developed 111 ML models for all 111 targets and 150,147 compounds in the extended DTI network. The cross-target binding affinity (BA) predictions were carried out using these 111 ML models, and the results are presented in Figure 2. The diagonal elements of the heatmap represent the Pearson correlation coefficient (R) obtained from 10-fold cross-validation for each ML model. The mean, maximum, and minimum values of R across the models are 0.77, 0.93, and 0.25, respectively. Notably, 53 models achieved R values greater than 0.8, indicating high predictive performance.

Furthermore, the root-mean-square deviation (RMSD) values of these models, as shown in Table S3 in the Supporting Information (available at http://links.lww.com/PAIN/B940), range from 0.43 to 1.15 kcal/mol. These values fall within a reasonable range, suggesting that the ML models exhibit excellent prediction accuracy and reliable performance for binding affinity predictions.

3.2.1. Cross-target binding affinity predictions for the extended drug–target interaction network—In this section, we conduct an analysis of compound cross-target interactions to estimate their side effects on other proteins in the protein–protein interaction (PPI) network, providing a better understanding of the extended DTI network. The off-diagonal elements of the heatmap in Figure 2 represent the maximum binding affinity (BA) values (ie, BA with the largest absolute values) of inhibitor compounds from one data set predicted by other ML models. The labels on the left side of the heatmap correspond to the 111 inhibitor data sets, whereas the labels on the top of the heatmap correspond to all the 111 ML models. Each column in the heatmap represents the predictions made by a specific model.

For instance, the *i*-th element in the *j*-th column indicates the prediction result of the *i*-th data set by the *j*-th model. These cross-target prediction results serve as indicators of the potential side effects of one inhibitor data set on other proteins. In our analysis, we use an inhibition threshold value of −9.54 kcal/mol (K_i = 0.1 μM) for the BA values.[19] If a compound has a BA value below this threshold, it is considered active in terms of its biological function. Otherwise, it is classified as an inactive compound.

According to our analysis, of the 12,210 cross-predictions, 9262 were found to exhibit side effects based on this threshold value because their predicted maximal BA values were below −9.54 kcal/mol. In addition, the remaining 2948 cross-prediction results showed weak side effects because their maximal BA values exceeded −9.54 kcal/mol. The color of the off-diagonal elements in the heatmap indicates the strength of the side effects, with closer proximity to green representing stronger side effects and closer proximity to yellow indicating weaker side effects.

It is worth noting that in Figure 2, several yellow vertical lines can be observed, suggesting very slight predicted side effects on these proteins. This could be because the majority of collected experimental BA labels being larger than −9.54 kcal/mol, which limits the predictive power of the ML models in such cases.

The reasons for side effects caused by drug candidates targeting a specific protein are often complex, and one possible factor is the presence of similar binding sites on off-target proteins. Proteins within the same family often share similar structures or sequences, leading to the existence of comparable binding sites. As a result, an inhibitor compound that is effective against one protein may also bind to another protein within the same family, giving rise to mutual side effects.

As observed in Figure 2, mutual side effects occur among the 3 targets CAMK2A, CAMK2B, and CAMK2D, which belong to the calmodulin-dependent protein kinase II (CAMK2) family and share similar 3D structural conformations or 2D sequences. This observation is further supported by the alignments of their 3D structures and 2D sequences, as shown in Fig. S1 of the Supporting Information (available at http://links.lww.com/PAIN/B940).

We can identify more examples of mutual side effects among proteins within the same family. For instance, the fibroblast growth factor target (FGFR) family, which includes

FGFR1, FGFR2, FGFR3, and FGFR4, as well as the mitogen-activated protein kinase (MAPK) family, which comprises MAPK3, MAPK8, MAPK9, and MAPK10, exhibit mutual side effects. These examples illustrate the occurrence of mutual side effects among proteins in the same family, emphasizing the importance of considering family-wide effects in drug development and analysis.

3.2.2. Predictions of side effects and repurposing potentials for the extended drug–target interaction network—Side effects occur when a drug candidate exhibits strong binding affinity to the intended target but inadvertently affects other proteins as potential off-target inhibitors. These side effects can be identified through cross-target predictions, as illustrated in Figure 3A, for the extended DTI network. Each panel in the figure represents a specific treatment target and 2 corresponding off-target proteins or side effect targets, indicated by the panel title, *x*-axis, and *y*-axis, respectively. The scattered points in the plot are color coded based on the experimental binding affinities (BAs) of the inhibitors for the treatment target. Red and green colors represent high and low binding affinities, respectively. The *x*-axis and *y*-axis values represent the predicted BAs obtained from 2 machine learning (ML) models constructed using inhibitor data sets for the 2 off-target proteins or side effect targets.

The blue frames in the 9 panels of Figure 3A indicate regions where no side effects are predicted on the 2 side effect targets. The 3 rows of the figure represent different scenarios for inhibitors targeting a specific treatment target, showing the presence of side effects on zero, one, or both of the given side effect targets. For instance, in the first panel of the first row, all active inhibitors for treatment target SCN10A are predicted to have weak inhibitory effects, with binding affinity (BA) values greater than −9.54 kcal/mol, on the 2 side effect targets CAMK2A and CACNA1C. In the first panel of the second row, a part of the active inhibitors for treatment target SCN9A is predicted to exhibit strong binding affinity to the hERG protein, whereas none of its active inhibitors are predicted to bind to the side effect target GAPDH. Furthermore, in the second panel of the third row, most active inhibitors of SCN10A are predicted to efficiently bind to both the FLT4 and FGFR1 proteins simultaneously.

The repurposing potential of inhibitors can also be determined through cross-target predictions. Drug candidates that exhibit weak binding affinity to their designated targets but exhibit potent inhibition of other proteins are defined to possess repurposing potential. Figure 3B displays 6 prediction cases of repurposing identified on 2 treatment targets SCN9A and SCN10A by our models. In the yellow frames, the inactive inhibitors for side effect target exhibit strong binding to one treatment target (ie, predicted BAs less than −9.54 kcal/mol) but weak binding to the other treatment target (ie, predicted BAs greater than −9.54 kcal/mol). For example, in the first panel of the first row in Figure 3B, many inactive inhibitors for side effect target P2RX3 are predicted to have repurposing potential for either SCN9A or SCN10A but not for the other one. Because both SCN9A and SCN10A are important treatment targets for drug design in pain treatment, it is crucial to identify more drug candidates for these 2 proteins through the virtual screening process. Carbamazepine, a voltage-dependent Nav1.7 sodium channel (SCN9A) blocker, has undergone a phase I clinical study in humans.[39] Our models can be employed to find more inhibitors that can

bind to SCN9A, similar to the mechanism of carbamazepine. The second and third rows in Figure 3 depict additional cases where inactive inhibitors for a given side effect target have repurposing potential for 2 treatment targets.

3.2.3. Protein similarity inferred by cross-target correlations in the drug–target interaction network—As side effects can arise when a drug candidate binds to proteins with similar 3D structures or sequences, the predicted BA values in cross-target BA prediction may exhibit correlation. In other words, correlated predicted BA values can serve as an indication of similar binding sites or 3D protein structures. Figure 4A illustrates a linear correlation between the predicted BAs of inhibitors for PTGS2 on CHRM1 and CHRM2 proteins, with a Pearson correlation coefficient R of up to 0.71. The high correlation is attributed to the high binding site similarity between CHRM1 and CHRM2 proteins, as validated by the alignments of 3D structures and 2D sequences in Figure 4A. The 3D structures of the 2 proteins were found to be quite similar, and the identity of the 2D binding site sequence reached as high as 63%.

Two additional examples can be observed in Figures 4B and C, demonstrating that the predicted BA correlation indicates similar 3D protein structures. The Pearson correlation coefficients are 0.82 and 0.72 for the cases in Figure 4B, corresponding to the predicted BAs for OPRM1 on CSNK2A2 and CSNK2A1, respectively. These alignments of 3D structures and 2D sequences validate the usefulness of cross-prediction in detecting protein similarity.

Furthermore, Figure 4C reveals a bilinear correlation relationship, where the predicted BAs of MAPK10 inhibitors not only linearly correlate with MAPK8 and MAPK9 proteins but also exhibit a linear correlation with their experimental BA values, as indicated by the color coding. This bilinear relationship is confirmed by the alignment of 3D structures and 2D sequences of the 3 proteins. This result suggests that a potent MAPK10 inhibitor is likely to be a strong binder for both MAPK8 and MAPK9 proteins simultaneously. The high structural similarities result in a drug-mediated trilinear target relationship. The observed bilinear or trilinear relationship indicates the possibility of developing inhibitors that can bind to multiple targets of major pain proteins simultaneously.

3.3. Druggable property screening

Evaluation of ADMET is of utmost importance in drug design and discovery. Absorption, distribution, metabolism, excretion, and toxicity encompasses several essential attributes that correlated with the pharmacokinetic study of a compound. A promising drug candidate should not only exhibit potency against the therapeutic target but also should possess favorable ADMET properties. Furthermore, hERG is a crucial potassium ion channel known for its contribution to the electrical activity of the heart. When this channel is blocked by a drug, it can lead to serious side effects on the heart. Therefore, the evaluation of hERG risk is indispensable in drug development and assessment.

In this section, we conducted the evaluation of ADMET using 6 indexes, namely, FDAMDD, $T_{1/2}$, $F_{20\%}$, logP, logS, and Caco-2, along with synthetic accessibility (SAS) and hERG risk assessment. FDAMDD represents the FDA maximum recommended daily dose, which aims to avoid toxicity in the human body. The half-life ($T_{1/2}$) refers to the

time it takes for the concentration of a drug in the body to decrease by half. A value of $T_{1/2}$ less than 3 hours indicates a shorter half-life. $F_{20\%}$ represents the probability of an administered drug reaching systemic circulation with less than 20% of the initial dose. This parameter is important for assessing the effectiveness, bioavailability, therapeutic efficacy, and potential side effects of a drug. LogP refers to the logarithm of the partition coefficient of a compound between a nonpolar solvent and water, providing information about its hydrophobicity. On the other hand, logS represents the logarithm of the aqueous solubility of a compound, which indicates its ability to dissolve in water. Caco-2 is a measure used to estimate the in vivo permeability of oral drugs. It provides valuable information about a drug candidate's interaction with efflux transporters, metabolism, and other factors that influence its absorption. Synthetic accessibility is employed to assess the feasibility of synthesizing a specific compound or molecule, taking into account its structural complexity and the availability of synthetic routes.

During the above estimation in this work, ADMETlab 2.0 (https://admetmesh.scbdd.com/) solvers were used for ML predictions and provided a set of optimal ranges for these ADMET properties.[56] The SAS assessment was implemented using Rdkit packages.[33] The optimal ranges of ADMET properties and SAS are listed in Table 1, in which a stricter threshold of −8.18 kcal/mol (K_i = 1 μM) is applied to exempt hERG side effects. Figure 5 illustrates the ADMET screening of 5 inhibitor data sets, including 3 important VGSCs, SCN5A, SCN9A, and SCN10A, as well as 2 important proteins CNR1 and steroid receptor coactivator (SRC), that play essential roles in pain treatment. Specifically, CNR1, a cannabinoid receptor, is involved in pain modulation through its influence on neurotransmitter release, anti-inflammatory effects, and potential effects on neuropathic pain. SRC protein, on the other hand, indirectly contributes to pain management by enhancing the transcription of anti-inflammatory genes in response to steroid hormone receptor activation. The first row of Figure 5 depicts the distributions of FDAMDD and hERG side effects of inhibitors from the 5 data sets. The blue frames represent the optimal domains of the 2 properties mentioned above. The colors of the points indicate the experimental BA values for targets. From this screening, all 5 data sets have sufficient compounds with optimal toxicity and hERG side effects. However, for the SCN10A data set, there are only a few potent inhibitors in the optimal domains. This suggests that ADMET properties and side effects should be taken into account before synthesizing a new compound.

The second row of Figure 5 displays the screening results on absorption properties: $T_{1/2}$ (half-life) and $F_{20\%}$ (bioavailability 20%). It is observed that for all 5 data sets, the optimal domain of $T_{1/2}$ and $F_{20\%}$ occupies only a small fraction of chemical space. This indicates a strict screening process, emphasizing the critical roles of these 2 properties in physicochemical assessment.

The third row of Figure 5 illustrates the screening for logP and logS, which are closely related to the distribution of chemicals in the human body. In all 5 data sets, only a small portion of potent inhibitors is found within the optimal domain, suggesting that a large number of inhibitors are not well absorbed in the human body.

The last row of Figure 5 presents the screening results for Caco-2 and SAS. These 5 plots demonstrate that almost all compounds from the 5 data sets are easy to synthesize, and approximately half of the compounds exhibit good cell permeability. Notably, a significant number of potent inhibitors fall within the optimal domain.

3.4. Side effect evaluations of existing medications for pain treatment

SCN3A, *SCN9A*, *SCN10A*, and *SCN11A* are genes that encode sodium channels in the Nav channels family. These channels play an important role in the generation and propagation of action potentials in neurons, including those involved in pain signaling. In addition, it has been found that blocking these channels could reduce pain hypersensitivity. There are several FDA-approved experimental medications available for the treatment of pain, which can be roughly classified into 4 classes: nonopioid analgesics, nonsteroidal anti-inflammatory drugs (NSAIDs), opioid medications, and others. In this study, we used our DTI-based ML models to predict the side effects of these medications.

Acetaminophen, commonly known as Tylenol or paracetamol, is a typical over-the-counter nonopioid analgesic used to temporarily relieve mild to moderate pain, such as headaches, muscular aches, backaches, toothaches, and premenstrual and menstrual cramps. It is a weak inhibitor of both cyclooxygenase (COX)-1 and COX-2 in vitro and eases pain by inhibiting the production of prostaglandins, which are chemicals that contribute to pain in the human body.

Our BA predictions for acetaminophen on SCN9A and SCN10A are −9.60 kcal/mol and −9.29 kcal/mol, respectively, indicating that acetaminophen is a good binder on SCN9A. Furthermore, the predicted BA value on hERG from our model is −7.39 kcal/mol, which is higher than the hERG side effect threshold of −8.18 kcal/mol, validating the safety profile of acetaminophen on hERG. This result agrees with the conclusion of the study by Su et al.[50]

Our predictions suggest that acetaminophen exhibits the highest inhibitory effect on the LATS2 protein, with a predicted BA value of −11.2 kcal/mol. LATS2 is a protein kinase that plays a significant role in cell growth regulation, apoptosis, and tumor suppression. It is associated with various diseases, including breast cancer, lung cancer, ovarian cancer, neurofibromatosis type 2 (NF2), and cardiovascular diseases. Inhibiting the LATS2 protein could lead to serious side effects, which might explain the potential reasons for the high side effects of acetaminophen, such as liver damage, allergic reactions, skin reactions, gastrointestinal issues, blood disorders, and kidney problems.

Nonsteroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen (Advil, Motrin), and naproxen (Aleve), are commonly used for the treatment of mild to moderate pain accompanied by swelling and inflammation. These medications can inhibit certain enzymes in the human body that are released due to tissue damage. Ibuprofen, a nonselective inhibitor of the enzyme COX, plays a crucial role in the synthesis of prostaglandins through the arachidonic acid pathway. Cyclooxygenase facilitates the conversion of arachidonic acid to prostaglandin H2 (PGH2) in the body, which is further transformed into other prostaglandins. By inhibiting COX, ibuprofen reduces the production of prostaglandins in the body, resulting in pain relief.

The predicted BA values of ibuprofen for SCN9A and SCN10A are −9.11 and −9.72 kcal/mol, respectively, indicating strong potency of ibuprofen on SCN10A. The predicted BA value for hERG is −7.13 kcal/mol, suggesting a safe hERG-blockade profile. In addition, ibuprofen is predicted to be a potent inhibitor of LATS2, USP9X, and MTOR, which are the top 3 proteins with the largest absolute predicted BA values (−11.17, −10.68, 210.46 kcal/mol). Furthermore, the predicted BA value of ibuprofen on TRPM8 is −10.04 kcal/mol, validating its strong binding affinity to TRPM8, a thermosensitive ion channel implicated in pain signaling, particularly in cold-induced pain or cold allodynia.

Despite its effectiveness, ibuprofen can cause a number of side effects, including nausea, constipation or diarrhea, and indigestion (dyspepsia).

Naproxen, like other NSAIDs such as ibuprofen, inhibits COX, leading to analgesic and anti-inflammatory effects. It is also a potent inhibitor of sodium channels, as validated by the predicted BA values of −9.02 and −9.6 kcal/mol for SCN9A and SCN10A, respectively. The predicted BA value of −6.55 kcal/mol for hERG confirms the safety profile of naproxen on hERG. Our predictions indicate that naproxen may have side effects on other targets, with the top 3 predicted BA values being −11.35, −11.32, and −11.13 kcal/mol for CSNK2A2, FGFR2, and LATS2, respectively. This aligns with the known fact that naproxen can cause a range of potential side effects, including dizziness, headache, bruising, allergic reactions, and stomach pain.[38] In addition, naproxen demonstrates strong inhibition of TRPM8 with a predicted BA value of −9.97 kcal/mol.

Opioids are powerful pain-relieving medications commonly prescribed for moderate to severe pain. Examples of opioid medications include oxycodone (OxyContin, Roxicodone), hydrocodone (Vicodin, Hysingla ER), fentanyl (Actiq, Fentora), and morphine (MS Contin), among others. They function by binding to opioid receptors in the brain, spinal cord, and other parts of the body, thereby reducing the perception of pain. Because of their potential for misuse, addiction, and overdose, these medications are subject to strict prescribing guidelines.

Oxycodone, a strong semisynthetic opioid, is used medically to treat moderate to severe pain. Its mechanism of action involves interacting with opioid receptors in the central nervous system. The predicted BA values of oxycodone for SCN9A and SCN10A are −9.75 and −10.62 kcal/mol, respectively. The predicted BA value for hERG is remarkably low at −7.8 kcal/mol, indicating a low potential for hERG side effects, which is consistent with the result of the study by Fanoe et al.[15] Oxycodone demonstrates strong binding potency to the top 3 proteins: ROS1, CSNK2A2, and OPRM1, with the largest predicted BA values being 211.77, −11.47, and −11.45 kcal/mol, respectively. In addition, our predictions suggest that oxycodone can inhibit the TRPA1 (transient receptor potential ankyrin 1) protein, with a predicted BA value of −10.09 kcal/mol. Transient receptor potential ankyrin 1 is a thermosensitive ion channel involved in the detection and transmission of pain signals. It is known for its role in mediating various types of pain, particularly in response to chemical irritants and inflammatory stimuli.

Hydrocodone is indicated for the relief of acute pain, sometimes in combination with acetaminophen or ibuprofen. It is also used for the symptomatic treatment of the common cold and allergic rhinitis, often in combination with decongestants, antihistamines, and expectorants. Hydrocodone inhibits pain signaling in both the spinal cord and brain. Its actions in the brain can also lead to euphoria, respiratory depression, and sedation.[51]

In our predictions, hydrocodone demonstrates good binding affinities for SCN9A and SCN10A, with BA values of −9.72 and −10.56 kcal/mol, respectively. The predicted BA value for hERG is −8.16 kcal/mol, suggesting a low potential for side effects on hERG. Hydrocodone has the potential to cause serious side effects on the top 3 proteins: ROS1, CSNK2A2, and TACR1, with predicted BA values of −11.98, −11.40, and −11.36 kcal/mol, respectively. In addition, our findings indicate that hydrocodone is a strong binder to the TRPA1 protein, with a predicted BA value of −9.94 kcal/mol.

Some medications prescribed to manage depression and prevent epileptic seizures have been found to relieve chronic pain. Tricyclic antidepressants used in the treatment of chronic pain include amitriptyline and nortriptyline (Pamelor). Antiseizure medications used for chronic nerve pain include gabapentin (Gralise, Neurontin, Horizant) and pregabalin (Lyrica).

Amitriptyline, a tricyclic antidepressant, has been used for decades to treat depression and has been investigated for its analgesic properties in pain-related conditions.[8] Our predicted BA values for SCN9A and SCN10A are −9.74 and −10.04 kcal/mol, respectively, validating the potency of amitriptyline in pain treatment according to our predictions. The predicted BA value of amitriptyline on hERG is −8.25 kcal/mol, indicating a potential side effect on hERG, which conforms to that amitriptyline has been known to induce QT prolongation and torsades de pointes, which causes sudden death.[27]

The 3 strongest predicted BA values are for LATS2, HRH1, and KCNA3 proteins, with values of −11.08, −11.01, and −10.61 kcal/mol, respectively. Gabapentin, a structural analogue of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA), was originally developed as an antiepileptic medication. It is now widely used to treat neuropathic pain.[31] Our predictions suggest that gabapentin has the potential to inhibit SCN9A and SCN10A, with BA values of −9.0 and −9.35 kcal/mol, respectively. Moreover, gabapentin is predicted to have no side effects on hERG, with a BA value of −6.85 kcal/mol. In addition, our predictions show that the 3 strongest predicted BA values are for LATS2, KCNA3, and FGFR2, with values of −10.94, −10.61, and −10.6 kcal/mol, respectively.

3.5. Nearly optimal lead compounds from screening and repurposing

We dedicate our efforts to finding more potential inhibitors of the 2 pain treatment targets, SCN9A and SCN10A, through the screening and repurposing processes in this section. In the process of screening and repurposing, we used 110 ML models to predict the cross-target binding affinity. In addition to considering potency, we also ensured that the optimal ranges for the ADMET properties and SAS (as listed in Table 1), as well as the hERG side effect, were all well satisfied. SCN9A and SCN10A are not only major pain targets but also key pharmacological targets in pain treatment. To identify more promising potent compounds for these 2 targets, we used the 110 inhibitor data sets as a source of inhibitor compounds.

During the screening process, we selected potent inhibitor compounds with experimental BA values below −9.54 kcal/mol from the inhibitor data sets of the 2 pain treatment targets, SCN9A and SCN10A. We then evaluated a series of other properties. It is important to note that if a designated inhibitor of one treatment target demonstrates high efficacy on the other treatment target, it is not considered a side effect. This is because it is common for an inhibitor to be potent on both major pain treatment targets simultaneously. However, we still need to evaluate the potential for side effects on the other 108 side effect targets, as well as hERG. We require predicted BA values greater than −9.54 kcal/mol to exclude side effects, except for hERG, which has a stricter requirement of BA values greater than −8.18 kcal/mol.

For repurposing, we assess the binding potency of all weak inhibitors in the other 108 data sets of side effect targets on the 2 pain treatment targets, SCN9A and SCN10A. Therefore, we select inactive inhibitors with experimental BA values greater than −9.54 kcal/mol and identify those with predicted BA values less than 29.54 kcal/mol on the 2 pain treatment targets. In our search for inhibitors with repurposing potential on the pain targets, these inhibitors should have no side effects on the other 107 side effect targets, as well as hERG. Furthermore, we also study the optimal range of ADMET properties and synthetic accessibility.

It is not easy to find inhibitors that satisfy all the aforementioned requirements. In the end, we identified 2 inhibitor compounds, CHEMBL1767278 from the MAPK8 data set and CHEMBL1453498 from the CASP3 data set, for repurposing. We evaluated additional ADMET properties of these 2 molecular compounds using the ADMETlab 2.0 prediction solver (https://admetmesh.scbdd.com/). Figures 6A and B show that the 2 compounds fall within the optimal ranges of these ADMET properties. For more details on the meaning and optimal ranges of the 13 ADMET properties, please refer to Table S4 in the Supporting Information (available at http://links.lww.com/PAIN/B940). The compound CHEMBL1767278 is predicted to have BA values of −8.13 and −9.68 kcal/mol on SCN9A and SCN10A, respectively, whereas the compound CHEMBL1453498 is predicted to have values of −9.68 and −8.04 kcal/mol, indicating their potency on SCN10A and SCN9A, respectively. Their predicted BA values on hERG are −7.13 and −7.92 kcal/mol, respectively, suggesting favorable side effect profiles. The representations of the 2 compounds and their side effect predictions are provided in Figures 6C and D, respectively. Furthermore, these 2 compounds are predicted to have no binding or side effects on the remaining 96 and 99 proteins, respectively.

Next, we investigated the molecular interactions between the 2 inhibitors and the 2 main pain treatment targets, SCN9A and SCN10A, using the software AutoDock Vina.[23] Figures 7A and C shows the 3D protein–ligand docking structures, and Figures 7B and D shows the 2D interaction diagrams of the 2 compounds, CHEMBL1767278 and CHEMBL1453498, respectively. Because of the structural complexity of SCN9A and SCN10A, we focused on the docking between the inhibitors and the central sites of the targets. AutoDock Vina generated 9 docking poses with different docking scores calculated from its scoring function. In our figures, we selected the pose with the highest affinity (kcal/mol), where hydrogen bonds are formed between the inhibitors and the 2 pain targets SCN9A and SCN10A. In the docking of compound CHEMBL1767278 (Fig. 7B),

one strong hydrogen bond with Asn312 (2.85 Å) is formed, whereas in the docking of compound CHEMBL1453498 (Fig. 7D), 3 hydrogen bonds with Tyr1696 (2.98 Å, 2.92 Å) and Arg1599 (3.22 Å) are formed. The predicted binding energies of these 2 compounds with SCN10A and SCN9A are both −9.68 kcal/mol. In addition, we found that neither of the 2 compounds formed a covalent bond with the side chains of the targets during the docking process, suggesting that hydrogen bonds play vital roles in the interaction between the atoms.

4. Conclusion

Pain is a complex sensory and emotional experience that serves as a protective mechanism in response to potential or actual tissue damage. Sodium channels, particularly Nav1.3, Nav1.7, Nav1.8, and Nav1.9, play a significant role in the generation and transmission of pain signals in various pain conditions. However, progress in drug design for pain treatment has been relatively slow, and there is a need for more treatment options to be investigated.

Sodium channels are attractive targets for the development of pain medications. Pain affects complex molecular and biological activities in the nervous system, involving significant protein–protein interactions (PPI) in different brain regions. The development of pain treatment medications must take into account the influence of drugs on the PPI networks of pain targets. In this study, we construct an extended DTI network informed by 4 pain-related sodium channels. We develop a machine learning framework to screen and propose additional drug candidates for pain reduction. We use 2 molecular fingerprints generated by advanced natural language processing (NLP) models based on transformer and autoencoder algorithms. These fingerprints are then used to build predictive machine learning models employing 3 common machine learning algorithms: SVM, GBDT, and RF. A consensus model combining the predictions from these algorithms is used to enhance the overall predictive performance. In addition, we apply these machine learning models to reevaluate the side effects of existing pain-relieving medications. Our ML models are also employed to analyze the repurposing potential of existing inhibitor compounds on major pain targets and screen for possible side effects associated with these inhibitors. Furthermore, we implement the assessment of ADMET properties using machine learning predictions. Finally, we identify a group of promising compounds for major pain targets. Further testing through in vitro or animal experiments is necessary to evaluate the toxicity and blood–brain barrier permeability characteristics of these candidate compounds.

Our machine learning–based framework provides a novel method for searching candidate compounds for pain relief and can be generalized for other diseases with neurological implications. Although the sodium channel genes studied in this work are associated with pain perception and pain disorders, it is important to note that pain is a complex and multifactorial phenomenon involving numerous other factors and pathways. Further research is needed to fully understand the roles of these sodium channels in pain processing and to explore their potential as therapeutic targets for pain management. We could also use the present methodology to carry out a detailed study of endorphin and enkephalin receptors in future work, which plays a pivotal role in pain modulation. In addition, future work will

be dedicated to stringent experimental validation and to providing robust evidence for the practical implications of our findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in part by NIH grants R01GM126189, R01AI164266, and R35GM148196, NSF grants DMS-2052983, DMS-1761320, and IIS-1900473, NASA grant 80NSSC21M0023, MSU Foundation, Bristol-Myers Squibb 65109, and Pfizer. The work of Jian Jiang and Bengong Zhang was supported by the National Natural Science Foundation of China under Grant No. 12371500, No.12271416, and No.11972266.

Data and code availability:

The related data sets studied in present work are available at https://weilab.math.msu.edu/Data-Library/2D/. Codes of the calculation of two molecular fingerprints are available via https://github.com/WeilabMSU/OUD-PPI.

References

[1]. Avram S, Bora A, Halip L, Curpan R. Modeling kinase inhibition using highly confident data sets. J Chem Inf Model 2018;58:957–67.

[2]. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. Brief Bioinformat 2021;22:247–69.

[3]. Bennett DL, Clark AJ, Huang J, Waxman SG, Dib-Hajj SD. The role of voltage-gated sodium channels in pain signaling. Physiol Rev 2019;99:1079–151.

[4]. Black JA, Liu S, Tanaka M, Cummins TR, Waxman SG. Changes in the expression of tetrodotoxin-sensitive sodium channels within dorsal root ganglia neurons in inflammatory pain. PAIN 2004;108:237–47.

[5]. Black JA, Nikolajsen L, Kroner K, Jensen TS, Waxman SG. Multiple sodium channel isoforms and mitogen-activated protein kinases are present in painful human neuromas. Ann Neurol 2008;64:644–53.

[6]. Bosselmann CM, Hedrich UB, Lerche H, Pfeifer N. Learning with phenotypic similarity improves the prediction of functional effects of missense variants in voltage-gated sodium channels. bioRxiv 2022. doi: 10.1101/2022.09.29.510111.

[7]. Breiman L Random forests. Machine Learn 2001;45:5–32.

[8]. Bryson HM, Wilde MI. Amitriptyline: a review of its pharmacological properties and therapeutic use in chronic pain states. Drugs Aging 1996; 8:459–76.

[9]. Cang Z, Wei G-W. Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction. Int J Numer Methods Biomed Eng 2018;34:e2914.

[10]. Chen D, Zheng J, Wei G-W, Pan F. Extracting predictive representations from hundreds of millions of molecules. J Phys Chem Lett 2021;12:10793–801.

[11]. Cortes C, Vapnik V. Support-vector networks. Machine Learn 1995;20:273–97.

[12]. Cox JJ, Reimann F, Nicholas AK, Thornton G, Roberts E, Springell K, Karbani G, Jafri H, Mannan J, Raashid Y, Al-Gazali L, Hamamy H, Valente EM, Gorman S, Williams R, McHale DP, Wood JN, Gribble FM, Woods CG. An SCN9A channelopathy causes congenital inability to experience pain. Nature 2006;444:894–8.

[13]. Dib-Hajj SD, Yang Y, Black JA, Waxman SG. The Nav1. 7 sodium channel: from molecule to man. Nat Rev Neurosci 2013;14:49–62.

[14]. Erickson A, Deiteren A, Harrington AM, Garcia-Caraballo S, Castro J, Caldwell A, Grundy L, Brierley SM. Voltage-gated sodium channels:(Nav) igating the field to determine their contribution to visceral nociception. J Physiol 2018;596:785–807.

[15]. Fanoe S, Jensen GB, Sjøgren P, Korsgaard MP, Grunnet M. Oxycodone is associated with dose-dependent QTc prolongation in patients and low-affinity inhibiting of hERG activity in vitro. Br J Clin Pharmacol 2009;67:172–179.

[16]. Feng H, Jiang J, Wei G-W. Machine-learning repurposing of drugbank compounds for opioid use disorder. Comput Biol Med 2023;160:106921.

[17]. Feng H, Wei G-W. Virtual screening of drugbank database for hERG blockers using topological laplacian-assisted AI models. Comput Biol Med 2023;153:106491.

[18]. Fertleman CR, Baker MD, Parker KA, Moffatt S, Elmslie FV, Abrahamsen B, Ostman J, Klugbauer N, Wood JN, Gardiner RM, Rees M. SCN9A mutations in paroxysmal extreme pain disorder: allelic variants underlie distinct channel defects and phenotypes. Neuron 2006;52:767–74.

[19]. Flower DR. Drug design: cutting edge approaches. Vol. 279. London, United Kingdom: Royal Society of Chemistry, 2002.

[20]. Han C, Yang Y, de Greef BT, Hoeijmakers JG, Gerrits MM, Verhamme C, Qu J, Lauria G, Merkies IS, Faber CG, Dib-Hajj S, Waxman SG. The domain Ii S4-S5 linker in Nav1.9: a missense mutation enhances activation, impairs fast inactivation, and produces human painful neuropathy. Neuromolecular Med 2015;17:158–69.

[21]. Han C, Yang Y, Te Morsche RH, Drenth JP, Politei JM, Waxman SG, Dib-Hajj SD. Familial gain-of-function Nav1.9 mutation in a painful channelopathy. J Neurol Neurosurg Psychiatry 2017;88:233–40.

[22]. Herrera-Bravo J, Farias JG, Contreras FP, Herrera-Belen L, Beltran JF. PEP-PRED^{Na1}: a web server for prediction of highly specific peptides targeting voltage-gated Na+ channels using machine learning techniques. Comput Biol Med 2022;145:105414,

[23]. Huey R, Morris GM, Forli S. Using autodock 4 and autodock vina with autodocktools: a tutorial. Scripps Res Inst Mol Graph Lab 2012;10550:1000.

[24]. Jenssen MDK, Bakkevoll PA, Ngo PD, Budrionis A, Fagerlund AJ, Tayefi M, Bellika JG, Godtliebsen F. Machine learning in chronic pain research: a scoping review. Appl Sci 2021;11:3205.

[25]. Jiang J, Wang R, Wang M, Gao K, Nguyen DD, Wei G-W. Boosting tree-assisted multitask deep learning for small scientific datasets. J Chem Inf Model 2020;60:1235–44.

[26]. Jiang J, Wang R, Wei G-W. GGL-Tox: geometric graph learning for toxicity prediction. J Chem Inf Model 2021;61:1691–1700.

[27]. Jo S-H, Youm JB, Lee CO, Earm YE, Ho W-K. Blockade of the herg human cardiac k^+ channel by the antidepressant drug amitriptyline. Br J Pharmacol 2000;129:1474–80.

[28]. Kalliokoski T, Kramer C, Vulpetti A, Gedeck P. Comparability of mixed IC_{50} data–a statistical analysis. PLoS One 2013;8:e61007.

[29]. Kong W, Huang W, Peng C, Zhang B, Duan G, Ma W, Huang Z. Multiple machine learning methods aided virtual screening of Nav1.5 inhibitors. J Cell Mol Med 2023;27:266–76.

[30]. Kong W, Tu X, Huang W, Yang Y, Xie Z, Huang Z. Prediction and optimization of Nav1.7 sodium channel inhibitors based on machine learning and simulated annealing. J Chem Inf Model 2020;60:2739–53.

[31]. Kukkar A, Bali A, Singh N, Jaggi AS. Implications and mechanism of action of gabapentin in neuropathic pain. Arch Pharmacal Res 2013;36:237–51.

[32]. Laedermann CJ, Abriel H, Decosterd I. Post-translational modifications of voltage-gated sodium channels in chronic pain syndromes. Front Pharmacol 2015;6:263.

[33]. Landrum G A software suite for cheminformatics, computational chemistry, and predictive modeling, 2013. Available at: https://rdkit.sourceforge.net/.

[34]. Leipold E, Hanson-Kahn A, Frick M, Gong P, Bernstein JA, Voigt M, Katona I, Oliver Goral R, Altmuller J, Nurnberg P, Weis J, Hubner CA, Heinemann SH, Kurth I. Cold-aggravated pain in humans caused by a hyperactive Nav1.9 channel mutant. Nat Commun 2015;6:10049.

[35]. Li X, Zhang Y, Li H, Zhao Y. Modeling of the hERG K+ channel blockage using online chemical database and modeling environment (OCHEM). Mol Inform 2017;36:1700074.

[36]. LoMartire R, Dahlström Ö, Bjork M, Vixner L, Frumento P, Constan L, Gerdle B, Äng BO. Predictors of sickness absence in a clinical population with chronic pain. J Pain 2021;22:1180–94.

[37]. Lotsch J, Ultsch A. Machine learning in pain research. PAIN 2018;159: 623–30.

[38]. Maniar KH, Jones IA, Gopalakrishna R, Vangsness CT Jr. Lowering side effects of nsaid usage in osteoarthritis: recent attempts at minimizing dosage. Expert Opin Pharmacother 2018;19:93–102.

[39]. Mann N, King T, Murphy R. Review of primary and secondary erythromelalgia. Clin Exp Dermatol 2019;44:477–82.

[40]. Matsangidou M, Liampas A, Pittara M, Pattichi CS, Zis P. Machine learning in pain medicine: an up-to-date systematic review. Pain Ther 2021;10:1067–84.

[41]. Miettinen T, Mantyselka P, Hagelberg N, Mustola S, Kalso E, Lotsch J. Machine learning suggests sleep as a core factor in chronic pain. PAIN 2021;162:109–23.

[42]. Mulcahy JV, Pa Jouhesh H, Beckley JT, Delwig A, Du Bois J, Hunter JC. Challenges and opportunities for therapeutics targeting the voltage-gated sodium channel isoform Nav1. 7. J Med Chem 2019;62:8695–710. [PubMed: 31012583]

[43]. Nguyen PT, Yarov-Yarovoy V. Towards structure-guided development of pain therapeutics targeting voltage-gated sodium channels. Front Pharmacol 2022;13:842032. [PubMed: 35153801]

[44]. Noda M, Ikeda T, Kayano T, Suzuki H, Takeshima H, Kurasaki M, Takahashi H, Numa S. Existence of distinct sodium channel messenger RNAS in rat brain. Nature 1986;320:188–92. [PubMed: 3754035]

[45]. Okuda H, Noguchi A, Kobayashi H, Kondo D, Harada KH, Youssefian S, Shioi H, Ka-bata R, Domon Y, Kubota K, Kitano Y, Takayama Y, Hitomi T, Ohno K, Saito Y, Asano T, Tominaga M, Takahashi T, Koizumi A. Infantile pain episodes associated with novel Nav1.9 mutations in familial episodic pain syndrome in Japanese families. PLoS One 2016;11:e0154827. [PubMed: 27224030]

[46]. Riechers RG, Walker MF, Ruff RL. Chapter 36—post-traumatic headaches. In: Grafman J, Salazar AM, editors. Traumatic brain injury, part II, volume 128 of handbook of clinical neurology. Amsterdam, the Netherlands: Elsevier, 2015. p. 567–78.

[47]. Robinson ME, O'Shea AM, Craggs JG, Price DD, Letzen JE, Staud R. Comparison of machine classification algorithms for fibromyalgia: neuroimages versus self-report. J Pain 2015;16:472–7. [PubMed: 25704840]

[48]. Rowe AH, Xiao Y, Rowe MP, Cummins TR, Zakon HH. Voltage-gated sodium channel in grasshopper mice defends against bark scorpion toxin. Science 2013;342:441–6. [PubMed: 24159039]

[49]. Steglitz J, Buscemi J, Ferguson MJ. The future of pain research, education, and treatment: a summary of the IOM report "Relieving pain in America: a blueprint for transforming prevention, care, education, and research". Transl Behav Med 2012;2:6–8. [PubMed: 24073092]

[50]. Su X, Young EW, Underkofler HA, Kamp TJ, January CT, Beebe DJ. Microfluidic cell culture and its application in high-throughput drug screening: cardiotoxicity assay for hERG channels. J Biomol Screen 2011;16:101–111. [PubMed: 21131594]

[51]. Trescot AM, Datta S, Lee M, Hansen H. Opioid pharmacology. Pain Physician 2008;11(suppl 2):S133–53. [PubMed: 18443637]

[52]. Tseng T-T, McMahon AM, Johnson VT, Mangubat EZ, Zahm RJ, Pacold ME, Jakobsson E. Sodium channel auxiliary subunits. Microb Physiol 2007;12:249–62.

[53]. von Buchholtz LJ, Lam RM, Emrick JJ, Chesler AT, Ryba NJ. Assigning transcriptomic class in the trigeminal ganglion using multiplex in situ hybridization and machine learning. PAIN 2020;161:2212–24. [PubMed: 32379225]

[54]. Wang R, Fang X, Lu Y, Yang C-Y, Wang S. The pdbbind database: methodologies and updates. J Med Chem 2005;48:4111–9. [PubMed: 15943484]

[55]. Winter R, Montanari F, Noe F, Clevert D-A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chem Sci 2019;10:1692–701. [PubMed: 30842833]

[56]. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, Wu C, Lu A, Chen X, Hou T, Cao D. Admetlab 2.0: an integrated online platform for accurate and comprehensive predictions of admet properties. Nucleic Acids Res 2021;49:W5–14. [PubMed: 33893803]

[57]. Zhang X, Mao J, Wei M, Qi Y, Zhang JZH. HergSPred: accurate classification of hERG blockers/nonblockers with machine-learning models. J Chem Inf Model 2022;62:1830–9. [PubMed: 35404051]

[58]. Zhu Z, Dou B, Cao Y, Jiang J, Zhu Y, Chen D, Feng H, Liu J, Zhang B, Zhou T, Wei G-W. Tidal: topology-inferred drug addiction learning. J Chem Inf Model 2023;63:1472–89. [PubMed: 36826415]

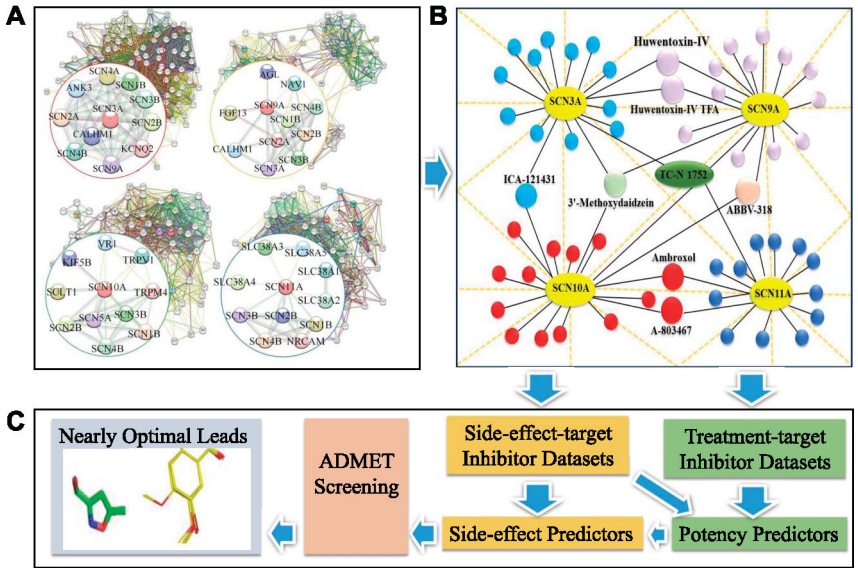

Figure 1.

The flowchart of screening nearly optimal lead compounds for inhibiting pain-related voltage-gated sodium channels (VGSCs). (A) Protein–protein interaction (PPI) networks of the 4 VGSCs involve more than 1000 proteins, including 4 treatment targets SCN3A, SCN9A, SCN10A, and SCN11A. Each of them has a core and global PPI network. Further details of the PPI networks are provided in the Table S1 in the Supporting Information (available at: http://links.lww.com/PAIN/B940). (B) The drug–target interaction (DTI) network involves 111 targets and 150,147 inhibitor compounds. Here, only 4 treatment targets (SCN3A, SCN9A, SCN10A, and SCN11A) with several compounds are displayed for simplicity. The yellow dashed lines indicate the connections among 111 targets. (C) Predictive models for side effect and repurposing evaluation, as well as ADMET screening. ADMET, absorption, distribution, metabolism, excretion, and toxicity.

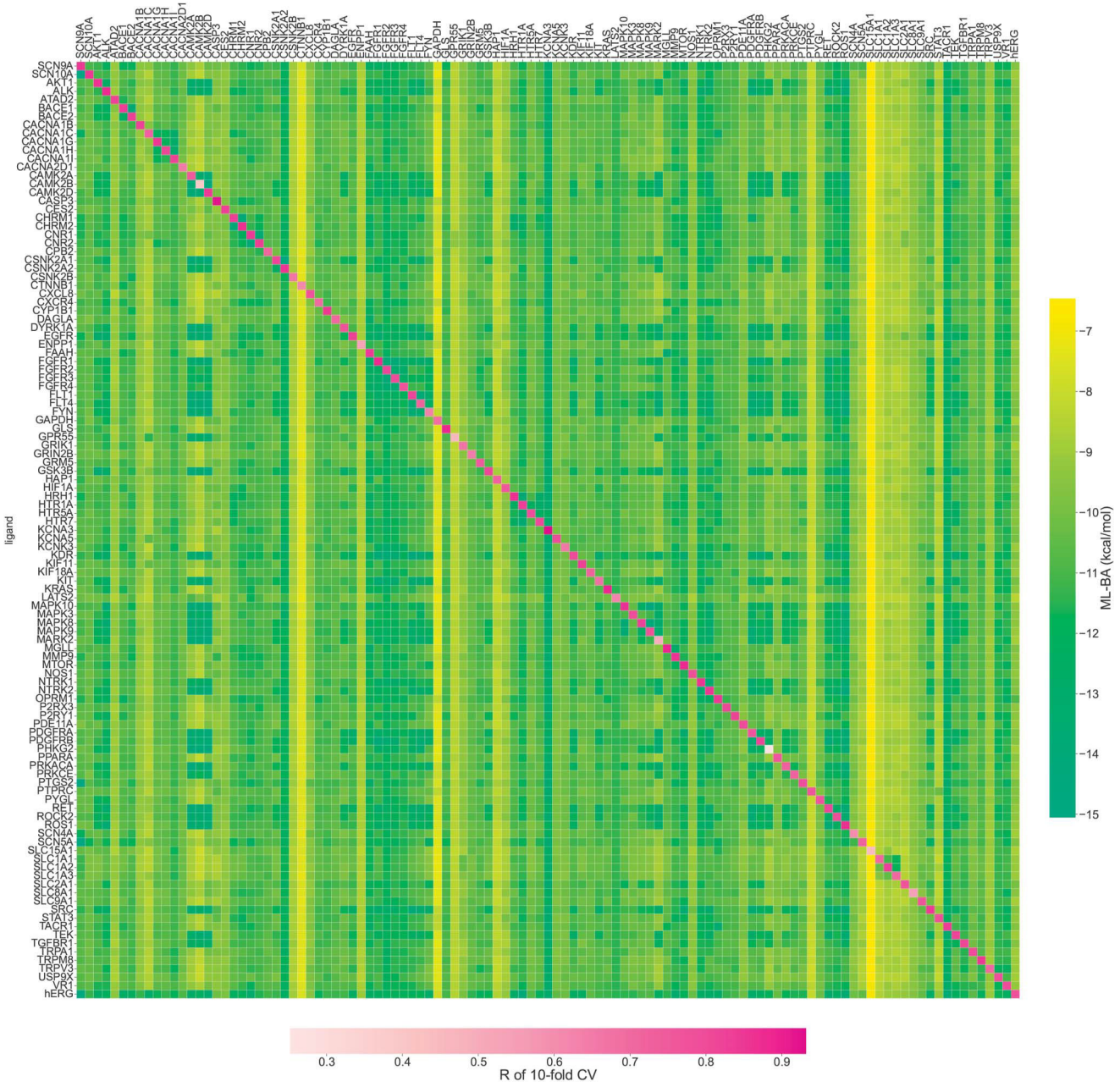

Figure 2.

The heatmap of cross-target binding affinities (BAs) predictions for the extended DTI networked informed by 4 pain-related voltage-gated sodium channels. The left labels of the heatmap represent all the inhibitor data sets and those above the heatmap mean the machine learning (ML) models. The diagonal elements in the heatmap denote the Pearson correlation efficient (R) of 10-fold cross-validation for all the ML models. The off-diagonal elements in each row indicate the highest BA values of inhibitors of one data sets predicted by 111 ML models. This heatmap is used to reveal the inhibitor specificity of each data set on other protein targets. DTI, drug–target interaction.

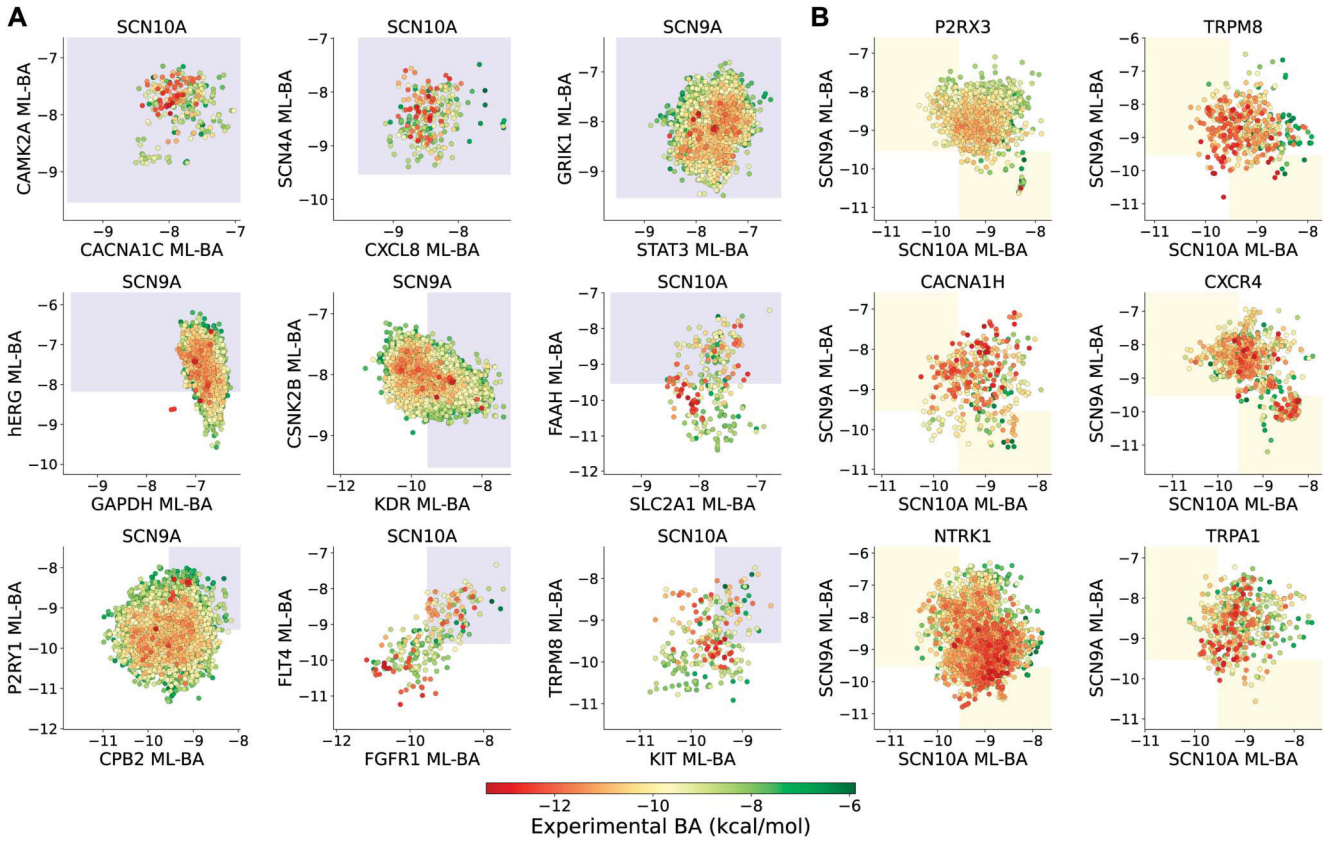

Figure 3.

Examples of predictions of side effects and repurposing potentials. (A) The first row, second row, and third row represent example inhibitor data sets of 2 treatment targets SCN9A and SCN10A that have side effects on none, 1, and 2 of the given 2 side effect targets, respectively. The blue frames indicate where there are no side effects. (B) Displays example inhibitor data sets of side effect targets that are equipped with repurposing potentials on treatment targets SCN9A and SCN10A. The yellow frames indicate that the inhibitors have repurposing potential for one treatment target but have no side effect on the other treatment target.

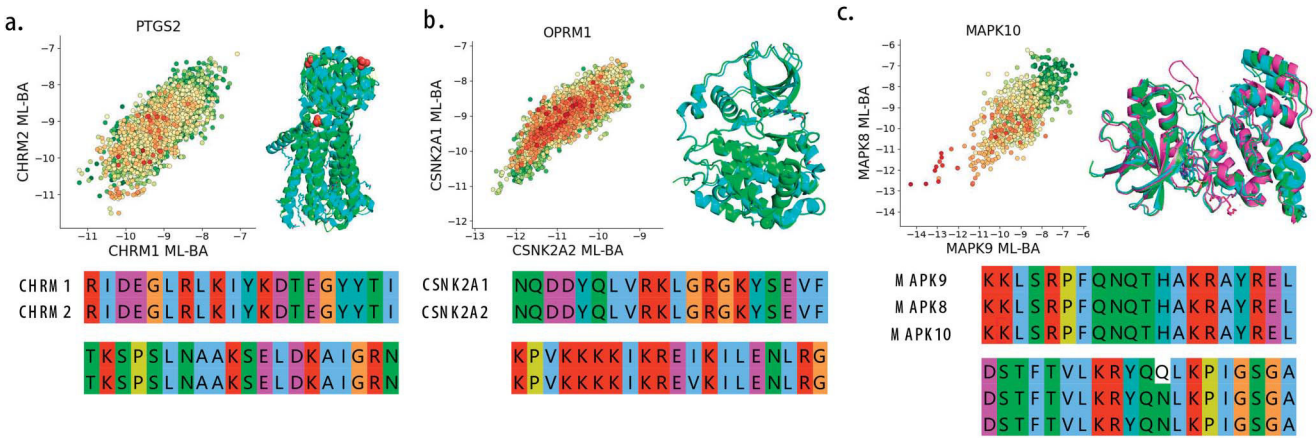

Figure 4.

Three examples of correlated predicted BA values suggesting the structure and/or sequence similarities of proteins. In each panel, the x-axis and y-axis represent the predicted BA values on 2 other proteins, and the scattered points with colors indicate the experimental labels of inhibitors of the target. The 3D structure alignment is shown in the right of the panel, and the 2D sequence alignment is shown below. In the 3D structure alignment, PDB 6ZG4 and 3UON are used for CHRM1 and CRMH2, PDB 6QY7 and 6QY9 for CSNK2A1 and CSNK2A2, PDB 3ELJ, 7N8T, and 3KVX for MAPK8, MAPK9, and MAPK10, respectively. BA, binding affinities.

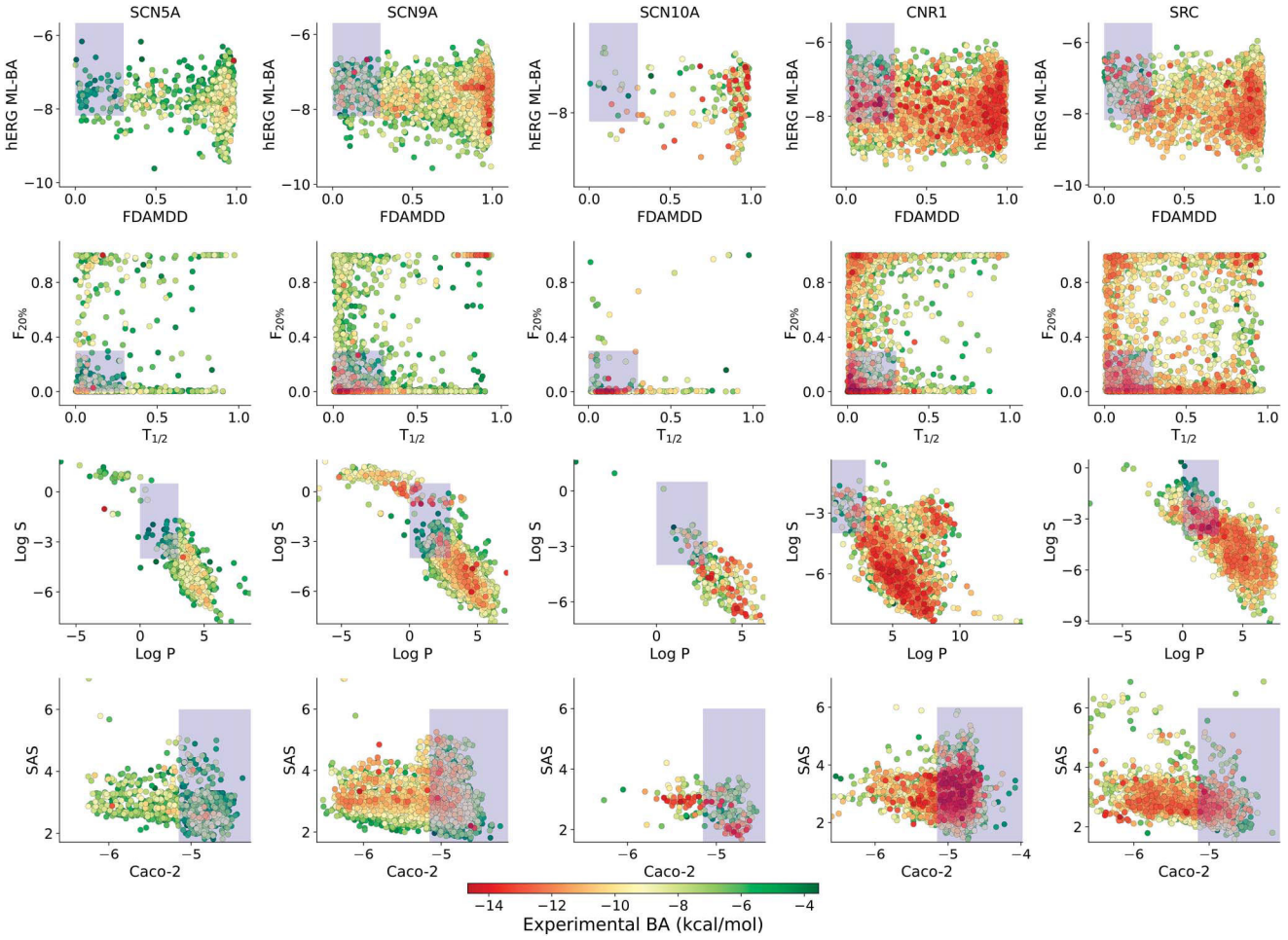

Figure 5.

Druggable property screening based on ADMET properties, synthesizability, and hERG side effects on compounds from 5 protein data sets: SCN5A, SCN9A, SCN10A, CNR1, and SRC. The colors of the points indicate the experimental BAs for these targets. The x- and y-axis represent various predicted ADMET properties, synthesizability, or hERG side effects. Blue frames highlight the optimal ranges of these properties and side effects. ADMET, absorption, distribution, metabolism, excretion, and toxicity.

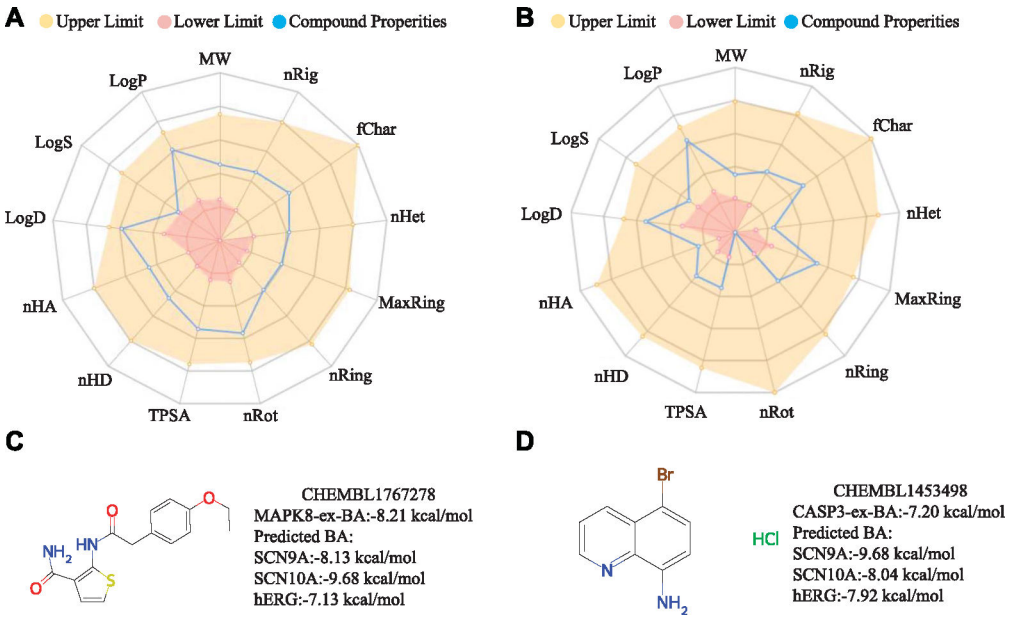

Figure 6.

Assessment of 13 ADMET properties for those molecular compounds with repurposing potentials. (A and B) indicate the evaluations of ADMET properties of 2 compounds CHEMBL1767278 and CHEMBL1453498, and C and D represent their chemical graphs and predictions of side effects, respectively. The boundaries of yellow and red regimes in A and B show the upper and lower limits of the optimal ranges for 13 ADMET properties, respectively. The blue curves suggest values of the specified 13 ADMET properties. The details of these property abbreviations are as following: MW, molecular weight; logP, log of octanol/water partition coefficient; logS, log of the aqueous solubility; logD, logP at physiological pH 7.4; nHA, number of hydrogen bond acceptors; nHD, number of hydrogen bond donors; TPSA, topological polar surface area; nRot, number of rotatable bonds; nRing, number of rings; MaxRing, number of atoms in the biggest ring; nHet, number of heteroatoms; fChar, formal charge; nRig, number of rigid bonds; ADMET, absorption, distribution, metabolism, excretion, and toxicity.

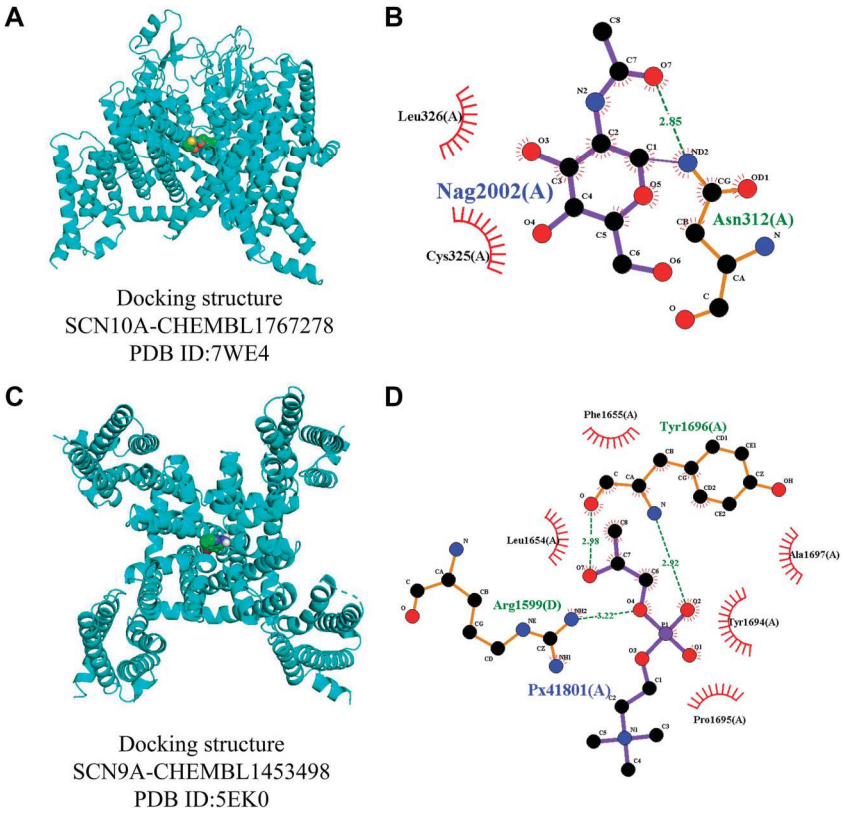

Figure 7.

The docking structure of our 2 optimal lead compounds bound to 2 pain targets SCN0A and SCN10A, and their 2D interaction diagrams. We use AutoDock Vina to implement the protein–ligand docking and find the hydrogen bonds generated during the docking of 2 compounds.

Table 1

The optimal ranges of selected absorption, distribution, metabolism, excretion, and toxicity properties and synthetic accessibility used for screening compounds in this work.

Property	Optimal ranges
FDAMDD	Excellent: 0–0.3; medium: 0.3–0.7; poor: 0.7–1.0
$F_{20\%}$	Excellent: 0–0.3; medium: 0.3–0.7; poor: 0.7–1.0
Log P	The proper range: 0–3 log mol/L
Log S	The proper range: −4–0.5 log mol/L
$T_{1/2}$	Excellent: 0–0.3; medium: 0.3–0.7; poor: 0.7–1.0
Caco-2	The proper range: > −5.15
SAS	The proper range: <6

SAS, synthetic accessibility.

FDAMDD represents the FDA maximum recommended daily dose. Caco-2 is a measure used to estimate the in vivo permeability of oral drugs.