Send Orders for Reprints to reprints@benthamscience.ae

Prediction of Ion Channels and Their Types from Protein Sequences: Comprehensive Review and Comparative Assessment

Jianzhao Gao[*a], Zhen Miao[b], Zhaopeng Zhang[a], Hong Wei[a] and Lukasz Kurgan[*c]

[a]*School of Mathematical Sciences and LPMC, Nankai University, Tianjin, China;* [b]*College of Life Sciences, Nankai University, Tianjin, China;* [c]*Department of Computer Science, Virginia Commonwealth University, Richmond, USA*

Abstract: Background: Ion channels are a large and growing protein family. Many of them are associated with diseases and consequently they are targets for over 700 drugs. Discovery of new ion channels is facilitated with computational methods that predict ion channels and their types from protein sequences. However, these methods were never comprehensively compared and evaluated. **Objective:** We offer first-of-its-kind comprehensive survey of the sequence-based predictors of ion channels. We describe eight predictors that include five methods that predict ion channels, their types, and four classes of the voltage-gated channels. We also develop and use a new benchmark dataset to perform comparative empirical analysis of the three currently available predictors. **Results:** While several methods that rely on different designs were published, only a few of them are currently available and offer a broad scope of predictions. Support and availability after publication should be required when new methods are considered for publication. Empirical analysis shows strong performance for the prediction of ion channels and modest performance for the prediction of ion channel types and voltage-gated channel classes. We identify a substantial weakness of current methods that cannot accurately predict ion channels that are categorized to multiple classes/types. **Conclusions:** Several predictors of ion channels are available to the end users. They offer practical levels of predictive quality. Methods that rely on a larger and more diverse set of predictive inputs (such as PSIONplus) are more accurate. New tools that address multi-label prediction of ion channels should be developed.

Keywords: ion channel; voltage-gated ion channel; ligand-gated ion channel; prediction.

1. INTRODUCTION

Ion channels are integral membrane proteins that form a water-filled pore. They control flow of ions and voltage potential across cell membranes [1, 2]. While some of these channels are selective for specific ions, such as sodium or potassium, others may facilitate passage of multiple ion types, typically sharing the same positive or negative charge. Ion channels are typically categorized based on their gate types into voltage-gated and ligand-gated ion channels [3, 4]. Upon binding of the ligand, the ligand-gated channels undergo a conformational change that leads to the opening of the channel gate and ion flux. The voltage-gated ion channels open and close depending on the voltage gradient across the cell membrane. They are further classified into several classes including potassium (K^+), sodium (Na^+), calcium (Ca^{2+}) and anion channels [4].

The ion channels are expressed in virtually all tissues and cell types. They were found to be associated with over 30 diseases. A few selected examples include retinal disease [5], deafness [6], renal cysts [7], cardiac arrhythmias [6, 8, 9] migraines [10], and epilepsy [6]. Correspondingly, many drugs target ion channels. One example is an antiarrhythmic drug lidocaine that acts as a voltage-gated sodium channel inhibitor [11]. Lidocaine's action affects the conduction system and muscle cells of the heart, raising the depolarization threshold of heart and making it less likely to initiate or conduct action potentials [12]. Another example is *ziconotide* that targets calcium channels and is used for pain relief [13]. This compound blocks calcium influx into nerve terminals, which results in reduced release of glutamate and neuropeptides effectively interrupting spinal transmission of pain signal [14].

Research shows that about 19% of human protein drug targets are in the ion channel family [15]. The number of ion channel-targeting drugs that have been approved by the US Food and Drug Administration (FDA) is steadily increasing since mid-1940s (Figure 1, these data were taken from Table S6 in [15]). However, ion channels are still underutilized in the context of drug discovery and some of the ion channel-targeting drugs have substantial toxicities and suboptimal efficacy [2]. Interestingly, the family of human ion channel proteins was estimated to account for 1% of the human protein coding genes, to be larger than the nuclear hormone receptor family, and to be as large as about half the size of the kinase and protease families [2]. Over 400 putative ion channels were found in the human genome alone [2] and over 300 types of

*Address correspondence to these authors: School of Mathematical Sciences and LPMC, Nankai University, Tianjin, China, Tel: 86-22-23501449, Email: gaojz@nankai.edu.cn; Department of Computer Science, Virginia Commonwealth University, Richmond, USA Tel: 804-827-3986, E-mail: lkurgan@vcu.edu.

XXX-XXX/14 $58.00+.00

ion channels were already discovered [16]. The number of annotated ion channel proteins has grown rapidly in recent years. Figure **1** estimates this growth based on UniProt records [17] that were deposited over the last three decades.

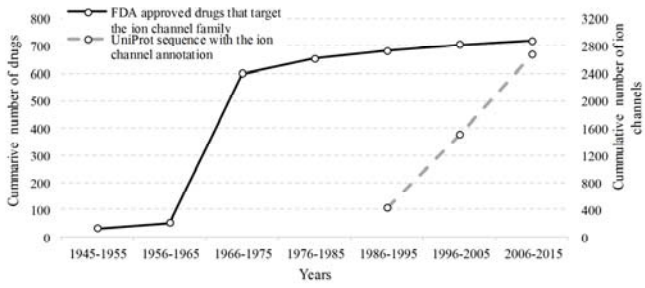

Figure 1. Number of ion channels annotated in UniProt and the number of drugs that target these proteins over the last seven decades. The cumulative number of drugs that target the ion channel family approved by the US Food and Drug Administration (FDA) are shown using solid line. The cumulative number of UniProt sequences is shown in dashed line.

The interest in the drugs that target ion channels and the need to expand the annotations of the ion channel proteins motivate the development of high-throughput computational tools that predict ion channels and their types directly from the protein sequences. Several computational methods were published over the past decade. However, to the best of our knowledge, they were never comprehensively surveyed and compared. This first-of-its-kind review article surveys the currently available computational tools and performs in-depth comparative assessment of their predictive performance. The unique aspects of this survey include the first empirical assessment of sequential prediction of the ion channels, their types and classes of voltage-gated channels; use of a novel benchmark dataset; and discussion of the predictions for ion channels that are classified into multiple types/classes. The latter aspect is overlooked by the current tools that assume that a given protein sequence can be classified into one specific ion channel type. The sequential prediction first classifies an input protein chain as ion channel vs. non-ion channel, following by the prediction of the ion channel type (voltage- vs. ligand-gated) for the putative ion channels, and finishing with prediction of the voltage-gated channel class (K^+, Na^+, Ca^{2+}, vs. anion channel) for the putative voltage-gated channels. This is contrast to the prior assessments that evaluate predictions in parallel. This means that the input protein sequence is predicted as ion channel vs. non-ion channel; only the native ion channels are predicted as voltage- vs. ligand-gated; and only proteins in the voltage-gated class are predicted for the native voltage-gated channels. The main drawback of this assessment is that it assumes knowledge of native annotations of channel types and voltage-gated channel classes, which arguably is not a realistic scenario. One exception is the VGIchan method that performs sequential prediction of the voltage-gated ion channels and their classes. However, VGIchan does not cover the ligand-gated channels and, to the best of our reading of the corresponding article, it was evaluated in the parallel mode [18]. We perform assessment in parallel and sequential modes.

2. SURVEY OF PREDICTORS OF ION CHANNELS, THEIR TYPES, AND CLASSES OF VOLTAGE-GATED CHANNELS

Eight computational methods have been developed to predict the ion channels, their types and classes of the voltage-gated channels. Table **1** summarizes the year of their publication, scope of their predictions, details of their predictive models, and availability.

The first two methods were released in 2006 and these are the only methods developed before 2011. One predictor was released nearly every year since 2011. The eight methods can be divided into two groups based on the scope of their predictions: three predictors of subfamilies of the K^+ channel [19-21] and five more generic methods that predict ion channels and their types [18, 22-25]. The three most recent predictors, which include IonchanPred2.0, PSIONplus and the predictor by Tiwari and Srivastava, are in the latter group. They share the same scope that includes prediction of ion channels, their types and classes of the voltage-gated channels. The only other tool with equally broad scope is IonchanPred1.0 that was released in 2011. The VGIchan method predicts ion channels and classes of the voltage-gated channels, while missing the prediction of ion channel types.

The predictive models utilized by the eight methods use a diverse range of inputs, from simple models that rely on a single input type to complex models that utilize up to six distinct input types. The simplest predictors include the method by Liu *et al* [20] and iVKC-OTC [21] that use dipeptide composition (frequency of dipeptides in the input protein chain) and tripeptide composition (frequency of tripeptides), respectively, as the only inputs. The most complex model is arguably PSIONplus that relies on the information about amino acid (AA) composition, dipeptide composition, physicochemical properties of AAs, putative intrinsic disorder, putative solvent accessibility and evolutionary profile in the form of PSSM [26]. The first three types of inputs are often used in other bioinformatics applications [27-35], likely because they are easy and fast to calculate. The broad range and number of inputs used by different predictors of ion channels is likely to have substantial impact on the predictive quality of the corresponding methods. This survey empirically tests this hypothesis. Interestingly, these diverse inputs are processed by similar machine learning algorithms to produce the predictions. Seven out of the eight methods use the Support Vector Machine (SVM) algorithm. This is consistent with popularity of this model in the prediction of various structural and functional characteristics of proteins [32, 34, 36-47]. The only divergent tool by Tiwari and Srivastava [24] uses the Random Forest (RF) algorithm. Moreover, VGIchan applies SVM in combination with hidden Markov model (HMM) [18] while PSIONplus combines SVM with PSI-BLAST [26]. Interestingly, deep learning models that were recently used in a number of related studies [29, 48-50] have not yet been used to predict ion channels.

Table 1. Computational predictors of ion channels, their types and classes of the voltage-gated channels. Webservers and implementations that are unavailable as of July 1, 2018 are given in italics in the URL column.

Method name or authors [ref]	Year	Scope of predictions[b]	Features[c]	Type of Model[d]	Availa-bility[a]	Support for batch predictions	URL[a]
Liu *et al.* [20]	2006	Subfamilies of K^+ channels	DPC	SVM	NA	NA	NA
VGIchan [18]	2006	Ion vs. non-ion; K^+, Ca^{2+}, Na^+, anion	AAC, DPC, HMM profile; PSI-BLAST	SVM; SVM+BLAST; SVM+HMM;	WS	Yes	http://crdd.osdd.net/raghava/vgichan/
IonchanPred1.0 [23]	2011	Ion vs non-ion; ligand- vs voltage- gated; K^+, Ca^{2+}, Na^+, anion	AAC, DPC	SVM	WS	Yes	*http://cobi.uestc.edu.cn/people/hlin/tools/IonchanPred/*
VKCPred [19]	2012	Subfamily of K^+ channel	AAC, DPC	SVM	SS	Unknown	*http://cobi.uestc.edu.cn/people/hlin/tools/VKCPred/*
iVKC-OTC [21]	2014	Subfamily of K^+ channels	TPC	SVM	WS	Yes	http://lin-group.cn/server/iVKC-OTC
Tiwari and Srivastava [24]	2015	Ion vs non-ion; ligand- vs. voltage-gated; K^+, Ca^{2+}, Na^+, Cl^-	AAC, DPC, correlation, transition and distribution, Pseudo AAC	RF	NA	NA	NA
PSIONplus [22]	2016	Ion vs non-ion; ligand- vs. voltage-gated; K^+, Ca^{2+}, Na^+, anion	AAC, DPC, PP, putative intrinsic disorder, RSA, PSSM profile	SVM+PSI-BLAST	SS	Yes	https://sourceforge.net/projects/psion/
IonchanPred2.0 [25]	2017	Ion vs non-ion; ligand- vs. voltage-gated; K^+, Ca^{2+}, Na^+, anion	DPC, PP	SVM	WS	Yes	http://lin-group.cn/server/IonchanPredv2.0/

[a] NA: not applicable, WS: webserver, SS: Standalone software. [b] K^+: potassium; Ca^{2+}: calcium; Na^+: sodium; Cl^-: Chloride. [c] AAC: amino acid composition, DPC: dipeptide composition, TPC: tripeptide composition, PP: physicochemical properties, RSA: relative solvent accessibility, HMM: hidden Markov model, PSSM: position specific scoring matrix. [d] RF: random forest; SVM: support vector machine.

XXX-XXX/14 $58.00+.00

Six methods are offered to end users as either webservers (VGIchan, IonchanPred1.0, iVKC-OTC and IonchanPred2.0) or standalone software (PSIONplus and VKCPred). The two other tools were not made available by the authors, rendering them impractical. Websites of the two initially available tools (IonchanPred1.0 and VKCPred) no longer work (Table **1**).

Interestingly, half of the predictors, (VGIchan, iVKC-OTC, PSIONplus and IonchanPred2.0) are still functional and offer a batch prediction mode where multiple proteins can be predicted together. This makes it easier for the users to perform predictions for a larger sets of proteins, such as protein families or proteomes.

Overall, we observe that while many tools were published, only a few of them are currently available and offer a broad scope of predictions. These tools include VGIchan, PSIONplus and IonchanPred2.0. They are capable of predicting ion channels, their types (expect for VGIchan) and classes of the voltage gated channels. Moreover, the lack of availability and support for the published tools after they were released is a major problem that should be addressed in the future.

3. COMPARATIVE ASSESSMENT OF THE ION CHANNEL PREDICTORS

We empirically compare predictive performance of the currently available and representative ion channel predictors. We consider the methods that are currently available as either webserver or standalone software since only these tools can be easily utilized by the end users. Moreover, we limit the considered methods to those that offer a broad scope of predictions, i.e., tools that predict ion channels and their types. Three of the eight surveyed methods meet these two practical conditions: IonchanPred2.0 [25], VGIchan [18] and PSIONplus [22] . These three tools cover a diverse range of models, from a simple IonchanPred2.0 that uses two types of inputs, through the intermediate VGIchan that utilizes four types of inputs, to the most complex PSIONplus that relies on six types of inputs. Three highlights of this comparative empirical analysis include: 1) assessment of two alternative ways to make these predictions: parallel and sequential; 2) a novel and high-quality benchmark dataset; and 3) analysis of the results for proteins that are annotated with single vs. multiple ion channel types/classes.

3.1. Materials and methods

3.1.1. Two ways to predict ion channels and their types/classes from protein sequence:

The outcomes generated by the considered predictors include: ion channel vs. non-ion channel; voltage-gated vs. ligand-gated ion channel; four classes of voltage-gated ion channels: potassium (K^+), calcium (Ca^{2+}), sodium (Na^+) and anion. There are two ways to make these prediction:

- Parallel prediction (Figure **2A**) where prediction of ion channel vs. non-ion channel, voltage- vs. ligand-gated ion channel, and the prediction of the four classes of voltage-gated ion channel (K^+, Ca^{2+}, Na^+ and anion) are performed individually. The inputs for these three predictions are a generic protein sequence, a native ion channel sequence (the outcomes do not include non-ion channels) and a native voltage-gated ion channel (the outcomes do not include non-voltage-gated ion channel), respectively.
- Sequential prediction (Figure **2B**) where in the first step a generic sequence is predicted as ion channel vs. non-ion channel. In the second step the prediction of the ion channel type (voltage- vs. ligand-gated) is performed for the putative ion channels generated in the first step. Finally, the prediction of the voltage-gated channel class (K^+, Na^+, Ca^{2+} vs. anion channel) is performed for the putative voltage-gated channels generated in the second step.

A substantial drawback of the parallel approach is that the predictions of the ion channel type and the voltage-gated channel class require that the input protein sequence is already known to be the ion channel and voltage-gated channel, respectively. The published methods were tested in the parallel way, where tests were performed on the entire test dataset for the prediction of ion channels, using only the ion channels from the test dataset for the prediction of ion channel types, and using only the voltage-gated channels from the test dataset for the prediction of the voltage-gated channel classes. To the best of our knowledge, we are the first to evaluate these predictors in a more realistic sequential way.

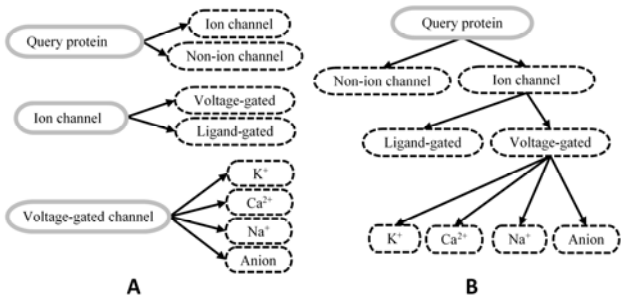

Figure 2. Two alternative ways to predict the ion channels and ion channel types. Solid gray lines denote inputs while dashed black lines denote putative annotations generated by the predictors. Panel A shows the parallel prediction while Panel B represents sequential prediction.

3.1.2. Benchmark dataset:

We developed a new benchmark dataset to comparatively evaluate predictive performance of the three predictors. The dataset is designed to share low identity to the training dataset of the three methods. This ensures that these predictions are beyond the capability of a simple sequence alignment and that the evaluation does not favor any of these tools. The overall protocol to collect this dataset is inspired by [18, 22, 23, 25]. The only differences are that we maintain a stricter sequence similarity cut-off at 30% when compared to 40% used in [23, 25] and 90% used in [18] ([22] uses the same 30% cut-off), and that similar to [18] we manually verify the ion channel annotations using the UniProt records. The dataset was collected on December 21, 2017 using the following five steps:

1. We collected the annotations of ion channels and their types/classes from UniProt using reviewed manual annotations and excluding annotated proteins fragments. We collected 473 ligand-gated ion channels by executing the following query: Gene Ontology (GO) term "ligand-gated ion channel activity [ligand-gated ion channel activity [15276]" AND Evidence level "Any manual assertion" AND Reviewed "Yes" AND Sequence "Fragment" "No". We found 660 voltage-gated ion channels by running the following query: GO term "voltage-gated ion channel activity [voltage-gated ion channel activity [5244]" AND Evidence level "Any manual assertion" AND Reviewed "Yes" AND Sequence "Fragment" "No". We also collected 13,657 high quality non-ion channels that include other types of membrane proteins with the following query: GO term "membrane [16020]" AND Evidence level "Any manual assertion" AND Reviewed "Yes" AND Sequence "Fragment" "No" AND Subcellular location Subcellular location[CC] Subcellular location term "Membrane[SL-0162] with Evidence "Any manual assertion" Type "Any". We retain the membrane proteins that do not include the keyword "channel" in GO molecular function annotation.
2. We cleaned up the dataset generated in step 1 to remove sequences with non-standard AAs and very short chain that are likely to be protein fragments. In particular, we excluded sequences that are shorter than 30 AA and that include X, B and Z characters. The resulting dataset includes 14,739 proteins.
3. We reduced the sequence similarity between the collected ion channels and non-ion channels and the training datasets of IonchanPred2.0, VGIchan and PSIONplus. First, we combined the 14,739 proteins with the training datasets of the three predictors. The combined dataset was processed with CD-HIT webserver [51, 52] using sequence identity cut-off 30% to generate 5,739 clusters. Clusters that contain any training sequences were removed. For each clusters that does not include training sequences we extracted sequences as follows:
   - We collected all ion channels sequences for each cluster that includes ion-channels
   - We collected a representative non-ion channel sequence identified by CD-HIT for the clusters that do not include ion-channels

   This resulted in selection of 4,819 proteins that share < 30% similarity with the training datasets of the three predictors.
4. We manually verified and extended the ion channel annotations. We also manually checked correctness of the ion channel annotations. Consequently, we removed 28 annotated proteins which are unlikely to be ion channels. The removed proteins include several GPCRs, scaffolding proteins and transcriptional repressors. We added annotations of the voltage-gated channel classes based on the GO molecular function annotation from UniProt using keywords "sodium", "potassium", "calcium" "anion" and "chloride". There are 29 proteins with multiple annotations. The sequences with multiple annotations were repeated once for each annotation to generate all corresponding sequence-label tuples.
5. We created the final balanced benchmark dataset. The output from steps 3 and 4 is the dataset with 4,639 non-ion channels and 175 ion channels. We selected at random 175 non-ion channels to create a balanced benchmark dataset with 350 proteins.

The composition of the benchmark dataset is summarized in Table **2**. The benchmark dataset that includes UniProt accession numbers, sequences and annotations is provided in the Supplement.

Table 2. Summary of benchmark dataset.

Protein type	Ion channel type	Voltage-gated channel class	Number of proteins
Ion-channel	Voltage-gated	Sodium (Na^+)	20
		Potassium (K^+)	29
		Calcium (Ca^{2+})	23
		Anion	22
	Ligand-gated		81
Non-ion channel			175

3.1.3. Evaluation of predictive performance:

The predictive performance of the three considered predictors was assessed using a broad selection of eight measures that were used in different combinations in previous related studies [18, 23-25, 36]. The same measures were also used in numerous related studies including prediction of crystallization propensity [33, 53], DNA and RNA binding proteins and residues [54, 55], disordered proteins [56], B-cell epitopes [36], beta-turns [57], secondary structure [58, 59], gamma-carboxylation sites [60], amidation sites [61], and outer membrane proteins [62, 63], to name a few. They include:

Accuracy (ACC) = (TP+TN)/ (TP+FP+TN+FN),

Matthews correlation coefficient (MCC) = (TP*TN+FP*FN)/SQRT((TP+FP)(TP+FN)(TN+FP)(TN+FN))

F1 = 2TP/ (2TP+FN+FP)

Sensitivity (SN) = TP/ (TP+FN)

Specificity (SP) = TN/ (FP+TN)

Precision (PRC) = TP / (TP+FP)

where TP, TN, FP and FN is the number of true positives, true negatives, false positives, and false negatives, respectively. This assessment was performed for each binary prediction (ion-channel vs. non-ion channels, voltage- vs. ligand-gated channels, sodium voltage-gated channel vs. other voltage-gated channels, etc.). Using ion channels vs. non-ion channel as an example, TP is number of native ion channels correctly predicted as ion channels, TN is number of native non-ion channels correctly predicted as non-ion channels, FP is number of native non-ion channels incorrectly predicted as ion channels, and FN is number of native ion channels incorrectly predicted as non-ion channels. We also computed the average accuracy, average MCC, average F1, and Q_4, Q_6 to evaluate the prediction over the four classes of the voltage-gated ion channels, and the six outcomes that are generated in the sequential prediction regime:

$$Q_4 = (TP_1+TP_2+\ldots TP_4)/N$$
$$Q_5 = (TP_1+TP_2+\ldots TP_5)/N$$
$$Q_6 = (TP_1+TP_2+\ldots+TP_6)/N$$

where N is total number of proteins in the benchmark dataset. We also assess statistical significance of the differences in the predictive quality measured with accuracy, MCC and F1 between the most accurate predictor and the other methods. We do not include the other measures (specificity, sensitivity and precision) since they assess only part of the dataset, e.g., only negative or only positive samples. We evaluate these differences when using a range of different test datasets by repeating the tests 10 times where each repetition is based on 50% of randomly chosen benchmark proteins. We use paired *t*-test to compare results over the ten pairs of datasets. We assume that the difference is statistically significant when *p*-value < 0.05.

3.2. Results

Similar to prior works [18, 22-25], we present results for the three types of the parallel predictions. This is followed by the novel assessment of the sequential prediction and the new evaluation for the proteins annotated with one type of ion channel vs. proteins annotated with multiple channel types.

3.2.1. Evaluation of predictive performance:

The predictive quality for the parallel prediction of the ion channels is summarized in Table **3**. The most accurate PSIONplus has achieved accuracy = 0.79, MCC = 0.59 and F1 = 0.80 and it outperforms the other two methods by a statistically significant margin (*p*-value < 0.05). The PSIONplus's MCC value suggests a strong correlation between the predicted and native annotations of ion channels, with the other two methods securing more modest levels of correlation at 0.43 and 0.48. While PSIONplus's sensitivity and specificity values are relatively balanced, IonchanPred2.0 secures the highest sensitivity = 0.87 coupled with relatively low specificity = 0.59. This suggests that IonchanPred2.0 tends to predict a large number of ion channels. The highest specificity = 0.89 is achieved by VGIchan, which reveals that this tool generates relatively few false positives. We also evaluated a random predictor that is implemented to generate the two outcomes at random while maintaining the correct proportion of the ion channels and non-ion channels labels. The three methods provide accuracy, MCC and F1 that are substantially higher than this random predictor. This shows that these methods provide useful predictions of the ion channels.

Table 3. Predictive performance of PSIONplus, IonchanPred2.0, VGIchan and a random predictor for the parallel prediction of ion channels (denoted as positive samples) vs. non-ion channel on the benchmark dataset.

Measure	PSIONplus	IonchanPred2.0	VGIchan	Random[a]
Accuracy	**0.79**	0.73*[b]	0.70*	0.47*
Sensitivity	0.86	**0.87**	0.51	0.47
Specificity	0.73	0.59	**0.89**	0.47
MCC	**0.59**	0.48*	0.43*	-0.05*
Precision	0.76	0.68	**0.82**	0.47
F1	**0.80**	0.76*	0.63*	0.47*

[a]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of the ion channels and non-ion channels labels.

[b]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing PSIONplus is significantly better and is not significantly different at *p*-value of 0.05 than the other method, respectively. The highest value for each specific measure is denoted with the bold font.

3.2.2. Parallel prediction of the voltage-gated and ligand-gated ion channels:

The predictive performance for the parallel prediction of voltage- vs. ligand-gated channels is evaluated in Table **4**. This test excludes VGIchan that does not offer this type of prediction. The most accurate predictions are generated by PSIONplus that secures accuracy = 0.67, MCC = 0.35 and F1 = 0.74. These results are significantly better than the results generated by IonchanPred2.0 and the random predictor (*p*-value < 0.05). In contrast to the prediction of ion channels, the parallel prediction of the ion channel types is characterized by modest levels of predictive quality. This is apparent based on the PSIONplus's MCC and accuracy values. However, these predictions are statistically significantly better than the results of the random predictor. Both PSIONplus and IonchanPred2.0 over-predict the voltage-gated channels. This claim is supported by their high sensitivity and low specificity, where the latter is particularly low for IonchanPred2.0.

Table 4. Predictive performance of PSIONplus, IonchanPred2.0 and a random predictor for the parallel prediction of voltage-gated (denoted as positive samples) vs. ligand-gated channels on a subset of the benchmark dataset with the 175 ion channel proteins.

Measure	PSIONplus	IonchanPred2.0	Random[a]
Accuracy	**0.67**	0.50*[b]	0.50*
Sensitivity	**0.85**	0.84	0.53
Specificity	**0.47**	0.11	0.46
MCC	**0.35**	-0.07*	-0.01*
Precision	**0.65**	0.52	0.53
F1	**0.74**	0.65*	0.53*

[a]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of the voltage- and ligand-gated channel labels.
[b]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing PSIONplus is significantly better and is not significantly different at *p*-value of 0.05 than the other method, respectively. The highest value for each specific measure is denoted with the bold font.

3.2.3. Parallel prediction of the four classes of the voltage-gated ion channels:

Table **5** summarizes results of the empirical assessment of the prediction of the four classes of the voltage-gated ion channels. These tests exclude VGIchan that makes predictions in the sequential way, i.e., it does not predict the voltages-gated channel classes for the sequence that it predicts as the non-ion channels. This prevented us from completing predictions for the entire set of voltage-gated ion channels from the benchmark dataset, i.e., since the non-voltage-gated ion channel outcome is not available for this assessment.

PSIONplus offers the average accuracy = 0.74, average MCC = 0.28, average F1 = 0.39 and Q_4 = 0.48. These are relatively modest levels of predictive quality. However, PSIONplus still statistically significantly outperforms IonchanPred2.0 and the random predictor (*p*-value < 0.05). The trends of the predictive quality over the four classes of channels are visualized in Figure **3**. The predictive quality is modest for the K^+, Ca^{2+} and anion channels (MCC is between 0.29 and 0.48 for PSIONplus) while none of methods can reliably predict the Na^+ channels (MCC = 0). In fact, sensitivity that equals 0 for the prediction of Na^+ channels reveals that they are never correctly predicted. Moreover, we observe that both PSIONplus and IonchanPred2.0 over-predict the K^+ channels. This is based on their high sensitivity and low specificity values. For PSIONplus this also results in lower MCC values compared to the predictions of Ca^{2+} and anion channels (solid green line in Figure **3**). The Q_4 value that quantifies the overall accuracy equals 0.48 for PSIONplus and 0.34 for IonchanPred2.0 when compared to 0.15 for the random predictor. This corresponds to a solid 3.2 and 2.3 folds improvement, respectively.

Table 5. Predictive performance of PSIONplus, IonchanPred2.0 and a random predictor for the parallel prediction of the voltage-gated channel classes on a subset of the benchmark dataset with the 94 voltage-gated ion channel proteins.

Method	Label	ACC[a]	SN[a]	SP[a]	MCC[a]	PRC[a]	F1	Q_4
PSIONplus	K^+	0.56	0.86	0.43	0.29	0.40	0.55	
	Ca^{2+}	0.81	0.61	0.87	0.48	0.61	0.61	
	Anion	0.80	0.27	0.96	0.33	0.67	0.39	
	Na^+	0.79	0.00	1.00	0.00	0.00	0.00	
	Average	0.74	0.44	0.82	0.28	0.42	0.39	0.48
IonchanPred2.0	K^+	0.52	0.69	0.45	0.13	0.36	0.47	
	Ca^{2+}	0.62	0.35	0.70	0.05	0.28	0.31	
	Anion	0.76	0.18	0.93	0.16	0.44	0.26	
	Na^+	0.79	0.00	1.00	0.00	0.00	0.00	
	Average	0.67*[b]	0.31	0.77	0.08*	0.27	0.26*	0.34
Random[c]	K^+	0.51	0.21	0.65	-0.15	0.21	0.21	
	Ca^{2+}	0.64	0.26	0.76	0.02	0.26	0.26	
	Anion	0.57	0.09	0.72	-0.19	0.09	0.09	
	Na^+	0.57	0.00	0.73	-0.27	0.00	0.00	
	Average	0.57*	0.14	0.71	-0.15*	0.14	0.14*	0.15

[a]ACC: accuracy, SN: sensitivity; SP: specificity; MCC: Matthews correlation coefficient; PRC: precision;
[b]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing PSIONplus is significantly better and is not significantly different at *p*-value of 0.05 than the other method, respectively.
[c]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of the four types of the voltage-gated channel classes.

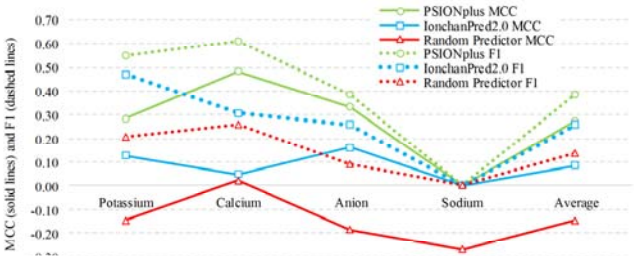

Figure 3. MCC and F1 scores for the prediction of the four voltage-gated ion channel classes for IonchanPred2.0, PSIONplus and the random predictor.

3.2.4. Sequential prediction ion channels, their types and classes of the voltage-gated ion channels:

The sequential prediction of the ion channels, their types and the four classes of the voltage-gated ion channels is summarized in Table **6**. This analysis considers methods that can predict the six outcomes: PSIONplus, IonchanPred2.0 and the random predictor. VGIchan that predicts five of the six outcomes (it does not predict the ligand-gated ion channels) is evaluated in the next paragraph. PSIONplus secures modest levels of predictive performance with the average accuracy = 0.85, average MCC = 0.29, average F1 = 0.35, and Q_6 = 0.55.

This Q_6 is 2.1 times better than the Q_6 of the random predictor. The predictions produced by PSIONplus are statistically significantly better than the predictions with IonchanPred2.0 and the random predictor (p-value < 0.05). However, the predictions from IonchanPred2.0 are also at practical levels with Q_6 = 0.4, which is 1.5 times better than the random predictor.

Table 6. Predictive performance of PSIONplus, IonchanPred2.0 and a random predictor for the sequential prediction of the ion channels, their types and the voltage-gated channel classes in the benchmark dataset.

Method	Label	ACC[a]	SN[a]	SP[a]	MCC[a]	PRC[a]	F1	Q_6
PSIONplus	Ion channel	0.79	0.86	0.73	0.59	0.76	0.80	
	Ligand-gated	0.79	0.40	0.91	0.36	0.58	0.47	
	K^+	0.79	0.86	0.78	0.40	0.26	0.40	
	Ca^{2+}	0.86	0.26	0.91	0.13	0.16	0.20	
	Anion	0.93	0.18	0.98	0.24	0.40	0.25	
	Na^+	0.94	0.00	1.00	0.00	0.00	0.00	
	Average	0.85	0.43	0.88	0.29	0.36	0.35	0.55
IonchanPred2.0	Ion channel	0.73	0.87	0.59	0.48	0.68	0.76	
	Ligand-gated	0.69	0.10	0.87	-0.04	0.19	0.13	
	K^+	0.67	0.66	0.67	0.19	0.15	0.25	
	Ca^{2+}	0.85	0.26	0.90	0.12	0.15	0.19	
	Anion	0.91	0.14	0.96	0.12	0.20	0.16	
	Na^+	0.94	0.00	1.00	-0.01	0.00	0.00	
	Average	0.80*[b]	0.34	0.83	0.14*	0.23	0.25*	0.40
Random[c]	Ion channel	0.45	0.45	0.45	-0.11	0.45	0.45	
	Ligand-gated	0.59	0.12	0.74	-0.14	0.12	0.12	
	K^+	0.85	0.10	0.92	0.02	0.10	0.10	
	Ca^{2+}	0.87	0.00	0.93	-0.07	0.00	0.00	
	Anion	0.88	0.05	0.94	-0.02	0.05	0.05	
	Na^+	0.89	0.00	0.94	-0.06	0.00	0.00	
	Average	0.75*	0.12	0.82	-0.06*	0.12	0.12*	0.26

[a]ACC: accuracy, SN: sensitivity; SP: specificity; MCC: Matthews correlation coefficient; PRC: precision;
[b]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing PSIONplus is significantly better and is not significantly different at p-value of 0.05 than the other method, respectively.
[c]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of annotations of the non-ion channels, ion channels, ion channel types and classes of the voltage-gated ion channels.

Table 7. Predictive performance of VGIchan and a random predictor for the sequential prediction of the ion channels and the voltage-gated channel classes in the benchmark dataset.

Method	Label	ACC[a]	SN[a]	SP[a]	MCC[a]	PRC[a]	F1	Q_5
VGIchan	Ion channel	0.81	0.67	0.89	0.58	0.77	0.72	
	Anion	0.92	0.05	1.00	0.13	0.50	0.08	
	Na^+	0.93	0.00	1.00	0.00	0.00	0.00	
	Ca^{2+}	0.91	0.00	1.00	0.00	0.00	0.00	
	K^+	0.80	0.93	0.78	0.48	0.34	0.50	
	Average	0.87	0.33	0.93	0.24	0.32	0.26	0.68
Random[b]	Ion channel	0.55	0.35	0.65	0.00	0.35	0.35	
	Anion	0.84	0.05	0.92	-0.04	0.05	0.05	
	Na^+	0.87	0.10	0.93	0.03	0.10	0.10	
	Ca^{2+}	0.86	0.17	0.92	0.10	0.17	0.17	
	K^+	0.81	0.14	0.90	0.03	0.14	0.14	
	Average	0.79*[c]	0.16	0.86	0.02*	0.16	0.16*	0.46

[a]ACC: accuracy, SN: sensitivity; SP: specificity; MCC: Matthews correlation coefficient; PRC: precision;
[b]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of annotations of the non-ion channels, ion channels and classes of the voltage-gated ion channels.
[c]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing VGIchan is significantly better and is not significantly different at p-value of 0.05 than the other method, respectively.

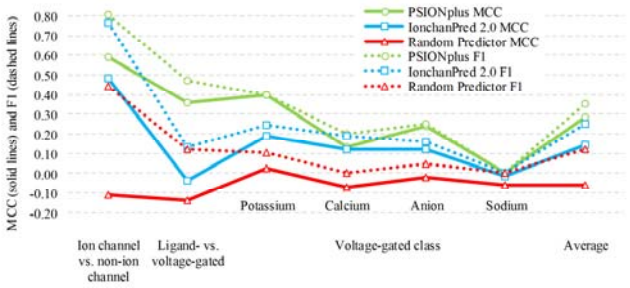

Figure 4. MCC and F1 scores for the sequential prediction with IonchanPred 2.0, PSIONplus and the random predictor.

Figure **4** visualizes the values of MCC and F1 for the three methods over the six outcomes. The lowest predictive quality is for the Na^+ channels (and for the ligand- vs. voltage-gated channel prediction for IonchanPred2.0), while the strongest predictions are observed for the prediction of ion channels. PSIONplus and IonchanPred2.0 fail to correctly predict any of the Na^+ voltage-gated channels (sensitivity = 0 and specificity = 1 in Table **6**). Except for this channel class, the PSIONplus predictions are universally better than the predictions from IonchanPred2.0, while the latter predictor is always better than the random predictions.

We also assess VGIchan and compare it to the random predictor in Table **7**. The overall predictive performance of VGIchan is modest with average MCC = 0.24, average accuracy = 0.87, average F1 = 0.26 and Q_5 = 0.68. VGIchan's Q_5 is 1.5 times better than the Q_5 of the random predictor and

its MCC, accuracy and F1 are significantly better (p-value < 0.05). The predictions for the Na^+ and Ca^{2+} channel classes are inaccurate, with sensitivity = 0 and specificity = 1. This reveal that VGIchan misses to correctly predict any of these two types of voltage-gated channel classes. It also predicts very few of the anion voltage-gated channels (sensitivity = 0.05). The average MCC, average F1 and ratio of Q_5 or Q_6 compared to the random predictor, which are arguably the most suitable to compare between the assessments in Tables 6 and 7 that cover different number of predictive outcomes, are slightly higher for PSIONplus. However, we conclude that both VGIchan and PSIONplus offer predictions that have practical value based on their substantial improvement when contrasted with the random predictor.

3.2.5. Comparison of results for proteins annotated with single vs. multiple ion channel types/classes:

There are 29 proteins with multiple annotations of the channel types/classes in the benchmark dataset. The average annotation count for these proteins is 2.5. They account for 73 samples (21% of the benchmark dataset) while the remaining 277 proteins/samples have a single annotation. We assess the sequential predictions on the proteins that have one annotation (79% of the benchmark dataset, Table **8**) and compare these results with Table **6** where the entire benchmark dataset was used. The results between the complete benchmark dataset and the set of single-annotation proteins are comparable. For instance, the average MCC, average F1, average accuracy and Q_6 values for the most accurate PSIONplus equal 0.29 vs. 0.27, 0.35 vs. 0.31, 0.85 vs. 0.87, and 0.55 vs. 0.61. Also, both PSIONplus and IonchanPred2.0 maintain substantial margin when compared to the random predictor on the benchmark dataset with the single-annotation proteins. Moreover, PSIONplus is statistically significantly better than IonchanPred2.0 (p-value < 0.05). This reveals that the considered predictors perform well on these proteins.

Next, we turn our attention to the 29 multi-annotation proteins. We note that none of the current predictors is equipped to make multiple different predictions for a single protein, and as such they are not capable to make correct predictions for these ion channels. We consider an option of combining predictions of the two methods that cover the entire set of six predictive outcomes (PSIONplus and IonchanPred2.0) to generate multiple predictions for these proteins. Table **9** compares the level of agreement between predictions generated by these two predictors between the proteins with single and multiple annotations. We observe that the two predictors are in agreement for roughly half of the proteins, which is substantially higher than the agreement between random predictors. The latter is expected given the overall significantly higher predictive performance of PSIONplus and IonchanPred2.0 when compared to the random predictors (Tables **6** and **8**). The agreement levels are slightly higher for the multi-annotation proteins. This reveals that combining PSIONplus and IonchanPred2.0 will fail to produce the correct predictions for 59% of the multi-annotation proteins. Table **10** directly evaluates predictive quality of combining these two methods to predict the multiple annotations for the 29 proteins. It reveals that while PSIONplus and IonchanPred2.0 correctly predict at least one of the annotations for 90% and 69% of these proteins, respectively, they correctly predict two annotations for only 10% of the multi-annotation proteins. While these rates of correct predictions are better than the corresponding rates for random predictors (Table **10**), they are short of levels that can find practical applications. Altogether, we conclude that the current tools cannot tackle the prediction of proteins with multiple annotations of ion channel types/classes and that combining these tools together is not going to yield satisfactory results.

Table 8. Predictive performance of PSIONplus, IonchanPred2.0 and a random predictor for the sequential prediction of the ion channels, their types and the voltage-gated channel classes for the proteins with a single annotation in the benchmark dataset.

Method	Label	ACC[a]	SN[a]	SP[a]	MCC[a]	PRC[a]	F1	Q_6
PSIONplus	Ion channel	0.75	0.79	0.73	0.50	0.63	0.70	
	Ligand-gated	0.81	0.44	0.92	0.40	0.60	0.51	
	Anion	0.92	0.19	0.98	0.24	0.40	0.26	
	Na^+	1.00	0.00	1.00	0.00	0.00	0.00	
	Ca^{2+}	0.87	0.27	0.90	0.11	0.10	0.15	
	K^+	0.87	1.00	0.86	0.37	0.16	0.28	
	Average	0.87	0.45	0.90	0.27	0.31	0.31	0.61
IonchanPred2.0	Ion channel	0.66	0.78	0.59	0.36	0.53	0.63	
	Ligand-gated	0.69	0.13	0.86	-0.02	0.21	0.16	
	Anion	0.89	0.10	0.96	0.07	0.15	0.12	
	Na^+	0.99	0.00	1.00	0.00	0.00	0.00	
	Ca^{2+}	0.88	0.18	0.91	0.07	0.08	0.11	
	K^+	0.74	0.57	0.75	0.11	0.06	0.10	
	Average	0.81*[b]	0.29	0.84	0.10*	0.17	0.19*	0.43
Random[c]	Ion channel	0.44	0.44	0.45	-0.11	0.32	0.37	
	Ligand-gated	0.60	0.13	0.74	-0.13	0.13	0.13	
	Anion	0.86	0.05	0.92	-0.03	0.05	0.05	
	Na^+	0.94	0.00	0.94	-0.02	0.00	0.00	
	Ca^{2+}	0.90	0.00	0.94	-0.05	0.00	0.00	
	K^+	0.91	0.29	0.92	0.12	0.09	0.13	
	Average	0.77*	0.15	0.82	-0.04*	0.10	0.11*	0.32

[a]ACC: accuracy, SN: sensitivity; SP: specificity; MCC: Matthews correlation coefficient; PRC: precision;

[b]* and = denote that predictive performance measured with accuracy, MCC and F1 of the best performing PSIONplus is significantly better and is not significantly different at p-value of 0.05 than the other method, respectively.

[c]The random predictor is based on a randomized assignment of the outcomes that maintains the correct proportion of annotations of the non-ion channels, ion channels, ion channel types and classes of the voltage-gated ion channels.

Table 9. Fraction of proteins for which predictions are in agreement for a pair of two random predictors and the pair of PSIONplus and IonchanPred2.0 predictors. The evaluation is performed for the single annotation and multi-annotation proteins in the benchmark dataset.

	PSIONplus and IonchanPred2.0	Two random predictors
Single-annotation proteins	0.48	0.32
Multi-annotation proteins	0.59	0.34

[a]The random predictors are based on a randomized assignment of the outcomes that maintains the correct proportion of annotations of the non-ion channels, ion channels, ion channel types and classes of the voltage-gated ion channels.

Table 10. Predictive performance of a combination of PSIONplus and IonchanPred2.0 and a combination of two random predictors for the 29 proteins that have multiple annotations of channel types/classes.

	Ratio of predictions where at least one outcome was correctly predicted
PSIONplus alone	0.90
IonchanPred2.0 alone	0.69
PSIONplus and IonchanPred2.0 combined	0.93
Random predictor 1 alone	0.28
Random predictor 2 alone	0.31
Two random predictors combined	0.48
	Ratio of predictions where at least two outcomes were correctly predicted
PSIONplus and IonchanPred2.0 combined	0.10
Two random predictors combined	0.03

4. DISCUSSION AND CONCLUSION

Ion channels play an important role in a wide range of cellular processes. Hundreds of ion channel were already annotated and many more await discovery. This protein family is associated with over two dozen human diseases [2]and is targeted by over 700 drugs [15]. Thus, the development of accurate computational tools that predict the ion channels from protein sequences is desirable.

We describe details of a comprehensive set of eight predictors of ion channels, three of which focus on the prediction of the subfamilies of K^+ channels, and five that cover the prediction of ion channels, their types and four classes of the voltage-gated channels. The latter five methods offer a comprehensive solution for this predictive task. They are characterized by diverse designs and only three of these tools (IonchanPred2.0 [25], VGIchan [18] and PSIONplus [22]) are available to the end user as either webservers or standalone software. The lack of support and availability for the majority of the published predictors is a major problem. This should be tackled by requiring the provision and support of the implementation/software at the time when the predictors are published. This is already required to publish in some venues, such as the *Nucleic Acids Research* journal.

Empirical comparative assessment of the three currently available and comprehensive predictors on a novel and carefully crafted benchmark dataset reveals several interesting observations. First, the three computational tools offer strong predictive performance for the prediction of ion channels, with PSIONplus providing the most accurate results. Second, performance for the prediction of ion channel types (voltage- vs. ligand gated) is modest and only PSIONplus should be used for this purpose. The three tools provide modest levels of predictive quality for the prediction of the four classes of the voltage-gated ion channels, where PSIONplus takes a slight lead ahead of VGIchan. However, none of these tools accurately predicts sodium (Na^+) channel class. First-of-its-kind assessment of the sequential prediction of the ion channels, their types and the four classes of the voltage-gated channels reveals that PSIONplus is slightly more accurate than VGIchan (the latter also does not predict the ligand-gated ion channels) and statistically significantly more accurate than IonchanPred2.0. We speculate that PSIONplus is the most accurate because it uses a more diverse and broader range of six types of inputs when compared to the other two methods, particularly when contrasted to IonchanPred2.0 that uses only two types of inputs (Table **1**). Moreover, the three methods are significantly more accurate than a random predictor, showing that end users would benefit from using these tools.

We are also the first to investigate prediction for proteins annotated with multiple type/classes of ion channels. The design of each of the current predictors precludes it from generating more than one prediction for a given input protein sequence. Combining outputs generated by multiple tools does not offer a feasible solution either. This is because they often produce the same prediction and because their combined predictions are inaccurate when they differ. New methods that can solve these problems are needed. One solution is motivated by the prediction of ligand-binding residues in proteins where some of the residues may interact with multiple types of ligands, such as DNA, RNA and proteins [35, 64, 65]. In this case specific methods are used to predict residues that bind to one type of ligand [54, 55, 66-69], and these methods are combined together to effectively annotate multiple types of ligands for the same residue [68]. In the context of this study, this would require building and combining multiple methods that would address prediction of specific types and classes of ion channels. The second option is to design one predictor that would rely on multi-label classification algorithms that directly produce multiple outcomes for the same input protein [70]. Such multi-labels predictors were recently designed to address prediction of protein and gene functions [71-73] and subcellular locations [74-76].

The results for the three parallel tests (ion channels, their types, and classes of voltage-gated channels) for VGIchan and IonchanPred2.0 are lower than these reported in [18] and [25], respectively. There are three potential reasons. First, we use a stricter similarity cut-off between training and benchmark

sequences at 30% vs. 90% [18] and 40% [25]. This results in a harder and arguably more practical assessment. Second, we use a benchmark dataset that was not used to develop/parametrize these tools. This is in contrast to the prior works that develop and evaluate their tools using cross validation [18] and jackknife tests [25] on the training datasets. Both of these test, and in particular the jackknife test, allow the predictive model to draw on the sequence similarity in the training dataset to secure higher levels of predictive performance. Third, VGIchan was evaluated using a dataset with voltage-gated ion channels and the negative sequences that include all other proteins [18], instead of the arguably harder set of negatives used in this review and in [22, 25]that includes other types of membrane proteins. Importantly, the results reported here for the prediction of ion channels and their types are consistent with results in [22] where a similar test protocol was used, i.e., test dataset with 30% similarity and membrane proteins as negatives. The results for the parallel prediction of the voltage-gated ion channel classes are, as expected, lower here compared to[22] where the corresponding test dataset shared much higher sequence similarity at 60% to the training proteins. The use of this high-similarity test set was a result of a data shortage at the time when [22]was published.

LIST OF ABBREVIATIONS

NA: not applicable; WS: webserver; SS: Standalone software; K^+: potassium; Ca^{2+}: calcium; Na^+: sodium; Cl^-: Chloride; AAC: amino acid composition; DPC: dipeptide composition; TPC: tripeptide composition; PP: physicochemical properties; RSA: relative solvent accessibility; HMM: hidden Markov model; PSSM: position specific scoring matrix; RF: random forest; SVM: support vector machine; SQRT: square root; ACC: accuracy; SN: sensitivity; SP: specificity; MCC: Matthews correlation coefficient; PRC: precision;

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work has been supported in part by the National Natural Science Foundation of China (NSFC) grant 11701296, the Natural Science Foundation of Tianjin grant 18JCQNJC09600, and the Fundamental Research Funds for the Central Universities to J.G. This work was also supported by the NSFC grant 61832019 and the Qimonda Endowment funds and to L.K.

SUPPLEMENTARY MATERIAL

Supplement.txt: benchmark dataset used in this article.

REFERENCES

[1] Domene, C., S. Haider, and M.S.P. Sansom, *Ion channel structures: a review of recent progress.* Current Opinion in Drug Discovery & Development, 2003. **6**(5): p. 611-619.

[2] Bagal, S.K., et al., *Ion channels as therapeutic targets: a drug discovery perspective.* J Med Chem, 2013. **56**(3): p. 593-624.

[3] Ger, M.F., et al., *Domain-Based Identification and Analysis of Glutamate Receptor Ion Channels and Their Relatives in Prokaryotes.* Plos One, 2010. **5**(10): p. e12827.

[4] Tabassum, N. and F. Ahmed, *Ion Channels and their Modulation.* European Journal of Pharmaceutical Sciences, 2011. **1**(1): p. 20-25.

[5] Bech-Hansen, N.T., et al., *Loss-of-function mutations in a calcium-channel alpha1-subunit gene in Xp11.23 cause incomplete X-linked congenital stationary night blindness.* Nat Genet, 1998. **19**(3): p. 264-7.

[6] Jentsch, T.J., *Neuronal KCNQ potassium channels: physiology and role in disease.* Nat Rev Neurosci, 2000. **1**(1): p. 21-30.

[7] Peters, D.J., et al., *Chromosome 4 localization of a second gene for autosomal dominant polycystic kidney disease.* Nat Genet, 1993. **5**(4): p. 359-62.

[8] Curran, M.E., et al., *A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome.* Cell, 1995. **80**(5): p. 795-803.

[9] Wang, Q., et al., *SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome.* Cell, 1995. **80**(5): p. 805-11.

[10] Lafreniere, R.G., et al., *A dominant-negative mutation in the TRESK potassium channel is linked to familial migraine with aura.* Nature Medicine, 2010. **16**(10): p. 1157-U1501.

[11] Kaczorowski, G.J., et al., *Ion channels as drug targets: The next GPCRs.* Journal of General Physiology, 2008. **131**(5): p. 399-405.

[12] Sheu, S.-S. and W. Lederer, *Lidocaine's negative inotropic and antiarrhythmic actions. Dependence on shortening of action potential duration and reduction of intracellular sodium activity.* Circulation research, 1985. **57**(4): p. 578-590.

[13] Skov, M.J., et al., *Nonclinical safety of ziconotide: an intrathecal analgesic of a new pharmaceutical class.* International journal of toxicology, 2007. **26**(5): p. 411-421.

[14] Schmidtko, A., et al., *Ziconotide for treatment of severe chronic pain.* The Lancet, 2010. **375**(9725): p. 1569-1577.

[15] Santos, R., et al., *A comprehensive map of molecular drug targets.* Nature Reviews Drug Discovery, 2017. **16**(1): p. 19.

[16] Gabashvili, I.S., et al., *Ion Channel Gene Expression in the Inner Ear.* Journal of the Association for Research in Otolaryngology, 2007. **8**(3): p. 305-328.

[17] Consortium, U., *UniProt: the universal protein knowledgebase.* Nucleic acids research, 2018. **46**(5): p. 2699.

[18] Saha, S., et al., *VGIchan: prediction and classification of voltage-gated ion channels.* Genomics, proteomics & bioinformatics, 2006. **4**(4): p. 253-258.

[19] Chen, W. and H. Lin, *Identification of voltage-gated potassium channel subfamilies from sequence information using support vector machine.* Computers in biology and medicine, 2012. **42**(4): p. 504-507.

[20] Liu, L.X., et al., *Local Sequence Information‐based Support Vector Machine to Classify Voltage‐gated Potassium Channels.* Acta biochimica et biophysica Sinica, 2006. **38**(6): p. 363-371.

[21] Liu, W., et al., *Identifying the subfamilies of voltage-gated potassium channels using feature selection technique.* International Journal of Molecular Sciences, 2014. **15**(7): p. 12940-12951.

[22] Gao, J., et al., *PSIONplus: Accurate Sequence-Based Predictor of Ion Channels and Their Types.* PLOS ONE, 2016. **11**(4).

[23] Lin, H. and H. Ding, *Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition.* Journal of theoretical biology, 2011. **269**(1): p. 64-69.

[24] Tiwari, A.K. and R. Srivastava, *An efficient approach for the prediction of ion channels and their subfamilies.* Computational biology and chemistry, 2015. **58**: p. 205-221.

[25] Zhao, Y.W., et al., *IonchanPred 2.0: A Tool to Predict Ion Channels and Their Types.* International Journal of Molecular Sciences, 2017. **18**(9).

[26] Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

[27] Cao, R. and J. Cheng, *Integrated protein function prediction by mining function associations, sequences, and protein-protein and gene-gene interaction networks.* Methods, 2016. **93**: p. 84-91.

[28] Cao, R. and J. Cheng, *Protein single-model quality assessment by feature-based probability density functions.* Sci Rep, 2016. **6**: p. 23990.

[29] Cao, R., et al., *ProLanGO: Protein Function Prediction Using Neural Machine Translation Based on a Recurrent Neural Network.* Molecules, 2017. **22**(10).

[30] Cao, R., et al., *SMOQ: a tool for predicting the absolute residue-specific quality of a single protein model with support vector machines.* BMC Bioinformatics, 2014. **15**: p. 120.

[31] Meng, F., V.N. Uversky, and L. Kurgan, *Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions.* Cell Mol Life Sci, 2017. **74**(17): p. 3069-3090.

[32] Hayat, M. and A. Khan, *Mem-PHybrid: Hybrid features-based prediction system for classifying membrane protein types.* Analytical Biochemistry, 2012. **424**(1): p. 35-44.

[33] Meng, F., C. Wang, and L. Kurgan, *fDETECT webserver: fast predictor of propensity for protein production, purification, and crystallization.* BMC Bioinformatics, 2018. **18**(1): p. 580.

[34] Mishra, N.K., J. Chang, and P.X. Zhao, *Prediction of Membrane Transport Proteins and Their Substrate Specificities Using Primary Sequence Information.* PLOS ONE, 2014. **9**(6): p. e100278.

[35] Peng, Z. and L. Kurgan, *High-throughput prediction of RNA, DNA and protein binding regions mediated by intrinsic disorder.* Nucleic Acids Research, 2015. **43**(18): p. e121-e121.

[36] Gao, J., et al., *BEST: improved prediction of B-cell epitopes from antigen sequences.* PLoS One, 2012. **7**(6): p. e40104.

[37] Xianfang, W., et al., *Predicting the Types of Ion Channel-Targeted Conotoxins Based on AVC-SVM Model %J BioMed Research International.* 2017. **2017**: p. 8.

[38] Nugent, T. and D.T. Jones, *Detecting pore-lining regions in transmembrane protein sequences.* BMC Bioinformatics, 2012. **13**: p. 169-169.

[39] Zheng, C. and L. Kurgan, *Prediction of beta-turns at over 80% accuracy based on an ensemble of predicted secondary structures and multiple alignments.* BMC Bioinformatics, 2008. **9**: p. 430-430.

[40] Yan, J., M. Marcus, and L. Kurgan, *Comprehensively designed consensus of standalone secondary structure predictors improves Q3 by over 3%.* Journal of Biomolecular Structure and Dynamics, 2014. **32**(1): p. 36-51.

[41] Yan, J., et al., *RAPID: Fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(8): p. 1671-1680.

[42] Kumar, R., B. Kumari, and M. Kumar, *Proteome-wide prediction and annotation of mitochondrial and sub-mitochondrial proteins by incorporating domain information.* Mitochondrion, 2018. **42**: p. 11-22.

[43] Hayat, S. and A. Elofsson, *BOCTOPUS: improved topology prediction of transmembrane β barrel proteins.* Bioinformatics, 2012. **28**(4): p. 516-522.

[44] Disfani, F.M., et al., *MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins.* Bioinformatics, 2012. **28**(12): p. i75-i83.

[45] Zhang, T., et al., *Accurate sequence-based prediction of catalytic residues.* Bioinformatics, 2008. **24**(20): p. 2329-2338.

[46] Kurgan, L., K. Cios, and K. Chen, *SCPRED: Accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences.* BMC Bioinformatics, 2008. **9**: p. 226-226.

[47] Chen, K., M.J. Mizianty, and L. Kurgan, *ATPsite: sequence-based prediction of ATP-binding residues.* Proteome Science, 2011. **9**(Suppl 1): p. S4-S4.

[48] Cao, R., et al., *DeepQA: improving the estimation of single protein model quality with deep belief networks.* BMC Bioinformatics, 2016. **17**(1): p. 495.

[49] Gao, J., Y. Yang, and Y. Zhou, *Grid-based prediction of torsion angle probabilities of protein backbone and its application to discrimination of*

protein intrinsic disorder regions and selection of model structures. BMC Bioinformatics, 2018. **19**(1): p. 29.

[50] Gao, J., Y. Yang, and Y. Zhou, *Predicting the errors of predicted local backbone angles and non-local solvent- accessibilities of proteins by deep neural networks.* Bioinformatics, 2016. **32**(24): p. 3768-3773.

[51] Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.* Bioinformatics, 2012. **28**(23): p. 3150-2.

[52] Huang, Y., et al., *CD-HIT Suite: a web server for clustering and comparing biological sequences.* Bioinformatics, 2010. **26**(5): p. 680-2.

[53] Wang, H., et al., *Critical evaluation of bioinformatics tools for the prediction of protein crystallization propensity.* Briefings in Bioinformatics, 2017. **18**(6): p. 1092-1092.

[54] Yan, J., S. Friedrich, and L. Kurgan, *A comprehensive comparative review of sequence-based predictors of DNA- and RNA-binding residues.* Brief Bioinform, 2016. **17**(1): p. 88-105.

[55] Zhao, H., Y. Yang, and Y. Zhou, *Prediction of RNA binding proteins comes of age from low resolution to high resolution.* Mol Biosyst, 2013. **9**(10): p. 2417-25.

[56] Peng, Z., M.J. Mizianty, and L. Kurgan, *Genome-scale prediction of proteins with long intrinsically disordered regions.* Proteins, 2014. **82**(1): p. 145-58.

[57] Zheng, C. and L. Kurgan, *Prediction of beta-turns at over 80% accuracy based on an ensemble of predicted secondary structures and multiple alignments.* BMC Bioinformatics, 2008. **9**: p. 430.

[58] Jiang, Q., et al., *Protein secondary structure prediction: A survey of the state of the art.* J Mol Graph Model, 2017. **76**: p. 379-402.

[59] Zhang, H., et al., *Critical assessment of high-throughput standalone methods for secondary structure prediction.* Brief Bioinform, 2011. **12**(6): p. 672-88.

[60] Gao, J., N. Zhang, and J. Ruan, *Prediction of protein modification sites of gamma-carboxylation using position specific scoring matrices based evolutionary information.* Comput Biol Chem, 2013. **47**: p. 215-20.

[61] Wang, T., et al., *PrAS: Prediction of amidation sites using multiple feature extraction.* Comput Biol Chem, 2017. **66**: p. 57-62.

[62] Mizianty, M.J. and L. Kurgan, *Improved identification of outer membrane beta barrel proteins using primary sequence, predicted secondary structure, and evolutionary information.* Proteins, 2011. **79**(1): p. 294-303.

[63] Tsaousis, G.N., S.J. Hamodrakas, and P.G. Bagos, *Predicting Beta Barrel Transmembrane Proteins Using HMMs.* Hidden Markov Models: Methods and Protocols, 2017. **1552**: p. 43-61.

[64] Miao, Z. and E. Westhof, *A Large-Scale Assessment of Nucleic Acids Binding Site Prediction Programs.* PLoS Comput Biol, 2015. **11**(12): p. e1004639.

[65] Zhang, J., Z. Ma, and L. Kurgan, *Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains.* Brief Bioinform, 2017.

[66] Ding, X.M., et al., *Computational prediction of DNA-protein interactions: a review.* Curr Comput Aided Drug Des, 2010. **6**(3): p. 197-206.

[67] Walia, R.R., et al., *Sequence-Based Prediction of RNA-Binding Residues in Proteins.* Prediction of Protein Secondary Structure, 2017. **1484**: p. 205-235.

[68] Yan, J. and L. Kurgan, *DRNApred, fast sequence-based method that accurately predicts and discriminates DNA- and RNA-binding residues.* Nucleic Acids Res, 2017. **45**(10): p. e84.

[69] Zhang, J. and L. Kurgan, *Review and comparative assessment of sequence-based predictors of protein-binding residues.* Brief Bioinform, 2017.

[70] Zhang, M.L. and Z.H. Zhou, *A Review on Multi-Label Learning Algorithms.* Ieee Transactions on Knowledge and Data Engineering, 2014. **26**(8): p. 1819-1837.

[71] Cerri, R., et al., *Reduction strategies for hierarchical multi-label classification in protein function prediction.* BMC Bioinformatics, 2016. **17**: p. 373.

[72] Wan, S., M.-W. Mak, and S.-Y. Kung, *Mem-ADSVM: A two-layer multi-label predictor for identifying multi-functional types of membrane proteins.* Journal of Theoretical Biology, 2016. **398**: p. 32-42.

[73] Stojanova, D., et al., *Using PPI network autocorrelation in hierarchical multi-label classification trees for gene function prediction.* BMC Bioinformatics, 2013. **14**: p. 285-285.

[74] Guo, X., et al., *Human Protein Subcellular Localization with Integrated Source and Multi-label Ensemble Classifier.* Scientific Reports, 2016. **6**: p. 28087.

[75] Xu, Y.-Y., F. Yang, and H.-B. Shen, *Incorporating organelle correlations into semi-supervised learning for protein subcellular localization prediction.* Bioinformatics, 2016. **32**(14): p. 2184-2192.

[76] Wan, S., Y. Duan, and Q. Zou, *HPSLPred: An Ensemble Multi-Label Classifier for Human Protein Subcellular Location Prediction with Imbalanced Source.* 2017. **17**(17-18): p. 1700262.

Received: August 16, 2018 Revised: August 16, 2018 Accepted: August 16, 2018