

Федеральное государственное бюджетное научное учреждение
**«Научно-исследовательский институт
биомедицинской химии
имени В.Н. Ореховича»**

Компьютерный прогноз аритмогенности лекарственных соединений на основе анализа молекулярных сетей

- СУХАЧЁВ ВЛАДИСЛАВ СЕРГЕЕВИЧ
Лаборант-исследователь ИБМХ, студент РТУ МИРЭА
e-mail: withstanding@yandex.ru
- ИВАНОВ СЕРГЕЙ МИХАЙЛОВИЧ
Научный сотрудник ИБМХ, к.б.н

Актуальность работы

- Побочным эффектом приема некоторых лекарственных веществ является желудочковая тахикардия, которая может привести к летальному исходу.
- Оценка аритмогенности лекарств должна проводиться на наиболее ранних этапах их разработки. (*Seth I. Berger et al. Systems Pharmacology of Arrhythmias , Sci Signal.; 3(118):ra30. DOI: 10.1126/scisignal.2000723.*)
- Соответствующая оценка может быть выполнена при помощи построения связей «структура-активность».
- Применение методов сетевой фармакологии к данным по мишеням лекарств может быть также использовано для оценки аритмогенности.
- Комбинирование методов построения связей «структура-активность» и методов сетевой фармакологии может повысить точность прогноза желудочковой тахикардии. (*Liang-Chin Huang et al., Predicting adverse drug reaction profiles by integrating protein interaction networks with drug structures, DOI: 10.1002/pmic.201200337*)

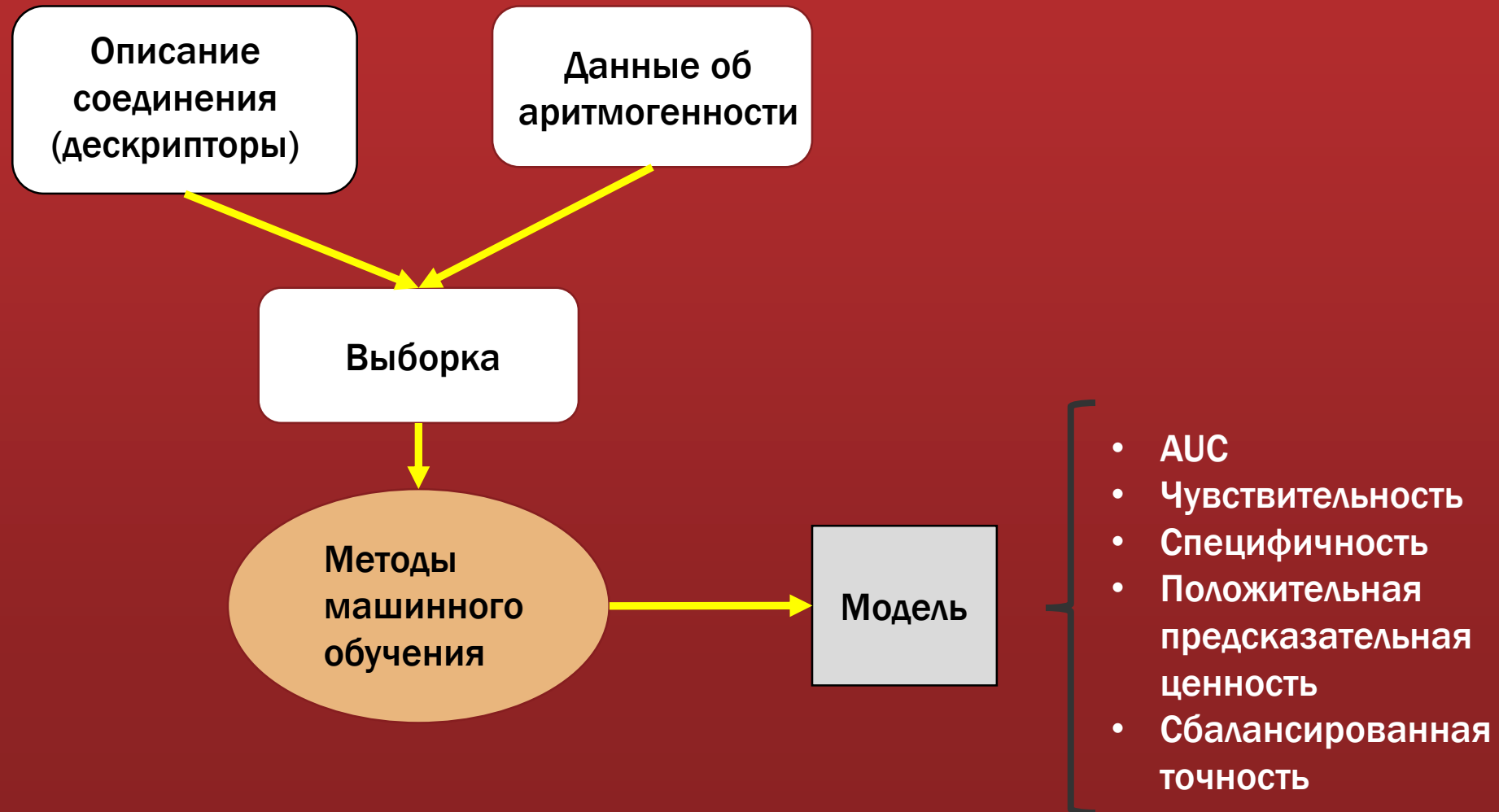
ЦЕЛЬ РАБОТЫ:

ПОСТРОИТЬ МОДЕЛЬ ДЛЯ ПРЕДСКАЗАНИЯ АРИТМОГЕННОСТИ ЛЕКАРСТВЕННЫХ СОЕДИНЕНИЙ ПУТЕМ КОМБИНИРОВАНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И АНАЛИЗА МОЛЕКУЛЯРНЫХ СЕТЕЙ.

ЗАДАЧИ ИССЛЕДОВАНИЯ:

- I. Построить модель для прогноза желудочковой тахикардии на основе связи «структура-активность» с использованием информации об известных и предсказанных белках-мишенях лекарств в качестве дескрипторов.
- II. Построить модель для прогноза желудочковой тахикардии на основе использования сети регуляции функций ионных каналов.
- III. Выполнить оценку точности прогноза комбинацией двух методов.

Прогноз желудочковой тахикардии на основе связи «структура-активность»



Дескрипторы

При проведении анализа методами машинного обучения, алгоритм получает на вход описания молекул (дескрипторы) и данные об аритмогенности лекарственных соединений.

В текущей работе были задействованы два типа дескрипторов:

Бинарные

- Представлены экспериментальными данными о связывании лекарственных соединений с белками человека.
- Качественные (вещество или связывается, или нет)
- Представляют собой бинарные значения (“1” и “0”).

Результаты прогноза PASS Targets

- Программа доступна в виде локальной версии и веб-сервиса (<http://www.way2drug.com/passtargets/>)
- Позволяет осуществлять прогноз взаимодействия с 1553 белками-мишенями человека.
- Для каждого соединения по каждой мишени рассчитываются оценки P_a и P_i – вероятности того, что соединение взаимодействует или не взаимодействует с мишенью.

Pogodin P.V. et al., SAR QSAR Environ Res. 2015, 26(10):783-93.

Данные об аритмогенности и методы машинного обучения

Данные об аритмогенности

- Взяты из научной статьи (Ivanov SM, Lagunin AA, Filimonov DA, Poroikov VV. Computer prediction of adverse drug effects on the cardiovascular system. *Pharmaceutical Chemistry Journal*. 2018; 52(9): 758–762. doi: 10.1007/s11094-018-1895-1).
- Среди лекарств были отобраны только те, которые вызывают или не вызывают желудочковую тахикардию с высокой степенью достоверности.

Метод машинного обучения

- В данной работе использован метод Random Forest, реализованный на языке R.
- Оценка точности прогноза желудочковой тахикардии была выполнена при помощи процедуры десятикратного скользящего контроля.

PASS Targets

Дескрипторы представлены разностями оценок вероятностей взаимодействия (P_a) и отсутствия взаимодействия (P_i) лекарственного вещества (Лек) с биологическими мишенями (M).

Вызывает аритмию ?	Лекарство	$P(a)-P(i)$							
		M1	M2	M3	M4	M5	M6	M7	M8
да	Лек1	-0,389	0,735	0,775	-0,345	-0,236	-0,457	-0,556	0,556
нет	Лек2	0,486	-0,375	-0,357	-0,357	-0,556	0,663	-0,421	0,856
да	Лек3	0,543	-0,267	-0,435	0,734	-0,556	-0,556	0,156	0,567

Комбинирование дескрипторов

Экспериментальные данные о взаимодействиях лекарств с мишенями могут быть неполны, поэтому они были дополнены результатами прогноза PASS Targets. Это позволило получить комбинированные дескрипторы (экспериментальные данные дополненные прогнозом).

Вызывает аритмию ?	белок Лекарство	STUB1	KCNH2	CACNA1C	SUM01	ITPR2
Да	DB01	+	???	+	-	+
нет	DB02	-	-	???	+	-
...	???



Вызывает аритмию ?	Р(a)-P(i) Лекарство	M1	M2	M3	M4
Да	Лек1	-0,3	-0,735	-0,775	-0,345
нет	Лек2	0,486	-0,375	-0,357	-0,357

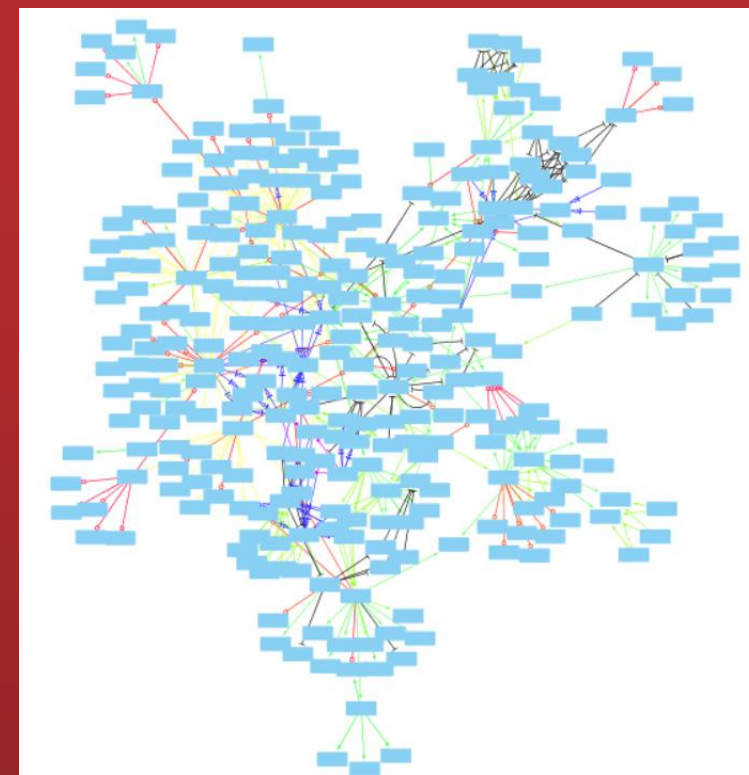
Вызывает аритмию ?	белок Лекарство	STUB1	KCNH2	CACNA1C	SUM01	ITPR2
Да	DB01	+	0,735	+	-	+
нет	DB02	-	-	-0,775	+	-
...	0,486

Точность прогноза желудочковой тахикардии при помощи Random Forest с использованием комбинированных дескрипторов

AUC	Чувствительность	Специфичность	Положительная предсказательная ценность	Сбалансированная точность
0.795	0.731	0.717	0.506	0.724
Площадь под ROC кривой (англ. Area under the ROC curve, AUC)	Доля аритмогенных соединений, которые были предсказаны как аритмогенные	Доля неаритмогенных соединений, которые были предсказаны как неаритмогенные	Доля аритмогенных соединений, среди соединений, которые были предсказаны как аритмогенные	Среднее значение между чувствительностью и специфичностью




Сеть регуляции функции ионных каналов в сердце

Блокада ионных каналов лекарственными соединениями вносит наибольший вклад в развитие желудочковой тахикардии, поскольку именно эти клеточные структуры отвечают за электростатический потенциал мембраны, который приводит к сокращению сердечной мышцы. Функционирование ионных каналов регулируется посредством посттрансляционных модификаций: например фосфорилирование, убиквитинирование.

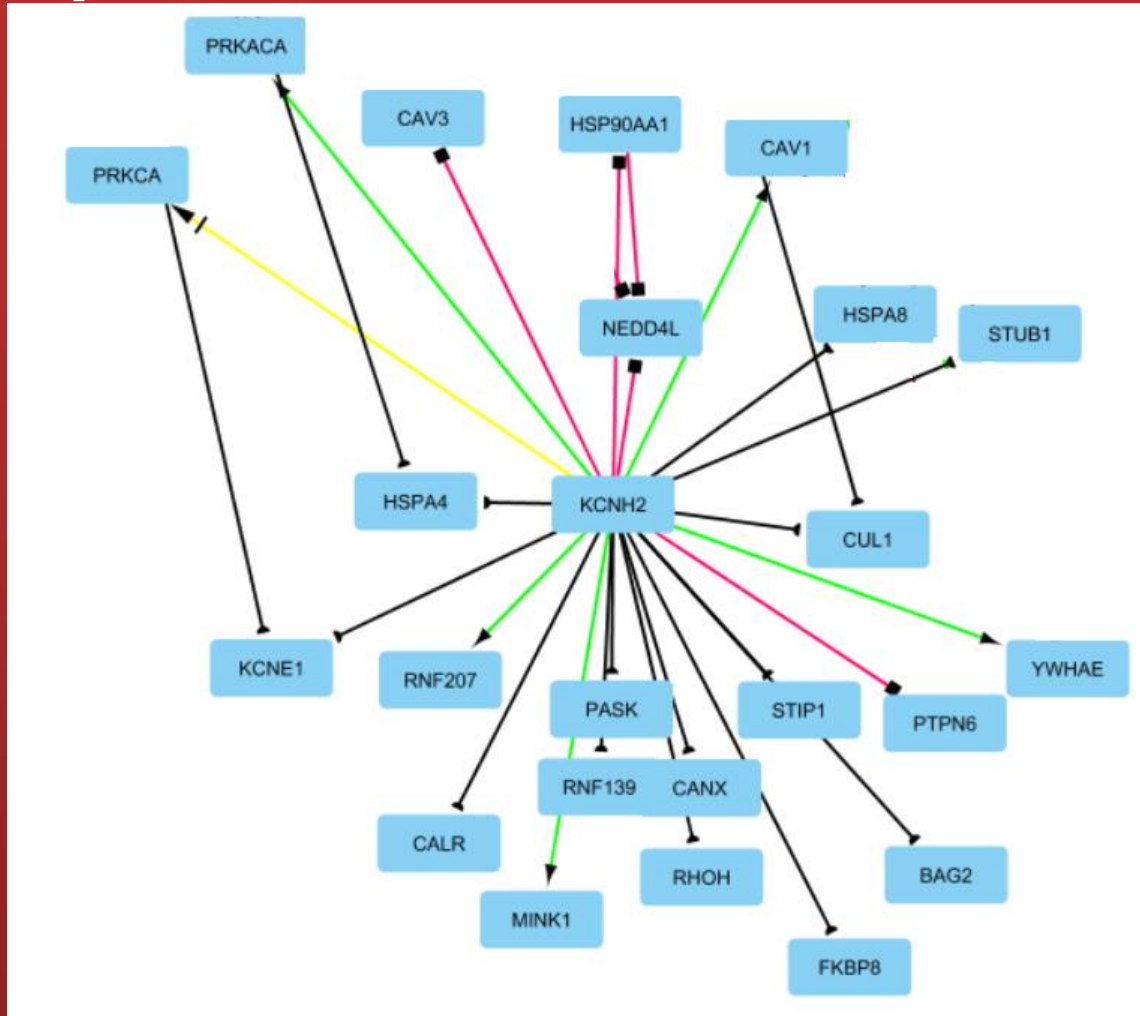


Лекарственные соединения могут вызывать желудочковую тахикардию, ингибируя ферменты, модифицирующие ионные каналы: протеин киназы, фосфатазы, убиквитин лигазы, белки регулирующие активность этих ферментов, и так далее, вплоть до рецепторов.

Информация о белок-белковых взаимодействиях для построения сети была получена из следующих баз данных:

	HIPPIE (Human Integrated Protein-Protein Interaction rEference)	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/	Открытая база данных белок-белковых взаимодействий
	KEGG (Kyoto Encyclopedia of Genes and Genomes)	https://www.genome.jp/kegg/	База данных геномов и генов от института Киото, включающая данные по сигнальным и регуляторным путям
Signor 2.0	Signor 2.0	https://www.signor.org	База данных по белок-белковым взаимодействиям с информацией об их функциональной роли
	PhosphositePlus	https://www.phosphosite.org/	База данных о сайтах фосфорилирования белков

Фрагмент сети



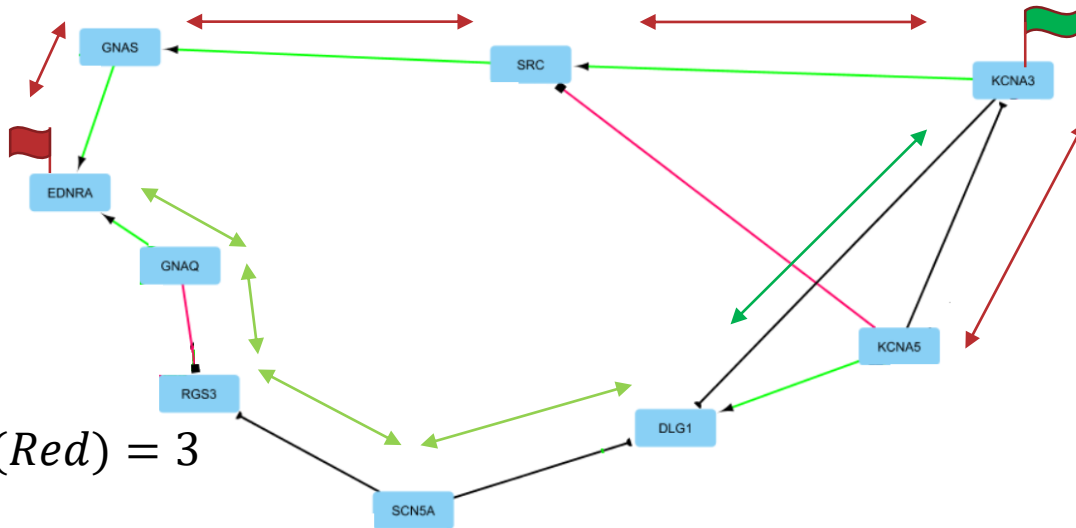
Зелёными стрелками указано положительное воздействие, красными отрицательное (ингибирование). Связывание выделено жёлтым цветом, фосфорилирование и дефосфорилирование синим и фиолетовым, соответственно.

Полученная сеть состоит из 250 узлов, являющихся белками и 484 рёбер, являющихся белок-белковыми взаимодействиями.

Слева представлен калиевый канал KCNH2 (HERG) и те белки, что взаимодействуют с ним непосредственно, а именно:

Название	Функция
PRKCA	Протеинкиназа
KCNE1	Ионный канал
HSP90AA1	Шаперон
CANX	Связывание ионов кальция

Построение прогностической модели на основе сети регуляции ионных каналов



$$\sum d(\text{Red}) = 3$$

$$\sum d(\text{Green}) = 5$$

Оценка аритмогенности

$$= \sum_i^N \sum_j^M \frac{1}{1 + \left(\frac{d(i,j)}{k}\right)^2}$$

Где i, j – вершины сети (т.е белки), между которыми рассматривается связь.

$d(i,j)$ – топографическое расстояние между вершинами i и j .

M – общее число ионных каналов

N – общее число белков в сети

Точность прогноза желудочковой тахикардии на основе анализа сети регуляции ионных каналов

AUC	Чувствительность	Специфичность	Положительная предсказательная ценность	Сбалансированная точность
0.733	0.674	0.678	0.450	0.676
Площадь под ROC кривой (англ. Area under the ROC curve, AUC)	Доля аритмогенных соединений, которые были предсказаны как аритмогенные	Доля неаритмогенных соединений, которые были предсказаны как неаритмогенные	Доля аритмогенных соединений, среди соединений, которые были предсказаны как аритмогенные	Среднее значение между чувствительностью и специфичностью

Точность прогноза желудочковой тахикардии на основе комбинации анализа сети регуляции ионных каналов и Random Forest

AUC	Чувствительность	Специфичность	Положительная предсказательная ценность	Сбалансированная точность
0.802	0.720	0.734	0.516	0.727
Площадь под ROC кривой (англ. Area under the ROC curve, AUC)	Доля аритмогенных соединений, которые были предсказаны как аритмогенные	Доля неаритмогенных соединений, которые были предсказаны как неаритмогенные	Доля аритмогенных соединений, среди соединений, которые были предсказаны как аритмогенные	Среднее значение между чувствительностью и специфичностью

Сравнение полученных результатов

показатели метод	AUC	Чувствитель- ность	Специфич- ность	Положительная предсказатель- ная ценность	Сбалансирован- ная точность
Связь «структура- активность»	0.795	0.731	0.717	0.506	0.724
Анализ сети	0.733	0.674	0.678	0.450	0.676
Комбинирование сетевых методов и анализа «структура-активность»	0.802	0.720	0.734	0.516	0.727

Выводы

- Построена модель для прогноза желудочковой тахикардии на основе связи «структура-активность» с использованием информации об известных и предсказанных белках-мишенях лекарств в качестве дескрипторов и Random Forest. Точность прогноза желудочковой тахикардии, вычисленная при помощи десятикратного скользящего контроля, составила 0.724.
- Построена модель для прогноза желудочковой тахикардии на основе использования сети регуляции функций ионных каналов. Ввиду сложности организации клетки и механизмов аритмогенности, точность прогноза, полученная с помощью анализа сети, оказалась более низкой (0.676), чем при использовании машинного обучения (0.724).
- Построенная молекулярная сеть позволяет наблюдать биохимические процессы, приводящие к возникновению желудочковой тахикардии, в то время как использование методов машинного обучения, работающих по принципу «черного ящика», включая Random Forest, не позволяет проводить оценку механизмов индукции желудочковой тахикардии.

Контактная информация

Сухачёв Владислав Сергеевич.

Моб.тел.: +7 (904)-419-42-73

e-mail: withstanding@yandex.ru