

Ефремов Р.Г.: Часть описанных в Выводах результатов не представлена в ходе доклада. В частности, что из себя представляют «спрогнозированы возможные вирусные мишени» и как именно они были получены? Как это согласуется с известными данными?

Никитина А.А.: Значительная часть данных о противовирусной активности представляет из себя результаты фенотипического скрининга - определения противовирусной активности соединений в экспериментах на клеточных линиях. Эксперименты такого типа не позволяют определить конкретную молекулярную мишень действия соединений. Однако, используя алгоритм SEA, можно выдвинуть гипотезу о том, обусловлена ли их активность взаимодействием с одним из белков вируса. Для этого было рассчитано подобие активных соединений для белков вирусов (viral target, tid) и фенотипических экспериментов по определению противовирусной активности (assay, assay_id) и получены пары "assay_id" - "tid" (assay_id из выборки ViralChEMBL - tid мишень из БД ViralChEMBL). На слайдах 18-19 выборочно представлены несколько обнаруженных пар эксперимент - мишень. Обозначение полей в таблице следующее:

"Описание эксперимента" - (assays.description)

"организм эксперимента" - (assays.assay_organism),

"мишень, организм мишени" - спрогнозированная молекулярная мишень действия (белок вируса) (tid, target_dictionary.organism)

"p_val" - рассчитанное р-значение, которое отражает подобие соответствующих молекулярных наборов.

Например, для записи эксперимента ChEMBL699363 в базе данных ChEMBL в качестве мишени указан вирус гриппа А без указания мишени. Согласно прогнозу с использованием метода SEA, мишенью действия данных соединений является нейраминидаза.

Ефремов Р.Г.: В докладе основное внимание уделено техническим деталям построения модели и результатам сравнения химических пространств соединений, однако калибровка модели по экспериментальным данным, вопросы устойчивости модели (чувствительность результатов к небольшим изменениям параметров и входных (часто неточных!) данных), точность прогноза и пр. не обсуждаются.

Никитина А.А.: На данном этапе нашей работе вопросы устойчивости и точности прогноза не анализировались. Однако эти вопросы были рассмотрены ранее в работах Keiser et.al.[1] и Wang et.al.[2], посвященных методу SEA. Например, было показано, что при небольших изменениях гиперпараметров модели (порогов структурного подобия и биологической активности, использовании различных распределений значений rawscore: нормальное и экстремальных значений), прогнозирующая способность модели меняется незначительно. В дальнейшем мы также планируем проанализировать влияние этих параметров на качество прогноза для выборки противовирусных соединений и при наличии достаточного количества данных провести time-split валидацию.

Ефремов Р.Г.: Существуют ли в литературе примеры противовирусных соединений и их мишеней, с помощью которых можно было бы сравнить эффективность предложенного подхода с уже имеющимися технологиями? Без этого нельзя оценить научную новизну и значимость работы.

Никитина А.А.: Особенность метода SEA по сравнению с другими методами заключается в том, что он изначально предназначен для оценки подобия мишеней (определяющегося подобием набора активных соединений), в отличие от многих других методов виртуального скрининга, основанных на структуре лиганда, в ходе которых производится прогноз какого-либо свойства для одного лиганда. В тоже время, метод SEA можно применить и в варианте прогнозирования мишени для индивидуального соединения. На данном этапе работы, мы не проводили валидацию метода на внешней выборке. Однако вопросы предсказательной способности алгоритма были рассмотрены в работах [1, 2], посвященных методу SEA, в том числе экспериментально. Кроме того, в работе [1] авторы сравнивали результаты SEA с наиболее

очевидным способом оценки подоби́я белков - анализом подоби́я их аминокислотных последовательностей. Авторами было показано, что данные методы, основанные на различных подходах к определению подоби́я мишеней, дополняют друг друга.

Ефремов Р.Г.: Чем объясняются столь значительные флуктуации значения «хи-квадрат» на слайде 11?

Никитина А.А.: На слайде 11 приведено изменение значения критерия с изменением порогового значения химического подоби́я, при этом максимальные значения критерия соответствуют низким порогам. С увеличением порога значение критерия быстро уменьшается от значений >20 до значений <1. Значение критерия при порогах более 0.7 колеблется в пределах ~0.01-0.1 незначительно в сравнении с максимальными значениями. При слишком высоких значениях порога подоби́я вклады в rawscore значительного количества пар соединений становятся равны нулю, что может приводить к потере релевантной информации о химическом подоби́и. Поэтому нами было выбрано пороговое значение индекса Танимото 0.65, которое очень близко к значению, использованному в оригинальной статье [1].

1. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotech* 25 (2), 197-206 (2007)
2. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. *J Cheminform.* 2016;8:20 (2016)