Предсказание антигенов по последовательностям белка

Антон Смирнов

Feb 27, 2023

Оглавление

# Предисловие

Данная книга является отчетом о проделанной исследовательской работе. Тема иследовательской работы является предсказание с помощью математических методов антигенных детерминант по исходным последовательностям белков.

# 1. Список сокращений

* МНС - Major histocompatibility complex - главный комплекс гистосовместимости
* ТКР - Т-клеточный рецептор
* CDR - complementarity determining region - регион, определяющий комплементарность

# 2. Введение

    Каждый организм стремится передать свой генетический материал потомкам. При этом у каждого организма должны быть механизмы защиты от конкурентов. Эти механизмы эволюционируют вместе с организмами. Примерно 500 миллионов лет назад, у хордовых скачкообразно возник ароморфоз: адаптивная иммунная система[2]. Иммунная система - специализированная система органов и тканей, обеспечивающих иммунитет[1]. Иммунитет - способ защиты организма от живых тел и веществ, несущих признаки чужеродной информации[1]. Вещества, которые несут такие признаки называются антигенами.

    Адаптивность “новинки” проявляется в том, что спектр антигенов, против которых защищает адаптивная иммунная система, не наследуется, а формируется в процессе жизни организма и определяется теми антигенами, с которыми встречался организм на протяжении жизни. Т- и В-лимфоциты, составляющие адаптивную иммунную систему, распознают участок молекулы антигена, которые называется эпитоп (антигенная детерминанта).

    Существуют разница в том, как и с помощью каких молекул распознается антиген Т- и В-лимофцитами. B-лимфоциты способны распознавать нативный антиген с помощью В-клеточного рецептора. Т-лимфоциты способны распознавать только антигены, презентированные на главном комплексе гистосовместимости (МНС), с помощью своих Т-клеточных рецепторов (ТКР). ТКР содержит 4 особых участка на обеих цепях, ответственные за распознавание комплекса эпитоп-МНС - регионы, определяющие комплементарность (CDR). В данной работе будет рассмотрено, как организм подгатавливает антигены для распознавания Т-лимфоцитами и как этот процесс можно предсказывать, и будет предложен новый подход к предсказанаию результата этого процесса.

    Математическая оценка возможного количества -цепей ТКР равно [3]. Количество возможных белков длины n из 20 основных протеиногенных аминокислот равняется и из этого разнообразия возможно появление ещё большего числа пептидов, которые могут быть эпитопами. Невозможно и нецелесообразно проверять все сочетания этого разнообразия на предмет наличия или отсутствия взаимодействия экспериментально. Однако за 50 лет активной работы был накомплен огромный массив данных о процессах подготовки антигенов, их презентации и распознавания целевыми молекулами, в нашем случае ТКР. Эти данные аккумулированы в различные базы данных и базы знаний, которые позволяют учёным строить математические модели для предсказания результатов разных этапов процессинга антигенов, наличия или отсутствия иммуногенности и аллергенности эпитопов.

    Над построением математических моделей работает большое количество ученых уже больше двадцати лет: первые программы начали появляться в начале нулевых годов этого века. Данные модели крайне полезны в вакцинологии и онкологии. В вакцинологии они используются для проектирования вакцин от различных заболеваний, в частности от лихорадки Эбола и COVID-19. Для онкологии они используются для предсказания неоантигенов. Неоантигены - новые антигены, которых нет в здоровых клетках и которые появились в раковых в процессе накопления новых генетических вариантов в сравнении с исходными, герминальными клетками. Без предсказания антигенов и неоантигенов невозможна персонализированная иммунотерапия опухолей. С течением времени качество моделей заметно росло и современные модели показывают хорошие результаты, но они ограничены небольшой выборкой людей, к которым они применимы. Существующие аналоги, как правило, позволяют предсказывать результат одного из этапов процессинга и работают на алгоритмах, требующих эмбеддинга, то есть представления последовательности белка в виде числовых векторов. Целью данной работы я строю целостную модель “структура-активность” процессинга антигена для МНС первого класса. Под целостной моделью я подразумеваю рассмотрение процессинга антигенов как единой системы, где каждый предыдущий результат подстроен и оптимизирован для следующего этапа. В моей модели структура белков будет использована в виде совокупности молекулярных фрагментов, а не числовых векторов. Для выполнения цели данной работы необходимо решить следующие задачи:

1. Исследовать литературу по процессингу антигенов для МНС первого класса, существующим аналогам и алгоритмам их работы;
2. Разработать концепт системы, определить алгоритмы и метрики, которые будут использованы для разработки частей системы;
3. Охарактеризовать имеющиеся необходимые источники данных для моделирования, подготовить выборки для обучения и валидации;
4. Построить модели, оценить их качество;
5. Разработать веб-сервис для общедоступного использования построенных моделей.

В дальнейшем я надеюсь, что построенная система будет пригодна к использованию в исследовательской и клинической практике.

# 3. Главный комплекс гистосовместимости

Главный комплекс гистосовместимости - комплекс тесно связанных генетических локусов, а также их белковых продуктов, отвечающих за развитие иммунного ответа и синтез трансплантационных антигенов.

# 4. Процессинг антигенов для MHC I

# 5. Процессинг антигенов для MHC II

# 6. Программы и сервисы для предсказания Т-клеточных антигенов

# Список литературы

1. Ганковская Л. В. Основы общей иммунологии. Учебно-методическое пособие для студентов медицинских вузов / Л. В. Ганковская, Л. С. Намазова-Баранова, Р. Я. Мешкова. – Москва : ПедиатрЪ, 2014.1. Ганковская Л. В. Основы общей иммунологии. Учебно-методическое пособие для студентов медицинских вузов / Л. В. Ганковская, Л. С. Намазова-Баранова, Р. Я. Мешкова. – Москва : ПедиатрЪ, 2014.

2. Ярилин А. А. Иммунология / А. А. Ярилин. – Москва : ГЭОТАР-Медиа, 2010.2. Ярилин А. А. Иммунология / А. А. Ярилин. – Москва : ГЭОТАР-Медиа, 2010.

3. Murugan A. [Statistical inference of the generation probability of T-cell receptors from sequence repertoires](https://doi.org/10.1073/pnas.1212755109) / A. Murugan, T. Mora, A. M. Walczak, C. G. Callan // Proceedings of the National Academy of Sciences. – 2012. – Vol. 109. – № 40. – P. 16161-161663. Murugan A. [Statistical inference of the generation probability of T-cell receptors from sequence repertoires](https://doi.org/10.1073/pnas.1212755109) / A. Murugan, T. Mora, A. M. Walczak, C. G. Callan // Proceedings of the National Academy of Sciences. – 2012. – Vol. 109. – № 40. – P. 16161-16166.