электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51038.

УДК 51-76

Web-приложение для анализа репертуаров Т-клеточных рецепторов

***Бабарыкин Д.С.**, студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»

Научный руководитель: Назаров В.И., ассистент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»;
биоинформатик-исследователь Лаборатории
Сравнительной и Фундаментальной Геномики,
Россия, 117997, г. Москва, ИБХ РАН
bauman@bmstu.ru

Введение

В организме человека существует два вида иммунитета: врождённый (*innate*) и приобретённый (*adaptive*). Врождённый иммунитет является эволюционно более древним и присутствует у всех видов растений и животных. Он полностью определяется геномом и способен обезвреживать чужеродный и потенциально опасный биоматериал до его первого попадания в организм. Однако из этого следует, что число патогенов, которые способна распознать эта часть иммунной системы ограничена и потому неспособна защитить носителя от всего разнообразия опасного биоматериала, представленного в окружающей среде. Именно поэтому в процессе эволюции появился более совершенный механизм распознавания патогенов – приобретённый иммунитет.

Важную роль в системе приобретённого иммунитета играют Т-клетки (*T cells*) или Т-лимфоциты (*T lymphocytes*). Они делятся на несколько типов, каждый из которых специализирован под один из трёх основных этапов приобретённой иммунной защиты:

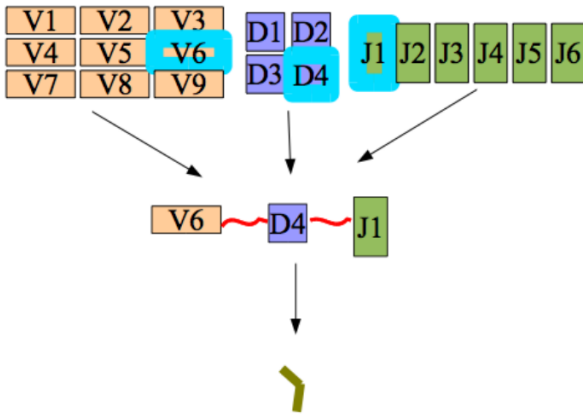
1) Распознавание антигенов и активация иммунного ответа
2) Иммунный ответ
3) Запоминание контакта с антигенами.

Для выполнения каждого из этих этапов лимфоциты должны иметь возможность распознавать конкретные антигены, с которыми они призваны бороться. Эту задачу

выполняют Т-клеточные рецепторы (*TCR*), расположенные на поверхности Т-клеток. Именно они и распознают антиген, т.е. определяют специфичность данной клетки к определенному антигену.

TCR представляет собой белковый комплекс, состоящий из двух субъединиц — α и β либо γ и δ, представленных на поверхности клетки (из-за малоизученности γ и δ, интерес представляют лишь αβ-цепи). Один Т-лимфоцит обладает одним Т-клеточным рецептором, который способен распознать только один определённый антиген. Таким образом, чтобы иметь возможность распознавать широкий спектр антигенов, организму необходимо иметь большое количество Т-клеток с большим диапазоном вариативности Т-клеточных рецепторов на них.

Правила сборки каждого рецептора записаны в гене этого рецептора. Ген – это участок *ДНК* и о нем можно думать просто как о последовательности, состоящей из четырехбуквенного алфавита. Уникальность гена Т-клеточного рецептора заключается в том, что он разный в разных клетках. Все остальные гены, которые есть в наших клетках – одинаковые. Составляется данный ген из нескольких основных кусков. Один кусок в Т-клеточных рецепторах всегда одинаковый. Затем в α-цепи идёт два участка, случайно выбранные из пулов (наборов) под названием V и J. Причём во время сборки этих двух участков между ними добавляются случайные буквы (порядка десяти), что приводит к формированию гигантского разнообразия таких рецепторов. β-цепь формируется аналогично, только между участками из V и J ещё вставляется короткая последовательность из пула с названием D. Участок получившейся последовательности, начинающийся с конца цепи из V и заканчивающийся в начале J, и распознает в белке антиген и называется *CDR3*. Схему вышеописанного процесса можно увидеть на рисунке.



CDR3 интересен с двух сторон: с одной стороны, он обуславливает специфичность, потому что как раз он кодирует ответственный за распознавание

антигена участок рецептора, а с другой стороны он удобен как идентификатор, потому что в нем сосредоточено все разнообразие. По его содержанию мы полностью можем определить, какой из набора этих сегментов был выбран, а также в нем полностью сосредоточены все случайные вставки. Поэтому, зная *CDR3*, мы можем реконструировать весь белок. Мы также можем его использовать как идентификатор, потому что, если *CDR3* одинаковый, то и весь ген тоже одинаковый.

Таким образом, анализ *CDR3* участков Т-клеток, как отдельно выбранного человека, так и большой выборки людей, открывает широкий спектр возможностей по диагностике больных: можно определить какой именно вирусной инфекцией поражен человек, становится возможным восстановить историю инфекционной болезни, можно определить к каким патогенам человек уязвим более всего и многое другое.

Обзор существующих средств анализа Т-клеточных рецепторов

На практике процесс получения информации о *TCR* проводится следующим образом. Сначала берётся образец крови, затем в лаборатории проводится специальная последовательность молекулярно-химических реакций, для того чтобы выделить из него последовательности *ДНК*, содержащие *CDR3*. Далее они загружаются на секвенатор (прибор для расшифровывания первичной структуры *ДНК*) и на выходе получаются нуклеотидные последовательности длиной в 100-250 символов. Следующей этапом является реконструирование из этих данных то, что было в самом начале т.е. узнать, сколько и каких клонов (Т-клеток с одинаковыми *CDR3*) было в первичном образце и т.п. Это уже сложная задача, выполняемая биоинформатическими средствами. Все осложняется тем, что в данных есть самые разные ошибки. Ошибки эти вносятся на всех этапах: пока происходит пробоподготовка, пока секвенируется и т.д. Темпы появления ошибок могут быть разными, в зависимости от архитектуры эксперимента и секвенатора, который был использован. Это могут быть вставки, выпадения и замены.

Для решения этой задачи на данный момент существует несколько программных продуктов с примерно одинаковым функционалом и используемыми методами. Из них самым популярным является *MiTCR*, распространяющийся под открытой лицензией. Он принимает на вход “сырые” данные с секвенатора и извлекает порядка 98% входных последовательностей, даёт информацию о составе *CDR3* и определяет количество клонов в образце. Однако вся эта информация записывается в громоздкую таблицу в текстовом файле (до 50000 строк), что делает её малопригодной для анализа человеком в силу

отсутствия наглядности данных. Чтобы человек смог провести анализ и сделать выводы из полученных данных, их необходимо грамотно визуализировать, что и подталкивает нас к необходимости создания веб-приложения, предоставляющего подобный функционал. Единственная сторонняя реализация этой идеи носит название *Vidjil* и так же распространяется под открытой лицензией. Однако данный программный продукт ещё далёк от стадии завершения, имеет неудобный пользовательский интерфейс и, несмотря на постепенное развитие, пока что мало пригоден для серьёзного использования. В связи с этим мы пришли к выводу о необходимости разработать собственное web-приложение для решения поставленной задачи, что и было сделано.

Функционал веб-приложения

Для осуществления анализа репертуаров Т-клеточных рецепторов были реализованы следующие интерактивные графики:

- Гистограмма распределения длин рецепторов (длины нуклеотидных последовательностей) для выбранного человека
- Гистограмма распределения V-сегментов для выбранного человека
- Гистограмма численности клонов для выбранной группы людей
- Теплокарта (*heatmap*) количества общих клонов по нуклеотидным и аминокислотным последовательностям CDR3 для выбранной пары людей
- Теплокарта распределения Дженсона – Шеннона V-сегментов для выбранной пары людей.

Структура приложения

Так как предметная область предполагает работу с базой данных Т-клеточных рецепторов множества пользователей, очевидно, что приложение должно обладать клиент-серверной архитектурой. В данном случае клиент расположен в Web-пространстве и представляет собой одностраничное приложение (*Single Page Application – SPA*). Это стало возможным благодаря использованию популярного *MVC* фреймворка на JavaScript – *AngularJS*. Он предоставляет широкие возможности по связыванию представлений с моделями при помощи контроллеров. Через контроллеры определена гибкая система маршрутизации. В данном случае получилось 7 контроллеров: один общий, один для работы с таблицей людей и по одному на каждую задачу по визуализации. Контроллеры отправляют данные на сервер с помощью *AJAX*-запросов в

архитектуре *RESTful* и принимают данные в формате *JSON*. Полученные данные визуализируются с помощью JS библиотеки для создания *Data-Driven* документов *D3.js*.

Серверная часть реализована с помощью технологии от *Microsoft – ASP.NET Web API 2*. Через *Entity Framework* на сервере была создана и заполнена база данных SQL с информацией из текстовых файлов, полученных из программы *MiTCR*. Сервер с помощью *Web API* получает запрос от клиента, затем через *Entity Framework* и *LINQ* обращается к базе данных, получает необходимую выборку, обрабатывает её, формирует ответ, сериализует его в JSON и отправляет обратно клиенту.

Заключение

В результате нами была рассмотрена предметная область приобретённого иммунитета и анализа репертуаров Т-клеточных рецепторов биоинформатическими методами; рассмотрены существующие в этой области программные продукты и создано собственное веб-приложение для визуализации репертуаров *TCR*.

Список литературы

1. Jorg J.A. Calis, Brad R. Rosenberg. Characterizing immune repertoires by high throughput sequencing: strategies and applications. Cambridge, Massachusetts, USA: CellPress, 2014. 10 p.
2. MiTCR – платформа для диагностики нового типа. Режим доступа: http://habrahabr.ru/company/yandex/blog/228003/ (дата обращения 26.03.2015).
3. Vidjil – Browser Manual. Available at: http://rbx.vidjil.org/browser/help/ accessed 26.03.2015.