

Федеральное государственное бюджетное учреждение науки  
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ  
им. академиков М.М. Шемякина и Ю.А. Овчинникова  
Российской академии наук

на правах рукописи

Погорелый Михаил Валерьевич  
**Анализ вклада генетических факторов и факторов окружающей среды в  
формирование репертуара Т-клеточных рецепторов монозиготных  
близнецов**

специальность – 03.01.03 – молекулярная биология

**АВТОРЕФЕРАТ**  
диссертации на соискание учёной степени  
кандидата биологических наук

Москва – 2018

Работа выполнена в лаборатории сравнительной и функциональной геномики Федерального государственного бюджетного учреждения науки Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук (ИБХ РАН).

**Научный руководитель:**

**Лебедев Юрий Борисович**, доктор биологических наук, профессор, главный научный сотрудник лаборатории сравнительной и функциональной геномики ИБХ РАН

**Официальные оппоненты:**

**Купраш Дмитрий Владимирович**, доктор биологических наук, профессор, член-корреспондент РАН; главный научный сотрудник Лаборатории передачи внутриклеточных сигналов в норме и патологии Федерального государственного бюджетного учреждения науки Института молекулярной биологии им. В. А. Энгельгардта Российской академии наук

**Базыкин Георгий Александрович**, кандидат биологических наук, заведующий сектором молекулярной эволюции Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук

**Ведущая организация:**

Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр гематологии» Министерства здравоохранения Российской Федерации

Защита диссертации состоится «10» октября 2018 года в 10:00 на заседании диссертационного совета Д 002.019.01 при Федеральном государственном бюджетном учреждении науки Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук по адресу: 117997, г. Москва, В-437, ул. Миклухо-Маклая, д. 16/10.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук и на сайте [www.ibch.ru](http://www.ibch.ru).

Автореферат разослан « » \_\_\_\_\_ 2018 года.

**Учёный секретарь диссертационного совета,**

**доктор физико-математических наук**



**В.А. Олейников**

## Общая характеристика работы

### Актуальность темы исследования

Развитие технологий высокопроизводительного секвенирования (High throughput sequencing, HTS) произвело настоящую революцию в изучении геномов и транскриптомов. Объемы данных растут экспоненциально, растет и число приложений этой технологии. Одна из новейших и наиболее перспективных областей применения HTS – направленное секвенирование генов гипервариабельных антигенраспознающих рецепторов Т- и В-лимфоцитов.

Иммунной системе требуется распознавать огромное разнообразие патогенов. Существуют два принципиальных подхода – распознавание паттернов патогенности (PAMP) и распознавание антигенов, определенных участков молекул патогена. В первом случае рецепторов нужно немного (в настоящее время известно несколько десятков), и их можно закодировать в геноме. Эти рецепторы играют ключевую роль в системе врожденного иммунитета. Распознавание антигенов осуществляется с помощью В-клеточных (BCR, B-cell receptor) и Т-клеточных (TCR, T-cell receptor) рецепторов. Разнообразие антигенов очень велико, и поэтому распознающие их рецепторы невозможно закодировать в геноме. Формирование генов TCR и BCR происходит в каждом лимфоците независимо из закодированных в геноме кассет сегментов в ходе процесса V(D)J рекомбинации. В результате этого стохастического процесса в некоторых лимфоцитах могут образовываться TCR, распознающие пептидные антигены собственного организма, удаление таких аутореактивных лимфоцитов происходит в ходе процесса негативной селекции в тимусе. После распознавания антигена Т-клеточным рецептором Т-лимфоцит приступает к пролиферации и дифференцировке, давая начало клону Т-клеток с одинаковой последовательностью и специфичностью Т-клеточного рецептора. Таким образом, три последовательных процесса определяют структуру репертуара Т-клеточных рецепторов: сборка, селекция и клональная экспансия в ответ на распознавание антигена. Все эти три процесса подвержены влиянию как генетических и индивидуальных биологических факторов, так и разнообразных факторов окружающей среды. Одним из наиболее эффективных подходов к изучению молекулярных механизмов такого влияния и к выяснению вклада отдельных факторов признаны исследования, проводимые на небольших когортах пар монозиготных близнецов.

В данной работе определены характерные особенности структуры репертуаров Т-клеточных рецепторов у группы монозиготных близнецов и исследованы вклады факторов окружающей среды, наследственности и своеобразия эмбрионального развития в формирование и функционирование репертуаров Т-лимфоцитов близнецовых пар.

### **Степень разработанности области исследования**

К настоящему моменту получены важные данные о структуре, особенностях рекомбинации и селекции TCR, что позволяет обучать модели сборки и селекции репертуара Т-клеточных рецепторов. Последовательное применение этих моделей позволяет симулировать репертуары функциональных и нефункциональных рецепторов *in silico*, что в свою очередь дает своего рода «нулевую гипотезу» для поиска интересных закономерностей в реальных данных массивированного секвенирования TCR. За последние годы накоплен массив экспериментальных данных о строении и онтогенезе репертуаров антиген-распознающих рецепторов человека. Часть подобных сведений содержится в первых работах, посвященных анализу репертуаров TCR монозиготных близнецов. Однако число таких работ незначительно, а особенности процессов формирования и развития адаптивного иммунитета близнецовых пар остаются малоизученными. Современные данные об онтогенезе адаптивной иммунной системы указывают в том числе на то, что в ранних эмбриональных В- и Т-клетках отличается механизм сборки Т-клеточного рецептора, что приводит к снижению разнообразия эмбриональных репертуаров. Тем не менее, вклад этих особенностей в строение зрелой иммунной системы не прослежен.

Сочетание биоинформатических и экспериментальных подходов анализа репертуаров (используемое в данной работе) открывает перспективы углубленного исследования ключевых закономерностей, включая возможное воздействие «эмбрионального» репертуара на строение и функционирование системы адаптивного иммунитета взрослого организма.

### **Цель и задачи работы**

Целью данной работы являлся анализ вклада генетических факторов и факторов окружающей среды в формирование репертуара Т-клеточных рецепторов монозиготных близнецов. Основной упор при этом делался на проверку гипотезы о внутриутробном обмене Т-клетками и углубленный анализ клонального состава периферических Т-лимфоцитов у близнецовых пар. Для достижения цели были поставлены следующие задачи:

1. Разработать усовершенствованную систему пробоподготовки к высокопроизводительному секвенированию кДНК библиотек альфа и бета цепей TCR и реконструировать серии индивидуальных репертуаров периферических T-лимфоцитов группы монозиготных близнецов.
2. Разработать эффективные алгоритмы и схемы биоинформатической обработки результатов секвенирования для глубокого профилирования, сравнительного анализа и прецизионной оценки динамики индивидуальных репертуаров TCR.
3. Провести сравнительный анализ общей структуры индивидуальных репертуаров периферических T-лимфоцитов для выявления и полной характеристики групп клонотипов, специфичных для каждой из обследуемых близнецовых пар.
4. Провести анализ распределения клонотипов предположительно эмбрионального происхождения в репертуарах образцов пуповинной крови, наивных T-клеток, T-клеток памяти и охарактеризовать возрастные изменения в части репертуара, представленной эмбриональными клонотипами.
5. Исследовать особенности развития противовирусного иммунного T-клеточного ответа на основе мониторинга изменений репертуара T-лимфоцитов при вакцинации монозиготных близнецов живой вакциной против вируса желтой лихорадки.
6. Разработать систему биоинформатического анализа индивидуальных репертуаров TCR для выявления клонотипов T-лимфоцитов ассоциированных с определенными заболеваниями.

### **Научная новизна**

В данной работе новые алгоритмы биоинформатического анализа и современные математические модели сборки и селекции TCR впервые применены к анализу уникальных массивов полученных данных высокопроизводительного секвенирования репертуаров TCR для проверки нескольких биологических гипотез.

В результате проведенного исследования были обнаружены свидетельства пренатального обмена лимфоцитами крови у близнецовых пар и определен вклад эмбрионального репертуара клонов T-клеток в систему адаптивного иммунитета взрослых доноров.

Впервые охарактеризован ряд принципиальных отличий между репертуарами наивных T-клеток и T-клеток памяти: показано, что наиболее высокопредставленные клонотипы репертуара наивных T-клеток часто не имеют вставочных нуклеотидов; определено, что доля

в репертуаре клонотипов, не имеющих вставочных нуклеотидов, снижается с возрастом; сформулирована гипотеза об образовании популяции Т-клеток со сниженным разнообразием TCR в раннем эмбриональном развитии организма, когда в механизме сборки Т-клеточных рецепторов не участвует фермент TdT.

Результативность практического применения нового биоинформатического подхода, использующего современные модели сборки Т-клеточных рецепторов, к анализу опубликованных данных секвенирования TCR продемонстрирована при оценке информативности маркеров Т-клеточных лейкозов и при поиске Т-клеточных клонов, ассоциированных с инфекционными и/или аутоиммунными заболеваниями.

Впервые описана динамика и охарактеризована перестройка репертуаров идентичных близнецов в модели острой вирусной инфекции (вакцинации против желтой лихорадки). Показано, что клональный состав ответа на вакцину уникален для любого индивида, за исключением монозиготных близнецов, в парах которых на вакцину чаще отвечают клоны с идентичной последовательностью TCR.

### **Практическая значимость**

Основные научные результаты, полученные при анализе структуры Т-клеточного репертуара и развитии иммунного ответа монозиготных близнецов, способствуют уточнению современного представлению о вкладе генетических факторов в формирование индивидуальных репертуаров Т-лимфоцитов и могут служить основой изучения роли эмбриональных клонотипов в функционировании развитой иммунной системы. Так, установление фактов пренатального обмена клонами Т-клеток и их длительного (десять лет) существования в репертуаре позволяет изучать продолжительность жизни клонов разных субпопуляций Т-клеток, а также возможный вклад этих эмбриональных клонов в иммунный и/или аутоиммунный ответ.

Усовершенствованные методы пробоподготовки и новые биоинформатические подходы к анализу данных секвенирования могут быть использованы в дальнейших исследованиях системы адаптивного иммунитета человека для глубокой реконструкции репертуаров TCR, в т.ч. для поиска клонов Т-клеток, ассоциированных с заболеваниями, или для решения других, сходных задач. Разработанный подход к поиску клонотипов TCR, значимо увеличивающих концентрацию после вакцинации, может найти применение в практике персонализированной медицины для оценки силы и клональной структуры Т-клеточного ответа на острые инфекции и вакцинацию.

### **Апробация работы и публикации**

По основным материалам диссертации было сделано 5 докладов на международных тематических конференциях, включая Systems Biology of Adaptive Immunity (SystImms), Аскона, Швейцария, 14-17 мая, 2017; International Congress of Immunology (ICI), Мельбурн, Австралия, 21-26 августа 2016; European Congress of Immunology (ECI), Вена, Австрия, 6-9 сентября 2015; Международную конференцию по биоорганической химии, биотехнологии и бионанотехнологии, Москва, 15-19 сентября 2014; XXVI Зимнюю молодежную научную школу «Перспективные направления физико-химической биологии и биотехнологии», Москва, 10-15 февраля 2014.

По теме диссертационной работы опубликовано 8 статей в рецензируемых журналах.

### **Объем и структура диссертации**

Диссертация изложена на 100 страницах текста и состоит из введения, обзора литературы, материалов и методов, результатов и их обсуждения, заключения, выводов, благодарностей, списка использованной литературы, включающего 96 ссылок. Диссертационная работа содержит 28 рисунков и 12 таблиц.

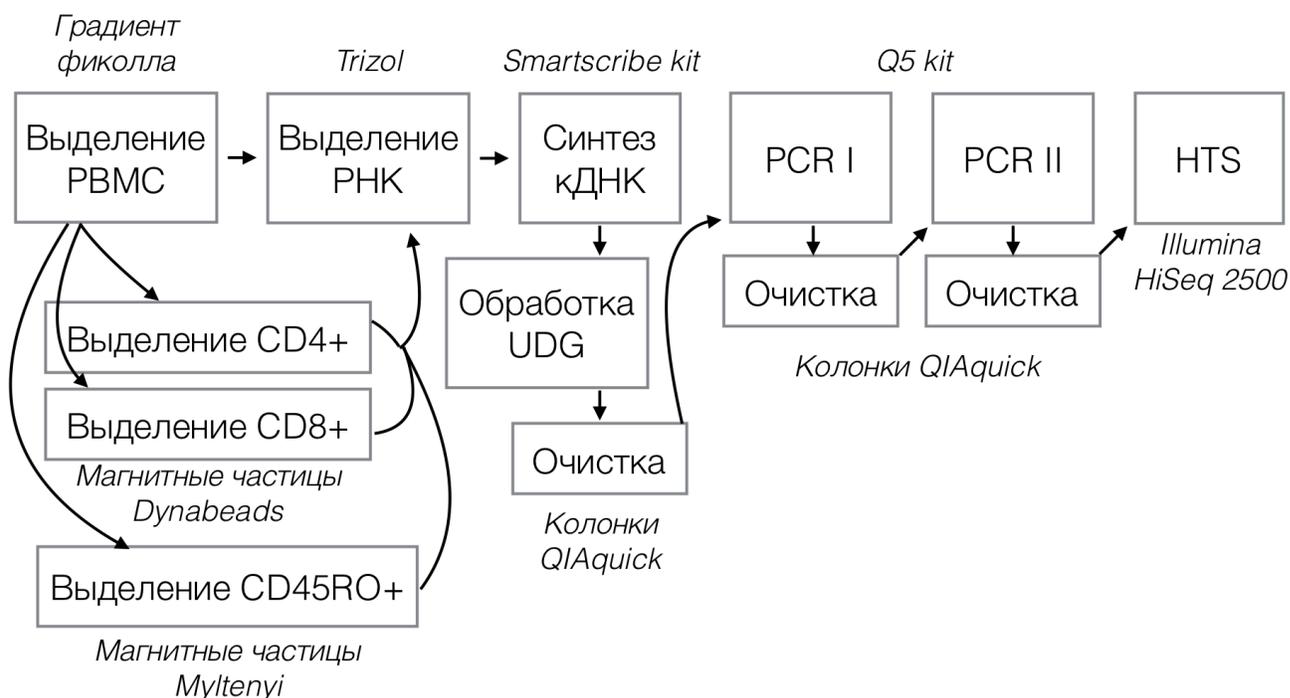
## Содержание работы

### 1. Разработка системы пробоподготовки библиотек кДНК TCR и подхода к анализу данных секвенирования репертуаров TCR

Для секвенирования кДНК альфа и бета цепей T-клеточных рецепторов на платформе Illumina необходимо решить несколько технических проблем, от выделения T-лимфоцитов из образца крови и введения в амплифицированную кДНК необходимых для секвенирования на Illumina адаптерных последовательностей до анализа полученных в ходе секвенирования больших объемов данных. Была разработана и оптимизирована система пробоподготовки библиотек кДНК TCRalpha и TCRbeta. Ниже перечислены ключевые особенности предлагаемой системы пробоподготовки:

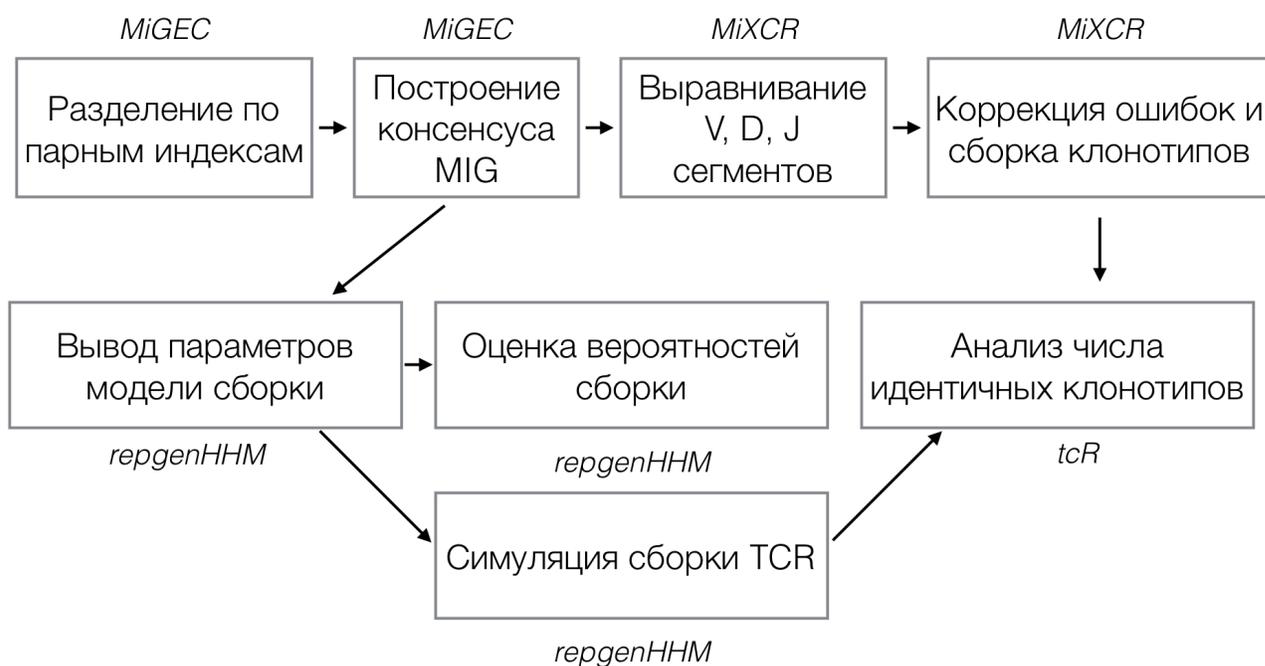
- 1) В качестве стартового материала используется тотальная РНК из фракции мононуклеарных клеток периферической крови (РВМС).
- 2) В ходе синтеза кДНК в каждую молекулу вводится уникальный молекулярный баркод (UMI, unique molecular identifier), а так же идентификатор образца (sample barcode).
- 3) ПЦР-амплификация кДНК происходит с универсальной пары праймеров (т.е. способной амплифицировать кДНК независимо от V- и J-сегментов TCR, который она кодирует).
- 4) В ходе ПЦР происходит введение всех необходимых для секвенирования на платформе Illumina адаптерных последовательностей.
- 5) Образцы содержат второй идентификатор образца, чтение которого происходит так же, как в системе Illumina TruSeq.

При длине чтения 100 нт. во втором «риде» целиком помещается участок CDR3, т.е. чтение начинается непосредственно с J-сегмента, а в первом «риде» находятся молекулярный баркод и идентификатор образца.



**Рис. 1. Диаграмма рабочего процесса подготовки библиотек.** На первом этапе с помощью градиента плотности фикола производилось выделение РВМС. Из части РВМС затем производилось иммуномагнитное выделение CD4+, CD8+, CD45RO+ субпопуляций. После этого из РВМС, а также CD4+, CD8+, CD45RO+ T-клеток выделяли тотальную РНК с помощью реагента Trizol. Тотальную РНК использовали для синтеза cДНК с помощью оригинальной системы праймеров, реакцию синтеза обрабатывали UDG, для расщепления содержащих dU олигонуклеотидов SMART и очищали на колонках для очистки PCR-продукта для удаления ревертазы и праймеров для синтеза. Далее очищенная cДНК амплифицировалась в двухстадийной ПЦР и секвенировалась на платформе Illumina HiSeq 2500.

Была разработана стратегия анализа данных секвенирования. Объем получаемых в результате высокопроизводительного секвенирования данных очень велик – десятки и сотни миллионов коротких нуклеотидных последовательностей. Анализ такого количества последовательностей невозможен без специализированных программных средств. Этапы анализа сырых данных секвенирования показаны на рис. 2.



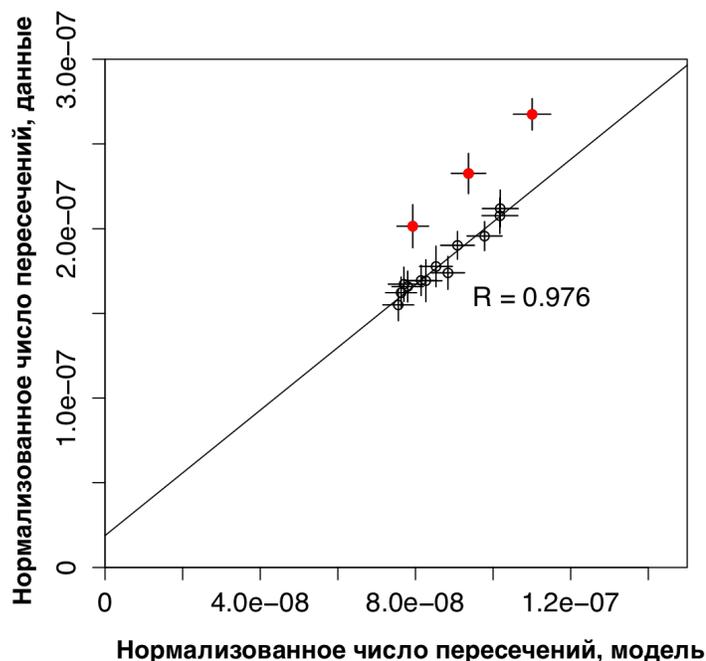
**Рис. 2.** Этапы анализа данных секвенирования репертуаров и используемые для этого программные пакеты.

## 2. Создание коллекции образцов и секвенирование библиотек кДНК TCR

Была собрана коллекция образцов периферической крови шести здоровых молодых близнецовых доноров. Из каждого образца с использованием описанной выше системы пробоподготовки были приготовлены библиотеки кДНК альфа- и бета- цепей TCR. Две пары доноров (пары S и P) в момент исследования проходили добровольную иммунизацию живой вакциной против желтой лихорадки. Для этих четырех доноров были также собраны образцы периферической крови на 15 день после иммунизации, из этих образцов были приготовлены библиотеки кДНК TCRbeta. Кроме того, по той же технологии были приготовлены библиотеки TCRbeta для полученных с помощью иммуномагнитной сепарации субпопуляций наивных T-клеток и T-клеток памяти, а также образца пуповинной крови. Все кДНК библиотеки были отсеквенированы на платформе Illumina HiSeq 2500. Полученные данные были депонированы в базу данных SRA (short read archive), идентификатор SRP078490. Репертуары TCR каждого образца были реконструированы с помощью описанного выше подхода. Дальнейшие разделы посвящены результатам анализа данных репертуаров.

### 3. Анализ числа идентичных клонотипов TCR между донорами

Если близнецовые доноры обмениваются Т-лимфоцитами до рождения, то это должно приводить к большему числу идентичных клонотипов между репертуарами близнецовых доноров, чем между репертуарами неродственными индивидов. Но как практически можно обнаружить этот эффект? И главное, как доказать, что он вызван именно физическим обменом Т-лимфоцитами, а не сходными параметрами машинерии сборки TCR и селекции в тимусе, или экспансией идентичных клонов в ответ на одинаковые патогены?



*Рис. 3. Нормализованное число идентичных клонотипов в близнецовых парах выше, чем у неродственных доноров. Число общих нефункциональных клонотипов альфа-цепи TCR для 15 возможных пар б доноров сравнивается с предсказанием модели сборки. Чтобы сравнить число идентичных клонотипов в выборках разного размера, число идентичных клонотипов нормализуют на произведение объемов выборок. Три красных круга - близнецовые пары. Теоретическое предсказание получено с помощью модели стохастической V(D)J рекомбинации, параметры модели выводились индивидуально для каждого донора, чтобы учесть особенности индивидуальных репертуаров. Модель хорошо предсказывает число идентичных клонотипов в парах неродственных индивидов, но систематически занижает это число для близнецов.*

Возможность обнаружить эффект определяется силой эффекта и чувствительностью метода. Чувствительность метода в нашем случае ограничивается глубиной секвенирования и небольшим объемом образца крови, который мы берем для исследования: нам удастся восстановить только малую часть существующего в человеке репертуара Т-клеточных рецепторов, поэтому существенную часть клонотипов, которыми близнецы обменялись до рождения мы можем просто не обнаружить. Под силой эффекта подразумевается приращение

числа общих клонотипов в близнецовой паре, которое вызвано физическим обменом Т-клетками до рождения, по сравнению с числом общих клонотипов, которое обусловлено другими причинами: конвергентной сборкой, селекцией и экспансией идентичных вариантов TCR в разных донорах.

Доказать возможность пренатального обмена Т-клетками между близнецовыми донорами можно используя простой принцип проверки статистических гипотез: сформулировать «нулевую гипотезу», и проверить, объясняет ли она наблюдаемые различия в числе общих клонов близнецовых доноров.

В качестве «нулевой гипотезы» были использованы симулированные репертуары TCR, которые можно получить, используя модели сборки и селекции TCR, при этом модели сборки и селекции были выведены для каждого донора отдельно, чтобы учесть индивидуальные особенности.

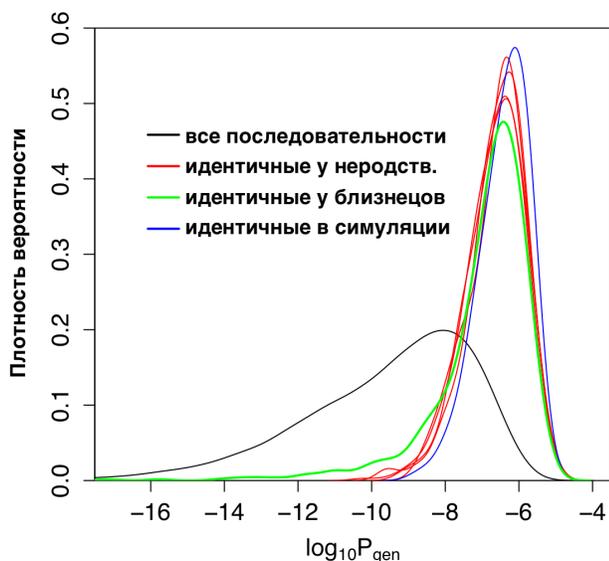
Анализ числа общих клонотипов проводили отдельно для функциональных и нефункциональных репертуаров альфа и бета цепи. Было показано, что число идентичных клонотипов между репертуарами близнецов превосходит теоретическое предсказание модели сборки для нефункциональных репертуаров TCRalpha (см. рис 3.), TCRbeta, и функциональных TCRbeta репертуаров. Большое число идентичных клонотипов в нефункциональных репертуарах не может быть объяснено случайной независимой сборкой этих клонотипов в близнецовых донорах и подтверждает гипотезу пренатального обмена.

#### **4. Анализ нефункциональных последовательностей TCR идентичных между парами близнецов**

В предыдущем разделе было показано, что у близнецов больше идентичных нуклеотидных последовательностей TCR между нефункциональными репертуарами.

Если гипотеза пренатального обмена верна, то помимо большего числа общих последовательностей, должны отличаться и сами последовательности. Последовательности TCR общие между неродственными донорами должны иметь высокую вероятность сборки, поскольку они два раза собрались независимо. Последовательности TCR, принадлежащие Т-клеткам, которыми близнецовые доноры обменялись до рождения могут иметь любую вероятность сборки. Из всех параметров рекомбинационного сценария, наибольшее влияние на вероятность сборки оказывает число инсерций. Ясно, что второй раз собрать последовательность с большим числом N-инсерций сложно, т.к. необходимо чтобы все инсерции совпали. Таким образом, можно ожидать большего среднего числа инсерций в

клонотипах идентичных между парами близнецов (поскольку у них может быть низкая вероятность сборки), чем между неродственными донорами. Мною были проверены эти гипотезы.



**Рис. 4. Последовательности TCR идентичные у монозиготных близнецов статистически отличаются от идентичных в неродственных индивидах.** Распределение  $\log_{10} P_{gen}$ , где  $P_{gen}$  это вероятность сборки последовательности T-клеточного рецептора, для клонотипов идентичных в репертуаре одного из индивидов и других пяти. Распределение вероятности сборки для клонов общих у неродственных доноров (красные кривые) хорошо предсказывается теоретически (синяя кривая). В то же время, распределение вероятностей идентичных клонов у близнецов имеет длинный хвост низковероятных последовательностей. Для сравнения черным показано распределение вероятностей сборки для всех клонов в репертуаре.

И действительно, было показано наличие группы идентичных между близнецами клонотипов с низкой вероятностью сборки (см. рис. 4, зеленая кривая). Было также подсчитано среднее число случайных инсерций, для клонотипов идентичных в парах близнецовых доноров, и было показано что оно выше, чем в клонотипах идентичных в парах неродственных доноров.

И наконец, если часть идентичных клонотипов близнецов по сути являются потомками одного клона Т-клеток, то в этом случае оба этих клонотипа будут принадлежать к одной субпопуляции: CD4+, либо CD8+ Т-клеток. Кроме того, можно предположить, что детектируемые спустя много лет после обмена пуповинной кровью идентичные клонотипы могут иметь фенотип клеток памяти. Для этого есть две причины. Во-первых, переход в популяцию Т-клеток памяти происходит после клональной экспансии, поэтому такие клоны имеют большую вероятность попасть в наши небольшие образцы крови. Во-вторых, клоны

клеток памяти могут иметь большую продолжительность жизни и сохраняться в организме десятилетиями.

Нами был проведен подробный анализ 84 нефункциональных клонотипов TCRbeta цепи идентичных в одной из близнецовых пар. Было показано неравномерное распределение данных клонотипов по фенотипам: 94% нефункциональных клонотипов имели идентичный в двух донорах фенотип, вместо примерно 50% ожидаемых теоретически по механизму конвергентной рекомбинации.

Таким образом, идентичные клонотипы близнецов имеют более низкую вероятность сборки, более высокое число инсерций и неравномерное распределение по фенотипам. Все эти три свойства согласуются с механизмом пренатального обмена и не согласуются с механизмом конвергентной рекомбинации.

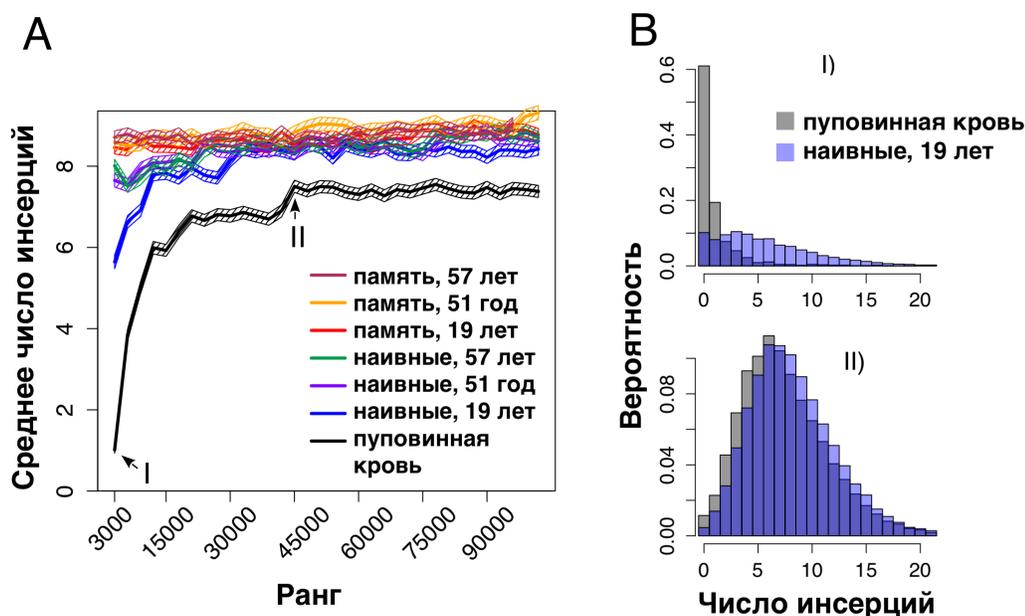
## **5. Анализ распределения числа инсерций в репертуарах наивных клеток и клеток памяти**

В предыдущем разделе нами были представлены доказательства гипотезы внутриутробного обмена Т-клетками между монозиготными близнецами. Из этого следует, что по крайней мере часть репертуара Т-клеточных рецепторов сформировавшегося еще до рождения сохраняется в организме как минимум несколько десятков лет. В этой главе будет сделана попытка сформулировать основные свойства эмбрионального репертуара Т-клеточных рецепторов и оценить его вклад в структуру репертуара TCR взрослых доноров.

Известно, что один из ключевых ферментов V(D)J-рекомбинации, терминальная дезоксирибонуклеотидилтрансфераза (TdT) отсутствует у плода в первом триместре беременности. Таким образом, первые Т-клетки не имеют N-вставок в последовательностях TCR. Были реконструированы репертуары TCRbeta клеток пуповинной крови, а также наивных Т-клеток и клеток памяти трех взрослых доноров 19, 51 и 57 лет соответственно. Я проанализировал количество и распределение клонов с нулем вставок в репертуаре. Для этого каждый репертуар был разбит на последовательные подгруппы клонов с убывающей численностью, по 3000 клонотипов в каждой. К каждой ранговой группе был независимо применен EM-алгоритм вывода параметров модели сборки.

Учитывая небольшой размер ранговых групп (2000-3000 последовательностей) полная модель сборки для каждой подгруппы не выводилась. Вместо этого были использованы ранее опубликованные параметры модели сборки бета-цепи Т-клеточного рецептора. Использовались частоты выбора V- и J-сегментов, а также профили делеций V- и J-сегментов из этой модели, а вероятности вставки определенного числа N-инсерций,  $P(\text{insVD})$  и  $P(\text{insDJ})$ ,

выводили из данных отдельно для каждой подгруппы. Выведенные распределения чисел инсерций в каждой подгруппе использовались для выявления зависимости среднего числа инсерций от ранга клона (рис. 5А).



**Рис. 5. Число N-вставок в функциональных клонотипах бета-цепи T-клеточного рецептора зависит от представленности этих клонотипов в репертуаре. А.** Среднее число инсерций было получено для последовательных групп из 3000 клонов, клоны в репертуаре отсортированы по представленности. Клонотипы пуповинной крови (показаны черным), показывают сильную зависимость от представленности, более высокопредставленные клоны имеют меньшее число N-вставок, чем менее высокопредставленные. Клонотипы из репертуара молодого донора демонстрируют тот же тренд, но менее ярко. Наивные клонотипы из пожилых доноров (фиолетовый и зеленый), демонстрируют еще более слабую зависимость среднего числа N-вставок в TCR от ранга. Интересно, что клонотипы из репертуаров клеток памяти не показывают связи между представленностью клона и числом N-вставок в нем (красная, желтая и каштановая кривая). Зашифрованная область показывает две стандартных ошибки среднего. **В.** Распределения вероятностей числа инсерций для клонов в двух ранговых группах (клоны с 1го по 3000 по представленности, и клоны с рангом от 45001 до 48000 по представленности) для молодого наивного репертуара и репертуара пуповинной крови. Для высокопредставленных последовательностей, вероятность иметь ноль вставок высока для наивных клонотипов и клонотипов пуповинной крови. Для группы малопредставленных клонотипов, вероятность сборки клонотипа с нулем вставок существенно ниже, и распределения близки для взрослого наивного репертуара и репертуара пуповинной крови.

На рис. 5А видно, что среднее число вставок существенно ниже в группе самых крупных клонов в репертуаре клеток пуповинной крови. Этот эффект достигается за счет возрастания доли клонов без N-вставок см рис. 5В. Интересно, что такая же зависимость

характерна для репертуаров наивных клеток взрослых доноров, но совершенно нехарактерна для клеток памяти.

Интересно, что эта зависимость ослабевает с возрастом - наивный репертуар девятнадцатилетнего донора гораздо ближе к репертуару клеток пуповинной крови, чем наивные репертуары 51- и 57-летних доноров. Причины подобной иерархии клонотипов в репертуаре пуповинной крови остаются невыясненными. Можно предложить две гипотезы, объясняющие повышенную представленность клонов без вставок в репертуаре: гипотезу гомеостатической пролиферации и гипотезу конвергентной рекомбинации. Первая гипотеза подразумевает, что у Т-клеток, которые появились раньше всех (в период отсутствия TdT, и, следовательно, они не имеют N-инсерций), было больше времени и свободного места для пролиферации, поэтому их клоны имеют больший размер. Гипотеза конвергентной рекомбинации подразумевает, что наблюдаемые крупные клонотипы на самом деле состоят из многих мелких, которые неразличимы по нуклеотидной последовательности бета-цепи Т-клеточного рецептора (но могут иметь разные альфа-цепи и фенотип). Недавно появившаяся технология массивированного секвенирования парных альфа-бета репертуаров, либо секвенирование большого числа отдельных клеток могут прояснить, какая из предложенных гипотез верна.

## **6. Анализ распределения числа наивных клонов в репертуаре в зависимости от возраста**

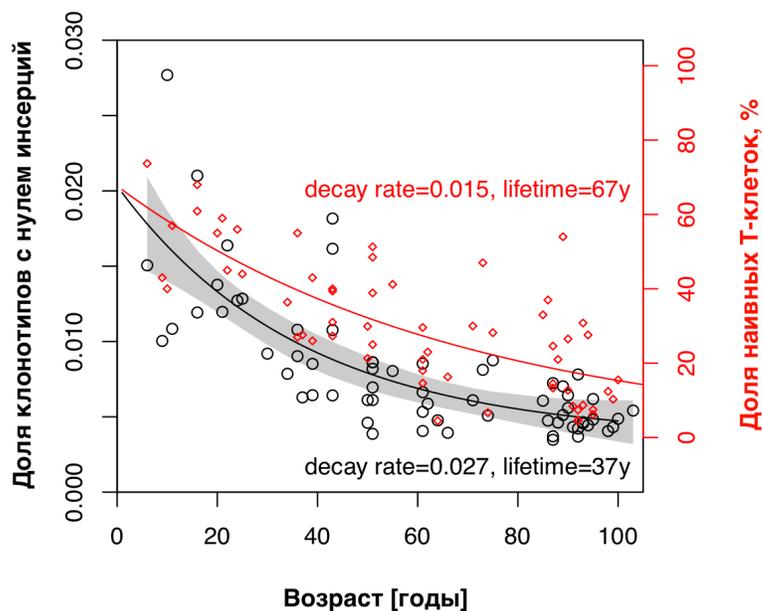
В предыдущем разделе было показано, что наиболее крупные клоны пуповинной крови и наивных Т-клеток взрослых доноров часто не имеют N-инсерций. Также было показано, что в репертуаре клеток памяти этот эффект не наблюдается.

Репертуар Т-клеток периферической крови можно рассматривать как смесь репертуара наивных Т-клеток и репертуара Т-клеток памяти. Известно, что доля наивных Т-клеток в репертуаре падает с возрастом. Можно предположить, что доля крупных клонотипов Т-клеток с нулем инсерций также должна снижаться с возрастом.

Чтобы проверить это предположение, был использован ранее опубликованный крупный массив данных репертуаров бета-цепи TCR доноров разных возрастов из предыдущих работ нашей лаборатории. Из репертуара каждого донора были взяты 2000 самых крупных клонотипов, и для них была подсчитана доля клонов с нулем N-инсерций (см рис. 6).

$$Z \approx c + a \exp(-bT)$$

где  $Z$  - доля клонов с нулем инсерций,  $1/b$  – это характеристическое время жизни, а  $a$  и  $c$  - свободные параметры.



**Рис. 6. Характеристическое время жизни функционального TCR-клонотипа без N-вставок.** Число клонов с нулем вставок в зависимости от возраста (черные кружки) может быть аппроксимировано законом радиоактивного распада (черная кривая). Этот распад происходит быстрее, чем снижение доли наивных клеток в организме по данным проточной цитофлуориметрии (красная кривая), что может свидетельствовать о постепенном замещении эмбриональных клонов более молодыми клетками внутри наивного репертуара.

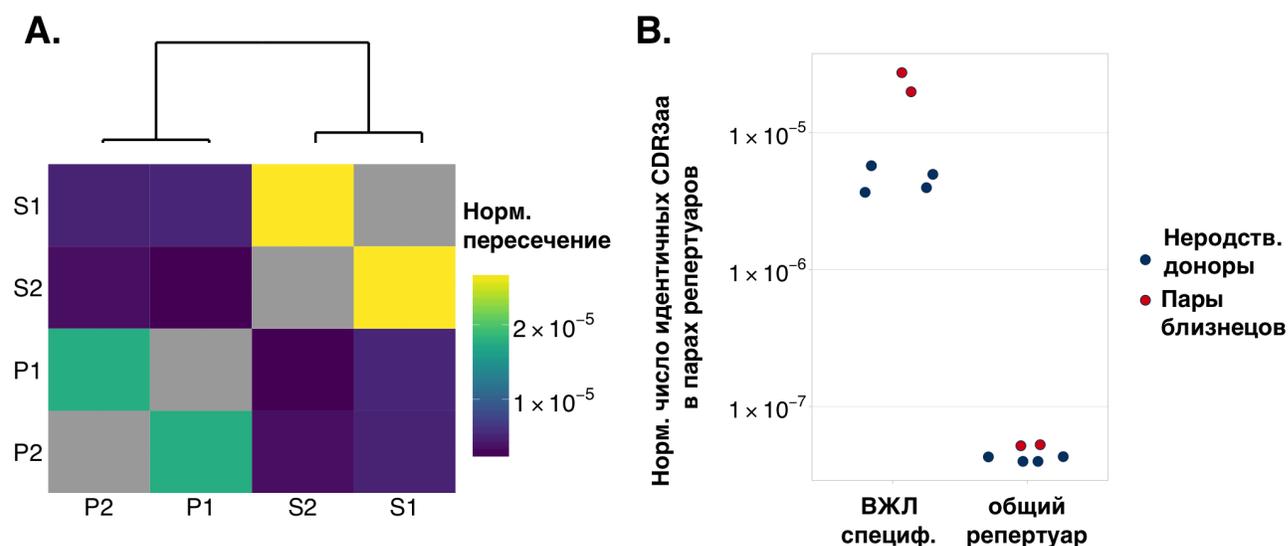
Выясняется, что характеристическое время жизни клона с нулем вставок (т.е. время, за которое его концентрация снижается в  $e$  раз) составляет 37 лет. Это снижение может быть вызвано двумя причинами - общим сокращением объема наивного репертуара и/или снижением концентрации клонов с нулем инсерций внутри наивного репертуара. Если обучить такую же регрессионную модель на данных о доле наивных Т-клеток для доноров разного возраста, то характеристическое время жизни наивной клетки составит 67 лет.

Представленные данные свидетельствуют о том, что наивные клоны без вставок, которые сформировались до рождения и занимают верхушку наивного репертуара постепенно вытесняются наивными клетками, которые генерируются в тимусе в течение жизни. Этот эффект вместе с общим падением доли наивных Т-клеток в репертуаре приводит к снижению доли клонов с нулем инсерций в группе высокопредставленных клонов TCR с возрастом.

## 7. Анализ реакции репертуара TCR монозиготных близнецов на мощный антигенный стимул

Две близнецовые пары, принимавшие участие в данном исследовании, S1, S2 и P1, P2 проходили добровольную вакцинацию против желтой лихорадки. Мы использовали эту возможность, чтобы оценить реакцию репертуара T-клеточных рецепторов на мощный иммунный стимул. Далее будет описана модель для поиска клонотипов TCRbeta проходящих клональную экспансию после иммунизации, а также дана краткая характеристика последовательностей этих клонотипов.

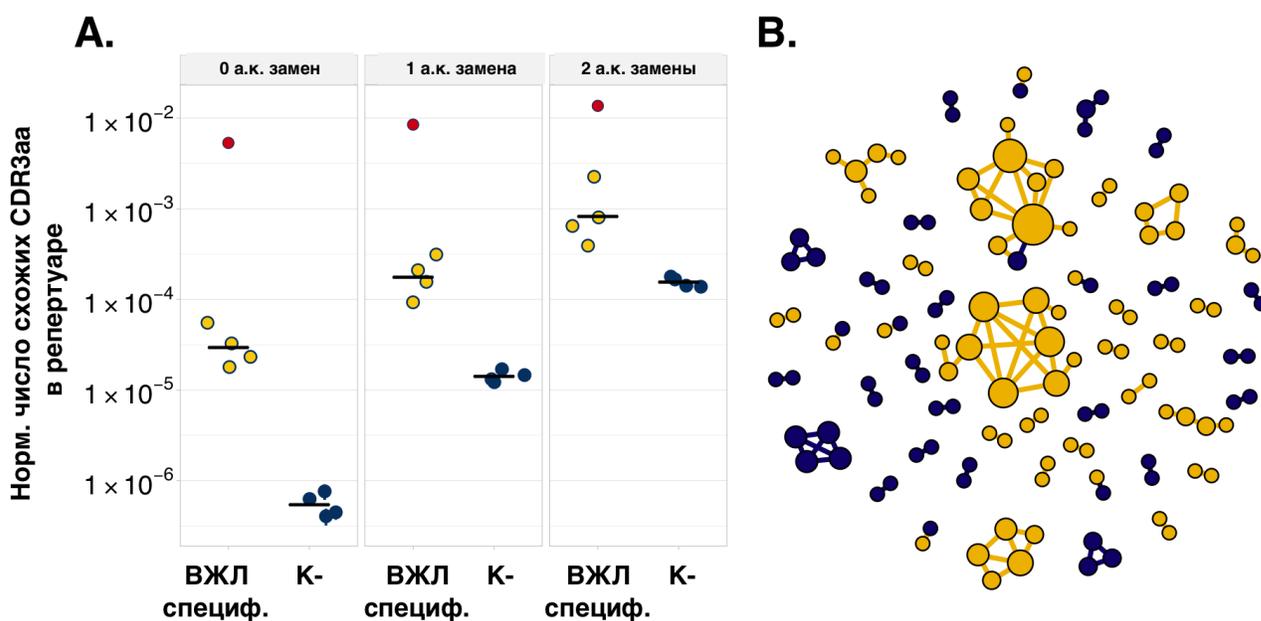
Нами были собраны образцы периферической крови четырех пар близнецовых доноров непосредственно перед (день 0) и на 15 день после иммунизации, где по данным литературы находится пик клеточного иммунного ответа. С помощью пакета *edgeR* широко используемого для дифференциальной экспрессии генов, нами были получены последовательности клонотипов T-клеточных рецепторов значимо ( $p\text{-value} < 0.001$ , qCML-test) и сильно ( $\log_2 \text{fold change} > 5$ ) увеличившие свою численность между днем 0 и днем 15. Число таких клонотипов составило 700-1100 в разных донорах. В дальнейшем для краткости такие клонотипы мы будем называть экспандированными клонотипами.



**Рис. 7.** Нормализованное число идентичных аминокислотных последовательностей TCRbeta значимо увеличившихся в ответ на иммунизацию против желтой лихорадки. Число идентичных экспандированных клонотипов в парах доноров нормализовано на произведение числа ответивших клонотипов в этих донорах. Близнецы оказываются в одном кластере, если применить к столбцам полученной матрицы алгоритм иерархической кластеризации.

Нами было подсчитано число идентичных аминокислотных и нуклеотидных последовательностей между экспандированными в ответ на вакцинацию против желтой лихорадки клонотипами TCRbeta. Интересно, что несмотря на в целом уникальный иммунный ответ, между близнецами обнаружилось больше идентичных аминокислотных и нуклеотидных последовательностей TCR, чем между неродственными донорами. При этом даже между неродственными донорами число идентичных ВЖЛ-специфичных а.к. последовательностей TCR выше, чем в среднем по репертуару (см. рис. 7B). Эти совпадения можно объяснить селекцией идентичных вариантов TCR на одинаковые антигены в репертуарах доноров.

Понятно, что те же самые процессы можно рассматривать и внутри каждого репертуара – много идентичных или очень похожих независимо сформировавшихся клонотипов TCR могут отвечать на одну и ту же вакцину. Для того, чтобы оценить число таких клонотипов, можно ввести индекс конвергенции: нормализованное число пар идентичных или очень похожих (одна-две аминокислотные замены) клонотипов в репертуаре. В качестве контроля можно взять случайно выбранные из репертуара последовательности TCR.



**Рис. 8. Конвергенция аминокислотных последовательностей, распознающих вакцину от желтой лихорадки TCR.** (А) число похожих клонов (идентичные аминокислотные последовательности, но разные нуклеотидные последовательности – левая панель; 1 аминокислотная замена – средняя панель, 2 аминокислотные замены – правая панель) в индивидуальных репертуарах TCR, нормализованное на возможное число пар последовательностей в каждой выборке. Число клонотипов с похожей последовательностью в специфичных к ВЖЛ репертуарах (желтые точки) выше, чем между случайной выборкой идентичного числа клонов из той же временной точки (К-, отрицательный контроль). Для сравнения, показано нормализованное число похожих ранее опубликованных клонотипов распознающих известный иммунодоминантный эпитоп. Показано, что между этими

точками наблюдаются значимые различия (односторонний парный  $t$ -тест,  $p=0.024, 0.0033, 0.016$  для 0, 1, 2 замен соответственно) (В) 1000 самых высокопредставленных TCR в доноре S1 на день 15, каждая вершина соответствует клонотипу TCR, они соединяются ребром, если между соответствующими последовательностями две или менее аминокислотные замены. Желтые точки соответствуют последовательностям, значимо увеличившим концентрацию после иммунизации против желтой лихорадки. Темно-синие точки соответствуют последовательностям значимо не изменившим свою численность. Видно, что синие и желтые точки кластеризуются отдельно – подавляющее большинство ребер (95 из 103) образуется между точками с одинаковым статусом (специфичных или неспецифичных к ВЖЛ).

Результат проведенного анализа представлен на рис 8. Видно, что TCR распознающие эпитопы вируса желтой лихорадки гораздо более похожи, чем случайно выбранные последовательности из TCR репертуара той же временной точки. Красные точки - нормализованное число схожих клонотипов (0,1 или 2 а.к. замены) в выборке опубликованных последовательностей TCR распознающих иммунодоминантный эпитоп вируса желтой лихорадки, NS4b<sub>214-222</sub>. Видно, что последовательности таких TCR еще больше похожи друг на друга, чем ВЖЛ-специфичные варианты TCR найденные в данной работе. Скорее всего это связано с тем, что наш метод идентифицировал несколько групп специфичных к различным иммунодоминантным эпитопам ВЖЛ последовательностей, которые не похожи друг на друга.

## 8. Оценка вероятности сборки TCR для поиска ассоциированных с заболеваниями вариантов TCR

Другое применение вероятности сборки возможно для поиска клонов TCR, ассоциированных с конкретными аутоиммунными или инфекционными заболеваниями. Подобные исследования стали особенно актуальны в последнее время в связи с появлением в международных базах данных (SRA, ENA) крупных массивов секвенированных репертуаров.

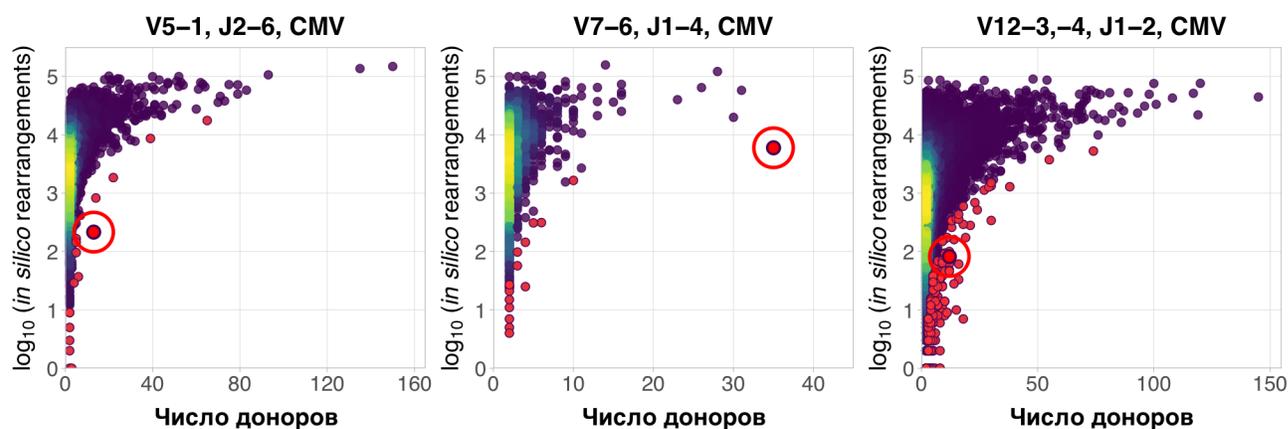


Рис. 9. Анализ вероятностей сборки клонотипов TCR, ассоциированных с цитомегаловирусной инфекцией (CMV). Красные круги – статистически значимые отклонения, найденные алгоритмом. Большим красным кругом обведены клонотипы с ранее доказанной CMV-специфичностью.

*Вероятность сборки для этих клонотипов ниже, чем для других клонотипов данных VJ-комбинаций общих у того же числа доноров.*

В рамках диссертационной работы мною был произведен анализ ранее опубликованных сырых данных когорты по цитомегаловирусной инфекции (CMV) доноров. Основной задачей была идентификация значимо ассоциированных с ответом на CMV клонотипов без использования контрольной когорты здоровых доноров.

Для каждой аминокислотной последовательности TCR идентичной в нескольких донорах проводили оценку вероятности сборки этой последовательности. Поскольку существующие алгоритмы не позволяют рассчитать вероятность сборки аминокислотной последовательности TCR, оценка вероятности сборки проводилась с помощью случайной симуляции. С помощью модели сборки генерировали  $2 \times 10^9$  нуклеотидных последовательностей, транслировали эти последовательности и подсчитывали число симулированных последовательностей с такой же аминокислотной последовательностью, как и у каждой последовательности из данных. Было показано, что у известных CMV-специфичных последовательностей TCR более низкая вероятность сборки, чем у клонов общих у аналогичного числа доноров см. рис. 9.

Таким образом, вероятность сборки можно использовать для поиска клонов значимо ассоциированных с заболеванием.

## **Заключение**

В данной работе мы использовали модели сборки и селекции TCR для доказательства гипотезы пренатального обмена Т-клетками между монозиготными близнецами. Функциональные последствия и распространённость этого эффекта еще предстоит выяснить.

Были рассмотрены характерные особенности онтогенеза репертуара Т-клеток: высказано предположение о большом вкладе в структуру репертуара особой популяции Т-клеток, которая образуется в раннем эмбриональном развитии и имеет сниженное разнообразие. Роль и функции этой популяции клеток еще предстоит прояснить.

Предложенные в работе подходы имеют и практическую ценность: модели сборки Т-клеточных рецепторов можно использовать для поиска мотивов в последовательностях TCR распознающих одни и те же антигены, по аналогии с тем, как модель случайной нуклеотидной последовательности используется для поиска мотивов в ДНК и РНК. Высокопроизводительное секвенирование, в частности новые методы реконструкции репертуаров из данных секвенирования транскриптомов, приведут к лавинообразному накоплению репертуаров TCR и важной задачей становится разработка новых статистических подходов для их анализа.

## Выводы

1. Разработана оригинальная система пробоподготовки библиотек кДНК альфа- и бета-цепей Т-клеточных рецепторов, обеспечивающая получение крупных массивов данных секвенирования и исчерпывающую реконструкцию полноразмерных репертуаров TCR.
2. С использованием разработанной системы пробоподготовки создана коллекция библиотек кДНК, проведено массивное секвенирование и реконструированы репертуары TCRalpha и TCRbeta периферических Т-лимфоцитов и различных субпопуляций Т-клеток для трех пар монозиготных близнецов в стационарном состоянии иммунной системы и при развитии иммунного ответа на противовирусную вакцинацию.
3. Разработаны программный пакет и серия биоинформатических алгоритмов для глубокого сравнительного анализа репертуаров Т-клеточных рецепторов и для определения динамики индивидуального репертуара TCR.
4. По результатам сравнительного анализа репертуаров TCR и определения свойств характеристических клонотипов:
  - выявлено, что число идентичных клонотипов в нефункциональных репертуарах TCRalpha, TCRbeta, а так же функциональных репертуарах TCRbeta монозиготных близнецов значительно превышает теоретически ожидаемое (для неродственных индивидов) число совпадений;
  - определены особенности идентичных в репертуарах близнецовых пар нефункциональных клонотипов TCR: относительно низкая вероятность сборки, повышенное число N-инсерций в структуре CDR3, неравномерное распределение клонотипов по популяциям Т-клеток с CD4 или CD8 фенотипами. Обе группы установленных фактов подтверждают гипотезу пренатального обмена пуповинной кровью между монозиготными близнецами.
5. С использованием глубокого клонального профилирования Т-лимфоцитов в образцах пуповинной крови и периферической крови выявлена фракция пренатальных долгоживущих клонотипов, которые сохраняются в репертуаре высокопредставленных

наивных Т-лимфоцитов взрослых доноров (но не Т-клеток памяти), и теряют клеточную численность с возрастом.

6. По результатам мониторинга индивидуальных репертуаров Т-лимфоцитов при противовирусной вакцинации выявлены и охарактеризованы особенности развития Т-клеточного иммунного ответа у пар монозиготных близнецов. В индивидуальных репертуарах пар монозиготных близнецов выявлены представительные группы ответивших на вакцину клонов Т-клеток, которые обладают идентичной или сходной структурой TCRbeta.
7. Разработана теоретическая модель и создан пакет программ для идентификации в репертуарах родственных и неродственных индивидов Т-клонов, ассоциированных с инфекционными и аутоиммунными заболеваниями. Эффективность работы ПО продемонстрирована при оценке информативности маркеров Т-клеточных лейкозов и при идентификации вирус-специфических субпопуляций Т-лимфоцитов.

## Список работ, опубликованных по теме диссертации

### Статьи

- 1) Shugay M, Bolotin DA, Putintseva EV, **Pogorelyy MV**, Mamedov IZ, Chudakov DM Huge overlap of individual TCR beta repertoires // Front. Immunol. 2013. Т. 4. № December. С. 466.
- 2) Putintseva EV, Britanova OV, Staroverov DB, Merzlyak EM, Turchaninova MA, Shugay M, Bolotin DA, **Pogorelyy MV**, Mamedov IZ, Bobrynina V, Maschan M, Lebedev YB, Chudakov DM Mother and child T cell receptor repertoires: deep profiling study // Front. Immunol. 2013. Т. 4. № December. С. 463.
- 3) Zvyagin IV, **Pogorelyy MV**, Ivanova ME, Komech EA, Shugay M, Bolotin DA, Shelonkov AA, Kurnosov AA, Staroverov DB, Chudakov DM, Lebedev YB, Mamedov IZ, Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing // Proc. Natl. Acad. Sci. 2014. Т. 111. № 16. С. 5980–5985.
- 4) Nazarov VI, **Pogorelyy MV**, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, Chudakov DM, Lebedev YB, Mamedov IZ tcR: an R package for T cell receptor repertoire advanced data analysis // BMC Bioinformatics. 2015. Т. 16. № 1. С. 175.
- 5) Nazarov VI, Minervina AA, Komkov AY, **Pogorelyy MV**, Maschan MA, Olshanskaya YV, Zvyagin IV, Chudakov DM, Lebedev YB, Mamedov IZ, Reliability of immune receptor rearrangements as genetic markers for minimal residual disease monitoring. // Bone Marrow Transplant. 2016. Т. 51. № 10. С. 1408–1410.
- 6) **Pogorelyy MV**, Elhanati Y, Marcou Q, Sycheva AL, Komech EA, Nazarov VI, Britanova OV, Chudakov DM, Mamedov IZ, Lebedev YB, Mora T, Walczak AM, Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. // PLOS Comput. Biol. 2017. Т. 13. № 7. С. e1005572.
- 7) **Pogorelyy MV**, Minervina AA, Chudakov DM, Mamedov IZ, Lebedev YB, Mora T, Walczak AM Method for identification of condition-associated public antigen receptor sequences. // Elife. 2018. Т. 7. С. 1–12.
- 8) Израельсон М, Касацкая С, **Погорелый М**, Киргизова В, Путинцева Е, Егоров ЕС, Британова ОВ, Чудаков ДМ Анализ индивидуальных репертуаров Т-клеточных рецепторов // Иммунология 2016 Т.37 №.6 С. 347-352

## **Тезисы докладов на конференциях**

- 1) **М.В. Погорельий**, Е.А. Комеч, В.И. Назаров, И.В. Звягин, Ю.Б. Лебедев, И.З. Мамедов  
Исследование репертуара TCR монозиготных и дизиготных близнецов, сборник тезисов XXVI Зимней молодежной научной школы “Перспективные направления физико-химической биологии и биотехнологии”, 2014, с. 113
- 2) Deep analysis of TCR repertoires of twins by next generation sequencing  
**Pogorelyy MV**, Zvyagin IV, Ivanova ME, Bolotin DA, Chudakov DM, Lebedev YB, Mamedov IZ, European Journal of Human genetics, 2014 T. 22 sup. 1,
- 3) Open software development for T-cell receptor repertoire data analysis  
**Pogorelyy MV**, Nazarov VI, Komech EA, Zvyagin IV, Mamedov IZ, Lebedev YB. Acta Naturae, Special issue №1, p. 38 2014
- 4) Persisting fetal clonotypes influence the structure and overlap of adult human TCR repertoires  
**Pogorelyy, M.**, Elhanati, Y., Marcou, Q., Sycheva, A., Komech, E., Britanova, O., Chudakov, D., Mamedov, I., Lebedev, Y., Mora, T., Walczak, A., European Journal of Immunology, Volume 46, Issue Supplement S1, 2016, p.52
- 5) High throughput sequencing of identical twins TCR repertoires after yellow fever vaccination  
**Pogorelyy M.V.**, Puelma Touzel M, Minervina A.A., Sycheva A.L., Chudakov D.M., Mamedov I.Z., Mora T., Walczak A.M., Lebedev Y.B., Systems Biology of Adaptive Immunity, Ascona, Switzerland, May 14-17, 2017