

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

диссертационная работа на соискание учёной степени  
кандидата биологических наук

научный руководитель:  
д.б.н. Лебедев Юрий Борисович

Москва – 2018

# Содержание

<b>1. Введение.....</b>	<b>4</b>
<b>2. Обзор литературы.....</b>	<b>10</b>
<b>2.1. Формирование разнообразия TCR .....</b>	<b>10</b>
2.1.1. Структура локусов TRA и TRB, сегменты .....	11
2.1.2. Последовательность событий при сборке генов TCRbeta и TCRalpha.....	12
<b>2.2. Положительная и отрицательная селекция TCR.....</b>	<b>14</b>
<b>2.3. Моделирование процесса сборки Т-клеточного рецептора.....</b>	<b>18</b>
<b>2.4. Моделирование селекции репертуара TCR.....</b>	<b>21</b>
<b>2.5. Особенности адаптивного иммунитета близнецов.....</b>	<b>22</b>
2.5.1. Сходство механизмов сборки и селекции TCR близнецов.....	23
2.5.2. Оценка вклада окружающей среды в структуру репертуара.....	24
2.5.3. Особенности эмбрионального развития близнецов .....	25
<b>2.6. Особенности раннего развития адаптивной иммунной системы.....</b>	<b>27</b>
<b>2.7. Мониторинг противовирусного иммунитета .....</b>	<b>28</b>
2.7.1. Вирус желтой лихорадки (YFV).....	28
2.7.2. Динамика иммунного ответа на вакцину против желтой лихорадки .....	29
2.7.3. Изучение иммунного ответа на вакцинацию с помощью HTS .....	31
<b>2.8. Заключение .....</b>	<b>32</b>
<b>3. Материалы и методы .....</b>	<b>33</b>
<b>3.1. Материалы .....</b>	<b>33</b>
3.1.1. Использованное оборудование и расходные материалы .....	33
3.1.2. Использованные реактивы.....	33
3.1.3. Буферные растворы .....	33
<b>3.2. Методы.....</b>	<b>34</b>
3.2.1. Доноры и образцы крови.....	34
3.2.2. Выделение РВМС.....	34
3.2.3. Выделение функциональных субпопуляций CD4, CD8, CD45RO Т-клеток.....	35
3.2.4. Выделение РНК из клеток с помощью реагента Trizol.....	35
3.2.5. Электрофорез в агарозном геле .....	36
3.2.6. Очистка ПЦР-продуктов.....	36
3.2.7. Выделение ДНК из агарозного геля .....	36
3.2.8. Приготовление библиотек кДНК альфа и бета цепей Т-клеточных рецепторов .....	37
3.2.9. Высокопроизводительное секвенирование (HTS).....	39
3.2.10. Обработка сырых массивов секвенирования .....	39
3.2.11. Обучение модели сборки .....	40
3.2.12. Вывод распределения инсерций для каждой ранговой группы бета-цепи TCR.....	41
3.2.13. Вывод факторов селекции .....	41
3.2.14. Предсказание совпадения нефункциональных последовательностей CDR3 между донорами .....	42
3.2.15. Предсказание числа общих функциональных последовательностей .....	42
3.2.16. Моделирование смеси фетальных и зрелых клонотипов .....	42
<b>4. Результаты и обсуждение .....</b>	<b>44</b>

<b>4.1. Разработка системы амплификации кДНК альфа- и бета-цепей Т-клеточного рецепторов.....</b>	<b>44</b>
<b>4.2. Разработка программного обеспечения для анализа Т-клеточных рецепторов.....</b>	<b>47</b>
<b>4.3. Создание коллекции образцов и секвенирование кДНК TCR.....</b>	<b>50</b>
<b>4.4. Анализ числа идентичных клонотипов в репертуарах доноров .....</b>	<b>53</b>
4.4.1. Анализ числа идентичных нефункциональных клонотипов TCRalpha между донорами	54
4.4.2. Анализ числа идентичных нефункциональных клонотипов TCRbeta между донорами..	56
4.4.3. Анализ числа идентичных клонотипов между функциональными репертуарами.....	59
4.4.4. Анализ числа идентичных функциональных клонотипов TCRalpha между донорами....	60
4.4.5. Анализ числа идентичных нефункциональных клонотипов TCRalpha между донорами	61
4.4.6. Анализ возможного вклада популяций Т-клеток со сниженным разнообразием .....	62
<b>4.5. Анализ последовательностей CDR3 клонов идентичных между донорами.....</b>	<b>64</b>
4.5.1. Анализ вероятности сборки идентичных клонотипов TCR .....	65
4.5.2. Анализ числа инсерций в идентичных клонотипах TCR.....	66
4.5.3. Анализ распределения идентичных между близнецовыми донорами клонотипов TCR по фенотипам.....	68
4.5.4. Некоторые следствия из гипотезы пренатального обмена Т-клетками .....	69
<b>4.6. Анализ распределения числа инсерций в репертуарах наивных клеток и клеток памяти .....</b>	<b>70</b>
4.6.1 Анализ распределения клонов предположительно эмбрионального происхождения в репертуаре в зависимости от возраста .....	76
4.6.2. Роль клонотипов с нулем инсерций в структуре репертуара .....	77
<b>4.7. Анализ реакции репертуара TCR монозиготных близнецов на мощный антигенный стимул .....</b>	<b>80</b>
4.7.1. Поиск клонотипов TCRbeta значимо увеличивающих численность в ответ на иммунизацию против желтой лихорадки .....	80
4.7.2. Анализ числа идентичных у разных доноров последовательностей TCR ВЖЛ-специфических клонов .....	81
4.7.3. Оценка числа идентичных нуклеотидных последовательностей TCRbeta с помощью вероятностной модели.....	82
4.7.4. Анализ сходства аминокислотных последовательностей вакциносpezifических клонотипов TCRbeta .....	84
<b>4.8. Практическое применение вероятностей сборки к анализу ранее опубликованных данных.....</b>	<b>86</b>
4.8.1. Оценка вероятности сборки TCR для оценки маркеров в задаче детекции минимальной остаточной болезни (МОБ).....	86
4.8.2. Оценка вероятности сборки TCR для поиска ассоциированных с заболеваниями вариантов TCR .....	87
<b>5. Заключение .....</b>	<b>89</b>
<b>6. Выводы .....</b>	<b>90</b>
<b>7. Благодарности .....</b>	<b>92</b>
<b>8. Список сокращений.....</b>	<b>93</b>
<b>9. Список литературы .....</b>	<b>94</b>

# 1. Введение

## **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 Т-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 Т-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 Т-клеток, Т-клеток памяти и охарактеризовать возрастные изменения в части репертуара, представленной эмбриональными клонотипами.
5. Исследовать особенности развития противовирусного иммунного Т-клеточного ответа на основе мониторинга изменений репертуара Т-лимфоцитов при вакцинации монозиготных близнецов живой вакциной против вируса желтой лихорадки.
6. Разработать систему биоинформатического анализа индивидуальных репертуаров TCR для выявления клонотипов Т-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 Т-клеток в систему адаптивного иммунитета взрослых доноров.

Впервые охарактеризован ряд принципиальных отличий между репертуарами наивных Т-клеток и Т-клеток памяти: показано, что наиболее высокопредставленные клонотипы репертуара наивных Т-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 T-клеточных рецепторов, к анализу опубликованных данных секвенирования TCR продемонстрирована при оценке информативности маркеров T-клеточных лейкозов и при поиске T-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 T-клеточного репертуара и развитии иммунного ответа монозиготных близнецов, способствуют уточнению современного представления о вкладе генетических факторов в формирование индивидуальных репертуаров T-лимфоцитов и могут служить основой изучения роли эмбриональных клонотипов в функционировании развитой иммунной системы. Так, установление фактов пренатального обмена клонами T-клеток и их длительного (десять лет) существования в репертуаре позволяет изучать продолжительность жизни клонов разных субпопуляций T-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 T-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 T-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 (SystImm), 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.

### 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, **Погорельый 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. Погорелый**, 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

## 2. Обзор литературы

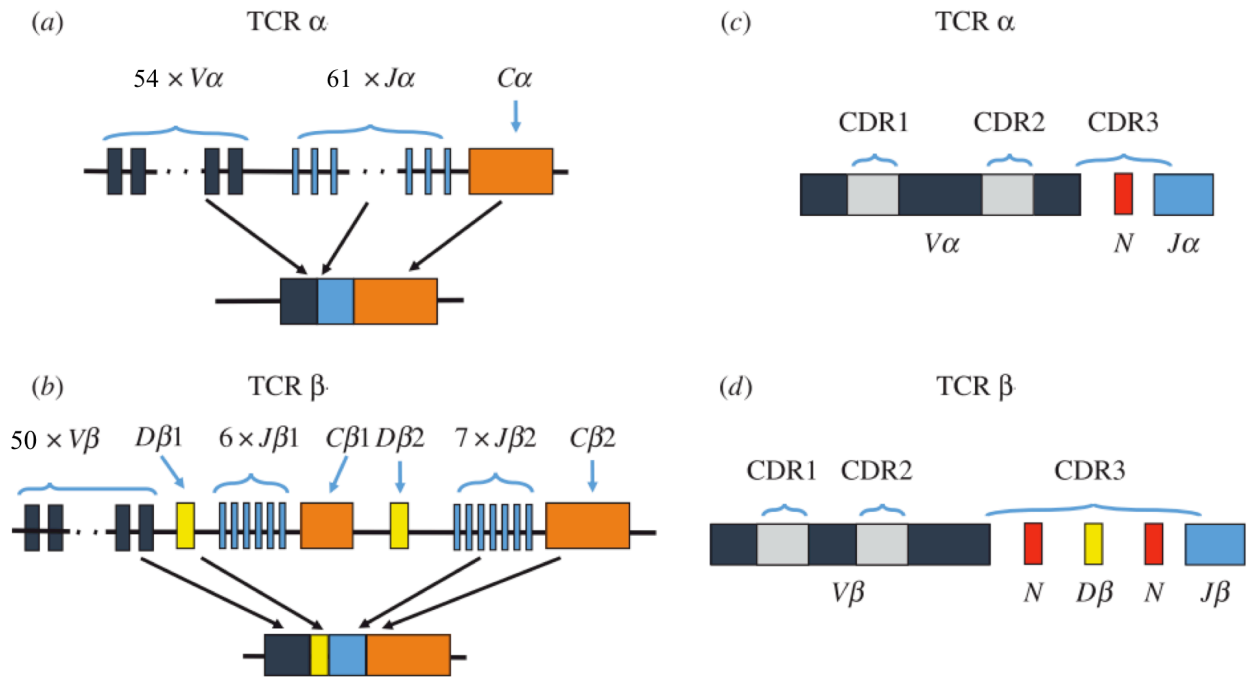
В первой части рассматривается формирование репертуара Т-клеточных рецепторов – последовательность событий при V(D)J-рекомбинации, во второй части обсуждается селекция Т-клеток в тимусе и ее вклад в формирование репертуара Т-клеток. Третья и четвертая части посвящена моделям рекомбинации и селекции соответственно. Пятая часть посвящена близнецовым моделям как способу выявить вклад генетических факторов в развитие репертуара. В шестой части рассматриваются некоторые особенности эмбрионального развития иммунной системы. В седьмой части кратко рассматривается динамика иммунного ответа на мощный антигенный стимул (на примере вакцинации от вируса желтой лихорадки) и подходы к исследованию иммунного ответа с помощью высокопроизводительного секвенирования репертуаров Т-клеточных рецепторов.

### 2.1. Формирование разнообразия TCR

TCR - гетеродимерные мембранные белки, состоящие из двух цепей: альфа и бета, или гамма и дельта. Соответствующие Т-клетки называют альфабета и гаммадельта Т-клетками. Принципы распознавания антигенов и роль гаммадельта Т-клеток в иммунном ответе сейчас активно изучаются [Vantourout, Hayday, 2013], альфа-бета TCR изучены гораздо лучше, и в дальнейшем речь пойдет именно о них.

TCR распознает короткие развернутые пептиды в комплексе с молекулами МНС (major histocompatibility complex). Структура комплекса МНС-пептид распознается как целое, и обе цепи TCR участвуют во взаимодействии [Rudolph, Stanfield, Wilson, 2006]. Последовательности генов альфа и бета цепей формируются при созревании тимоцитов в ходе процесса V(D)J рекомбинации. Сборка осуществляется из закодированных в геноме сегментов V (variable), D (diversity, в случае альфа цепей отсутствует), J (joining) и C (constant). У каждого типа сегментов (кроме C-сегментов альфа цепей) есть несколько различающихся вариантов, из которых для сборки рецептора выбирается один. Также сборка может сопровождаться вставками и делециями случайных нуклеотидов на стыках между сегментами. Этот процесс происходит независимо в разных лимфоцитах, что приводит к появлению большого разнообразия Т-клеток, каждая из которых несет свой тип альфа-бета TCR.

### 2.1.1. Структура локусов TRA и TRB, сегменты



**Рис 1.** Организация локусов TRA(a) и TRB(b) человека. Синим, желтым, голубым и оранжевым обозначены V, D, J и C сегменты соответственно (для каждого семейства сегментов указана численность). На (c) и (d) приведена структура зрелого гена альфа и бета-цепи TCR. Цит. по [Laydon, Bangham, Asquith, 2015] с изменениями.

Правильно перестроенный ген бета цепи имеет структуру V(N)D(N)JC, ген альфа цепи имеет структуру V(N)JC. Локусы, содержащие сегменты, которые используются для построения альфа (TRA локус) и бета (TRB локус) цепей TCR, находятся на 14 и 7 хромосоме соответственно (см. Рис. 1).

Сохраняя общий принцип организации и расположение на синтеничных участках двух разных хромосом локусы TRA(D) и TRB разных видов млекопитающих значительно отличаются по внутреннему строению. Отличия касаются, в основном, числа и взаиморасположения V сегментов, формирующихся в ходе эволюции генома путем внутривидовых дупликаций. Сравнение нуклеотидных последовательностей V сегментов обнаружило существование линии- и видоспецифических семейств V сегментов [Olivieri, Gambón-Deza, 2015]. При этом некоторые V сегменты могут содержать стоп-кодоны и нарушенные RSS участки, что может сказываться на их функциональности [Dean и др., 2015].

TRA locus расположен в 14q11.2 и перекрывается с локусом для дельта цепей гамма-дельта TCR. Это перекрытие существенно для однозначной дифференцировки T-клеток на альфа-бета и гамма-дельта типы. В составе локуса 54 варианта V-сегментов (TRAV), 61 вариант J-сегментов и единственный вариант C-сегмента (TRAC). Интересно, что пять TRAV могут использоваться при построении как альфа, так и дельта цепей [Nishana, Raghavan, 2012], [Lefranc, 2014]. Так же в локусах находятся транскрипционные энхансеры, удаление которых блокирует рекомбинацию [Bassing, Swat, Alt, 2002].

TRB locus находится в локусе 7q34 и содержит 67 TRBV сегментов (из них функциональных 50), 2 TRBD сегмента, 14 (из них 13 функциональных) TRBJ сегментов и 2 TRBC сегмента.

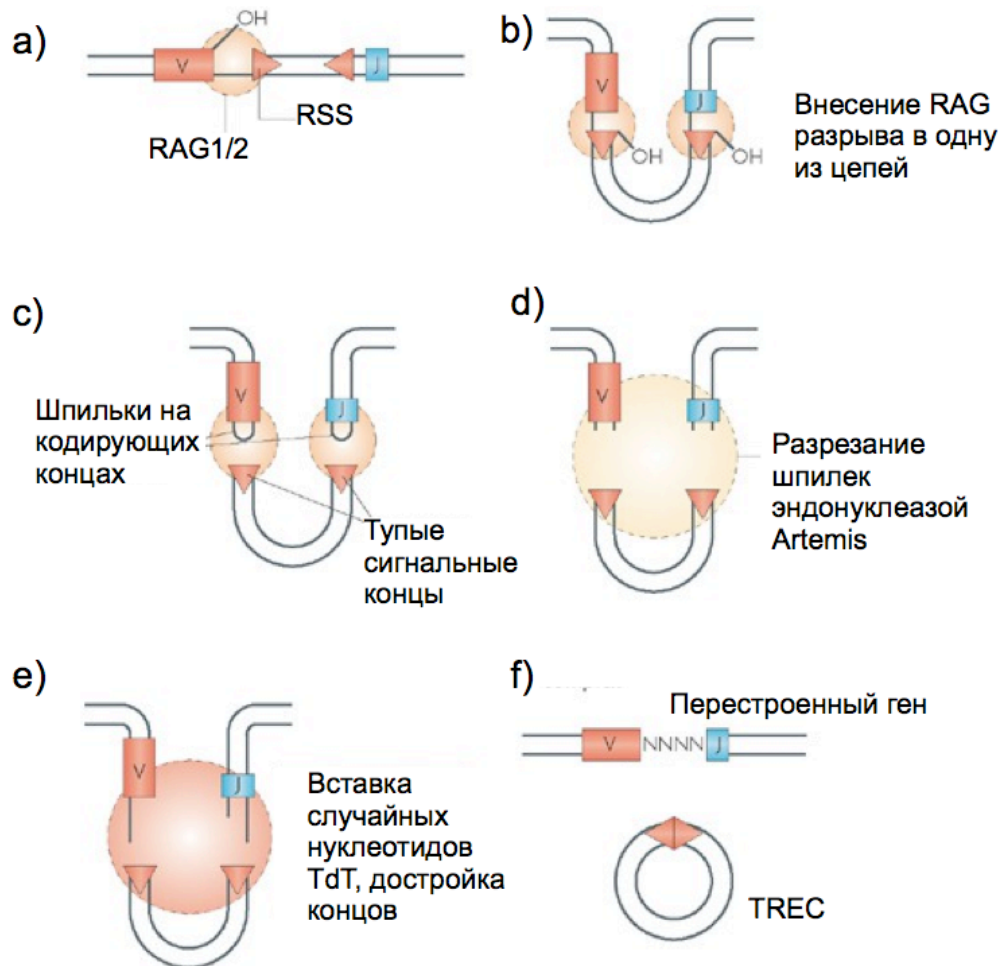
Количество различных комбинаций сегментов примерно равно для альфа и для бета цепей, при этом разнообразие бета-цепей существенно выше за счет большего количества случайных нуклеотидов, которые могут вставляться при сборке. Потенциальное разнообразие репертуара TCR, оцененное теоретически с помощью математической модели V(D)J рекомбинации составляет  $10^{14}$ - $10^{39}$  для бета цепи [Mora, Walczak, 2016; Murugan и др., 2012] и  $10^9$ - $10^{21}$  для альфа цепи [Elhanati и др., 2015a; Mora, Walczak, 2016].

Сборка гена происходит в ходе сайт-специфической рекомбинации, и все сегменты фланкированы узнаваемыми рекомбиназами RAG сигналами (RSS). Структура этих сигналов играет важную роль для корректной сборки генов бета и альфа цепей. Последовательности RSS консервативны, и состоят из трех частей: гептамера CACAGTG, нонамера ACAAAAACC и спейсера консервативной длины (12 или 23 нт.) между ними. Наличие спейсеров двух различных длин важно, RSS со спейсером длиной 12 будет сближаться и рекомбинировать только с RSS с другим спейсером, длиной 23. Эту закономерность называют “правилом 12/23”, она обеспечивает правильное направление рекомбинации. Так, в локусе TRA 23RSS находятся в 3'-концевых областях TRAV, а 12RSS лежат в 5' части TRAJ, и поэтому RSS различных TRAV взаимодействуют с RSS TRAJ, а не друг с другом, что приводило бы к формированию нефункциональных генов альфа-цепей.

### 2.1.2. Последовательность событий при сборке генов TCRbeta и TCRalpha

Центральную роль в сайт-специфической V(D)J рекомбинации играют белки RAG1 и RAG2. Пространственная структура комплекса DNA/RAG1-RAG2 и мол.механизм рекомбинации были определены совсем недавно [Kim и др., 2015; Ru и др., 2015].

Предполагается, что гены, кодирующие эти белки, происходят от древнего ДНК-транспозона [Agrawal, Eastman, Schatz, 1998]. Белковый комплекс RAG1 и RAG2 вносит разрыв точно между RSS и кодирующим сегментом (см. рис. 2a). Освободившийся 3'-ОН конец сегмента атакует фосфоэфирную связь комплементарной цепи, со стороны кодирующего сегмента образуется шпилька, а со стороны RSS образуются тупые концы (см. рис. 2c). Таким образом, в ДНК вносится двуцепочечный разрыв, который должен быть репарирован. Репарация происходит с помощью ферментов системы нехомологичного соединения концов NHEJ. Тупые концы, которые образуются со стороны RSS могут лигироваться сразу, с образованием небольших кольцевых ДНК TREC (T-cell receptor excision circles, см. рис 2f). Белки HMGB1 и HMGB2 (high mobility group protein B) способствуют сближению концов с сигнальными последовательностями.



**Рис. 2.** Последовательность событий при V(D)J-рекомбинации. a) – связывание RAG1/2 с RSS, внесение одноцепочечного разрыва b) - сближение рекомбинирующих участков, образование синатса c) – атака свободным 3'ОН фосфатной связи комплементарной цепи и образование шпилек (d-e) открытие шпилек и модификация концов f) – лигирование концов и образование TREC. Цит. по [Weill, Reynaud, 2008] с изменениями

Перед лигированием кодирующие сегменты концов происходит их модификация (см. рис. 2d-e). Шпильки разрезаются с помощью белка Artemis, который в комплексе с DNA-ПК (DNA-dependent protein kinase, ДНК-зависимая протеинкиназа) приобретает активность эндонуклеазы. Место, в которое вносится одноцепочечный разрыв на шпильке может варьировать, в результате чего могут образовываться выступающие концы, и более короткая цепь потом будет комплементарно достраиваться на матрице более длинной с образованием палиндромных вставок (P-вставок), свободные выступающие концы могут быть модифицированы терминальной дезокинуклеотидилтрансферазой, которая обеспечивает нематричный синтез коротких участков ДНК, N-вставок. Кроме N и P вставок возможна деградация концов с образованием делеций, вероятно, осуществляемая неизвестной экзонуклеазой [Bertocci и др., 2006]. Таким образом, перед лигированием кодирующие концы подвергаются существенной модификации, которая сильно увеличивает разнообразие TCR. В альфа-цепи эти процессы происходят на стыке V и J сегмента, в бета-цепях на стыке V-D и D-J, из-за чего теоретическое разнообразие бета-цепей существенно выше [Wang и др., 2010]. Лигирование осуществляется с помощью аппарата негомологичной рекомбинации концов: гетеродимером лигазы IV/XRCC4 при участии белка XLF и гетеродимера Ku70/Ku80 [Prochazkova, Loizou, 2016].

## 2.2. Положительная и отрицательная селекция TCR

Принципиальная задача V(D)J рекомбинации - получить большое разнообразие рецепторов, часть из которых будет специфично узнавать антигены, с которыми организм еще не встречался. С генерацией рецепторов неизвестной специфичности связано две проблемы:

- 1) могут сформироваться рецепторы, которые будут узнавать антигены собственного организма и вызывать аутоиммунную реакцию.
- 2) могут формироваться дефектные рецепторы, которые в принципе не способны узнавать антигены.

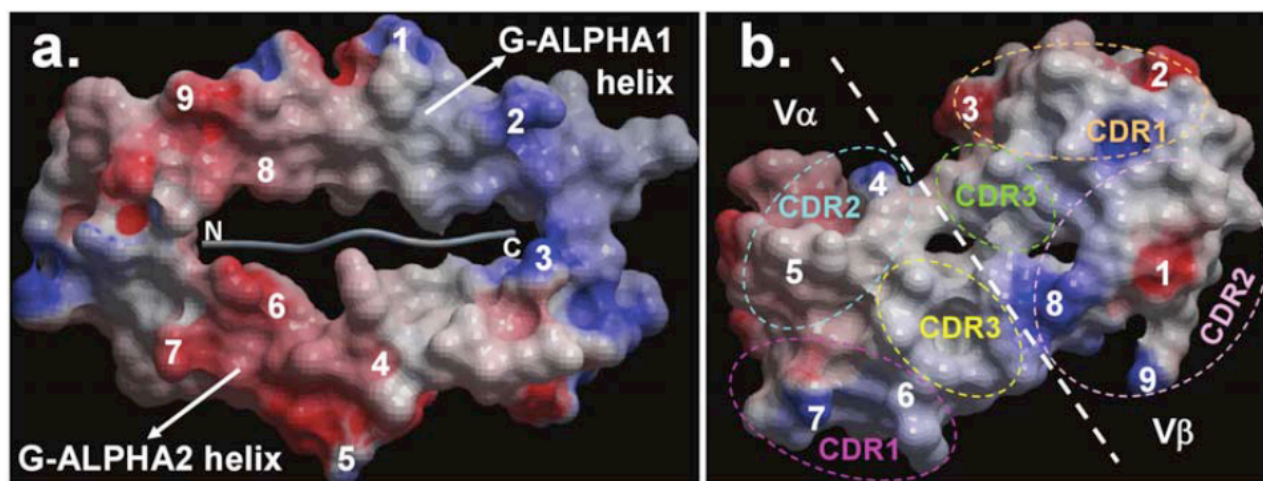
Обе эти задачи для TCR решаются с помощью селекции в тимусе.

TCR узнают короткие развернутые пептиды в комплексе с молекулами главного комплекса гистосовместимости (МНС), находящимися на клеточной мембране. Пептиды образуются при расщеплении белков синтезируемых в клетке, в том числе дефектных продуктов трансляции [Yewdell, Anton, Vennink, 1996]. Такие пептиды включаются в МНС класса I, и имеют стандартную длину 9-11 а.о. Этот процесс происходит почти во всех клетках организма. Профессиональные антигенпрезентирующие клетки (АПК): макрофаги и дендритные клетки, способны так же

презентировать эндоцитированные белки внеклеточного происхождения, при этом пептиды включаются в МНС класса II. Принципиальное отличие МНС II от МНС I заключается в длине включаемых пептидов - в первом случае щель, в которой лежит связанный пептид, открыта с двух сторон, и длина связываемого пептида варьирует от 13 до 25 а.о. [Wang и др., 2008]. Молекулы МНС I кодируются генами HLA-A, HLA-B и HLA-C, молекулы МНС II генами HLA-DRA, HLA-DRB, HLA-DQA, HLA-DQB, HLA-DPA, HLA-DPB. МНС - самый полиморфный ген в человеческой популяции, известно свыше 6000 аллельных вариантов для МНС I и 1500 для МНС II [Robinson и др., 2013]. Таким образом, в каждом человеке может быть представлено до 10 различных видов МНС II, и до 6 разновидностей МНС I. В полиморфности МНС есть биологический смысл, поскольку различные варианты отличаются по репертуару представляемых пептидов, а также по количеству различных пептидов, которые в принципе могут быть представлены [Kořmrlj и др., 2010].

Взаимодействие TCR-МНС опосредовано корецепторами Т-клеток, молекулами CD4 и CD8, TCR CD4+ лимфоцитов (Т-хелперов) специфически связывает МНС-II, а TCR CD8+ лимфоцитов (или Т-киллеров) – МНС-I [Wang, Reinherz, 2012].

TCR распознает комплекс МНС-пептид как целое, и долгое время было неясно, как может быть организовано специфическое белок-белковое взаимодействие между такими разнообразными группами молекул.



**Рис 3.** Участвующие в белок-белковых взаимодействиях поверхности МНС I (а.) и TCR (б.), вид сверху. Цифрами и цветом обозначены заряженные аминокислотные остатки, обеспечивающие TCR-МНС узнавание (друг с другом взаимодействуют области, обозначенные одной цифрой). [Khan, Ranganathan, 2011]

К настоящему времени с помощью рентгеноструктурного анализа решены структуры более сотни комплексов TCR-pMHC (TCR-peptide-MHC) [Kass, Buckle, Borg, 2014]. Части TCR, взаимодействующие с MHC называют CDR (complementarity determining region), их на каждой из цепей выделяют три (см. Рис 3.). CDR1 и CDR2 кодируются последовательностью V-сегмента и отвечают, в основном, за взаимодействие с MHC [Rudolph, Stanfield, Wilson, 2006]. Последовательности CDR3 гораздо более разнообразны, поскольку кодируются на стыке между сегментами, где происходят вставки и делеции случайных нуклеотидов, считается, что эта часть взаимодействует с презентуемым пептидом. TCR в комплексе ориентирован по диагонали для MHC-I, а для MHC-II – почти параллельно [Rossjohn и др., 2015] относительно пептид-связывающей щели, при этом бета-цепь взаимодействует с C-концом пептида и альфа-2 спиральным доменом MHC I или бета-1 спиралью для MHC II, а альфа цепь с N-концом и альфа-1 спиральными доменами для MHC обоих классов [Garcia и др., 2009; Josephs, Grant, Gras, 2017].

Такая стандартная ориентация белков в комплексе TCR-pMHC может быть результатом селекции TCR, либо генетически закодированных в последовательностях V-сегментов и MHC консервативных взаимодействующих участков. Сейчас появляется все больше свидетельств в пользу второй гипотезы, с той поправкой, что взаимодействия кодируемых V-сегментами частей TCR с MHC не универсальны. Каждый MHC может взаимодействовать с различными V-сегментами, используя разные структурные мотивы, и обратное тоже верно - V-сегменты используют разные участки для взаимодействия с разными MHC. При этом показано, что эти взаимодействия специфичны, TCR, полученные в организме мышей с нарушенной системой селекции, ориентируются относительно MHC так же, как прошедшие селекцию [Dai, Huseby, Rubtsova, 2008]. Предполагается существование набора структурных «кодонов» - специфических парных взаимодействий отдельных участков кодируемых V-сегментами частей TCR и участков MHC, и разные «кодоны» могут использоваться в разных парах TCR-MHC, что и делает взаимодействие двух исключительно разнообразных белковых комплексов возможным [Garcia и др., 2009]. Эта система взаимодействий гибкая, было показано, что изменение последовательности CDR3 может существенно менять набор взаимодействующих с MHC участков [Deng и др., 2012].

CDR3 является самой вариабельной частью TCR и взаимодействует с презентуемыми MHC пептидами. Было показано, что это взаимодействие специфично, т.е. внесение точечных аминокислотных замен в пептид препятствует узнаванию [Huseby и др., 2006]. Удивительно, что при этом взаимодействие вырожденное, отдельный TCR может узнавать несколько сильно отличающихся по последовательности пептидов [Borbulevych и др., 2009]. Предполагается, что оба этих эффекта являются следствием работы системы селекции в тимусе. Благодаря работе



транскрипционного фактора AIRE, в тимусе экспрессируется большинство генов, продукты которых дальше нарезаются на короткие фрагменты, которые презентуются специальными типами клеток CD4+CD8+ лимфоцитам [Murphy, Weaver, 2016]. CD4+CD8+ лимфоциты экспрессируют корцепторы CD4 и CD8, обеспечивающие связывание TCR с MHC II и MHC I соответственно. Для того, чтобы пройти селекцию, TCR должен соответствовать двум условиям: он должен провзаимодействовать хотя бы с одним комплексом pMHC с некоторой достаточной силой (положительная селекция), и он не должен взаимодействовать ни с каким из pMHC слишком сильно (негативная селекция). Первое условие необходимо, чтобы отфильтровать TCR, которые плохо взаимодействуют с MHC, а значит, не смогут распознать и комплекс MHC с пептидом вирусного или бактериального происхождения. Второе условие защищает от потенциально аутореактивных Т-клеток, так как порог негативной селекции близок к порогу активации Т-клетки [Daniels и др., 2006]. Была предложена простая математическая модель селекции TCR, объясняющая роль негативной селекции в формировании специфичных, но в то же время вырожденных взаимодействий TCR с пептидами патогенных организмов [Košmrlj, Jha, 2008].

Энергия взаимодействия TCR с pMHC представлена в модели суммой постоянной компоненты  $E_c$ , отражающей взаимодействие кодируемых V-сегментами CDR1 и CDR2 TCR с MHC, и суммой парных взаимодействий аминокислот гипервариабельного CDR3 TCR и пептида:

$$E = E_c + \sum_{i=1}^N J(l_i, j_i) \quad (1)$$

Энергия взаимодействия пары любых аминокислотных остатков  $J(l_i, j_i)$  в модели описывается матрицей Миязавы-Джернигана [Miyazawa, Jernigan, 1996]. Особенность этой матрицы в том, что можно выделить аминокислотные остатки, которые дают сильные взаимодействия с любыми другими (такие как L, F, I, M, V), и наоборот, те, которые довольно слабо связываются со всеми остальными (например, R, S, N). После *in silico* селекции репертуар CDR3 TCR обогащался аминокислотными остатками, которые слабо взаимодействуют с любыми другими аминокислотными остатками. Участки CDR3, в последовательностях которых присутствовали сильновзаимодействующие по Миязаве-Джернигану аминокислоты, отфильтровывались, поскольку при достаточном количестве презентуемых пептидов в тимусе найдется такой, с которым произойдет взаимодействие достаточное для отрицательной селекции. Если в тимусе презентуется мало пептидов, то отрицательная селекция CDR3 с несколькими сильновзаимодействующими аминокислотными остатками может не произойти, и в этом случае образуется неспецифичный TCR, узнавание которым антигенного пептида обеспечено

несколькими сильными взаимодействиями, а остальные аминокислотные остатки могут быть заменены. Образование таких кроссреактивных TCR ранее наблюдалось у линии мышей с единственным презентруемым в тимусе пептидом [Huseby и др., 2006]. Интересно, что с аллелями MHC I, презентрующими минимально разнообразный репертуар пептидов, ассоциирована устойчивость к развитию СПИДа у ВИЧ-серопозитивных индивидов [Emi и др., 2008]. Выдвинута гипотеза, что эта устойчивость связана с формированием менее специфичных TCR, которые успешно распознают пептиды с заменами, возникающими из-за неточной работы вирусных полимераз [Kořmrlj и др., 2010].

### 2.3. Моделирование процесса сборки T-клеточного рецептора

В разделе 2.1 мы описали последовательность событий сборки T-клеточного рецептора. Отдельные части этого процесса можно моделировать *in silico*, и получать ценные биологические выводы, если модель составлена адекватно.

Один из сложных вопросов сборки T-клеточного рецептора - причины неравномерного использования V и J сегментов - одни сегменты используются гораздо чаще других [Livak и др., 2000]. Один из гипотетических факторов, определяющий частоту использования сегментов - особенности конформации хроматина, в результате которой одни сегменты оказываются менее доступны для белков рекомбинации, чем другие. В работе [Ndifon и др., 2012] механическая модель хроматина была использована для предсказания частот J-сегментов, образующих пары с сегментом TRBD1. Хроматин представлен в модели как изогнутая упругая гладкая нить (такую модель в химии высокомолекулярных соединений называют персистентной), для которой задана гибкость и радиус кривизны. Частота комбинации сегментов (в данном случае TRBD1-J) определяется частотой случайных столкновений соответствующих участков, на которую влияет расстояние между участками в локусе. Полученная модель удовлетворительно ( $R^2=0.73$ ,  $p<0.01$ ) предсказывала частоты J-сегментов в репертуаре мыши и человека, притом в нефункциональных и функциональных клонотипах частоты оказывались схожими. Однако для V-сегментов аналогичная модель не сработала совсем, что авторы связывают с более сложной топологией части локуса, где находятся V-сегменты. В этой же работе экспериментально показана независимость спаривания V-сегмента с парой D и J сегментов для бета-цепи.

Рекомбинация бета-цепи отличается от рекомбинации альфа-цепи тем, что может происходить только один раз на каждой хромосоме. В альфа-цепи на каждой хромосоме может происходить много раундов рекомбинации. Из-за отсутствия аллельного исключения несобранных

генов альфа-цепи это часто приводит к формированию Т-клеток с одной бета-цепью и двумя разными альфа-цепями. На одной цепи каждый V и J сегмент может участвовать в рекомбинации только один раз, поскольку в ходе рекомбинации удаляются сигнальные последовательности RSS. Рекомбинация может продолжиться с участием более внешней пары сегментов. Этот процесс моделировался в работе [Warmflash, Dinner, 2006]. В модели предполагается, что в рекомбинации могут принимать участие V и J сегменты, находящиеся на небольшом удалении друг от друга (в окнах фиксированного размера в правой и левой частях кассет V- и J-сегментов соответственно). Авторами предложена комбинаторная формула, определяющая частоту каждой комбинации сегментов. В целом последовательная многократная рекомбинация приводит к тому, что пары V и J сегментов выбираются для рекомбинации неслучайно - частота наблюдения пары дистальных или пары проксимальных сегментов выше, чем пар дистального и проксимального. Экспериментально этот факт был показан в работе [Elhanati и др., 2015a] на данных из работы [Zvyagin и др., 2014].

Все эти особенности рекомбинации (неравномерное использование сегментов, зависимость выбора V и J сегментов альфа-цепи, а также D и J сегментов бета-цепи) должны быть учтены при разработке механистической модели сборки генов TCR. Модели, описанные выше предсказывали определенный аспект рекомбинации: частоты отдельных сегментов, зависимость выбора V и J сегментов при многократной рекомбинации в альфа-локусе и т.д. Если учесть все остальные события сборки, то полученная модель может использоваться для симуляции как отдельных последовательностей TCR, так и целых репертуаров, а также для расчета теоретических частот сборки различных вариантов TCR. Впервые такая модель для бета-цепи Т-клеточного рецептора была предложена в работе [Murugan и др., 2012], для альфа-цепи аналогичная модель описана в [Elhanati и др., 2015a], для В-клеточного рецептора в работе [Elhanati и др., 2015b].

Основное понятие этой модели - это представление о *сценарии сборки*, то есть конкретной комбинации событий V(D)J рекомбинации: выбора V-, D-, J-сегментов, определенного числа N и P-инсерций и делеций. Вероятность сценария сборки рассчитывается как произведение вероятностей этих отдельных событий. Отметим, что есть взаимно однозначное соответствие между сценарием сборки и разметкой нуклеотидной последовательности TCR с выделением нуклеотидов, закодированных в V-сегменте, N- и P-вставок, D-сегмента и J-сегмента. Важная часть модели - это определение того, какие события сборки являются независимыми друг от друга В работе была подтверждена независимость выбора V-сегмента и комбинации DJ, и зависимость выбора D и J сегментов. Число делеций и P-инсерций (которые в этой модели рассматриваются как делеции отрицательной длины) оказалось зависимым от V и J сегмента, а число N-инсерций нет.

Последовательность инсерций также оказалась неслучайной - вероятность вставки конкретного нуклеотида зависит от предыдущего нуклеотида.

Сценарий сборки однозначно задает нуклеотидную последовательность рецептора, и вероятность каждого сценария легко рассчитывается. Но при этом для любой наблюдаемой в эксперименте последовательности реальный сценарий сборки неизвестен, и одной нуклеотидной последовательности рецептора могут отвечать тысячи возможных сценариев. Поскольку механизм V(D)J рекомбинации включает делеции и вставку произвольного числа нуклеотидов, то произвольная нуклеотидная последовательность может быть собрана из любых сегментов, в худшем случае потребуется удалить полную последовательность сегментов и затем вставить нужную последовательность N-инсерций. Но вероятность такой сборки будет очень низкой, как минимум из-за того, что все вставленные N-инсерции должны совпасть с последовательностью существующего в геноме сегмента. Таким образом вероятность сборки любой наблюдаемой нуклеотидной последовательности TCR представляет собой интересную вычислительную проблему: необходимо перебрать и просуммировать вероятности всех сценариев сборки, которые приводят к данной последовательности. В работе [Muguan и др., 2012] это реализуется простым перебором сценариев, в работе [Elhanati и др., 2015a] используются более эффективные алгоритмы из области скрытых Марковских моделей (НММ, Hidden Markov models).

Пока мы рассмотрели расчет вероятности сборки рекомбинационного сценария и нуклеотидной последовательности для случая, когда параметры модели сборки (вероятность выбора каждого V-сегмента, вероятность определенного числа делеций при условии выбора этого V-сегмента и пр.) известны. Отдельную проблему представляет вывод этих параметров из реальных данных секвенирования, то есть набора нуклеотидных последовательностей TCR, для которых истинные сценарии сборки неизвестны. Кроме сборки на репертуар влияет селекция лимфоцитов в тимусе и на периферии, поэтому вывод параметров модели осуществляют на нефункциональных клонотипах, содержащих сдвиг рамки считывания или стоп кодон внутри гипервариабельного участка CDR3.

Если бы был размеченный набор последовательностей для обучения (то есть для каждой последовательности TCR был бы известен точный сценарий сборки), то задача вывода параметров модели сборки была бы тривиальной. Например, оценка вероятности трех делеций в сегменте TRBV9 совпала бы с долей сценариев с тремя делециями от всех сценариев с TRBV9, и так далее для всех остальных параметров. Истинной разметки для настоящих экспериментальных данных у нас нет, но глядя на конкретную нуклеотидную последовательность можно, задавшись определенными параметрами модели сборки, сказать, какая вероятность у каждой из возможных

разметок. Далее для каждого события модели сборки (таких как выбор V-сегмента, либо 3'-делеции при условии выбора конкретного V-сегмента) считается сумма вероятностей сценариев, в которых данное событие происходит, это дает новое значение этого параметра модели сборки. То есть, чтобы вывести параметры модели сборки, нужно сначала взвесить по вероятности возможные сценарии, а для этого необходимо знать параметры модели сборки. Таким образом, процесс происходит итеративно, в самом начале можно выбрать равномерное распределение для параметров модели, получить новые параметры и использовать их для следующей итерации алгоритма. С каждой итерацией возрастает правдоподобие (произведение полных вероятностей сборки) для данных обучающей выборки, и эту статистику можно использовать для остановки выполнения алгоритма – когда значение правдоподобия меняется незначительно между итерациями, найдено оптимальное значение всех параметров модели сборки. Описанный алгоритм называется EM-алгоритмом, его реализация предложена в [Mugan и др., 2012] и требует перебора всех сценариев сборки для каждой последовательности. В работе [Elhanati и др., 2015a] для обучения модели используется более эффективный алгоритм Баума-Велша, обычно применяемый для обучения скрытых Марковских моделей.

## 2.4. Моделирование селекции репертуара TCR

Модели сборки, рассмотренные в предыдущей главе, удовлетворительно описывают число общих клонотипов в нефункциональном репертуаре TCR (состоящем из последовательностей со сбитой рамкой считывания или стоп-кодоном внутри).

Эти последовательности можно наблюдать в периферической крови, поскольку в несущих их Т-клетках есть прошедший положительную и отрицательную селекцию в тимусе нормальный аллель TCR. Большое количество Т-клеток погибает в тимусе, поскольку их рецепторы не могут пройти селекцию [McDonald и др., 2015]. Это значит, что вероятность сборки рецептора должна быть систематически меньше, чем вероятность встретить этот рецептор в прошедшем селекцию репертуаре, поскольку объем последнего гораздо меньше, чем объем репертуара TCR до селекции. Поэтому для расчета  $P_{\text{post}}$ , вероятности найти рецептор в селектированном репертуаре, вводится поправка  $Q$  к вероятности сборки  $P_{\text{pre}}$ :

$$Q(\vec{\tau}, V, J) = \frac{P_{\text{post}}(\vec{\tau}, V, J)}{P_{\text{pre}}(\vec{\tau}, V, J)} = \frac{1}{Z} q_L q_{VJ} \prod_{i=1}^L q_{i;L}(a_i), \quad (2)$$

где  $Z$ -нормализационная константа (чтобы интеграл от  $P_{\text{post}}$  по всем возможным последовательностям TCR был равен единице),  $q_L$ - фактор селекции длин,  $q_{VJ}$ - фактор селекции VJ-комбинации,  $q_{i:L}(a_i)$  – фактор селекции аминокислоты  $a$  в положении  $i$  для CDR3 длиной  $L$ .

В работе [Elhanati и др., 2014], поправка  $Q$  рассчитывается в зависимости от аминокислотной последовательности CDR3, она является произведением отдельных факторов  $q$ , которые описывают влияние определенной аминокислоты в определенном положении CDR3aa, длины CDR3aa, VJ-комбинации на селекцию. Известно, что некоторые аминокислоты (например, цистеин) практически никогда не встречаются внутри CDR3aa участка, хотя соответствующие им кодоны вполне могут возникнуть при VDJ рекомбинации. Предполагается, что это происходит из-за того, что включение цистеина внутрь CDR3aa приведет к образованию дисульфидных связей и неправильному сворачиванию белка, такой рецептор не пройдет селекцию в тимусе. Поэтому большинство  $q$  факторов для цистеина очень малы, наличие этой аминокислоты в CDR3 сводит шансы TCR пройти селекцию к минимуму. Оценка всех  $q$ -факторов производится методом градиентного спуска, для оценки необходимо предварительно вывести модель сборки для данного донора, так как обучающийся алгоритм сравнивает симулированные с помощью модели сборки функциональные последовательности TCR с функциональными последовательностями из репертуара конкретного донора. Интересно, что авторами была обнаружена значимая корреляция вероятности сборки последовательности TCR  $P_{\text{gen}}$  с  $Q$ , то есть чаще генерируемые последовательности имеют большую вероятность пройти селекцию. Авторы выдвигают гипотезу, что параметры машинерии сборки были оптимизированы в ходе эволюции, чтобы производить меньше плохо сворачивающихся или узнающих собственные пептиды последовательностей.

## 2.5. Особенности адаптивного иммунитета близнецов

Монозиготные близнецы очень интересный объект для исследования репертуаров адаптивной иммунной системы по нескольким причинам. Во-первых, у них идентичные геномы. Из этого следует, что у них одинаковый набор МНС, ключевых молекул системы презентации антигенов. МНС кодируются самыми полиморфными генами в человеческой популяции и в существенной степени определяют репертуар Т-клеток. Это демонстрируют классические эксперименты по совместному культивированию лимфоцитов разных доноров (MLR - mixed lymphocyte culture). Оказывается, что в среднем 1-10% всех Т-клеток одного донора распознают комплексы рМНС на клетках другого донора [D'Orsogna и др., 2012; Emerson и др., 2014]. Вопрос совместимости наборов МНС между донорами является ключевым в трансплантации любых

органов и клеток (за редкими исключениями вроде слабо экспрессирующих МНС эритроцитов и тромбоцитов [Schenkel-Brunner, 2013]). МНС имеет решающее значение в формировании репертуара в ходе селекции, как в тимусе так и на периферии. Но кроме МНС, у близнецов идентичны и все остальные гены, которые могут иметь значения при сборке TCR (например, у близнецов идентичны аллели V-сегментов), а также процессинге антигенов (ERAAP, TAP и другие).

Во-вторых - это идентичная или близкая окружающая среда. При совместном проживании близнецы сталкиваются с идентичными антигенами, что может приводить к клональной экспансии идентичных вариантов T-клеточных рецепторов.

В-третьих - особенности эмбрионального развития близнецов. В большинстве случаев монозиготные близнецы присоединены к одной плаценте, что может приводить к обмену кровью между двумя плодами (но не с кровеносной системой матери).

В следующих разделах мы разберем, что удалось установить к настоящему моменту по каждой из перечисленных особенностей близнецовых доноров.

### 2.5.1. Сходство механизмов сборки и селекции TCR близнецов

Для того чтобы оценить особенности сборки T-клеточных рецепторов, необходимо исследовать репертуар до селекции в тимусе. Самый простой способ это сделать, сосредоточиться на нефункциональном репертуаре TCR, который содержит последовательности со стоп-кодоном и/или сбитой рамкой считывания.

В работе [Zvyagin и др., 2014] сравнивались функциональные и нефункциональные репертуары трех пар монозиготных близнецов. Оказалось, что нефункциональные репертуары близнецов имеют схожие частоты V-сегментов бета-цепи, что можно оценить, используя дивергенцию Дженсена-Шеннона, симметричную меру расстояния между дискретными распределениями, широко используемую в теории информации. Схожий результат был показан в работе [Rubelt и др., 2016], но использовались другие меры расстояния между распределениями V-сегментов - корреляция Пирсона и введенный авторами индекс RDI (repertoire dissimilarity index, индекс различия репертуаров). Авторы показали большее сходство и некоторых других событий сборки между близнецами, к примеру распределение числа N-инсерций. В то же время, анализ функционального репертуара T-клеточных рецепторов также выявляет сходство между распределениями V-сегментов репертуаров близнецов, и эта тенденция выражена даже сильнее, чем при сравнении нефункциональных репертуаров. Так, в [Zvyagin и др., 2014] показано, что не

только TRBV (как в нефункциональных клонотипах), но также распределения TRAV и TRAJ более близки у функциональных клонотипов монозиготных близнецов. Т.е. машинерия сборки близнецов настроена одинаково, и этот эффект усиливается в результате селекции более схожих Т-клеточных рецепторов в тимусе. По всей видимости усиление сходства связано с идентичной системой презентации: у близнецов одинаковые аллели МНС, и идентичные геномы, а значит и одинаковый репертуар презентруемых в тимусе и на периферии пептидов. Интересно отметить, что к последним относятся и пептиды бактерий комменсалов – микробиоты кишечника и кожи. Их влияние на иммунную систему и организм велико, но только начинает изучаться. Так было показано, что пересадка кишечной микробиоты близнецов дискордантных по ожирению приводит к изменению метаболизма и ожирению в мышинной модели [Ridauga и др., 2014]. В другом близнецовом исследовании было показано, что состав микробных сообществ кишечника частично определяется генетическими факторами [Goodrich и др., 2016]. И наконец, накапливаются свидетельства сильного Т-клеточного иммунного ответа на кишечную микробиоту при различных заболеваниях [Hand и др., 2012].

### 2.5.2. Оценка вклада окружающей среды в структуру репертуара

После сборки и селекции в тимусе наивные Т-клетки могут встретиться со своими антигенами. Узнавшие антигенные пептиды клоны проходят клональную экспансию и меняют наивный фенотип на фенотип эффекторных клеток или клеток памяти. В работе [Zvyagin и др., 2014], обнаружилось большее количество идентичных клонотипов между репертуарами близнецов среди самых высокопредставленных (т.е. клонов, имеющих наибольшее число клеток) клонов в репертуаре Т-клеток. Авторы предполагают, что высокая представленность этих общих клонотипов - результат клональной экспансии в ответ на одни и те же антигены. Аналогичные результаты были получены в работе [Qi и др., 2016], где изучался репертуар Т-клеток специфичных к вакцине против ветряной оспы, *Varicella zoster*. Перед секвенированием Т-клетки несколько дней инкубировали с лизатом зараженных вирусом клеток, что приводило к размножению и отбору специфично узнающих вирусные пептиды Т-клеток. Оказалось, что у близнецов больше идентичных клонотипов отвечают на вирус, что может объяснить большее количество общих клонов среди высокопредставленных клонотипов, которое наблюдали в работе [Zvyagin и др., 2014].

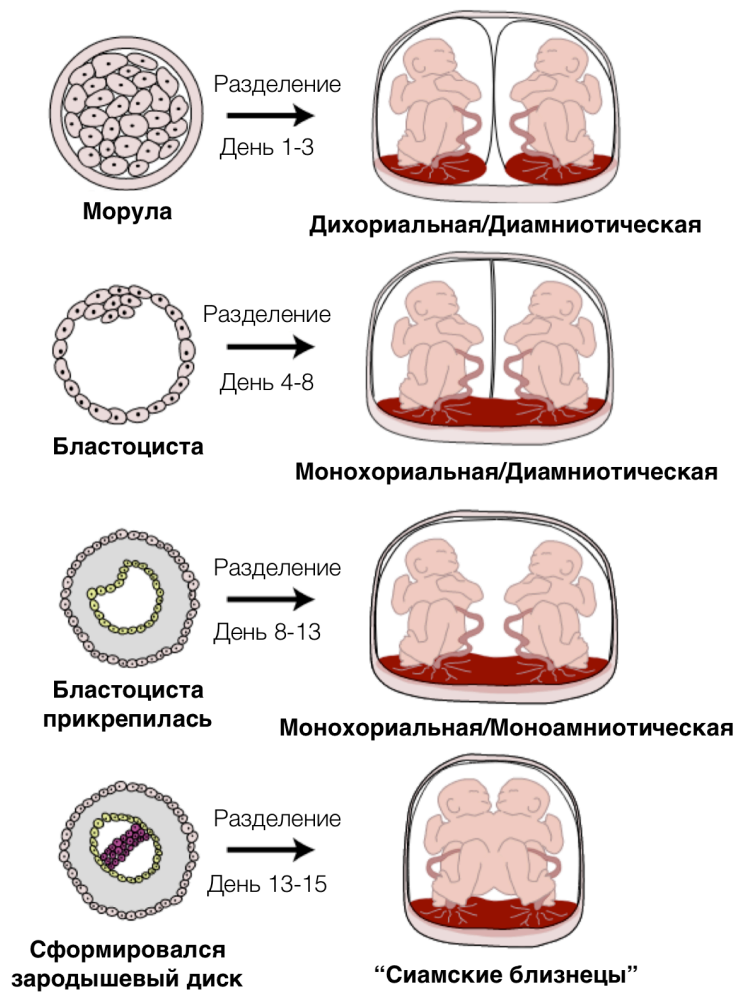
Таким образом большая схожесть репертуаров близнецов проявляется на всех этапах - от сборки и селекции до иммунного ответа на конкретные антигены. При этом общее число



идентичных аминокислотных последовательностей TCR у близнецов незначительно выше, чем между репертуарами неродственных доноров [Zvyagin и др., 2014], то есть репертуар каждого близнеца в большой степени уникален, как и репертуар каждого человека. Одно из возможных практических применений этой информации возможно в идентификации личности: близнецов невозможно отличить по каким бы то ни было полиморфным маркерам в геноме, в то время как набор клонов Т-клеток позволит однозначно установить их личности.

### 2.5.3. Особенности эмбрионального развития близнецов

Монозиготные близнецы образуются при разделении эмбриона на одной из стадий развития [Dias, Thilaganathan, Bhide, 2012]. От момента деления зависят особенности внутриутробного развития монозиготных близнецов (см. рис. 4). Если деление происходит на стадии двух бластомеров, то у каждого плода будет своя плацента и свой амнион (т.н. диамниотическая дихориальная беременность). В большинстве случаев деление происходит позже, на стадии ранней бластоцисты. В этом случае хорион (в том числе плацента) у плодов общий, а амниона два (диамниотическая монохориальная беременность). В редких случаях деление происходит позже, на стадии двухслойного эмбрионального диска. В этом случае у близнецов общий амнион и хорион (моноамниотическая монохориальная беременность). И наконец, деление на более поздних стадиях может приводить к формированию «сиамских» близнецов и другим подобным патологиям.



**Рис. 4** Типы близнецовых беременностей в зависимости от стадии на которой разделяется эмбрион. Пояснения в тексте. Цит. <https://en.wikipedia.org/wiki/Twin> с изменениями.

Во всех вариантах монохориальных беременностей описаны осложнения, связанные с кровотоком через общую плаценту. Достаточно распространен синдром фето-фетальной трансфузии, вызванный неравномерным распределением крови между плодами. Причиной фето-фетальной трансфузии является образование анастомозов между кровеносными системами плодов. Наличие этих анастомозов является нормальным и характерно для 80-90 процентов монохориальных беременностей [Lewi, Deprest, Hecher, 2013], но при их неудачной конфигурации или большом количестве может происходить неравномерное распределение крови между кровеносными системами эмбрионов. Наличие анастомозов показано экспериментально в опытах с инъекциями красителей разных цвета в соответствующие двум плодам части плаценты [Lorigioe и др., 2011]. В то же время, ни в одной из опубликованных ранее работ не рассматривается возможный вклад пренатального обмена кровью в развитие адаптивного иммунитета близнецов.

Также имеются некоторые данные о случаях микрохимеризма между организмами матери и плода [Jonsson и др., 2008]. Однако, в работе [Putintseva и др., 2013] было показано, что вклад этого обмена в репертуар Т-клеточных рецепторов незначителен - большая часть идентичных клонотипов появляется в результате независимой конвергентной рекомбинации. Тем не менее, оказывается, что можно однозначно установить пару мать-ребенок, хорошим предсказывающим фактором оказывается доля пересечений аминокислотных последовательностей CDR1+CDR2+CDR3 от пересечений CDR3, та же характеристика, которая отличала близнецовые пары в работе [Zvyagin и др., 2014].

## 2.6. Особенности раннего развития адаптивной иммунной системы

Итак, идентичные близнецы отличаются особенностями эмбрионального развития - в большей части случаев их кровеносные системы сообщаются между собой. В этом разделе приводятся краткие сведения об онтогенезе иммунной системы в эмбриональном развитии.

Первые В-лимфоциты обнаруживаются в печени плода (которая является важным эмбриональным кроветворным органом) примерно на 8ой неделе беременности, а к 12й неделе они начинают циркулировать в кровеносной системе [Mold и др., 2008]. Т-лимфоциты развиваются немного медленнее - к девятой неделе они начинают мигрировать в тимус эмбриона и обнаруживаются в кровотоке с 15й недели беременности. Интересно, что репертуар антигенраспознающих рецепторов первых лимфоцитов плода по всей видимости отличается от репертуара лимфоцитов, которые генерируются в конце беременности и после рождения: в частности показано что отличаются частоты использования V-сегментов [Rechavi и др., 2015]. Кроме того, до 19 недели практически неактивна терминальная дезоксирибонуклеотидил трансфераза (TdT), фермент отвечающий за формирование N-инсерций [George, Schroeder, 1992]. Таким образом, разнообразие раннего репертуара лимфоцитов почти полностью обеспечивается комбинированием геномных сегментов и делециями на стыках между ними. Функциональный смысл такого снижения разнообразия точно неизвестен. Интересно, что в модельных экспериментах на трансгенных мышах ранняя активация TdT приводила к ослабленному иммунному ответу на бактерию *Streptococcus pneumoniae* [Benedict и др., 2000]. Авторы предполагают, что сниженное разнообразие эмбрионального репертуара может быть важно для формирования ответа на определенные распространенные иммунные стимулы. Интересно, что кроме особенного репертуара TCR ранние лимфоциты имеют и фенотипические отличия от поздних. В работе [Mold и др., 2010] было показано, что лимфоциты плода можно отличить от

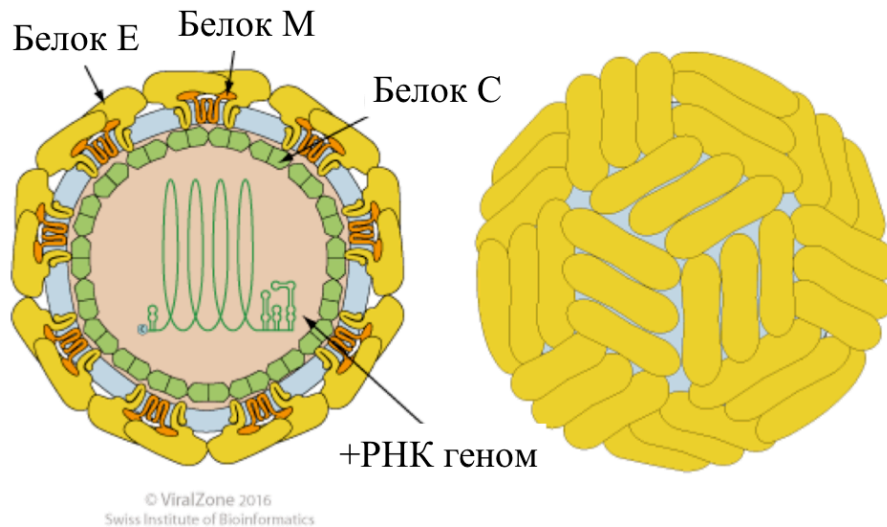
наивных лимфоцитов взрослого по экспрессии определенных генов, а также по пролиферативной активности. Авторы объясняют эти различия происхождением фетальных лимфоцитов и большинства лимфоцитов взрослого из разных популяций кроветворных клеток: на ранних этапах источником лимфоцитов плода является печень, а затем костный мозг, как и у взрослого человека.

Таким образом, формирование адаптивной иммунной системы происходит в несколько этапов: в районе 11 недели [Tavian, Réault, 2005] происходит переключение с эмбрионального репертуара Т-клеток, источником которых служит печень и которые не имеют N-вставок к лимфоцитам производимым в костном мозгу и имеющим схожий с нормальным репертуар.

## 2.7. Мониторинг противовирусного иммунитета

### 2.7.1. Вирус желтой лихорадки (YFV)

Желтая лихорадка (YFV) относится к *Flaviviridae*, куда так же входят гепатит С (HCV), лихорадка Денге (DENV), вирус Зика (ZIKV) и вирус клещевого энцефалита (TBEV). Это +РНК вирус, имеющий липидную оболочку и среднего размера (около 10 кб) геном. Липидная оболочка модифицирована димерами вирусного белка Е и белком М, под ней находится икосаэрический капсид из С-белка (см. рис 5). Вирусная частица в диаметре около 50 нм.



**Рис. 5.** Вирсион желтой лихорадки, слева – в разрезе, показаны структурные белки, справа – вид сбоку, видна  $T=3$ -подобная икосаэрическая пространственная организация димеров белка Е. Цит. по [viralzone.expasy.org](http://viralzone.expasy.org) с изменениями.

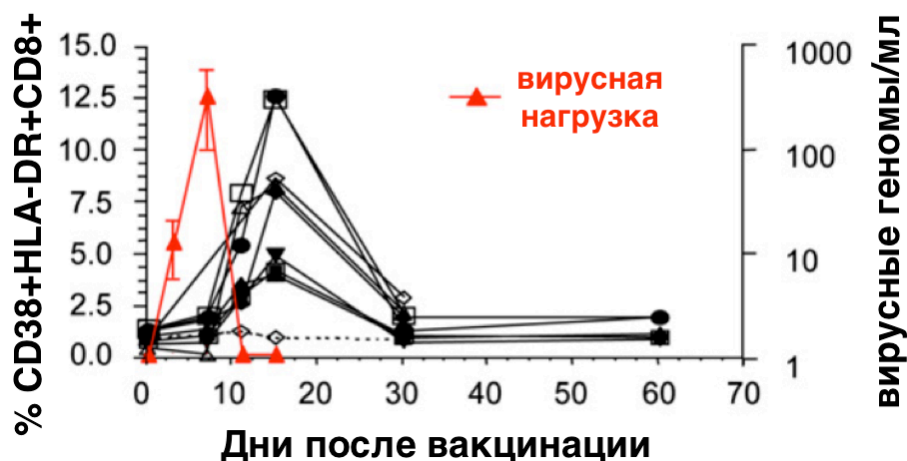
Желтая лихорадка распространена в Южной Америке и Африке, переносится комарами [Pulendran и др., 2013]. Заболевание в тяжелых случаях протекает как геморрагическая лихорадка со смертельным исходом.

### 2.7.2. Динамика иммунного ответа на вакцину против желтой лихорадки

Вакцина от желтой лихорадки – одна из лучших по эффективности и безопасности вакцин, была создана в 1930х годах южноафриканским иммунологом Максом Тейлером. Вакцина представляет собой аттенуированный последовательными пассажами на куриных эмбрионах штамм Асиби вируса желтой лихорадки. Живой вакцинный штамм YFV17D полученный Тейлером используется до сих пор [Pulendran и др., 2013]. После инъекции живой вакцины вирус размножается в организме вакцинируемого, и вызывает массивный иммунный ответ. Побочные эффекты от использования этой вакцины не связаны с реверсией вакцинного штамма к дикому типу, а вызваны неадекватной реакцией иммунной системы.

В целом клеточный иммунный ответ на вакцинацию можно разделить на несколько фаз: в первой фазе происходит взаимодействие зараженных клеток и вирусных частиц с системой врожденного иммунитета [Bocharov, Argilaguet, Meyerhans, 2015; Le, Miller, Ganusov, 2015]. Эффективная активация врожденного иммунитета необходима для эффективной презентации антигена системе адаптивного иммунитета. После презентации, специфичные к вирусу вакцины от желтой лихорадки CD8+ клетки начинают пролиферировать и занимают существенную часть репертуара Т-клеток (фаза клональной экспансии), после полного уничтожения вируса и зараженных клеток ненужные эффекторные клетки погибают (фаза сокращения репертуара) и остается небольшое число клеток памяти, которые активируются при повторной инфекции [Kongsgaard и др., 2017].

Клеточный иммунный ответ на вакцину от желтой лихорадки в целом хорошо охарактеризован. Первые данные по динамике ответа были получены с помощью метода проточной цитофлуориметрии с антителами на активационные маркеры CD8+ Т-клеток [Miller и др., 2008]. Было показано, что пик ответа приходится на 15 день, когда до 12% CD8+ клеток донора имеют активированный CD38+HLA-DR+ фенотип (см. рис 6). При этом аналогичный параметр для вируса осповакцины оказался еще выше – около 40% активированных лимфоцитов на пике инфекции. К 45 дню после вакцинации у всех доноров концентрация активированных клонов снизилась до нормальных значений.



**Рис. 6.** Динамика клеточного иммунного ответа на вакцину от желтой лихорадки. Красным показан титр вакцинного штамма вируса в плазме крови, черными линиями – доля CD38+ HLA-DR+ CD8+ активированных клеток, определенная по данным поточной цитофлуориметрии. Цит. по [Miller и др., 2008] с изменениями.

Другой метод оценки концентрации специфических к вирусу желтой лихорадки CD8+ Т-лимфоцитов – с помощью окраски МНС-тетрамерами. Т-клетки со специфичными к определенному рМНС Т-клеточными рецепторами можно окрасить связанными с флуоресцентным красителем рМНС, после чего определить их концентрацию методом проточной цитофлуориметрии [Akondy и др., 2009; Akondy и др., 2017; Blom и др., 2013; Bovay и др., 2017]. Но для создания таких рМНС комплексов необходимо знать иммунодоминантные пептиды вируса вакцины от желтой лихорадки, а также МНС-аллели доноров. Эта задача была решена с помощью инкубации перекрывающихся пулов пептидов из геномного полипротеина вируса желтой лихорадки с Т-клетками вакцинированных доноров в работе [Akondy и др., 2009]. Авторы сосредоточились на HLA-A\*02+ донорах и обнаружили пептид – LLWNGPMAV (позиции 214-222 неструктурного белка NS4B) который вызывал ответ у всех протестированных доноров. Тетрамер HLA-A\*02-LLWNGPMAV использовался ими для оценки динамики ответа специфических клонов Т-клеток. Были получены те же выводы, что и в работе [Miller и др., 2008] – пик ответа приходится на 15-30 день, когда свыше 4% клеток специфично узнают рМНС комплекс. Ответ CD4+ клеток анализировался в работе [Blom и др., 2013] по другим активационным маркерам (Ki67+CD38+), было показано, что он предвещает ответ CD8+ клеток. Авторы также показали, что ответ на вирус желтой лихорадки достаточно разнообразен – из 78 предсказанных с помощью алгоритма NetCTL эпитопов 57 вызывали ответ CD8+ клеток у хотя бы одного из 13 вакцинированных доноров. При

этом 12 эпитопов вызывали ответ более чем у четырех доноров, в их числе и найденный ранее LLWNGPMAV эпитоп (представляется на HLA-A\*02) [Akondy и др., 2009], а так же доминантный эпитоп для другого HLA-аллеля RPIDDRFGL (представляется на HLA-B\*07). На пике ответа на вакцину (15 день) – доля специфичных к этим эпитопам CD8+ Т-клеток превышала 7 процентов от всех CD8+ клеток.

### 2.7.3. Изучение иммунного ответа на вакцинацию с помощью HTS

В одной из первых работ по изучению репертуара TCR вакцинированного индивида с помощью HTS, был определен репертуар и частоты использования V-сегментов в различных временных точках после вакцинации [Li и др., 2013], но из-за использования платформы Roche 454 общее количество сиквенсов оказалось недостаточным для воспроизводимого количественного анализа: авторы не делают никаких выводов о значимом изменении концентрации клонов или отдельных V-J комбинаций в результате противогриппозной вакцинации.

В работе [Qi и др., 2016] изучалась вакцинация близнецовых доноров живой вакциной от ветряной оспы (VZV, Varicella Zoster Virus). Авторы секвенировали репертуар TCRbeta CD4+ наивных клеток и клеток памяти, а также репертуара TCR VZV-специфичных культур CD4+ клеток, полученных в результате стимуляции CD4+ клеток лизатом VZV. Было показано большее количество общих последовательностей CDR3 между репертуарами CD4+ клеток памяти и VZV-специфичными CD4+ репертуарами близнецовых доноров по сравнению с аналогичными репертуарами неродственных доноров.

В работе [DeWitt и др., 2015] авторы использовали основанную на секвенировании ДНК TCR технологию реконструкции репертуара на примере вакцинации от желтой лихорадки. Для оценки статистической значимости изменений концентраций клонотипов использовался точный тест Фишера. В работе рассматривались пары образцов из двух временных точек, без биологических повторностей. Для валидации значимо изменивших концентрацию клонов авторы также секвенировали HLA-DR+CD38+CD8+ активированную фракцию Т-лимфоцитов характерную для клеток специфичных к вакцине от желтой лихорадки [Miller и др., 2008]. Было показано, что на вакцину отвечает большое число клонов (>2000 клонов CD8+ лимфоцитов), но только 5-6% этих клонов переходят в субпопуляцию клеток памяти после окончания первичного иммунного ответа на инфекцию.

## 2.8. Заключение

Таким образом, к настоящему моменту получены важные данные о структуре, особенностях рекомбинации и селекции TCR, что позволяет обучать модели сборки и селекции репертуара T-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. С другой стороны, накоплен массив эмпирических данных о строении и онтогенезе репертуаров антиген-распознающих рецепторов, при этом часть существенных сведений получена при анализе репертуаров монозиготных близнецов. Было выяснено, что машинерия сборки близнецов настроена схожим образом, а также показано большее количество идентичных рецепторов в репертуарах близнецовых доноров. Тем не менее, ни одна из опубликованных ранее работ не учитывала особенности эмбрионального развития близнецов, а именно возможность обмена кровью между эмбрионами.

С другой стороны, накоплены интересные данные об онтогенезе адаптивной иммунной системы, в том числе показано, что в ранних эмбриональных В- и Т-клетках отличается механизм сборки T-клеточного рецептора, что приводит к снижению разнообразия эмбриональных репертуаров. Тем не менее, вклад этих особенностей в строение зрелой иммунной системы не прослежен.

Таким образом, возникает интересная возможность использовать новые подходы к моделированию репертуаров TCR для экспериментальной проверки двух взаимосвязанных гипотез: о пренатальном обмене кровью близнецами донорами и роли эмбрионального репертуара Т-клеток в структуре репертуаров Т-клеточных рецепторов взрослых доноров.



## 3. Материалы и методы

### 3.1. Материалы

#### 3.1.1. Использованное оборудование и расходные материалы

Автоматические микропипетки (Gilson, Франция); центрифуги Eppendorf 5415R, 5804R (Eppendorf, Германия); термостат Термит (ДНК-технология, Россия); камера для электрофореза (Хеликон, Россия); источник питания Эльф-4 (ДНК-технология, Россия); автоматический термоциклер DNA-engine PTC-200 (Bio-Rad, США); качалка Bio shaker 3D (Biosan, Латвия); миксер лабораторный Bio Vortex V1 (Biosan, Латвия); 0.5мл пробирки для ПЦР (SSI, США), 1.5 мл пробирки для микроцентрифугирования (Sarstedt, Германия); пластиковые пробирки на 15 мл (Corning, США) и 50 мл (Greiner, Австрия), пробирки для сбора крови с ЭДТА Vacuette (Greiner, Австрия), пластиковые наконечники для микропипеток (SSI, США), бокс для ПЦР с УФ-лампой (Ламинарные системы, Россия).

#### 3.1.2. Использованные реактивы

Агароза для электрофореза (Promega, США), изопропиловый спирт, этиловый спирт, глицерин (Sigma, США), смесь dNTP (Евроген, Россия), смесь ДНК-полимераз Q5 (NEB, США), обратная транскриптаза SmartScribe (Clontech, США), урацил-ДНК-гликозилаза (NEB, США), Trizol (Invitrogen, США), набор для выделения ДНК из геля и очистки продуктов ПЦР (Qiagen, США), набор для позитивной селекции CD4<sup>+</sup> клеток Dynabeads (Invitrogen, США), набор для позитивной селекции CD8<sup>+</sup> клеток Dynabeads (Invitrogen, США), набор для селекции CD45RO<sup>+</sup> клеток (Myltenyi Biotec, Германия), набор для выделения CD8<sup>+</sup> наивных клеток (Myltenyi Biotec, Германия), раствор фиколла (Панэко, Россия).

#### 3.1.3. Буферные растворы

Буфер PBS (1x) - 137 mM NaCl, 2.7 mM KCl, 1.47mM KH<sub>2</sub>PO<sub>4</sub>, 4.29mM Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, pH 7.3-7.5 (Amresco, США)

Буфер TBE (5×) - 45 mM Трис, 45 mM H<sub>3</sub>BO<sub>3</sub>, 10 mM ЭДТА (pH 8.3)

Буфер для нанесения образцов ДНК на агарозный гель – 0.1% бромфеноловый синий, 0.1 % ксиленцианол, 30 % глицерин.

## 3.2. Методы

### 3.2.1. Доноры и образцы крови

Образцы венозной крови получали из 3 пар здоровых монозиготных близнецов женского пола. Доноры были близки по возрасту: 23 года (доноры S1, S2), 23 года (доноры P1 и P2), 25 лет (доноры Q1 и Q2). В каждой из близнецовых пар доноры жили вместе большую часть своей жизни. Перед взятием крови каждый донор был протестирован на опасные инфекции в диагностической лаборатории (гепатит С, ВИЧ, сифилис). Образцы наивных клеток и клеток памяти были взяты у двух доноров мужского пола (19 и 57 лет), и донора женского пола (51 год). Все доноры были европеоидами, без хронических заболеваний, кровь была собрана после информированного согласия.

Генетическая идентичность близнецов проверялась с помощью генотипирования полиморфных локусов Alu-инсерций [Mamedov и др., 2010].

Фракция периферических мононуклеарных клеток крови (PBMC), была выделена методом центрифугирования на градиенте плотности раствора фиколла (Панэко, Россия). Треть выделенных клеток использовалась для выделения тотальной РНК с помощью реагента Trizol (Invitrogen, США) согласно протоколу производителя. Остальные клетки были использованы для выделения функциональных популяций CD4<sup>+</sup> (Т-хелперов), CD8<sup>+</sup> (Т-киллеров) и CD45RO<sup>+</sup> (клетки центральной памяти).

### 3.2.2. Выделение PBMC

Периферическую кровь собирали в вакуумные пробирки с ЭДТА (Vacuette, Австрия). До начала выделения фракции периферических мононуклеаров пробирки с кровью держали на ротаторе. Фиколл нагревали до комнатной температуры и по 7мл наливали в пробирки объемом 15 мл. Далее разведённую в четыре раза стерильным буфером PBS кровь медленно наслаивали сверху, не допуская смешивания фаз. После этого центрифугировали при 400g 32 минуты при комнатной температуре на центрифуге с отключенным тормозом. Затем отбирали белесую интерфазу в отдельные пробирки, и разводили в три раза стерильным PBS. Затем

центрифугировали 15 минут с ускорением 450g и удаляли супернатант. Осадок ресуспендировали в 15 мл PBS и снова центрифугировали 15 минут с ускорением 450g. Осадок ресуспендировали в 100 мкл PBS и заливали 1 мл реагента Trizol, либо использовали для выделения CD4+, CD8+ либо клеток памяти с помощью наборов для позитивной селекции соответствующих субпопуляций.

### 3.2.3. Выделение функциональных субпопуляций CD4, CD8, CD45RO T-клеток

CD4 и CD8 T-клетки выделяли из фракции периферических мононуклеарных клеток с помощью наборов для позитивной селекции CD4+ и CD8+ T-клеток соответственно (Invitrogen, США). CD45RO клетки выделяли с помощью системы позитивной селекции CD45RO+ клеток (Myltenyi Biotec, Германия). Наивные T-клетки выделяли с помощью набора для выделения наивных CD8+ клеток без финального этапа обогащения CD8+ клеток (Myltenyi Biotec, Германия).

### 3.2.4. Выделение РНК из клеток с помощью реагента Trizol

Выделение РНК реагентом Trizol (Invitrogen), имеет тот же принцип, что и выделение нуклеиновых кислот классическим фенол-хлороформным методом. Все операции при выделении РНК производили в перчатках для защиты препарата от РНКаз. Замороженный образец Trizol с лизированными клетками размораживали и инкубировали 5 минут при комнатной температуре. Далее на каждый объем Trizol добавляли 1/5 объема хлороформа. Затем смесь интенсивно встряхивали в течение 15-30 секунд до образования непрозрачной эмульсии и инкубировали при комнатной температуре 3 минуты. Затем смесь центрифугировали 15 минут при 12000g и температуре 4°C. После этого верхнюю водную фазу осторожно отбирали и добавляли 0.7 мкл соосадителя Satellite Red (Евроген, Россия), а также изопропиловый спирт, половину объема реагента Trizol. Смесь инкубировали при комнатной температуре 10 минут, и затем центрифугировали 15 минут при 12000g и температуре 4°C. Затем супернатант сливали, к осадку добавляли 1 мл 80% этилового спирта и центрифугировали 10 минут при 12000g и температуре 4°C. После этого супернатант снова удаляли и повторно добавляли 1 мл 80% этилового спирта к осадку и центрифугировали 10 минут при 12000g и температуре 4°C. После этого супернатант удаляли, осадок сушили на 45°C 10 минут, после чего растворяли в 20 мкл деионизованной воды и прогревали при 50°C 5 минут. После этого 1 мкл раствора РНК отбирали для измерения концентрации на флуориметре Qubit и электрофереза на 1% агарозном геле.

### 3.2.5. Электрофорез в агарозном геле

Электрофорез проводили в 1% агарозном геле на 0.5x TBE с бромистым этидием, камеру для горизонтального электрофореза заполняли 0.5x TBE. Перед нанесением в лунки образцы смешивали с буфером для нанесения. Режим электрофореза: напряжение 120В, 40 мин. Для определения длины фрагментов использовали маркер длин фрагментов ДНК.

### 3.2.6 Очистка ПЦР-продуктов

Очистка ДНК проводилась на спин-колонках QIAquick (Qiagen, США). ПЦР продукт переносили в пробирку на 1.5 мл и смешивали с 5 объемами буфера РВ. Смесь переносили в спин-колонку и центрифугировали 1 мин при 12000g, центрифугат удаляли. Для промывки вносили в колонку 0.75 мл буфера РЕ центрифугировали 1 мин. при 12000g, центрифугат удаляли. Для удаления остатков РЕ центрифугировали пустую колонку 1 мин 12000g и переносили колонку в чистую собирательную пробирку. Элюцию ДНК с колонки производили буфером ЕВ: 50 мкл буфера вносили в колонку, инкубировали 1 минуту и центрифугировали 1 мин 12000g.

### 3.2.7. Выделение ДНК из агарозного геля

Выделение ДНК из геля производилось набором для экстракции ДНК из геля QIAquick (Qiagen, США). Вырезанный кусок геля помещали в пробирку на 1.5 мл, добавляли к 1 объему геля 3 объема буфера QG, и инкубировали 20 минут на качалке при комнатной температуре, после чего добавляли 1 объем изопропанола. Смесь переносили в спин-колонку и центрифугировали 1 мин при 12000g, центрифугат удаляли. Затем вносили в колонку 0.5 мл QG, центрифугировали 1 мин 12000g, центрифугат удаляли. Для промывки вносили в колонку 0.75 мл буфера РЕ, инкубировали при комнатной температуре 3 минуты, и центрифугировали 1 мин. при 12000g. Промывку повторяли с 0.25 мл буфера РЕ, центрифугат удаляли. Для удаления остатков РЕ центрифугировали пустую колонку 1 мин 12000g и переносили колонку в чистую собирательную пробирку. Элюцию ДНК с колонки производили буфером ЕВ: 50 мкл буфера вносили в колонку, инкубировали 1 минуту и центрифугировали 1 мин 12000g.

### 3.2.8. Приготовление библиотек кДНК альфа и бета цепей Т-клеточных рецепторов

Первая цепь кДНК синтезировалась с помощью набора для обратной транскрипции SmartScribe (Clontech, США), праймер для синтеза был комплементарен константным сегментам альфа и бета цепей Т-клеточного рецептора. Для внесения на 3'-конец первой цепи кДНК универсального сайта отжига праймера использовали эффект смены матрицы обратной транскриптазой. Был разработан набор олигонуклеотидов для смены матрицы (SMART - адаптеры), содержащие уникальные молекулярные идентификаторы (UMI) и индивидуальные баркоды для различных образцов.

Каждая пробирка содержала:

РНК 500нг,

5x SmartScribe буфера 2 мкл,

dTT 1мкл

dNTP (1mM каждого нуклеотида) 1 мкл,

праймер BCuniR4vvshort 10 пкмоль и

праймер TRACR2 10 пкмоль

10 пкмоль SMART-олигонуклеотида.

mQ до объема 10 мкл.

5 мкг РНК (соответствует 5 млн РВМС) использовалось для синтеза кДНК для каждого образца (10 реакций синтеза на образец).

Продукт синтеза кДНК далее обрабатывался 1 мкл 5 ед./мкл дезоксиурацилгликозилазой (UDG, NEB, США) 45 мин, 37 градусов для расщепления SMART-адаптера. После обработки UDG реакция синтеза очищалась с помощью набора для очистки продуктов ПЦР (Quigen, США) по протоколу производителя.

После синтеза кДНК проводили две стадии ПЦР для амплификации кДНК и введения необходимых для секвенирования на платформе Illumina последовательностей. Первая стадия ПЦР проходила 16 циклов при следующем температурном режиме:

- 1) 94 градуса, 20 секунд,
- 2) 60 градусов, 15 секунд,
- 3) 72 градуса, 60 секунд.

Каждая ПЦР реакция проходила в объеме 15 мкл и содержала:

5x Q5 буфер (NEB, США), 3 мкл

5 пкмоль каждого из праймеров Sm1msq, RPacj, RPbcj1, RPbcj2

dNTP (0.125mM каждого нуклеотида)

0.15 мкл Q5 полимеразы

mQ до объема 15 мкл.

Затем 1 мкл продукта первой ПЦР реакции использовался в качестве матрицы во второй ПЦР-реакции. Температурный режим второй реакции (9 циклов):

- 1) 94 градуса - 20 секунд,
- 2) 60 градусов - 15 секунд,
- 3) 72 градуса - 40 секунд.

Вторая ПЦР-реакция проходила в объеме 25 мкл и содержала:

5xQ5 буфер 5 мкл

по 5 пкмоль праймеров Sm-out-msq

5 пкмоль IL-acj-ind (для библиотек альфа-цепей)

5 пкмоль IL-bcj-ind (для библиотек бета-цепей),

0.25 мкл Q5 полимеразы.

Отбор фрагментов необходимой длины (500-800 п.н.) проводили с помощью электрофореза в 1% агарозном геле. Выделение РНК из геля проводили с использованием набора для выделения ДНК из геля (Qiagen, США), согласно протоколу производителя.

**Таблица 1.** Последовательности олигонуклеотидов, использованных в данной работе

SMART-Мк олигонуклеотиды	
МК-108	CAGUGGUAUCAACGCAGAGUACNNNNNNUAATGCUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-248	CAGUGGUAUCAACGCAGAGUACNNNNNUNNTGGCANNUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-253	CAGUGGUAUCAACGCAGAGUACNNNNNUNNTTATGNNUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-103	CAGUGGUAUCAACGCAGAGUACNNNNNNUAACGGUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-257	CAGUGGUAUCAACGCAGAGUACNNNNNUNNTTGCNNUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-143	CAGUGGUAUCAACGCAGAGUACNNNNNNUCAGATUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-135	CAGUGGUAUCAACGCAGAGUACNNNNNNUATGCAUNNNNNNUCTT (rG) (rG) (rG) (rG)
МК-227	CAGUGGUAUCAACGCAGAGUACNNNNNUNNTAACCNUNNNNNNUCTT (rG) (rG) (rG) (rG)

Праймеры для синтеза кДНК	
BC_uni_R4vvsh ort	TGGAGTCATTGA
TRAC_R2	ACACATCAGAATCCTTACTTTG
Праймеры PCR I	
Sm1msg	GAGATCTACACGAGTCAGCAGTGGTATCAACGCAG
RPbcj1	CGACTCAGATTGGTACACCTTGTTCAGGTCCTC
RPbcj2	CGACTCAGATTGGTACACGTTTTTTCAGGTCCTC
RPacj	CGACTCAAGTGTGTGGGTCAGGGTCTGGATAT
Праймеры PCR II (XXXXX стандартную означает последовательность TruSeq index)	
Sm-out-msg	AATGATACGGCGACCACCGAGATCTACACGAGTCA
Il-bcj-indX	CAAGCAGAAGACGGCATAACGAGATXXXXXXXXCGACTCAGATTGGTAC
Il-acj-indX	CAAGCAGAAGACGGCATAACGAGATXXXXXXXXCGACTCAAGTGTGTGG
Праймеры для секвенирования	
IL-AIRP	ATATCCAGAACCCTGACCCACACACTTGAGTCG
IL-IRP-b1	GAGGACCTGAAAAACGTGTACCAATCTGAGTCG
IL-IRP-b2	GAGGACCTGAACAAGGTGTACCAATCTGAGTCG
IL-RP1-msg	ACACGAGTCAGCAGTGGTATCAACGCAGAGTAC
IL-RP2-b1	CGACTCAGATTGGTACACGTTTTTTCAGGTCCTC
IL-RP2-b2	CGACTCAGATTGGTACACCTTGTTCAGGTCCTC
IL-ARP2	CGACTCAAGTGTGTGGGTCAGGGTCTGGATAT

### 3.2.9. Высокопроизводительное секвенирование (HTS)

Секвенирование полученных фрагментов проводили на платформе Illumina HiSeq 2500, чтение с двух сторон по 100 нт.

### 3.2.10. Обработка сырых массивов секвенирования

Массивы сиквенсов обрабатывали с помощью программы MiGEC (для кластеризации ридов по уникальным молекулярным идентификаторам). Были отброшены уникальные молекулярные идентификаторы с числом ридов менее 2, поскольку эта группа обогащена сиквенсами с ошибками. Затем массивы секвенирования обрабатывали с помощью программы MiXCR для определения последовательности гипервариабельного участка CDR3 и выравнивания V и J сегментов. Число сиквенсов, уникальных молекулярных идентификаторов и уникальных клонотипов по данным MiXCR представлено в таблице 2 и таблице 3.

### 3.2.11. Обучение модели сборки

На массиве данных каждого донора была выведена своя индивидуальная модель сборки. Метод вывода параметров модели был подробно описан ранее [Elhanati и др., 2015a; Murugan и др., 2012]. Модель сборки каждой последовательности рассчитывалась как сумма вероятностей сборки рекомбинационных сценариев, которые могут давать данную нуклеотидную последовательность. Для альфа-цепи Т-клеточного рецептора была предложена следующая формула для расчета вероятности рекомбинационного сценария:

$$P_{\text{rearr}}^{\alpha}(\tau) = P(V, J)P(\text{del}V|V)P(\text{del}J|J)P(\text{ins}). \quad (3)$$

где  $P(V, J)$  вероятность выбора комбинации V и J сегмента,  $P(\text{del}V|V)$  – вероятность определенного числа делеций при условии выбора V-сегмента,  $P(\text{del}J|J)$  – вероятность определенного числа делеций при условии выбора J-сегмента,  $P(\text{ins})$  – вероятность определенного числа N-инсерций.

Вывод параметров модели проводили с помощью EM-алгоритма, который итеративно подбирает параметры модели для максимизации правдоподобия набора данных. Для альфа-цепей использовали программу `gergenNHM` [Elhanati и др., 2015a], которая выводит параметры модели сборки с помощью алгоритма Баума-Велша.

Для бета-цепи использовалась более сложная модель:

$$P_{\text{rearr}}^{\beta}(\tau) = P(V, D, J)P(\text{del}V|V)P(\text{ins}VD) \\ \times P(\text{del}Dl, \text{del}Dr|D)P(\text{ins}DJ)P(\text{del}J|J) \quad (4)$$

где  $P(V, D, J)$  - вероятность выбора комбинации V, D и J сегментов,  $P(\text{del}V|V)$  – вероятность определенного числа делеций при условии выбора V-сегмента,  $P(\text{del}J|J)$  – вероятность определенного числа делеций при условии выбора J-сегмента,  $P(\text{ins}VD)$  – вероятность определенной последовательности N-инсерций на стыке между V и D сегментами,  $P(\text{ins}DJ)$  – вероятность определенной последовательности N-инсерций на стыке между D и J сегментами,  $P(\text{del}Dl, \text{del}Dr|D)$  – вероятность определенного числа 3' и 5' делеций D при условии выбора определенного D-сегмента).

Параметры модели выводились напрямую с использованием EM-алгоритма, путем перебора всех возможных сценариев сборки для каждой последовательности, как описано в [Murugan и др., 2012]. Эта процедура позволяет получить вероятности отдельных событий сборки с большой точностью, несмотря на то, что конкретный сценарий сборки невозможно однозначно восстановить для каждой отдельной последовательности из данных - как правило каждая отдельная последовательность может собираться многими способами.



### 3.2.12. Вывод распределения инсерций для каждой ранговой группы бета-цепи TCR

Мы применили описанную в предыдущем разделе процедуру независимо к каждой группе рангов репертуара Т-клеточных рецепторов. Однако учитывая небольшой размер ранговых групп (2000-3000 тысячи последовательностей) мы не выводили полную модель сборки для каждой группы. Вместо этого была использована опубликованная ранее [Murugan и др., 2012] модель сборки бета-цепи Т-клеточного рецептора. Мы использовали частоты выбора V- и J-сегментов, а также профили делеций из этой модели, а  $P(\text{insVD})$  и  $P(\text{insDJ})$  выводили из данных отдельно для каждой группы. Выведенные распределения числа инсерций в каждом классе использовались для выявления зависимости среднего числа инсерций от ранга клона (рис. 19) и возраста донора (рис. 21-23).

### 3.2.13. Вывод факторов селекции

Функциональный репертуар отличается от нефункционального репертуара последовательностей со сдвигом рамки считывания. Очевидное отличие - в распределении длин последовательностей гипервариабельного участка CDR3, нефункциональному репертуару соответствует больший диапазон длин участка CDR3, в то время как для функциональных последовательностей его длина всегда кратна трем, по свойству триплетности генетического кода. Но есть еще одно существенное отличие, которое дополнительно снижает потенциальное разнообразие функционального репертуара - селекция в тимусе. Для того, чтобы учесть селекцию в работе [Elhanati и др., 2014] было предложено ввести фактор селекции  $Q$ , зависящий от последовательности CDR3.  $Q$  определяется как отношение вероятности увидеть последовательность в функциональном репертуаре к вероятности сборки этой последовательности, при этом параметры вероятности сборки выводятся на нефункциональном репертуаре (см. выше). Фактор селекции  $Q$  представляется как произведение факторов  $q_i$ , при этом величина  $q_i$  зависит от аминокислоты и ее положения в CDR3 определенной длины:

$$Q \propto q_L \prod_{i=1}^L q_{i;L}(a_i) \quad (5)$$

### 3.2.14. Предсказание совпадения нефункциональных последовательностей CDR3 между донорами

Сначала для каждого донора была выведена индивидуальная модель сборки, см. 3.2.11, своя для репертуаров бета и альфа-цепей. Далее с помощью этой модели сборки были сгенерированы наборы симулированных последовательностей, в которых средствами пакета *tcR* [Nazarov и др., 2015] производился поиск идентичных вариантов. Чтобы сравнить это число с реальными данными его нужно нормировать на произведение размеров симулированных репертуаров.

### 3.2.15. Предсказание числа общих функциональных последовательностей

Для того, чтобы точно предсказать число общих функциональных последовательностей необходимо ввести дополнительную поправку на селекцию в тимусе. Симуляция последовательностей происходила так же, как и в случае нефункциональных последовательностей. Далее для каждой последовательности рассчитывалось значение фактора селекции  $Q$ . Предсказанное нормализованное число общих последовательностей рассчитывалось по формуле:

$$\frac{1}{|S_1| \cdot |S_2|} \sum_{s \in S_1 \cap S_2} Q^{(1)}(s)Q^{(2)}(s), \quad (6)$$

где  $S_1$  и  $S_2$  это два симулированных с использованием индивидуальных моделей сборки репертуара,  $Q^{(1)}$  и  $Q^{(2)}$  - индивидуальные модели селекции.  $|S_1|$  и  $|S_2|$  обозначают размеры двух массивов. Суммирование идет только по последовательностям присутствующим в обоих симулированных репертуарах.

### 3.2.16. Моделирование смеси фетальных и зрелых клонотипов

Мы выдвинули гипотезу, что большее число идентичных последовательностей среди наиболее крупных клонотипов нефункциональных репертуаров неродственных людей связано с большей долей клонотипов с нулем инсерций, возникших в эмбриональном развитии. Для проверки этой гипотезы мы предложили простую модель такой смеси.

Отсортируем клонотипы в репертуаре по представленности и разобьем полученный список на последовательные неперекрывающиеся группы одинаковой численности.

Мы предполагаем, что в каждой группе  $C$  есть  $F(C)$  зародышевых клонотипов, имеющих ноль инсерций и  $1-F(C)$  обычных зрелых клонотипов. Измерить  $F(C)$  непросто, поскольку зрелые клонотипы тоже могут иметь ноль инсерций. В дополнение, число инсерций в принципе невозможно измерить точно. Например, делеция, которая сопровождается инсерцией той же самой буквы будет ошибочно интерпретирована как отсутствие инсерций.

Чтобы преодолеть это ограничение, мы вводим величину, которую измерить гораздо проще:  $F_0(C)$  долю клонотипов которые *не противоречат* отсутствию инсерций, т.е. таких клонотипов, в которых вся последовательность гипервариабельного участка CDR3 может быть выровнена на геномные последовательности V, D и J сегментов без пропусков. Долю таких клонотипов в экспериментальных данных можно рассчитать после выравнивания. Необходимо отметить, что  $F_0(C)$  не равно  $F(C)$ . Однако,  $F_0(C)$  зависит от  $F(C)$  линейно, т.е.  $F_0(C)=A+BF(C)$ . Таким образом, если мы можем сгенерировать такие последовательности, что их  $F_0(C)$  сойдется с данными, то их  $F(C)$  так же сойдется с данными. Чтобы получить эту смесь, мы сгенерировали множество сиквенсов с помощью модели сборки. Для определения, какие сгенерированные последовательности не противоречат нулю инсерций, мы выравнивали их на возможные V и J сегменты. Затем мы отделили последовательности, не противоречащие гипотезе о нуле инсерций от остальных. Для каждой группы  $C$  мы далее создали смесь  $F_0(C)$  симулированных клонотипов не противоречащих нулю инсерций с  $1-F_0(C)$  клонотипов противоречащих нулю инсерций. Далее мы рассчитали число общих клонотипов между такими синтетическими массивами и сравнили результат с экспериментальными данными.

## 4. Результаты и обсуждение

Для секвенирования кДНК альфа и бета-цепей Т-клеточных рецепторов на платформе Illumina необходимо решить несколько технических проблем, от выделения Т-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. Следующие разделы посвящены различным аспектам пробоподготовки и анализа сырых данных высокопроизводительного секвенирования.

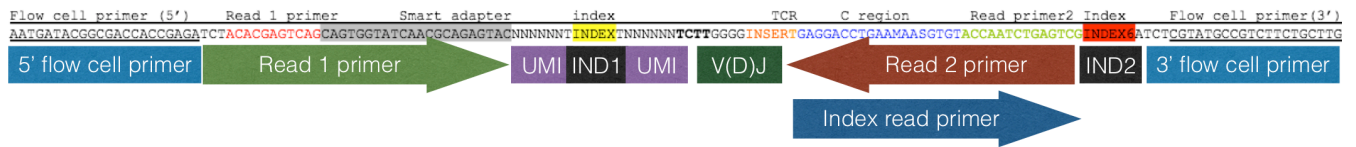
### 4.1. Разработка системы амплификации кДНК альфа- и бета-цепей Т-клеточного рецепторов

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 периферических мононуклеаров.
- 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.
- 6) При длине чтения 100 нт во втором ряде целиком помещается участок CDR3, т.е. чтение начинается непосредственно с J-сегмента, а в другом ряде находится молекулярный баркод и идентификатор образца.

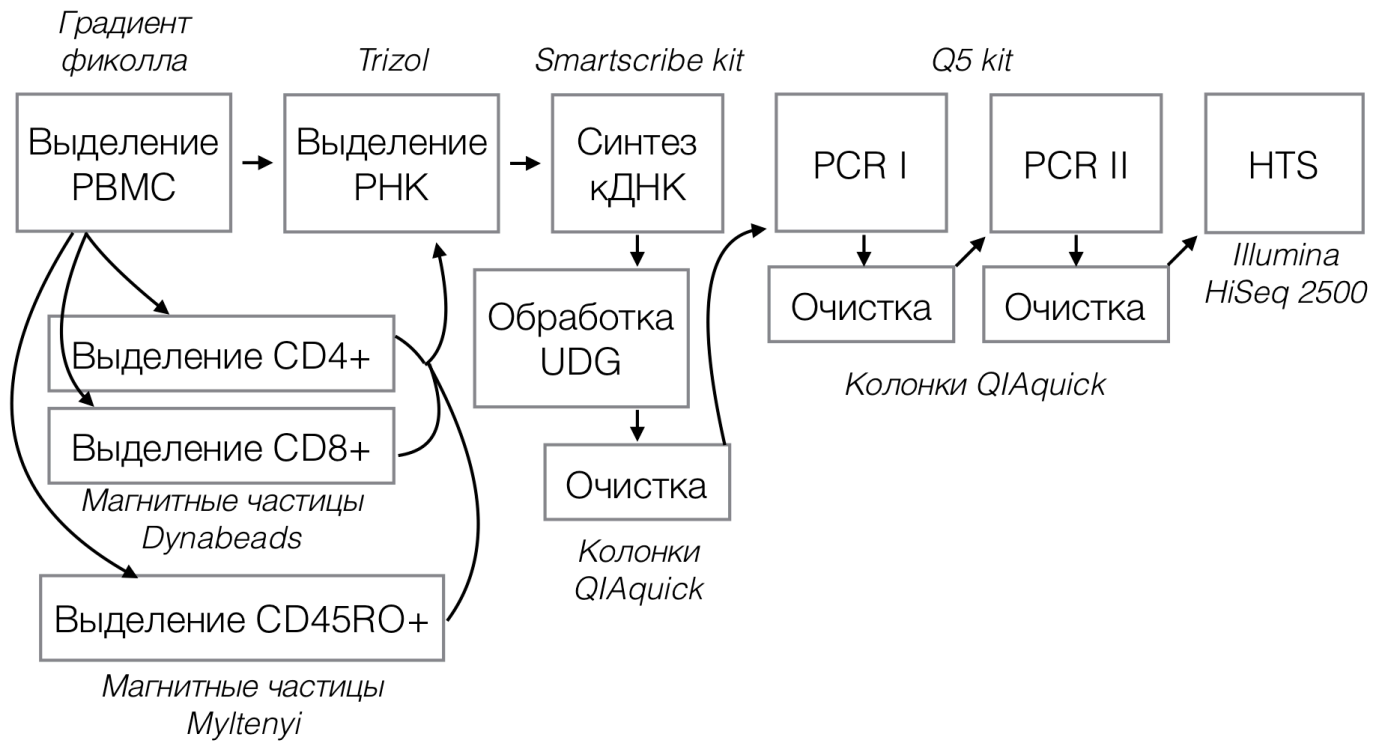
Итоговая схема библиотеки со всеми перечисленными выше особенностями представлена на рис. 7



**Рис. 7.** Схематическая структура библиотеки. Слева направо: сайт отжига flow cell праймера для мостиковой амплификации на чипах Illumina HiSeq и MiSeq (синий), сайт отжига для Read праймера первого рида (зеленый), вырожденная последовательность уникального молекулярного идентификатора (UMI, unique molecular identifier, фиолетовый), разбитая идентификатором образца (черный), собственно последовательность кДНК TCR от 5'UTR V-сегмента до начала J-сегмента (темно-зеленый), сайт отжига праймера второго рида (красный) и индекс-рида (синий), второй идентификатор образца (черный), второй сайт отжига праймера для мостиковой амплификации на чипах (3' flow cell primer, синий).

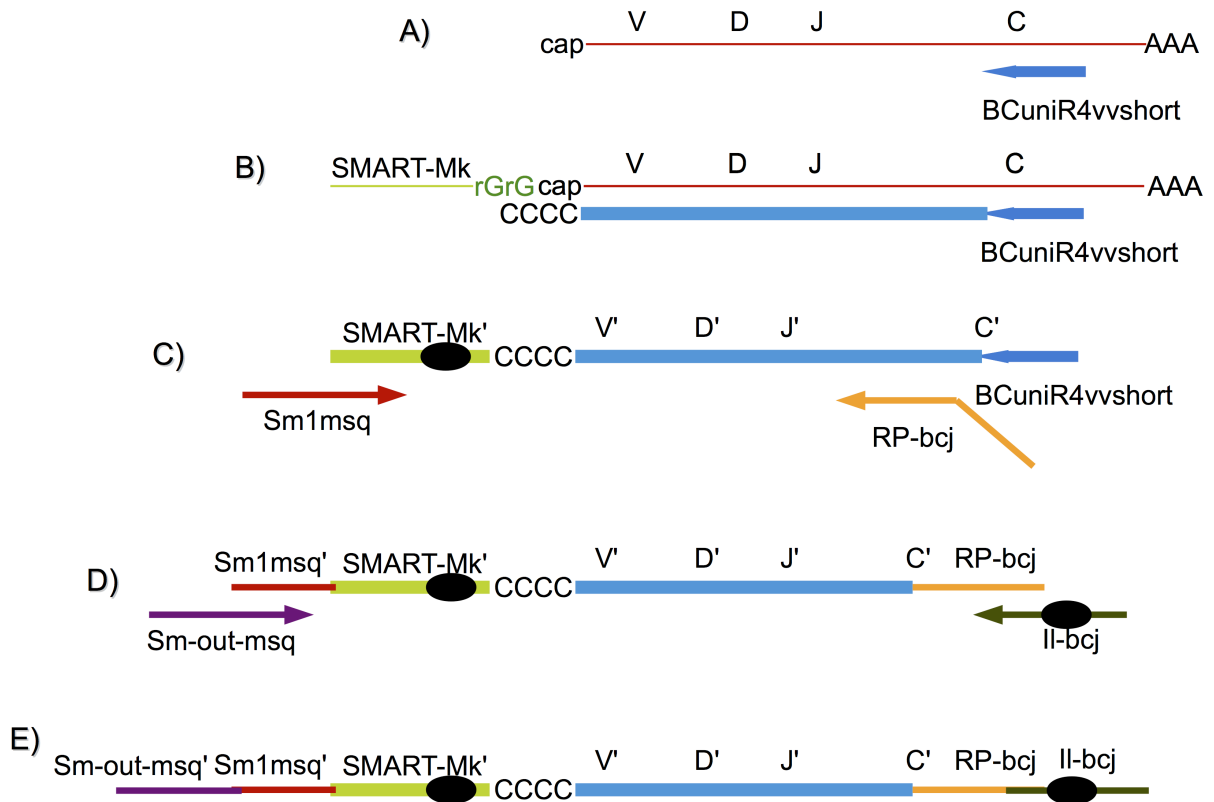
Необходимо отметить, что для секвенирования данной библиотеки необходимо использовать нестандартные праймеры для чтения (см. таблицу 1.), при этом один из праймеров (read primer 2) содержит в себе участок последовательности С-сегмента, что позволяет начать чтение напрямую с J-сегмента. Это необходимо для того, чтобы CDR3 участок был целиком покрыт ридом длиной 100 нт, от триплета кодирующего консервативный остаток цистеина в V-сегменте до триплета кодирующего консервативный остаток фенилаланина в J-сегменте.

Синтез кДНК в системе идет со специфических праймеров, комплементарных С-сегментам бета или альфа-цепи соответственно. За счет эффекта смены матрицы обратной транскриптазой вводятся молекулярный баркод и универсальный сайт отжига праймеров для первой реакции ПЦР. Затем в ходе двух реакций ПЦР вводятся необходимые адаптерные последовательности и второй идентификатор образца. Диаграмма пробоподготовки представлена на рис. 8, подробное описание синтеза кДНК и ПЦР-амплификации с названиями используемых олигонуклеотидов представлено на рис. 9, последовательности олигонуклеотидов представлены в таблице 1.



**Рис. 8.** Диаграмма рабочего процесса подготовки библиотек. На первом этапе с помощью градиента плотности фиколла 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 кДНК 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 кДНК амплифицировалась в двухстадийной ПЦР (см рис. 9) и секвенировалась на платформе Illumina HiSeq 2500.

Подробный протокол описан в разделе 3.2.8. главы Материалы и методы. Отметим, что система достаточно удобна и позволяет получить готовую библиотеку с образца тотальной РНК поставив одну реакцию синтеза и две последовательные реакции ПЦР, освобождая исследователя от трудоемкой реакции лигирования адаптерных последовательностей.

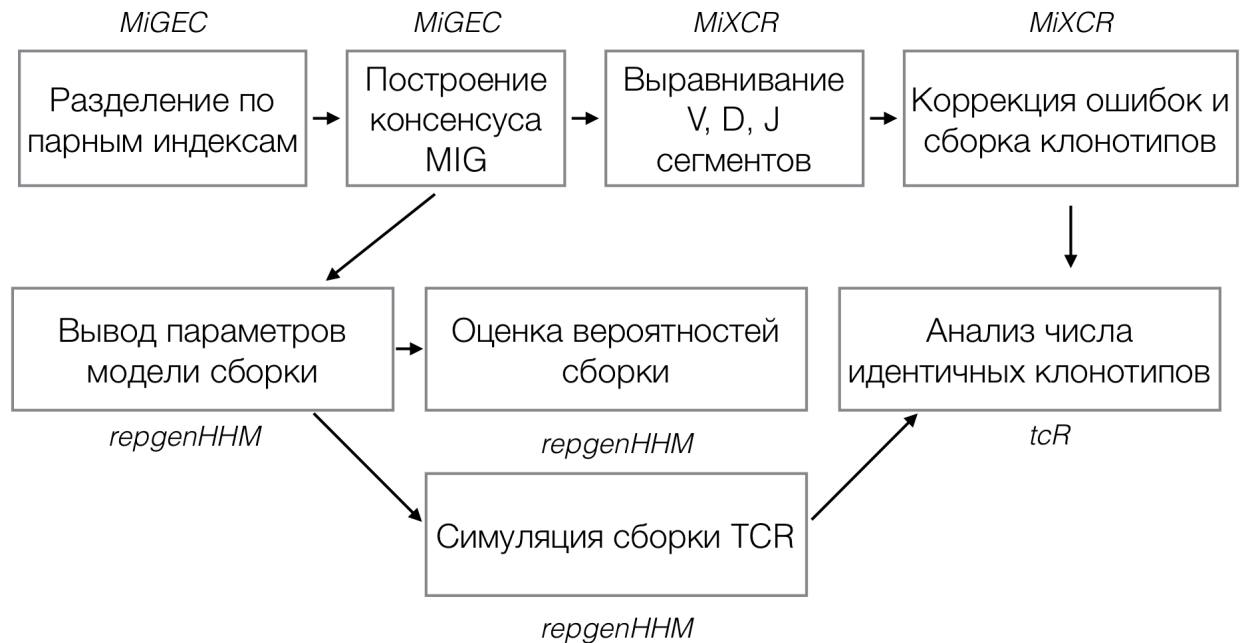


**Рис. 9.** Схема синтеза  $\kappa$ ДНК и ПЦР (на примере бета-цепи TCR). А) Синтез первых цепей  $\kappa$ ДНК проходит со специфических праймеров, комплементарных C-сегменту TCRB. В) Для ввода универсального сайта отжига праймеров образца использовался эффект смены матрицы обратной транскриптазой. Последовательность олигонуклеотида SMART-Mk содержит также идентификатор образца и UMI (черный эллипс). С) и D) Две стадии ПЦР нужны для ввода необходимых для секвенирования на платформе Illumina последовательностей, а также второго идентификатора образца (черный эллипс). Е) Полученная молекула  $\kappa$ ДНК имеет идентификаторы образца с двух сторон, UMI и может быть отсеквенирована на платформе Illumina с использованием специальных праймеров.

## 4.2. Разработка программного обеспечения для анализа Т-клеточных рецепторов

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. 10.



**Рис. 10.** Этапы анализа данных и используемые программные пакеты.

Первый шаг в анализе данных – очистка и разделение ридов по парным идентификаторам, соответствующим отдельным образцам (т.н. «демультиплекс»). Несмотря на то, что для каждой библиотеки используется уникальная последовательность второго идентификатора, этот шаг является необходимым. Было показано [Sinha и др., 2017], что на платформе Illumina при мостиковой амплификации часто происходит обмен концами молекул, что может приводить к формированию химерных ридов, где баркод соответствует одной библиотеке, а вся остальная последовательность рида – другой. Чтобы отфильтровать такие химерные последовательности необходимо использовать оба идентификатора образца, которые находятся на разных концах библиотеки. Для этого используется программа MiGEC [Shugay и др., 2014]. Результат этого шага – файлы .fastq для первого и второго рида, где и левый и правый баркоды соответствуют образцу. Следующий шаг – группировка ридов, которые имеют общий молекулярный идентификатор (UMI unique molecular identifier) в группы (MIG, molecular identifier group). Все риды внутри группы имеют идентичный UMI и соответствуют одной исходной молекуле кДНК, амплифицированной в ходе ПЦР. Риды внутри MIG используются для создания консенсусной последовательности, где на каждой позиции находится наиболее часто встречаемый в соответствующей позиции ридов нуклеотид. Результат этого шага для каждой библиотеки – два .fastq файла, в которых записаны консенсусы первого и второго ридов для каждой MIG. Общее число записей в этих файлах соответствует количеству уникальных молекул кДНК после синтеза первых цепей и является



важным показателем богатства библиотеки. На следующем шаге используется программа MiXCR [Volotin и др., 2015; Volotin и др., 2017]. Это очень производительный алгоритм выравнивания и кластеризации последовательностей. Он позволяет на каждую консенсусную последовательность полученную на предыдущем шаге выровнять V,D,J сегменты, а также определить границы и последовательность участка CDR3. Понятно, что в результате клональной экспансии клеток отдельных клонов в образце гораздо больше, чем других, соответственно многие молекулы РНК будут кодировать идентичные последовательности CDR3. Для удобства анализа MiXCR позволяет собрать эти идентичные последовательности в «клонотипы», попутно исправляя оставшиеся после предыдущего шага ошибки ПЦР и секвенирования. Результатом этого шага является массив данных («клонсет»), содержащий последовательности найденный в образце клонотипов TCR (определенных как уникальные последовательности CDR3), информацию о выравнивании последовательностей и V(D)J-сегментов, абсолютное число и доля MIG (а значит и молекул кДНК), которое соответствует каждому клонотипу в данных. Визуализация одного из полученных массивов представлена на рис. 11.

Seq. Count (pe... ▾	AA Sequence	N Sequence	CDR3 Le...	V Segment	J Segment
249909 (17.823%)	CASSSGQGMNTEAFF	TGTGCCAGCAGCTCGGGACAGGGTATGAACACTGAAGCTTTCCTT	45	TRBV7-9	TRBJ1-1
77851 (5.552%)	CSAGPGGSEQYF	TGCAGCGCCGGGCCCGGGGAAGCGAGCAGTACTTC	36	TRBV29-1	TRBJ2-7
50042 (3.569%)	CASNGSSYEQYF	TGTGCCAGCAACGGCGGGAGCTCTACGAGCAGTACTTC	39	TRBV28	TRBJ2-7
44288 (3.159%)	CSATRTARNQPQHF	TGCAGTGCCTAGGACGGCGCAATCAGCCCCAGCATTTT	42	TRBV20-1	TRBJ1-5
30133 (2.149%)	CASSTSGGSTDTQYF	TGTGCCAGTACAACAGCGGGGTTCCACAGATACGAGTATTTT	45	TRBV3-1	TRBJ2-3
25095 (1.79%)	CASSLAGQGFNYGYTF	TGTGCCAGCAGCTTAGCCGACAGGGATTTAACTATGGCTACACCTTC	48	TRBV7-9	TRBJ1-2
23128 (1.649%)	CASSLDSSGNTIYF	TGTGCCAGCAGTTAGGGGACAGCTCTGGAACACCATATATTTT	45	TRBV28	TRBJ1-3
22884 (1.632%)	CASSLLAGYNEQFF	TGTGCCAGCAGCCTACTAGCGGGGTACAATGAGCAGTCTTC	42	TRBV7-3	TRBJ2-1
15746 (1.123%)	CASQDPYEQYF	TGTGCCAGTGGTCAGGACCCCTACGAGCAGTACTTC	36	TRBV6-5	TRBJ2-7
14319 (1.021%)	CSATRGF-NQPQHF	TGCAGTGTACGAGGGGCTTCAATCAGCCCCAGCATTTT	40	TRBV20-1	TRBJ1-5
14266 (1.017%)	CASSGQGMNTEAFF	TGTGCCAGCAGCTCGGGACAGGGTATGAACACTGAAGCTTTCCTT	45	TRBV7-9	TRBJ1-1

**Рис. 11.** Пример визуализации обработанного массива данных секвенирования репертуара («клонсета»). Первая колонка – численность (число уникальных последовательностей кДНК) и доля клонотипа TCR в репертуаре. Вторая колонка – аминокислотная последовательность участка CDR3, третья колонка – нуклеотидная последовательность CDR3 участка (для краткости именно ее мы будем обозначать термином «клонотип»). Цветами обозначены лучшие выравнивания V(красный), D(синий) и J(зеленый) сегментов. Четвертая колонка – длина CDR3 участка, не кратные трем длины соответствуют нефункциональным клонотипам со сбитой рамкой считывания. Следующие две колонки – лучшие по результатам выравнивания V и J-сегменты по результатам выравнивания.

С этими данными работать гораздо проще, чем с сырыми данными секвенирования, но они все еще имеют большой объем, как правило размер одного массива превышает 100 тысяч строк, что делает анализ таких данных стандартными офисными пакетами (вроде Open Office Calc или

Microsoft Excel) практически невозможным. Поэтому при моем активном участии был разработан пакет для языка статистического программирования R – *tcR* [Nazarov и др., 2015]. В пакете содержатся функции для выполнения простых операций с клонсетами – подсчета числа идентичных клонотипов между массивами, определение частот использования V- и J- сегментов, разделение функциональных и нефункциональных клонотипов и так далее. Пакет также включает функции для визуализации результатов анализа – построения графиков и диаграмм.

Вторая важная часть этой работы проводилась с помощью программы *repergenNHM* [Elhanati и др., 2015a]. Эта программа позволяет выводить модели сборки TCR и BCR симулировать репертуар с помощью этих моделей, а также оценивать вероятности сборки произвольных последовательностей. Результатом этого шага являются массивы симулированных функциональных и нефункциональных последовательностей с использованием индивидуальной модели сборки для каждого донора. Анализ симулированных данных затем проводился с помощью пакета *tcR*, как и анализ экспериментальных данных.

Таким образом, разработан подход к количественному анализу сырых данных секвенирования репертуаров, включающий коррекцию ошибок и подсчет молекул с помощью UMI, расчет вероятности сборки и выравнивание V,D,J сегментов, анализ числа идентичных клонотипов, числа N-инсерций, распределений частот использования V, D, J - сегментов.

### 4.3. Создание коллекции образцов и секвенирование 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 трех других здоровых доноров (19 лет, 51 год, 57 лет), а также образца пуповинной крови. Все кДНК библиотеки были отсеквенированы на платформе Illumina HiSeq. Полученные данные были депонированы в базу данных SRA (short read archive), идентификатор SRP078490. Число полученных для каждой библиотеки ридов, уникальных молекулярных идентификаторов (UMI) и уникальных клонотипов для каждого из полученных образцов см. в

таблицах 2 (для альфа-цепи) и 3 (для бета-цепи). Обработка сырых данных секвенирования проводилась с использованием пакетов MiGEC и MiXCR, анализ данных производился с использованием пакета *tcR*. Детали пробоподготовки и анализа данных описаны в разделе Материалы и методы.

Следующие разделы посвящены результатам анализа реконструированных репертуаров.

**Таблица 2.** Сводная таблица объема данных секвенирования репертуаров *TCRalpha*

Идентификатор образца	Число ридов	Число UMI	Число уникальных клонотипов CDR3
P1_CD4	6566952	430915	248457
P1_CD8	4620425	378044	162607
P1_unpart	9571058	574439	348419
P1_45RO	4099026	431529	173883
P2_CD4	4269624	941176	432476
P2_CD8	4040615	561437	204094
P2_unpart	8213565	873546	471850
P2_45RO	4608991	653326	228429
Q1_CD4	3894188	653649	277621
Q1_CD8	3201067	589757	147918
Q1_unpart	8360990	1091786	456024
Q1_45RO	3587344	687916	201218
Q2_CD4	3877893	828573	315922
Q2_CD8	3880048	825539	158954
Q2_unpart	9159719	1215155	473672
Q2_45RO	3890664	834828	224276
S1_CD4	4655514	734158	360161
S1_CD8	1009038	219433	105232
S1_unpart	3191701	621723	351923
S1_45RO	4977466	495057	189739
S2_CD4	11727155	761495	348109
S2_CD8	12436797	468345	190534
S2_unpart	11135704	610105	336177
S2_45RO	9064981	633362	228579

*Таблица 3. Сводная таблица объема данных секвенирования репертуаров TCRbeta*

Идентификатор образца	Число ридов	Число UMI	Число уникальных клонотипов CDR3
P1_CD4	3757755	759270	235040
P1_CD8	3565384	517737	204963
P1_unpart	7429601	955106	444708
P1_45RO	4036708	695379	195023
P2_CD4	3042278	449048	475545
P2_CD8	3438238	477696	241048
P2_unpart	8144134	817306	624074
P2_45RO	4598733	578663	249001
Q1_CD4	3694288	673037	386005
Q1_CD8	4586088	758201	237511
Q1_unpart	6511237	1060251	581114
Q1_45RO	3171012	664732	216879
Q2_CD4	3066472	605062	351640
Q2_CD8	3389029	691438	174552
Q2_unpart	7256515	1241753	644594
Q2_45RO	3110044	667997	214628
S1_CD4	3510759	722883	423689
S1_CD8	3162597	489393	248236
S1_unpart	7019324	1181194	673755
S1_45RO	3363725	574876	218185
S2_CD4	4034384	717023	410283
S2_CD8	4267632	546529	258832
S2_unpart	7093628	875357	521882
S2_45RO	2848644	526765	209807
S1_day15_unpart_r1	1982705	1123254	730553
S2_day15_unpart_r1	5467758	1673973	897395
P1_day15_unpart_r1	3542804	1836565	982154
P2_day15_unpart_r1	5157347	2248792	1278149
S1_day15_unpart_r2	2054370	1083361	715091

S2_day15_unpart_r2	5187862	1606488	861442
P1_day15_unpart_r2	3461032	1737376	961138
P2_day15_unpart_r2	5534486	2443931	1352051
Memory_aged19	7486248	424156	149292
Naïve_aged19	9166800	932396	697091
Memory_aged51	4376542	366646	104477
Naïve_aged51	4115592	602950	348005
Memory_aged57	5743372	476395	245092
Naïve_aged57	5227973	358245	422545
Cord_blood	8015355	1803557	1119000

#### 4.4. Анализ числа идентичных клонотипов в репертуарах доноров

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?

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 вызвано 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» мы будем использовать симулированные репертуары TCR, которые можно получить, используя модели сборки и селекции TCR, при этом модели сборки и

селекции будут выведены для каждого донора отдельно, чтобы учесть индивидуальные особенности.

В следующих подразделах отдельно описаны результаты моделирования числа общих функциональных и нефункциональных клонотипов альфа и бета-цепи TCR, и сравнение предсказанных значений с наблюдаемыми в реальных данных.

#### 4.4.1. Анализ числа идентичных нефункциональных клонотипов TCRalpha между донорами

Для предсказания числа идентичных нуклеотидных последовательностей TCR между нефункциональными репертуарами TCRalpha для каждого донора была выведена механистическая модель сборки TCR. Формулировка модели взята из работы [Elhanati и др., 2015a]. На массиве данных каждого донора была выведена своя индивидуальная модель сборки. Метод вывода параметров модели был подробно описан ранее [Elhanati и др., 2015a; Murugan и др., 2012]. Вероятность сборки каждой последовательности в модели рассчитывается как сумма вероятностей сборки рекомбинационных сценариев, которые могут давать данную нуклеотидную последовательность. Для альфа-цепи T-клеточного рецептора использовалась следующая формула для расчета вероятности рекомбинационного сценария:

$$P_{\text{реagg}}^{\alpha}(\tau) = P(V, J)P(\text{del}V|V)P(\text{del}J|J)P(\text{ins}). \quad (7)$$

где  $P(V, J)$  вероятность выбора комбинации V и J сегмента,  $P(\text{del}V|V)$  – вероятность определенного числа делеций при условии выбора V-сегмента,  $P(\text{del}J|J)$  – вероятность определенного числа делеций при условии выбора J-сегмента,  $P(\text{ins})$  – вероятность определенного числа N-инсерций.

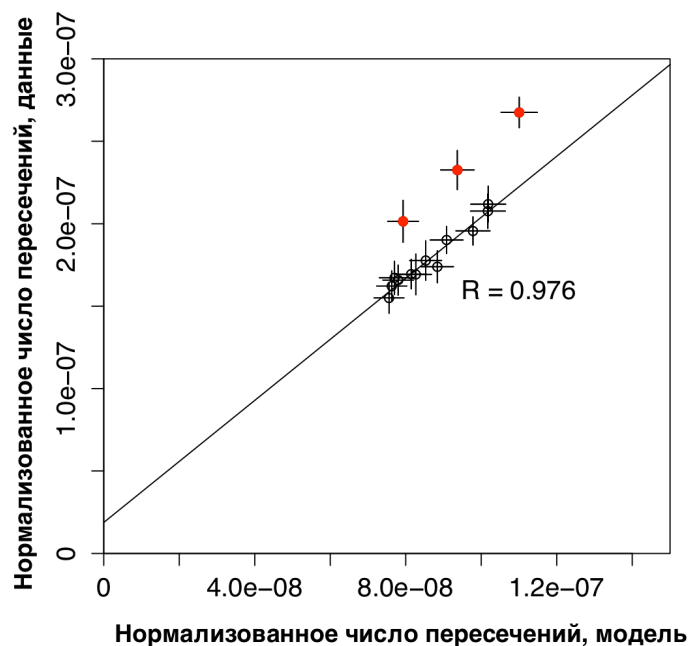
Вывод параметров модели проводили с помощью EM-алгоритма, который итеративно подбирает параметры модели для максимизации правдоподобия набора данных. Для альфа-цепей использовали программу `gergenNHM` [Elhanati и др., 2015a], которая выводит параметры модели сборки с помощью алгоритма Баума-Велша. Параметры модели сборки для каждого донора были выведены на нефункциональном репертуаре TCRalpha.

Известно, что число идентичных клонотипов в двух репертуарах пропорционально произведению размеров этих репертуаров. Коэффициент пропорциональности мы будем называть нормализованным числом общих клонотипов. Предсказание нормализованного числа общих клонотипов проводили с помощью метода Монте-Карло – с помощью модели сборки для каждого человека симулировалась сборка 100000 нефункциональных последовательностей TCR, и

подсчитывалось число идентичных клонотипов в парах таких симулированных нефункциональных репертуаров. Затем нормализованное число общих клонотипов в модели сравнивали с наблюдаемым в реальных данных. Результат этого сравнения показан на рисунке 12, абсолютное число совпадающих клонотипов представлено в таблице 4.

**Таблица 4.** Число идентичных нефункциональных нуклеотидных последовательностей TCRalpha клонотипов (или мощность попарных пересечений) во всех возможных парах репертуаров. На диагонали таблицы показан объем каждого репертуара (число уникальных клонотипов).

	S1	S2	P1	P2	Q1	Q2
S1	33366	250	188	276	280	310
S2	250	37195	220	320	334	361
P1	188	220	33288	383	382	396
P2	276	320	383	49470	524	539
Q1	280	334	382	524	54152	830
Q2	310	361	396	539	830	57291



**Рис. 12:** Нормализованное число идентичных клонотипов в близнецовых парах выше, чем у неродственных доноров. Число общих нефункциональных клонотипов альфа-цепи TCR для 15 возможных пар 6 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.*

Модель систематически недооценивает нормализованное число пересечений в 1.7 раза для неродственных доноров, при этом коэффициент корреляции Пирсона составляет  $r=0.8$  между данными и предсказанием модели.

В репертуарах близнецовых доноров обнаруживается гораздо большее число идентичных нефункциональных клонотипов TCRalpha в данных, чем в предсказании модели сборки. Поскольку использовалась индивидуальная модель сборки для каждого донора, наблюдаемый избыток идентичных клонотипов в парах близнецов не может быть вызван большей схожестью параметров модели сборки.

#### 4.4.2. Анализ числа идентичных нефункциональных клонотипов TCRbeta между донорами

Для последовательностей бета-цепей предсказание числа идентичных нефункциональных последовательностей представляет более сложную задачу. Из-за аллельного исключения большинство нефункциональных бета-клонотипов не экспрессируется и отсутствует в кДНК библиотеках. В то же время средняя вероятность сборки ниже для бета-цепей из-за их большего разнообразия: в генах бета-цепи между V и J присутствует D-сегмент, и есть два участка на которых могут происходить инсерции случайных нуклеотидов – между V и D, и между D и J-сегментами. Вместе эти два фактора (большее разнообразие бета-цепей и меньший объем доступных нефункциональных репертуаров) приводят к очень низкому числу наблюдаемых совпадений клонотипов между нефункциональными репертуарами.

Для бета-цепи использовалась более сложная модель сборки, чем для альфа-цепи. Вероятность рекомбинационного сценария в ней представлена формулой:

$$P_{\text{rearr}}^{\beta}(r) = P(V, D, J)P(\text{del}V|V)P(\text{ins}VD) \times P(\text{del}Dl, \text{del}Dr|D)P(\text{ins}DJ)P(\text{del}J|J) \quad (8)$$

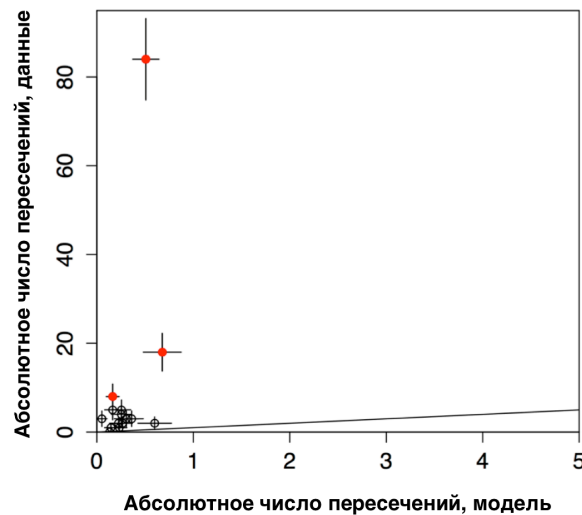
где  $P(V, D, J)$  - вероятность выбора комбинации V, D и J сегментов,  $P(\text{del}V|V)$  – вероятность определенного числа делеций при условии выбора V-сегмента,  $P(\text{del}J|J)$  – вероятность определенного числа делеций при условии выбора J-сегмента,  $P(\text{ins}VD)$  – вероятность определенной последовательности N-инсерций на стыке между V и D сегментами,  $P(\text{ins}DJ)$  – вероятность определенной последовательности N-инсерций на стыке между D и J сегментами,



$P(\text{del}D_l, \text{del}D_r|D)$  – вероятность определенного числа 3' и 5' делеций  $D$  при условии выбора определенного  $D$ -сегмента).

Параметры модели выводились напрямую с использованием EM-алгоритма, путем перебора всех возможных сценариев сборки для каждой последовательности, как описано в [Mugan и др., 2012]. Эта процедура позволяет получить вероятности отдельных событий сборки с большой точностью, несмотря на то, что конкретный сценарий сборки невозможно однозначно восстановить для каждой отдельной последовательности из данных - как правило каждая отдельная последовательность может собираться многими способами.

Для каждого донора было сгенерировано 100000 симулированных последовательностей TCRbeta. Далее было подсчитано нормализованное число идентичных клонотипов. Нормализованное число идентичных клонотипов позволяет предсказать абсолютное число идентичных клонотипов между репертуарами известного размера, что и было сделано: для каждой пары доноров нормализованное число идентичных клонотипов в модели было умножено на произведение объемов наблюдаемых нефункциональных репертуаров этой пары доноров. Сравнение получившейся величины с наблюдаемым числом пересечений между нефункциональными репертуарами показано на рис. 13, абсолютное число совпадающих клонотипов представлено в таблице 5.



**Рис. 13.** Абсолютное число общих нефункциональных клонотипов бета-цепи TCR (ось  $y$ ) для 15 возможных пар 6 доноров сравнивается с предсказанием модели сборки (ось  $x$ ). Три красных круга - близнецовые пары. Линии у каждой из точек показывают одно стандартное отклонение. Диагональная линия  $y=x$ .

Видно, что предсказание модели во всех случаях меньше единицы. При этом для одной из близнецовых пар наблюдается особенно много идентичных последовательностей – более 80. Абсолютное число общих клонотипов между парами нефункциональных репертуаров TCRbeta (см. таблицу 5) гораздо меньше, чем для нефункциональных репертуаров альфа-цепей (см. таблицу 4). Это приводит к тому, что возрастает вклад разного рода артефактов и постороннего шума в наблюдаемое число общих клонотипов в экспериментальных данных. Из данных были удалены 26 нефункциональных последовательностей, которые встретились более чем в двух донорах. Эти последовательности скорее всего появляются из-за воспроизводимых ошибок алгоритма выравнивания или представляют собой артефакты пробоподготовки. Например, некоторые удаленные последовательности содержат фрагменты интронов, которые должны отсутствовать в кДНК библиотеке.

**Таблица 5.** Число идентичных нефункциональных нуклеотидных последовательностей TCRbeta клонотипов (т.е. мощность попарных пересечений) во всех возможных парах репертуаров. На диагонали таблицы показан объем каждого репертуара (число уникальных клонотипов).

	S1	S2	P1	P2	Q1	Q2
S1	22137	5	3	2	4	5
S2	5	7880	0	0	0	2
P1	3	0	15109	10	5	1
P2	2	0	10	16272	4	1
Q1	4	0	5	4	16697	84
Q2	5	2	1	1	84	17763

Несмотря на меньшее абсолютное число пересечений, сила изучаемого эффекта, т.е. относительный избыток числа идентичных клонотипов по сравнению с предсказанием индивидуальных моделей сборки здесь больше, чем в нефункциональных репертуарах альфа-цепей – 84 клонотипа вместо в среднем приблизительно одного оказываются идентичными у одной из близнецовых пар. То есть всего несколько из этих 84 клонотипов могут быть объяснены конвергентной сборкой (и это в целом соответствует наблюдаемому числу идентичных нефункциональных TCRbeta клонотипов у неродственных доноров).

Таким образом, и для нефункциональных клонотипов бета-цепи число идентичных клонотипов в парах близнецов превосходит как число идентичных клонотипов между неродственными людьми, так и предсказание индивидуальных моделей сборки.

#### 4.4.3. Анализ числа идентичных клонотипов между функциональными репертуарами

Функциональный репертуар отличается от нефункционального репертуара последовательностей со сдвигом рамки считывания. Очевидное отличие - в распределении длин последовательностей гипервариабельного участка CDR3, нефункциональному репертуару отвечает больший диапазон длин участка CDR3, в то время как для функциональных последовательностей его длина всегда кратна трем, по свойству триплетности генетического кода. Но есть еще одно существенное отличие, которое дополнительно снижает потенциальное разнообразие функционального репертуара - селекция в тимусе. Для того, чтобы учесть селекцию в работе [Elhanati и др., 2014] было предложено ввести фактор селекции  $Q$ , зависящий от последовательности CDR3.  $Q$  определяется как отношение вероятности увидеть последовательность в функциональном репертуаре к вероятности сборки этой последовательности, при этом параметры вероятности сборки выводятся на нефункциональном репертуаре (см. выше). Фактор селекции  $Q$  представляется как произведение факторов  $q_i$ , при этом величина  $q_i$  зависит от аминокислоты и ее положения в CDR3 определенной длины:

$$Q \propto q_L \prod_{i=1}^L q_{i;L}(a_i) \quad (9)$$

Для альфа- и бета-цепи каждого донора была выведена своя модель селекции, т.е. свой набор факторов  $q_i$ . Модель селекции позволяет рассчитывать коэффициент селекции  $Q$ , на который нужно поправить вероятность сборки.

Для того, чтобы точно предсказать число общих функциональных последовательностей необходимо ввести дополнительную поправку на селекцию в тимусе. Симуляция последовательностей TCR происходила так же, как и в случае нефункциональных последовательностей. Далее для каждой последовательности рассчитывалось значение фактора селекции  $Q$ . Предсказанное нормализованное число общих последовательностей рассчитывалось по формуле:

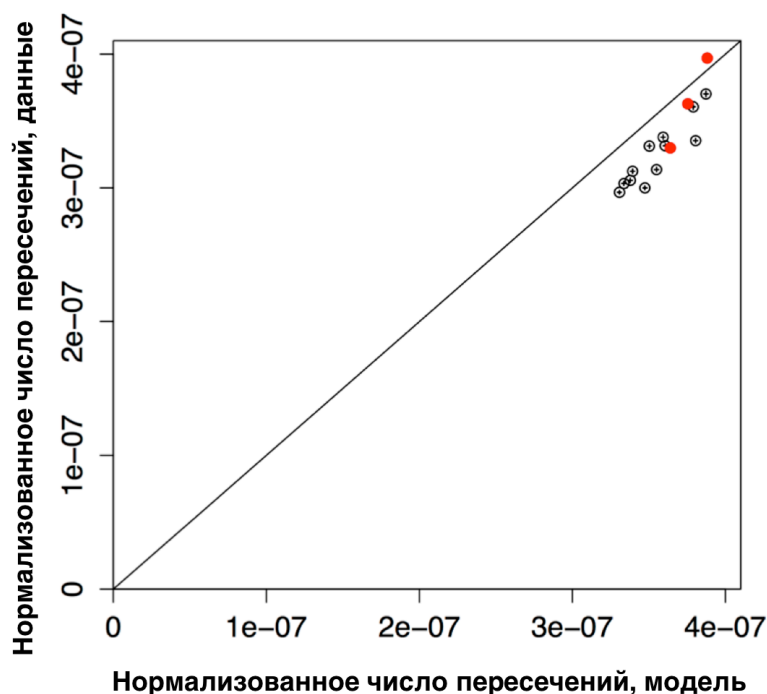
$$\frac{1}{|S_1| \cdot |S_2|} \sum_{s \in S_1 \cap S_2} Q^{(1)}(s) Q^{(2)}(s), \quad (10)$$

где  $S_1$  и  $S_2$  это два симулированных с использованием индивидуальных моделей сборки репертуара,  $Q_1$  и  $Q_2$  - индивидуальные модели селекции.  $|S_1|$  и  $|S_2|$  обозначают размеры двух

массивов. Суммирование проходит только по последовательностям присутствующим в обоих симулированных репертуарах.

#### 4.4.4. Анализ числа идентичных функциональных клонотипов TCRalpha между донорами

Нормализованное число идентичных клонотипов предсказанных описанным выше образом сравнивается с наблюдаемым в реальных данных на рис. 14.



*Рис. 14. Нормализованное число идентичных функциональных альфа-клонотипов в близнецовых парах такое же, как у неродственных доноров. Число общих нефункциональных клонотипов альфа-цепи TCR для 15 возможных пар 6 доноров сравнивается с предсказанием модели сборки. Чтобы сравнивать число идентичных клонотипов в выборках разного размера, число идентичных клонотипов нормализуют на произведение объемов выборок. Три красных круга - близнецовые пары. Диагональная линия  $y=x$ .*

Предсказание хорошо сходится с данными, однако число идентичных клонотипов в парах близнецов здесь не больше, чем в парах неродственных доноров. Это может объясняться большим числом функциональных последовательностей TCRalpha совпадающих по механизму конвергентной рекомбинации. Эффект пренатального обмена становится несущественным при

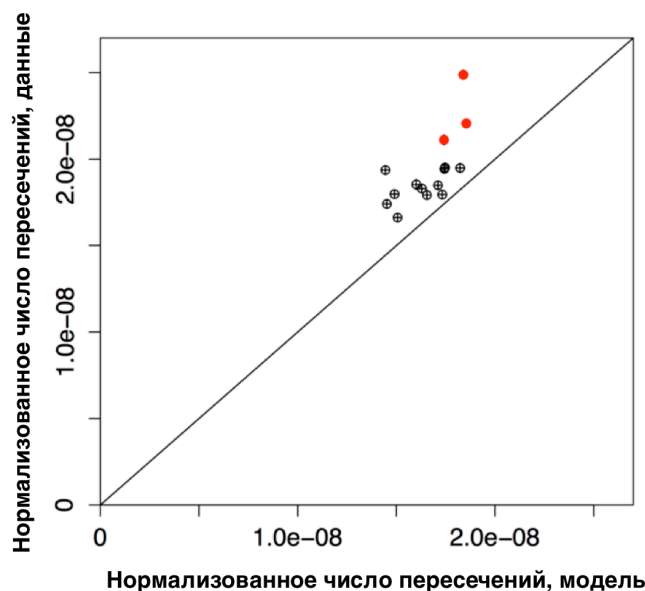
таком уровне случайных совпадений. Абсолютное число идентичных последовательностей приведено в таблице 6.

**Таблица 6.** Число идентичных функциональных нуклеотидных последовательностей TCRalpha клонотипов (или мощность попарных пересечений) во всех возможных парах репертуаров. На диагонали таблицы показан объем каждого репертуара (число уникальных клонотипов).

	S1	S2	P1	P2	Q1	Q2
S1	338442	34283	32779	40351	35366	36581
S2	34283	323226	31612	38658	34158	35380
P1	32779	31612	335107	43773	38449	39539
P2	40351	38658	43773	454194	46129	47466
Q1	35366	34158	38449	46129	436633	50366
Q2	36581	35380	39539	47466	50366	453978

#### 4.4.5. Анализ числа идентичных нефункциональных клонотипов TCRalpha между донорами

Для функциональных клонотипов бета-цепи сравнение предсказания модели с реальными данными приведено на рис. 15.



**Рис. 15.** Нормализованное число идентичных функциональных клонотипов бета-цепи в близнецовых парах выше, чем у неродственных доноров. Число общих нефункциональных клонотипов бета-цепи TCR для 15 возможных пар 6 доноров сравнивается с предсказанием модели сборки. Чтобы сравнивать число идентичных клонотипов в выборках разного размера, число идентичных клонотипов нормализуют на произведение объемов выборок. Три красных круга - близнецовые пары. Диагональная линия  $y=x$ .

Предсказание также неплохо сходится с данными, но видно, что в близнецовых парах наблюдается некоторый избыток идентичных клонотипов. Стоит отметить, что нормализованное число пересечений между неродственными донорами, соответствующее числу совпадений за счет конвергентной рекомбинации здесь почти на порядок меньше, чем при сравнении функциональных репертуаров альфа-цепей. Можно предположить, что из-за этого эффект пренатального обмена здесь обнаружить проще, так как его относительный вклад в число идентичных клонотипов в парах близнецов здесь выше, чем для функциональных репертуаров альфа-цепи. Абсолютное число идентичных клонотипов приведено в таблице 7.

*Таблица 7. Число идентичных функциональных нуклеотидных последовательностей TCRbeta клонотипов (или мощность множеств попарных пересечений) во всех возможных парах репертуаров. На диагонали таблицы показан объем каждого репертуара (число уникальных клонотипов).*

	S1	S2	P1	P2	Q1	Q2
S1	673755	1491	1174	1084	1515	1522
S2	1491	521882	1106	1113	1430	1524
P1	1174	1106	444708	1187	1271	1349
P2	1084	1113	1187	624074	1223	1315
Q1	1515	1430	1271	1223	581114	1880
Q2	1522	1524	1349	1315	1880	644594

Таким образом, наблюдается избыток числа идентичных клонотипов между функциональными репертуарами бета-цепи, но не альфа-цепи, что может быть связано с большим вкладом конвергентной рекомбинации в совпадение функциональных последовательностей TCRalpha близнецовых доноров.

#### 4.4.6. Анализ возможного вклада популяций Т-клеток со сниженным разнообразием

Ранее было показано, что существуют субпопуляции Т-клеток (MAIT и iNKT), которые имеют инвариантную альфа-цепь с очень низким разнообразием. Для этих альфа-цепей всегда выбираются определенные V-J комбинации (TRAV10/TRAJ18 для iNKT и TRAV1-2/TRAJ33 для MAIT), а кроме того отсутствуют вставочные нуклеотиды. Чтобы проверить, могут ли подобные клонотипы повлиять на результаты приведенного выше анализа, мы провели поиск известных последовательностей TCR iNKT-лимфоцитов и MAIT-лимфоцитов собранных в работе [Greenaway

и др., 2013] в наших данных. 25 из 27 известных MAIT последовательностей были найдены в массивах как минимум однажды (21 из них обнаружилось во всех 6 индивидах), и 8 из 13 известных iNKT последовательностей (1 из них во всех шести индивидах). Ранги найденных в репертуарах наших доноров известных [Greenaway и др., 2013] клонотипов MAIT и iNKT представлены в таблице 8 и таблице 9 соответственно. Интересно, что некоторые MAIT-, а также iNKT-клонотипы достаточно многочисленны, и часто входят в десятку самых высокопредставленных клонотипов репертуара TCRalpha.

MAIT и iNKT последовательности действительно присутствуют в наших данных, но только несколько общих последовательностей соответствуют им, поэтому мы не исключали эти последовательности из анализа. Большинство общих последовательностей с нулем инсерций наблюдаемых в наших данных не могут быть отнесены к известным клонотипам MAIT или iNKT-клеток.

**Таблица 8.** Ранг известных последовательностей клонотипов TCRalpha MAIT-лимфоцитов в репертуарах TCRalpha

CDR3aa	S1	S2	P1	P2	Q1	Q2
CAVMDSNYQLIW	1	1	8	2	3	8
CAVRDSNYQLIW	2	2	6	3	4	14
CAVLDSNYQLIW	3	9	10	6	9	28
CAVTDSNYQLIW	7	12	23	22	23	65
CAAMDSNYQLIW	9	10	18	19	24	36
CAVKDSNYQLIW	11	7	27	27	36	35
CAVVDSNYQLIW	16	16	19	38	19	41
CAGMDSNYQLIW	17	70	65	59	87	559
CAALDSNYQLIW	21	20	26	26	76	233
CASMDSNYQLIW	33	30	36	81	57	239
CAAVDSNYQLIW	48	243	3527	535	722	2424
CAVGDSNYQLIW	72	25	67	111	208	901
CAFMDSNYQLIW	99	75	1881	1768	149	263
CAVQDSNYQLIW	133	54	209	70	95	384
CAPLDSNYQLIW	168	171	189	229	2714	5980
CAAIDSNYQLIW	246	21060	9531	16233	15147	4078
CAVRDRDYQLIW	465	342	2101	4825	2877	100233
CAGLDSNYQLIW	587	45	196	172	789	8966
CAVIDSNYQLIW	671	87	45	325	12	5491
CAAEDSNYQLIW	675	7222	4094	12808	96961	24958
CATMDSNYQLIW	929	57	565	1450	1831	1155
CAIMDSNYQLIW	1512	320	9531	394	18484	24958

CALLDSNYQLIW	2824	21060	9531	428	2301	1505
CAGWDSNYQLIW	22894	14861	66396	108398	NA	NA
CARSDSNYQLIW	NA	NA	NA	NA	NA	NA
CASIDSNYQLIW	NA	NA	66396	7358	NA	NA
CAVMDDNYQLIW	NA	NA	NA	NA	NA	NA

*Таблица 9. Ранг известных последовательностей клонотипов TCRalpha iNKT-лимфоцитов в репертуарах TCRalpha.*

CDR3aa	S1	S2	P1	P2	Q1	Q2
CVVSDRGSTLGRLYF	12	11	2	1	10	106
CVVTDRGSTLGRLYF	NA	NA	18426	79	NA	NA
CVVGDRGSTLGRLYF	1626	73883	66396	2831	NA	19747
CVVFDRGSTLGRLYF	NA	NA	309	7358	6107	NA
CVVNDRGSTLGRLYF	NA	NA	2101	10435	NA	NA
CVVARDRGSTLGRLYF	NA	NA	NA	108398	NA	NA
CVVSRGSTLGRLYF	NA	34402	NA	NA	96961	NA
CVVRADRGSTLGRLYF	NA	NA	NA	NA	NA	100233
CVASDRGSTLGRLYF	NA	NA	NA	NA	NA	NA
CVSVDRGSTLGRLYF	NA	NA	NA	NA	NA	NA
CVVSDKGSTLGRLYF	NA	NA	NA	NA	NA	NA
CVVSATNRGSTLGRLYF	NA	NA	NA	NA	NA	NA
CVVPGRLYF	NA	NA	NA	NA	NA	NA

#### 4.5. Анализ последовательностей CDR3 клонов идентичных между донорами

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.

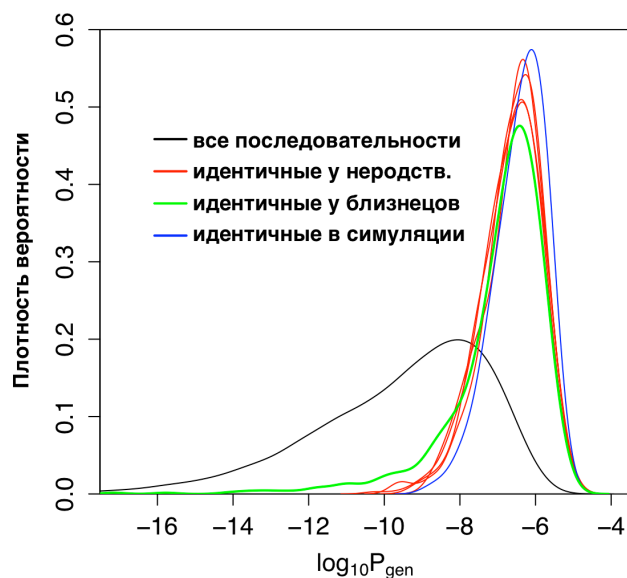


И наконец, если часть идентичных клонотипов близнецов по сути являются потомками одного клона Т-клеток, то в этом случае оба этих клонотипа будут принадлежать к одной субпопуляции: CD4+, либо CD8+ Т-клеток. Кроме того, можно предположить, что детектируемые спустя много лет после обмена пуповинной кровью 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.

В следующих подразделах будет проведена проверка этих гипотез.

#### 4.5.1. Анализ вероятности сборки идентичных клонотипов TCR

Помимо возможности симулировать репертуар TCR, модель предоставляет оценку вероятности сборки для каждого нуклеотидного варианта Т-клеточного рецептора.



**Рис. 16:** Последовательности TCR идентичные у монозиготных близнецов отличаются от идентичных в неродственных индивидах. Распределение  $\log_{10} P_{gen}$ , где  $P_{gen}$  это вероятность сборки последовательности Т-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 маловероятных последовательностей. Для сравнения черным показано распределение вероятностей сборки для всех клонов в репертуаре.

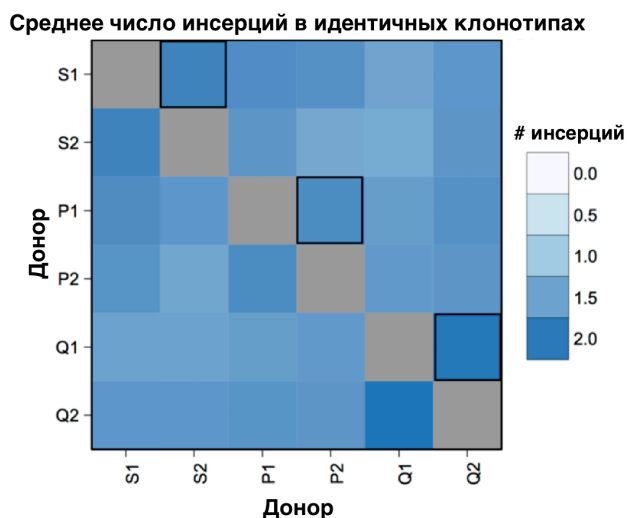
Распределения вероятностей сборки для нефункционального альфа-репертуара Т-клеточных рецепторов одного из доноров приведены на рис. 16. Черная кривая - плотность распределения вероятности сборки всех рецепторов в секвенированном репертуаре. Очевидно, что у рецепторов, совпавших между донорами по механизму конвергентной рекомбинации (т.е. собравшихся независимо два раза), частота сборки должна быть гораздо выше, чем в среднем по репертуару, что и наблюдаем: красные кривые обозначают вероятность сборки у клонов общих с неродственными людьми. С помощью модели сборки можно получить теоретическое предсказание для красных кривых. Синяя кривая - это плотность распределения вероятностей для клонов общих в двух симулированных репертуарах, для симуляции использовались модели сборки близнецов. Зеленая кривая - клоны общие с близнецовым донором. Она отличается от теоретического предсказания и от данных по неродственным индивидам наличием длинного хвоста клонов с низкой вероятностью сборки. Таким образом, у близнецов совпадают последовательности Т-клеточных рецепторов, которые должны редко собираться независимо.

#### 4.5.2. Анализ числа инсерций в идентичных клонотипах TCR

Результаты, описанные в предыдущем разделе, были получены с использованием сложной модели сборки, которую мы обучили на индивидуальных репертуарах нефункциональных последовательностей. Чтобы подтвердить это наблюдение способом независимым от обучаемой модели сборки, мы использовали свойство последовательностей TCR, которое негативно коррелирует с вероятностью сборки - число инсерций в гипервариабельном участке CDR3alpha. Репертуары альфа-цепей использовались потому, что общее число инсерций там подсчитать гораздо проще – участок с инсерциями не разделен коротким D-сегментом, для которого сложно получить однозначное выравнивание.

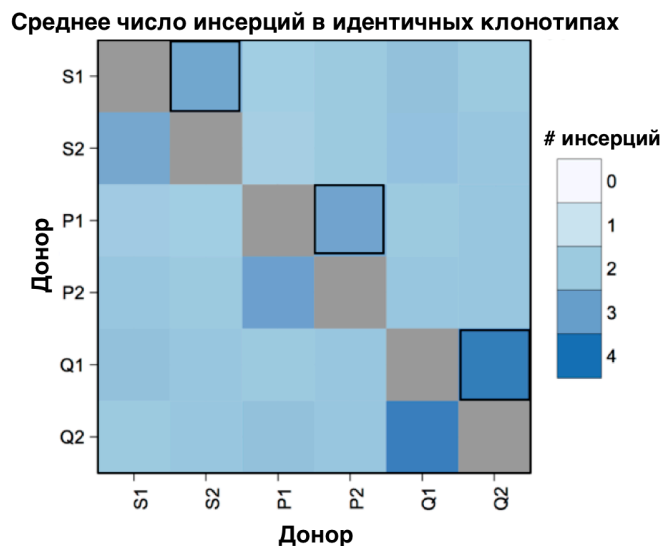
Было подсчитано среднее число инсерций в нефункциональных клонотипах TCRalpha идентичных между всеми возможными парами доноров. В графическом виде эти величины представлены на рис. 17.

Как и ожидалось, среднее число инсерций в гипервариабельном участке CDR3 гораздо меньше для общих клонотипов, чем в среднем в репертуаре каждого индивида. В то же время среднее число инсерций в общих клонотипах было значимо выше для близнецовых пар ( $p=1.83 \times 10^{-8}$ , двусторонний t-тест).



**Рис. 17.** Среднее число инсерций нефункциональных клонотипов альфа-цепи идентичных в парах репертуаров доноров. Значения для пар близнецов выделены черной рамкой.

Такая же процедура была проведена для нефункциональных репертуаров альфа-цепи CD45RO+ Т-клеток памяти. Аналогичный и даже более сильный эффект наблюдался для клеток памяти (CD45RO+),  $p < 10^{-16}$ , двусторонний t-тест, см. рис 18.



**Рис. 18.** Среднее число инсерций нефункциональных клонотипов альфа-цепи идентичных в парах репертуаров CD45RO+ клонотипов доноров. Значения для пар близнецов близнецов выделены черной рамкой.

Таким образом, клонотипы общие между близнецовыми донорами в среднем имеют более высокое число инсерций, чем клонотипы общие между неродственными индивидами, и этот эффект сильнее выражен в репертуарах Т-клеток памяти.

#### 4.5.3. Анализ распределения идентичных между близнецовыми донорами клонотипов TCR по фенотипам

В одной из близнецовых пар количество совпадающих клонотипов в бета-цепях было особенно велико: 84, притом, что модель сборки предсказывала менее одного идентичного клонотипа между выборками такого размера. Мы можем предположить, что почти все эти клонотипы являются результатом обмена Т-клетками до рождения, а значит у них должны быть идентичные фенотипы в обоих донорах.

Мы провели поиск этих 84 общих клонов в отсековированных репертуарах CD4+, CD8+ и CD45RO+ клеток обоих доноров, чтобы установить возможный фенотип этих общих клонов.

44 CDR3 были найдены в CD8 субпопуляциях обоих доноров, и 5 последовательностей в CD4 репертуарах обоих доноров. 25 последовательностей из 44 CD8+ и 3 из 5 CD4+ были также обнаружены в CD45RO+ субпопуляции. Только 3 последовательности имели дискордантный в двух близнецах фенотип CD4-CD8 (к примеру, CD4 в одном близнеце и CD8 в другом), а 2 последовательности вообще не удалось обнаружить в CD4, CD8 или CD45RO субпопуляциях ни одного из близнецов. Для оставшихся 32 общих последовательностей CD4/CD8 фенотип был определен только для одного из доноров из-за ограниченной глубины секвенирования.

**Таблица 10.** Фенотип нефункциональных TCRbeta клонов идентичных между донорами Q1 и Q2.

Q1\Q2	CD4	CD8	N/A
CD4	5	2	6
CD8	1	44	9
N/A	3	6	8

Если бы данные нефункциональные последовательности совпали бы из-за конвергентной рекомбинации, совпадение фенотипов также было бы случайным, и доля клонотипов с идентичным фенотипом была бы гораздо ниже в паре доноров (около 50%, а в данном случае наблюдаем около  $(44+5)/(44+5+3)=94\%$ ). Получить такой результат невозможно при случайном совпадении последовательностей TCR по механизму конвергентной рекомбинации: согласно этому механизму, совпадение последовательности CDR3 в нефункциональных клонотипах случайно, и мы можем ожидать равномерного распределения идентичных нефункциональных последовательностей по фенотипам. Например, в одном доноре данный нефункциональный вариант мог собраться в предшественнике CD8+ клетки, а в другом доноре – в предшественнике CD4+ клетки, и наоборот. Интересно также, что существенная часть обнаруженных общих клонов имеет CD45RO+ фенотип, характерный для Т-клеток центральной памяти (T<sub>CM</sub>).

Таким образом, идентичные нефункциональные клонотипы TCRbeta в одной паре близнецов в большинстве случаев имеют идентичные фенотипы, что не согласуется с механизмом конвергентной рекомбинации.

#### 4.5.4. Некоторые следствия из гипотезы пренатального обмена Т-клетками

Итак, у близнецов больше общих нефункциональных клонотипов и эти клонотипы имеют необычные с точки зрения конвергентной рекомбинации свойства – низкую вероятность сборки, большое число вставок и идентичные фенотипы в разных донорах. Совокупность этих фактов позволяет предположить, что причина совпадения последовательностей этих аномальных Т-клеточных рецепторов не в конвергентной рекомбинации. Было показано, что в большей части беременностей монозиготными близнецами эмбрионы имеют общую плаценту. При этом между кровеносными системами эмбрионов могут формироваться анастомозы, которые обнаруживаются в большом числе случаев. Это делает возможным внутриутробный обмен кровью между монозиготными близнецами.

Иногда этот обмен настолько интенсивен, что приводит к неравномерному развитию патологий, таких как TTFS (Twin-to-twin transfusion syndrome). Такой обмен кровью - особенность именно монозиготных близнецов. В случаях, когда он происходит у dizиготных близнецов [Viran и др., 2011], это приводит к микрохимеризму и реакциям подобным классической реакции трансплантат против хозяина в случае несовместимости HLA-гаплотипов. У монозиготных близнецов этот обмен не должен приводить к конфликту: из-за схожей системы селекции их иммунные системы должны быть полностью совместимы. Можно предположить, что обнаруживаемые во взрослых близнецовых донорах идентичные клонотипы с низкой вероятностью сборки образовались еще до рождения в одном из близнецов и попали в другого через анастомоз между плацентами вместе с пуповинной кровью. Эта гипотеза объясняет совпадение фенотипов общих клонов – т.к. обмен происходил уже зрелыми наивными Т-клетками, которые экспрессируют CD4 или CD8 маркер. Понятно, что у таких идентичных клонов может быть низкая вероятность сборки, как у большинства клонов в репертуаре. Удивительно то, что эти клоны просуществовали в организме 25 лет. Это может быть связано с приобретением этими клонами фенотипа клеток памяти в раннем детстве. Действительно, существенная часть обнаруженных нами клонотипов имеет фенотип клеток памяти в обоих донорах. С другой стороны, приобретение фенотипа памяти происходит после клональной экспансии, что приводит к повышению концентрации клона в репертуаре, и соответственно большей вероятности обнаружить его нашим методом. Иными словами, нам сложно обнаружить общий наивный клон, которым

обменялись близнецовые доноры до рождения, и легко обнаружить размножившийся в обоих донорах после инфекции клон Т-клеток памяти.

Возможны и другие объяснения появления одинаковых клонотипов в репертуарах разных доноров. Предполагалось, что часть клонов Т-клеток из репертуара матери могут проникать через плацентарный барьер и приживаться в иммунной системе плода [Burlingham, Nelson, 2012; Götherström и др., 2005; Mold и др., 2008], но в нашей предыдущей работе не было найдено никаких подтверждений этой гипотезы [Putintseva и др., 2013].

Какое значение обмен Т-клетками может иметь для иммунной системы близнецов?

Он может приводить к тому, что иммунный ответ близнецов будет идентичным, при том что отвечающий клон будет иметь низкую вероятность сборки, которая исключает его появление в других индивидах. Таким образом близнецы могут развивать идентичный друг с другом, но уникальный для данной близнецовой пары иммунный ответ. Мы наблюдаем последствия этого эффекта в наших данных: часть клонов имеет идентичный фенотип клеток памяти (CD45RO+), что говорит об их участии в иммунном ответе в прошлом.

С другой стороны, может произойти обмен клонами, которые в дальнейшем станут причиной аутоиммунного или онкологического заболевания (лейкоза). Последнее несколько раз наблюдали экспериментально, близнецы конкордантные по детскому лейкозу имели идентичную перестройку в злокачественном клоне [Ford и др., 1993]. Для большинства аутоиммунных заболеваний известно, что вероятность заболевания обоих близнецов высокая [Bogdanos и др., 2012], но связь с возможной внутриутробной передачей аутоиммунных клонов не прослежена.

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

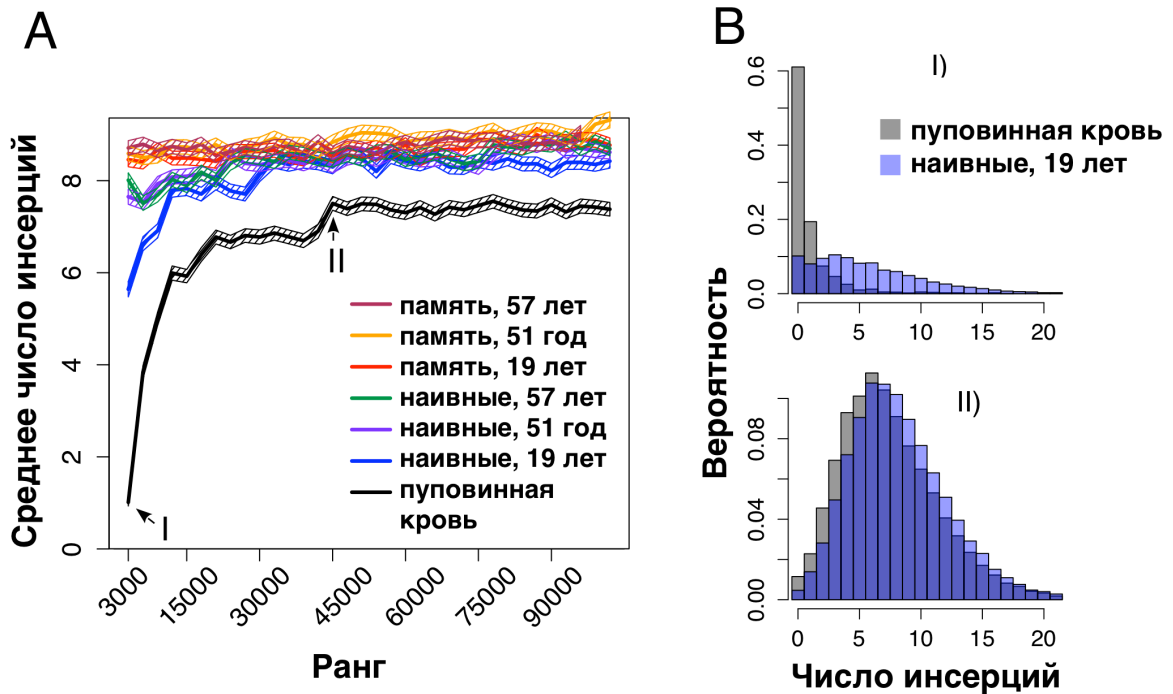
В предыдущей главе нами были представлены доказательства гипотезы внутриутробного обмена Т-клетками между монозиготными близнецами. Из этого следует, что по крайней мере часть репертуара Т-клеточных рецепторов сформировавшегося еще до рождения сохраняется в организме как минимум несколько десятков лет.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

Было показано, что один из ключевых ферментов V(D)J-рекомбинации, терминальная дезоксирибонуклеотидилтрансфераза (TdT) отсутствует у плода в первом триместре беременности [George, Schroeder, 1992]. Таким образом, первые Т-клетки не имеют N-вставок в

последовательностях TCR. Нами были реконструированы репертуар TCRbeta клеток пуповинной крови, а также наивных T-клеток и клеток памяти трех взрослых доноров 19, 51 и 57 лет соответственно. Мы проанализировали количество и распределение клонов с нулем вставок в репертуаре. Для этого мы разбили каждый репертуар по последовательные подгруппы клонов с убывающей численностью, по 3000 клонотипов в каждой.

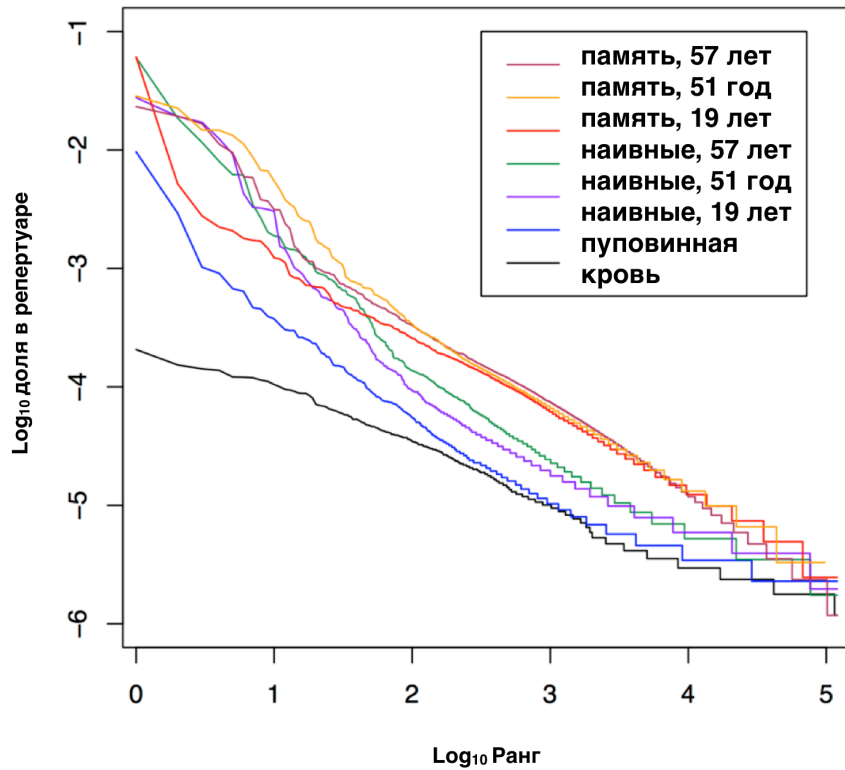
Из-за небольшого размера ранговых групп (2000-3000 тысячи последовательностей) полная модель сборки для каждой группы не выводилась. Вместо этого были использованы ранее опубликованные параметры модели сборки бета-цепи T-клеточного рецептора [Murgan и др., 2012]. Использовались частоты выбора V- и J-сегментов, а также профили делеций V- и J-сегментов из этой модели, а  $P(\text{insVD})$  и  $P(\text{insDJ})$  выводили из данных для каждой группы отдельно. Выведенные распределения чисел инсерций в каждой группе использовались для выявления зависимости среднего числа инсерций от ранга клона (рис. 19А).

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. 19В. Интересно, что такая же зависимость характерна для репертуаров наивных клеток взрослых доноров, но совершенно нехарактерна для клеток памяти. Зависимость частоты клона от его ранга представлена на рис. 20. Клоны репертуаров клеток памяти в среднем имеют большие концентрации, чем наивные клоны того же ранга, что является последствием клональной экспансии. Похожие эффекты наблюдались ранее в работе [Quigley и др., 2010]. Высокие концентрации нескольких наивных клонов могут объясняться невысокой точностью разделения популяций клеток на магнитных колонках.



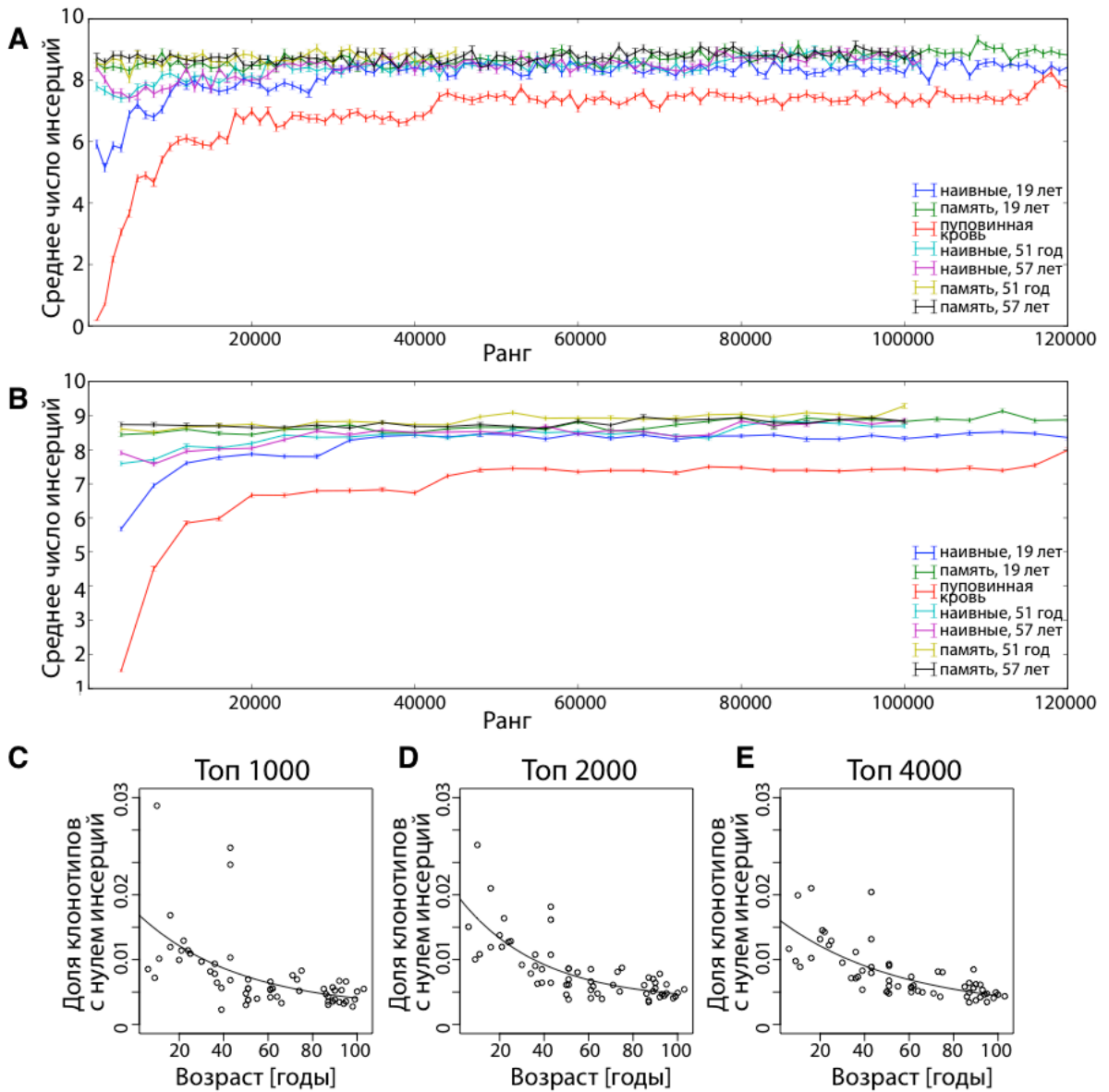
**Рис. 19.** Число *N*-вставок в функциональных клонотипах бета-цепи *T*-клеточного рецептора зависит от представленности этих клонотипов. *A.* Среднее число инсерций было получено для последовательных групп из 3000 клонов, клоны в репертуаре отсортированы по представленности. Клонотипы пуповинной крови (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). Заштрихованная область показывает две стандартных ошибки среднего. *B.* 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. Аналогичный анализ для других опубликованных репертуаров пуповинной крови и других размеров групп показан на рис. 21 и рис. 22.





**Рис. 20.** Зависимость ранга клона (ось  $x$ ) от концентрации клона (ось  $y$ ). Клоны клеток памяти как правило имеют большие концентрации чем наивные клоны и клоны из образца пуповинной крови такого же ранга, что может быть связано с историей клональных экспансий.

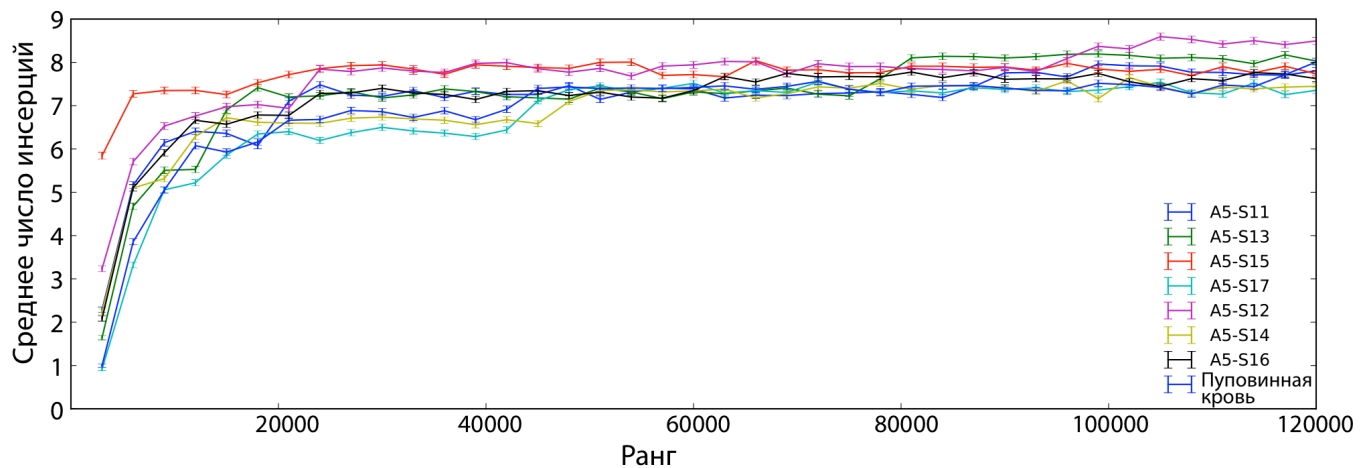
Важно отметить, что размер наблюдаемого эффекта (среднее число инсерций гораздо ниже в группе высокопредставленных клонотипов) будет зависеть от выбранного разбиения рангов по группам. Анализ был повторен с другими близкими разбиениями и был обнаружен тот же качественный эффект (см. рис. 21).



**Рис. 21.** Зависимость среднего числа инсерций от ранга и возраста работает для разных диаметров разбиения. Среднее число инсерций для ранговых групп размером 1000 (А) и 4000 (В). Метод построения такой же как на рис. 19. С, D, E – варианты разбиений для рис. 23, 1000, 2000 и 4000 наиболее представленных клонотипов из каждого репертуара соответственно, черные кривые соответствуют наложенным на данные уравнениям экспоненциального распада. Частота распада для 1000 наиболее представленных и 4000 наиболее представленных  $0.0218 \text{ z}^{-1}$  и  $0.0184 \text{ z}^{-1}$ , что находится в пределах одной стандартной ошибки от значения для 2000 наиболее высокопредставленных клонотипов  $0.0272 \pm 0.0091 \text{ z}^{-1}$

Разбиение на более крупные группы ослабляет наблюдаемый эффект. Также было показано, что наблюдаемое различие в числе инсерций не является отличительной особенностью

конкретного образца пуповинной крови. На рис. 22 представлены результаты для 7 опубликованных репертуаров пуповинной крови из [Britanova и др., 2016]. Среднее число инсерций в каждом из них значительно ниже для крупных клонотипов, чем для небольших.



**Рис. 22.** Зависимость между рангом клона и средним числом инсерций для семи опубликованных репертуаров TCR образцов пуповинной крови из работы [Britanova и др., 2016]. Среднее число инсерций получено при анализе ранговых групп размером 3000 клонотипов, как на рис. 19А. Аналогичный тренд на увеличение числа инсерций с увеличением ранга наблюдается для всех образцов.

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 пуповинной крови неясны. 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 [Howie и др., 2015], либо секвенирование большого числа отдельных клеток могут прояснить этот вопрос.

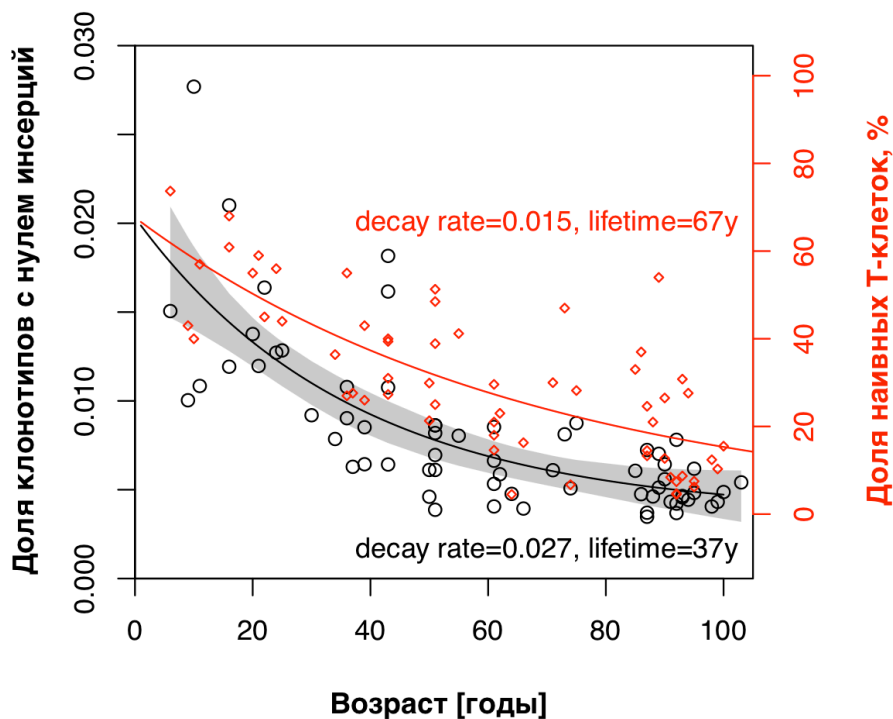
В целом, из этих данных можно сделать вывод, что наиболее высокопредставленные клонотипы в репертуаре пуповинной крови и наивных Т-клеток (но не клеток памяти) часто не имеют вставочных нуклеотидов, что может быть связано с их эмбриональным происхождением.

#### 4.6.1 Анализ распределения клонов предположительно эмбрионального происхождения 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 [Nikolich-Zugich, 2008]. 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 работ [Britanova и др., 2014; Britanova и др., 2016]. Из репертуара каждого донора были взяты 2000 самых крупных клонотипов, и для них была подсчитана доля клонов с нулем N-инсерций (см рис. 23).



**Рис. 23:** Характеристическое время жизни функционального TCR-клонотипа без N-вставок. Число клонотипов с нулем вставок в зависимости от возраста (черные кружки) может быть

*аппроксимирована 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 клонотипов более молодыми клетками внутри наивного репертуара.*

На рис. 23 видно, что доля клонов с нулем инсерций в топе репертуара падает с возрастом. Для оценки скорости этого падения можно использовать простую модель типа закона радиоактивного распада:

$$Z \approx c + a \exp(-bT), \quad (11)$$

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

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.

Эти 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 клонотипов TCR с возрастом.

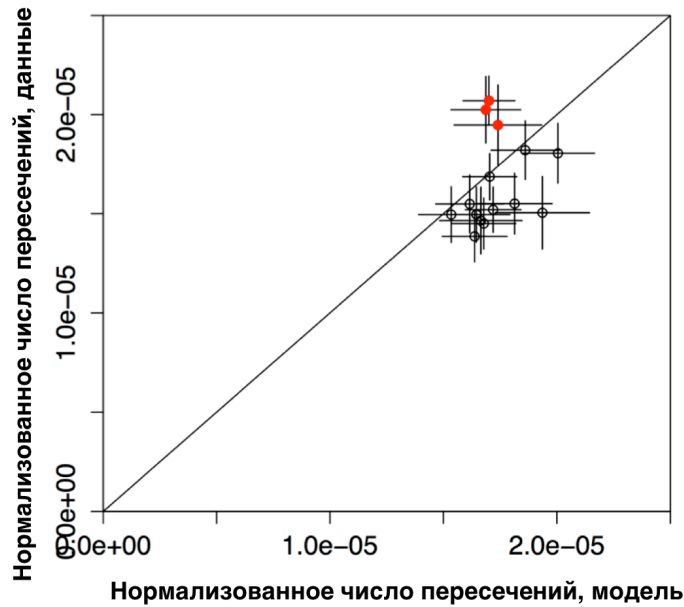
#### 4.6.2. Роль клонотипов с нулем инсерций в структуре репертуара

В предыдущем разделе было показано, что клонотипы с нулем N-инсерций неравномерно распределены в репертуаре (ими обогащена группа высокопредставленных клонотипов наивных клеток), и есть зависимость числа этих клонотипов в репертуаре от возраста.

Из этих наблюдений есть несколько следствий для совпадения клонотипов между репертуарами близнецовых и неродственных доноров.

Во-первых, можно предположить, что у близнецов должен быть избыток общих клонотипов с нулем инсерций, если их сравнивать с неродственными донорами. Для того, чтобы проверить эту гипотезу, мы сравнили нормализованное число совпадений для клонотипов с нулем инсерций в

данных с предсказанием модели сборки (см. рис. 24). Несмотря на то, что мы наблюдаем большее число совпадающих клонотипов с нулем инсерций в парах близнецов, этот эффект не так ярко выражен на фоне большого числа пересечений между неродственными донорами. Поскольку клоны с нулем инсерций имеют меньшее разнообразие, нормализованное число пересечений между неродственными донорами гораздо выше, чем представленное на рис. 12. Другими словами, конвергентная рекомбинация таких клонов гораздо более вероятна, что скрывает эффект обмена кровью в эмбриональном развитии.

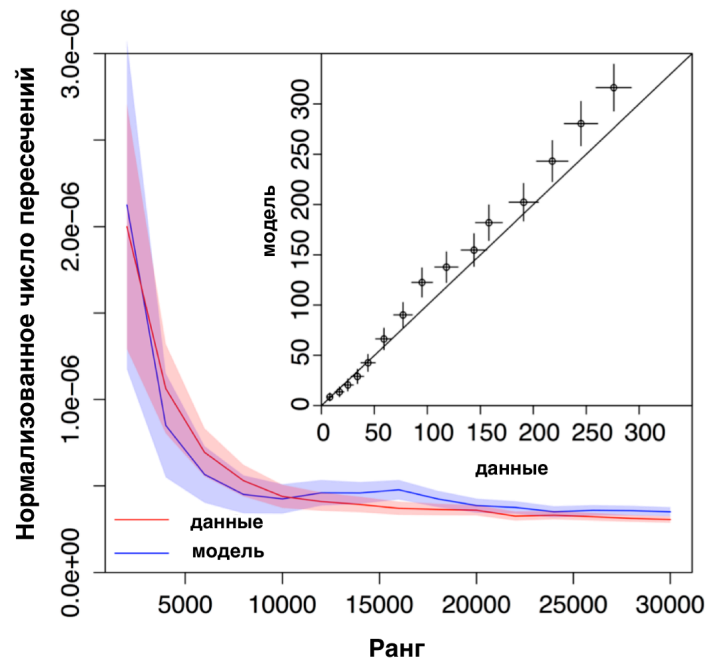


**Рис. 24.** Нормализованное число пересечений клонов с нулем инсерций. Число общих нефункциональных CDR3 альфа-цепи TCR с нулем инсерций для 15 возможных пар 6 доноров сравнивается с предсказанием модели сборки. Три красными кругами обозначены пары монозиготных близнецов. Диагональ – линия  $y=x$ , показано одно стандартное отклонение.

Во-вторых, поскольку клонотипы с нулем инсерций сосредоточены в топе репертуара, можно ожидать большего числа идентичных клонотипов в паре репертуаров при сравнении самых высокопредставленных клонотипов у неродственных доноров. Для функциональных клонотипов это было показано в работе [Zvyagin и др., 2014]. Авторы объясняют это ответом идентичных клонотипов TCR на идентичные патогены. В ходе иммунного ответа происходит клональная экспансия распознающих патоген клонов, они занимают место в топе репертуара. Если идентичные клоны ответят на идентичный распространённый патоген (к примеру, вирус гриппа) в разных донорах, то это приведет к большему количеству общих клонов в топе репертуара функциональных клонотипов TCR, что и наблюдаем.

Однако, за счет того, что топ репертуара обогащен клонами с нулем инсерций можно

ожидать большего числа общих клонов и без селекции клонов на идентичный патоген, поскольку вероятность совпасть у клонотипов без N-инсерций очень высокая. Чтобы проверить эту гипотезу, посмотрим на нормализованное число идентичных клонотипов в нефункциональных репертуарах неродственных доноров. На рис. 25 по оси ординат отложено нормализованное число пересечений в различных по размеру выборках высокопредставленных клонотипов, а на оси абсцисс показаны размеры выборок: 1000 самых высокопредставленных клонотипов, 2000 самых высокопредставленных клонотипов и так далее. Наблюдается снижение числа нормализованных клонотипов по мере продвижения вглубь репертуара.



**Рис. 25:** Число идентичных нефункциональных клонотипов альфа-цепи TCR как функция представленности этого клонотипа. Нормализованное число идентичных последовательностей в паре репертуаров неродственных доноров представлено как функция от ранга сравниваемых клонотипов (т.е. По оси ординат отложено - нормализованное число идентичных клонотипов в 2000 самых крупных клонотипов, в 4000 тысячах самых крупных и так далее). Синяя кривая - теоретическое предсказание числа идентичных клонотипов с учетом неравномерного распределение клонов с нулем инсерций в репертуаре. Врезка в правом верхнем углу показывает абсолютное число идентичных клонотипов, предсказание модели точно описывает данные.

Можно предположить, что этот эффект связан с большим числом клонов без вставочных нуклеотидов среди самых высокопредставленных клонов репертуара. Для проверки этой гипотезы нужно симулировать репертуар Т-клеточных рецепторов, учитывая неравномерное распределение

клонов с нулем вставок по репертуару. Трудность в том, что неизвестно, какой клон из репертуара имеет ноль вставок, а в каком вставочные нуклеотиды в действительности есть, но они совпали с геномными последовательностями V-сегментов. Поэтому проще использовать другую характеристику - долю клонов, последовательность которых может быть собрана без вставочных букв, только с помощью записанных в геноме последовательностей сегментов. Для репертуара двух неродственных индивидов было подсчитана доля клонов, которые могут быть собраны без вставочных букв в первой тысяче клонов репертуара, во второй тысяче клонов и так далее. Затем с помощью модели сборки были сгенерированы последовательности Т-клеточных рецепторов, которые поделили на две группы - те, которые могут быть реализованы без вставочных нуклеотидов, и те, для сборки которых эти нуклеотиды необходимы. Далее две эти группы смешивались исходя из доли клонов без вставок, которая была получена на реальных данных. Для полученных искусственных репертуаров повторили ту же процедуру, что и для данных, таким образом была получена красная кривая теоретического предсказания (см. рис. 25). Эти кривые очень схожи, т.е. большее число идентичных клонотипов среди самых высокопредставленных клонов нефункциональных репертуаров может быть объяснено повышенным содержанием клонов с нулем вставок, которые скорее всего имеют эмбриональное происхождение.

#### 4.7. 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 Т-клеточных рецепторов на мощный иммунный стимул. В следующих разделах будет описана модель для поиска клонотипов TCRbeta проходящих клональную экспансию после иммунизации, а также дана краткая характеристика последовательностей этих клонотипов.

##### 4.7.1. Поиск клонотипов TCRbeta значимо увеличивающих численность в ответ на иммунизацию против желтой лихорадки

Нами были собраны образцы периферической крови четырех пар близнецовых доноров непосредственно перед (день 0) и на 15 день после иммунизации, где по данным литературы [Akondy и др., 2009; Blom и др., 2013; Miller и др., 2008] находится пик клеточного иммунного ответа. С помощью пакета edgeR [Robinson, McCarthy, Smyth, 2009] широко используемого для анализа дифференциальной экспрессии генов, нами были получены последовательности



клонотипов Т-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 представлено в табл. 11. В дальнейшем для краткости эти клоны мы будем называть экспандированными клонотипами.

На пике ответа экспандированные клоны занимают 4-6% репертуара, см. таблицу 11, при этом почти все они отсутствовали до иммунизации, либо были представлены в крайне небольших количествах.

Таким образом все доноры сильно ответили на вакцинацию, и она вызвала существенные изменения в репертуаре.

**Таблица 11.** Число и общая доля значимо (по результатам анализа пакетом edgeR) увеличивших свою концентрацию клонов до вакцинации и на пике ответа.

	P1	P2	S1	S2
Число эксп. клонов	1056	1098	742	1042
Доля репертуара на день 0	$1.15 \cdot 10^{-5}$	$2.40 \cdot 10^{-5}$	$1.46 \cdot 10^{-5}$	$3.58 \cdot 10^{-5}$
Доля репертуара на день 15	0.045	0.043	0.047	0.055

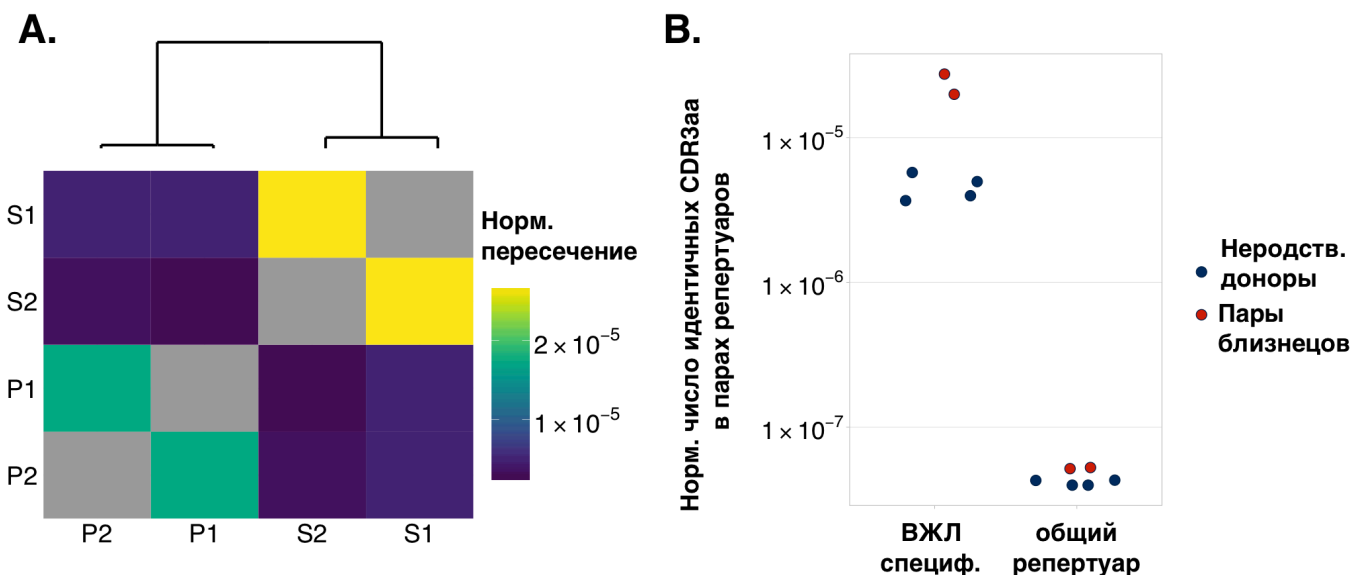
#### 4.7.2. Анализ числа идентичных у разных доноров последовательностей TCR ВЖЛ-специфических клонов

Нами было подсчитано число идентичных аминокислотных и нуклеотидных последовательностей между экспандированными в ответ на вакцинацию против желтой лихорадки клонотипами TCRbeta. Абсолютное число идентичных клонотипов приведено в таблице 12.

**Таблица 12.** Абсолютное число идентичных нуклеотидных/аминокислотных последовательностей TCR отвечающих на желтую лихорадку в различных донорах.

	S2	S1	P2	P1
S2	1042/1022	6/20	0/4	0/3
S1	6/20	742/737	0/4	0/4
P2	0/4	0/3	1098/1089	1/20
P1	0/3	0/4	1/20	1056/1049

Было показано, что у близнецовых доноров гораздо большее число идентичных клонотипов отвечает на вакцинацию (см. рис. 26). В работе [Zvyagin и др., 2014] было показано, что у близнецов гораздо больше идентичных клонотипов встречается между самыми высокопредставленными клонами в репертуаре. Здесь показан возможный механизм этого явления – у близнецов на идентичные стимулы чаще отвечают идентичные клонотипы. В ходе ответа клон Т-клеток проходит клональную экспансию и оказывается среди самых высокопредставленных клонов в репертуаре.



**Рис. 26** Нормализованное число идентичных аминокислотных последовательностей TCRbeta значимо увеличившихся в ответ на иммунизацию против желтой лихорадки. Число идентичных экспандированных клонотипов в парах доноров поделено на произведение числа ответивших клонотипов в этих донорах. Близнецы оказываются в одном кластере, если применить к полученной матрице алгоритм иерархической кластеризации.

#### 4.7.3. Оценка числа идентичных нуклеотидных последовательностей TCRbeta с помощью вероятностной модели

Интересно, что несмотря на в целом уникальный иммунный ответ, между близнецами обнаружилось больше идентичных аминокислотных и нуклеотидных последовательностей TCR.

Надо отметить, что все это функциональные последовательности, которые прошли селекцию в тимусе и узнали на периферии один из антигенов из вакцины. Понятно, что идентичные аминокислотные клонотипы с различными нуклеотидными последовательностями сформировались в двух близнецах независимо. Что касается идентичных нуклеотидных последовательностей, то здесь возможны два принципиальных варианта: либо это потомки одного

клонотипа, клетки которого участвовали в пренатальном обмене, либо это частосборные нуклеотидные варианты, которые отобрались в близнецах случайно.

Точно на этот вопрос ответить сложно, поскольку нужно как-то исключить фактор селекции на антиген. Однако можно попробовать оценить, насколько вероятно появление такого числа идентичных клонотипов близнецов при случайном выборе нуклеотидного варианта. То есть мы исходим из того, что для заданной аминокислотной последовательности, все кодирующие ее нуклеотидные варианты эквивалентны, и должны появляться в соответствии с вероятностью сборки.

Можно предложить следующую простую симуляцию:

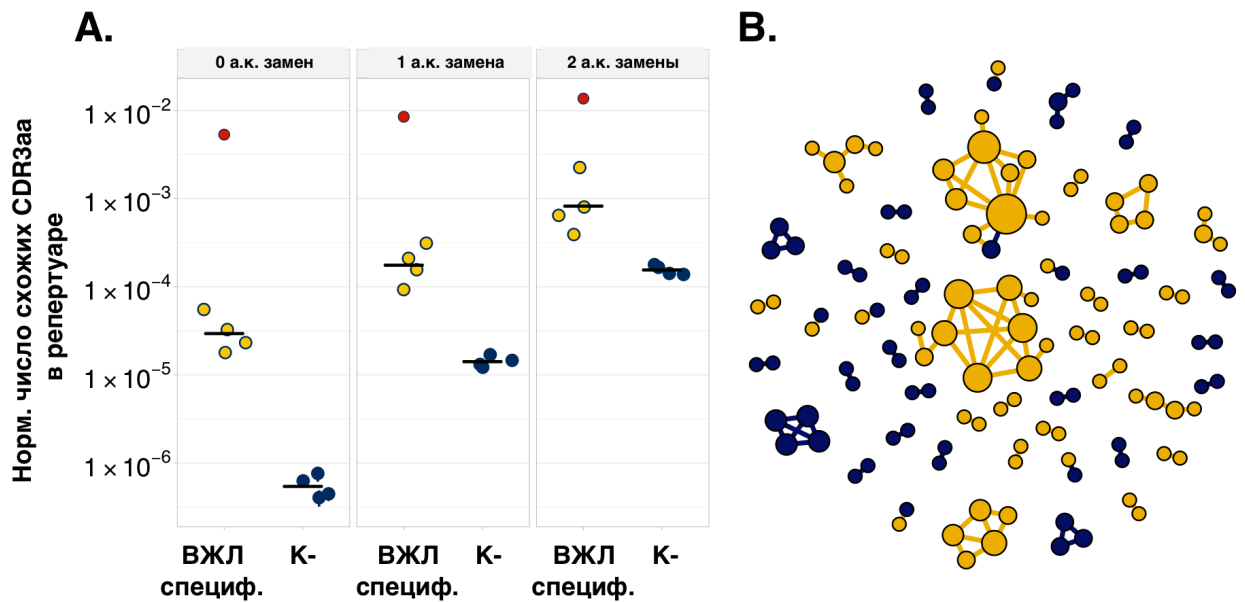
- 1) Для каждой аминокислотной последовательности каждого человека, для которой имеется идентичная последовательность в другом человеке (близнеце или нет) случайно выбирается кодирующий ее нуклеотидный вариант. При этом частота выбора конкретного нуклеотидного варианта пропорциональна его вероятности сборки.
- 2) Выбранные нуклеотидные варианты попарно сравниваются между всеми донорами, подсчитывается число идентичных пар в разных донорах
- 3) Симуляция повторяется много раз, и затем число идентичных клонотипов между донорами в этой случайной симуляции сравнивается с наблюдаемым в данных.

Нами было проведено 1000 таких симуляций. Оказалось, что в 984 из них между близнецами было меньше идентичных нуклеотидных последовательностей, чем в реальных данных.

Таким образом, конвергентная сборка не очень хорошо может объяснить совпадение нуклеотидных последовательностей специфичных к вакцине против желтой лихорадки клонотипов близнецов, возможно часть этих клонотипов соответствует клонам Т-клеток, участвовавшим в пренатальном обмене. Однако, стоит отметить, что большая часть идентичных аминокислотных последовательностей отвечающих на вакцину против желтой лихорадки клонотипов кодируется разными нуклеотидными последовательностями (см. табл. 12) – совпадение этих клонотипов нельзя объяснить ничем, кроме независимой конвергентной рекомбинации с последующей селекцией на одинаковый антиген.

#### 4.7.4. Анализ сходства аминокислотных последовательностей вакциноспецифичных клонотипов TCRbeta

В прошлом разделе было показано, что нормализованное число идентичных функциональных последовательностей TCR отвечающих на вакцинацию против желтой лихорадки выше, у близнецов, чем у неродственных доноров. При этом даже у неродственных доноров оно выше, чем в среднем по репертуару. Эти совпадения можно объяснить селекцией идентичных вариантов TCR на одинаковые антигены в разных донорах. 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. Эту статистику проще всего определить в терминах теории графов. Если каждый клонотип TCR это вершина, а ребрами соединены клонотипы с идентичными или очень похожими аминокислотными последовательностями, то число ребер в полном графе будет равно  $n(n-1)/2$ . Понятно, что в целом число ребер в подобных графах будет расти вместе с числом вершин, поэтому вместо абсолютного числа ребер  $E$  нужно сравнивать нормализованные  $E/(N(N-1)/2)$ , где  $E$ -абсолютное число ребер,  $N$ -абсолютное число вершин. В качестве контроля можно взять случайно выбранные из репертуара последовательности TCR. Результат этого анализа представлен на рис. 27. Видно, что последовательности TCR против желтой лихорадки гораздо более похожи, чем случайно выбранные последовательности TCR из репертуара. Красные точки - нормализованное число схожих клонотипов (0,1 или 2 а.к. замены) в выборке опубликованных последовательностей TCR распознающих иммуннодоминантный 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 эпитопам ВЖЛ последовательностей, которые не похожи друг на друга.



**Рис. 27** 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 похоже последовательностью в специфичных к YFV репертуарах (желтые точки) выше, чем между случайной выборкой идентичного числа клонов из той же временной точки. Для сравнения, показано нормализованное число ранее 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 мисматчей 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 YFV).

## 4.8. Практическое применение вероятностей сборки к анализу ранее опубликованных данных

В предыдущих разделах было подробно рассмотрено применение модели стохастической VDJ рекомбинации к конкретной биологической задаче: были обнаружены трудносорные зрелые гены TCR, которые предположительно принадлежат T-клеткам, которыми близнецовые доноры обменялись внутриутробно. Но эти модели можно с пользой применить в широком разнообразии ситуаций. Задача этого раздела определить перспективы и роль методов подсчета вероятности сборки в других задачах анализа репертуара TCR, используя опубликованные данные секвенирования репертуаров.

### 4.8.1. Оценка вероятности сборки TCR для оценки маркеров в задаче детекции минимальной остаточной болезни (МОБ)

В острых лимфобластных лейкозах (ОЛЛ), как и при терапии других злокачественных новообразований, важна оценка эффективности уничтожения опухолевых клеток в ходе терапии. Если часть клеток опухоли переживет химио- или радиотерапию, это может привести к рецидиву заболевания. Чтобы отличать опухолевые клетки от здоровых можно использовать молекулярные маркеры: в острых лимфобластных лейкозах такими маркерами служат последовательности перестроенных локусов TCR и BCR. Часто в таких лейкозах в организме пациента интенсивно размножается один или несколько злокачественных клонов, каждый клон имеет определенный набор маркеров – конкретных перестроек TCRbeta, TCRalpha, TCRgamma, TCRdelta, а также тяжелых и легких цепей BCR. Эти перестройки определяют до химиотерапии с помощью секвенирования [Campana, 2010; Wu и др., 2012], после терапии проводят поиск этих же перестроек в образцах костного мозга. Если такие перестройки находятся, это может свидетельствовать о наличии т.н. минимальной остаточной болезни (МОБ), которая может привести к рецидиву. В этом случае назначаются дополнительные сеансы терапии.

Проблема заключается в том, что механизм сборки перестроек злокачественных клонов такой же, как у нормальных клонов. Таким образом, теоретически возможно получить в лимфобластном клоне перестройку с высокой вероятностью сборки, которая затем будет обнаружена в репертуаре пациента после терапии, но будет принадлежать нормальному, не злокачественному клону. Совпадение перестроек происходит по тому же механизму, что и совпадение последовательностей TCR в репертуарах неродственных доноров: происходит независимая сборка одной и той же последовательности при VDJ рекомбинации. Для того, чтобы проверить реалистичность подобного сценария был проведен анализ 56 ранее опубликованных

перестроек TCRbeta из злокачественных клонов. Было обнаружено, что 9 из них имеют достаточно высокую вероятность сборки, чтобы совпасть в репертуарах двух доноров. Более того, поиск этих перестроек в 145 опубликованных репертуарах здоровых доноров показал, что 5 из них находятся в как минимум в одном (самая распространенная перестройка была обнаружена в 20 донорах) донорах с точным совпадением последовательности, а 18 в как минимум в одном (самая распространенная перестройка была найдена в 36 донорах) донорах с одной нуклеотидной заменой. Эти перестройки нужно использовать для детекции МОБ с осторожностью, так как возможно их независимое появление в здоровом репертуаре из-за высокой вероятности сборки.

Таким образом, вероятность сборки может использоваться для оценки качества маркеров для детекции МОБ. В случае, если вероятность сборки высокая, необходимо одновременно отслеживать перестройки нескольких локусов опухолевого клона, чтобы избежать ошибочного определения МОБ после терапии.

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

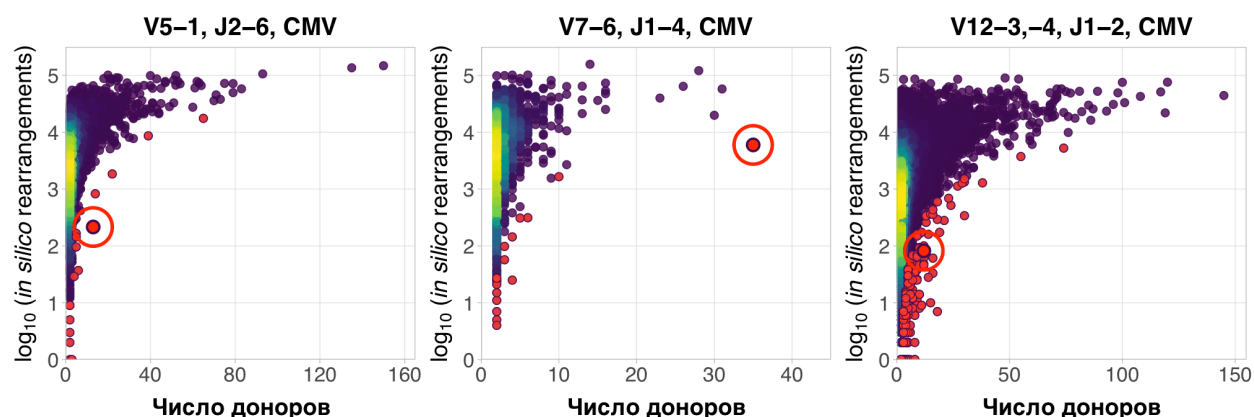
Другое применение вероятности сборки возможно для поиска клонов TCR ассоциированных с конкретными аутоиммунными или инфекционными заболеваниями. Подобные исследования стали актуальны в последнее время в связи с появлением крупных массивов репертуаров. В ходе нескольких таких исследований авторы описывали общие между многими донорами клоны T-клеточных рецепторов, которые были неравномерно распределены между когортами здоровых контролей и пациентов [Emerson и др., 2014; Faham и др., 2017]. Неравномерность распределения клона между двумя группами оценивали с помощью точного теста Фишера (доля пациентов, имеющих данный клон сравнивалась с долей доноров контрольной группы, имеющих данный клон). Этот тест выполняется для каждого клона общего между несколькими донорами, поэтому авторы вводят поправку р-значений на множественное тестирование.

Оценка вероятности сборки позволяет существенно улучшить этот подход.

Был произведен повторный анализ данных из работы [Emerson и др., 2017], где описанным выше методом сравнивались когорты доноров серопозитивных и серонегативных по цитомегаловирусной инфекции (CMV). Для анализа была использована только когорта серопозитивных доноров.

Для каждой аминокислотной последовательности TCR идентичной в нескольких донорах проводили оценку вероятности сборки этой последовательности. Поскольку существующие

алгоритмы не позволяют оценить вероятность сборки аминокислотной последовательности, оценка вероятности сборки аминокислотной последовательности TCR 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. 28.



**Рис. 28** Анализ вероятностей сборки клонотипов TCR ассоциированных с цитомегаловирусной инфекцией (CMV), по данным [Emerson и др., 2017]. Красные круги – статистически значимые отклонения, найденные алгоритмом. Большим красным кругом обведены клонотипы с доказанной в [Emerson и др., 2017] CMV специфичностью. Вероятность сборки для этих клонотипов ниже, чем для других клонотипов общих у того же числа доноров.

Это происходит из-за того, что клоны Т-клеток несущие данную последовательность TCR участвуют в иммунном ответе против цитомегаловируса (CMV), проходят клональную экспансию и их относительная доля в репертуаре повышается. Это приводит к тому, что они гораздо чаще оказываются в небольших образцах периферической крови, которые используются для секвенирования.

Таким образом, вероятность сборки можно использовать для поиска клонов ассоциированных с заболеванием, даже если когорта контрольных доноров отсутствует.



## 5. Заключение

Любая наука в своем развитии проходит путь от естествоиспытательства и простых экспериментов к полному теоретическому описанию, когда вычислить и предсказать можно почти все, что угодно. Науки о простых объектах (к примеру физика и химия) продвинулись на этом пути дальше биологии и социальных наук. Количественные методы и математические модели будут играть все большую роль в биологии и в том числе иммунологии, одной из самых молодых и сложных ее областей.

В этой работе мы совместили современные математические модели с большими данными секвенирования репертуаров TCR. Были обнаружены свидетельства пренатального обмена кровью между близнецовыми донорами, функциональные последствия и распространённость этого эффекта еще предстоит выяснить.

С другой стороны, были показаны принципиальные отличия между репертуарами наивных T-клеток и T-клеток памяти: наиболее высокопредставленные клонотипы наивного репертуара часто не имеют вставочных нуклеотидов, при этом этот эффект ослабевает с возрастом.

Было высказано предположение, что эта популяция клеток со сниженным разнообразием образуется в раннем эмбриональном развитии, когда в механизме сборки T-клеточных рецепторов не участвует TdT, фермент, нематрично добавляющий в последовательности TCR случайные нуклеотиды. Роль и функции этой популяции клеток еще предстоит прояснить.

Было так же показано несколько практических применений моделей сборки T-клеточных рецепторов к анализу ранее опубликованных данных: для оценки информативности маркеров T-клеточных лейкозов и для поиска клонов T-клеточных рецепторов ассоциированных с инфекционными и/или аутоиммунными заболеваниями.

Этот подход можно расширить и дальше – модели сборки T-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 репертуаров TCR, а в особенности новые методы реконструкции репертуаров из данных секвенирования транскриптомов, приведет к лавинообразному накоплению новых данных и важной задачей становится разработка новых статистических подходов для их анализа.

## 6. 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.

## 7. Благодарности

Эта работа была бы невозможна без помощи многих сотрудников отдела геномики адаптивного иммунитета ИБХ РАН и лаборатории сравнительной геномики в особенности. Я благодарю своего научного руководителя, Юрия Борисовича Лебедева, за спокойствие и терпение. Ильгара Мамедова, Екатерину Комеч, Анастасию Сычеву за помощь в подготовке экспериментов. Анастасию Минервину за постоянную мотивацию и обсуждение рукописи. Ольгу Британову и Дмитрия Чудакова за помощь с поиском ранее опубликованных данных.

Я также хочу поблагодарить зарубежных коллабораторов, авторов лучших на данный момент моделей сборки T- и B-клеточных рецепторов, за помощь с анализом данных, это Тьерри Мора (Thierry Mora), Александра Вальчак (Aleksandra Walczak), Квентин Маркоу (Quentin Marcou) и Ювал Эльханати (Yuval Elhanati) - сотрудники лаборатории теоретической физики Высшей Нормальной Школы (ENS) в Париже.

Отдельное спасибо хочется сказать всей моей семье за постоянную моральную и материальную поддержку.

## 8. Список сокращений

п.н. – пары нуклеотидов

н., нт. – нуклеотид

а.о. – аминокислотный остаток

НК – нуклеиновые кислоты

ПЦР – полимеразная цепная реакция

TCR (T cell receptor) – Т-клеточный рецептор

BCR (B cell receptor) – В-клеточный рецептор

RSS – recombination signal sequence

МНС (major histocompatibility complex) – главный комплекс гистосовместимости

CDR (complementarity determining region) – участок, определяющий комплементарность

ВИЧ – вирус иммунодефицита человека

СПИД – синдром приобретенного иммунодефицита

АПК - антигенпрезентирующая клетка

TREC – T-cell recombination excision circle

HMGB – High mobility group protein B

TCR-рМНС – комплекс TCR-МНС-пептид

NGS – next generation sequencing, секвенирование нового поколения

HTS – high-throughput sequencing, высокопроизводительное секвенирование

НММ – Hidden Markov Models, скрытые Марковские модели

RDI – repertoire dissimilarity index, индекс различия репертуаров

UMI – unique molecular identifier, уникальный молекулярный идентификатор

MIG – molecular identifier group, группа молекулярных идентификаторов

## 9. Список литературы

1. Agrawal A., Eastman Q., Schatz D. Implications of transposition mediated by V(D)J-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity // *Nature*. 1998. Т. 394. № 20. С. 744–751.
2. Akondy R.S. и др. The yellow fever virus vaccine induces a broad and polyfunctional human memory CD8+ T cell response. // *J. Immunol.* 2009. Т. 183. № 12. С. 7919–30.
3. Akondy R.S. и др. Origin and differentiation of human memory CD8 T cells after vaccination // *Nature*. 2017. Т. 552. № 7685. С. 362–367.
4. Bassing C., Swat W., Alt F. The mechanism and regulation of chromosomal V (D) J recombination // *Cell*. 2002. Т. 109. № D. С. 45–55.
5. Benedict C.L. и др. Terminal deoxynucleotidyl transferase and repertoire development. // *Immunol. Rev.* 2000. Т. 175. № 4. С. 150–157.
6. Bertocci B. и др. Nonoverlapping functions of DNA polymerases mu, lambda, and terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination in vivo. // *Immunity*. 2006. Т. 25. № 1. С. 31–41.
7. Biran V. и др. A long-term competent chimeric immune system in a dizygotic dichorionic twin. // *Pediatrics*. 2011. Т. 128. С. e458–e463.
8. Blom K. и др. Temporal dynamics of the primary human T cell response to yellow fever virus 17D as it matures from an effector- to a memory-type response. // *J. Immunol.* 2013. Т. 190. № 5. С. 2150–8.
9. Bocharov G., Argilaguet J., Meyerhans A. Understanding Experimental LCMV Infection of Mice: The Role of Mathematical Models // *J. Immunol. Res.* 2015. Т. 2015.
10. Bogdanos D.P. и др. Twin studies in autoimmune disease: Genetics, gender and environment // *J. Autoimmun.* 2012. Т. 38. № 2–3. С. J156–J169.
11. Bolotin D.A. и др. MiXCR: software for comprehensive adaptive immunity profiling // *Nat. Methods*. 2015. Т. 12. № 5. С. 380–381.
12. Bolotin D.A. и др. Antigen receptor repertoire profiling from RNA-seq data // *Nat. Biotechnol.* 2017. Т. 35. № 10. С. 908–911.
13. Borbulevych O.Y. и др. T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. // *Immunity*. 2009. Т. 31. № 6. С. 885–96.
14. Bovay A. и др. T cell receptor alpha variable 12-2 bias in the immunodominant response to Yellow

fever virus // *Eur. J. Immunol.* 2017.

15. Britanova O. V. и др. Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling // *J. Immunol.* 2014. Т. 192. № 6. С. 2689–2698.
16. Britanova O. V. и др. Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians // *J. Immunol.* 2016. Т. 196. № 12. С. 5005–5013.
17. Burlingham W.J., Nelson J.L. Microchimerism in cord blood: Mother as anticancer drug // *Proc. Natl. Acad. Sci.* 2012. Т. 109. № 7. С. 2190–2191.
18. Campana D. Minimal Residual Disease in Acute Lymphoblastic Leukemia // *Hematology.* 2010. Т. 2010. № 1. С. 7–12.
19. D’Orsogna L.J. и др. TCR cross-reactivity and allorecognition: New insights into the immunogenetics of allorecognition // *Immunogenetics.* 2012. Т. 64. № 2. С. 77–85.
20. Dai S., Huseby E., Rubtsova K. Crossreactive T cells spotlight the germline rules for  $\alpha\beta$  T cell receptor interactions with MHC molecules // *Immunity.* 2008. Т. 28. № 3. С. 324–334.
21. Daniels M. и др. Thymic selection threshold defined by compartmentalization of Ras/MAPK signalling // *Nature.* 2006. Т. 444. № December.
22. Dean J. и др. Annotation of pseudogenic gene segments by massively parallel sequencing of rearranged lymphocyte receptor loci // *Genome Med.* 2015. Т. 7. № 1. С. 1–8.
23. Deng L. и др. Structural insights into the editing of germ-line–encoded interactions between T-cell receptor and MHC class II by  $V\alpha$  CDR3 // *Proc. Natl. Acad. Sci. U. S. A.* 2012. Т. 109. № 37. С. 14960–14965.
24. DeWitt W.S. и др. Dynamics of the Cytotoxic T Cell Response to a Model of Acute Viral Infection // *J. Virol.* 2015. Т. 249. № February. С. JVI.03474-14.
25. Dias T., Thilaganathan B., Bhide A. Monoamniotic twin pregnancy // *Obstet. Gynaecol.* 2012. Т. 14. № 2. С. 71–78.
26. Elhanati Y. и др. Quantifying selection in immune receptor repertoires // *Proc. Natl. Acad. Sci. U. S. A.* 2014. Т. 111. № 27. С. 9875–80.
27. Elhanati Y. и др. rengenHMM : a dynamic programming tool to infer the rules of immune receptor generation from sequence data // *Bioinformatics.* 2015a. Т. 32. № 13. С. 1943–1951.
28. Elhanati Y. и др. Inferring processes underlying B-cell repertoire diversity // *Philos. Trans. R. Soc. B Biol. Sci.* 2015b. Т. 370. № 1676. С. 20140243.
29. Emerson R.O. и др. Defining the alloreactive T cell repertoire using high-throughput sequencing of mixed lymphocyte reaction culture // *PLoS One.* 2014. Т. 9. № 11. С. 1–7.
30. Emerson R.O. и др. Immunosequencing identifies signatures of cytomegalovirus exposure history and

HLA-mediated effects on the T cell repertoire // *Nat. Genet.* 2017. Т. 49. № 5. С. 659–665.

31. Emu В. и др. HLA class I-restricted T-cell responses may contribute to the control of human immunodeficiency virus infection, but such responses are not always necessary for long-term virus control. // *J. Virol.* 2008. Т. 82. № 11. С. 5398–407.

32. Faham M. и др. Discovery of T Cell Receptor  $\beta$  Motifs Specific to HLA-B27-Positive Ankylosing Spondylitis by Deep Repertoire Sequence Analysis // *Arthritis Rheumatol.* 2017. Т. 69. № 4. С. 774–784.

33. Ford A.M. и др. In utero rearrangements in the trithorax-related oncogene in infant leukaemias // *Nature.* 1993. Т. 363. № 6427. С. 358–360.

34. Garcia K.C. и др. The molecular basis of TCR germline bias for MHC is surprisingly simple. // *Nat. Immunol.* 2009. Т. 10. № 2. С. 143–7.

35. George J.F., Schroeder H.W. Developmental regulation of D beta reading frame and junctional diversity in T cell receptor-beta transcripts from human thymus. // *J. Immunol.* 1992. Т. 148. № 4. С. 1230–9.

36. Goodrich J.K. и др. Genetic determinants of the gut microbiome assessed in UK Twins // *Cell Host Microbe.* 2016. Т. 19. № 5. С. 731–743.

37. Götherström С. и др. Identification of maternal hematopoietic cells in a 2nd-trimester fetus // *Fetal Diagn. Ther.* 2005. Т. 20. № 5. С. 355–358.

38. Greenaway H.Y. и др. NKT and MAIT invariant TCR $\alpha$  sequences can be produced efficiently by VJ gene recombination. // *Immunobiology.* 2013. Т. 218. № 2. С. 213–24.

39. Hand T.W. и др. Acute Gastrointestinal Infection Induces Long-Lived Microbiota-Specific T Cell Responses // *Science (80-. ).* 2012. Т. 337. № 6101. С. 1553–1556.

40. Howie В. и др. High-throughput pairing of T cell receptor and sequences // *Sci. Transl. Med.* 2015. Т. 7. № 301. С. 301ra131-301ra131.

41. Huseby E.S. и др. Interface-disrupting amino acids establish specificity between T cell receptors and complexes of major histocompatibility complex and peptide. // *Nat. Immunol.* 2006. Т. 7. № 11. С. 1191–9.

42. Jonsson A.M. и др. Maternal microchimerism in human fetal tissues // *Am. J. Obstet. Gynecol.* 2008. Т. 198. № 3. С. 1–6.

43. Josephs T.M., Grant E.J., Gras S. Molecular challenges imposed by MHC-I restricted long epitopes on T cell immunity // *Biol. Chem.* 2017. Т. 398. № 9. С. 1027–1036.

44. Kass I., Buckle A.M., Borg N.A. Understanding the structural dynamics of TCR-pMHC interactions // *Trends Immunol.* 2014. Т. 35. № 12. С. 604–612.

45. Khan J.M., Ranganathan S. Understanding TR binding to pMHC complexes: how does a TR scan



- many pMHC complexes yet preferentially bind to one. // *PLoS One*. 2011. T. 6. № 2. C. e17194.
46. Kim M.-S. и др. Crystal structure of the V(D)J recombinase RAG1–RAG2 // *Nature*. 2015. T. 518. № 7540. C. 507–511.
47. Kongsgaard M. и др. Adaptive immune responses to booster vaccination against yellow fever virus are much reduced compared to those after primary vaccination // *Sci. Rep.* 2017. T. 7. № 1. C. 1–14.
48. Košmrlj A. и др. Effects of thymic selection of the T cell repertoire on HLA-class I associated control of HIV infection // *Nature*. 2010. T. 465. № 7296. C. 350–354.
49. Košmrlj A., Jha A. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences // *Proc. ....* 2008. T. 105. № 43.
50. Laydon D.J., Bangham C.R.M., Asquith B. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach // *Philos. Trans. R. Soc. B Biol. Sci.* 2015. T. 370. № 1675. C. 20140291.
51. Le D., Miller J.D., Ganusov V. V. Mathematical modeling provides kinetic details of the human immune response to vaccination // *Front. Cell. Infect. Microbiol.* 2015. T. 4. № January. C. 1–13.
52. Lefranc M.-P. Immunoglobulin and T Cell Receptor Genes: IMGT and the Birth and Rise of Immunoinformatics // *Front. Immunol.* 2014. T. 5.
53. Lewi L., Deprest J., Hecher K. The vascular anastomoses in monozygotic twin pregnancies and their clinical consequences // *Am. J. Obstet. Gynecol.* 2013. T. 208. № 1. C. 19–30.
54. Li S. и др. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. // *Nat. Commun.* 2013. T. 4. № May. C. 2333.
55. Livak F. и др. Genetic modulation of T cell receptor gene segment usage during somatic recombination. // *J. Exp. Med.* 2000. T. 192. № 8. C. 1191–6.
56. Lopriore E. и др. Accurate and simple evaluation of vascular anastomoses in monozygotic placenta using colored dye. // *J. Vis. Exp.* 2011. № 55. C. e3208.
57. Mamedov I.Z. и др. A new set of markers for human identification based on 32 polymorphic Alu insertions. // *Eur. J. Hum. Genet.* 2010. T. 18. № 7. C. 808–14.
58. McDonald B.D. и др. Crossreactive  $\alpha\beta$  T Cell Receptors Are the Predominant Targets of Thymocyte Negative Selection // *Immunity*. 2015. T. 43. № 5. C. 859–869.
59. Miller J.D. и др. Human effector and memory CD8<sup>+</sup> T cell responses to smallpox and yellow fever vaccines. // *Immunity*. 2008. T. 28. № 5. C. 710–22.
60. Miyazawa S., Jernigan R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. // *J. Mol. Biol.* 1996. T. 256. № 3. C. 623–44.

61. Mold J.E. и др. Maternal Alloantigens Promote the Development of Tolerogenic Fetal Regulatory T Cells in Utero // *Science* (80-. ). 2008. Т. 322. № 5907. С. 1562–1565.
62. Mold J.E. и др. Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. // *Science*. 2010. Т. 330. № 6011. С. 1695–1699.
63. Mora T., Walczak A. Quantifying lymphocyte receptor diversity // 2016. С. 1–10.
64. Murphy K., Weaver C. *Janeway's Immunobiology*, 9th edition. : CRC Press, 2016.
65. Murugan A. и др. Statistical inference of the generation probability of T-cell receptors from sequence repertoires // *Proc. Natl. Acad. Sci.* 2012. Т. 109. № 40. С. 16161–16166.
66. Nazarov V.I. и др. tcR: an R package for T cell receptor repertoire advanced data analysis. // *BMC Bioinformatics*. 2015. Т. 16. № 1. С. 175.
67. Ndifon W. и др. Chromatin conformation governs T-cell receptor J $\beta$  gene segment usage. // *Proc. Natl. Acad. Sci. U. S. A.* 2012. Т. 109. № 39. С. 15865–70.
68. Nikolich-Zugich J. Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections // *Nat. Rev. Immunol.* 2008. Т. 8. № 7. С. 512–522.
69. Nishana M., Raghavan S.C. Role of recombination activating genes in the generation of antigen receptor diversity and beyond. // *Immunology*. 2012. Т. 137. № 4. С. 271–81.
70. Olivieri D.N., Gambón-Deza F. V genes in primates from whole genome sequencing data // *Immunogenetics*. 2015. Т. 67. № 4. С. 211–228.
71. Prochazkova J., Loizou J.I. Programmed DNA breaks in lymphoid cells: Repair mechanisms and consequences in human disease // *Immunology*. 2016. Т. 147. № 1. С. 11–20.
72. Pulendran B. и др. Immunity to viruses: learning from successful human vaccines. // *Immunol. Rev.* 2013. Т. 255. № 1. С. 243–55.
73. Putintseva E. V. и др. Mother and Child T Cell Receptor Repertoires: Deep Profiling Study // *Front. Immunol.* 2013. Т. 4. № DEC.
74. Qi Q. и др. Diversification of the antigen-specific T cell receptor repertoire after varicella zoster vaccination // *Sci. Transl. Med.* 2016. Т. 8. № 332. С. 332ra46-332ra46.
75. Quigley M.F. и др. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. // *Proc. Natl. Acad. Sci. U. S. A.* 2010. Т. 107. № 45. С. 19414–9.
76. Rechavi E. и др. Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. // *Sci. Transl. Med.* 2015. Т. 7. № 276. С. 276ra25.
77. Ridaura V.K. и др. Cultured gut microbiota from twins discordant for obesity modulate adiposity and metabolic phenotypes in mice // *Science* (80-. ). 2014. Т. 341. № 6150. С. 1–22.
78. Robinson J. и др. The IMGT/HLA database. // *Nucleic Acids Res.* 2013. Т. 41. № Database issue. С.

D1222-7.

79. Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data // *Bioinformatics*. 2009. T. 26. № 1. С. 139–140.
80. Rossjohn J. и др. T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules // *Annu. Rev. Immunol.* 2015. T. 33. № 1. С. 169–200.
81. Ru H. и др. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures // *Cell*. 2015. T. 163. № 5. С. 1138–1152.
82. Rubelt F. и др. Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells. // *Nat. Commun.* 2016. T. 7. С. 11112.
83. Rudolph M., Stanfield R., Wilson I. How TCRs bind MHCs, peptides, and coreceptors // *Annu. Rev. Immunol.* 2006.
84. Schenkel-Brunner H. HLA (Human Leucocyte-Associated) Class I Antigens on Red Cells // *Human Blood Groups*. Oxford, UK: Wiley-Blackwell, 2013. С. 512–514.
85. Shugay M. и др. Towards error-free profiling of immune repertoires. // *Nat. Methods*. 2014. T. 11. № 6. С. 653–5.
86. Sinha R. и др. Index Switching Causes “Spreading-Of-Signal” Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing // *bioRxiv*. 2017. С. 125724.
87. Tavian M., Péault B. The changing cellular environments of hematopoiesis in human development in utero // *Exp. Hematol.* 2005. T. 33. № 9 SPEC. ISS. С. 1062–1069.
88. Vantourout P., Hayday A. Six-of-the-best: unique contributions of  $\gamma\delta$  T cells to immunology // *Nat. Rev. Immunol.* 2013. T. 13. № 2. С. 88–100.
89. Wang C. и др. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. // *Proc. Natl. Acad. Sci. U. S. A.* 2010. T. 107. № 4. С. 1518–23.
90. Wang J., Reinherz E.L. The structural basis of  $\alpha\beta$  T-lineage immune recognition: TCR docking topologies, mechanotransduction, and co-receptor function // *Immunol. Rev.* 2012. T. 250. № 1. С. 102–119.
91. Wang P. и др. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. // *PLoS Comput. Biol.* 2008. T. 4. № 4. С. e1000048.
92. Warmflash a., Dinner a. R. A Model for TCR Gene Segment Use // *J. Immunol.* 2006. T. 177. № 6. С. 3857–3864.
93. Weill J.-C., Reynaud C.-A. DNA polymerases in adaptive immunity. // *Nat. Rev. Immunol.* 2008. T. 8. № 4. С. 302–12.

94. Wu D. и др. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia // *Sci. Transl. Med.* 2012. Т. 4. № 134.
95. Yewdell J., Anton L., Bennink J. Defective Ribosomal Products (DRiPs) A Major Source of Antigenic Peptides for MHC Class I molecules? // *J. Immunol.* 1996. Т. 7. С. 1823–1826.
96. Zvyagin I. V и др. Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing // *Proc. Natl. Acad. Sci.* 2014. Т. 111. № 16. С. 5980–5985.