

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

На правах рукописи

Егоров Евгений Станиславович  
**Возрастные изменения в структуре репертуаров Т-клеточных рецепторов  
наивных Т-лимфоцитов**

Специальность – 03.01.03 – молекулярная биология

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
кандидата биологических наук

Москва 2019

Работа выполнена в Группе структурной организации Т-клеточного иммунитета Отдела геномики адаптивного иммунитета в Федеральном государственном бюджетном учреждении науки Институте биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук (ИБХ РАН).

**Научный руководитель:**

Британова Ольга Владимировна, кандидат биологических наук

**Научный консультант:**

Чудаков Дмитрий Михайлович, доктор биологических наук

**Официальные оппоненты:**

**Казанский Дмитрий Борисович**, д.б.н., профессор, заведующий Лабораторией механизмов регуляции иммунитета НИИ канцерогенеза Федерального государственного бюджетного учреждения «Национального медицинского исследовательского центра онкологии имени Н.Н. Блохина» Министерства здравоохранения Российской Федерации.

**Касьянов Артем Сергеевич**, к.ф.-м.н., старший научный сотрудник в лаборатории №19 Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН).

**Ведущая организация:**

Федеральное государственное бюджетное учреждение “Государственный научный центр “Институт иммунологии” Федерального медико-биологического агентства России.

Защита диссертации состоится в 26 февраля 2020 года в 10:00 часов на заседании Диссертационного совета Д 002.019.01 при Федеральном государственном бюджетном учреждении науки Институте биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук по адресу: 117997, г. Москва, ул. Миклухо-Маклая, д.16/10.

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

Автореферат разослан \_\_\_\_\_ 20\_\_ г.

**Учёный секретарь диссертационного совета,**  
**доктор физико-математических наук**

**В.А. Олейников**

## Общая характеристика работы

### Актуальность проблемы

Отличительной особенностью нашего иммунитета является способность эффективно и с высокой специфичностью распознавать инфекционные и опухолевые антигены и вырабатывать на них сбалансированный ответ. Однако с возрастом эффективность иммунитета значительно снижается, что зачастую может приводить к развитию серьезных инфекционных, аутоиммунных и онкологических заболеваний. Ключом к пониманию механизмов старения иммунитета может стать популяция наивных Т-клеток, характеризующаяся крайне разнообразным репертуаром Т-клеточных рецепторов (англ. *T-cell receptor, TCR*). Каждая Т-клетка несет на своей поверхности определенный вариант Т-клеточного рецептора, потенциально способный распознать молекулу антигена в составе главного комплекса гистосовместимости (англ. *major histocompatibility complex, MHC*) на поверхности антигенпрезентирующих клеток (англ. *antigen-presenting cell, APC*) и определяющий специфичность иммунного ответа. Таким образом, именно наивный репертуар Т-клеток в значительной степени определяет диапазон распознаваемых антигенов и эффективность иммунитета в целом.

В настоящее время технологии высокопроизводительного секвенирования (англ. *high-throughput sequencing, HTS*) нашли широкое применение в современных исследованиях адаптивного иммунитета и активно используются для глубокого анализа репертуаров Т-клеточных рецепторов. Сегодня эти технологии востребованы для изучения закономерностей функциональной организации адаптивного иммунитета в норме и патологии, оценки последствий того или иного терапевтического вмешательства, затрагивающего иммунную систему (иммуносупрессия, трансплантация гемопоэтических клеток), изучения механизмов и повышения эффективности вакцинации, изучения закономерностей старения адаптивного иммунитета, идентификации терапевтически применимых вариантов TCR, и других практических приложений.

Настоящая работа посвящена исследованию изменений в тотальных и сортированных наивных популяциях Т-клеток доноров разного возраста. Данная работа включает три экспериментальные части.

Первая часть посвящена разработке метода с применением технологии молекулярного баркодирования или уникальных молекулярных идентификаторов (англ. *unique molecular*

*identifier, UMI*) для качественного нормированного анализа репертуаров Т-клеточных рецепторов.

Во второй части рассматриваются возрастные изменения в тотальных Т-клеточных репертуарах с применением разработанных методологических подходов.

Третья часть посвящена анализу изменений в TCR репертуарах наивных CD4 и CD8 Т-клеток и внутри двух субпопуляций наивных CD4 Т-клеток: CD31<sup>+</sup> Т-клеток, условно определяемых как “недавние эмигранты из тимуса” (англ. *recent thymic emigrant, RTE*), и зрелых наивных (CD31<sup>-</sup>) CD4 Т-лимфоцитов.

### **Цель работы**

Цель настоящей работы заключалась в исследовании возрастных изменений в репертуарах  $\beta$  цепей Т-клеточных рецепторов наивных Т-клеток человека. Для достижения цели были поставлены следующие задачи:

1. Разработать метод, позволяющий проводить качественный анализ репертуаров Т-клеточных рецепторов, в том числе для небольшого стартового числа (несколько сотен) Т-клеток.
2. Подготовить библиотеки и с помощью высокопроизводительного секвенирования получить данные по тотальным репертуарам  $\beta$  цепей Т-клеточных рецепторов периферической крови здоровых доноров нескольких возрастных групп.
3. Провести сравнительный анализ структуры репертуаров  $\beta$  цепей Т-клеточных рецепторов доноров разного возраста.
4. Провести сортировку популяций CD4 и CD8 наивных Т-клеток для взрослых здоровых доноров разного возраста.
5. Подготовить кДНК библиотеки генов Т-клеточных рецепторов и с помощью высокопроизводительного секвенирования получить информацию о разнообразии репертуаров наивных Т-клеток для доноров нескольких возрастных групп.
6. Проанализировать возрастные изменения в структуре репертуаров Т-клеточных рецепторов нескольких популяций наивных Т-клеток для доноров разного возраста.

### **Научная новизна и практическая ценность работы**

На сегодняшний день технологии высокопроизводительного секвенирования нашли широкое применение в исследованиях адаптивного иммунитета в норме и при различных патологиях и дали возможность провести качественный и количественный анализ разнообразия репертуаров Т-клеточных рецепторов и антител на недоступном до недавнего времени уровне. На основе технологий 5'-RACE и молекулярного баркодирования (UMI) мы

разработали простой и надежный метод анализа репертуаров Т-клеточных рецепторов, позволяющий корректно и нормировано сравнивать их характеристики. Также наш метод оказался крайне эффективным при работе с небольшими стартовыми количествами клеток. Использование в анализе UMI продемонстрировало высокую эффективность метода – позволило качественно нормировать данные секвенирования и точно устранять ошибки, возникшие в ходе амплификации и секвенирования, сохранив при этом естественное разнообразие анализируемой библиотеки репертуара TCR.

В данной работе мы проанализировали возрастные изменения репертуаров как на уровне общего пула Т-клеток, так и на уровне функциональных субпопуляций. Мы показали значительные изменения в наивных репертуарах CD4 Т-клеток (RTE и зрелых наивных CD4) с возрастом, которые могут являться ключом к пониманию старения иммунитета.

Мы уверены, что разработанный нами метод будет полезен в дальнейших исследованиях репертуаров иммунных клеток, а полученные нами результаты представляют интерес с фундаментальной и практической точек зрения. Мы рассчитываем, в частности, что они будут полезны для развития методов поддержания адаптивного иммунитета в пожилом возрасте и при создании новых эффективных стратегий вакцинации.

### **Структура диссертации**

Диссертационная работа изложена на 115 страницах и состоит из введения, обзора литературы, описания цели и поставленных задач, материалов и методов исследования, результатов и их обсуждения, выводов, заключения, списка сокращений и списка литературы, включающего 183 источника. Диссертация содержит 21 рисунок и 12 таблиц.

### **Апробация работы**

Основные результаты диссертации были изложены во время научного семинара Skoltech Systems Biology Seminar, 2019, Москва.

### **Публикации**

По результатам работы опубликовано 7 статей в рецензируемых зарубежных и отечественных журналах.

## Содержание работы

### 1. Разработка качественного метода анализа иммунных репертуаров

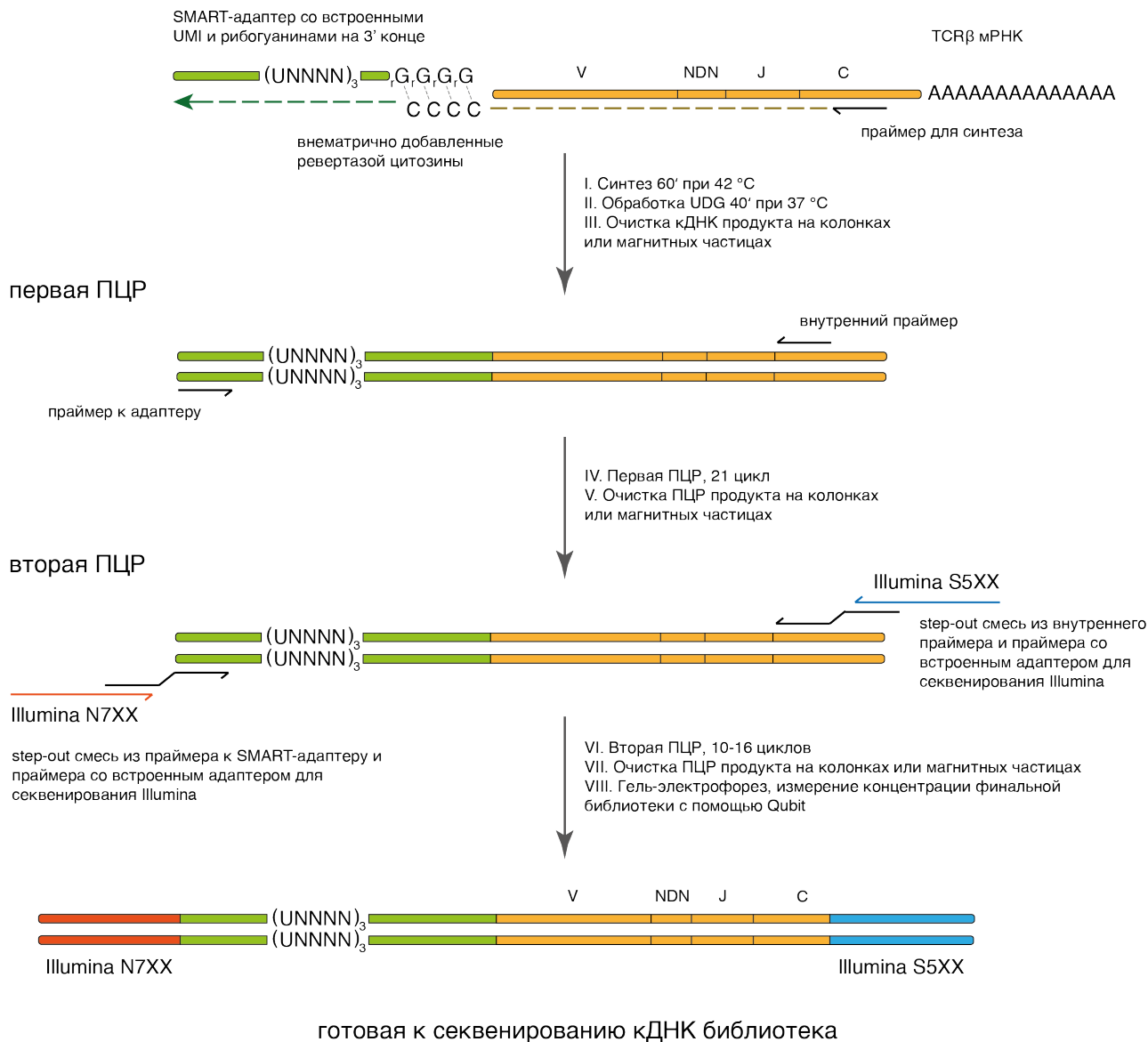
Для приготовления библиотек иммунных рецепторов к настоящему моменту были разработаны разнообразные технические подходы, использующие в качестве стартовой материи геномную ДНК или РНК и основанные на мультиплексной ПЦР амплификации переменных генных сегментов или методе RACE (англ. *rapid amplification of cDNA ends*), соответственно. При наличии достаточного количества исходного материала данные методы хорошо себя зарекомендовали. Однако для случаев, когда исследователям приходится иметь дело с небольшим количеством целевых клеток (это может быть редкая популяция отсортированных лимфоцитов или образец ткани, содержащий ограниченное число Т-клеток), подобные подходы имеют ряд серьезных ограничений. На качество анализа начинают отрицательно сказываться искажения, вызванные неравномерной амплификацией генных сегментов, а также ошибки, возникающие в ходе ПЦР-амплификации и секвенирования.

В настоящей работе мы стремились разработать простой и надежный подход, позволяющий анализировать репертуары иммунных клеток в том числе для экспериментов, когда в качестве исходного материала исследователю доступно от нескольких сотен до тысяч клеток и классические методы приготовления библиотек малоприспособлены, поскольку связаны со значительными потерями исходного материала (РНК или ДНК) в процессе приготовления библиотек для высокопроизводительного секвенирования.

#### 1.1. Анализ репертуаров иммунных клеток на основе UMI

Разработанный нами метод пробоподготовки кДНК библиотек (Рисунок 1) основывается на технологии 5'-RACE со сменой матрицы (англ. *template switch*), при котором на 3' и 5' концах библиотеки вводятся универсальные сайты отжига праймеров. При таком подходе дальнейшая амплификация вместо сложной мультиплексной смеси праймеров на 5' конце библиотеки проводится с использованием всего двух олигонуклеотидов – комплементарных последовательности введенного адаптера и консервативному участку TCR, соответственно. Это позволяет минимизировать количественные искажения, связанные с разной эффективностью амплификации переменных генных сегментов TCR. Также в реакции обратной транскрипции каждая новая синтезированная молекула кДНК маркируется специальной последовательностью – уникальным молекулярным идентификатором (англ. *unique molecular identifier, UMI*) или баркодом, состоящим из 12 встроенных в SMART-адаптер случайных нуклеотидов.

## синтез первой цепи кДНК



**Рисунок 1.** Схема пробоподготовки кДНК библиотеки для секвенирования на платформе Illumina. (UNNNN)<sub>3</sub> – молекулярный баркод (UMI) внутри SMART-адаптера из 12 случайных нуклеотидов, разделенных U в три кассеты из 4 нуклеотидов. Фрагменты Illumina S5XX и Illumina N7XX – последовательности адаптеров для секвенирования (Illumina), встроенные в праймер для второй ПЦР.

Такой подход по сравнению с обычными методами приготовления библиотек генов иммунных рецепторов обладает рядом преимуществ.

Во-первых, молекулярное баркодирование UMI позволяет оценить количество молекул, успешно прошедших все стадии синтеза кДНК и амплификации, что крайне важно для качественной нормировки и сравнения данных при анализе результатов секвенирования.

Во-вторых, анализ на основе UMI устраняет количественные искажения, возникающие из-за неравномерной амплификации разных генных сегментов TCR, а также из-за

неоднородной эффективности секвенирования. Для образцов, содержащих сотни и тысячи клеток и разным распределением клонотипов по частоте, подобные искажения способны привести к серьезному изменению естественного разнообразия.

В-третьих, использование UMI позволяет эффективно исправлять ошибки ПЦР и секвенирования с применением алгоритмов Safe-Sequencing или Molecular Identifier-Based-Error Error Correction (MIGEC). Оба алгоритма группируют вместе последовательности с одинаковыми UMI – представляющие копии одной исходной молекулы ДНК или кДНК. В процессе анализа информация об этих последовательностях усредняется, ошибочно возникшие в ходе амплификации варианты клонотипов отфильтровываются, таким образом устраняется искусственно завышенное разнообразие исходных библиотек.

## **1.2. Разработка метода пробоподготовки кДНК библиотек для небольшого числа Т-клеток**

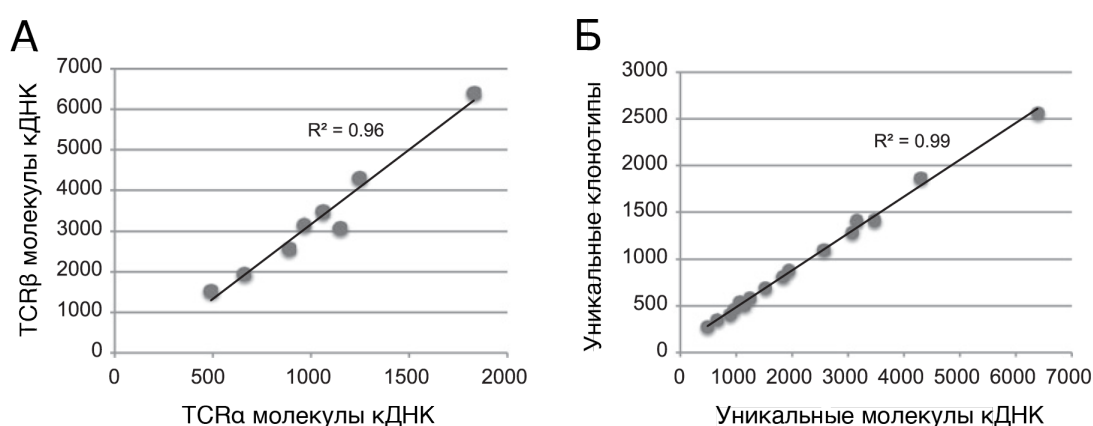
Нами был значительно улучшен и оптимизирован имеющийся протокол пробоподготовки кДНК библиотек генов Т-клеточных рецепторов человека, который подтвердил свою эффективность в том числе и при работе с небольшими количествами стартового материала. При работе с малым числом клеток, чтобы избежать потерь материи в ходе пробоподготовки, мы проводили синтез кДНК непосредственно с клеток, которые лизировали при определенных условиях. Для разрушения клеточной мембраны мы использовали несколько разных техник: кратковременно замораживали клеточную суспензию на  $-70\text{ }^{\circ}\text{C}$  или проводили серию кратковременных нагревов до  $50\text{ }^{\circ}\text{C}$ , а также инкубировали клетки вместе с детергентом Tween 20. Все эти методики позволили воспроизводимо получать кДНК библиотеки Т-клеточных рецепторов после 30-36 циклов ПЦР амплификации, имея на старте несколько сотен сортированных Т-лимфоцитов или несколько тысяч моноклеарных клеток периферической крови (англ. *peripheral blood mononuclear cells, PBMC*).

Мы подготовили и отсековировали кДНК библиотеки  $\alpha$  и  $\beta$  цепей TCR для 8 реплик по 4 000 PBMC (приблизительно соответствующие 2 000 Т-клеток) одного взрослого здорового донора. При первичном биоинформатическом анализе без учета UMI было идентифицировано от 61 860 до 702 195 CDR3-содержащих последовательностей Т-клеточных рецепторов.

При анализе с помощью MIGEC последовательности секвенирования с одинаковыми молекулярными идентификаторами (представляющие одну и ту же исходную молекулу кДНК) были сгруппированы. В анализе учитывались только UMI, независимо отсековированные не менее 4 раз. Варианты CDR3, содержащие стоп-кодон или находящиеся вне рамки считывания, а также представленные низким покрытием секвенирования, были



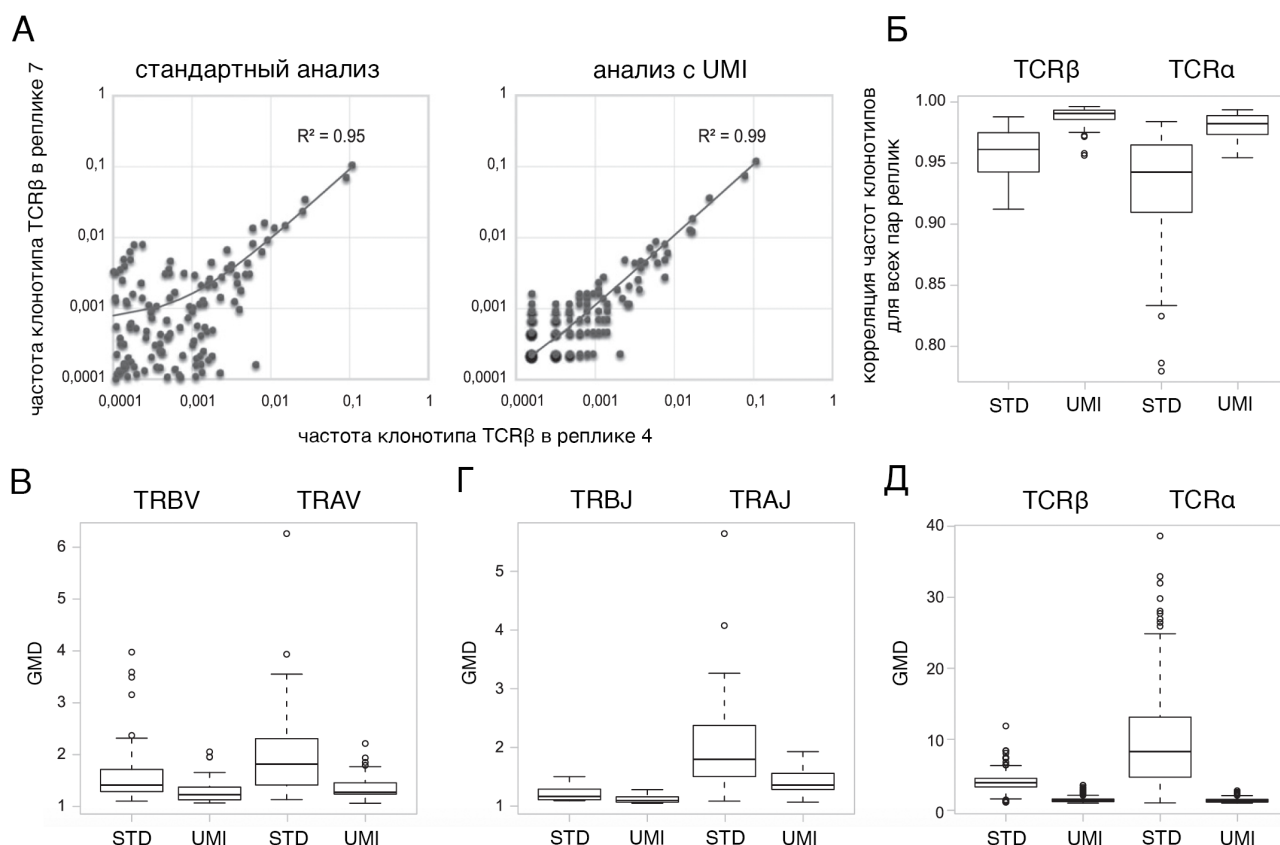
удалены из анализа. В результате было выявлено от 1 520 до 6 395 молекул кДНК на реплику для библиотек TCR $\beta$  ( $3\,302 \pm 1\,522$ ) и от 489 до 1 830 молекул на реплику для библиотек TCR $\alpha$  ( $1\,035 \pm 406$ ). Таким образом, на каждую из 2 000 Т-клеток в среднем на реплику было проанализировано по 1,7 TCR $\beta$  и 0,5 TCR $\alpha$  молекул мРНК, вошедших в реакцию синтеза кДНК. Количество полученных молекул кДНК TCR $\alpha$  и TCR $\beta$  точно коррелировало между репликами ( $R^2 = 0,96$ ), что указывает на неравное число Т-клеток, попавших в каждую реакцию, при высокой воспроизводимости метода (Рисунок 2.А). Разнообразие последовательностей CDR3, то есть число уникальных клонотипов Т-клеток, составило от 690 до 2 558 вариантов для TCR $\beta$  и от 276 до 812 вариантов для TCR $\alpha$  библиотек на реплику и хорошо коррелировало с количеством анализируемых молекул кДНК (Рисунок 2.Б).



**Рисунок 2.** (А) Соотношение полученных TCR $\alpha$  и TCR $\beta$  молекул между репликами и (Б) корреляция между количеством кДНК молекул и числом уникальных клонотипов для каждой реплики (совместно для TCR $\alpha$  и TCR $\beta$  цепей).

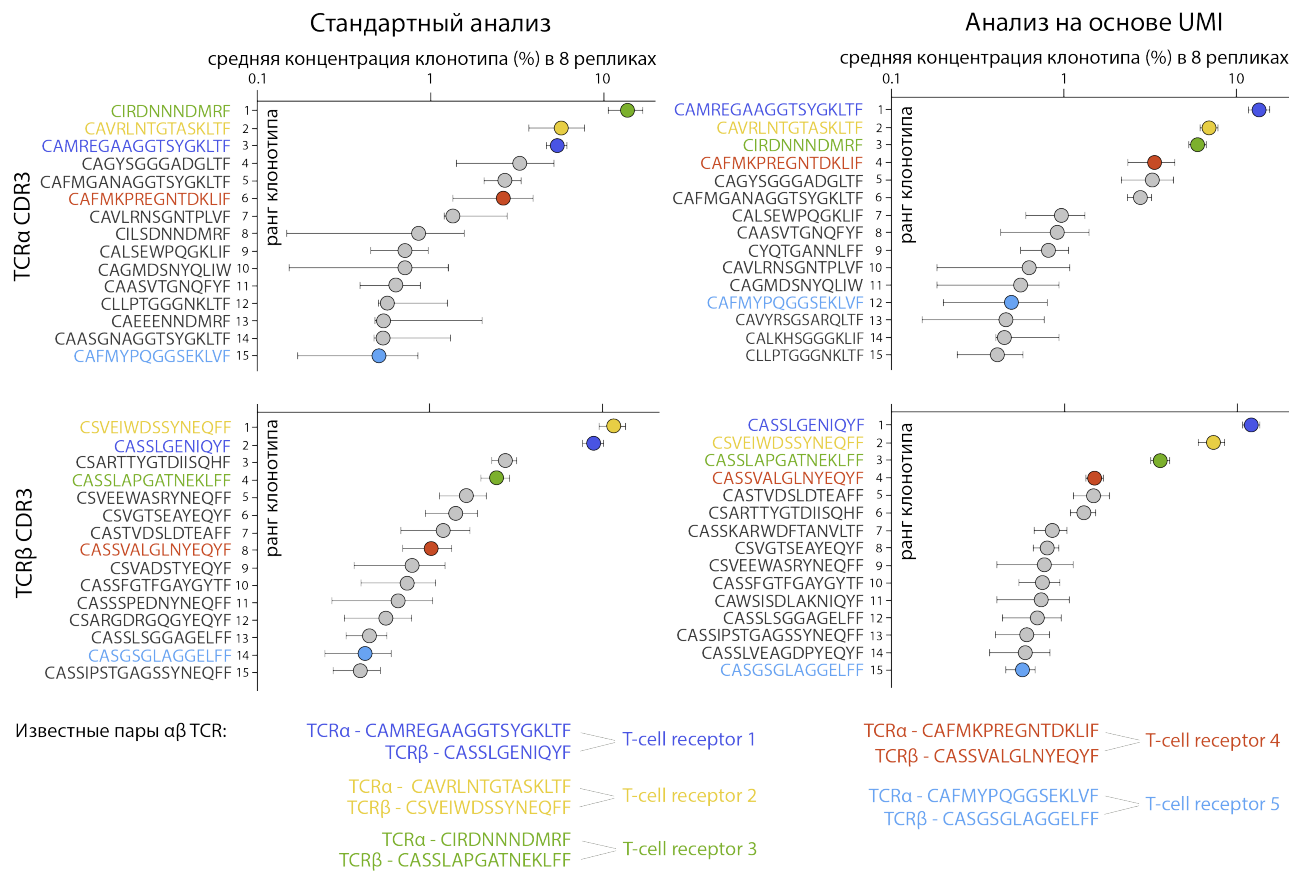
Сравнение двух подходов анализа данных показало, что учет UMI существенно улучшает количественную оценку клонотипов и генных сегментов. Частоты идентифицированных клонотипов лучше коррелировали при сравнении парных реплик в методе молекулярного баркодирования по сравнению со стандартным способом анализа данных (Рисунок 3.А-Б). Использование UMI также позволило получить более точные частоты V- и J-генных сегментов TCR с минимальным разбросом между репликами (Рисунок 3.В и Г).

Также при стандартном анализе данных относительная концентрация основных клонотипов была идентифицирована с заметной дисперсией между репликами, тогда как подсчет молекул кДНК с помощью UMI показал значительно более воспроизводимые данные о концентрации каждого клонотипа (Рисунок 3.А и Д).



**Рисунок 3.** Сравнение эффективности стандартного метода анализа иммунных репертуаров и метода на основе UMI. **(А)** Корреляция частот CDR3 TCR $\beta$  клонотипов между двумя репликами при стандартном анализе и анализе с применением UMI. **(Б)** Корреляция частот CDR3 TCR $\alpha$  и TCR $\beta$  клонотипов между парами всех 8 реплик. Дисперсия частот V- **(В)** и J-генных сегментов **(Г)** для CDR3 TCR $\alpha$  и TCR $\beta$  клонотипов для 8 биологических реплик. Частота для каждого генного сегмента была рассчитана как сумма частот клонотипов, содержащих этот генный сегмент в последовательности CDR3. **(Д)** Дисперсия частот для клонотипов, идентифицированных во всех 8 репликах. UMI – метод обработки данных с использованием молекулярных баркодов, STD – стандартный метод (без учета UMI), GMD – геометрическое стандартное отклонение.

Для нескольких крупных и стабильных клонов Т-клеток этого донора ранее нами были идентифицированы пары  $\alpha$  и  $\beta$  цепей функциональных TCR. Используя эту информацию, мы проверили, можно ли на основании ранга клонотипов внутри репертуара идентифицировать нативные пары  $\alpha$  и  $\beta$  цепей Т-клеточных рецепторов. Было показано, что в сложной смеси клеток периферической крови при анализе на основе UMI точно могут быть идентифицированы пары  $\alpha$  и  $\beta$  цепей Т-клеточных рецепторов наиболее крупных клонов (Рисунок 4, справа). И, напротив, при стандартном анализе, то есть без использования молекулярных идентификаторов, для известных клонотипов были получены плохо согласующиеся данные (Рисунок 4, слева).



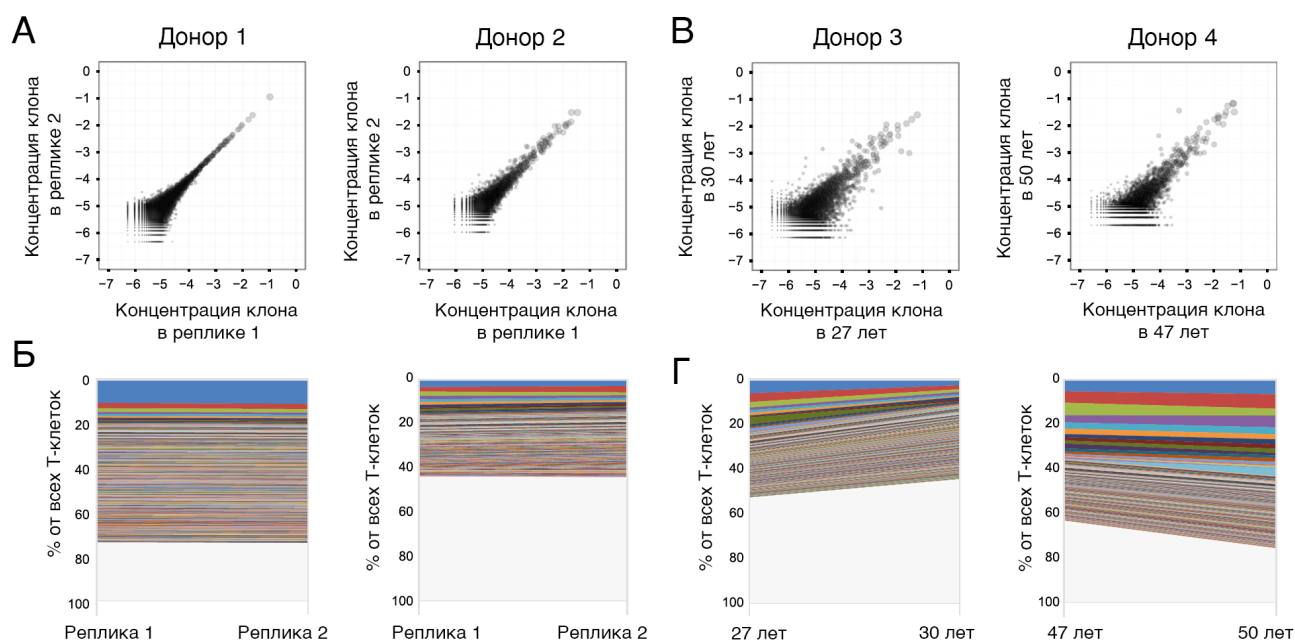
**Рисунок 4.** Идентификация пар αβ TCR с помощью ранжирования наиболее крупных клонотипов при стандартном и UMI анализах. Показаны 15 наиболее крупных клонов для α и β цепей Т-клеточных рецепторов после стандартной обработки данных (слева) и анализа на основе UMI (справа). Индивидуальным цветом отображены известные αβ пары функциональных TCR.

### 1.3. Эффективность метода при глубоком секвенировании иммунных репертуаров

Мы продемонстрировали высокую эффективность разработанного метода при глубоком секвенировании иммунных репертуаров на примере парных реплик тотальных TCR репертуаров для двух взрослых здоровых доноров.

Из имеющихся на старте от 3 до 10 миллионов РВМС после всех процедур пробоподготовки и секвенирования для каждого образца после анализа на основе UMI было получено от 1 000 000 до 2 600 000 уникальных молекул кДНК TCR $\beta$ , которые включали от 400 000 до 600 000 различных клонотипов. Сравнение частот всех пересекающихся клонотипов, обнаруженных в двух репликах, продемонстрировало высокую степень корреляции двух образцов –  $R^2 = 1$  для первого донора и  $R^2 = 0,92$  для второго (Рисунок 5.А).

Гомеостатическое пространство, занимаемое общими для обеих реплик клонотипами, и размер наиболее представленных клонотипов были практически идентичны в обеих репликах для каждого донора, что также подтвердило высокую точность метода (Рисунок 5.Б).



**Рисунок 5.** Стабильность Т-клеточного репертуара на примере парных реплик нескольких доноров. Точечные диаграммы частот клонотипов для двух технических реплик (А) или временных точек (В). Гистограмма изменения частот наиболее представленных клонотипов для двух доноров в репликах (Б) или с временным промежутком в три года (Г).

## 2. Возрастные изменения в тотальных Т-клеточных репертуарах

### 2.1. Стабильность Т-клеточного репертуара на протяжении трёх лет

Достаточно давно было установлено, что антиген-специфичные клоны Т-клеток памяти способны накапливаться в организме и поддерживать свою численность на протяжении длительного времени. Однако не было показано, насколько сильно меняется репертуар Т-клеточных рецепторов в целом в течение определенного времени. Для того чтобы непосредственно оценить стабильность пула Т-клеток, мы провели глубокий анализ TCR $\beta$  репертуаров для двух взрослых здоровых доноров с разницей в три года. Для первого донора мы обнаружили 78 009 TCR $\beta$  общих клонотипов, встречающихся в репертуарах этого донора в возрасте 27 и 30 лет. Для второго донора мы нашли 36 854 TCR $\beta$  общих для обеих временных точек (47 и 50 лет) клонотипов.

По прошествии трех лет основная часть клонотипов сохранила свои относительные частоты ( $R^2 = 0,861$  и  $0,905$  для доноров 3 и 4 соответственно; Рисунок 5.В). Хотя среди 1 000 наиболее крупных клонов наблюдалось изменение концентрации в 2 и 1,8 раз. Из 1 000 наиболее крупных клонов, наблюдаемых в первый момент времени, 952 и 882 были повторно идентифицированы через три года в репертуарах для доноров 3 и 4, соответственно. Аналогично из 1 000 наиболее крупных клонов, наблюдаемых во второй временной точке, 962 и 966 клонов также были идентифицированы в первой временной точке для индивидуумов 3 и 4, соответственно.

Таким образом, мы наблюдали минимальные различия в тотальном Т-клеточном репертуаре за период в три года, сопоставимые с разницей между биологическими репликами. Обнаруженные в обеих временных точках клоны составляли в совокупности от 45 до 75% от общего количества Т-клеток в периферической крови (Рисунок 5.Г). И процент таких долгоживущих клонов хорошо коррелировал с долей Т-клеток памяти, измеренных с помощью проточной цитометрии. Данное наблюдение подтверждает информацию о том, что значительная часть клональных экспансий Т-лимфоцитов сохраняется в организме на протяжении многих лет.

## 2.2. Динамика Т-клеточного репертуара на протяжении жизни

Для того чтобы охарактеризовать длительные изменения, затрагивающие Т-клеточный репертуар в течение жизни, мы проанализировали тотальные TCR $\beta$  репертуары для 8 образцов пуповинной крови (англ. *umbilical cord blood, UCB*) и 65 образцов периферической крови здоровых доноров разного возраста, которые были разделены на 4 возрастные группы:

1. Молодые доноры (в среднем – 16 лет).
2. Взрослые доноры (в среднем – 40 лет).
3. Пожилые доноры (в среднем – 60 лет).
4. Долгожители (в среднем – 92 года).

Проведенный сравнительный анализ выявил определенные признаки, характерные для Т-клеточного иммунитета в разные периоды жизни.

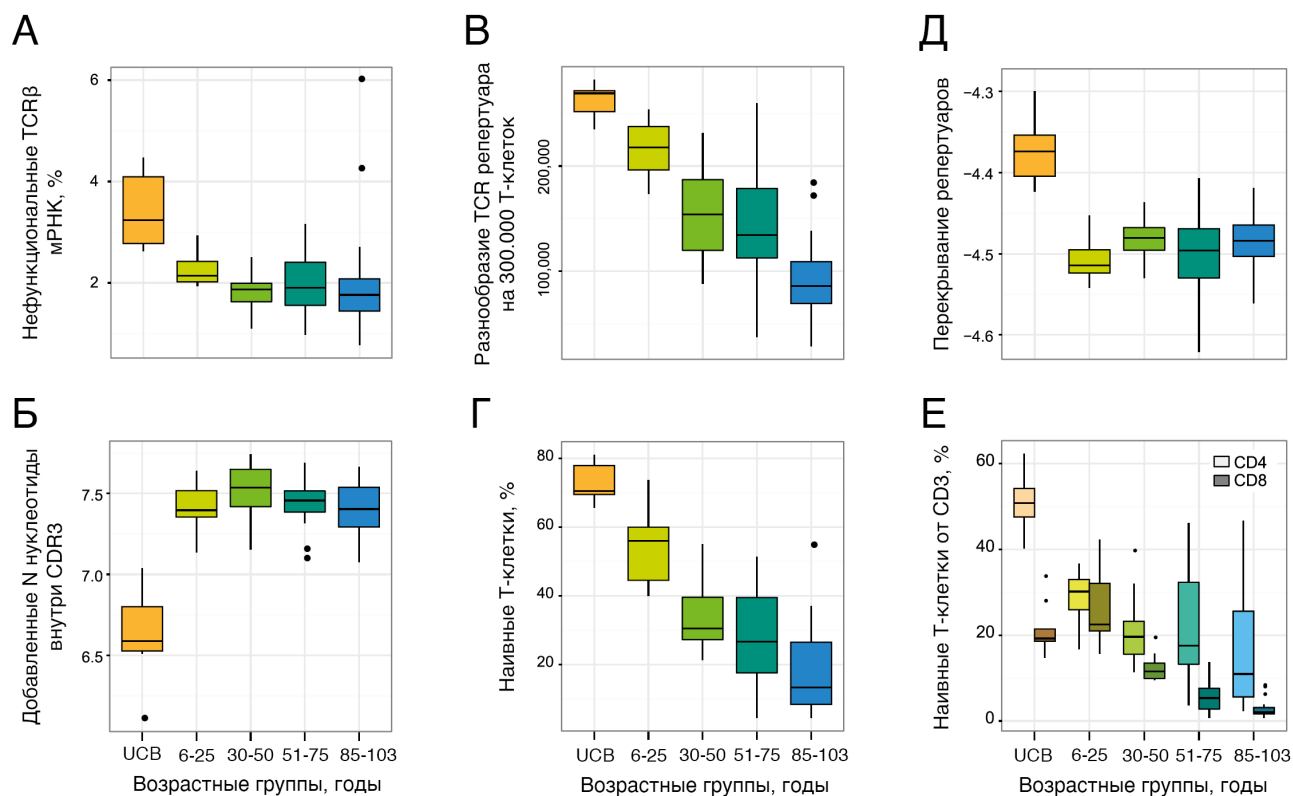
*Репертуары образцов пуповинной крови отличались большим количеством нефункциональных TCR*

Проанализировав полученные TCR $\beta$  репертуары доноров разных возрастных групп, мы обнаружили высокий процент нефункциональных молекул мРНК TCR $\beta$  для образцов пуповинной крови (Рисунок 6.А). Данные нефункциональные варианты содержали в своей последовательности CDR3 региона либо стоп-кодон, либо сдвиг в рамке считывания. Полученный результат, по нашему предположению, может быть следствием подавления системы нонсенс-опосредованного распада мРНК (англ. *nonsense-mediated mRNA decay, NMD*), непосредственно отвечающей за количество нефункциональной мРНК в клетке, в течение пренатального периода.

*Репертуары UCB отличались небольшим числом N нуклеотидов*

Мы обнаружили, что образцы UCB характеризовались значительно меньшим количеством добавленных случайных нуклеотидов в CDR3 регион (в среднем по клонотипу), чем в других возрастных группах (Рисунок 6.Б). Примечательно, что при нормировке на 300 000 Т-клеток такая особенность пренатальных репертуаров не влияла на разнообразие последовательностей TCR $\beta$ , которое было самым высоким в образцах UCB по сравнению с образцами периферической крови взрослых доноров (Рисунок 6.В). Почти каждая отсеквенированная молекула кДНК кодировала уникальный вариант TCR $\beta$ , а общее разнообразие приближалось к 250 000 различных клонотипов на образец UCB.

Наибольшее индивидуальное разнообразие последовательностей CDR3  $\beta$ -цепей Т-клеточных рецепторов, наблюдаемое в образцах пуповинной крови, коррелировало с самым высоким процентом наивных Т-клеток, и оба параметра равномерно снижались с возрастом (Рисунок 6.Г).



**Рисунок 6.** Отличия TCR $\beta$  репертуаров доноров разных возрастных групп. **(А)** Процент нефункциональных последовательностей TCR $\beta$  мРНК. **(Б)** Среднее число добавленных нуклеотидов внутри CDR3 TCR $\beta$ . **(В)** Уникальное разнообразие Т-клеточных рецепторов на нормированные 300 000 Т-клеток. **(Г)** Процент наивных клеток внутри репертуара от всех Т-клеток (на основании данных проточной цитометрии). **(Д)** Нормированное перекрытие репертуаров между донорами внутри собственной возрастной группы. **(Е)** Доли наивных CD4 (светлый боксплот) и CD8 (тёмный боксплот) Т-клеток внутри репертуара, измеренные с помощью проточной цитометрии.

### *Межиндивидуальное сходство репертуаров*

Мы проанализировали среднее число общих клонотипов между неродственными донорами в каждой возрастной группе с нормировкой данных на разнообразие репертуара в образце. Данное сравнение отражало общее сходство в разнообразии репертуаров Т-клеток и не учитывало размер анализируемых клонов. Перекрытие репертуаров было высоким между образцами UCB, где низкая активность TdT ограничивала теоретическое разнообразие полученных вариантов TCR $\beta$ . Однако в дальнейшем оно существенно снижалось и не изменялось на протяжении всего старения (Рисунок 6.Д).

### *Динамика наивных CD4 и CD8 Т-клеток внутри репертуара*

Для большей части доноров (68 из 73) мы провели окрашивание поверхностными маркерами небольшой аликвоты цельной крови, чтобы оценить долю наивных CD4 ( $CD3^+CD4^+CD27^{high}CD45RA^{high}$ ) и наивных CD8 ( $CD3^+CD4^-CD27^{high}CD45RA^{high}$ ) Т-клеток внутри тотальной популяции Т-клеток. Наивные CD4 Т-клетки представляли более половины всех Т-клеток в образцах пуповинной крови, в то время как наивные CD8 – около 20%. В молодом возрасте соотношение наивных популяций CD4 и CD8 понемногу выравнивалось (Рисунок 6.Е). Процент наивных Т-клеток уменьшался с возрастом (Рисунок 6.Г), причем наибольшее снижение наблюдалось среди CD8 Т-клеток (Рисунок 6.Е). Учитывая общее сокращение количества Т-клеток в крови с возрастом, уменьшение абсолютных количеств наивных Т-клеток на микролитр периферической крови было еще более существенным.

Далее мы проанализировали отношение доли наивных CD4 Т-клеток к доле наивных CD8 Т-клеток. Данный параметр не зависел от относительного содержания всех CD4 и CD8 клеток и хорошо отражал степень наивности CD4 репертуара по отношению к наивному CD8 репертуару. Проведенный анализ подтвердил наше предыдущее наблюдение, что популяция CD4 Т-клеток у самых пожилых доноров характеризуется более высокой долей наивных клеток сравнению с CD8 популяцией.

### *Функциональное сходство репертуаров*

Анализ корреляции долей общих для двух доноров TCR репертуаров показал, что функционально репертуары TCR более похожи в детстве, но с увеличением возраста постепенно расходятся с этим общим базовым репертуаром (Рисунок 7.А-Б). Высокая доля пересекающихся клонов в образцах пуповинной крови и у детей, скорее всего была обусловлена множеством публичных клонотипов из пренатального репертуара с высокой вероятностью сборки (Рисунок 7.В) при отсутствии активности TdT в эмбриональном развитии (Рисунок 6.Б).

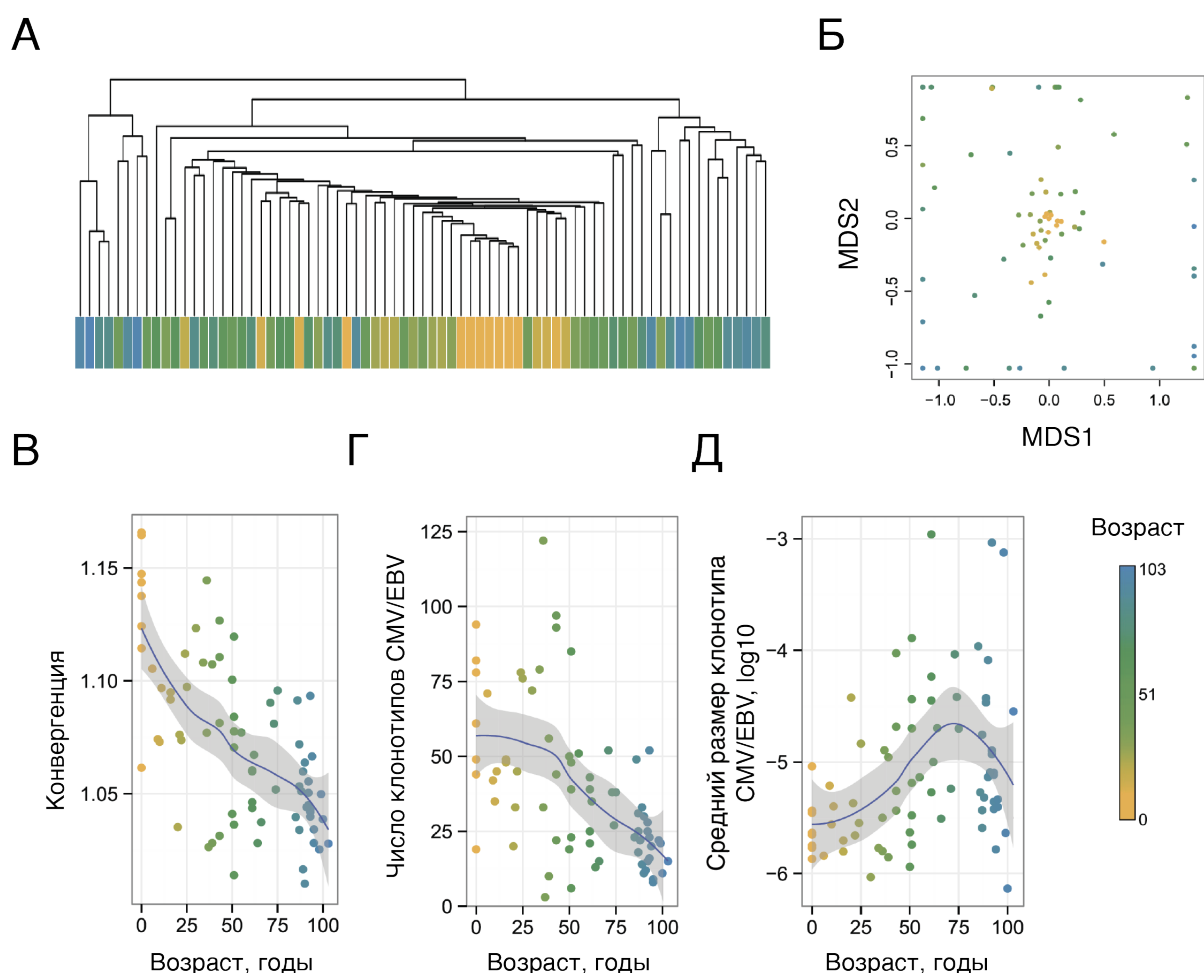
### *Динамика CMV и EBV специфичных клонов*

Используя собранный из литературных источников список известных CMV и EBV клонов, мы оценили изменение с возрастом количества и размера таких клонотипов в репертуарах наших доноров (Рисунок 7.Г-Д). Мы обнаружили снижение количества CMV и EBV специфичных вариантов TCR у пожилых доноров. В то же время средний размер клонотипов с возрастом увеличивался, причем в когорте от 51 до 75 лет наблюдался 7-кратный рост по сравнению с образцами пуповинной крови (Рисунок 7.Д). Мы предположили, что в молодости в наивном репертуаре присутствует большое количество вирус-специфичных



вариантов TCR. В процессе роста и развития организма, его коммуникации с окружающей средой происходит отбор и пролиферация определенных CMV и EBV специфичных клонов, которые в ходе своей дальнейшей экспансии вытесняют первоначальное разнообразие антиген-специфичного репертуара.

В глубокой старости мы наблюдали некоторое уменьшение размера CMV и EBV специфичных клонотипов (Рисунок 7.Д), что могло быть связано как с общим истощением антиген-специфичного репертуара Т-клеток в старости, так и с наличием возрастного барьера. Ранее было показано, что отсутствие CMV инфекции ассоциировано с долгожительством. Таким образом, условный возрастной барьер в 75 лет преимущественно преодолевают люди без CMV инфекции, что может вносить искажение в представленность CMV-специфичных клонотипов среди когорты долгожителей в нашем исследовании.



**Рисунок 7.** Расхождение репертуаров с возрастом. (А, Б) Кластерный анализ, где дистанция между донорами пропорциональна сумме частот перекрывающихся клонотипов. (А) Дендрограмма на основе иерархической кластеризации. (Б) MDS анализ. (В) Конвергенция VDJ рекомбинации. Рассчитана на основании среднего числа нуклеотидных последовательностей CDR3, соответствующих одной аминокислотной последовательности CDR3. (Г) Общее число CMV и EBV специфичных клонотипов, обнаруженных в репертуарах. (Д) Средний размер CMV/EBV специфичного клонотипа.

### 3. Изменение характеристик наивных репертуаров с возрастом

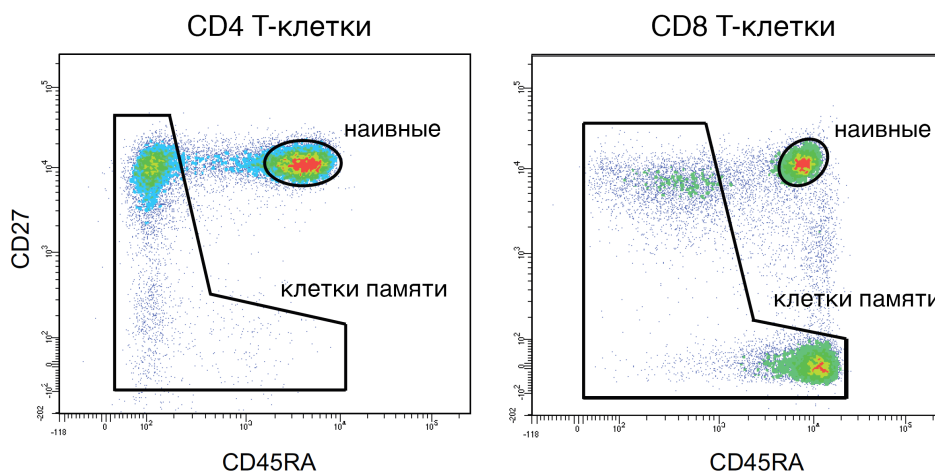
В нашем исследовании возрастных изменений в структуре TCR $\beta$  репертуаров наивных Т-клеток принимали участие 13 системно здоровых доноров, разделенных на две возрастные группы:

1. Молодые доноры в возрасте от 25 до 32 лет (в среднем 30 лет)
2. Взрослые/пожилые доноры в возрасте от 51 года и до 88 лет (в среднем 65 лет)

Наши данные мы дополнили данными из недавно опубликованной работы Qi et al. 2014. В сформированные когорты было добавлено 4 молодых (от 20 до 35 лет) и 5 пожилых (от 70 до 85 лет) здоровых доноров. Таким образом, суммарно в работе принимали участие 22 донора в возрасте от 20 до 88 лет.

#### 3.1. Характеристики TCR репертуаров наивных CD4 и CD8 Т-клеток

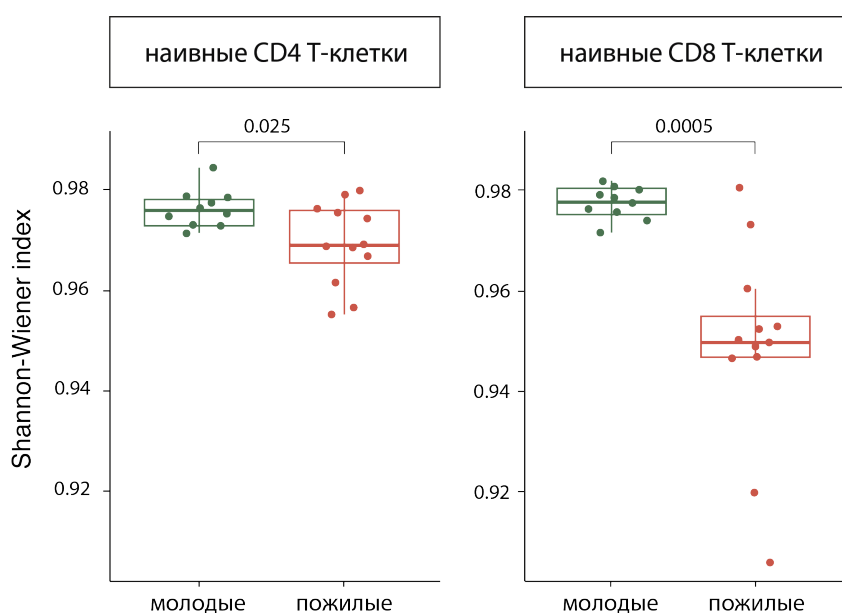
Для оценки возрастных изменений в характеристиках наивных TCR репертуаров мы отсортировали (Рисунок 8) наивные CD4 (CD3<sup>+</sup>CD4<sup>+</sup>CD27<sup>high</sup>CD45RA<sup>high</sup>) и наивные CD8 Т-клетки (CD3<sup>+</sup>CD4<sup>-</sup>CD27<sup>high</sup>CD45RA<sup>high</sup>) молодых и пожилых доноров. Библиотеки CDR3 последовательностей Т-клеточных рецепторов были приготовлены с использованием описанного в части 1 метода.



**Рисунок 8.** Пример выделения популяций наивных и Т-клеток памяти для CD4 и CD8 лимфоцитов одного молодого донора.

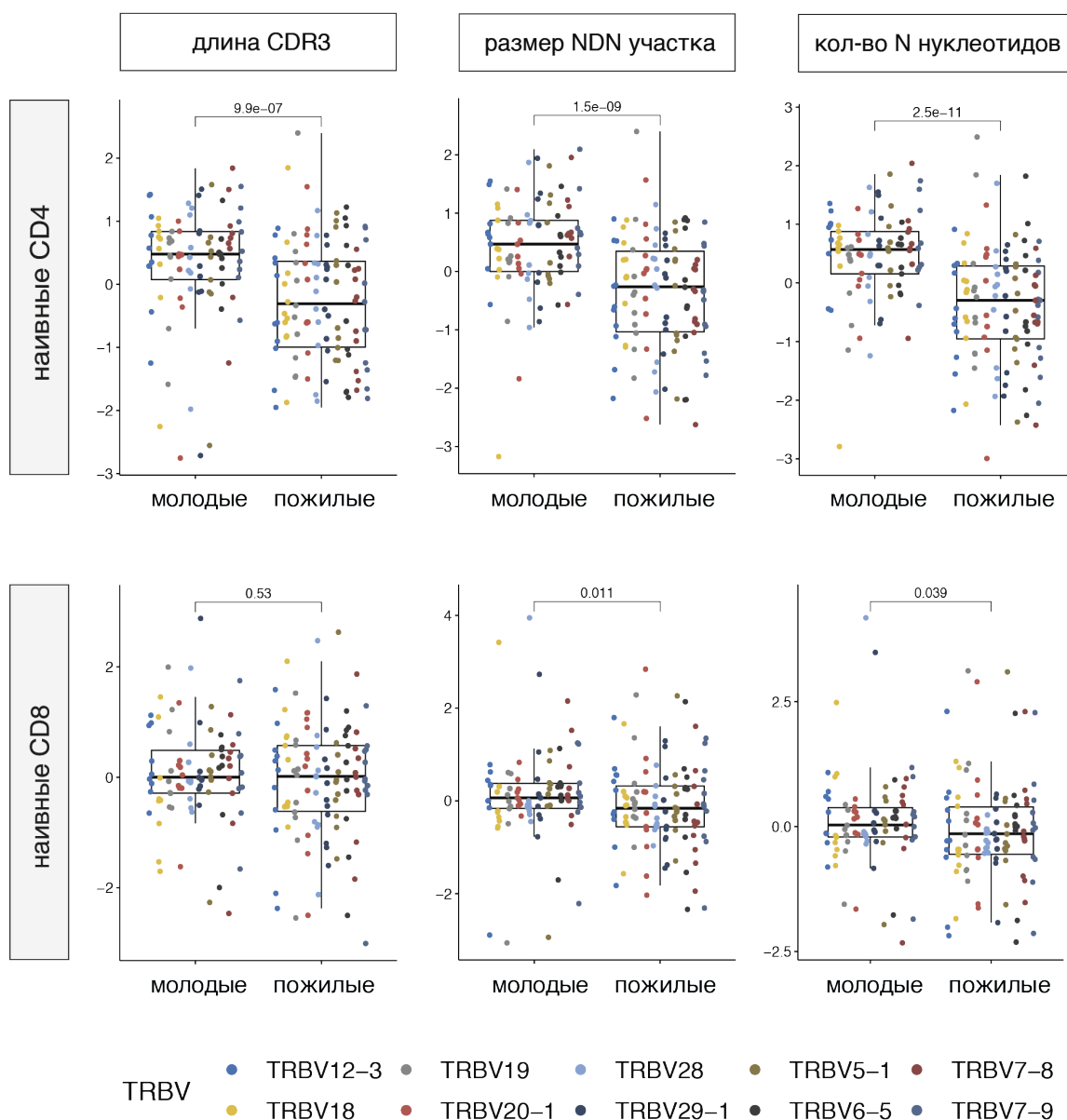
Результаты анализа объединенных данных с использованием нормализованного индекса разнообразия Шеннона-Винера продемонстрировали, что с возрастом наивные CD4 и CD8 Т-клетки накапливают клональные экспансии (Рисунок 9).

Далее мы рассмотрели изменение стандартных характеристик TCR репертуаров (средняя длина CDR3 участка в нуклеотидах, размер вставки NDN и количество случайно добавленных (N) нуклеотидов) для 10 наиболее представленных TRBV генных семейств: TRBV5-1, TRBV6-5, TRBV7-8, TRBV7-9, TRBV12-3, TRBV18, TRBV19, TRBV20-1, TRBV28 и TRBV29-1. Под вставкой NDN подразумевался участок внутри последовательности CDR3 между V и J сегментами и включающий в себя D сегмент (при наличии) и добавленные N нуклеотиды. N нуклеотиды – добавленные случайные нуклеотиды между генными сегментами за счет работы фермента TdT при перестройке генов T-клеточных рецепторов во время созревания наивных T лимфоцитов в тимусе. Именно за счет данных нуклеотидов, их количества (соответственно и размера NDN региона), без учета комбинации пар V и J генных сегментов, во многом обеспечивается огромное разнообразие репертуаров TCR.



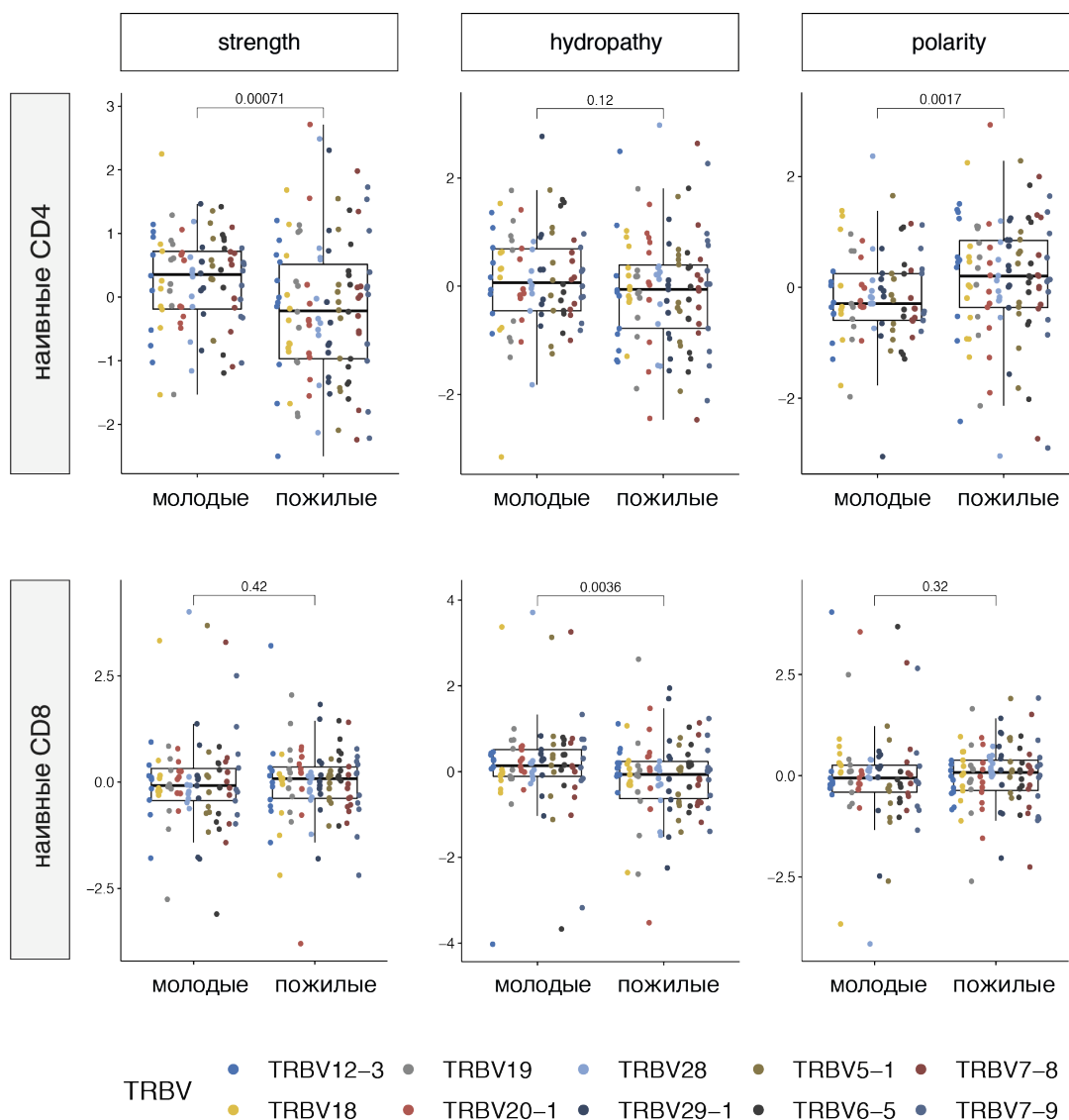
**Рисунок 9.** Наивные CD4 и CD8 Т-клетки накапливают клональные экспансии с возрастом. Показан нормализованный индекс разнообразия Шеннона-Винера для TCR $\beta$  CDR3 репертуаров. Объединенные данные из нашего исследования и Qi et al. 2014.

Мы обнаружили уменьшение с возрастом средней длины CDR3, размера вставки NDN и количества добавленных нуклеотидов для наивных CD4 Т-клеточных репертуаров. В то же время для наивных CD8 репертуаров изменения носили минимальных характер, и статистически значимые различия были получены для размера вставки и среднего числа N нуклеотидов, количество которых как в случае с наивными CD4 клетками уменьшалось у пожилых доноров (Рисунок 10).



**Рисунок 10.** Базовые характеристики репертуаров Т-клеточных рецепторов для наивных CD4 и CD8 Т-клеток. Анализ проведен для 10 наиболее представленных TRBV генных сегментов. Характеристики приведены после z-score нормализации. Объединенные данные из нашего исследования и Qi et al. 2014.

Кроме того, мы рассмотрели биофизические характеристики аминокислотных остатков в центральной части CDR3 региона (5 аминокислот), используя функцию “Annotate” в VDJTools. Как и в случае со стандартными характеристиками репертуаров TCR значения для всех использованных генных сегментов были выровнены с учетом z-score нормализации. Были проанализированы следующие показатели: strength, hydrophathy, polarity. Рассчитав указанные характеристики для наших TCR $\beta$  CDR3 репертуаров, мы наблюдали значимое изменение параметров “strength” и “polarity” у наивных CD4 и “hydrophathy” у наивных CD8 Т-лимфоцитов с возрастом (Рисунок 11).



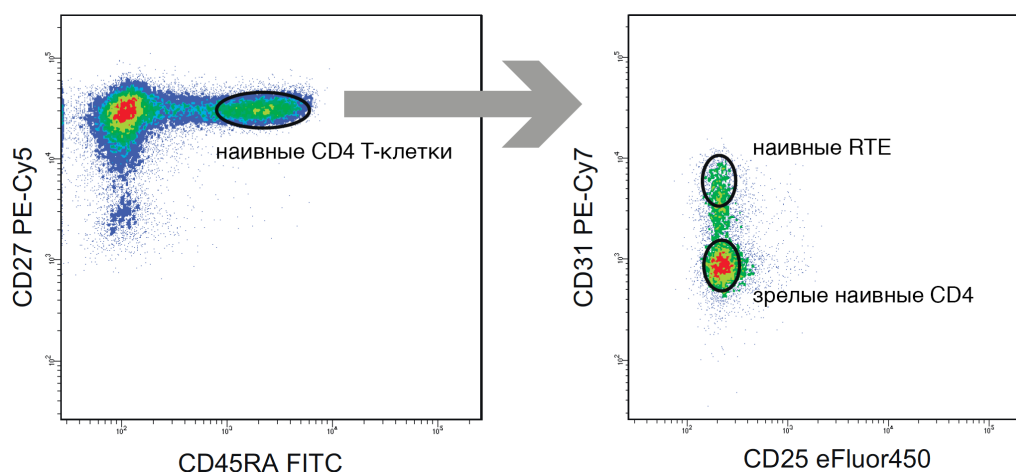
**Рисунок 11.** Биофизические характеристики репертуаров Т-клеточных рецепторов наивных CD4 и CD8 Т-клеток. Характеристики приведены с учетом z-score нормализации. Для сравнения биофизических свойств репертуаров молодых (25-35 лет) и пожилых (51-88) доноров были выбраны следующие характеристики аминокислотных остатков внутри CDR3 региона: “strength” (сила взаимодействия), “hydropathy” (гидрофильность/гидрофобность) и “polarity” (полярность). Анализ проведен для 10 наиболее крупных TRBV генных семейств. Объединенные данные из нашего исследования и Qi et al. 2014.

Данные наблюдения указывают на присутствие определенной периферической селекции наивного Т-клеточного репертуара при старении иммунной системы. Уменьшение с возрастом относительного содержания сильно взаимодействующих аминокислотных остатков и средней длины CDR3 в TCRβ репертуаре наивных Т-клеток может свидетельствовать о значительно более быстром истощении клонов наивных Т-клеток с высокой аффинностью к собственным рМНС. Это может являться результатом более эффективной передачи тонических сигналов и, как правило, более быстрой пролиферации, а затем истощения пролиферативной способности и исчезновения таких наивных Т-клеток из общего репертуара.

### 3.2. Изменения характеристик в репертуарах RTE и зрелых наивных CD4 Т-клеток

Обе субпопуляции наивных CD4 Т-клеток – зрелые наивные CD4 Т-клетки ( $CD45RA^+CD31^-$ ) и RTE ( $CD45RA^+CD31^+$ ) – способны поддерживать свою численность за счет периферического деления. При этом делящиеся клетки из популяции RTE способны продолжительное время сохранять экспрессию CD31 маркера. Тем не менее, со временем количество  $CD45RA^+CD31^+$  наивных CD4 Т-клеток в организме заметно уменьшается. Следовательно, можно предположить, что характеристики зрелых  $CD31^-$  наивных CD4 Т-клеток изменяются более существенно, чем характеристики RTE. А свойства наивных CD4 Т-клеток в целом могут меняться как за счет внутренних различий между свойствами RTE и зрелых наивных Т-клеток, так и за счет уменьшения доли  $CD31^+$  клеток с возрастом в популяции наивных CD4 Т-клеток.

Для проверки последнего предположения мы сравнили характеристики  $TCR\beta$  репертуаров отсортированных субпопуляций RTE ( $CD4^+CD45RA^{high}CD27^{high}CD31^+CD25^-$ ) и зрелых наивных CD4 Т-клеток ( $CD4^+CD45RA^{high}CD27^{high}CD31^-CD25^-$ ) для 6 молодых и 4 пожилых здоровых доноров (Рисунок 12). Сортировка  $CD25^-$  наивных CD4 Т-клеток позволила исключить вклад в анализ регуляторных Т-клеток ( $CD4^+CD25^+$ , англ. *Treg*), характеристики которых существенно отличаются от канонических CD4 Т-клеток.

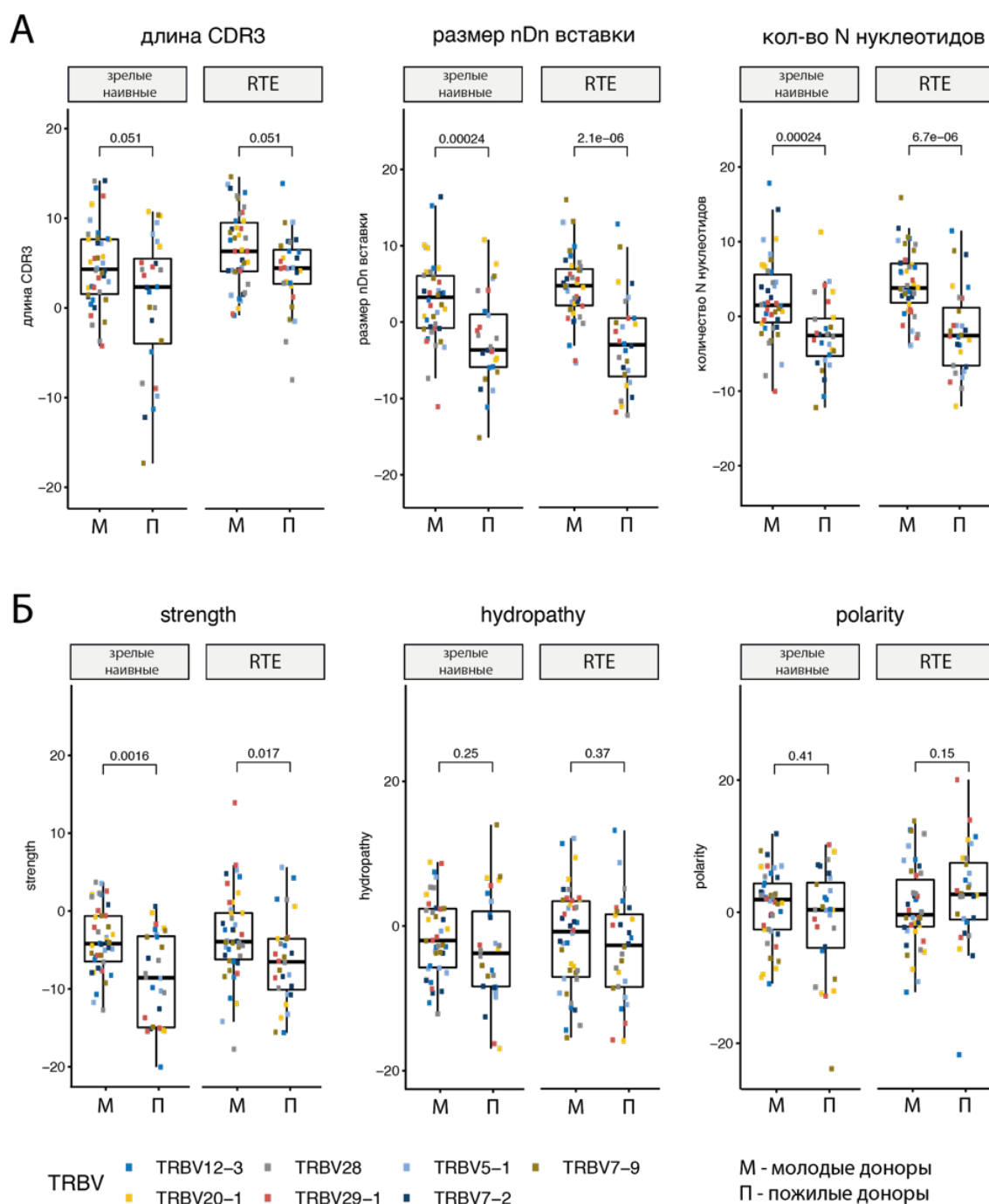


**Рисунок 12.** Стратегия выделения субпопуляций RTE и зрелых наивных CD4 Т-клеток при сортировке с помощью FACS на примере одного молодого донора.

Анализ полученных CDR3 последовательностей показал, что похожие изменения наблюдались в  $TCR\beta$  репертуарах зрелых наивных CD4 и RTE Т-лимфоцитов (Рисунок 13). Характеристики обеих субпопуляций наивных CD4 Т-клеток были практически идентичны внутри каждой возрастной группы, но заметно различались между молодыми и пожилыми донорами как при рассмотрении стандартных параметров CDR3  $TCR\beta$  репертуаров – средняя длина CDR3 участка, количество добавленных нуклеотидов и размер вставки между генными

сегментами – (Рисунок 19.А), так и при анализе биофизических свойств аминокислотных остатков, образующих центр CDR3 региона (Рисунок 13.Б).

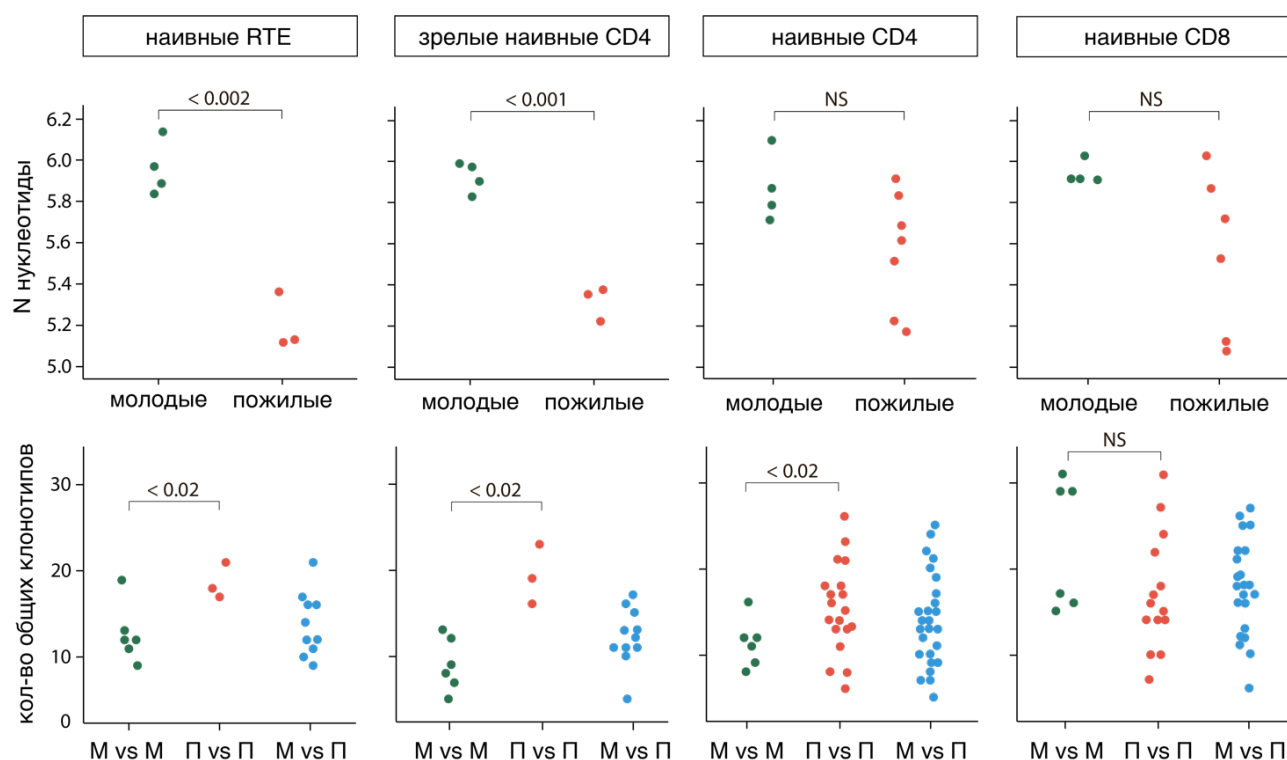
Таким образом мы пришли к выводу, что наблюдаемые изменения в пуле наивных CD4 Т-клеток затрагивают обе субпопуляции и RTE, и зрелых наивных клеток и не зависят от падения доли RTE с возрастом или изменения соотношения CD31<sup>+</sup>/CD31<sup>-</sup> клеток в целом.



**Рисунок 13.** Свойства репертуаров для зрелых наивных CD4 Т-клеток и RTE для 7 наиболее широко представленных семейств наивных CD4 Т-клеток. Z-score нормализация выполнена для объединенного сравнения TRBV генных семейств. (А) Средний размер участка CDR3, размер NDN вставки и количество добавленных N нуклеотидов. (Б) Биофизические характеристики для 5 центральных аминокислот внутри CDR3 региона. М – когорта молодых, П – пожилых доноров.

### 3.3. Рост публичности репертуара наивных CD4 Т-клеток с возрастом

Короткие CDR3 последовательности и небольшое число добавленных нуклеотидов часто оказываются ассоциированы с высокой публичностью таких репертуаров. Поскольку мы наблюдали значительное укорочение длины CDR3 как у тотальных наивных CD4 (Рисунок 10), так и у субпопуляций RTE и зрелых наивных CD4 Т-клеток (Рисунок 13) с возрастом, мы предположили рост публичности наивных репертуаров у пожилых доноров. Для того чтобы оценить изменение публичности наивного CD4 репертуара с возрастом, мы извлекли из данных по 3 000 наиболее крупных клонотипов для каждого донора. Сравнив перекрытие TCR $\beta$  CDR3 репертуаров наших доноров, мы обнаружили увеличение публичности наивного CD4, RTE и зрелого наивного репертуаров с возрастом (Рисунок 14). Умеренное перекрытие репертуаров было показано для молодых и взрослых/пожилых доноров среди наивных CD4, RTE и зрелых наивных CD4 Т-клеток. Однако, для наивных репертуаров CD8 Т-клеток мы не обнаружили изменение публичности с возрастом.



**Рисунок 14.** Относительная публичность наивных Т-клеточных репертуаров, рассчитанная для 3 000 наиболее крупных клонотипов внутри каждого образца. (А) Среднее число добавленных нуклеотидов внутри каждого образца. Анализ учитывал только CDR3 последовательности, где был определен D сегмент. Анализ выполнен для 4 молодых и 3 пожилых доноров – наивные RTE и зрелых наивные CD4 и для 4 молодых и 7 пожилых доноров – наивные CD4 и CD8 Т-клетки. (Б) Количество перекрывающихся аминокислотных последовательностей CDR3 для каждой пары доноров: М – молодой, П – пожилой донор.



## Выводы

1. На основе технологий 5'-RACE и молекулярного баркодирования (UMI) был разработан метод анализа репертуаров Т-клеточных рецепторов, применимый в том числе для небольших стартовых количеств клеток (несколько сотен), позволяющий корректно и нормировано сравнивать их характеристики.
2. С применением разработанного метода была продемонстрирована возможность предсказания нативных пар  $\alpha\beta$  цепей Т-клеточных рецепторов.
3. Была показана высокая стабильность периферического Т-клеточного репертуара человека: значительная часть крупных клонотипов, занимающих не менее половины репертуара взрослого здорового донора, сохраняет свои частоты на протяжении как минимум трех лет.
4. С использованием разработанного метода были проанализированы тотальные TCR $\beta$  репертуары периферической крови доноров разного возраста (от 6 до 103 лет) и образцов пуповинной крови (UCB). Показано, что:
  - a. репертуары UCB отличаются большим количеством нефункциональных последовательностей CDR3 TCR, небольшим числом добавленных N нуклеотидов и высоким межиндивидуальным сходством;
  - b. процент наивных CD4 и CD8 клеток практически линейно падает с возрастом и напрямую коррелирует с сокращением разнообразия репертуара TCR. При этом популяция CD4 Т-клеток у самых пожилых доноров (со средним возрастом 92 года) характеризуется более высокой долей наивных клеток сравнению с CD8 популяцией;
  - c. с возрастом уменьшается общее количество CMV и EBV специфичных клонотипов, при этом наблюдается увеличение среднего размера подобных клонотипов.
  - d. возрастная группа долгожителей (средний возраст 92 года) по сравнению с группой пожилых доноров (средний возраст 60 лет) отличается меньшим количеством и размером CMV и EBV специфичных клонотипов, что вместе с относительно высокой долей наивных CD4 Т-клеток может указывать на наличие возрастного отбора и быть связано с долголетием.

5. Проанализированы CDR3 TCR $\beta$  репертуары наивных CD4 и CD8 клеток периферической крови для доноров двух возрастных групп. Было показано, что:
  - a. у пожилых доноров наблюдается рост клональных экспансий наивных CD4 и CD8 Т-клеток;
  - b. репертуары наивных CD4 Т-клеток пожилых доноров характеризуются уменьшением длины CDR3 участка, размера NDN вставки и количества N нуклеотидов. Для CD8 клеток эти изменения не наблюдаются, либо носят слабо выраженный характер;
  - c. происходит изменение биофизических характеристик центральной части CDR3 региона у наивных CD4 Т-клеток, что в совокупности с изменением базовых характеристик может быть следствием периферической селекции наивного Т-клеточного репертуара с возрастом;
  - d. с возрастом наблюдается рост публичности наивного CD4 репертуара, что может быть ассоциировано с преимущественной селекцией коротких вариантов CDR3 с меньшим количеством добавленных нуклеотидов.
  
6. Наблюдаемые изменения в пуле наивных CD4 Т-клеток затрагивают как субпопуляцию недавних эмигрантов из тимуса (RTE), так и зрелых наивных клеток, и не зависят от падения доли RTE с возрастом или изменения соотношения CD31<sup>+</sup>/CD31<sup>-</sup> клеток в целом.

## Список работ, опубликованных по теме диссертации

### Публикации

1. Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, Greenshields-Watson A, Attaf M, **Egorov ES**, Zvyagin IV, Babel N, Cole DK, Godkin AJ, Sewell AK, Kesmir C, Chudakov DM, Luciani F, Shugay M. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* 2019 Oct 7. pii: gkz874. doi: 10.1093/nar/gkz874.
2. Li N, van Unen V, Abdelaal T, Guo N, Kasatskaya SA, Ladell K, McLaren JE, **Egorov ES**, Izraelson M, Chuva de Sousa Lopes SM, Höllt T, Britanova OV, Eggermont J, de Miranda NFCC, Chudakov DM, Price DA, Lelieveldt BPF, Koning F. Memory CD4<sup>+</sup> T cells are generated in the human fetal intestine. *Nature Immunology.* 2019 Mar;20(3):301-312. doi: 10.1038/s41590-018-0294-9. Epub 2019 Jan 21.
3. **Egorov ES**, Kasatskaya SA, Zubov VN, Izraelson M, Nakonechnaya TO, Staroverov DB, Angius A, Cucca F, Mamedov IZ, Rosati E, Franke A, Shugay M, Pogorelyy MV, Chudakov DM, Britanova OV. The Changing Landscape of Naive T Cell Receptor Repertoire With Human Aging. *Front Immunol.* 2018 Jul 24; 9:1618. doi: 10.3389/fimmu.2018.01618. eCollection 2018.
4. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, Komech EA, Sycheva AL, Koneva AE, **Egorov ES**, Eliseev AV, Van Dyk E, Dash P, Attaf M, Rius C, Ladell K, McLaren JE, Matthews KK, Clemens EB, Douek DC, Luciani F, van Baarle D, Kedzierska K, Kesmir C, Thomas PG, Price DA, Sewell AK, Chudakov DM. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* 2018 Jan 4;46(D1): D419-D427. doi: 10.1093/nar/gkx760.
5. Britanova OV, Shugay M, Merzlyak EM, Staroverov DB, Putintseva EV, Turchaninova MA, Mamedov IZ, Pogorelyy MV, Bolotin DA, Izraelson M, Davydov AN, **Egorov ES**, Kasatskaya SA, Rebrikov DV, Lukyanov S, Chudakov DM. Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians. *J Immunol.* 2016 Jun 15;196(12):5005-13. doi: 10.4049/jimmunol.1600005. Epub 2016 May 13.
6. **Egorov ES**, Merzlyak EM, Shelenkov AA, Britanova OV, Sharonov GV, Staroverov DB, Bolotin DA, Davydov AN, Barsova E, Lebedev YB, Shugay M, Chudakov DM. Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J Immunol.* 2015 Jun 15;194(12):6155-63. doi: 10.4049/jimmunol.1500215. Epub 2015 May 8.

7. **Егоров Е.С.**, Израельсон М.А., Касацкая С.А., Чудаков Д.М., Лукьянов С.А. Количественный и безошибочный анализ данных массированного секвенирования с использованием молекулярного баркодирования. *Вестник РГМУ*. 2015. № 4. С. 4-9.

**Конференции и семинары:**

1. Егоров Е.С., Израельсон М.А., Касацкая С.А., Наконечная Т.О., Британова О.В. Возрастные изменения в структуре Т-клеточных репертуаров. *Всероссийская конференция "Высокопроизводительное секвенирование в геномике"*, 18-23 июня 2017, Новосибирск.
2. Egorov E.S. Age-associated alterations in the structure of naive TCR repertoire. *Skoltech Systems Biology Seminar*, May 28, 2019. Moscow.