# ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР ИНСТИТУТ ЦИТОЛОГИИ И ГЕНЕТИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»

На правах рукописи

# СМИРНОВ АЛЕКСАНДР ВАСИЛЬЕВИЧ

# ИССЛЕДОВАНИЕ МЕХАНИЗМОВ, ОБЕСПЕЧИВАЮЩИХ ИНТЕГРАЦИЮ ГЕНЕТИЧЕСКИХ КОНСТРУКЦИЙ ПРИ ПОЛУЧЕНИИ ТРАНСГЕННЫХ МЫШЕЙ МЕТОДОМ ПРОНУКЛЕАРНОЙ МИКРОИНЪЕКЦИИ

03.02.07 - генетика

Диссертация на соискание ученой степени кандидата биологических наук

> Научный руководитель: кандидат биологических наук Баттулин Нариман Рашитович

Новосибирск 2020

# ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ	4
введение	5
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	10
1.1. Конкатемеры и пронуклеарная микроинъекция	10
1.1.1. Основы пронуклеарной микроинъекции	10
1.1.2. Структура конкатемеров: модели конкатемеризации	12
1.1.3. Структура конкатемеров: встройка в геном	15
1.1.4. Особенности пронуклеарной микроинъекции: мозаицизм	17
1.1.5. Особенности пронуклеарной микроинъекции: уровень экспрессии	
1.1.6. Особенности пронуклеарной микроинъекции: копийность	19
1.1.7. Сборка протяженных конструкций из фрагментов	20
1.2. Современные методы секвенирования встроек конкатемеров	21
1.2.1. Методы, базирующиеся на NGS анализе	22
1.2.2. Таргетная амплификация локуса (TLA)	25
1.2.3. Секвенирование третьего поколения (LRS)	28
1.3. Механизмы репарации ДНК и конкатемеризация	30
1.3.1. Негомологичное соединение концов (non-homologous end-joining, NHEJ)	
1.3.2. Гомологичная рекомбинация (HR)	
1.3.3. Взаимодействие NHEJ и HR	36
1.3.4. Стабильность палиндромов в геноме и генетических конструкциях	
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ	44
2.1. Клонирование баркодированной плазмидной библиотеки	44
2.2. Микроинъекция баркодированной линеаризованной библиотеки.	45
2.3. ПЦР-анализ трансгенных конструкций и генотипирование эмбрионов	46
2.4. Определение копийности трансгена методом droplet digital PCR (ddPCR)	46
2.5. Приготовление библиотек для NGS	47
2.6. Секвенирование РасВіо SMRT	50
ГЛАВА 3. РЕЗУЛЬТАТЫ	52

3.1. Получение трансгенных эмбрионов с конкатемерами	
3.1.1. Дизайн и клонирование барколированных генетических конструкций	
3.1.2. Пронуклеарная микроинъекция и генотипирование эмбрионов	53
3.1.3. Определение количества копий трансгена в конкатемерах.	54
3.1.4.NGS анализ баркодов в конкатемерах	57
3.1.5. Независимое подтверждение данных NGS	60
3.2. Анализ карт конкатемеров	65
3.2.1. Конкатемеры образуются без участия <i>de novo</i> амплификации	65
3.2.2. Концы трансгенов активно процессируются HR	67
3.2.3. Механизмы рекомбинации, основанные на анализе паттернов связей	68
3.2.4. Индикаторы кроссинговера и dHJ среди паттернов связей	71
3.2.5. Роль NHEJ в конкатемеризации	73
3.2.6. Локализация встроек конкатемеров в эмбрионах и трансгенных линиях	76
ГЛАВА 4. ОБСУЖДЕНИЕ	79
выводы	90
СПИСОК ЛИТЕРАТУРЫ	91
ПРИЛОЖЕНИЕ 1	111
ПРИЛОЖЕНИЕ 2	112
ПРИЛОЖЕНИЕ 3	113
ПРИЛОЖЕНИЕ 4	114
ПРИЛОЖЕНИЕ 5	115
ПРИЛОЖЕНИЕ 6	116
ПРИЛОЖЕНИЕ 7	117

# СПИСОК СОКРАЩЕНИЙ

NGS – next generation sequencing, секвенирование второго поколения (в т.ч. применительно к Illumina HiSeq 2500)

dHJ – double Holliday junction, двойная холидеевская структура

DSB – double-strand break, двухцепочечный разрыв ДНК

ddPCR – droplet digital PCR, капельная цифровая ПЦР

EBOBE – elongation beyond original broken end, удлинение обломанного конца трансгена

NHEJ – non-homologous end-joining, негомологичное соединение концов

HR – homologous recombination, гомологичня рекомбинация

LRS – single molecule long-read sequencing, секвенирование третьего поколения

BAC – bacterial artificial chromosome, искусственная бактериальная хромосома

LOH – loss of heterozygosity, потеря гетерозиготности

SDSA – synthesis-dependent strand annealing

**DSBR** – double-strand break repair

MMEJ - microhomology-mediated end joining

**BIR** – break-induced replication

MIR – multi-invasion (MI)-induced rearrangement

**SSA** – single strand annealing

MAR – matrix attachment region

CDK – cyclin-dependent kinase, циклин-зависимая киназа

**PEV** – position-effect variegation

HIRI – homologous illegitimate random integration

YAC – yeast artificial chromosome, искусственная хромосома дрожжей

MMR – mismatch repair, мисматч репарация

NFP – nucleoprotein filament

**RIGS** – repeat-induced gene silencing

#### введение

Актуальность проблемы. Микроинъекция ДНК в пронуклеус зиготы способом получения трансгенных животных. Впервые является основным эффективность этого метода была продемонстрирована почти 40 лет назад (Gordon et al., 1980; Brinster et al., 1981). Еще тогда исследователи отмечали удивительное свойство пронуклеарной микроинъекции - трансгены встраивались в единый локус, образуя большие повторенные районы ("конкатемеры") ИЗ копий, ориентированных в одном направлении, - "голова-к-хвосту". Это явление подверглось пристальному изучению co стороны групп, занимающихся трансгенезом (Folger et al., 1982; McKnight et al., 1983; Brinster et al., 1985). Стало понятно, что копийность конкатемеров коррелирует с числом инъецированных молекул; что линеаризованные плазмиды чаще, чем циклические, рекомбинируют и встраиваются в геном; что большая часть копий конкатемера (>90%) встраивается тандемно, а не в случайной ориентации.

В ходе последующего десятилетия было предложено несколько объяснений феномену "конкатемеризации" (concatenation). Одна из моделей, которую можно назвать "модель de novo амплификации" или "модель катящегося кольца" (rolling circle amplification), гласит, что конкатемер возникает из одной или нескольких циклических копий за счет многократной репликации (Rohan *et al.*, 1990). В пользу этой модели говорит высокая эффективность конкатемеризации (в конкатемер могут включаться тысячи копий). Второй гипотетический механизм конкатемеризации можно обозначить, как "модель перекрывающихся фрагментов" (Bishop, 1996; Smith, 2001). Как известно, линеаризованные молекулы трансгенов способны образовывать циклические молекулы после инъекции. Случайная фрагментация таких молекул приведет к появлению большого пула ДНК, перекрывающихся гомологичных участков которые эффективно рекомбинируют В зиготе. Реальность этого процесса была косвенно продемонстрирована в новаторских, но забытых статьях по сборке больших (>35000 п.о.) трансгенов из инъецированных фрагментов непосредственно в зиготе (Pieper et al., 1992; Tacken et al., 2001). Наконец, самая простая гипотеза постулирует, что конкатемеры образуются путем рекомбинации между линейными

концами молекул трансгенов с некоторым вкладом циклических копий (Hamada et al., 1993; Brinster et al., 1985). Эта модель наиболее реалистична с точки зрения механизма, но не объясняет, почему линейные концы более предпочтительны для факторов резекции и гомологичной рекомбинации, а не для лигирования по механизму негомологичного соединения концов (non-homologous end-joining, NHEJ). Удивительно, но несмотря на стремительный прогресс в областях связанных с трансгенезом И репарацией ДНК. биологии. однозначного подтверждения ни одна из предложенных моделей конкатемеризации так и не получила, даже спустя десятки лет после их появления.

Современные методы секвенирования геномов, такие как TLA (Targeted Locus Amplification), или секвенирование третьего поколения (Pacific Biosciences, Oxford Nanopore), позволяют с высокой точностью установить место встройки и структуру конкатемеров у трансгенных животных. К сожалению, все копии трансгенов в конкатемере идентичны, поэтому данные секвенирования не могут объяснить, как проходит процесс внехромосомной рекомбинации трансгенов, или однозначно установить их порядок в итоговом конкатемере. Чтобы проследить судьбу трансгенов в зиготе, мы приняли решение пометить каждую из инъецированных молекул уникальным ДНК-баркодом, представленным случайной последовательностью нуклеотидов.

**Целью диссертационной работы** является выяснение механизма формирования конкатемеров в зиготе при использовании метода пронуклеарной микроинъекции.

Для достижения этой цели были поставлены следующие задачи:

- Разработать дизайн оптимальной последовательности баркодов и выбрать участки их расположения в трансгене. Собрать и просеквенировать библиотеку плазмидных конструкций, содержащих парные концевые баркоды с достаточным разнообразием (10<sup>5</sup>-10<sup>6</sup> уникальных молекул).
- 2. Получить трансгенных животных со встройкой конкатемеров в геноме.
- Просеквенировать концевые баркоды трансгенов в конкатемерах у животных с различной копийностью. Оценить представленность и копийность

6

индивидуальных баркодов в животных и сравнить их с исходной плазмидной библиотекой. Сделать вывод об основном механизме конкатемеризации в пронуклеарной микроинъекции.

4. Построить карты конкатемеров на основе анализа взаимного расположения баркодов в конкатемерах. Изучить паттерны связей между баркодированными трансгенами и оценить с их помощью вклад различных путей репарации двухцепочечных разрывов ДНК в конкатемеризацию.

#### Научная новизна:

- 1. Получена и охарактеризована уникальная баркодированная плазмидная библиотека, пригодная для исследований механизмов рекомбинации ДНК на клетках и эмбрионах.
- 2. Впервые применены различные молекулярные и биоинформатические ДНК-баркодов тандемно-повторенных методы анализа В Опробованы последовательностях трансгенов. различные методы интерпретации данных NGS баркодированных визуализации И ДЛЯ трансгенов.
- 3. Впервые применен современный метод секвенирования третьего поколения (PacBio) ДЛЯ анализа ДНК-баркодов В геномной ДНК. Также продемонстрировано, что ланный метод детектировать позволяет "идеальные" палиндромные слияния трансгенов.
- 4. Описан новый продукт рекомбинации трансгенной ДНК (elongation beyond original broken end, EBOBE), который может быть распространенным элементом конкатемеров, характерным для метода пронуклеарной микроинъекции.

Теоретическая и практическая значимость исследования. Баркодированная плазмидная библиотека, охарактеризованная в данной работе, может применяться для изучения различных аспектов рекомбинации ДНК в зиготе и культурах клеток. Основываясь на результатах, полученных в данной работе, можно представить рекомендации по улучшению эффективности трансгенеза в пронуклеарной микроинъекции. Например, временное блокирование активности NHEJ совместно с добавлением в ядро циклических копий трансгена может увеличить копийность трансгена. Использование перекрывающихся длинных одоцепочечных фрагментов может стимулировать сборку составных двухцепочечных конструкций за счет активности гомологичной рекомбинации.

#### Основные положения, выносимые на защиту:

1. Формирование конкатемеров в пронуклеарной микроинъекции происходит за счет гомологичной рекомбинации между концами линейных молекул трансгенов.

2. Баркодирование трансгенов выявляет характерные сигнатуры активности разных путей репарации двухцепочечных разрывов ДНК.

Апробация результатов и публикации. Результаты докладывались на международном конгрессе CRISPR-2018 (Новосибирск). По теме диссертации были опубликованы 4 работы. Основные результаты были опубликованы в рецензируемом журнале "Nucleic Acids Research". Дополнительные материалы с характеристиками конкатемерных встроек и сайтов интеграции были опубликованы ранее в 3 статьях в рецензируемом журнале "Transgenic Research".

- Smirnov A., Fishman V., Yunusova A., Korablev A., Serova I., Skryabin B.V., Rozhdestvensky T.S., Battulin N. DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote // Nucleic Acids Res. - 2020. - Vol. 48, № 2. - P. 719-735.
- Smirnov A.V., Kontsevaya G.V., Feofanova N.A., Anisimova M.V., Serova I.A., Gerlinskaya L.A., Battulin N.R., Moshkin M.P., Serov O.L. Unexpected phenotypic effects of a transgene integration causing a knockout of the endogenous Contactin-5 gene in mice // Transgenic Res. - 2018. - Vol. 27, № 1. -P. 1-13.
- 3. Burkov I.A., Serova I.A., Battulin N.R., **Smirnov A.V.**, Babkin I.V., Andreeva L.E., Dvoryanchikov G.A., Serov O.L. Expression of the human granulocytemacrophage colony stimulating factor (hGM-CSF) gene under control of the 5'regulatory sequence of the goat alpha-S1-casein gene with and without a MAR

element in transgenic mice // Transgenic Res. - 2013. - Vol. 22, № 5. - P. 949-964.

Serova I.A., Dvoryanchikov G.A., Andreeva L.E., Burkov I.A., Dias L.P., Battulin N.R., Smirnov A.V., Serov O.L. A 3,387 bp 5'-flanking sequence of the goat alpha-S1-casein gene provides correct tissue-specific expression of human granulocyte colony-stimulating factor (hG-CSF) in the mammary gland of transgenic mice // Transgenic Res. - 2012. - Vol. 21, № 3. - P. 485-498.

Вклад автора. Основные результаты получены автором самостоятельно. Баркодированные плазмидные библиотеки были сконструированы и охарактеризованы под руководством м.н.с., к.б.н. А.М. Юнусовой (ИЦиГ СО РАН, Новосибирск). Процедура пронуклеарной микроинъекции выполнялась м.н.с А.Н. Кораблевым (ИЦиГ СО РАН, Новосибирск). Биоинформатический анализ данных выполнялся в.н.с., к.б.н. В.С. Фишманом (ИЦиГ СО РАН, Новосибирск).

Структура и объем работы. Диссертация состоит из введения, трех глав, выводов, списка литературы и 7 страниц приложений. Работа изложена на 117 страницах, иллюстрирована 33 рисунками и содержит 2 таблицы.

Благодарности. Автор благодарен всем!

# ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ 1.1. Конкатемеры и пронуклеарная микроинъекция 1.1.1. Основы пронуклеарной микроинъекции

В прежние времена создание и изучение различных генетических моделей ограничивалось отсутствием способа целенаправленного изменения генома. До 1974 года животных с мутантными фенотипами получали с помощью радиации или химических агентов (мутагенез), эффект которых был во многом случайным. Два близких по времени открытия имели огромные последствия для развития биотехнологии И трансгенеза В целом. В 1974 году Рудольф Яниш продемонстрировал, что микроинъекция ДНК вируса SV40 в полость бластоцисты мышей может приводить к появлению потомков, имеющих встройку генетической конструкции в геноме (Jaenisch, Mintz, 1974). Однако такие животные отличались сильным мозаицизмом, поэтому Гордон с коллегами (Gordon et al., 1980) несколькими годами позже предложили осуществлять трансформацию на стадии одноклеточного эмбриона, инъецируя модифицированные плазмиды в один из пронуклеусов. В первом эксперименте эффективность трансформации была очень низкой (2 из 187 животных), хотя и сравнима с другими способами того времени для доставки генов в клетки, а сами трансгены оказались сильно перестроенными. Тем не менее, в этой работе были заложены основы современного метода микроинъекции в пронуклеус. За короткий период с 1980 по 1985 год было получено множество трансгенных линий мышей и отработана методика (Gordon et al., 1980; Brinster et al., 1981; McKnight et al., 1983; Wagner et al., 1983; Brinster et al., 1985). Созданный в то время протокол и спустя 40 лет почти не претерпел изменений.

Для начала выбирают линии мышей, наиболее подходящие для проведения эксперимента. Обычно это гибридные самки C57BL/6 x CBA как доноры яйцеклеток, и гибридные самки C57BL/6 x DBA/2 (или CD-1) как реципиенты эмбрионов. Далее, 7-10 самок (примерно 200 яйцеклеток), у которых была вызвана суперовуляция, спариваются с самцами. Альтернативным подходом является *in* 

vitro оплодотворение яйцеклеток. Через 12 часов после оплодотворения яйцеводы извлекают из оплодотворенной самки и переносят в рабочий раствор. Ооциты вымываются из яйцеводов с помощью шприца и очищаются от кумулюсных и жировых клеток в растворе с гиалуронидазой. После этого, все собранные яйцеклетки переносятся в культуральную среду под слой масла, предохраняющего от контаминации и испарения. Под микроскопом выбирают подходящую яйцеклетку, не имеющую повреждений и без признаков начавшегося деления. Ее удерживают вакуумной пипеткой и делают инъекцию 1-2 пиколитра раствора ДНК внутрь одного из пронуклеусов (Рис. 1). Как правило, выбирают мужской пронуклеус, так как он крупнее. Для инъекции используют очищенный фильтрованный раствор ДНК (1-10 нг/мкл) в ТЕ буфере (0.01 M Tris-HCl, 0.25 мМ EDTA pH 7.4). Это соответствует нескольким сотням или тысячам копий трансгена на зиготу (в зависимости от размера конструкции). Проколотые зиготы трансплантируются псевдобеременной суррогатной матери, которую ссадили с вазэктомированным (стерильным) самцом. Через месяц после рождения все потомки анализируются с помощью ПЦР-генотипирования или Саузерн блоттинга. Обычно примерно 10-30% родившихся животных имеют встройку трансгена (Brown, Corbin, 2002).



Рисунок 1. Микроинъекция ДНК в пронуклеус оплодотворенной яйцеклетки. Заметны два пронуклеуса, в один из которых инъецируется раствор ДНК. Вокруг яйцеклетки видна *zona pellucida*. Фотография И.А. Серовой.

#### 1.1.2. Структура конкатемеров: модели конкатемеризации

При микроинъекции в ядро попадают несколько тысяч копий трансгена. До момента интеграции они объединяются в конкатемеры, содержащие до нескольких сотен копий трансгена и, в подавляющем числе ситуаций, направленных тандемно ("голова-к-хвосту"). То, что почти во всех случаях трансгенная ДНК обнаруживается в виде многократно повторенных мономеров (тандемов), встроенных в едином локусе (Nakanishi *et al.*, 2002), предсказывает наличие внехромосомных механизмов конкатемеризации.

Если бы трансгены соединялись по механизму обычного лигирования негомологичных концов (NHEJ), ориентация трансгенных копий в конкатемере была бы случайной, или, во всяком случае, отличной от традиционной "голова-кхвосту". Поэтому на первый план выходит другая гипотеза: копии трансгенов становятся участниками внехромосомной рекомбинации с помощью системы гомологичной рекомбинации (HR). Замечено, что доставленные в клетку конструкции образуют циклические молекулы и рекомбинируют между собой, в результате образуются конкатемеры из трансгенов с одинаковой ориентацией (Bishop, 1996: Smith, 2001). Альтернативной гипотезой возникновения конкатемеров может быть амплификация по "механизму катящегося кольца". У трансгенных животных иногда наблюдается низкая вариативность нуклеотидных делеций в слияниях между копиями, когда на 20 копий приходится всего 3 разных варианта слияний (Rohan et al., 1990). Существующие клеточные механизмы в теории способны проводить такую амплификацию. Например, амплификация по механизму катящегося кольца была экспериментально показана у дрожжей (Watanabe, Horiuchi, 2005; Watanabe et al., 2018). Для этого процесса, названного double rolling circle replication (DRCR), необходимо наличие повторенных участков генома, между которыми происходит взаимная рекомбинация по механизму breakinduced replication (BIR), что приводит к замыканию участка в кольцо (Watanabe, Horiuchi, 2005). Также, на концах теломер эукариот были обнаружены короткие повторенные участки (t-loops), которые могут поддерживать свою длину в отстутствие теломеразы, используя механизм катящегося кольца и стандартный набор факторов гомологичной рекомбинации (Tomaska *et al.*, 2019). Однако прямых доказательств активности подобных механизмов в зиготе нет.

Хронология изучения конкатемеров перекликается с технологическим развитием молекулярной биологии. Так, на ранних этапах для изучения трансгенных животных использовался Саузерн блот. Чувствительность этого метода невысока, но позволяет примерно оценить копийность (относительная яркость бэнда на мембране) и определить наличие перестроек и палиндромов. Подчет копийности конкатемеров показал, что она широко варьирует (от одной копии до сотен копий), а рестрикционный анализ ориентаций копий выявил доминирование ориентации "голова-к-хвосту" (более 90% всех копий в сотнях трансгенных линий) (Palmiter *et al.*, 1982; Brinster *et al.*, 1983; Khillan *et al.*, 1985; Overbeek *et al.*, 1986).

Несмотря на ограниченные технологические возможности того времени, пожалуй, самая информативная статья по механизму формирование конкатемеров вышла именно в начальный период трансгенеза. Эта работа из лаборатории будущего нобелевского лауреата Марио Капекки вышла почти 40 лет назад (Folger et al., 1982), но часто цитируется и в настоящее время. Исследователи инъецировали ядра клеток (в культуре) разным количеством молекул ДНК, как линеаризованными, так и циклическими. Количество молекул варьировало от 2 до 1000. Одним из главных выводов работы было то, что копийность встроек коррелирует с концентрацией ДНК, что свидетельствует в пользу сборки конкатемеров из отдельных молекул. Присутствие линеаризованных концов также стимулировало конкатемеризацию. Для изучения механизма сборки конкатемера своеобразными "баркодами" пометили инъецируемые молекулы они селективными маркерами (ген HSVtk) в двух ориентациях (Folger et al., 1982). Молекулы с альтернативными вариантами HSVtk (А/В варианты) давали разные сигналы на Саузерн блоте после рестрикции. Анализ конкатемеров из клеточных сублонов показал, что они состоят из двух перемешанных вариантов молекул. Это подтверждает гипотезу о сборке конкатемера из линейных копий, хотя разрешение метода не позволяет однозначно отвергнуть гипотезу амплификации (Рис. 2).



Рисунок 2. Одна из первых схем процесса конкатемеризации в пронуклеарной микроинъекции из знаковой статьи Folger *et al.*, 1982.

#### 1.1.3. Структура конкатемеров: встройка в геном

Принято выделять два самостоятельных механизма, с помощью которых трансген может инкорпорироваться в геном: это гомологичная рекомбинация (HR) с определенным участком в геноме и "незаконная", случайная встройка. Последняя происходит гораздо чаще (в 1000-10000 раз) (Smith, 2001).

Так сложилось, что все случаи интеграции экзогенной ДНК, не имеющей целенаправленно сконструированных гомологичных районов, попадают в собирательную группу "гомологичной неканонической случайной интеграции" (homologous illegitimate random integration, HIRI). Логику названия стоит объяснить поподробнее: подавляющее число встраивающейся ДНК (в том числе и трансгенов) имеет короткие гомологичные участки на обоих концах, но из-за небольшой длины гомологичного района такая ДНК встраивается в случайных и непредсказуемых местах. О причинах этого и выборе пути, по которому пойдет встраивание трансгена, рассказано в разделе о репарации ДНК.

Одно из убедительных свидетельств привел Бринстер с коллегами (Brinster *et al.*, 1989). Они пробовали скорректировать делецию в одном из генов *MHC*-комплекса, инъецируя в пронуклеус трансген, несущий большой гомологичный участок, который перекрывал район делеции. Только одна трансгенная мышь из 506 восстановила нормальный фенотип, в остальных случаях конструкции интегрировались в случайных местах.

Инкорпорация экзогенной ДНК вызывает перестройки возле сайта интеграции. Это делеции генетического материала (иногда до сотен тысяч п.о.), транслокации фрагментов хромосомной ДНК (Würtele *et al.*, 2003; Goodwin *et al.*, 2019). Нарушения могут быть весьма сложными: комбинацией из делеций и вставок внутри трансгенного тандема (Mark *et al.*, 1982). В редких случаях, трансген бывает окружен теломерными последовательностями, добавленными в его концам теломеразой (Murnane, Yu, 1993).

В начале 2010 года вышла статья по картированию бакмидного трансгена (ВАС) после пронуклеарной микроинъекции (Le Saux *et al.*, 2010). Стоит отметить,

что ВАС- и YAC-последовательности стали популярным объектом исследований, так как протяженный участок родного для трансгена окружения (в этой работе - 30-145 тыс. п.о.), способен обеспечивать высокий уровень экспрессии целевого белка в геноме хозяина. В эксперименте попытались локализовать встройку ВАСтрансгена в 11 линиях мышей, применив TAIL-PCR. Удалось им это только в трех случаях. Оказалось, что почти все трансгены были сильно перестроены: от делеций концевых участков, затрудняющих анализ слияний трансгенов и геномной ДНК, до вставок хромосомных участков внутрь последовательностей ВАС-трансгенов.

МакФарлан с сотрудником (McFarlane, Wilson, 1996) описали случай очень точной интеграции трансгена после микроинъекции. Их конструкция (как оказалось, тандем из двух копий) встроилась в участок Х-хромосомы, потеряв меньше десятка нуклеотидов с 5'-конца, а 3'-конец остался неповрежденным. При этом не было замечено никаких перестроек возле точки встройки. Интересно, что рядом с местом интеграции был обнаружен консенсусный сайт рекомбинации (т.н. горячая точка рекомбинации: Chi-сайт, human minisatellite core, murine E<sub>B</sub> hotspot). Проанализировав участки слияния трансгена и хромосомной ДНК, они нашли небольшой участок гомологии (9 из 13 нуклеотидов) на 5'-конце. Авторы считают, что такого сходства достаточно для инициации процесса гомологичной "незаконной рекомбинации".

Масштабная работа по изучению распределения сайтов встройки была проделана Наканиши с коллегами (Nakanishi *et al.*, 2002). Они изучили локализацию трансгена (*EGFP*) у 142 особей, полученных в результате пронуклеарной микроинъекции. В 82.4% случаев трансген интегрировался в единственном месте генома. Было обнаружено, что, в общем, каких-либо предпочтительных хромосомных сайтов для встройки не существует, хотя некоторые хромосомы все же служили мишенью чаще других. Вполне возможно, что "непредсказуемость" интеграции напрямую зависит от вероятности появления двухцепочечного разрыва ДНК в конкретном районе.

К сегодняшнему дню накопилось огромное количество описаний конкатемерных встроек в линиях мышей. Большой список таких работ (31 статья) можно найти в дополнительных материалах к статье (Smirnov *et al.*, 2020). К этим данным можно также добавить новые описания десятков линий, полученные

16

современными высокопроизводительными методами секвенирования (см. следующие разделы обзора).

#### 1.1.4. Особенности пронуклеарной микроинъекции: мозаицизм

Когда родившихся трансгенных животных ("фаундеров"), имеющих уникальные сайты интеграции генетической конструкции, берут в скрещивание с обычными животными для созданий отдельных линий, нередко наблюдаются отклонения от менделевского типа наследования, при котором трансген передается, словно гемизиготный или гетерозиготный аллель.

В норме, если интеграция ДНК произойдет до первого деления зиготы, все клетки взрослого животного, в том числе и половые, будут генетически модифицированными. В этом случае, половина потомков фаундера будет трансгенными. Бывает, что трансген интегрируется в несколько геномных участков на разных хромосомах, тогда процент позитивных по встройке потомков будет выше стандартного (75%, если у фаундера два независимых сайта). Половые хромосомы также иногда становятся мишенью для трансгена. Интеграция в Y-хромосому легко определяется после изучения потомства – в нем только самцы являются трансгенными. Аналогично, самцы с модифицированной X-хромосомой в помете будут давать исключительно трансгенных самок.

Сложности возникают при запоздавшей интеграции трансгена после начала эмбриона. Высокой уровень мозаицизма отражается деления В крайне маловероятном рождении трансгенного потомства от фаундера, так как популяция родителя лишь в небольшой степени половых клеток содержит трансформированные клоны. Тем не менее, родившийся потомок будет иметь встройку уже во всех тканях и станет основателем нормальной линии.

Количество мозаиков после пронуклеарной микроинъекции иногда может превышать 60%, и здесь данный метод трансгенеза уступает некоторым другим (Moreira *et al.*, 2007).

# 1.1.5. Особенности пронуклеарной микроинъекции: уровень экспрессии

В нашей и многих других работах отмечено, что экспрессия трансгена сильно варьирует между линиями и даже животными из одной трансгенной линии. Это объясняется несколькими причинами, о которых речь идет ниже.

Эффект положения гена - одна из причин огромного разнообразия уровней экспрессии целевого белка у трансгенных животных. Обсуждаемое явление впервые было замечено еще Стертевантом и Меллером в их классических работах по индукции мутагенеза у дрозофилы. Например, траслокация гена white на хромосоме приводила к изменению количества мутантных белых фасеток в глазе дрозофилы. Изменения в экспрессии гена, вызванные его перемещением в геноме, с тех пор называются эффектом положения гена (position-effect variegation, PEV). Именно у дрозофилы и, позже, у дрожжей было детально изучено влияние гетерохроматина и его модификаций на степень экспрессии гена (Girton, Johansen, 2008). Хромосомный контекст сильно влияет на экспрессию встроившейся ДНК, в частности, перицентромерные и теломерные гетерохроматиновые участки в значительной доле случаев репрессируют активность Трансгены, генов. содержащие энхансеры и подобные им в функциональном плане элементы, избегают эффекта положения, так как энхансер не только стимулирует транскрипцию, но и поддерживает открытое состояние хроматина (Ramírez et al., 2001).

Есть и другой фактор, который ведет к гетерогенности клеточной популяции – так называемая мозаичность экспрессии гена. В большинстве случаев, свойства линий, созданных одним способом, с одинаковой конструкцией и даже с интеграцией в идентичном месте, все равно отличаются друг от друга (Williams *et al.*, 2008). Более того, внутри трансгенной линии клетки особей уже со стадии нескольких бластомеров начинают различаться по экспрессии чужеродного белка. Например, эпителий молочных желез в трансгенных линиях характеризуется разной степенью мозаицизма - от полного отсутствия или очень низкого уровня экспрессии (менее 0.01% клеток) до максимально возможного уровня (>95% экспрессирующих клеток) (Krepulat *et al.*, 2005). Точные причины гетерогенности клеточной экспрессии пока не ясны. На настоящий момент к ним относят эффект

положения гена, влияние локального хроматина трансгена (Krepulat *et al.*, 2005) и импринтинг (Swain *et al.*, 1987). В некоторых задокументированных случаях, гены экспрессировались в клетках с разной эффективностью при раке (Hanash, 2004).

Вдобавок, зависеть от мозаицизм может того. ИЗ каких клетокпредшественников развился орган. Серова и сотрудники изучали клеточный мозаицизм в экспрессии трансгена, помещенного под тканеспецифичный промотор, который активировался только в секретирующих клетках молочной железы (Серова и др., 2009). В двух линиях окрашенные клетки с целевым белком были равномерно распределены по молочной железе. Напротив, в одной из линий позитивными оказали отдельные дольки (lobuli) - они были расположены среди полностью негативных долек-соседей. Каждая долька развивается из одной клеткипредшественницы, так что характер мозаицизма в ткани трансгенного животного может регулироваться и какими-то скрытыми эпигенетическими механизмами.

#### 1.1.6. Особенности пронуклеарной микроинъекции: копийность

Считается, что копийность трансгена слабо связана с уровнем его экспрессии. Линии, число копий в которых отличается в сотни раз, могут иметь сходный уровень экспрессии, и, напротив, трансгенные животные с одним экземпляром чужеродного гена сильно отличаются между собой (Ramírez *et al.*, 2001; Kwaks *et al.*, 2003; Burkov *et al.*, 2013). Тем не менее, феномен высококопийных встроек можно использовать для повышения продуктивности клеточных линий, если включать в конструкцию специальные регуляторные элементы (matrix attachment regions (MARs)) (Grandjean *et al.*, 2011; Kostyrko *et al.*, 2017; Asoshina *et al.*, 2019).

С высокой копийностью конкатемеров связан эффект индуцированного сайленсинга (repeat-induced gene silencing, RIGS). Показано, что уровень метилирования промотора растет пропорционально количеству копий, негативно влияя на экспрессию трансгенов. Справедливости ради, нужно сказать, что RIGS наблюдается как в больших конкатемерах (>100 копий), так и в более компактных (<20 копий) (Garrick *et al.*, 1998; McBurney *et al.*, 2002). Тандемные участки также менее чувствительны к ДНКазе I и несут метки неактивного хроматина (H3K9me3), вероятно, из-за гетерохроматинизации всего района (Ye, Signer, 1996; Fukuma *et al.*,

2016). Механизм сайленсинга многократно повторенных последовательностей у позвоночных пока не выяснен (Garrick *et al.*, 1998; Rosser, An, 2010; Fukuma *et al.*, 2016).

# 1.1.7. Сборка протяженных конструкций из фрагментов

Эффект конкатемеризации представляет из себя интересный инструмент для сборки протяженных трансгенов непосредственно в зиготе. Первое время что получение трансгенных животных с бакмидными считалось. (BAC) конструкциями (50-300 тыс п.о.) может быть проблематично, так как высокомолекулярная ДНК нестабильна и часто аккумулирует перестройки во время подготовки и при конкатемеризации. Через пронуклеарную микроинъекцию были созданы трансгенные линии мышей с экспрессирующимися генами альбумина человека (hSA) (фрагменты 7 + 13 + 17 тыс. п.о.) или аденозиндеаминазы человека (ADA) (фрагменты 29 + 39 тыс. п.о.) (Pieper et al., 1992; Migchielsen et al., 1996). В первом случае перекрывающиеся районы фрагментов составляли около 2-3 тыс. п.о., во втором – 19.5 тыс. п.о. Анализ Саузерн блотом подтвердил, что у большинства трансгенных потомков присутствовали правильные копии гена, хотя встречались и неполные копии или перестройки (Pieper et al., 1992). Обращает на себя внимание тот факт, что конкатемеризации между одинаковыми фрагментами не было. В схожей по замыслу работе успешно провели сборку гена рецептора липопротеина человека (VLDLR) из двух фрагментов длиной 40 и 35 тыс. п.о. с перекрыванием в 8 тыс. п.о. (Tacken et al., 2001). FISH анализ на распластанных хромосомах (Fiber-FISH) позволил визуализировать структуру конкатемеров в геноме животных. Как видно на Рис. 3, встройка из множества копий (около 50) имеет тандемную структуру. Красный и зеленый зонды гибридизуются с двумя исходными фрагментами и чередуются в конкатемере, а место перекрывания двух фрагментов дает оранжевый цвет.



Рисунок 3. Анализ конкатемеров в трансгенной линии мышей методом Fiber-FISH. Для гибридизации использованы два зонда: зеленый (первый фрагмент 40 тыс. п.о.) и красный (второй фрагмент 35 тыс. п.о.). Зона перекрывания между фрагментами имеет оранжевый цвет. На врезке указана одна полная копия. Хорошо видно тандемную структуру конкатемера (отдельные копии отмечены белыми точками). Рисунок из статьи Tacken *et al.*, 2001.

Метод сборки конструкции за счет перекрывающихся участков активно изучался в 90х годах, но оказался неактуален, так как микроинъекции бакмид в пронуклеус доказали свою эффективность (van Keuren *et al.*, 2009). Более того, даже при использовании трех бакмид они интегрируются в составе конкатемера в один локус (но с разной копийностью) (Dougherty *et al.*, 2012). Возможно, что описанный подход получит новый импульс с появлением методов синтеза длинных одноцепочечных ДНК-фрагментов (Minev *et al.*, 2019), которые меньше подвержены перестройкам и проще в приготовлении.

#### 1.2. Современные методы секвенирования встроек конкатемеров

В предыдущем разделе были описаны традиционные методы анализа конкатемерных встроек: Саузерн блот, inverse PCR, TAIL-PCR, FISH и некоторые другие. Иногда в исследованиях встречаются комбинации экзотических методов,

например, применение droplet digital PCR (ddPCR) с многочисленными зондами, покрывающими протяженные районы трансгена, в сочетании с inverse PCR. Так, к примеру, были картированы и изучены ВАС-конструкции (>100 тыс. п.о. и их фрагменты) в линиях мышей (Nakagaki et al., 2018; Nakagaki et al., 2018-2). Данные молекулярные методы очень трудоемки и дают лишь ограниченную информацию о структуре трансгенной встройки. Можно констатировать, что они не отвечают современным вызовам трансгенеза. Так, по некоторым подсчетам, лишь 433 из 8715 трансгенных линий мышей в базе informatics.jax.org имеют охарактеризованный сайт встройки (Nicholls et al., 2019). В настоящее время внимание исследователей смешается В пользу высокопроизводительных технологий секвенирования, таких как NGS (next generation sequencing) и LRS (long read sequencing, секвенирование третьего поколения), которые постепенно сравниваются с описанными выше молекулярными методами по совокупности затрат, при этом значительно опережая их в информативности и простоте.

Для понимания процесса конкатемеризации важно также изучать палиндромные варианты слиний между копиями, их частоту и стабильность в геноме животных. Стандартные молекулярные методы, основанные на ПЦР, не подходят для секвенирования палиндромов. Это. вероятно, связано с формированием шпильки в ПЦР-продукте, препятствующей амплификации, как было продемонстрировано в недавней статье Mikhailov et al., 2019. Современные методы NGS и LRS решают эту проблему (Masumura et al., 2015; Nicholls et al., 2019; Smirnov et al., 2020).

# 1.2.1. Методы, базирующиеся на NGS-анализе

Полногеномное секвенирование коротких фрагментов (NGS) малоэффективно для глубокого анализа строения конкатемера, но оно способно дать информацию о сайте встройки конкатемера в геноме. Как правило, перед использованием NGS проводят таргетное обогащение библиотеки коротких последовательностей трансгена разными методами (Kozarewa *et al.*, 2015). Для этого существует три причины: во-первых, трансгенные участки составляют очень малый процент от геномной ДНК (в районе 0.00001%); во-вторых, трансген встраивается в участок генома гемизиготно, поэтому сигнал с аллеля дикого типа может помешать правильно установить наличие делеций или перестроек; в-третьих, целевой геном может иметь участки гомологии с некоторыми элементами генетической конструкции, что затрудняет анализ данных NGS (Kozarewa *et al.*, 2015).

Как демонстрируют некоторые исследования секвенировать конкатемеры можно без обогащения, но необходимо получить высокое покрытие генома (>10х) (Zhang et al., 2012; Masumura et al., 2015; Yong et al., 2015; Yong et al., 2015-2). B одной из этих работ, сайты интеграции генетической конструкции lambda EG10 (74 тыс. п.о.) были изучены в двух линиях (мышь и крыса) с помощью секвенирования на платформе SOLiD (Masumura et al., 2015). Было обнаружено, что в одной линии встройка произошла без делеций в геноме. При этом конкатемер состоял из 40 копий, ориентированных в типичном порядке "голова-к-хвосту". В дополнение к нормальным слияниям было два палиндромных слияния и 14 перестроенных слияний. Во второй линии конкатемер состоял из 15 копий с одним палиндромом. Также стоит отметить, что авторы не смогли подтвердить наличие палиндромных слияний обычным ПЦР (Masumura et al., 2015). В другой работе, секвенировали трансгенную линии коров с бакмидной встройкой гена лактоферрина человека *hLF* (150 тыс. п.о.) (paired end Illumina HiSeq2000) (Zhang et al., 2012). Покрытие бакмидной ДНК составило 20-50х, что указывает на высокую копийность конкатемера (от 2 до 8 копий). В данных было 6 вариантов слияний, часть из которых соответствовала палиндромным слияниям. В исследованных с трансгеном Her2 (7 тыс. п.о.), который вызвал летальный нокаут по гену Pbs (кофактор когезина) за счет встройки большого конкатемера из 162 копий в интрон гена (Yong et al., 2015). В аналогичной работе с другой трансгенной линией (конструкция vavCar, 11.5 тыс п.о.) был охарактеризован протяженный конкатемер из 270 копий (paired end Illumina HiSeq2500) (Yong et al., 2015-2) Интересно, что в обеих трансгенных линиях авторы не нашли в данных NGS палиндромных или перестроенных слияний, несмотря на высокую копийность.

Альтернативой глубокому секвенированию генома может быть таргетное обогащение участков трансгена в библиотеках NGS (Рис. 4А). К примеру, в одной из работ было просеквенировано 5 трансгенных линий овец, полученных микроинъекцией конструкции 11.5 тыс. п.о. (ген человека *HTT*), и две линии

мышей с трансгеном длиной 1.9 тыс. п.о. (Chiang *et al.*, 2012). Геномные библиотеки для NGS обогащались с помощью биотинилированных зондов: PHKзонды длиной 120 п.о., покрывающие обе комплементарные нити генетической конструкции, гибридизовались с фрагментированной геномной ДНК и отбирались на магнитных частицах со стрептавидином (Рис. 4А). Оказалось, что встройки имели невысокую копийность, от 1 до 4 копий, при этом в каждой из 7 трансгенных линий наблюдались перестройки между копиями. В местах интеграции конкатемеров обнаруживались небольшие делеции и одна инверсия (80 тыс. п.о.). Эти результаты контрастировали с типичной ориентацией копий в конкатемерах, что может говорить о существовании нескольких альтернативных механизмов конкатемеризации.

В другой работе в пронуклеус был инъецирован циклический ВАС со специфическим аллельным вариантом гена *LMNA* человека (Dubose *et al.*, 2013). Размер конструкции составлял более 170 тыс. п.о., длина кодирующей части гена *LMNA* – около 25 тыс. п.о. При инъекции таких протяженных конструкций необходимо секвенировать встройку трансгена с высоким покрытием, чтобы увидеть возможные перестройки, поэтому авторы воспользовались технологией гибридизации на чипе с зондами, комплементарными бакмидной ДНК (Puc. 4A). Обогащение библиотек этим методом увеличило число целевых ридов в данных NGS в 3000-26000 раз. Оказалось, что конкатемер состоит из 5 фрагментированных бакмид, из которых только две сохранили нужный участок гена. Интересно, что два района из ВАС (2.1 и 31.7 тыс. п.о.) не встречались ни в одной из ВАС-копий. Возможно, это было связано с неравномерной гибридизаций данных участков бакмиды на чипе.

Интересным способом таргетного обогащения трансгенных библиотек для NGS является захват целевых участков генома с применением системы CRIPSR/Cas9 (Slesarev *et al.*, 2019). В этом методе геномная ДНК фрагментируется на длинные участки ~20 тыс. п.о., которые связываются dCas9 (Cas9 без нуклеазной активности) через гайдовые РНК (гРНК). Сайты узнавания гРНК должны быть распределены по последовательности трансгена через каждые 2-3 тыс. п.о. Отбор целевых фрагментов осуществляется при помощи биотина, который вносится либо в гРНК, либо в белок dCas9. К сожалению, данный метод также достаточно

24

трудоемок и не дает высокого обогащения в сравнении с другими описанными методами (Slesarev *et al.*, 2019).

Недавно был предложен новый метод, названный "таргетная амплификация локуса" (targeted locus amplification, TLA), который направлен на коммерциализацию и удешевление процедуры локализации трансгенов (Cain-Hom *et al.*, 2017).

# 1.2.2. Таргетная амплификация локуса (TLA)

Метод TLA был разработан в лаборатории Воутера де Лаата, известного специалиста по 3С (chromosome conformation capture) (de Vree et al., 2014). TLA, как и 3С-методы, основан на принципе лигирования сближенных в пространстве молекул ДНК (Баттулин и др., 2012) (Рис. 4В). На первой стадии геномную ДНК в клетках фиксируют формальдегидом для создания сшивок белок-ДНК (crosslinking). В результате, районы ДНК, которые располагаются рядом в ядре, включая и конкатемер с фланкирующими геномными участками, оказываются фиксированы вместе. Затем фиксированную ДНК обрабатывают рестриктазами, лигируют сближенные в пространстве концы, получая кольцевые молекулы, и очищают ДНК (reverse cross-linking). На финальной стадии проводится inverse PCR с праймерами на трансген и амплификация лигированных фрагментов (Рис. 4B). TLA позволяет повысить число ридов рядом с районом, на который были выбраны праймеры, при этом обогащение эффективнее других методов, требующих дорогостоящих китов (xGen (IDT); SureSelect (Agilent); и другие) (Slesarev et al., 2019). В ТLА для эффективной амплификации трансген-геномных границ не обязательно знать строение всей конструкции или последовательность линеаризованных перед микроинъекцией концов, достаточно использовать пару праймеров на известную часть трансгена. Это значительно облегчает геномный анализ для "архивных" трансгенных линий, полученных с плохо описанными генетическими конструкциями (Cain-Hom et al., 2017).

Впервые эффективность этого метода была продемонстрирована на 7 трансгенных линиях мышей со встройками популярных Сге-рекомбиназ под тканеспецифичными промоторами ("Сге-делитеры") (Cain-Hom *et al.*, 2017). Специфичность экспрессии – важный фактор для Сге-делитеров, поэтому локализация встройки важна для объяснения особенностей экспрессии трансгена. Авторы работы не знали полную фланкирующую последовательность генетических конструкций и использовали праймеры на ген Сге-рекомбиназы. Копийность, определенная по числу ридов, была не высока (1-16 копий). Из 7 линий, лишь в одной трансген был локализована в интроне гена. В трех случаях из семи, интеграция конкатемера привела к дупликации фланкирующего участка генома (Cain-Hom *et al.*, 2017).



Рисунок 4. Современные методы секвенирования встроек конкатемеров. (А) Методы таргетного обогащения библиотек для NGS. Для примера показано два метода обогащения: биотинилированные зонды или гибридизация на чипах с комплементарными зондами. (В) Схема метода TLA (подробности в тексте) (de Vree *et al.*, 2014). (С) Принцип секвенирования Oxford Nanopore (Deamer *et al.*, 2016). (D) Принцип работы секвенирования PacBio (Ardui *et al.*, 2018).

Еще более впечатляющей демонстрацией возможностей TLA стала систематическая локализация трансгенных встроей в мышах из вивария JAX Repository (Goodwin *et al.*, 2019). Авторы описали трансгенные локусы для 40 трансгенных линий, выбрав линии с Сге-трансгенами и моделями болезней Паркинсона и Альцгеймера. Тип и размер конструкций варьировали от небольших трансгенов и лентивирусных встроек до ВАС-конструкций и космид. Выборка из сорока линий позволяет оценить аспекты интеграции трансгена, для которых ранее не хватало статистических данных.

В 30 линиях из 40 интеграции трансгена спровоцировали перестройки, в основном, делеции (размером от <100 п.о. до >500 тыс. п.о.) (Goodwin *et al.*, 2019). В 21 линии из 40 авторы обнаружили интеграцию трансгена в кодирующий район генома, которая затрагивала экзоны или локализовалась в интрон. Очевидно, что эти данные придется учитывать при использовании описанных трансгенных моделей. В целом, эти результаты подтвердили предыдущие догадки о повышенной частоте встроек в кодирующие участки генома (Burkov et al., 2013). С другой стороны, частая интеграция в транскрипционно-активные участки генома может быть следствием отбора животных на высокий уровень экспрессии трансгена. Ранее в обзоре литературы обсуждалась работа по FISH-картированию встроек трансгена EGFP в 142 линиях мышей, полученных методом пронуклеарной микроинъекции (Nakanishi et al., 2002). Исходя из этих данных, трансгены чаще встраиваются в районы, бедные генами (Nakanishi et al., 2002). Другие исследования, использовавшие TLA-картирование, также не увидели высокой частоты встройки внутрь кодирующих участков (Cain-Hom et al., 2017; Laboulaye et al., 2018). Так что вопрос о предпочтительных сайтах интеграции трансгенов в пронуклеарной микроинъекции остается открытым.

Авторами также было обнаружено 6 случаев дупликации в районе встройки трансгена (Goodwin *et al.*, 2019). Любопытно, что и первой работе (Cain-Hom *et al.*, 2016) были описаны значительные дупликации фланкирующих геномных участков (в 3 линиях из 7, диапазон от 1.5 до 87.6 тыс. п.о.). Это явление было замечено и ранее, и объяснялось интеграцией трансгена в репликативную вилку, когда часть генома уже удвоена (Wilkie, Palmiter, 1987). По-видимому, дупликации участков в

районе интеграции также являются специфической чертой пронуклеарной микроинъекции, но конкретный механизм этого процесса не изучен.

Таким образом, TLA является относительно простым и эффективным методом картирования трансгенов, особенно для линий, где информация о внутреннем строении конкатемера не представляет интерес. Для изучения же конкатемерных встроек лучше подходят методы секвенирования третьего поколения (LRS).

# 1.2.3. Секвенирование третьего поколения (LRS)

Методы секвенирования третьего поколения, основанные на получении длинных ридов, стали доступны в 2011-2014 годах и значительно облегчили анализ геномов (van Dijk *et al.*, 2018). В настоящий момент их применяют для сборки сложных бактериальных геномов и метагеномов (De Maio *et al.*, 2019; Sevim *et al.*, 2019), анализа транскриптомов (Weirather *et al.*, 2017) и генетики человека (секвенирование повторов, аллельных вариантов, псевдогенов) (Ardui *et al.*, 2018; Mantere *et al.*, 2019).

Среди коммерчески доступных методов LRS лидерство делят два альтернативных подхода: Oxford Nanopore (ON) и Pacific Biosciences SMRT (PacBio). Идеи об использовании мембран с белковыми нанопорами для секвенирования ДНК высказывались давно (Kasianowicz et al., 1996; Deamer, 2010), но первые приборы фирмы Oxford Nanopore, основанные на этом принципе, стали доступны всего 5 лет назад (см. историю метода у Mantere et al., 2019). Принцип секвенирования основан на использовании мембраны с включенной в нее нанопорой, к которой присоединен белок (полимераза, геликаза), осуществляющий связывание молекулы ДНК или РНК для подачи в пору. Мембрана разделяет камеру на две части. К частям камеры прикладывают напряжение, из-за чего возникает ток ионов через пору. При прохождении одноцепочечной молекулы ДНК через пору, ток ионов характерно замедляется (в зависимости от типа основания) и создает детектируемое изменение силы тока (Рис. 4С). Скорость "чтения" составляет примерно 30 оснований в секунду (Deamer *et al.*, 2016). Частота ошибок около 15%, в особенности при секвенировании гомополимерных участков (Mantere et al., 2019).

Секвенирование методом РасВіо стало коммерчески доступным в 2011 и в настоящий момент активно развивается (Mantere et al., 2019). Технология PacBio SMRT (Single Molecule, Real-Time) основана на микроскопических ячейках, каждая одну молекулу ДНК-полимеразы. Секвенирование ИЗ которых содержит происходит параллельно в тысячах ячеек (от 150000 до 1 миллиона, в зависимости от прибора). Для приготовления библиотек необходимо большое количество высокомолекулярной ДНК (>5 мкг). Фрагментированная ДНК лигируется с одноцепочечными адаптерами-шпильками, которые замыкают фрагмент ДНК в кольцо (Рис. 4D). Когда специально подготовленная циклическая молекула ДНК с адаптерами попадает в ячейку, полимераза связывается с одноцепочечным районом адаптере и начинает синтезировать комплементарную цепь, используя в флуоресцентные нуклеотиды. Включение нуклеотида в цепь вызывает всплеск сигнала, который измеряет лучом лазера, направленным на дно ячейки (Рис. 4D). Ошибки секвенирования составляют порядка 13-20%, но, в отличие от Oxford Nanopore, носят абсолютно случайный характер, поэтому их можно избежать, повысив глубину секвенирования (Weirather et al., 2017). К тому же, благодаря циклической форме секвенируемого фрагмента, ДНК-полимераза в ячейке может прочитать его несколько раз (до 10). Эти "сабриды" объединяются в консенсус (circular consensus sequence (CCS)), что в итоге дает очень высокую точность (>99%) (Ardui et al., 2018). Средняя длина ридов в методах Oxford Nanopore и РасВіо составляет от 10 до 100 тыс. п.о., в зависимости от прибора и условий секвенирования (Jain et al., 2018; van Dijk et al., 2018).

В связи с новизной LRS-методов пока опубликована лишь одна статья, посвященная секвенированию конкатемеров в трансгенной линии мышей методом Oxford Nanopore (Nicholls *et al.*, 2019). В этой публикации авторы провели секвенирование генома популярной линии мышей *Oct4:EGFP*, в которой структура и место встройки трансгена до сих пор были не охарактеризованы (Szabó *et al.* 2002). Ими было получено 32 рида, содержащего последовательности *EGFP*. Медианная длина ридов составила 28297 п.о. при двухкратном покрытии генома. В статье было описано место интеграции трансгена: участок на хромосоме 9, встройка в который привела к делеции 686 п.о. Копийность встройки была оценена косвенно, по количеству ридов, содержащих трансген (18 копий), а также

альтернативным способом, qPCR с праймерами на *EGFP* (26 копий). В работе не удалось определить полную внутреннюю структуру конкатемера из-за его повторенной природы. Тем не менее, авторы описали две перестройки: палиндромное слияние между копиями и встройку фрагмента генома *E. coli* (6.2 тыс. п.о.), который также лежит в палиндромном слиянии копий. Этот результат показывает, что детальный анализ конкатемеров без применения баркодов пока еще невозможен, несмотря на огромную длину ридов Oxford Nanopore.

Методы LRS имеют высокую стоимость, поэтому для них также разрабатываются методы таргетного обогащения участков трансгена. Обычно через биотинилированные зонды можно добиться обогащения в  $10^3$ - $10^5$  раз, как и для библиотек NGS (Li *et al.*, 2019). Альтернативные способы обогащения для Oxford Nanopore и PacBio строятся на разрезании целевых участков ДНК нуклеазой CRISPR/Cas9 и лигировании адаптеров для таргетной селекции (Gabrieli *et al.*, 2018; Hafford-Tear *et al.*, 2019; Watson *et al.*, 2020). В скором времени, когда стоимость технологий снизится, LRS станут основными инструментами изучения трансгенных животных.

# 1.3. Механизмы репарации ДНК и конкатемеризация

Встройка экзогенной ДНК и конкатемеризация происходят во время репарации двухцепочечных разрывов (DSB), когда клетка воспринимает концы трансгена как альтернативные DSB. Каждый конкретный случай интеграции нужно рассматривать отдельно, а для этого важно понимать, какие механизмы восстановления целостности ДНК представлены в клетке. По имеющимся расчетам, в клетке обычно возникает от 10 до 50 DSB каждый день, в зависимости от клеточного типа и конкретной фазы клеточного цикла (Vilenchik and Knudson, 2003), причем большая часть этих DSB возникает за счет нарушений репликации и во время транскрипции (Ranjha *et al.*, 2018). Существует два независимых способа, с помощью которых может быть восстановлена целостность генома.

## **1.3.1.** Негомологичное соединение концов (non-homologous end-joining, NHEJ)

30

NHEJ является основным путем репарации DSB в клетке. По сравнению с гомологичной рекомбинацией, описанной ниже, этот путь не требует сходства последовательностей на концах разрыва, работает во всех фазах клеточного цикла и опирается на относительно небольшое число консервативных факторов (Yano et al., 2009; Chang et al., 2017). NHEJ может соединять любые типы концов: тупые, комплементарные выступающие 3'- и 5'-концы, некомплементарные выступающие концы. В последнем случае, негомологичное соединение концов зачастую приводит к потере концевых участков ДНК или добавлению неправильных нуклеотидов (Chang et al., 2017; Schimmel et al., 2018). Помимо поддержания NHEJ целостности ДНК, также участвует В рекомбинации генов иммуноглобулинов (V(D)J-рекомбинация) (Rooney et al., 2004).

Белковый NHEJ каталитической комплекс представлен Ки-белком, субъединицей DNA-PKcs для ДНК-зависимой протеин-киназы DNA-PK, лигазным комплексом XRCC4-LigIV и белком XLF. Ки – гетеродимер, состоящий из двух субъединиц (Ku70+Ku80), и имеющий симметричную тороидальную структуру. Он окружает ДНК-дуплекс в месте разрыва, стерически взаимодействует с малой и большой бороздками ДНК и стабилизирует DSB (Walker et al., 2001; Pannunzio et al., 2018). Этот белок, без преувеличения, является главным агентом NHEJ. Будучи одним из самых распространенных белков в ядре (помимо гистонов) и обладая очень высоким сродством к DSB, Ки способен эффективно и быстро связываться с поврежденной ДНК, инициируя начало репарационного процесса (Blier et al., 1993; Grundy et al., 2016). После связывания Ки с DSB, он служит остовом для остальных белков NHEJ-комплекса (в клетках без Ки сборки комплекса не происходит) (Downs et al., 2004; Williams et al., 2014). Наблюдения за сборкой белков на ДНК после индукции DSB лазером показывают, что Ки начинает накапливаться возле поврежденного участка уже через несколько секунд и остается там на протяжении двух часов, пока идет репарация (Mari et al., 2006).

Белок Ки является катализатором сборки репарационного комплекса. Остальные факторы собираются на нем независимо друг от друга и дополнительно стабилизируют соседей за счет белковых взаимодействий и фосфорилирования. DNA-PKcs связывается с Ки и ДНК (ДНК-зависимое взаимодействие) и образует полноценный фермент. Далее активная DNA-PK фосфорилирует белки NHEJ-

31

комплекса: стабилизирует XRCC4, XLF и себя по сайтам автофосфорилирования (Yu *et al.*, 2008). Комплекс XRCC4-LigIV, который соединяет концы разрыва, тоже рекрутируется к месту DSB Ku-белком. Раньше считалось, что для посадки XRCC4 нужна и DNA-PKcs (Drouet *et al.*, 2005; Jiang *et al.*, 2015), однако сейчас ее роль сводят лишь к стабилизации XRCC4 в NHEJ-комплексе. Последний фактор, недавно описанный гомодимер XLF (или Cernunnos), как и протеин-киназа, связывается с Ku ДHK-зависимым образом. XLF тесно взаимодействует с XRCC4-LigIV и стимулирует лигазную активность этого комплекса (Hentges *et al.*, 2006).

Двухцепочечные тупые концы в месте разрыва лигируются сразу же ДНКлигазой, входящей в комплекс NHEJ. Одноцепочечные некомплементарные концы достраиваются до тупых ДНК полимеразой, или процессируются клеточными нуклеазами (Рис. 5).



ДНК-полимераза заделывает брешь

Рисунок 5. Разные типы концов (А - С) и их лигирование в NHEJ (Pastwa, Błasiak, 2003).

#### 1.3.2. Гомологичная рекомбинация (HR)

Механизм гомологичной рекомбинации позволяет очень точно восстанавливать поврежденный участок с помощью гомологичных последовательностей ДНК. Гомологичная рекомбинация также используется поддержания длины теломер (в отсутствие теломеразы) клеткой для И мейотического кроссинговера, обеспечивающего рекомбинацию генетического материала и правильное расхождение хромосом во время деления (Hedges, Deininger, 2007; Ranjha et al., 2018).

Отличительная черта этого пути репарации – резекция (resection) свободных концов ДНК, которые образовались после DSB, в 5'- 3' направлении. В ходе этого процесса одна из нитей деградирует, и образуется одноцепочечная ДНК, которая служит субстратом для связывания факторов, инициирующих рекомбинацию. Резекция – крайне сложный процесс, и именно она определяет, какой путь изберет клетка при репарации разрыва (Cejka, 2015). Чтобы не возникло путаницы с названиями ортологичных белков (исследования HR ведутся на дрожжах и клетках млекопитающих), основные участники резекции указаны через черту: дрожжевой белок и его ортолог у млекопитающих.

Главный компонент резекции – MRX/MRN-белковый комплекс, он состоит из факторов Mre11, Rad50 и Xrs2/Nbs1 (Ranjha *et al.*, 2018). Этот комплекс первым из белков HR связывается с DSB и начинает экзо- и эндонуклеазное расщепление концов, его компоненты также осуществляют взаимодействие с белками чекпойнтов, сообщая клетке о имеющемся разрыве. Второй фактор, Sae2/CtIP – эндонуклеаза со слабой активностью, которая кооперирует с MRX/MRN для начального расщепления концов (Sartori *et al.*, 2007; Wang *et al.*, 2012). Резекция продолжительных участков производится двумя экзонуклеазами: Exo1 и Sgs1/BLM (Osman *et al.*, 2016; Ranjha *et al.*, 2018).

Выступающий после резекции 3'-конец покрывается белками RPA (replication protein A) и затем Rad51, формируя 3'-филамент (nucleoprotein filament, NPF) (Taylor *et al.*, 2015). У млекопитающих, замена RPA на Rad51 - один из ключевых

этапов HR, на котором блокируются потенциально мутагенные RPA-зависимые пути SSA и MMEJ (см. Рис.6). В этом процессе участвует несколько важных факторов HR (основные - Rad52, Brca2, Palb1) (Kan et al., 2017; Sullivan, Bernstein, 2018). З'-филамент сканирует геном в поисках гомологичных донорных участков (например, сестринскую хроматиду), и по этим шаблонам восстанавливает (Son, целостность поврежденного района Hasty, 2019). При инвазии двухцепочечных гомологичных районов образуется D-петля (D-loop) - особая структура, характерная для HR (Рис. 6, стадия "поиск гомологов"). В D-петле 3'-ОН конец филамента, связанный комплементарно с донорной молекулой, служит затравкой для синтеза ДНК специализированными полимеразами б и є (McVey et al., 2016).

Репарация ДНК после формирования D-петли может происходить по-разному - сегодня известно множество альтернативных механизмов HR (обзор в Ranjha *et al.*, 2018) (Рис. 6).

#### Классическая penapaция DSB (DSB repair, DSBR)

Классическая модель, которая была сформулирована Шостаком, так описывает этот механизм. Оба резектированных 3'-конца одновременно атакуют гомологичные районы на другой хромосоме и начинается синтез ДНК по шаблонам используемой хроматиды или гомологичной хромосомы. Во время рекомбинации образуются двойные структуры Холидея (dHJ). В зависимости от способа их разрешения может происходить кроссинговер (как, например, в мейозе) или простая рекомбинация (Szostak et al., 1983).



Рисунок 6. Различные механизмы репарации двухцепочечных разрывов. Красным цветом показаны гомологичные ДНК-доноры, голубым – синтезированные участки ДНК (Huertas, 2010).

#### Отжиг синтезируемых цепей (synthesis-dependent strand annealing, SDSA)

SDSA считается важным путем репарации DSB в клетке, хотя его вклад в HR совместно с DSBR варьирует в зависимости от условий (McMahill *et al.*, 2007; Bzimek *et al.*, 2010; Zapotoczny, Sekelsky, 2017). В отличие от DSBR, лишь один из 3'-концов атакует гомологичный ДНК-дуплекс и достраивает поврежденный участок. Второй резектированный конец восстанавливается с использованием вновь синтезированной цепи (Рис. 6) (Ranjha *et al.*, 2018). Для SDSA необходимо быстрое исчезновение D-петли (D-loop dissolution), так как эта структура может спровоцировать захват второй цепи, что приведет к DSBR и потенциальному кроссинговеру (Daley *et al.*, 2013).

# Penликация, вызванная разрывом (break-induced replication, BIR)

В случае, если присутствует только один поврежденный двухцепочечный конец ДНК (к примеру, из-за деградации теломеры или коллапса репликативной

35

вилки), подвергшийся резекции 3'-конец внедряется в гомологичный район на другой хромосоме и восстанавливает всю утраченную последовательность до конца нити (Llorente *et al.*, 2008; Verma, Greenberg, 2016).

# Отжиг одноцепочечных концов (single-strand annealing, SSA)

Если разрыв случился между двумя гомологичными участками, резектированные одноцепочечные 3'-концы могут отжечься в этих местах и стянуть два конца разрыва. При этом последовательность ДНК, находившаяся между районами гомологии, будет удалена нуклеазами, что приведет к потере фрагмента (Ramakrishnan et al., 2018). Этот способ появился в арсенале высших эукариот, как упрощенная альтернатива другим путям для поддержания целостности генома, содержащего огромное число повторенных элементов источников гомологичных участков. Тем не менее, он зачастую приводит к потере важных последовательностей ДНК или интеграции экзогенных фрагментов, у которых случайно обнаружился гомологичный район (McVey, Lee, 2008).

# 1.3.3. Взаимодействие NHEJ и HR

NHEJ и HR конкурируют друг с другом за одну мишень (DSB). Ясно, что белок Ки из-за своей доступности легко опережает факторы Rad52 и MRX/MRN, имеющие ключевые функции для HR. Ки не мешает их доступу, но замедляет резекцию и процессинг концов. Защитники «мирной» гипотезы кооперации двух путей предполагают, что NHEJ – попытка быстрой и грубой репарации DSB, а в случае неудачи, за дело принимается точная и основательная HR-система. Такие выводы строятся на кинетике связывания факторов репарации с DSB: Ки-белок рекрутируется через несколько секунд, но находится в точке разрыва недолго (несколько часов), уступая свое место более стабильным MRX/MRN и Rad51 (Kim *et al.*, 2005; Biehs *et al.*, 2017). В частности, этим и объясняется более высокая частота репарации разрывов с помощью NHEJ. В отсутствие белков NHEJ-комплекса (особенно Ku) наблюдается увеличение частоты резекции и HR (Pierce *et al.*, 2001; Zelensky *et al.*, 2017). Напротив, повышенная экспрессия Ku ингибирует резекцию. Известны случаи тонкой регуляции двух путей, которая может
эксплуатироваться клеткой. К примеру, гистоны H2AX, служащие меткой DSB, подвергаются полиубиквитинилированию убиквитин-лигазой RNF8 - так в точке разрыва создается конформация хроматина, благоприятствующая посадке HRфакторов (Rad9/53BP1 и BRCA1) (Mailand *et al.*, 2007; Ochs *et al.*, 2016). Наоборот, факторы ремоделлинга хроматина ACF1 и SNF2H, служащие одними из самих ранних сигналов DSB, взаимодействуют с Ku и привлекают его к месту разрыва (Lan *et al.*, 2010). В другой работе показано, что полиубиквитинилирование уже самого Ku-белка ДHK лежит в основе его диссоциации с DSB (Postow *et al.*, 2008; Ismail *et al.*, 2015).

Второй фактор, оказывающий влияние на выбор механизма репарации, стадия клеточного цикла (Takata et al., 1998; Orthwein et al., 2015; Biehs et al., 2017). Логично предположить, что HR наиболее активна в S/G2-фазах, когда есть источник идентичной ДНК в виде сестринской хроматиды. Они используются гораздо чаще, чем гомологичные хромосомы, потому что расположены поблизости (Johnson, Jasin, 2000). Репарация с помощью гомологичной хромосомы может приводить к генной конверсии и потере гетерозиготности (loss of heterozygosity, LOH). Сейчас ясно, что деятельность большинства факторов резекции регулируется циклин-зависимыми киназами (cyclin-dependent kinase, CDK), работающими именно на этих стадиях клеточного цикла. Во-первых, CDK модифицируют Rad9/53BP1 (большой хроматин-ассоциированный белок, мешающий процессингу концов ДНК) так, что он перестает блокировать резекцию (Lazzaro et al., 2008). Во-вторых, CDK активируют Sae2 и CtIP, фосфорилируя их по Ser267 и Thr847, соответственно. Мутация этих аминокислот затрудняет резекцию, что отражается в увеличенном числе хромосомных перестроек (Sartori et al., 2007; Huertas, Jackson, 2009). Наконец, CDK-зависимое фосфорилирование CtIP по другому сайту (Ser327) улучшает связывание данного белка с BRCA1, который доставляет CtIP к месту двухцепочечного разрыва (Yu et al., 2004; Chen et al., 2008). Замечено, что резекция в S-фазе протекает быстрее, чем в G1 и даже G2. Возможно, что репликативный комплекс сам может рекрутировать факторы резекции через взаимодействие белка репликации PCNA с мотивом на CtIP (Chinnadurai, 2009), т.к. нарушение этого процесса вызывает повреждение ДНК, дестабилизацию вилки репликации и клеточный арест в S/G2 (Gu et al., 2009).

#### 1.3.4. Стабильность палиндромов в геноме и генетических конструкциях

Палиндромные последовательности – это симметричные участки генома, в которых последовательности две идентичных ориентированы зеркально друга (Рис. 7А). Палиндромы относительно друг иногда встречаются В трансгенных конкатемерах, хотя систематически их частота не оценивалась (Masumura et al., 2015; Nicholls et al., 2019; Smirnov et al., 2020). После интеграции в геном конкатемеры стабильно наследуются потомками трансгенного животногофаундера, но судьба палиндромных слияний между копиями, которые встречаются в конкатемерах, почти не изучена. В геномах высших организмов палиндромы достаточно распространены и встречаются как виде коротких микросаттелитных повторов, вроде GAA/TTC (связан с атаксией Фридрейха) (Ditch *et al.*, 2009), так и в длинных палиндромных повторах, к примеру, Alu (300 п.о.) или LINE-элементах (6 тыс. п.о.) (Ramakrishanan et al., 2018); протяженных MSY-палиндромах на Yхромосоме (Trombetta, Cruciani, 2017); а также в *de novo* перестроенных районах раковых клеток (Tanaka et al., 2005; Hasty, Montagna, 2014).

Известно, что присутствие палиндромов в геноме провоцирует перестройки, но общего консенсуса по механизму их возникновения нет. Согласно самой популярной модели, репликация и транскрипция участков генома, индуцируют появление одноцепочечных районов ДНК, что приводит к образование особых структур (крестоформов, cruciform) (Рис. 7А) или шпилек (Voineagu et al., 2008; Wang, Vasquez, 2017). Напротив, некоторые эксперименты демонстрируют, что разрушение палиндромов происходит даже в в пост-митотических нейронах (Ditch *et al.*, 2009), и что добавление в репортерный вектор ориджина репликации SV40 не повышает частоту разрывов в палиндроме (Holkers et al., 2012). В любом случае, образование крестоформов И шпилек В палиндромах было показано экспериментально (Holkers et al., 2012; Inagaki et al., 2013) (Рис. 8А). Эта двухмерная структура затем узнается и разрезается некоторыми эндонуклеазами клетки, превращаясь в DSB (Рис. 7А). Возможными кандидатами на эту роль являются многие факторы репарации (Brázda et al., 2016), включая эндонуклеазы Mre11 (Lobachev *et al.*, 2002), GEN1 + Artemis (Rass *et al.*, 2010), ERCC4 (Xpf) (Kirschner, Melton, 2010) или эндонуклеазный комплекс MutL (Halabi *et al.*, 2018).

Длительные наблюдения за палиндромным слиянием в трансгенной линии мышей помогли лучше оценить стабильность этой структуры. В лаборатории Марии Ясин (Maria Jasin) проводились эксперименты по получению мышей с репортерной конструкцией из двух последовательных копий гена LacZ (репортер рекомбинации) (Zhou et al., 2001). В одной из линии трансгены сформировали палиндром, состоящий из двух зеркальных копий общим размером 15.4 тыс. п.о. (Рис. 7В). Данная линия мышей, получившая названия "Линия 78" представляет из себя редкий случай "идеального" палиндрома без делеций между копиями и стабильно интегрированного в геном (Akgün et al., 1997; Zhou et al., 2001; Cunningham et al., 2003). Создатели линии изучали передачу трансгена по наследству и задокументировали высокую частоту простых делеций и более сложных продуктов рекомбинации в точке слияния палиндрома у потомков (около 35%) (Akgün et al., 1997; Zhou et al., 2001). Из мышей линии 78 была получена иммортализованная культура клеток (Cunningham et al., 2003). Длительное культивирование клеток с палиндромной конструкцией показало, что усредненная частота перестроек внутри палиндрома составила 5.5x10<sup>-3</sup>, что составляет примерно 0.5% клеток за один пассаж. Большая часть перестроек представляла из себя небольшие делеции в палиндромном слиянии, которые нарушали симметрию (из 118 проанализированных перестроек -93% были делециями) (Рис. 7В). Разрушение центральной симметрии повышало стабильность палиндрома в субклоне в 25 раз (Cunningham et al., 2003). В целом, частота и структура перестроек внутри палиндрома в мышах и клетках были очень похожи.

Еще один хорошо охарактеризованный палиндромный район встречается в геноме человека и ассоциирован с частой реципрокной транслокацией участков хромосом 11 и 22 (Inagaki *et al.*, 2013) (Рис. 8А). Это АТ-богатый палиндромный повтор (PATRR), который имеет 200-400 п.о. в длину и 90% АТ-состав. Такие повторы разбросаны по разным хромосомам, включая хромосомы 11 и 22, и имеют внутреннюю симметрию, но не гомологичны другим PATRR (Inagaki *et al.*, 2016). Транслокации между PATRR спонтанно возникают при сперматогенезе (частота около 1:10000 у мужчин) (Inagaki *et al.*, 2016).

Рекомбинацию палиндромных слияний также изучают на специальных репортерных конструкциях (Ramakrishnan *et al.*, 2018). Авторы получили встройки в геноме дрожжей с двумя типами палиндромных слияний: с большим (1 тыс п.о.) или коротким (12 п.о.) спейсером. Спейсер – это уникальный участок ДНК, разделяющий два инвертированных повтора (палиндром). Сами палиндромные районы имели размер около 2 тыс. п.о (Ramakrishnan et al., 2018). DSB рядом с палиндромной конструкцией инициирует резекцию и появление одноцепочечных районов ДНК, образующих шпильки (Рис. 8В) (Chen et al., 2013). В нашем конкатемерном эксперименте похожие структуры могут появляться при копировании палиндромного слияния 3'-резектированным концом молекул (Smirnov et al., 2020).



Рисунок 7. Рекомбинация палиндромов в геноме и возникновение перестроек. (А) Общая схема появления крестоформов и их превращение в DSB. Причины локальной денатурации могут быть разные: торсионный стресс, репликация, транскрипция, близкий DSB. Рисунок из статьи Cunningham *et al.*, 2003. (В) Строение палиндромного слияния трансгенов в линии мышей 78 (Zhou *et al.*, 2001). Сверху представлена общая схема идеального слияния. Каждый из мономеров содержит две последовательные копии *LacZ* (черные прямоугольники). Ниже изображены

основные пути репарации DSB в палиндроме (93% случаев), которые обычно приводят к асимметричным делециям (Cunningham *et al.*, 2003).

В другой работе использовали эписомную репортерную конструкцию, в которой ген *GFP* был прерван встройкой палиндромного участка (40 + 40 п.о.) Эксперименты велись на нескольких типах клеток (HeLa, HEK293T и COS-7) (Holkers *et al.*, 2012). Позитивные события рекомбинации (разрвы в палиндроме) детектировались, как результат восстановления гена *GFP* с помощью донорной плазмиды (полноценный ген *GFP* без старт-кодона). В качестве контроля использовали мегануклеазу I-SceI и аналогичную репортерную плазмиду, где вместо палиндромного участка был вставлен сайт для I-SceI. Было обнаружено, что наличие палиндромного участка в плазмиде провоцирует DSB с частотой, сравнимой с таргетным разрезанием нуклеазой (1-2% клеток). При этом вставка небольшого 35bp-спейсера между инвертированными копиями резко снижала частоту разрывов до фонового уровня (Holkers *et al.*, 2012). В схожем исследовании с палиндромными *Alu*-повторами были показаны аналогичные результаты, и оценен размер спейсера, который блокиирует перестройки (52 п.о.) (Voineagu *et al.*, 2008).



Рисунок 8. Роль крестоформов и обратных петель в разрушении палиндромов. (А) Атомная силовая микроскопия фрагментов ДНК, содержащих АТ-богатый повтор (PATRR) (*Kurahashi et al.*, 2004). Размер молекулы ДНК – 1 тыс. п.о., размер палиндрома – 204 п.о.. Снизу: Схема возникновения перестройки t(11,22) в клетках человека. Два района PATRR на разных хромосомах имеют палиндромную природу. Эндонуклеаза GEN1 узнает крестофромы, которые появляются в AT-богатых участках генома и разрезает эти структуры. Второй разрыв вносится нуклеазой Artemis, чтобы вскрыть шпильки и сформировать DSB. Эти хромосомные концы лигируются друг с другом через NHEJ (Inagaki *et al.*, 2013), вызывая транслокацию. (С) Резекция может провоцировать рекомбинацию в палиндроме (Ramakrishan *et al.*, 2018). В случае межхромосомной рекомбинации появляется дицентрическая хромосома, которая теряется в анафазе. Во втором случае, обратная петля (folding back) процессируется эндонуклеазами и превращается в DSB (не показано), формируя односторонний разрыв, который лигируется с подходящим районом, но уже с потерей палиндромного участка.

Таким образом, палиндромные слияния в конкатемерах гипотетически могут удаляться, как на стадии резекции концов во время экстрахромосомной рекомбинации, так и в ходе клеточных делений, после интеграции конкатемера в геном.

## Заключение: процесс конкатемеризации и репарация DSB

Рекомбинация трансгенной ДНК в зиготе идет в уникальных условиях, которые накладывают отпечаток на процессы репарации. В клетках раннего

эмбриона клеточный цикл сдвинут в сторону S/G2-фазы, при этом G1-фаза почти отсутствует. Это создает благоприятные условия HR, для за счет пострансляционных модификаций многих факторов HR циклин-зависимыми киназами (Orthwein et al., 2015). Эксперименты на стволовых клетках и эмбрионах (Essers et al., 2000; Bohrer et al., 2018) подтверждают повышенную активность HR. В дополнение к этому факту, по новым данным репарация DSB идет в специальных компартментах, где объединяются несколько DSB по принципу фазовой сепарации (Kilic et al., 2019). Исходя из этого, одновременная инъекция сотен копий гомологичных молекул в пронуклеус должна облегчать их взаимодействие и стимулировать рекомбинацию по механизму HR на раннем этапе конкатемеризации. К сожалению, несмотря на наличие огромного массива данных по строению конкатемеров y трансгенных животных (несколько сотен охарактеризованных линий), каких-то общих молекулярных механизмов выявлено не было. Для этого необходимо переходить от описательных работ к анализу вклада путей репарации DSB молекулярно-генетическими методами, но на зиготах это сделать очень сложно.

## ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

## 2.1. Клонирование баркодированной плазмидной библиотеки

Плазмидная библиотека была создана на основе вектора pcDNA3-Clover (Addgene #40259). На первом шаге в вектор по сайту РсіІ был лигирован линкерный фрагмент с сайтами узнавания рестриктаз SbfI и NheI. Затем по этим сайтам был встроен ПЦР-фрагмент, который в дальнейшем сыграл роль баркодированных концов трансгена. Это фрагмент длиной 500 п.о. был амплифицирован с района 6 хромосомы генома человека с помощью пары (Chr6:109287850-109288443). праймеров Затем было проведено два дополнительных цикла ПЦР с удлиненными праймерами (Рис. 9), несущими баркоды на 5'-концах (см. таблицу праймеров: Primer F/R (Приложение 1)). Для этого была поставлена ПЦР-реакция с Q5-полимеразой (NEB), удлиненными баркодированными праймерами и большим количеством первичного ПЦРфрагмента (2 мкг) в соответствии с протоколом производителя полимеразы. После двух циклов (Рис. 9А-С) удлиненный ПЦР-продукт (+ 60 п.о.) был выделен из геля и использован для лигирования с плазмидным остовом по сайтам NheI и SbfI. Ограничение в два цикла помогло избежать перемешивания баркодов при ПЦР, когда один баркод объединяется с несколькими другими (Рис. 9D).



Рисунок 9. Схема внесения баркодов в ПЦР-продукты для клонирования библиотеки. (А) ПЦРпродукт предварительно нарабатывается с геномной ДНК9 человека с помощью стандартных праймеров. Затем ПЦР-продукт нужной длины выделяется из геля и используется для двух дополнительных раундов ПЦР. (В) В первом раунде ПЦР перввичный ПЦР-продукт денатурирует и связывается комплементарным баркодированным праймером. Появляются ПЦР-продукты, фланкированные одним из баркодов. (С) Во втором раунде, ПЦР-продукты фланкируются вторым баркодом. Получаются пары баркодов, которые встречаются только в уникальной комбинации. (D) Если увеличивать число циклов ПЦР, баркодированные продукты денатурируют и комплементарно связываются с новыми баркодами, что приводит к накоплению химерных, премешанных пар баркодов, когда один баркод связан сразу с несколькими партнерами.

Последовательность баркодированных праймеров:

# 5'-CCTGCAGG<u>NNCGANNGCANNTGCNN</u>CTTGAATGACAACTAGTGCTCCAGG-3', 5'-GCTAGC<u>NNACTNNGATNNGGTNN</u>CTATCCTGACCCTGCTTGGCT-3'.

В середине амплифицированного фрагмента располагается два сайта для рестриктазы BsmBI, которая использовалась в дальнейшем для линеаризации плазмидной библиотеки. Общая схема плазмиды показана на Рис. 9. Лигазная смесь была электропорирована в клетки Top10 и высажена на большие чашки Петри 24x24 см. Всего было насчитано около 10 тыс. уникальных колоний. Эти колонии были смыты с чашки, и содержащаяся в них плазмидная библиотека была наработана и очищена с использованием кита GeneJet Midiprep (Thermo Fischer).

## 2.2. Микроинъекция баркодированной линеаризованной библиотеки

Баркодированная плазмидная библиотека (10 мкг) была обработана рестриктазой BsmBI для создания несовместимых 5'-выступающих концов (Рис. 9). Фрагмент ДНК ожидаемой длины (~6700 п.о.) был выделен из агарозного геля (diaGene, Диа-М) и дополнительно очищен с помощью ДНК-связывающих магнитных частиц Agencourt Ampure XP (Beckman Coulter<sup>тм</sup>). Очищенную ДНК элюировали в ТЕ буфере для микроинъекций (0.01 М Tris-HCl, 0.25 мМ EDTA pH 7.4). Трансгенных животных получали стандартным методом пронуклеарной микроинъекции (Brown, Corbin, 2002). Эта работа проводилась сотрудниками лаборатории генетики развития ИЦИГ СО РАН м.н.с А.Н. Кораблевым и с.н.с, к.б.н. И.А. Серовой. Раствор ДНК (1000-2000 копий в 1-2 пиколитрах) инъецировали в мужской пронуклеус зигот. В качестве доноров оплодотворенных яйцеклеток использовали суперовулированных гибридных самок (F1 CBA × C57BL/6), скрещенных с самцами C57BL/6. Выжившие проколотые зиготы подсаживали псевдобеременным самкам линии CD-1. Эмбрионы собирались от беременных самок на 13.5 день развития для выделения ДНК. Работы с животными проводились в Конвенциональном виварии ИЦИГ СО РАН в соответствии с рекомендациями этического комитета ИЦИГ СО РАН.

# 2.3. ПЦР-анализ трансгенных конструкций и генотипирование эмбрионов

Геномная ДНК выделялась из тканей эмбрионов на 13.5 день развития с помощью метода фенол-хлороформной экстракции. Куски тканей лизировались в буфере с протеиназой К (100 mM NaCl, 10 mM Tris pH 8.0, 25 mM EDTA pH 8.0, 0.5% SDS, 0.5 мкг/мкл протеиназа К) в течение 4 часов при 56°С. Затем к лизату добавлялся 3 М ацетат натрия (1/10 объема) и равный объем смеси фенолхлороформ (1:1). После центрифугирования верхняя фаза раствора отбиралась и ДНК высаждалась изопропанолом (1 объем). Затем осадок ДНК дважды промывался 80% этанолом и растворялся в воде. Для ПЦР-генотипирования эмбрионов использовались праймеры на трансген (участок гена Clover) и на слияния копиями В конкатемере (ориентация "голова-к-хвосту") между (Приложение 1) (Рис. 10). Для картирования места встройки трансгена проводился TAIL-PCR по стандартной методике (Liu, Chen, 2007) с использованием праймеров, указанных в Приложении 1. В зависимости от целей ПЦР (генотипирование, анализ перестроек в конкатемере или ПЦР-амплификация протяженного участка ДНК) использовались hot-start полимеразы Taq, LongAmp или Q5 (NEB), в соответствии с рекомендациями производителя.



Рисунок 10. Номера и расположение праймеров, использованных в работе. Последовательности и конкретное назначение праймеров указаны в таблице в Приложении 1. bc – баркод.

## 2.4. Определение копийности трансгена методом droplet digital PCR (ddPCR)

ddPCR проводилась на приборе QX100 ddPCR Systems (Bio-Rad) по протоколу производителя. Реакционная смесь объемом 20 мкл содержала 1x ddPCR Supermix

for Probes (No dUTP), по 0.9 мкМ прямого и обратного праймеров, 0.25 мкМ зондов, и 3-60 нг геномной ДНК. Количество ДНК зависело от копийности встройки и уровня мозаицизма в конкретном эмбрионе, и подбиралось экспериментально. Геномная ДНК предварительно обрабатывалась рестриктазами HindIII или DpnII, которые имеют сайт узнавания в трансгене, для фрагментации ДНК и надежного разделения копий в конкатемерах. Мы протестировали различные условия инкубации реакции рестрикции и остановились на длительной инкубации (16 часов на 37°С) в буфере CutSmart (NEB) (Приложение 2). Условия ddPCR включали исходную денатурацию ДНК при 95°С в течение 10 мин с последующей амплификацией (40 циклов): денатурация при 95°С – 30 сек, элонгация при 61°С – 1 мин. Финальный шаг ddPCR: 98°С – 7 мин. Скорость изменения температуры (гатр гаtе) равнялась 2°С в секунду. ddPCR проводилась в двух технических репликах. Последовательности для праймеров и зондов доступны в Приложении 1.

Анализ результатов ddPCR проводился в приложении QuantaSoft (Bio-Rad). Копийность оценивалась для каждого эмбриона и двух рестриктаз (HindIII-HF/DpnII) как соотношение сигнала с контрольного участка (ген *Emid1*) и целевого участка трансгена (ген Clover). Для уточнения степени мозаицизма в некоторых эмбрионах с известным сайтом встройки проводился дополнительный анализ ddPCR с зондами на трансген-геномную границу. В этом случае подсчитывалось соотношение трансгенных копий (Clover) относительно трансген-геномных границ вместо контрольного гена *Emid1* (Приложение 2).

## 2.5. Приготовление библиотек для NGS

Для баркодов было анализа конкатемерах приготовлено В И просеквенировано NGS библиотеки: три баркодированная плазмидная библиотека (№1), ПЦР-библиотека слияний между копиями (№2) и inverse PCRбиблиотека концевых баркодов (iPCR) (№3) (Рис. 15). Библиотека №1 была получена после рестрикции баркодированной плазмидной библиотеки по сайтам NheI и SbfI. Обработка плазмид рестриктазами дает фрагмент длиной 640 п.о.,

который содержит пары концевых баркодов. Это фрагмент был выделен из геля и оправлен на NGS (см. ниже).

Для приготовления библиотеки №2 использовался ПЦР-продукт района слияний между копиями, амплифицированный с помощью hot-start Q5 полимеразы (NEB). Мы протестировали разные пары праймеров и условия ПЦР, чтобы снизить неспецифичный сигнал, так как для приготовления библиотек образцы ПЦР использовались без выделения специфического бэнда из геля (мы хотели отсвеквенировать максимальный спектр слияний). Условия ПЦР: начальная денатурация ДНК – 95°С в течение 30 сек с последующей амплификацией (25-30 циклов), включая денатурацию при 98°С – 15 сек, отжиг праймеров при 64°С – 30 сек, и элонгацию при 72°С – 1 мин. Финальная элонгация: 98°С – 3 мин. Число циклов варьировало в зависимости от копийности и степени мозаицизма эмбриона. Реакции ПЦР (25 мкл) были очищены на магнитных частицах AMPure XP (Beckman Coulter) и отправлены на NGS (см. ниже).

В дополнение к ПЦР-продуктам слияний, полученных с помощью праймеров снаружи от баркодов (NGS праймеры, Рис. 12), нами были использованы праймеры на внутреннюю часть слияний (праймеры №1 или 14 на Рис. 10). 100 п.о. от точки рестрикции BsmBI. Это было необходимо, чтобы отсеквенировать слияния в многокопийных эмбрионах (#2, #3, #7) (Рис. 23).

Для создания библиотеки №3 был проведен inverse PCR (iPCR) (Рис. 15). Геномная ДНК из 10 эмбрионов была обработана рестриктазой PciI (37°C, >16 часов), которая разделяет копии в конкатемере, внося разрыв внутри слияния между копиями (Рис. 12). После рестрикции геномная ДНК очищалась на магнитных частицах AMPure XP (Beckman Coulter). Затем ДНК обрабатывалась фрагментом Кленова (СибЭнзим), чтобы достроить липкие концы, образованные PciI, до тупых. Это позволяло отличать *de novo* продукты лигирования от непорезанной ДНК в данных NGS. После достройки концов и инактивации фермента (75°C, 20 мин) геномная ДНК сильно разбавлялась (300 нг в 100 мкл лигазной реакции) и лигировалась при 16°C в течение 16 часов. Такие условия провоцируют внутримолекулярное лигирование разделенных копий трансгенов. Затем реакции лигирования были очищены на магнитных частицах и сконцентрированы. Для ПЦР-амплификации пар концевых баркодов из

лигированных циклических молекул использовались те же праймеры, что и при приготовлении библиотеки №2 (NGS праймеры, Рис. 12). Уже после приготовления ПЦР-библиотек с адаптерами Illumina для NGS, ПЦР-продукты снова обрабатывались РсіІ и выделялись из геля (размер 500-1200 п.о.), чтобы окончательно удалить все непорезанные слияния (интересующие нас *de novo* слияния лигировались по тупым концам с разрушенным сайтом PciI). По нашим оценкам, примерно 10-50% ПЦР-продукта были порезаны PciI. Эти фрагменты не попали в итоговые данные NGS.

Суммарно для библиотеки №3 с помощью метода iPCR было приготовлено 10 образцов для каждого из эмбрионов, а также дополнительная библиотека, в которой порезанная геномная ДНК двух эмбрионов (#1 и #4) была смешана в соотношении 1:1 (150 нг + 150 нг в 100 мкл лигазной реакции). Эта контрольная библиотека был необходима для оценки степени случайного лигирования между молекулами, которое могло создать химерные комбинации баркодов. Секвенирование этой библиотеки показало, что метод iPCR высокоспецифичен и пересечений между баркодами двух эмбрионов нет (Рис. 11). Это означает, что связи между баркодами, которые мы видим в NGS данных для библиотеки №3, действительно являются концевыми баркодами копии после игтрамолекулярного лигирования.



Рисунок 11. Результаты анализа данных NGS (карта конкатемеров) для контрольной библиотеки iPCR из смеси геномной ДНК из эмбрионов #1 и #4. Подробности обработки данных NGS доступны в дополнительных материалах к статье (Smirnov *et al.*, 2020).

NGS библиотеки №1-3 были подготовлены с помощью кита NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) и просеквенированы на платформе Illumina HiSeq 2500 (Illumina). Предварительно, качество библиотек оценивалось с помощью Agilent 2100 bioanalyzer (Agilent) и Qubit dsDNA HS assay kit (Life Technologies) в ЦКП функциональной геномики ИЦИГ СО РАН.

## 2.6. Секвенирование РасВіо SMRT

Кусок тканей (50 мг) от эмбриона #8 был отправлен в сервис Novogene (Hong-Kong), где из него была выделена высокомолекулярная геномная ДНК и подготовлены библиотеки для секвенирования на платформе PacBio Sequel. Полученные данные PacBio были конвертированы в формат fasta (bam2fasta). Данные PacBio выравнивались на последовательность трансгена в NCBI

megablast. Было обнаружено 113 ридов, часть из которых происходила из одной циклической молекулы (subreads), эти последовательности были объединены. В итоге мы получили 76 уникальных ридов разной длины (см. Результаты). Аннотация и анализ финальных последовательностей ридов были проведены вручную, с помощью выравниваний ридов в NCBI blast и поиска баркодов в библиотеках №1-3. Покрытие генома мыши расчитывалось путем выравнивания данных РасВіо на геном mm10 в NCBI blast (GATK DepthOfCoverage). Анализ данных РасВіо выполнялся сотрудником лаборатории генетики развития ИЦИГ СО РАН в.н.с., к.б.н. В.С. Фишманом.

## 2.7. Биоинформатический анализ данных NGS

Анализ данных NGS проходил в несколько этапов. На первом шаге риды триммировались, чтобы удалить константные последовательности, фланкирующие баркоды (cutadapt). На втором шаге концевые баркоды сравнивались с соответствующими паттернами, NN CGA NN GCA NN TGC NN и NN ACT NN GAT NN GGT NN, для того, чтобы оценить идентичность баркода и объединить риды с одинаковыми баркодами вместе. На третьем шаге баркоды соединялись в пары и отбирались с разными порогами фильтрации по числу ридов. Финальные пары баркодов были визуализированы на картах конкатемеров, используя модуль Network (http://visjs.org/). Расчеты производились на вычислительных кластерах НГУ. Полное описание методов анализа NGS данных приведено в статье Smirnov et al., 2020. Анализ данных NGS выполнялся сотрудником лаборатории генетики развития ИЦИГ СО РАН В.С. Фишманом.

# ГЛАВА З. РЕЗУЛЬТАТЫ

# 3.1. Получение трансгенных эмбрионов с конкатемерами

# 3.1.1. Дизайн и клонирование баркодированных генетических конструкций

При дизайне генетической конструкции для изучения конкатемеров мы руководствовались несколькими соображениями. Во-первых, размер и строение трансгенам, конструкции были соответствовать стандартным должны используемым в пронуклеарной микроинъекции. Поэтому для встройки баркодов мы выбрали конструкцию с флуоресцентым репортерным белком и размером около 7000 п.о. (Рис. 12). Во-вторых, баркоды должны были быть расположены на некотором расстоянии от концов линеаризованных трансгенов, чтобы избежать их потери возможном укорочении концов из-за активности при нуклеаз. Последовательность баркодов вида NNCGANNGCANNTGCNN была выбрана так, чтобы обеспечить достаточное разнообразие (10<sup>4</sup>-10<sup>5</sup>) и выравненность по нуклеотидному составу за счет включения константных тринуклеотидов. Для того, чтобы недопустить рекомбинацию между концами трансгена и какими-либо геномными последовательностями, мы решили фланкировать трансген участками из генома человека. Фрагмент человеческой ДНК, содержащий район с двумя сайтами рестрикции BsmBI, был амплифицирован с помощью ПЦР с хромосомы 6. Для ПЦР использовались два типа праймеров: обычные (на первом этапе) и баркодированные (несли баркоды на 5'-конце). Баркоды вносились в двух дополнительных раундах ПЦР, чтобы избежать появления большого числа химерных молекул, то есть ПЦР-продуктов, в которых один баркод может иметь несколько партнеров В разных молекулах Методы, Рис. 9A-D). (см. Баркодированный ПЦР-продукт был выделен из геля и встроен в плазмиду pcDNA3-Clover (Addgene #40259) (Рис. 12). NGS анализ баркодированной плазмидной библиотеки показал, что она состоит из 12657 уникальных молекул.



Рисунок 12. Схематическое изображение баркодированного вектора, который был использован для получения конкатемеров. Размер вектора – 6774 п.о. Дополнительные обозначения: PciI, BsmBI – сайты рестрикции, стрелки – позиции праймеров, которые использовались для ПЦР слияний и приготовления NGS библиотек ("NGS праймеры").

#### 3.1.2. Пронуклеарная микроинъекция и генотипирование эмбрионов

Плазмидная библиотека была линеаризована рестриктазой BsmBI, очищена выделением из агарозного геля и растворена в стандартном ТЕ буфере для микроинъекций. Как видно на Рис. 12, в линеаризованных плазмидах баркоды располагаются на расстоянии примерно 270 п.о. от концов. Линеаризованную библиотеку (длина молекулы - 6719 п.о.) инъецировали в пронуклеус и подсаживали эмбрионы псевдобеременным самкам. В каждый эмбрион было инъецировано 1000-2000 копий трансгена. На 13.5 день развития выжившие эмбрионы собирались и анализировались с помощью ПЦР. Для генотипирования использовались праймеры на плазмидную часть и на стык между трансгенами (слияние "голова-к-хвосту"). На Рис. 13 показана идентификация трансгенных эмбрионов из одной самки (12 эмбрионов). Суммарно было прогенотипировано 50 эмбрионов, из которых 10 оказались позитивны по встройке трансгена (20%). Такая эффективность является типичной (10-30%) для пронуклеарной микроинъекции на мышах (Brown, Corbin, 2002).



Рисунок 13. Генотипирование эмбрионов из одной из самок. (А) ПЦР с праймерами на внутреннюю часть плазмиды (ген Clover) (350 п.о.). (В) ПЦР с праймерами на слияния между копиями (NGS праймеры) (около 700 п.о.). Обозначения: М – ДНК-маркер (100 п.о.); wt – ДНК дикого типа; pl – плазмидная библиотека, стрелка указывает на продукт ожидаемого размера (с BsmBI линкером); #5-9 – трансгенные эмбрионы, которые использовались в дальнейшей работе под данными номерами.

### 3.1.3. Определение количества копий трансгена в конкатемерах

Для определения копийности трансгенов в геноме эмбрионов использовался метод droplet digital PCR (ddPCR). Этот метод позволяет подсчитать абсолютное и относительное число копий целевого участка в растворе геномной ДНК. В ходе дизайна праймеров и зондов для ddPCR были выбраны два участка: центральная часть гена Clover (трансген) и референсный ген мыши *Emid1* (контроль), который присутствует в гаплоидном геноме мыши в единственной копии. Этот контрольный локус был охарактеризован ранее (Codner *et al.*, 2016). Серьезное отличие данного эксперимента от стандартных приложений метода ddPCR

заключалось в том, что копии внутри конкатемера физически сцеплены. Поэтому нами были протестированы условия рестрикции в оптимальных условиях, и выбраны две рестриктазы, которые вносят разрыв внутри трансгена (см. Методы). Оба варианта рестрикции показали схожие цифры копийности, что говорит об эффективном разделении копий. Финальные данные копийности для 10 эмбрионов представлены на Рис. 14 (синие столбцы). Число копий сильно варьировало между эмбрионами: от менее чем 1 копии трансгена в эмбрионах #5 и #6 до сотен копий (эмбрионы #2, #3, #7), как и во многих экспериментах с пронуклеарной микроинъекцией (Brinster *et al.*, 1985; Burkov *et al.*, 2013).

Значения копийности менее 1 объясняются мозаицизмом эмбриона (эмбрионы #5, #6). Если встройка трансген в геном произошла на стадии двух бластомеров, лишь половина клеток эмбриона будет позитивна по встройке. Геномная ДНК, выделенная из тканей такого мозаичного эмбриона, даст искаженный сигнал при подсчете копийности ddPCR (в 2 раза больше контрольного участка и соответственно сниженная копийность трансгена). Для того, чтобы оценить вклад мозаицизма в данные копийности мы попробовали картировать встройки трансгенов методом Thermal Asymmetric Interlaced PCR (TAIL-PCR), которые использовался в нашей лаборатории ранее (Serova et al., 2012; Burkov et al., 2013; Smirnov et al., 2018). Этот метод основан на отжиге случайных и трансгенспецифичных праймеров в условиях асимметричного температурного профиля ПЦР. Результатом TAIL-PCR являются ПЦР-продукты, которые содержат участки генома, фланкирующие трансген (Liu, Chen, 2007). К сожалению, его применение для картирования конкатемеров не всегда эффективно, так как они несут много перестроек, в том числе и в районе трансген-геномных границ. Нам удалось локализовать сайты встройки конкатемеров в четырех эмбрионах (эмбрионы #4, #5, #9, #10) (Smirnov et al., 2020). Эти трансген-геномные границы уникальны для каждого эмбриона и использовались в качестве контрольного локуса для ddPCR вместо гена *Emid1*. По этим данным реальная копийность оказалась выше: эмбрион #4 - 45 копий вместо 23 копий, эмбрион #5 - 1 вместо 0.4, эмбрион #9 - 22 вместо 5, эмбрион #10 – 4 вместо 3.5 (Рис. 14, Данные ddPCR "граница").



Рисунок 14. Определение копийности трансгенов в конкатемерах. Синие столбцы – данные ddPCR для двух рестриктаз. Представлены подсчеты копийности трансгена Clover относительно контрольного гена *Emid1* (оранжевый фон) или границы геном-трансген (там, где она была картирована) (голубой фон). NA – данных нет. Зеленые столбцы – оценки копийности, исходя из NGS данных (см. далее). Для значений копийности ddPCR указаны станд. ошибки.

Стоит отметить, что копийность также определялась независимым способом, исходя из числа уникальных баркодных связей в данных NGS (см. Рис. 14 и объяснения далее в тексте).

# 3.1.4. NGS анализ баркодов в конкатемерах

Для секвенирования концевых баркодов в конкатемерах из 10 эмбрионов нами библиотеки ПЦР-продуктов. были приготовлены две Во-первых, ΜЫ амплифицировали учатки слияний между копиями в ориентации "голова-к-хвосту" (NGS библиотека слияний). Во-вторых, мы использовали метод inverse PCR (iPCR), чтобы получить информацию о баркодах на концах копий (NGS библиотека концевых баркодов). Геномная ДНК эмбрионов была обработана рестриктазой PciI, которая имеет один сайт узнавания в трансгене, внутри потенциального слияния между копиями (Рис. 12). Рестрикция РсіІ позволила разделить конкатемеры на отдельные копии, которые затем лигировались в большом реакционном объеме. В таких условиях внутримолекулярное лигирование фрагментов замыкает копии трансгенов в кольцо, поэтому ПЦР-амплификакия с NGS праймерами (Рис. 12) позволяет прочитать баркоды на концах копий. Создание iPCR библиотеки было необходимо из-за высокой рекомбинации между копиями, которая разрушила исходные пары концевых баркодов. ПЦР библиотеки были подготовлены по стандартному протоколу и просеквенированы на платформе Illumina HiSeq 2500.



Рисунок 15. Схема экспериментов по секвенированию баркодов в конкатемерах. (А) Схема клонирования баркодированной плазмидной библиотеки и микроинъекции в пронуклеус. (В) Секвенирование баркодов на концах копий в конкатемерах. NGS библиотека ПЦР-продуктов создана с помощью inverse PCR (iPCR). (С) Секвенирование баркодов в слияниях между копиями. (D) Элемент карты конкатемеров – графического способа визуализации данных NGS. Каждой NGS библиотеке соответствует свой цвет.

Данные NGS для трех баркодированных библиотек (исходная плазмидная библиотека, ПЦР-библиотека слияний, iPCR-библиотека концевых баркодов) были объединены вместе, обработаны (см. Методы) и визуализированы в графическом формате (Рис. 15D). На Рис. 16 представлена карта связей баркодов в эмбрионе #9 в качестве наглядной демонстрации. Карты конкатемеров состоят из двух элементов: узлы-баркоды под уникальными номерами (зеленый и серый цвет соответствует баркодам с 5' и 3'-концов трансгена), а также разноцветные линии-связи между баркодами. Связи соединяют баркоды, которые были обнаружены вместе в NGS библиотеках. Цвет связи определяется NGS библиотекой, где она была найдена. Так, синие связи соответствуют исходной плазмидной библиотеке, то есть баркодам в инъецированных молекулах. Зеленые связи объединяют баркоды из iPCR-библиотеки концевых баркодов, то есть те баркоды, которые мы видим в эмбрионах после конкатемеризации. Понятно, что в норме баркодные связи из плазмидной библиотеки должны обнаруживаться и в эмбрионах, однако таких сине-зеленых связей оказалось мало из-за рекомбинации баркодов (см. ниже). Трансгены, которые потеряли исходную комбинацию баркодов имеют только зеленую связь. Третий тип связей, короткие красные, обозначает слияния "головак-хвосту" между копиями из ПЦР-библиотеки слияний.

Всего нами было обнаружено 1135 уникальных трансгенных копий (*зеленые* связи) в 10 эмбрионах. Объединение данных из трех библиотек позволяет построить достаточно протяженные карты баркодных связей, однако одной из проблем является наличие разрывов между узлами (Рис. 16). Разрывы вызваны отсутствием баркодной связи в ПЦР-библиотеке по причине делеции в трансгене (потеря сайта для праймера) или палиндромного сляиния копий. Выборочное секвенирование ПЦР-продуктов методом Сэнгера выявило множество перестроек и концевых делеций, причем их количество коррелировало с копийностью конкатемера (Smirnov *et al.*, 2020).



Рисунок 16. Принципы визуализации данных NGS на примере карты конкатемера для эмбриона #9. Эллипсы соответствуют баркодам, связи между ними – молекулам трансгенов (*зеленый* и *синий* цвета) или слияниям между трансгенами (*красный* цвет). Розовая связь обозначает трансген-геномную границу, картированную методом TAIL-PCR. Разрывы между цепочками баркодов вызваны делециями концевых участков в трансгенах или палиндромными слияниями. Некоторые из них были отсеквенированы методом Сэнгера (Smirnov *et al.*, 2020).

## 3.1.5. Независимое подтверждение данных NGS

Предложенный нами метод анализа состава конкатемеров был применен впервые, поэтому одной из основных наших задач была оценка "слабых мест" предложенной методики. По нашему мнению, таким слабым местом могла бы быть чувствительность к артефактам ПЦР, так как в основе анализа лежит амплификация пар баркодов на концах молекул ДНК. Если в результате ПЦР в реакционной смеси возникает химерная молекула, несущая неправильные комбинации баркодов, на наших схемах мы будем видеть связи, которых на самом деле нет в геномах трансгенных животных.

Методы, основанные на ПЦР, могут вносить ошибки в данные NGS на нескольких этапах: точечные мутации в баркодах при стартовой ПЦР-

60

амплификации; появление химерных баркодированных молекул при стартовой ПЦР-амплификации (polymerase template switching); появление химерных ПЦРпродуктов И библиотек при лигировании адаптеров Illumina: ошибки секвенирования на чипе Illumina (Potapov, Ong, 2017; Omelina et al., 2019). Для проверки наших NGS данных мы сделали два контрольных эксперимента. Вопровели секвенирование генома одного из первых, ΜЫ эмбрионов (#8) независимым методом - Pacific Biosciences Single Molecule Real-Time (SMRT) в сервисе Novogene (Hong Kong). Этот метод секвенирования третьего поколения позволяет прочитывать фрагменты ДНК длиной в несколько десятков тысяч п.о. без использования ПЦР-амплификации при подготовке проб и в процессе секвенирования. Геном эмбриона #8 был просеквенирован с трехкратным покрытием. Этого было недостаточно для полного секвенирования конкатемера в эмбрионе, но позволило обнаружить протяженные цепочки трансгенов и комплексные перестройки. После фильтрации геномных данных мы нашли 76 уникальных ридов, содержащих последовательнось трансгена (суммарно 1.16 миллиона п.о.). Медианная длина ридов (N50) составила 20387 п.о., а разброс длины ридов – от 239 до 51644 п.о. С учетом того, что длина линеаризованного траснгена составляет около 6700 п.о., большинство ридов содержали 2-4 копии из конкатемера. К сожалению, методы секвенирования третьего поколения не отличаются большой точностью (частота ошибок ~15%), поэтому последовательности баркодов в ридах РасВіо часто несли мутации. Это затруднило автоматический анализ данных, поэтому анализ и аннотация всех элементов конкатемера производилась вручную. Всего мы смогли получить 31 контиг (объединение нескольких перекрывающихся ридов между собой), которые содержали информацию о внутренней структуре конкатемеров и участках трансген-геномных границ в эмбрионе #8 (Рис. 17), описание которых представлено далее в тексте. В конкатемере этого эмбриона было аннотировано примерно 56 копий трансгена, ~20 делеций, 8 палиндромных слияний (Рис. 17, оранжевые связи) и 4 паттерна "elongation beyond original broken end" (EBOBE) (Рис. 17 A,H,K). Важно отметить, что обнаруженные баркодные связи подтверждали наши NGS карты, что видно на примере нескольких контигов (Рис. 18).



Рисунок 17. Аннотированные контиги (объединенные риды РасВіо) после секвенирования генома эмбриона #8. Обозначения: frag – небольшой фрагмент трансгена; del – концевая делеция; H-H и T-T – палиндромные слияния ("head-to-head" и "tail-to-tail"), обозначены оранжевым цветом. "EBOBE" и "dHJ pattern 1" – паттерны рекомбинации (см. описание далее в тексте). Amplified сору – участок амплификации трансгена. bc\* - неизвестные баркоды. Цвета связей аналогичны NGS картам конкатемеров. Рисунок в высоком разрешении и оригинальные последовательности с номерами ридов (PacBio reads) доступны в статье (Smirnov *et al.*, 2020).

Другим контрольным экспериментом было смешивание геномной ДНК из двух независимых трангсгенных эмбрионов при приготовлении iPCR библиотеки концевых баркодов. Появление химерных связей между баркодами из разных эмбрионов указывало бы на наличие артефактов в методе iPCR. Мы объединили равное количество геномной ДНК (150 нг + 150 нг) из двух эмбрионов (#1 и #4) и провели анализ библиотеки iPCR в стандартных условиях, начиная со стадии лигирования и ПЦР-амплификации, и заканчивая секвенированием и фильтрацией данных. Пересечений между картами конкатемеров из двух эмбрионов обнаружено не было (Рис. 11). Подробное описание анализа можно найти в Методах.

Результаты контрольных экспериментов подтвердили, что лигирование ДНК, ПЦР-амплификация, и подготовка библиотек для NGS, не создают ошибочных комбинаций баркодов в наших данных. Таким образом, сгенерированные нами карты связей между баркодами верно отражают взаимное расположение ДНКбаркодов в геномах эмбрионов и могут быть использованы для поиска закономерностей в строении конкатемеров.



Рисунок 18. Данные секвенирования РасВіо подтверждают результаты NGS. (А) Одна из трансгенных копий была амплифицирована (#6037-#152462). На карте конкатемера этот участок (снизу) присутствует в виде кольцевой связи с увеличенным числом ридов (см. копийность ридов на Рис. 19). (В) Развилки связей между баркодами детектируются в нескольких независимых ридах РасВіо. Баркод #3680 имеет пять связей на карте конкатемера (снизу). В данных РасВіо эти связи встречаются в трех контигах. "dHJ паттерн" – паттерн кроссинговера (dHJ паттерн 1) (Рис. 22). (С) Длинный контиг из нескольких ридов РасВіо, наложенная на фрагмент карты конкатемера. Пунктирные стрелки – новые связи, которых не было на карте из-за отсутствия данных NGS для

двух районов (трансген-геномная граница и делеция). Двойные связи соответствуют dHJ паттернам (Рис. 18) и также подтверждаются PacBio.

### 3.2. Анализ карт конкатемеров

# 3.2.1. Конкатемеры образуются без участия de novo амплификации

Одной из целей нашей работы было определить, каким образом линейные молекулы трансгенов объединяются вместе в специфической ориентации. В особенности нас интересовало есть ли вклад *de novo* амплификации по механизму катящегося кольца в конкатемеризацию. Мы сопоставили данные копийности для 10 эмбрионов, расчитанные с помощью ddPCR, и данные, полученные путем подсчета зеленых связей на картах конкатемеров (Рис. 14). В случае de novo амплификации значительная часть копий в конкатемерах должна нести одинаковые баркоды из амплифицированной молекулы. ddPCR оценивает общую копийность трансгена, в то время как зеленые связи отражают количество копий с уникальными баркодами. Исходя из наших данных, во всех эмбрионах количество копий трансгена (синие столбцы на Рис. 14) приблизительно соответствовало числу уникальных трансгенов (зеленые столбцы на Рис. 14). Хотя в некоторых эмбрионах эти значения отличаются в 2 раза (эмбрионы #2, #6, #7), это, скорее всего, вызвано мозаицизмом, а не амплификацией. Ни в одном из эмбрионов копийность ddPCR не превышала значительно число уникальных копий, что говорит об отсутствии амплификации.



Рисунок 19. Распределение относительного числа ридов для отдельных копий (*зеленые* связи). Относительное число ридов = общее число NGS ридов в эмбрионе / копийность конкатемера (N). Копии с увеличенным числом ридов (>1) были исследованы дополнительно (помечены стрелками). Делеции в слияниях - одна из причин более эффективной ПЦР-аплификации баркодных слияний, что вызывает отклонение от среднего числа ридов в финальной NGS библиотеке. "Амплификация" – повторенный участок конкатемера с копией #6037-#152462 (Рис. 18А). Снизу: копийность встройки (N) – число копий (*зеленые* связи) в конкатемере.

Альтернативный способ для поиска случаев амплификации в конкатемере заключается в анализе числа ридов каждой копии (*зеленые* связи) относительно среднего числа трансгенных ридов в образце. Как видно на Рис. 19, большинство уникальных *зеленых* связей группировалось вокруг 1 (нет отличий от среднего). Всего мы обнаружили 12 случаев превышения среднего числа ридов. Например, в эмбрионе #2 было 3 случая, в которых копийность *зеленой* связи была повышена в 5-7 раз (Рис. 19). Однако детальный анализ последовательностей ридов для этих копий показал, что они содержат концевые делеции (~190 п.о.) в месте слияния, которое попадает в ПЦР-продукт для NGS. Так что превышение среднего числа ридов связано с более эффективной ПЦР-амплификацией таких молекул, а не реальной амплификацией участка. Похожие ситуации были и в других эмбрионах. Единственный доказанный случай амплификации был обнаружен в эмбрионе #8 (в 5 раз больше ридов). Этот же участок присутствует и в независимых данных

66

РасВіо (Рис. 18А). В данном случае, трансгенная копия с баркодами #6037 и #152462 была повторена минимум два раза (Рис. 18А). Таким образом, из 1135 уникальных трансгенов (зеленые связи) лишь несколько молекул могли пройти через стадию амплификации, и не более, чем в несколько раз. Это убедительно отсутствие значительной амплификации формировании доказывает при большинство конкатемеров. Следовательно, копий трансгенов после микроинъекции объединяются между собой напрямую, причем это справедливо как для малокопийных (<20 копий), так и для многокопийных встроек (50-300 копий).

#### 3.2.2. Концы трансгенов активно процессируются HR

При взгляде на карты конкатемеров (Рис. 16; Приложение 3, 4) (ссылка на файлы доступна в статье Smirnov et al., 2020) в первую очередь обращает на себя большое количество зеленых связей-трансгенов, потерявших парную синюю связь. Это означает, что баркодная метка на одном или обоих концах инъецированной молекулы была изменена. Всего нами было проанализировано 1135 копий, среди которых только 222 (20%) сохранили свои исходные комбинации концевых (Таблица 1). баркодов (сине-зеленые связи) Вероятнее всего, частота рекомбинации еще выше, так как баркоды расположены на удалении от концов трансгена (270 п.о.) и резекция достигает не баркода В каждой ИЗ процессированных молекул. Можно заключить, что большинство трансгенных молекул (>80%) не лигируются друг с другом напрямую за счет активности NHEJ ММЕЈ, а связываются и процессируются или факторами гомологичной рекомбинации (HR).

Название	Паттерн на карте	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Σ
Копии, сохранившие баркоды		3 (19%)	24 (7%)	37 (23%)	6 (15%)	0	0	124 (30%)	18 (16%)	9 (45%)	1 (25%)	222 (20%)
Копии с рекомбинацией баркодов		11 (69%)	323 (89%)	114 (71%)	35 (85%)	0	1 (50%)	281 (68%)	85 (75%)	11 (55%)	3 (75%)	864 (76%)
Слияния между копиями		14	278	137	32	1	1	518	93	19	2	1195
dHJ паттерн 1	$\frown$	2 (12%)	16 (4%)	8 (5%)	0	0	1 (50%)	7 (1,5%)	10 (9%)	0	0	44 (3,5%)
dHJ паттерн 2	$\rightarrow$	0	0	1 (1%)	0	0	0	4 (0,5%)	0	0	0	5 (0,5%)
dHJ паттерн 3	$\frown$	0	5	2	0	0	0	1	1	0	2	11
Копийность		16 (100%)	363 (100%)	160 (100%)	41 (100%)	0	2 (100%)	416 (100%)	113 (100%)	20 (100%)	4 (100%)	1135 (100%)
Развилки (2 и более связей)		0	167	54	17	0	0	148	35	0	0	421

Таблица 1. Частота различных паттернов связей на картах конкатемеров

#1-10 – номера эмбрионов. ∑ - суммарное число связей для данного паттерна в 10 эмбрионах. Процент в скобках определялся от числа всех копий (копийность, *зеленые* связи) в эмбрионе.

## 3.2.3. Механизмы рекомбинации, основанные на анализе паттернов связей

Чем обусловлена частая рекомбинация и обмен баркодами на концах молекул? По нашим предположениям, это является следствием процессинга баркодированных концов факторами HR по механизмам synthesis-dependent strand annealing (SDSA) и double-strand break repair (DSBR) (Ranjha *et al.*, 2018). Оба этих пути имеют одинаковый начальный этап и не различимы с точки зрения паттернов, поэтому в данной работе мы рассматриваем их совокупный вклад. На Рис. 20 представлена схема рекомбинации концов. Гомологичная рекомбинация инициируется резекцией 3'-концов с формированием 3'-филаментов, покрытых рекомбиназой Rad51 (Ranjha et al., 2018). З'-филамент активно ищет область гомологии (гомологичный дуплекс), которой является слияние между двумя копиями (Рис. 20А). После инвазии дуплекса и синтеза в D-петле на конце трансгена появляются несовпадающие баркодные участки, которые исправляются системой мисматч репарации (см. врезку на Рис. 20) (Chakraborty et al., 2016). Таким образом, атакующий трансген всегда копирует баркод в результате репарации мисматча в новосинтезированной цепи (Рис. 20). Это можно сравнить с механизмом генной конверсии, когда инвазия атакующей молекулой района на хромосоме приводит к копированию донорной последовательности в атакующую молекулу и потере прежней информации. Наглядной иллюстрацией этого процесса служит фрагмент карты конкатемера в эмбрионе #1 (Приложение 7). В этом участке две копии сформировали развилку, с альтернативными парами баркодов (Приложение 7А). Скорее всего, это произошло из-за незаконченной мисматч репарации в гетеродуплексе баркодов после рекомбинации. Репликация удвоила нити ДНК с каждым из баркодов, что привело к сегрегации баркодов по разным бластомерам эмбриона. Судя по количеству ридов копий из этого участка (50% для каждой копии), сегрегация произошла сразу после первого деления (Приложение 7С). Это был единственный такой случай в наших данных.

Заметной чертой карт конкатемеров является наличие огромного числа развилок, когда из одного узла-баркода выходит несколько связей (Рис. 21С). Это характерно в первую очередь для многокопийных конкатемеров из эмбрионов #2, #3, #4, #7, #8, карты которых выглядят как запутанные сети и совсем не похожи на линейные цепочки из малокопийных эмбрионов (Приложение 3 и 4). Появление развилок вызвано тем, что одно слияние с парой баркодов может копироваться несколькими независимыми атакующими молекулами по механизму, представленному на Рис. 21. Атакующие молекулы могут скопировать пару баркодов вокруг слияния целиком, или скопировать лишь один из баркодов (это зависит от степени резекции атакующего конца молекулы). Затем молекулы со скопированными баркодами могут встроиться в другом районе конкатемера (Рис. 21А-С). Например, гипотетическое слияние "голова-к-хвосту" с баркодами А<sup>3</sup> и В<sup>5</sup> может быть атаковано тремя независимыми линейными молекулами с баркодами  $X^{3'}$ ,  $Y^{3'}$  и  $Z^{3'}$  на 3'-конце (Рис. 21А). Если продукты рекомбинации будут встроены в конкатемер (Рис. 21В), мы увидим на карте связи для слияний А<sup>3'</sup>-В<sup>5'</sup>, Х<sup>3'</sup>-В<sup>5'</sup>, Ү<sup>3'</sup>-В<sup>5</sup>', Z<sup>3'</sup>-В<sup>5'</sup>, при этом узел-баркод В<sup>5'</sup> в центре развилки будет общим для четырех связей (Рис. 21С). На наших картах развилки встречаются часто (всего 421 случай) (Таблица 1), причем некоторые узлы-баркоды имеют по 4-5 связей, в редких случаях даже больше – 7-8 связей, что говорит о высокой интенсивности рекомбинации в зиготе.



Рисунок 20. Два основных механизма HR (SDSA/DSBR), которые осуществляют сборку трансгенов в тандем. (А) Общая стадия для двух путей: 3'-резекция двухцепочечных концов линейных молекул (красная и розовая) и последующий поиск гомологичного района для инвазии. Этим районом должен быть участок слияния между двумя копиями (синяя молекула). (В) Шаги и финальный результат пути DSBR. После захвата второго резектированного конца формируется двойная холидеевская структура (dHJ), которая может быть разрешена двумя способами (с кросинговером или без него). (С) Шаги и финальный результат пути SDSA. Цифры в серых и зеленых эллипсах символизируют баркоды. Снизу указаны паттерны связей, как они выглядят на картах конкатемеров. На врезке указаны шаги мисматч репарации, приводящие к смене баркода после копирования (баркод #7 заменяет баркод #3) (этот процесс одинаков для всех путей HR).

70



Рисунок 21. Механизм формирования развилок при копировании баркодов. (А) Три копии последовательно атакуют слияние А-В. (В) После копирования участка слияния, включающего и баркод В (синий пунктир), атакующие копии объединяются с другими копиями (розовые) по механизмам SDSA/DSBR (Рис. 20). Образуются новые слияния и новые пары баркодов А-В, Х-В, Y-B, Z-B. (С) В данных NGS и на картах конкатемеров связи баркода В будут выглядеть как развилки.

Как показывает анализ РасВіо данных для эмбриона #8, развилки не являются артефактом нашего метода (Рис. 18). К примеру, один из баркодов #3680 встречается с разными партнерами три раза, а многие пары баркодов были скопированы целиком в разных участках ридов РасВіо (#147233-#3447, #150859-#3172, #149759-#13397) (Рис. 17). Все это говорит о высокой интенсивности HR в зиготах, когда за один клеточный цикл единственная молекула может быть донором несколько раз.

# 3.2.4. Индикаторы кроссинговера и dHJ среди паттернов связей

В большинстве случаев продукты SDSA и DSBR имеют одинаковые паттерны связей на картах конкатемеров (Рис. 20), поэтому разделить их вклад в рекомбинацию невозможно. Однако косвенно на активность DSBR указывает появление продуктов кроссинговера, которые образуются при разрешении двойной холидеевской структуры (dHJ) (Рис. 22). Мы выделили три паттерна связей между баркодами, которые могут объясняться кроссинговером. dHJ паттерн 1 (Рис. 22A):

71

Когда два конца одной молекулы атакуют слияние между копиями с образованием dHJ, разрешение этой структуры должно привести к интеграции атакующей молекулы внутрь слияния. Тогда оба исходных баркода атакующей копии будут изменены (потеря синей связи). На картах конкатемеров этот паттерн представлен красно-зеленой связью. Всего мы насчитали 44 таких паттерна в 10 эмбрионах (Таблица 1). Если же dHJ разрешается без кроссинговера, атакующая молекула лигируется в кольцо и, возможно, даже участвует в рекомбинации далее (Рис. 18А). Еще один вариант связи (dHJ паттерн 2) возникает, когда циклическая копия взаимодействует с двумя независимыми атакующими молекулами (Рис. 18В), причем атакующие молекулы копируют баркоды у циклической формы трансгена. В случае кроссинговера, все три молекулы объединяются вместе, образуя тройную красно-сине-зеленую связь на карте конкатемеров. Мы нашли 5 таких паттернов (Таблица 1). Если dHJ для такой структуры разрешается без кроссинговера, но с копированием баркодов у циклической молекулы, мы увидим красно-синие связи (упрощенное описание паттерна: две линейных копии скопировали слияние у третьей, циклической копии, объединившись с его помощью) – это dHJ паттерн 3. Гипотетически, этот паттерн (красная + синяя связь) также соответствует самолигировавшейся копии, которая после разрыва в случайном месте встроилась концами В конкатемер. Такие молекулы упоминались уже В модели конкатемеризации №2. Частота этого паттерна была невысока - только 11 случаев на более чем тысячу копий (Таблица 1). Из этого можно заключить, что самолигированные копии редко встраиваются в конкатемер и поэтому модель конкатемеризации №2 не соответствует действительности.

Хотелось бы отметить, что описанный здесь уровень кроссинговера (около 5% всех копий) (сумма dHJ паттернов 1-3 в Таблице 2) достаточно высок (см. обсуждение), но реальный уровень кроссинговера еще выше, так как многие сложные паттерны разрушаются из-за множественной рекомбинации, которая идет при конкатемеризации. Например, когда dHJ формируется не между отдельными копиями, как в примерах на Рис. 22, а между протяженными многокопийными участками конкатемера, наличие кроссинговера можно будет предсказать по "боковой петле" на картах конкатемеров. Такие "петли" заметны на карте многокопийного эмбриона #2 (Приложение 4).


Рисунок 22. Разрешение dHJ может приводить к кроссинговеру и характерным паттернам связей. (А) Кроссинговер между двумя атакующими концами трансгена и слиянием в конкатемере (dHJ паттерн 1). (В) Кроссинговер между атакующими концами двух линейных копий и циклической молекулой (dHJ паттерны 2 и 3).

# 3.2.5. Роль NHEJ в конкатемеризации

Чтобы оценить вклад NHEJ в соединение копий мы выбрали три эмбриона с копийностью (#2, #3, #7) и просеквенировали ПЦР-продукты, высокой соответствующие району слияния "голова-к-хвосту" (Рис. 23). Все сигнатуры NHEJ были привязаны к близлежащим баркодам (1803 баркода). Интересно, что три самых представленных сигнатуры NHEJ составляли 80% всех слияний и обнаруживались в каждом из эмбрионов (Рис. 23В). Эти варианты, изображенные (Var1-3), представляли собой небольшие делеции выступающих на одноцепочечных 5'-концов, которые образуются после линеаризации трансгена рестриктазой (Рис. 12). Остальные минорные варианты сигнатур NHEJ также несли разнообразные делеции и, в редких случаях, вставки или замены нуклеотидов. Перед началом этого анализа мы планировали использовать уникальные сигнатуры

73

NHEJ в слияниях для того, чтобы понять, как копируются баркоды и "наследуются" ли делеции вместе с баркодами при рекомбинации. К сожалению, низкое разнообразие сигнатур NHEJ не позволило сделать такой анализ.

Несмотря на то, что предпочтительные варианты делеций не дают подсчитать число "уникальных" концевых делеций при внутримолекулярном точное лигировании, мы оценили минимальное значение таких молекул. На Рис. 23А видно, что всего было обнаружено 103 уникальных варианта слияний. Учитывая, что в данных для этих эмбрионов присутствует 1803 уникальных баркода (~900 уникальных молекул трансгенов), можно сказать, что в среднем 1 из 10 инъецированных молекул участвует в лигировании через NHEJ (скорее всего, до начала активной конкатемеризации через HR). На самом деле это значение должно быть гораздо выше, так как многие варианты слияний повторяют друг друга (Var1-3), а какие-то варианты не попадают в финальный конкатемер. К тому же, часть молекул лигируется друг с другом в палиндромной ориентации ("голова-к-голове" или "хвост-к-хвосту") и не попадает в наш анализ из-за технических ограничений. В целом же, по совокупности данных о рекомбинации (зеленые копии) и сигнатур NHEJ, мы можем заключить, что более 10% инъецированных трансгенных молекул предварительно обрабатываются NHEJ и лигируются друг с другом или внутримолекулярно, а уже затем служат "затравкой" для формирующих конкатемер через HR резектированных копий (Рис. 20А, синее слияние). Таким образом, NHEJ важен на первом шаге формирования конкатемера, но он почти не присоединяет копии к растущему конкатемеру "физически" (см. Обсуждение).



Рисунок 23. Сигнатуры NHEJ в слияниях. (А) Все варианты слияний (Var1-Var93), выравненные на линейные концы конструкции. Справа – частота каждого слияния. (В) Частота разных слияний в каждом из эмбрионов, Var1-3 – доминирующие слияния, "другие" – остальные минорные типы слияний.

# 3.2.6. Локализация встроек конкатемеров в эмбрионах и трансгенных линиях

В геноме эмбриона #8, который был просеквенирован РасВіо, мы обнаружили сразу четыре сайта встройки на разных хромосомах. Один из сайтов на хромосоме 8 состоял из короткого фрагмента трансгена (Рис. 24В). Остальные три хромосомных участка находились рядом с протяженным районом конкатемера. Это говорит от том, что в эмбрионе #8 есть как минимум три независимых встройки конкатемеров. Мы проанализировали риды с трансген-геномными границами, дополнив их данными NGS, чтобы получить "суперконтиги" – перекрывания нескольких контигов PacBio, в которых для заполнения разрывов используются конкатемеров. Эти участки представлены на Рис. 24 карты вместе с соответствующими хромосомными сайтами локализации.

Трансген-геномная граница на хромосоме 1 (1:104244373) лежит рядом с укороченной копией трансгена, которая далее связана с конкатемером. Установить дальнейшие связи контига (Рис. 24А) невозможно, потому что комбинация баркодов #6900-#145218 лежит в развилке и имеет несколько партнеров (Smirnov et al., 2020).

Трансген-геномные границы на хромосоме 3 (3:66073367-66074320) фланкируют протяженный конкатемер неустановленной длины (Рис. 24С). Интеграция конкатемера привела к делеции 1 тыс. п.о. генома. Заметны также паттерны рекомбинации на одной из границ конкатемера (dHJ паттерн + фрагментированные копии). Вторая граница скорее всего встроилась через прямое лигирование конца конкатемера.

Трансген-геномные границы на хромосоме 10 (10:101680335-101681703) также фланкируют протяженный конкатемер неустановленной длины (Рис. 24D). Интеграция конкатемера привела к делеции 1.4 тыс. п.о. генома. Безусловно, интересной находкой является присутствие участка локальной амплификации, который обсуждался ранее (баркоды #6037-#152462), прямо в трансген-геномной границе (Рис. 24D). Установить размер амплифицированного района невозможно из-за его повторенной структуры (как минимум две полных копии и одна укороченная).

К сожалению, одним из недостатков данных PacBio является множество ошибок в сиквенсах, поэтому мы не проводили анализ нуклеотидной последовательности в трансген-геномных слияниях у эмбриона #8.



Рисунок 24. Трансген-геномные границы в эмбрионе #8, обнаруженные в данных PacBio. (A-D) Схемы встроек выполнены в стиле карт конкатемеров (обозначения те же).

Для некоторых из эмбрионов встройка была картирована с помощью TAIL-ПЦР (см. описание метода ранее). Всего было локализовано 7 границ (Рис. 25), их строение мы изучили более детально и нашли участки микрогомологии почти в каждом эмбрионе (голубые нуклеотиды на Рис. 25). Такая же информация была получена и для трансгенных линий из архива нашей лаборатории (Serova et al., 2012; Burkov et al., 2013; Smirnov et al., 2018). Проанализировав 8 трансгенгеномных границ в 6 трансгенных линиях мышей, мы также обнаружили частое присутствие микрогомологии и высокую частоту интеграции в интроны генов (3 случая из 6) (Приложение 6). В итоге, примерно половина трансген-геномных границ несла признаки ММЕЈ, при этом средний размер микрогомологии составил около 3 нуклеотидов. Это соответствует общепринятой точки зрения на роль ММЕЈ в интеграции трансгенной ДНК в геном. В большинстве статей, где проводили секвенирование и анализ трансген-геномных границ (Hamada et al., 1993; Yan et al., 2010; Masumura et al., 2015), наблюдаются такие участки микрогомологии размером 3-4 п.о. В работе Yan et al., 2013, где проанализировали трансген-геномные границы в 30 трансгенных линиях мышей, был выявлен очень высокий процент микрогомологичных нуклеотидов (76% линий, средний размер микрогомологии - 6 п.о.) (Yan et al., 2013). Таким образом, MMEJ в зиготах работает гораздо активнее, чем в других типах клеток, где обычно служит "запасным" механизмом репарации DSB (Schimmel et al., 2017).



Рисунок 25. Схема трансген-геномных границ в некоторых из эмбрионов, где конкатемер был картирован методом TAIL-PCR. Голубые нуклеотиды – участки микрогомологии в месте встройки. Зеленые нуклеотиды – инсерции нуклеотидов.

#### ОБСУЖДЕНИЕ

Зачем изучать феномен конкатемеризации? Это процесс имеет несколько важных следствий для трансгенеза. Во-первых, конкатемеризация линейных молекул указывает на высокую частоту гомологичной рекомбинаций (HR) в эмбрионе, в отличие от соматических клеток (Fung, Weinstock, 2011). Известно, что NHEJ активен в эмбриональных стволовых клетках мыши и эмбрионах Danio rerio (Dai et al., 2010; Schimmel et al., 2017), но, по-видимому, подавляется при конкатемеризации трангенов в зиготе. Важно понять, является ли это лишь следствием доминирующей S-фазы клеточного цикла, когда активна HR, или есть специальные зиготические механизмы регуляции двух путей репарации DSB. Вовторых, рекомбинация между молекулами трансгенов увеличивает копийность встройки, что может повлиять на экспрессию трансгена. Для некоторых областей биотехнологии, таких как продукция белков в культурах клеток СНО, предпочтительна высокая копийность. В сочетании с инсуляторными элементами конкатемеризация помогает достичь высокой экспрессии рекомбинантных белков (Kostyrko et al., 2017). Однако для стандартных задач по получению трансгенных моделей мышей высокая копийность не влияет на ожидаемый уровень активности трансгена трансгена или даже имеет негативный эффект на экспрессию (McBurney et al., 2002; Williams et al., 2008). Так, есть данные, что многокопийные тандемные встройки подвергаются эффекту RIGS и гетерохроматинизации (Garrick et al., 1998; Williams et al., 2008). В наших работах мы отмечали, что копийность никак не коррелирует с экспрессией (Serova et al., 2012; Burkov et al., 2013). К примеру, в одной из линий с 200 копиями трансгена его итоговая экспрессия была ниже, чем в малокопийных линиях (2-5 копий). Наличие палиндромных слияний В многокопийных линиях также провоцирует дополнительные перестройки (Akgün et al., 1997; Würtele et al., 2003, Inagaki et al., 2016), что может привести к делециям и транслокациям в трансгенном локусе у потомков. Наконец, для CRISPR/Cas9опосредованного трансгенеза активная рекомбинация между копиями трансгенов приводит к встройке нежелательных тандемных вариантов трансгена, несмотря на позитивные результаты генотипирования (Skryabin *et al.*, 2019).



Рисунок 26. Модели конкатемеризации (A-C) и их визуализация в формате карт конкатеров. (A) Модель №1: линейные молекулы трансгенов используют гомологичные участки в циклических копиях для инвазии после резекции концов. Линейные молекулы копируют участки ДНК в районе слияния (розовые), а затем объединяются вместе по механизму SDSA/DSBR. На итоговой карте конкатемеров это выглядит как линейная цепочка трансгенов (в нашем эксперименте рекомбинация на концах приводила к частой смене баркодов, однако это не влияет на общую картину). (B) Модель №2: Большинство линейных молекул трансгена проходят через внутримолекулярное лигирование, а затем фрагментируется в случайных точках. Сборка конкатемера идет за счет перекрывающихся участков фрагментов (альтернатива розовым участкам из (A)). На итоговой карте конкатемеров это отражается в большом количестве *красно-синих* связей (исходные пары баркодов защищены от рекобинации). (C) Модель №3: тандемная структура конкатемера задается циклической амплификацией, которая синтезирует участок из одной или нескольких копий по механизму "катящегося кольца". На итоговой карте конкатемеров это выглядит как кольцевая связь с многократно увеличенным числом ридов.

### Карты связей баркодов объясняют механизм конкатемеризации

Наши результаты позволили однозначно ответить на основную загадку происхождения конкатемеров (Рис. 26). Изначально мы рассматривали три наиболее вероятных механизма конкатемеризации: рекомбинация за счет линейных

концов (модель №1) (Рис. 26А); рекомбинация за счет перекрывающихся фрагментов (модель №2) (Рис. 26В); синтез путем *de novo* амплификации через циклическую копию (модель №3) (Рис. 26С). Еще одним маловероятным сценарием может быть случайное лигирование копий С выбраковкой палиндромных слияний, однако это бы привело к множественным делециям и разрывам в картах конкатемеров, которые мы не наблюдаем. Наши данные баркодирования однозначно указывают на то, что конкатемеризация происходит по модели №1, без значительного вклада других механизмов. Это можно оценить визуально, сравнив карты конкатемеров в эмбрионах #1 (Приложение 3) и #9 (Рис. 13) с ожидаемыми паттернами связей для разных моделей (Рис. 26А-С). Об этом также говорит преобладание рекомбинации на концах трансгена (80% зеленых рекомбинировавших копий) (Таблица 1) и отсутствие амплифицированных баркодных связей (Рис. 14, 19). Конкатемеризация по модели №2 должна была бы привести К большому числу красно-синих связей. когда быстрое внутримолекулярное лигирование концов трансгена "защищает" баркодную связь от рекомбинации (Рис. 26В), но таких связей очень мало (Таблица 1). Конкатемеризация по модели №3 показала бы низкое число уникальных баркодов относительно копийности и циклические паттерны баркодных связей (Рис. 26С), этого также не наблюдается даже в многокопийных линиях (Рис. 14, 19).

Конверсия баркодов (зеленые связи с потерей синих) говорит об активной резекции, которая в большинстве копий (80%) доходит до баркодов (270 п.о. от конца). Резектированные концы объединяются вместе с помощью SDSA или DSBR, используя синтезированные комплементарные участки (розовые фрагменты на Рис. 26А). В редких случаях DSBR разрешается с кроссинговером, что приводит к обмену между участками конкатемера (Рис. 22). Мы выделили несколько паттернов связей, которые указывают на кроссинговер (Таблица 1). Интересно, что разрешение двойной холидеевской структуры (dHJ) с кроссинговером – нежелательная мера в клетке, так как приводит к перестройкам. Вне мейоза система репарации DSB избегает кроссинговера, благодаря активности геликазы Sgs1 (BLM), которая устраняет dHJ (Bizard, Hickson, 2014; Daley *et al.*, 2014). В норме частота кроссинговера в эмбриональных стволовых клетках мыши составляет менее 1.5% и 2.5%, в зависимости от репортерной конструкции

(Dronkert *et al.*, 2000; LaRocque *et al.*, 2011). По нашим данным, частота кроссинговера в эмбрионах была в несколько раз выше (>5%).

# Новый тип гибридных продуктов HR/NHEJ

Мы также обнаружили и другой интересный признак активности HR. Наше внимание привлекли участки дупликации рядом с баркодами, которые иногда встречаются в конкатемерах (Рис. 27). По нашим предположения, эти паттерны могут появляться, когда одноцепочечный резектированный 3'-конец трансгена атакует слияние, копирует участок до баркода, и покидает D-петлю, имея на конце баркодную последовательность (Рис. 27). Такой гетерологичный район на 3'-конце не дает копии вернуться для нового раунда инвазии, поэтому он присоединяется к конкатемеру через NHEJ, образуя дуплицированный участок. Мы назвали этот паттерн EBOBE (от "elongation beyond original broken end"). По нашему мнению, наличие баркодов в трансгенных молекулах провоцирует эти случаи (и позволяет их детектировать). Мы не ставили себе целью точно оценить частоту ЕВОВЕ, так как для этого понадобилось бы секвенировать слияния в поисках минидупликаций, однако в имеющихся данных мы нашли ЕВОВЕ в эмбрионе #6 и на границах трансген-геном в эмбрионах #5 и #10. Мы также нашли 4 ЕВОВЕ участка в данных РасВіо для эмбриона #8. Одна из копий (Рис. 17К) даже фланкирована этими паттернами с двух сторон. Мы заметили похожие паттерны в ранее опубликованных статьях, однако в тех случаях авторы объяснили эти паттерны случайной фрагментацией трансгенов (Hamada et al., 1993; Yan et al., 2010) или неполным копированием трансгена при встройке (Paix et al., 2017). По нашим данным, баркоды рядом с ЕВОВЕ не несут синей связи, то есть это не случайно фрагментированная циклическая молекула; при этому все ЕВОВЕ одинаково прерываются на баркоде на расстоянии 270 п.о. от середины слияния (Рис. 12). Мы EBOBE представляют собой считаем. что результат взаимодействия терминированных производных HR с путями репарации NHEJ (или MMEJ). Такие механизмы известны для некоторых других производных HR (non-canonical HR termination) (Hartlerode et al., 2016). В некоторых случаях (эмбрион #10), EBOBE имеют участки микрогомологии в стыках (Приложение 5), что указывает на участие MMEJ. К сожалению, мы детектируем только те случаи EBOBE, которые случаются рядом с баркодам, в реальности их может быть больше. Гипотетически, большие делеции и обломки копий в конкатемерах также могут возникать по механизму EBOBE, когда 3'-филамент покидает D-петлю и не ре-инициирует инвазию. В дальнейшем важно будет изучить, насколько этот патологический путь репарации активен в зиготах и других системах.



Рисунок 27. Гипотетический механизм формирования паттерна EBOBE (elongation beyond original broken end).

Некоторые механизмы репарации ДНК, основанные на HR, приводят к патологическим перестройкам в геноме человека. К ним относят single-strand

annealing (SSA), microhomology-mediated break-induced replication (MM-BIR) и multi-invasion recombination (MIR). Мы не нашли значительного вклада этих путей в конкатемеризацию. SSA соединяет одноцепочечные концы ДНК напрямую, что могло бы отразиться в объединении фрагментированных циклических копий на первом этапе конкатемеризации (красно-синие связи), но таких связей было мало (Таблица 1). MM-BIR происходит при инвазии негомологичных районов генома за счет микрогомологии на 3'-конце молекулы. ММ-BIR привел бы к копированию протяженного участка конкатемера (от нескольких тысяч до многих сотен тысяч п.о.), что привело бы к дупликации большого района (Verma, Greenberg, 2016). Мы всего один подходящий случай в наших данных (эмбрион #3, нашли неподписанная точка на Рис. 19) (Smirnov et al., 2020), поэтому данный путь не играет особой роли. MIR – это новый потенциально патогенный путь репарации DSB (Piazza, Heyer, 2019). В недавних работах группы Вольфа-Дитриха Хейера (Wolf-Dietrich Heyer) было продемонстрировано, что резектированные 3'филаменты могут использовать внутренние одноцепочечные участки далеко от 3'конца филамента для поиска гомологии. Это приводит к тому, что один филамент может произвести инвазию двух независимых участков генома и объединить их в dHJ (Piazza et al., 2017), что при разрешении dHJ с кроссинговером приводит к транслокациям хромосомных фрагментов. В нашем случае, этот путь может давать дополнительный вклад в рекомбинацию трансгенов из-за наличия большого числа потенциальных доноров. К сожалению, оценить этот вклад сложно из-за интенсивной рекомбинации трансгенов. Скорее всего эти мутагенные пути (ММ-BIR, MIR) подавлены при конкатемеризации. Известно, что для их инициации нужна протяженная резекция (>1000 п.о.) (Verma, Greenberg, 2016; Piazza, Heyer, 2019). В условиях же избытка гомологичных участков для инвазии SDSA/DSBR должны эффективно конкурировать с другими путями, так как не требуют протяженной резекции.

#### Вклад NHEJ в конкатемеризацию

Роль NHEJ в формировании конкатемеров по-прежнему не ясна. Очевидно, что на первой стадии конкатемеризации необходимо соединение линеаризованных

копий в кольцо для предоставления донора для рекомбинации (розовые фрагменты на Рис. 26А). По нашим данным, более 10% копий (сотни молекул) лигируются интармолекулярно (или другими копиями) внутри ядра, что говорит об активности NHEJ (Рис. 23). Однако по каким-то причинам вклад NHEJ в формирование финального конкатемера путем прямого лигирования копий не высок. При случайном лигировании копий за счет NHEJ можно ожидать расщепления 25%:50%:25% для трех типов ориентации копий (50% соответствует ориентации "голова-к-хвосту"). Многочисленные описания конкатемеров в пронуклеарной микроинъекции убедительно доказывают, что ориентация "голова-к-хвосту", доминирует над случайной ориентацией (наблюдаемая частота более 90% против 50% ожидаемой) (см. приложение в статье Smirnov et al., 2020). Есть и случаи очень длинных конкатемеров (~267 копий) с единственной ориентацией трансгенов (Yong et al., 2015-2). Согласно ранним гипотезам и нашим экспериментальным данным, такая предрасположенность определяется гомологичной рекомбинацией. На всех картах конкатемеров из нашей работы можно заметить длинные цепочки трансгенов из десятков копий в типичной ориентации, что было бы невозможно при активном вкладе NHEJ в "строительство" конкатемера (см. Приложения 3,4 и карты конкатемеров в статье Smirnov et al., 2020). Тем не менее, известно, что делеции и палиндромные слияния копий все-таки встречаются в конкатемерах (Chiang et al., 2012; Nicholls et al., 2018). Предполагается, что число таких копий увеличивается с концентрацией ДНК (Folger et al., 1982). В нашем случае, конкатемер из эмбриона #8 состоял из десятков перестроенных копий и нескольких палиндромных слияний (Рис. 17, оранжевые). В основном, они происходили из укороченных копий, однако было и одно идеальное палиндромное слияние с потерей лишь 4 нуклеотидов (Рис. 17J). Вопрос стабильности этих структур очень интересен и может быть изучен новыми методами секвенирования. Некоторые палиндромные слияния в эмбрионе #8 имели внутреннюю делецию (Рис. 17К, Р, Q). Почему мы не видим больше идеальных палиндромных слияний, хотя 80% слияний "голова-к-хвосту" были "идеальными" (несли лишь короткие делеции в 3-5 нуклеотидов) (Рис. 23). Возможно, это указывает на то, что в зиготе есть активный механизм удаления палиндромов из конкатемеров. Происходит ли это до или после интеграции в геном? В одной работе проследили судьбу идеального

палиндромного слияния, полученного при встройке трансгенной конструкции в геном мыши (Akgün *et al.*, 1997). Перестройки в точке палиндромного слияния происходили часто (15-56% потомков), а большинство событий были делециями (100-3000 п.о.), которые нарушали симметрию палиндрома (Akgün *et al.*, 1997).

Активная конкатемеризация происходит экстрахромосомно, когда линейные концы молекул рекомбинируют в ядре. Вероятно, на этом этапе важную роль в разрушении палиндромных слияний играет резекция. В дрожжах был обнаружен процесс, при котором резектированные одноцепочечные концы могут формировать петлю после копирования палиндромного участка (folding back) (Chen *et al.*, 2013). Такие молекулы исключались бы из конкатемеризации из-за заблокированных концов. При другом гипотетическом механизме атака резектированным концом трансгена палидромного участка приведет к образованию нестабильной D-петли. В D-петле одноцепочечные нити ДНК сформируют нестабильные структуры, вроде шпилек или крестов, которые будут удаляться системой репарации (Smirnov *et al.*, 2020).

#### Вклад NHEJ/MMEJ в интеграцию конкатемеров в геном

Интеграция конкатемеров в геном тоже происходит за счет активности NHEJ/MMEJ. Инактивация двух этих путей через ингибирование Ku70/Ku80/Lig4 (NHEJ) и полимеразы  $\theta$  (TMEJ, основной путь MMEJ) снижает частоту случайной встройки трансгенов на пять порядков – от  $10^{-2}$ - $10^{-3}$  до  $10^{-7}$ - $10^{-8}$  (Zelensky *et al.*, **MMEJ** 2017). Сигнатуры легко заметны при анализе нуклеотидных последовательностей трансген-геномных границ. Эксперименты с генетическими репортерами в эмбриональных стволовых клетках мыши показывают, что в отсутствие классического NHEJ (Ku80<sup>-/-</sup>, LigIV<sup>-/-</sup>) 93% разрывов репарируются с микрогомологиями, от 2 и более нуклеотидов. В норме это значение в 2-3 раза меньше в зависимости от конкретного локуса (Schimmel et al., 2017).

Мы изучили сиквенсы для трансген-геномных границ в 7 сайтах в трансгенных эмбрионах (Рис. 25) (Smirnov *et al.*, 2020) и в 8 сайтах из более ранних работ по локализации встроек трансгенов (Приложение 6) (Serova *et al.*, 2012; Burkov *et al.*, 2013; Smirnov *et al.*, 2018). Как видно на схемах сайтов,

последовательности микрогомологии встречаются очень часто (10/15 сайтов) и в некоторых случаях эти участки достигают 6-7 нуклеотидов. Это подтверждает главную роль MMEJ в интеграции конкатемеров, учитывая, что этот путь в норме конкурирует с NHEJ (Schimmel *et al.*, 2017).

Насколько часто происходит интеграция конкатемеров в кодирующие районы генома? Существуют разные оценки (Chiang et al., 2012; Goodwin et al., 2019). Суммарно мы картировали 15 независимых сайтов интеграции: 4 для эмбриона #8, 5 для других эмбрионов, а также 6 для описанных ранее трансгенных линий. ИЗ них, 7 встроек произошли в интроны генов и еще 3 сайта были локализованы рядом с генами (менее 20 тыс. п.о. от кодирующей области). Интроны составляют около 20-25% генома млекопитающих, поэтому повышенная частота интеграция в гены (47% (7/15) в наших данных), может указывать на ассоциацию DSB с этими районах генома. Таким образом, при получении трансгенных животных методом пронуклеарной микроинъекции необходимо секвенировать геномы трансгенных линий. Интеграция конкатемеров почти всегда приводит к геномным делециям в сайтах встройки, причем размер делеций варьирует в широких пределах, от 1 п.о. до сотен тысяч п.о. Нужно учитывать также, что даже маленькие фрагменты (Рис. 24В) могут встроиться в геном и потенциально приводить к нарушениям. Интересно, что, судя по накопленным данным локализации трансгенов, концы геномного DSB почти всегда удерживаются вместе до интеграции конкатемера (не теряются), но при этом не защищены от резекции или повторных близлежащих разрывов, что часто выражается в огромных делециях >1M п.o. (Goodwin et al., 2019).

Встройка трансгенов в интроны, как правило, приводит к нарушениям экспрессии затронутого гена (гипоморфные мутации), хотя в редких случаях интеграция большого конкатемера в интрон может никак не сказываться на экспрессии (Zhang *et al.*, 2012; Tosh *et al.*, 2017). При анализе трансгенных линий, полученных методом пронуклеарной микроинъекции в нашей лаборатории, мы обнаружили редкий случай, когда встройка трансгена в интрон гена приводит к появлению химерного транскрипта (Приложение 6) (Smirnov *et al.*, 2018). В этой трансгенной линии конструкция с геном человека *hGM-CSF* встроилась внутрь интрона гена *Cntn5*, вызвав делецию 132 тыс. п.о. При экспрессии гена *Cntn5* в

мозге у мутантных мышей возникал химерный транскрипт, в котором собственные экзоны *Cntn5* фланкировали экзоны, кодируемые трансгенной конструкцией, сохраняя правильный сплайсинг гена человека *hGM-CSF*. Уровень экспрессии химерных транскриптов составил 50% от активности нормального аллеля *Cntn5*.

#### Карты баркодов и линейная последовательность трансгенов

Из-за высокой частоты делеций и рекомбинации баркодов мы смогли построить непротиворечивые линейные карты конкатемеров лишь для эмбрионов с невысокой копийностью. Например, в эмбрионе #1 все трансгены (15 копий) выстраиваются в цепочки типичной ориентации "голова-к-хвосту" с одним разрывом (Приложение 3). Интересные данные были получены после анализа данных секвенирования РасВіо для эмбриона #8. Нам удалось восстановить длинные цепочки трансгенов, разделенные фрагментированными довольно копиями или палиндромными слияниями (Рис. 17, 24). Сложный характер рекомбинации демонстрирует эмбрион #10 с низкой копийностью (4 копии) (Приложение 5). Несмотря на наличие полной информации о структуре конкатемера, включая данные о месте интеграции, мы не смогли объяснить происхождение трансгенной встройки в этом эмбрионе. Скорее всего, конкатемер образовался через дупликацию двух объединившихся уникальных копий, за которой последовал кроссинговер с еще одним фрагментом, который перемешал баркоды (Приложение 5). Карты конкатемеров для 10 эмбрионов и исходные htmlфайлы можно найти в статье (Smirnov et al., 2020).

Таким образом, предложенный метод баркодирования позволяет устанавливать примерный порядок копий в конкатемере. Наш опыт показывает, что данные NGS (мало ошибок, но очень короткие риды) необходимо дополнять данными секвенирования третьего поколения (много ошибок, но длинные риды). Один из изъянов нашего метода – нарушенная "хронология" рекомбинации, то есть последовательности событий, которая привела к конкретным перестройкам. Для более качественного анализа вклада путей репарации необходимо равномерное распределение SNP или баркодов по всей длине трансгена, что позволит картировать районы кроссинговера и генной конверсии между копиями. Этот

подход вероятно также стимулирует образование EBOBE (много участков с баркодами – выше шанс возникновения гетерогенности на 3'-конце при инвазии) и поможет оценить процессивность и вклад этого гибридного пути репарации HR + NHEJ/MMEJ.

#### Заключение

Причины конкатемеризации трансгенов в пронуклеарной микроинъекции оставались невыясненными более 30 лет. Мы использовали новый метод баркодирования трансгенов для объяснения механизма рекомбинации между трансгенами. Секвенирование и анализ баркодов в трансгенных эмбрионах однозначно указывают на то, что конкатемеры образуются путем объединения линейных концов молекул по механизму гомологичной рекомбинации. Анализируя данные, мы также обнаружили несколько новых сигнатур рекомбинации в зиготах (свидетельства кроссинговера и дупликации трансгенов, а также EBOBE) Таким образом, предложенный нами метод баркодирования представляет перспективный инструмент для дальнейших исследований конкатемеризации в культурах клеток и эмбрионов.

# выводы

- Применение баркодированной библиотеки трансгенов с разнообразием 12.7 тыс. молекул позволяет оценивать процессы конкатемеризации в зиготе.
- Негомологичное соединение концов участвует в лигировании как минимум 10% инъецированных копий, которые служат в качестве матриц для HR, но почти не встраиваются в конкатемер.
- Как минимум 5% копий являются продуктами митотического кроссинговера при репарации DSBR, что говорит о высокой частоте этого процесса в зиготе.
- 4. Некоторые копии вероятно встраиваются в конкатемер по гибридному механизму, когда метаболиты HR соединяются с помощью MMEJ (паттерн EBOBE).
- 5. Конкатемеризация введенных в пронуклеус молекул ДНК происходит через рекомбинацию исходных концов молекул по механизмам SDSA/DSBR. Амплификации копий в конкатемере по механизму катящегося кольца не происходит.

# СПИСОК ЛИТЕРАТУРЫ

- Akgün E., Zahn J., Baumes S., Brown G., Liang F., Romanienko P.J., Lewis S., Jasin M. Palindrome resolution and recombination in the mammalian germ line // Mol Cell Biol. - 1997. - Vol. 17, № 9. - P. 5559-5570.
- Ardui S., Ameur A., Vermeesch J.R., Hestand M.S. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics // Nucleic Acids Res. - 2018. - Vol. 46, № 5. - P. 2159-2168.
- Asoshina M., Myo G., Tada N., Tajino K., Shimizu N. Targeted amplification of a sequence of interest in artificial chromosome in mammalian cells // Nucleic Acids Res. - 2019. -Vol. 47, № 11. - P. 5998-6006.
- Biehs R., Steinlage M., Barton O., Juhász S., Künzel J., Spies J., Shibata A., Jeggo P.A., Löbrich M. DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination // Mol Cell. 2017. Vol. 65, № 4. P. 671-684.
- Bishop J.O. Chromosomal insertion of foreign DNA // Reprod Nutr Dev. 1996. -Vol. 36. - P. 607-618.
- Bizard A.H., Hickson I.D. The dissolution of double Holliday junctions // Cold Spring Harb Perspect Biol. - 2014. - Vol. 6, № 7. - P. a016477.
- Blier P.R., Griffith A.J., Craft J., Hardin J.A. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks // J Biol Chem. - 1993. - Vol. 268, № 10. - P. 7594-7601.
- Bohrer R.C., Dicks N., Gutierrez K., Duggavathi R., Bordignon V. Double-strand DNA breaks are mainly repaired by the homologous recombination pathway in early developing swine embryos // FASEB J. - 2018. - Vol. 32, № 4. - P. 1818-1829.
- Brinster R.L., Chen H.Y., Trumbauer M., Senear A.W., Warren R., Palmiter R.D. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs // Cell. - 1981. - Vol. 27. - P. 223-231.

- 10. Brinster R.L., Chen H.Y., Trumbauer M.E., Yagle M.K., Palmiter R.D. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs // Proc Natl Acad Sci. - 1985. - Vol. 82. - P. 4438-4442.
- Brinster R.L., Ritchie K.A., Hammer R.E., O'Brien R.L., Arp B., Storb U. Expression of a microinjected immunoglobulin gene in the spleen of transgenic mice // Nature. 1983. Vol. 306, № 5941. P. 332.
- 12. Brown G.A.J., Corbin T.J. Oocyte injection in the mouse. Transgenesis Techniques. 2002; 180: New Jersey Humana Press, 39–70.
- 13. Burkov I.A., Serova I.A., Battulin N.R., Smirnov A.V., Babkin I.V., Andreeva L.E., Dvoryanchikov G.A., Serov O.L. Expression of the human granulocyte-macrophage colony stimulating factor (hGM-CSF) gene under control of the 5'-regulatory sequence of the goat alpha-S1-casein gene with and without a MAR element in transgenic mice // Transgenic Res. 2013. Vol. 22, № 5. P. 949-964.
  - Bzymek M., Thayer N.H., Oh S.D., Kleckner N., Hunter N. Double Holliday junctions are intermediates of DNA break repair // Nature. - 2010. - Vol. 464, № 7290. - P. 937-941.
  - 15. Cain-Hom C., Splinter E., van Min M., Simonis M., van de Heijning M., Martinez M., Asghari V., Cox J.C., Warming S. Efficient mapping of transgene integration sites and local structural changes in Cre transgenic mice using targeted locus amplification // Nucleic Acids Res. 2017. Vol. 45, № 8. P. e62.
  - Cejka P. DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination // J Biol Chem. - 2015. - Vol. 290, №38. -P. 22931-22938.
  - Chaible L.M., Corat M.A., Abdelhay E., Dagli M.L. Genetically modified animals for use in research and biotechnology // Genet Mol Res. - 2010. - Vol. 9, № 3. - P. 1469-1482.
  - Chakraborty U., George C.M., Lyndaker A.M., Alani E. A delicate balance between repair and replication factors regulates recombination between divergent DNA Sequences in Saccharomyces cerevisiae // Genetics. - 2016. - Vol. 202. - P. 525–540.

- Chang H.H.Y., Pannunzio N.R., Adachi N., Lieber M.R. Non-homologous DNA end joining and alternative pathways to double-strand break repair // Nat Rev Mol Cell Biol. - 2017. - Vol. 18. - P. 495–506.
- Chen H., Lisby M., Symington L.S. RPA coordinates DNA end resection and prevents formation of DNA hairpins // Mol Cell. - 2013. - Vol. 50, № 4. - P. 589-600.
- Chen L., Nievera C.J., Lee A.Y., Wu X. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair // J Biol Chem. - 2008. - Vol. 283, № 12. - P. 7713-7720.
- 22. Chiang C., Jacobsen J.C., Ernst C., Hanscom C., Heilbut A., Blumenthal I., Mills R.E., Kirby A., Lindgren A.M., Rudiger S.R. et al. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration // Nat Genet. 2012. Vol. 44. P. 390–397.
- 23. Chinnadurai G. Joint surveillance of the replication foci by PCNA and CtIP // Cell Cycle. 2009. Vol. 8, № 9. P. 1306-1307.
- 24. Codner G.F., Lindner L., Caulder A., Wattenhofer-Donzé M., Radage A., Mertz A., Eisenmann B., Mianné J., Evans E.P., Beechey C.V., et al. Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction // BMC Cell Biol. 2016. Vol. 17. P. 30.
- 25. Cunningham L.A., Coté A.G., Cam-Ozdemir C., Lewis S.M. Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism // Mol Cell Biol. 2003. Vol. 23, № 23. P. 8740-8750.
- Dai J., Cui X., Zhu Z., Hu W. Non-homologous end joining plays a key role in transgene concatemer formation in transgenic zebrafish embryos // Int. J. Biol. Sci. 2010. Vol. 6. P. 756–768.
- Daley J.M., Gaines W.A., Kwon Y., Sung P. Regulation of DNA pairing in homologous recombination // Cold Spring Harb. Perspect. Biol. - 2014. - Vol. 6. -P. 1–15.
- De Maio N., Shaw L.P., Hubbard A., George S., Sanderson N.D., Swann J., Wick R., AbuOun M., Stubberfield E., Hoosdally S.J., Crook D.W., Peto T.E.A., Sheppard A.E., Bailey M.J., Read D.S., Anjum M.F., Walker

A.S., Stoesser N., On Behalf Of The Rehab Consortium. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes // Microb Genom. - 2019. - Vol. 5, №9.

- 29. de Vree P.J., de Wit E., Yilmaz M., van de Heijning M., Klous P., Verstegen M.J., Wan Y., Teunissen H., Krijger P.H., Geeven G., Eijk P.P., Sie D., Ylstra B., Hulsman L.O., van Dooren M.F., van Zutven L.J., van den Ouweland A., Verbeek S., van Dijk K.W., Cornelissen M., Das A.T., Berkhout B., Sikkema-Raddatz B., van den Berg E., van der Vlies P., Weening D., den Dunnen J.T., Matusiak M., Lamkanfi M., Ligtenberg M.J., ter Brugge P., Jonkers J., Foekens J.A., Martens J.W., van der Luijt R., van Amstel H.K., van Min M., Splinter E., de Laat W. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping // Nat Biotechnol. 2014. Vol. 32, № 10. P. 1019-1025.
- 30. Deamer D. Nanopore analysis of nucleic acids bound to exonucleases and polymerases // Annu Rev Biophys. 2010. Vol. 39. P. 79-90.
- Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing // Nat Biotechnol. - 2016. - Vol. 34, № 5. - P. 518-524.
- Ditch S., Sammarco M.C., Banerjee A., Grabczyk, E. Progressive GAA/TTC repeat expansion in human cell lines // PLoS Genet. 2009. Vol. 5. P. e1000704.
- Dougherty J.D., Zhang J., Feng H., Gong S., Heintz N. Mouse transgenesis in a single locus with independent regulation for multiple fluorophores // PLoS One. 2012. Vol. 7, № 7. P. e40511.
- 34. Downs J.A., Jackson S.P. A means to a DNA end: the many roles of Ku // Nat Rev Mol Cell Biol. 2004. Vol. 5, № 5. P. 367-378.
- 35. Dronkert M.L., Beverloo H.B., Johnson R.D., Hoeijmakers J.H., Jasin M., Kanaar R. Mouse RAD54 affects DNA double-strand break repair and sister chromatid exchange // Mol Cell Biol. 2000. Vol. 20, № 9. P. 3147-3156.
- 36. Drouet J., Delteil C., Lefrançois J., Concannon P., Salles B., Calsou P. DNAdependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks // J Biol Chem. - 2005. - Vol. 280, № 8. -P. 7060-7069.

- 37. Dubose A.J., Lichtenstein S.T., Narisu N., Bonnycastle L.L., Swift A.J., Chines P.S., Collins F.S. Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene // Nucleic Acids Res. 2013. Vol. 41, № 6. P. e70.
- 38. Essers J., van Steeg H., de Wit J., Swagemakers S.M., Vermeij M., Hoeijmakers J.H., Kanaar R. Homologous and non-homologous recombination differentially affect DNA damage repair in mice // EMBO J. 2000. Vol. 19. P. 1703–1710.
- Folger K.R., Wong E.A., Wahl G., Capecchi M.R. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules // Mol Cell Biol. 1982.
   Vol. 2. P. 1372-1387.
- 40. Fukuma M., Ganmyo Y., Miura O., Ohyama T., Shimizu N. Cloning and Characterization of a Human Genomic Sequence that Alleviates Repeat-Induced Gene Silencing // PLoS One. 2016 Vol. 11, № 4. P. e0153338.
- 41. Fung H., Weinstock D.M. Repair at single targeted DNA double-strand breaks in pluripotent and differentiated human cells // PLoS One. 2011. Vol. 6, № 5. P. e20514.
- 42. Gabrieli T., Sharim H., Fridman D., Arbib N., Michaeli Y., Ebenstein Y. Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH) // Nucleic Acids Res. 2018. Vol. 46, № 14. P. e87.
- 43. Garrick D., Fiering S., Martin D.I., Whitelaw E. Repeat-induced gene silencing in mammals // Nat Genet. 1998. Vol. 18, № 1. P. 56-59.
- 44. Girton J.R., Johansen K.M. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila // Adv Genet. - 2008. - Vol. 61. - P. 1-43.
- 45. Goodwin L.O., Splinter E., Davis T.L., Urban R., He H., Braun R.E., Chesler E.J., Kumar V., van Min M., Ndukum J., Philip V.M., Reinholdt L.G., Svenson K., White J.K., Sasner M., Lutz C., Murray S.A. Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis // Genome Res. - 2019. - Vol. 29, № 3. - P. 494-505.

- 46. Gordon J.W., Scangos G.A., Plotkin D.J., Barbosa J.A., Ruddle F.H. Genetic transformation of mouse embryos by microinjection of purified DNA // Proc Natl Acad Sci USA. - 1980. - Vol. 77, № 12. - P. 7380-7484.
- 47. Grandjean M., Girod P.A., Calabrese D., Kostyrko K., Wicht M., Yerly F., Mazza C., Beckmann J.S., Martinet D., Mermod N. High-level transgene expression by homologous recombination-mediated gene transfer // Nucleic Acids Res. 2011. Vol. 39, № 15. P. e104.
- 48. Grundy G.J., Rulten S.L., Arribas-Bosacoma R., Davidson K., Kozik Z., Oliver A.W., Pearl L.H., Caldecott K.W. The Ku-binding motif is a conserved module for recruitment and stimulation of non-homologous end-joining proteins // Nat Commun. - 2016. - Vol. 7. - P. 11242.
- 49. Gu B., Chen P.L. Expression of PCNA-binding domain of CtIP, a motif required for CtIP localization at DNA replication foci, causes DNA damage and activation of DNA damage checkpoint // Cell Cycle. 2009. Vol. 8, № 9. P. 1409-1420.
- 50. Hafford-Tear N.J., Tsai Y.C., Sadan A.N., Sanchez-Pintado B., Zarouchlioti C., Maher G.J., Liskova P., Tuft S.J., Hardcastle A.J., Clark T.A., Davidson A.E. CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy-associated TCF4 triplet repeat // Genet Med. 2019.
  Vol. 21, № 9. P. 2092-2102.
- 51. Halabi A., Fuselier K.T.B., Grabczyk E. GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutLγ and depends upon the endonuclease domain in MLH3 isoform one // Nucleic Acids Res. 2018. Vol., 46, № 8. P. 4022-4032.
- 52. Hamada T., Sasaki H., Seki R., Sakaki Y. Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: sequence analysis of genometransgene and transgene-transgene junctions at two loci // Gene. - 1993. - Vol. 128. - P. 197-202.
- 53. Hanash S. Integrated global profiling of cancer // Nat Rev Cancer. 2004. -Vol. 4, № 8. - P. 638-644.
- 54. Hartlerode A.J., Willis N.A., Rajendran A., Manis J.P., Scully R. Complex Breakpoints and Template Switching Associated with Non-canonical Termination

of Homologous Recombination in Mammalian Cells // PLoS Genet. - 2016. - Vol. 12, № 11. - P. e1006410.

- 55. Hasty P., Montagna C. Chromosomal Rearrangements in Cancer: Detection and potential causal mechanisms // Mol Cell Oncol. 2014. Vol. 1, № 1. P. e29904.
- 56. Hedges D.J., Deininger P.L. Inviting instability: Transposable elements, doublestrand breaks, and the maintenance of genome integrity // Mutat Res. - 2007. -Vol. 616, № 1-2. - P. 46-59.
- 57. Hentges P., Ahnesorg P., Pitcher R.S., Bruce C.K., Kysela B., Green A.J., Bianchi J., Wilson T.E., Jackson S.P., Doherty A.J. Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos // J Biol Chem. 2006. Vol. 281, № 49. P. 37517-37526.
- 58. Holkers M., de Vries A.A., Gonçalves M.A. Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells // Nucleic Acids Res. - 2012. - Vol. 40, № 5. - P. 1984-1999.
- 59. Huertas P. DNA resection in eukaryotes: deciding how to fix the break // Nat Struct Mol Biol. 2010. Vol. 17, № 1. P. 11-16.
- 60. Huertas P., Jackson S.P. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair // J Biol Chem. 2009. Vol. 284, № 14.
   P. 9558-9565.
- 61. Inagaki H., Kato T., Tsutsumi M., Ouchi Y., Ohye T., Kurahashi H. Palindrome-Mediated Translocations in Humans: A New Mechanistic Model for Gross Chromosomal Rearrangements // Front Genet. - 2016. - Vol. 12, №7. - P. 125.
- Inagaki H., Ohye T., Kogo H., Tsutsumi M., Kato T., Tong M., Emanuel B.S., Kurahashi H. Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations // Nat Commun. - 2013. -Vol. 4. - P. 1592.
- 63. Ismail I.H., Gagné J.P., Genois M.M., Strickfaden H., McDonald D., Xu Z., Poirier G.G., Masson J.Y., Hendzel M.J. The RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair pathway choice // Nat Cell Biol. - 2015. - Vol. 17, № 11. - P. 1446-1457.

- 64. Jaenisch R., Mintz B. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA // Proc Natl Acad Sci USA. - 1974. - Vol. 71, № 4. - P. 1250-1254.
- 65. Jain M., Koren S., Miga K.H., Quick J., Rand A.C., Sasani T.A., Tyson J.R., Beggs A.D., Dilthey A.T., Fiddes I.T., Malla S., Marriott H., Nieto T., O'Grady J., Olsen H.E., Pedersen B.S., Rhie A., Richardson H., Quinlan A.R., Snutch T.P., Tee L., Paten B., Phillippy A.M., Simpson J.T., Loman N.J., Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads // Nat Biotechnol. - 2018. - Vol. 36, № 4. - P. 338-345.
- 66. Jiang W., Crowe J.L., Liu X., Nakajima S., Wang Y., Li C., Lee B.J., Dubois R.L., Liu C., Yu X., Lan L., Zha S. Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining // Mol Cell. 2015. Vol. 58, № 1. P. 172-185.
- 67. Johnson R.D., Jasin M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells // EMBO J. 2000. Vol. 19, № 13. P. 3398-3407.
- 68. Kan Y., Batada N.N., Hendrickson E.A. Human somatic cells deficient for RAD52 are impaired for viral integration and compromised for most aspects of homology-directed repair // DNA Repair (Amst). - 2017. - Vol. 55. - P. 64-75.
- Kasianowicz J.J., Brandin E., Branton D., Deamer D.W. Characterization of individual polynucleotide molecules using a membrane channel // Proc Natl Acad Sci USA. - 1996. - Vol. 93, № 24. - P. 13770-13773.
- 70. Khillan J.S., Overbeek P.A., Westphal H. Drosophila P element integration in the mouse // Dev Biol. 1985. Vol. 109, № 1. P. 247-250.
- 71. Kilic S., Lezaja A., Gatti M., Bianco E., Michelena J., Imhof R., Altmeyer M. Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments // EMBO J. 2019. Vol. 38, № 16. P. e101379.
- 72. Kim J.S., Krasieva T.B., Kurumizaka H., Chen D.J., Taylor A.M., Yokomori K. Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells // J Cell Biol. - 2005. - Vol. 170, № 3. - P. 341-347.

- 73. Kirschner K., Melton D.W. Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs // Anticancer Res. 2010. Vol. 30, № 9. P. 3223-3232.
- 74. Kostyrko K., Neuenschwander S., Junier T., Regamey A., Iseli C., Schmid-Siegert E., Bosshard S., Majocchi S., Le Fourn V., Girod P.A., Xenarios I., Mermod N. MAR-Mediated transgene integration into permissive chromatin and increased expression by recombination pathway engineering // Biotechnol Bioeng. 2017. Vol. 114, № 2. P. 384-396.
- Kozarewa I., Armisen J., Gardner A.F., Slatko B.E., Hendrickson C.L. Overview of Target Enrichment Strategies // Curr Protoc Mol Biol. - 2015. - Vol. 112. - P. 7.21.1-7.21.23.
- 76. Krepulat F., Löhler J., Heinlein C., Hermannstädter A., Tolstonog G.V., Deppert W. Epigenetic mechanisms affect mutant p53 transgene expression in WAP-mutp53 transgenic mice // Oncogene. 2005. Vol. 24, № 29. P. 4645-4659.
- 77. Kurahashi H., Inagaki H., Yamada K., Ohye T., Taniguchi M., Emanuel B.S., Toda T. Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations // J Biol Chem. - 2004. - Vol. 279, № 34. - P. 35377-35383.
- 78. Kwaks T.H., Barnett P., Hemrika W., Siersma T., Sewalt R.G., Satijn D.P., Brons J.F., van Blokland R., Kwakman P., Kruckeberg A.L., Kelder A., Otte A.P. Identification of anti-repressor elements that confer high and stable protein production in mammalian cells // Nat Biotechnol. 2003. Vol. 21, № 5. P. 553-558.
- 79. Laboulaye M.A., Duan X., Qiao M., Whitney I.E., Sanes J.R. Mapping Transgene Insertion Sites Reveals Complex Interactions Between Mouse Transgenes and Neighboring Endogenous Gene // Front Mol Neurosci. - 2018. - Vol. 11. - P. 385.
- 80. Lan L., Ui A., Nakajima S., Hatakeyama K., Hoshi M., Watanabe R., Janicki S.M., Ogiwara H., Kohno T., Kanno S., Yasui A. The ACF1 Complex Is Required for DNA Double-Strand Break Repair in Human Cells // Mol Cell. 2010. Vol. 40, № 6. P. 976-987.
- LaRocque J.R., Stark J.M., Oh J., Bojilova E., Yusa K., Horie K., Takeda J., Jasin
   M. Interhomolog recombination and loss of heterozygosity in wild-type and

Bloom syndrome helicase (BLM)-deficient mammalian cells // Proc Natl Acad Sci USA. - 2011. - Vol. 108, № 29. - P. 11971-11976.

- Lazzaro F., Sapountzi V., Granata M., Pellicioli A., Vaze M., Haber J.E, Plevani P., Lydall D., Muzi-Falconi M. Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres // EMBO J. 2008. Vol. 27, № 10. P. 1502-1512.
- 83. Le Saux A., Houdebine L.M., Jolivet G. Chromosome integration of BAC (bacterial artificial chromosome): evidence of multiple rearrangements // Transgenic Res. - 2010. - Vol. 19, № 5. - P. 923-931.
- 84. Li S., Jia S., Hou L., Nguyen H., Sato S., Holding D., Cahoon E., Zhang C., Clemente T., Yu B. Mapping of transgenic alleles in soybean using a nanoporebased sequencing strategy // J Exp Bot. - 2019. - Vol. 70, № 15. - P. 3825-3833.
- 85. Liu Y.G., Chen Y. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences // Biotechniques. - 2007. - Vol. 43, № 5. - P. 649-654.
- 86. Llorente B., Smith C.E., Symington L.S. Break-induced replication: what is it and what is it for? // Cell Cycle. 2008. Vol. 7, № 7. P. 859-864.
- 87. Lobachev K.S., Gordenin D.A., Resnick M.A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements // Cell. - 2002. - Vol. 108, № 2. - P. 183–193.
- 88. Mailand N., Bekker-Jensen S., Faustrup H., Melander F., Bartek J., Lukas C., Lukas J. RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins // Cell. - 2007. - Vol. 131, № 5. - P. 887-900.
- Mantere T., Kersten S., Hoischen A. Long-Read Sequencing Emerging in Medical Genetics // Front Genet. - 2019. - Vol. 7, №10. - P. 426.
- 90. Mari P.O., Florea B.I., Persengiev S.P., Verkaik N.S., Brüggenwirth H.T., Modesti M., Giglia-Mari G., Bezstarosti K., Demmers J.A., Luider T.M., Houtsmuller A.B., van Gent D.C. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4 // Proc Natl Acad Sci USA. -2006. - Vol. 103, № 49. - P. 18597-18602.

- 91. Mark W.H., Signorelli K., Blum M., Kwee L., Lacy E. Genomic structure of the locus associated with an insertional mutation in line 4 transgenic mice // Genomics. - 1992. - Vol. 13, № 1. - P. 159-166.
- 92. Masumura K., Sakamoto Y., Kumita W., Honma M., Nishikawa A., Nohmi T. Genomic integration of lambda EG10 transgene in gpt delta transgenic rodents // Genes Environ. - 2015. - Vol. 37. - P. 24.
- 93. McBurney M.W., Mai T., Yang X., Jardine K. Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes // Exp Cell Res. 2002. Vol. 274, № 1. P. 1-8.
- 94. McFarlane M., Wilson J.B. A model for the mechanism of precise integration of a microinjected transgene // Transgenic Res. 1996. Vol. 5, № 3. P. 171-177.
- 95. McKnight G.S., Hammer R.E., Kuenzel E.A., Brinster R.L. Expression of the chicken transferrin gene in transgenic mice // Cell. - 1983. - Vol. 34, № 2. - P. 335-341.
- 96. McMahill M.S., Sham C.W., Bishop D.K. Synthesis-dependent strand annealing in meiosis // PLoS Biol. 2007. Vol. 5, № 11. P. e299.
- 97. McVey M., Khodaverdian V.Y., Meyer D., Cerqueira P.G., Heyer W.D. Eukaryotic DNA Polymerases in Homologous Recombination // Annu Rev Genet.
  2016. Vol. 50. P. 393-421.
- 98. McVey M., Lee S.E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings // Trends Genet. - 2008. \_ Vol. 24, № 11. - P. 529-538.
- 99. Migchielsen A.A., Breuer M.L., Hershfield M.S., Valerio D. Full genetic rescue of adenosine deaminase-deficient mice through introduction of the human gene // Hum Mol Genet. - 1996. - Vol. 5, № 10. - P. 1523-1532.
- 100. Mikhailov K.V., Efeykin B.D., Panchin A.Y., Knorre D.A., Logacheva M.D., Penin A.A., Muntyan M.S., Nikitin M.A., Popova O.V., Zanegina O.N., Vyssokikh M.Y., Spiridonov S.E., Aleoshin V.V., Panchin Y.V. Coding palindromes in mitochondrial genes of Nematomorpha // Nucleic Acids Res. - 2019. - Vol. 47, № 13. :6858-6870.
- 101. Minev D., Guerra R., Kishi J.Y., Smith C., Krieg E., Said K., Hornick A., Sasaki H.M., Filsinger G., Beliveau B.J., Yin P., Church G.M., Shih W.M. Rapid in vitro

production of single-stranded DNA // Nucleic Acids Res. - 2019. - Vol. 47, № 22. -P. 11956-11962.

- 102. Moreira P.N., Pozueta J., Pérez-Crespo M., Valdivieso F., Gutiérrez-Adán A., Montoliu L. Improving the generation of genomic-type transgenic mice by ICSI // Transgenic Res. - 2007. - Vol. 16, № 2. - P. 163-168.
- 103. Murnane J.P., Yu L.C. Acquisition of telomere repeat sequences by transfected DNA integrated at the site of a chromosome break // Mol Cell Biol. 1993. Vol. 13, № 2. P. 977-983.
- 104. Nakagaki A., Hirano S., Urakawa A., Mitake M., Kishino T. Transgenic mice with a tandem duplication of the Necdin gene overexpress Necdin // Mamm Genome. - 2018. - Vol. 29, № 9-10. - P. 680-689. (2)
- 105. Nakagaki A., Urakawa A., Hirano S., Anami T., Kishino T. Application of droplet digital PCR in the analysis of genome integration and organization of the transgene in BAC transgenic mice // Sci Rep. 2018. Vol. 8, № 1. P. 6638. (1)
- 106. Nakanishi T., Kuroiwa A., Yamada S., Isotani A., Yamashita A., Tairaka A., Hayashi T., Takagi T., Ikawa M., Matsuda Y., Okabe M. FISH analysis of 142 EGFP transgene integration sites into the mouse genome // Genomics. 2002. -Vol. 80, № 6. P. 564-574.
- 107. Nicholls P.K., Bellott D.W., Cho T.J., Pyntikova T., Page D.C. Locating and Characterizing a Transgene Integration Site by Nanopore Sequencing // G3 (Bethesda). - 2019. - Vol9, №5. - P. 1481-1486.
- 108. Ochs F., Somyajit K., Altmeyer M., Rask M.B., Lukas J., Lukas C.. 53BP1 fosters fidelity of homology-directed DNA repair // Nat Struct Mol Biol. 2016. Vol. 23, № 8. P. 714-721.
- 109. Omelina E.S., Ivankin A.V., Letiagina A.E., Pindyurin, A.V. Optimized PCR conditions minimizing the formation of chimeric DNA molecules from MPRA plasmid libraries // BMC Genomics. - 2019. - Vol. 20. - P. 536.
- 110. Orthwein A., Noordermeer S.M., Wilson M.D., Landry S., Enchev R.I., Sherker A., Munro M., Pinder J., Salsman J., Dellaire G., Xia B., Peter M., Durocher D. A mechanism for the suppression of homologous recombination in G1 cells // Nature.
  2015. Vol. 528, № 7582. P. 422-426.

- 111. Osman F., Ahn J.S., Lorenz A., Whitby M.C.. The RecQ DNA helicase Rqh1 constrains Exonuclease 1-dependent recombination at stalled replication forks // Sci Rep. 2016. Vol. 6. P. 22837.
- 112. Overbeek P.A., Lai S.P., Van Quill K.R., Westphal H. Tissue-specific expression in transgenic mice of a fused gene containing RSV terminal sequences // Science. -1986. -Vol. 231, № 4745. - P. 1574-1577.
- 113. Paix A., Folkmann A., Goldman D.H., Kulaga H., Grzelak M.J., Rasoloson D., Paidemarry S., Green R., Reed R.R., Seydoux G. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks // Proc Natl Acad Sci USA. - 2017. - Vol. 114, № 50. - P. e10745.
- 114. Palmiter R.D., Chen H.Y., Brinster R.L. Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring // Cell. - 1982. - Vol. 29, № 2. - P. 701-710.
- 115. Pannunzio N.R., Watanabe G., Lieber M.R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks // J Biol Chem. 2018. Vol. 293, № 27. P. 10512-10523.
- 116. Pastwa E., Błasiak J. Non-homologous DNA end joining // Acta Biochim Pol. 2003. Vol. 50, №4. P. 891-908.
- 117. Piazza A., Heyer W.-D. Homologous recombination and the formation of complex genomic rearrangements // Trends Cell Biol. - 2019. - Vol. 29. - P. 135– 149.
- 118. Piazza A., Wright W.D., Heyer W.-D. Multi-invasions are recombination byproducts that induce chromosomal rearrangements // Cell. - 2017. - Vol. 170. - P. 760–773.
- 119. Pieper F.R., de Wit I.C., Pronk A.C., Kooiman P.M., Strijker R., Krimpenfort P.J., Nuyens J.H., de Boer H.A. Efficient generation of functional transgenes by homologous recombination in murine zygotes // Nucleic Acids Res. 1992. Vol. 20, № 6. P. 1259-64.
- 120. Pierce A.J., Hu P., Han M., Ellis N., Jasin M. Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells // Genes Dev. - 2001. - Vol. 15, № 24. - P. 3237-3242.

- 121. Postow L., Ghenoiu C., Woo E.M., Krutchinsky A.N., Chait B.T., Funabiki H. Ku80 removal from DNA through double strand break-induced ubiquitylation // J Cell Biol. 2008. Vol. 182, № 3. P. 467-479.
- 122. Potapov V., Ong J.L. Examining sources of error in PCR by single-molecule sequencing // PLoS One. 2017. Vol. 12, № 7. P. e0181128.
- 123. Ramakrishnan S., Kockler Z., Evans R., Downing B.D., Malkova A. Single-strand annealing between inverted DNA repeats: Pathway choice, participating proteins, and genome destabilizing consequences // PLoS Genet. - 2018. - Vol. 14, № 8. - P. e1007543.
- 124. Ramírez A., Milot E., Ponsa I., Marcos-Gutiérrez C., Page A., Santos M., Jorcano J., Vidal M. Sequence and chromosomal context effects on variegated expression of keratin 5/lacZ constructs in stratified epithelia of transgenic mice // Genetics. 2001. Vol. 158, № 1. P. 341-350.
- 125. Ranjha L., Howard S.M., Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes // Chromosoma.
  2018. Vol. 127. P. 187–214.
- 126. Rass U., Compton S.A., Matos J., Singleton M.R., Ip S.C., Blanco M.G., Griffith J.D., West S.C. Mechanism of Holliday junction resolution by the human GEN1 protein // Genes Dev. 2010. Vol. 24, № 14. P. 1559-1569.
- 127. Rohan R.M., King D., Frels W.I. Direct sequencing of PCR-amplified junction fragments from tandemly repeated transgenes // Nucleic Acids Res. - 1990. - Vol. 18. - P. 6089-6095.
- 128. Rooney S., Chaudhuri J., Alt F.W. The role of the non-homologous end-joining pathway in lymphocyte development // Immunol Rev. - 2004. - Vol. 200. - P. 115-131.
- 129. Rosser J.M., An W. Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation // Gene. 2010. Vol. 456, № 1-2. P. 15-23.
- 130. Sartori A.A., Lukas C., Coates J., Mistrik M., Fu S., Bartek J., Baer R., Lukas J., Jackson, S.P. Human CtIP promotes DNA end resection // Nature. - 2007. - Vol. 450. - P. 509–514.

- 131. Schimmel J., Kool H., Schendel R., Tijsterman M. Mutational signatures of non- homologous and polymerase theta- mediated end- joining in embryonic stem cells // EMBO J. - 2017. - Vol. 36. - P. 3634–3649.
- 132. Serova I.A., Dvoryanchikov G.A., Andreeva L.E., Burkov I.A., Dias L.P., Battulin N.R., Smirnov A.V., Serov O.L. A 3,387 bp 5'-flanking sequence of the goat alpha-S1-casein gene provides correct tissue-specific expression of human granulocyte colony-stimulating factor (hG-CSF) in the mammary gland of transgenic mice // Transgenic Res. - 2012. - Vol. 21, № 3. - P. 485-498.
- 133. Sevim V., Lee J., Egan R., Clum A., Hundley H., Lee J., Everroad R.C., Detweiler A.M., Bebout B.M., Pett-Ridge J., Göker M., Murray A.E., Lindemann S.R., Klenk H.P., O'Malley R., Zane M., Cheng J.F., Copeland A., Daum C., Singer E., Woyke T. Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies // Sci Data. - 2019. - Vol. 26, No1. - P. 285.
- 134. Skryabin B.V., Gubar L., Seeger B., Kaiser H., Stegemann A., Roth J., Meuth S.G., Pavenstädt H., Sherwood J., Pap T. et al. Pervasive head-to-tail insertions of DNA templates mask desired CRISPR/Cas9-mediated genome editing events // bioRxiv. 2019. <u>https://doi.org/10.1101/570739</u>. preprint: not peer reviewed.
- 135. Slesarev A., Viswanathan L., Tang Y., Borgschulte T., Achtien K., Razafsky D., Onions D., Chang A., Cote C. CRISPR/CAS9 targeted CAPTURE of mammalian genomic regions for characterization by NGS // Sci Rep. - 2019. - Vol. 9, № 1. - P. 3587.
- 136. Smirnov A., Fishman V., Yunusova A., Korablev A., Serova I., Skryabin B.V., Rozhdestvensky T.S., Battulin N. DNA barcoding reveals that injected transgenes are predominantly processed by homologous recombination in mouse zygote // Nucleic Acids Res. - 2020. - Vol. 48, № 2. - P. 719-735.
- 137. Smirnov A.V., Kontsevaya G.V., Feofanova N.A., Anisimova M.V., Serova I.A., Gerlinskaya L.A., Battulin N.R., Moshkin M.P., Serov O.L. Unexpected phenotypic effects of a transgene integration causing a knockout of the endogenous Contactin-5 gene in mice // Transgenic Res. - 2018. - Vol. 27, № 1. - P. 1-13.
- 138. Smith K. Theoretical mechanisms in targeted and random integration of transgene DNA // Reprod Nutr Dev. - 2001. - Vol. 41, № 6. - P. 465-485.

- 139. Son M.Y., Hasty P. Homologous recombination defects and how they affect replication fork maintenance // AIMS Genet. 2019. Vol. 5, №4. P. 192-211.
- 140. Sullivan M.R., Bernstein K.A. RAD-ical New Insights into RAD51 Regulation // Genes (Basel). - 2018. - Vol. 9, № 12. - P. e629.
- 141. Swain J.L., Stewart T.A., Leder P. Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting // Cell. 1987. Vol. 50, № 5. P. 719-727.
- 142. Szabó P.E., Hübner K., Schöler H., Mann J.R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells // Mech Dev. 2002. Vol. 115, №1-2. P. 157-160.
- 143. Szostak J.W., Orr-Weaver T.L., Rothstein R.J., Stahl F.W. The double-strandbreak repair model for recombination // Cell. - 1983. - Vol. 33, № 1. - P. 25-35.
- 144. Tacken P.J., Zee A.V.D., Beumer T.L., Florijn R.J., Gijpels M.J.J., Havekes L.M., Frants R.R., Dijk K.W.V., Hofker M.H. Effective generation of very low density lipoprotein receptor transgenic mice by overlapping genomic DNA fragments: high testis expression and disturbed spermatogenesis // Transgenic Res. - 2001. - Vol. 10.
  - P. 211-221.
- 145. Takata M., Sasaki M.S., Sonoda E., Morrison C., Hashimoto M., Utsumi H., Yamaguchi-Iwai Y., Shinohara A., Takeda S. Homologous recombination and nonhomologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells // EMBO J. - 1998. - Vol. 17, № 18. - P. 5497-5508.
- 146. Tanaka H., Bergstrom D.A., Yao M.C., Tapscott S.J. Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification // Nat Genet. 2005. Vol. 37, № 3. P. 320-327.
- 147. Taylor M.R.G., Špírek M., Chaurasiya K.R., Ward J.D., Carzaniga R., Yu X., Egelman E.H., Collinson L.M., Rueda D., Krejci L., Boulton S.J. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination // Cell. - 2015. - Vol. 162, № 2. - P. 271-286.
- 148. Tomaska L., Nosek J., Kar A., Willcox S., Griffith J.D. A New View of the T-Loop Junction: Implications for Self-Primed Telomere Extension, Expansion of

Disease-Related Nucleotide Repeat Blocks, and Telomere Evolution // Front Genet. - 2019. - Vol. 10. - P. 792.

- 149. Tosh J.L., Rickman M., Rhymes E., Norona F.E., Clayton E., Mucke L., Isaacs A.M., Fisher E.M.C., Wiseman F.K. The integration site of the APP trans- gene in the J20 mouse model of Alzheimer's disease // Wellcome Open Res. 2017. Vol. 2. P. 84.
- 150. Trombetta B., Cruciani F. Y chromosome palindromes and gene conversion // Hum Genet. - 2017. - Vol. 136, № 5. - P. 605-619.
- 151. van Dijk E.L., Jaszczyszyn Y., Naquin D., Thermes C. The Third Revolution in Sequencing Technology // Trends Genet. 2018. Vol. 34, № 9. P. 666-681.
- 152. Van Keuren M.L., Gavrilina G.B., Filipiak W.E., Zeidler M.G., Saunders T.L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes // Transgenic Res. 2009. Vol. 18, № 5. P. 769-785.
- 153. Verma P., Greenberg R.A. Noncanonical views of homology-directed DNA repair// Genes Dev. 2016. Vol. 30. P. 1138–1154.
- 154. Vilenchik M.M., Knudson A.G. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer // Proc Natl Acad Sci USA. -2003. - Vol. 100, № 22. - P. 12871-12876.
- 155. Voineagu I., Narayanan V., Lobachev K.S., Mirkin S.M. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins // Proc. Natl Acad. Sci. USA. - 2008. - Vol. 105. - P. 9936–9941.
- 156. Wagner EF, Covarrubias L, Stewart TA, Mintz B. Prenatal lethalities in mice homozygous for human growth hormone gene sequences integrated in the germ line. Cell. 1983 Dec;35(3 Pt 2):647-55.
- 157. Walker J.R., Corpina R.A., Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair // Nature. 2001. Vol. 412, № 6847. P. 607-614.
- 158. Wang G., Vasquez K.M. Effects of Replication and Transcription on DNA Structure-Related Genetic Instability // Genes (Basel). - 2017. - Vol. 8, №1. - P. e17.

- 159. Wang H., Shao Z., Shi L.Z., Hwang P.Y., Truong L.N., Berns M.W., Chen D.J., Wu X. CtIP protein dimerization is critical for its recruitment to chromosomal DNA double-stranded breaks // J Biol Chem. - 2012. - Vol. 287, № 25. - P. 21471-21480.
- 160. Watanabe T., Horiuchi T. A novel gene amplification system in yeast based on double rolling-circle replication // EMBO J. 2005. Vol. 24, № 1. P. 190-198.
- 161. Watanabe T., Tanaka H., Horiuchi T. Complex repeat structure promotes hyperamplification and amplicon evolution through rolling-circle replication // Nucleic Acids Res. - 2018. -Vol. 46, № 10. - P. 5097-5108.
- 162. Watson C.M., Crinnion L.A., Hewitt S., Bates J., Robinson R., Carr I.M., Sheridan E., Adlard J., Bonthron D.T. Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications // Lab Invest. -2020. - Vol. 100, № 1. - P. 135-146.
- 163. Weirather J.L., de Cesare M., Wang Y., Piazza P., Sebastiano V., Wang X.J., Buck D., Au K.F. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. Version 2 // F1000Res. - 2017. - Vol. 6. - P. 100.
- 164. Wilkie T.M., Palmiter R.D. Analysis of the integrant in MyK-103 transgenic mice in which males fail to transmit the integrant // Mol Cell Biol. 1987. Vol. 7, № 5. P. 1646-1655.
- 165. Williams A., Harker N., Ktistaki E., Veiga-Fernandes H., Roderick K., Tolaini M., Norton T., Williams K., Kioussis D. Position effect variegation and imprinting of transgenes in lymphocytes // Nucleic Acids Res. 2008. Vol. 36, № 7. P. 2320-2329.
- 166. Williams G.J., Hammel M., Radhakrishnan S.K., Ramsden D., Lees-Miller S.P., Tainer J.A. Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time // DNA Repair (Amst). - 2014. - Vol. 17. - P. 110-120.
- 167. Würtele H., Little K.C., Chartrand P. Illegitimate DNA integration in mammalian cells // Gene Ther. 2003. Vol. 10, № 21. P. 1791-1799.
- 168. Xu L., Seki M. Recent advances in the detection of base modifications using the Nanopore sequencer // J Hum Genet. - 2020. - Vol. 65, № 1. - P. 25-33.
- 169. Yan B., Li D., Gou K. Homologous illegitimate random integration of foreign DNA into the X chromosome of a transgenic mouse line // BMC Mol Biol. - 2010. -Vol. 11. - P. 58.
- 170. Yan B.W., Zhao Y.F., Cao W.G., Li N., Gou K.M. Mechanism of random integration of foreign DNA in transgenic mice // Transgenic Res. 2013. Vol. 22, № 5. P. 983-992.
- 171. Yano K., Morotomi-Yano K., Adachi N., Akiyama H. Molecular mechanism of protein assembly on DNA double-strand breaks in the non-homologous end-joining pathway // J Radiat Res (Tokyo). - 2009. - Vol. 50, № 2. - P. 97-108.
- 172. Ye F., Signer E.R. RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration // Proc Natl Acad Sci USA. -1996. - Vol. 93, № 20. - P. 10881-10886.
- 173. Yong C.S., Sharkey J., Duscio B., Venville B., Wei W.Z., Jones R.F., Slaney C.Y., Mir Arnau G., Papenfuss A.T., Schröder J., Darcy P.K., Kershaw M.H. Embryonic Lethality in Homozygous Human Her-2 Transgenic Mice Due to Disruption of the Pds5b Gene // PLoS One. 2015. Vol. 10, № 9. P. e0136817.
- 174. Yong C.S., Westwood J.A., Schröder J., Papenfuss A.T., von Scheidt B., Moeller M., Devaud C., Darcy P.K., Kershaw M.H. Expression of a Chimeric Antigen Receptor in Multiple Leukocyte Lineages in Transgenic Mice // PLoS One. 2015. Vol. 10, № 10. P. e0140543. (2)
- 175. Yu X., Chen J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains // Mol Cell Biol. 2004. V. 24. N. 21. P. 9478-86.
- 176. Yu Y., Mahaney B.L., Yano K., Ye R., Fang S., Douglas P., Chen D.J., Lees-Miller S.P. DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks // DNA Repair (Amst). - 2008. -Vol. 7, № 10. - P. 1680-1692.
- 177. Zapotoczny G., Sekelsky J. Human Cell Assays for Synthesis-Dependent Strand Annealing and Crossing over During Double-Strand Break Repair. Human Cell Assays for Synthesis-Dependent Strand Annealing and Crossing over During Double-Strand Break Repair // G3 (Bethesda). - 2017. - Vol. 7, № 4. - P. 1191-1199.

- 178. Zelensky A.N., Schimmel J., Kool H., Kanaar R., Tijsterman M. Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA // Nat Commun. - 2017. - Vol. 8, № 1. - P. 66.
- 179. Zhang R., Yin Y., Zhang Y., Li K., Zhu H., Gong Q., Wang J., Hu X., Li N. Molecular characterization of transgene integration by next-generation sequencing in transgenic cattle // PLoS One. - 2012. - Vol. 7, № 11. - P. e50348.
- 180. Zhou Z.H., Akgūn E., Jasin M. Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences // Proc Natl Acad Sci USA. -2001. - Vol. 98, № 15. - P. 8326-8333.
- 181. Баттулин Н.Р., Фишман В.С., Орлов Ю.Л., Мензоров А.Г., Афонников Д.А., Серов О.Л. 3с-методы в исследованиях пространственной организации генома // Вавиловский журнал генетики и селекции. - 2012. - Т. 16, № 4/2.
- 182. Серова И.А., Андреева Л.Е., Хайдарова Н.В., Диас Л.П.Б., Дворянчиков Г.А., Бурков И.А., Багинская Н.В. Мозаичная экспрессия репортерного гена lacZ под контролем 5'-регуляторных последовательностей альфа-S1-казеина у трансгенных мышей // Цитология. 2009. Т. 51, вып. 11. С. 917-923.

Праймеры (5'-3')	Назначение
1 TCCAAAACTGAAAGTCTGTGTCCA	NGS, перестройки
2a CGGCCTTTTTACGGTTCCTG	NGS, TAIL-PCR
2b GAGCCTATGGAAAAACGCCA	NGS
2c TGGAAAAACGCCAGCAACG	NGS, перестройки
3 GATAGTTACCGGATAAGGCGC	Перестройки
4 GCCCCAGTGCTGCAATGATA	Перестройки
5 GCTGAAGATCAGTTGGGTGCAC	Перестройки
6 CGCAAAACGCCTAACCCTAAG	Перестройки
7 ACATCAATGGGCGTGGATAG	Перестройки
8 GAACTTCAGGGTCAGCTTGC	Перестройки
9 GTCCAGGAGCGCACCATCTC	Перестройки, ddPCR
10 ATGTGTGTCAGTTAGGGTGTGG	Перестройки
11 CCGCTTTTCTGGATTCATCGAC	Перестройки
12 GCGAGTCAGTGAGCGAGGAA	Перестройки
13a TATCAGCTCACTCAAAGGCGG	NGS. перестройки. TAIL-PCR
13b ACGGTTATCCACAGAATCAGGG	NGS
13c ACTCAAAGGCGGTAATACGGT	NGS
	Перестройки
	NGS
	ТАП - PCR (случайный праймер)
LADI ACCATGGACTCCAGAGCGGCCGCCNNNNNGGTT	ТАЦ-РСВ (случайный праймер)
LAD2 ACGATGGACTCCAGAGCGGCCGCVNVNNNTATG	ТАЦ-РСВ (случайный праймер)
	ТАЦ РСР (случайный праймер)
	ТАЦ ВСР (случайный праймер)
	ТАЦ DCD (отучанный праимер)
	ТАЦ РСК (случаиный праимер)
	IAIL-PCK
	ddPCR (Clover) F
	ddPCR (Clover) R
	ddPCR (Clover) Зонд
	ddPCR (Emid1) F
	ddPCR (Emid1) R
Emili Pr CCIGOUICAICIGAGICC (FAM-BHQ)	ddPCR (Emid1) Зонд
	ddPCR (Emb#4 граница) F
N4 R GUAUAUAU IGAAGGAAAGGU	ddPCR (Emb#4 граница) R
N4 PT CLAGGGACGAAAAACCCCAACCAA	аарск (Етв#4 граница) зонд
	ddPCR (Emb#5 border) F
	ddPCR (Emb#5 border) R
N5 Pr ICCATAGGCICCGCIAATAIGCCIGCIAAA (FAM-BHQ)	ddPCR (Етоня граница) Зонд
N9 F TCCAAAACIGAAAGICIGIGICCA	ddPCR (Emb#9 border) F
N9 R CIAAACIGACCCCAGGIGGG	ddPCR (Emb#9 border) R
N9 Pr TCCICCGICICCCTGGCTTTCCTIG (FAM-BHQ)	ddPCR (Emb#9 border) Probe
NI0 F ATGGCTAAAATGGCTGAGAGGT	ddPCR (Emb#10 border) F
NIO R CCAGGGATGCAGGGATGGT	ddPCR (Emb#10 border) R
N10 Pr AGGGTCAGGATAGATACCGAAGGGTTTGTTTA (FAM-BHQ)	ddPCR (Emb#10 border) Probe
N10 bc F TATCAGCTCACTCAAAGGCGG	ddPCR (Emb#10 barc) F
N10 bc R CCCTGGAGCACTAGTTGTCAT	ddPCR (Emb#10 barc) R
N10 bc 6409 Pr CAGGTTCGATTGCATGTGCCACTT (HEX-BHQ)	ddPCR (Emb#10 barc 6409) Probe
N10 bc 4523 Pr CAGGTTCGACTGCATATGCACCTT (HEX-BHQ)	ddPCR (Emb#10 barc 4523) Зонд
N10 bc 146980 Pr CATGGCTAGCTAACTCCGATGGGGTTACTATCC (HEX-BHQ)	ddPCR (Emb#10 barc 146980) Зонд
N10 bc 149645 Pr CATGGCTAGCTTACTACGATTCGGTATCTATCC (HEX-BHQ)	ddPCR (Emb#10 barc 149645) Зонд
Primer F CTTGAATGACAACTAGTGCTCCAGG	Клонирование конструкций
Primer F (barc) CCTGCAGGNNCGANNGCANNTGCNNCTTGAATGACAACTAGTGCTCCAGG	Клонирование конструкций
Primer R CTATCCTGACCCTGCTTGGCT	Клонирование конструкций
Primer R (barc) GCTAGCNNACTNNGATNNGGTNNCTATCCTGACCCTGCTTGGCT	Клонирование конструкций

Таблица праймеров, использованных в работе.





Ch2: HEX зонд (трансген Clover)

В

С

Условия рестрикции	Копийность (относительно гена <i>Emid1</i> ), ± станд. ошибка
Без рестрикции	6.3 ± 0.3
Рестрикция HindIII-HF, 20 мин в ddPCR смеси	$12.7\pm0.5$
Рестрикция HindIII-HF, 16 часов (CutSmart buffer)	14.7 ± 0.8
Рестрикция DpnII, 16 часов (DpnII buffer)	$14.4\pm0.9$



Приложение 2. Анализ копийности трансгенов методом droplet digital PCR (ddPCR). (A) 2D плот ddPCR измерений копийности участка Clover (трансген) относительно контрольного гена *Emid1*. (B) Подбор условий рестрикции для эффективного разделения копий в конкатемере в геномной ДНК (ДНК из эмбриона #1). (C) Схема зондов для учета мозаицизма эмбрионов в ddPCR (в эмбрионах с известной трансген-геномной границей).



Приложение 3. Карта конкатемера в эмбрионе #1. В этом эмбрионе 15 копий, которые распределены по двум цепочкам. В нижней части карты можно заметить раздвоенный участок цепочки (вокруг баркода #185100), который появился в результате неполной репарации баркодных мисматчей (см. подробные объяснения в статье Smirnov *et al.*, 2020).



Приложение 4. Карта конкатемера в многокопийном эмбрионе #2. В этом эмбрионе насчитывается 365 уникальных копий (*зеленые* связи), при этом копийность по ddPCR в два раза ниже (скорее всего, из-за мозацизма эмбриона). Видно большое число развилок, которые затрудняют анализ связей в эмбрионе.



## Карта конкатемера для эмбриона #10

A



Приложение 5. Строение конкатемера в эмбрионе #10. Копийность конкатемера – 4 копии (ddPCR). (A) Карта конкатемера для эмбриона #10. (B) Схематичное изображение реальной структуры конкатемера, установленной по итогам дополнительного ПЦР-анализа и секвенирования. Показаны нуклеотидные последовательности трансген-геномных границ. Серый цветом выделены участки баркодов, которые находятся в месте стыков. Черными рамками выделены последовательности микрогомологии.





Приложение 6. Локализация встроек конкатемера в линиях трансгенных мышей из ранних работ (Serova *et al.*, 2012; Burkov *et al.*, 2013; Smirnov *et al.*, 2018). (А) Схема трансгенгеномных границ в 6 линиях мышей. Голубые нуклеотиды – участки микрогомологии в месте встройки. (В) Схема химерного транскрипта в линии GM9. Трансген встроился в интрон гена мыши контактин-5 (*Cntn5*), что вызвало делецию размером 132 тыс. п.о., которая затронула несколько экзонов. В мутантных мышах экспрессируется химерный транскрипт из фланкирующих экзонов *Cntn5* (E1-E4 + E9-E24) и экзонов генетической конструкции (human *GMCSF*) (Smirnov *et al.*, 2018). Цифрами указаны длины экзонов.



Приложение 7. Сегрегация баркодов, указывающая на роль мисматч репарации в обмене баркодами на концах трансгенов. (А) Участок карты конкатемера в эмбрионе #1 (Приложение 3). Видно, что участок из двух копий (копии В\* и С\*) имеет альтернативные варианты баркодов, которые выглядят как развилки. (В) Гипотетическая схема формирования района. После рекомбинации по механизму SDSA/DSBR возникают гетеродуплексы баркодов, которые репарируются мисматч репарацией (см. Рис. 20). Если участок не успеет отрепарироваться до S-фазы, репликация может удвоить район, созранив оба баркода (движение реплисомы показано стрелкой). (С) Число ридов для каждой копии трансгена (зеленая связь). Копии из участка с сегрегировавшими баркодами имеют число ридов примерно 50%, что указывает на мозаицизм данного района (предположительно, расхождение произошло после первого деления).