

На правах рукописи

Смирнов Александр Васильевич

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

03.02.07 - генетика

**АВТОРЕФЕРАТ**

диссертации на соискание учёной степени  
кандидата биологических наук

Новосибирск – 2021

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН) в лаборатории генетики развития, г. Новосибирск.

**Научный руководитель** **Баттулин Нариман Рашитович**  
к.б.н., заведующий лабораторией генетики развития  
Института цитологии и генетики СО РАН, г. Новосибирск.

**Официальные оппоненты** **Кантидзе Омар Леванович**  
д.б.н., заведующий лабораторией стабильности генома  
Института биологии гена РАН, г. Москва.

**Кулемзин Сергей Викторович**  
к.б.н., с.н.с. лаборатории иммуногенетики Института  
молекулярной и клеточной биологии СО РАН, г.  
Новосибирск.

**Ведущее учреждение** Федеральное государственное бюджетное учреждение  
Институт молекулярной генетики Национального  
исследовательского центра «Курчатовский институт», г.  
Москва.

Защита диссертации состоится «\_\_»\_\_\_\_\_2021 г. на утреннем заседании диссертационного совета Д 003.011.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» в конференц-зале Института по адресу: пр. ак. Лаврентьева 10, г. Новосибирск, 630090, тел +7 (383) 3634906, факс +7(383) 3331278. e-mail: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института: [www.bionet.nsc.ru](http://www.bionet.nsc.ru).

Автореферат разослан «\_\_»\_\_\_\_\_2021 г.

Ученый секретарь  
диссертационного совета,  
доктор биологических наук

Хлебодарова Т.М.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность исследования.** Микроинъекция ДНК в пронуклеус зиготы является основным способом получения трансгенных животных в течение последних 40 лет. Одним из примечательных свойств микроинъекции является то, что трансгены встраиваются в единый локус, образуя большие повторенные районы из копий, ориентированных в направлении «голова-к-хвосту» (т.н. «конкатемеры») (Rohan *et al.*, 1990). Предполагается, что в этом процессе участвуют два пути репарации двухцепочечных разрывов ДНК (double-strand breaks, DSB): гомологичная рекомбинация (homologous recombination, HR) и негомологичное соединение концов (non-homologous end-joining, NHEJ) (Huertas, 2010). Для объяснения феномена конкатемеризации было предложено несколько моделей. Согласно модели *de novo* амплификации, конкатемер возникает из одной или нескольких циклических копий за счет многократной репликации (Bishop, 1996). Другая модель утверждает, что конкатемеры происходят из фрагментированных копий трансгена (Bishop, 1996; Smith, 2001). Как известно, часть линейаризованных трансгена образует циклические молекулы сразу после инъекции. Случайная фрагментация таких молекул привела бы к появлению большого пула перекрывающихся гомологичных участков ДНК, которые могут рекомбинировать в зиготе. Третья гипотеза постулирует, что конкатемеры образуются путем рекомбинации между исходными 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 HR, а не для случайного лигирования по механизму NHEJ. Несмотря на стремительный прогресс в области трансгенеза, однозначного подтверждения ни одна из предложенных моделей конкатемеризации до сих пор так и не получила. Чтобы проследить, как проходит процесс внехромосомной рекомбинации трансгенов в зиготе, мы приняли решение пометить каждую из инъецированных молекул уникальным ДНК-баркодом, представленным случайной последовательностью нуклеотидов.

### Цели и задачи диссертационной работы

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

### **Задачи:**

1. 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 ( $10^4$ - $10^6$  уникальных молекул).
2. Получить трансгенных животных со встройкой конкатемеров в геноме.
3. Просеквенировать концевые баркоды трансгенов в конкатемерах у животных с различной копийностью. Оценить представленность и копийность 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.
5. Изучить строение трансген-геномных границ в месте встройки конкатемера.

### **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 методы анализа ДНК-баркодов в tandemно-повторенных последовательностях трансгенов. Опробованы различные методы визуализации и интерпретации данных секвенирования нового поколения (NGS) для баркодированных трансгенов.
3. Впервые применен современный метод секвенирования третьего поколения (PacBio) для анализа баркодов в геномной ДНК. Также продемонстрировано, что данный метод позволяет детектировать «идеальные» палиндромные слияния трансгенов.
4. Описан новый продукт рекомбинации трансгенной ДНК – «удлинение обломанного конца» (elongation beyond original broken end, EBOBE), который может быть характерен для метода пронуклеарной микроинъекции.

## **Теоретическая и практическая значимость работы**

С теоретической точки зрения, изучение конкатемеризации поможет обнаружить эволюционные особенности процесса репарации двухцепочечных разрывов в зиготе млекопитающих. Основываясь на результатах, полученных в данной работе, можно представить рекомендации по улучшению эффективности трансгеназа в пронуклеарной микроинъекции. Например, временное блокирование активности негомологичного соединения концов (NHEJ), совместно с добавлением в ядро циклических копий трансгена, может увеличить копийность трансгена. Использование перекрывающихся длинных одноцепочечных фрагментов может стимулировать сборку составных двухцепочечных конструкций за счет активности гомологичной рекомбинации. Баркодированная плазмидная библиотека, охарактеризованная в данной работе, может применяться для изучения различных аспектов рекомбинации ДНК в зиготе и культурах клеток.

### **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.

**Вклад автора.** Основные результаты получены автором самостоятельно. Баркодированные плазмидные библиотеки были сконструированы и охарактеризованы под руководством 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). Биоинформатический анализ данных выполнялся в.н.с., к.б.н. В.С. Фишманом (ИЦиГ СО РАН).

**Апробация результатов.** Результаты докладывались на международном конгрессе CRISPR-2018 (Новосибирск). По теме диссертации было опубликовано 4 работы в рецензируемых научных журналах, входящих в международные базы цитирований (Web of Science, Scopus).

**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.

## МАТЕРИАЛЫ И МЕТОДЫ

**Клонирование баркодированной плазмидной библиотеки.** Плазмидная библиотека была создана на основе вектора pcDNA3-Clover (Addgene #40259), методом рестрикции-лигирования. В вектор были встроены два баркода вида NNCGANNGCANNTGCNN и NNACTNNGATNNGGTNN, разделенные парой сайтов рестрикции BsmBI.

**Микроинъекция баркодированной линейризованной библиотеки.** Библиотека была линейризована рестриктазой BsmBI, очищена из геля, и инъецирована в пронуклеусы по протоколу (Brown, Corbin, 2002).

**ПЦР-генотипирование эмбрионов.** Геномная ДНК была выделена из эмбрионов на 13.5 день развития. Генотипирование проводили с праймерами на внутренний плазмидный участок и на участок слияния между трансгенами в конкатемере.

**Определение копийности методом droplet digital PCR.** Замешивание реакций и анализ результатов проводились в соответствии с рекомендациями производителя (Bio-Rad). Для определения концентрации копий трансгена в растворе геномной ДНК применялись зонды на контрольный ген (*Emid1*) и трансген (участок гена *Clover*).

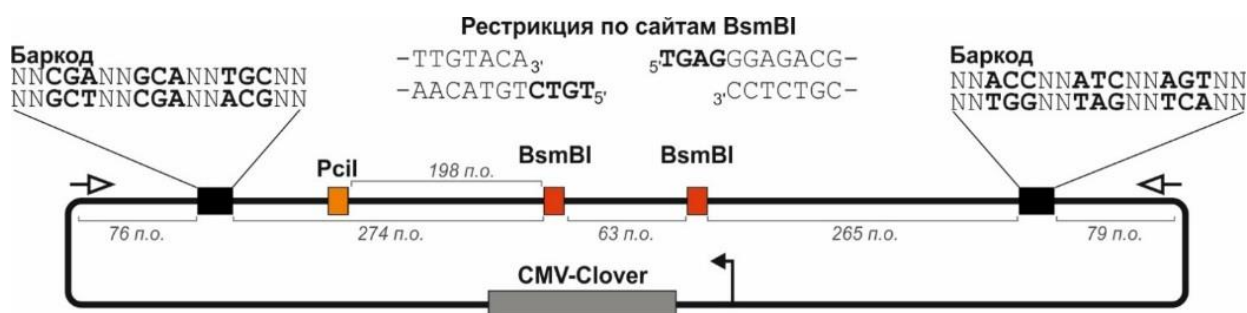
**Приготовление библиотек для NGS.** Для анализа баркодов в конкатемерах было просеквенировано три NGS библиотеки: баркодированная плазмидная библиотека (№1), ПЦР-библиотека слияний между копиями (№2) и inverse PCR-библиотека концевых баркодов (iPCR) (№3). NGS библиотеки №1-3 были подготовлены с помощью кита NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) и просеквенированы на платформе Illumina HiSeq 2500.

**Биоинформатический анализ данных NGS.** На первом шаге риды триммировались, чтобы удалить последовательности, фланкирующие баркоды (cutadapt). На втором шаге концевые баркоды сравнивались с соответствующими паттернами, NNCGANNGCANNTGCNN и NNACTNNGATNNGGTNN, и риды с одинаковыми баркодами объединялись вместе. На третьем шаге баркоды соединялись в пары и отфильтровывались по числу ридов. Финальные пары баркодов были визуализированы модулем Network (<http://visjs.org/>).

**Секвенирование PacBio SMRT.** Высокомолекулярная ДНК выделялась из фрагмента тканей (50 мг) от эмбриона #8 и секвенировалась на платформе PacBio (Novogene, Гонконг). Данные PacBio выравнивались на последовательность транскрипта в NCBI megablast. Было обнаружено 76 уникальных ридов разной длины. Аннотация и анализ финальных последовательностей ридов были проведены вручную, с помощью выравниваний ридов в NCBI blast и поиска баркодов в библиотеках №1-3.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**Получение трансгенных эмбрионов со встройками конкатемеров.** Генетическая конструкция состояла из плазмиды с репортерным геном *Clover* и спейсера с парой баркодов, разделенных сайтом рестрикции *BsmBI*. Для ПЦР баркодированного спейсера 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. Баркодированный ПЦР-продукт был выделен из геля и встроен в плазмиду pcDNA3-Clover (Addgene #40259) (6774 п.о.) (Рис. 1). NGS анализ баркодированной плазмидной библиотеки показал, что она состоит из 12657 уникальных молекул.



**Рисунок 1.** 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.

Плазмидная библиотека была линейаризована рестриктазой *BsmBI*, так что баркоды оказались на расстоянии около 270 п.о. от концов транскрипта (Рис. 1). Линейаризованную библиотеку 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. Суммарно было прогенотипировано 50 эмбрионов, из которых 10 оказались позитивны по встройке трансгена (20%).

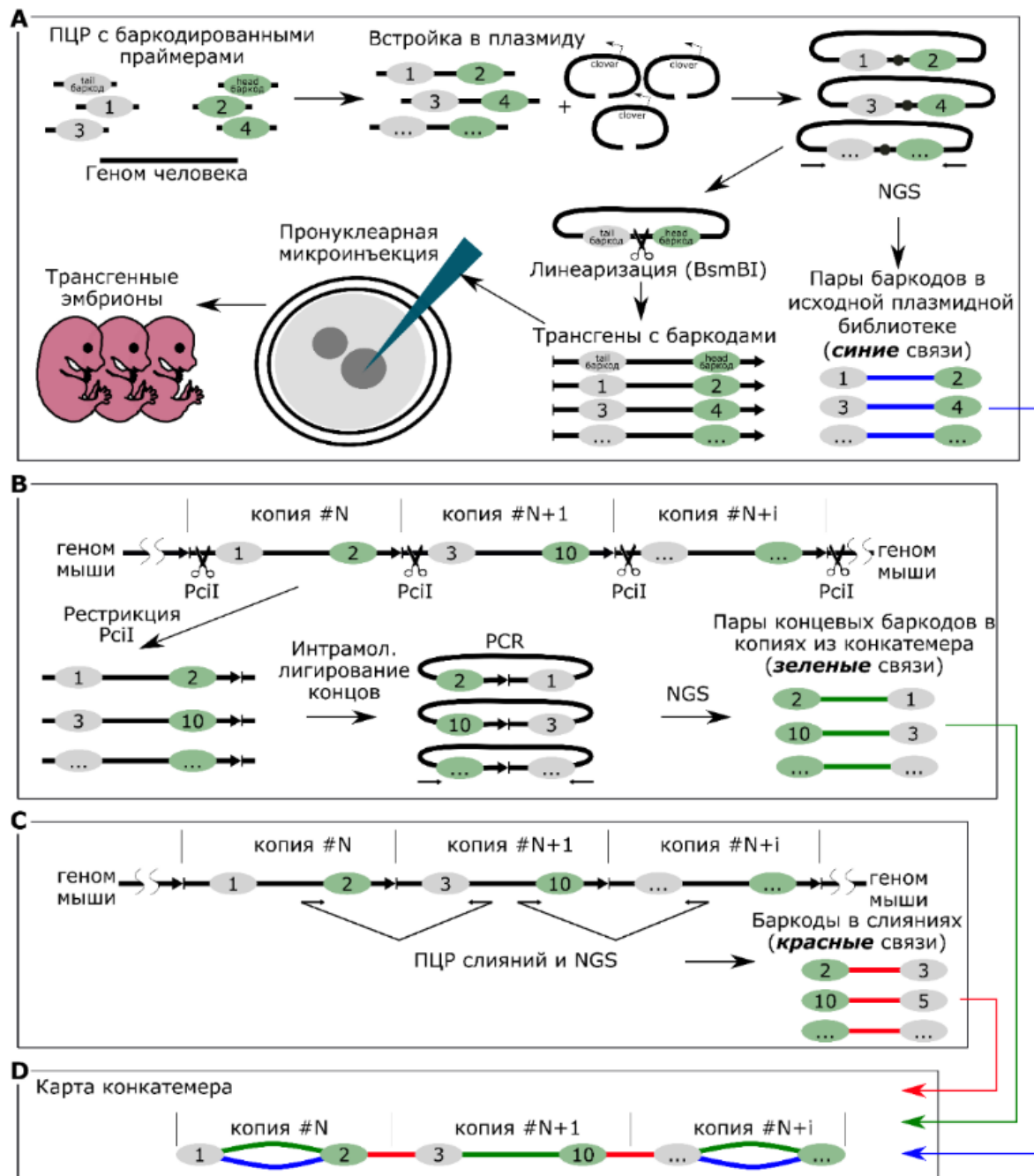
Для определения копийности трансгенов в геноме эмбрионов использовался метод droplet digital PCR (ddPCR). Этот метод позволяет подсчитать абсолютное и относительное число копий целевого участка в растворе геномной ДНК. В ходе дизайна праймеров и зондов для ddPCR были выбраны два участка: центральная часть гена Clover (трансген) и референсный ген мыши *Emid1* (контроль), который присутствует в гаплоидном геноме мыши в единственной копии. Копийность конкатемеров варьировала от одной до нескольких сотен копий.

**NGS анализ баркодов в конкатемерах.** Для секвенирования концевых баркодов в конкатемерах из 10 эмбрионов нами были приготовлены две библиотеки ПЦР-продуктов (Рис. 2). Во-первых, мы амплифицировали участки слияний между копиями в ориентации «голова-к-хвосту» (библиотека слияний). Во-вторых, мы использовали метод inverse PCR (iPCR), чтобы получить информацию о баркодах на концах копий (библиотека концевых баркодов). Рестрикция с PciI позволила разделить конкатемеры на отдельные копии, которые затем лигировались в кольцо. ПЦР-библиотеки были подготовлены по стандартному протоколу и просеквенированы на платформе Illumina HiSeq 2500. Данные NGS для трех баркодированных библиотек (исходная плазмидная библиотека, ПЦР-библиотека слияний, iPCR-библиотека концевых баркодов) были объединены вместе и визуализированы в графическом формате (Рис. 2).

На Рис. 3 представлены карты связей баркодов в эмбрионе #1 (15 копий) в качестве наглядной демонстрации. Карты конкатемеров состоят из двух элементов: узлы-баркоды под уникальными номерами, а также разноцветные линии-связи между баркодами. Связи соединяют баркоды, которые были обнаружены вместе в NGS-библиотеках. **Синие** связи соответствуют исходной плазмидной библиотеке, то есть баркодам в инъецированных молекулах. **Зеленые** связи объединяют баркоды из iPCR-библиотеки, то есть те баркоды, которые мы видим в эмбрионах. Понятно,

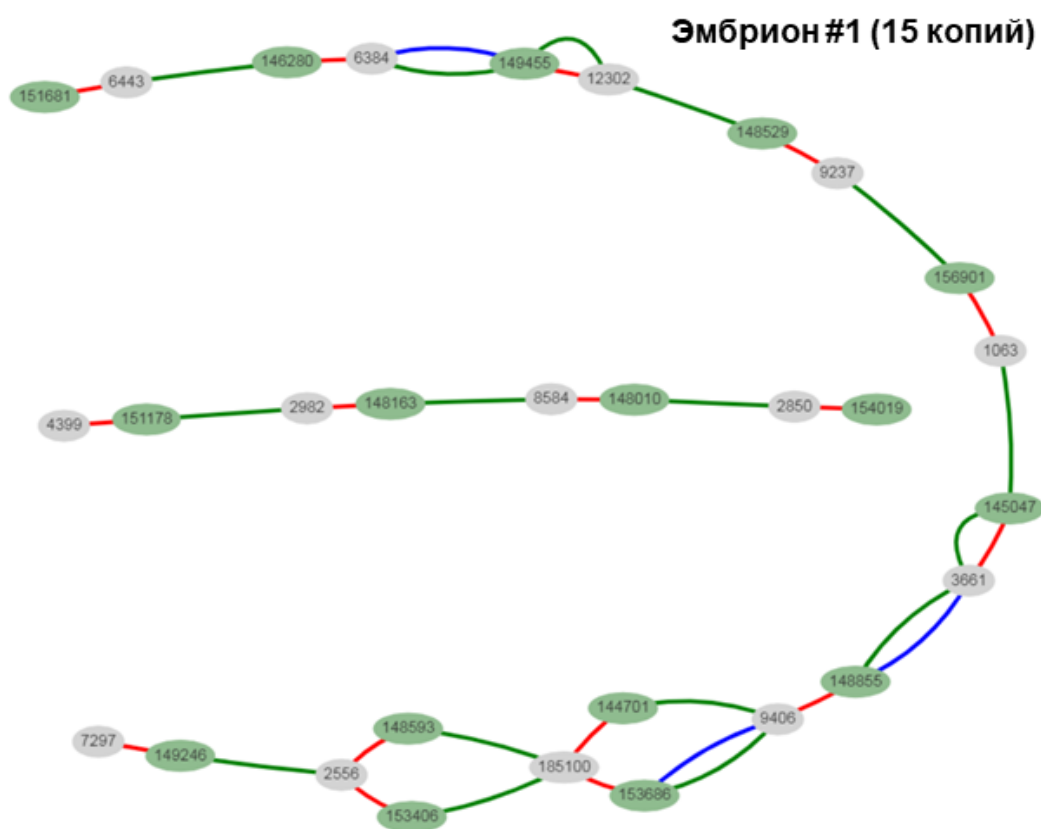


что в норме баркодные пары из плазмидной библиотеки должны обнаруживаться и в эмбрионах, однако таких **сине-зеленых** связей оказалось мало из-за рекомбинации баркодов (см. ниже). Третий тип связей, короткие **красные**, обозначает слияния «голова-к-хвосту» между копиями из ПЦР-библиотеки слияний.



**Рисунок 2.** Схема секвенирования баркодов в конкатемерах. (A) Схема клонирования плазмидной библиотеки и ее микроинъекции в пронуклеус. (B) Секвенирование баркодов на концах копий в конкатемерах. NGS библиотека ПЦР-продуктов создана с помощью inverse PCR (iPCR). (C) Секвенирование баркодов в слияниях между копиями. (D) Итоговый элемент карты конкатемеров.

Всего нами было обнаружено 1135 уникальных трансгенных копий (*зеленые* связи) в 10 эмбрионах (Таблица 1). Объединение данных из трех библиотек позволило нам построить достаточно протяженные карты баркодных связей (Рис. 3).



**Рисунок 3.** Карта связей баркодов в конкатемере из эмбриона #1 (15 копий).








Для валидации результатов NGS мы провели секвенирование генома эмбриона #8 независимым методом - Pacific Biosciences Single Molecule Real-Time (SMRT) sequencing. Этот метод секвенирования третьего поколения позволяет прочитывать фрагменты ДНК длиной в несколько десятков килобаз без использования ПЦР-амплификации при подготовке проб и в процессе секвенирования. Геном эмбриона #8 был просеквенирован с трехкратным покрытием. После фильтрации геномных данных мы нашли 76 уникальных ридов, содержащих последовательность 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 линейаризованного трансгена составляет около 6700 п.о., большинство ридов содержали 2-4 копии из конкатемера. В конкатемере

этого эмбриона было аннотировано примерно 56 копий трансгена, ~20 делеций, 8 палиндромных слияний и 4 паттерна «elongation beyond original broken end» (EBOBE) (Рис. 6). Важно, что обнаруженные баркодные связи подтверждали наши NGS карты.

Таким образом, сгенерированные нами карты связей между баркодами верно отражают взаимное расположение ДНК-баркодов в геномах эмбрионов и могут быть использованы для поиска закономерностей в строении конкатемеров.

**Анализ карт конкатемеров: конкатемеры образуются без участия *de novo* амплификации.** Баркодирование трансгенов помогло выяснить, участвует ли *de novo* амплификация по механизму катящегося кольца в конкатемеризации. Мы сопоставили данные копийности для 10 эмбрионов, рассчитанные с помощью ddPCR, и данные, полученные путем подсчета уникальных *зеленых* связей на картах конкатемеров (Таблица 1). В случае амплификации значительная часть копий в конкатемерах должна нести одинаковые баркоды. Исходя из наших данных, во всех эмбрионах количество копий трансгена приблизительно соответствовало числу уникальных *зеленых* связей. Ни в одном из эмбрионов копийность ddPCR не превышала значительно число уникальных копий, что говорит об отсутствии амплификации (Smirnov et al., 2020).

**Таблица 1.** Частота различных паттернов связей на картах конкатемеров.

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
dNJ паттерн 1		2 (12%)	16 (4%)	8 (5%)	0	0	1 (50%)	7 (1,5%)	10 (9%)	0	0	44 (3,5%)
dNJ паттерн 2		0	0	1 (1%)	0	0	0	4 (0,5%)	0	0	0	5 (0,5%)
dNJ паттерн 3		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

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

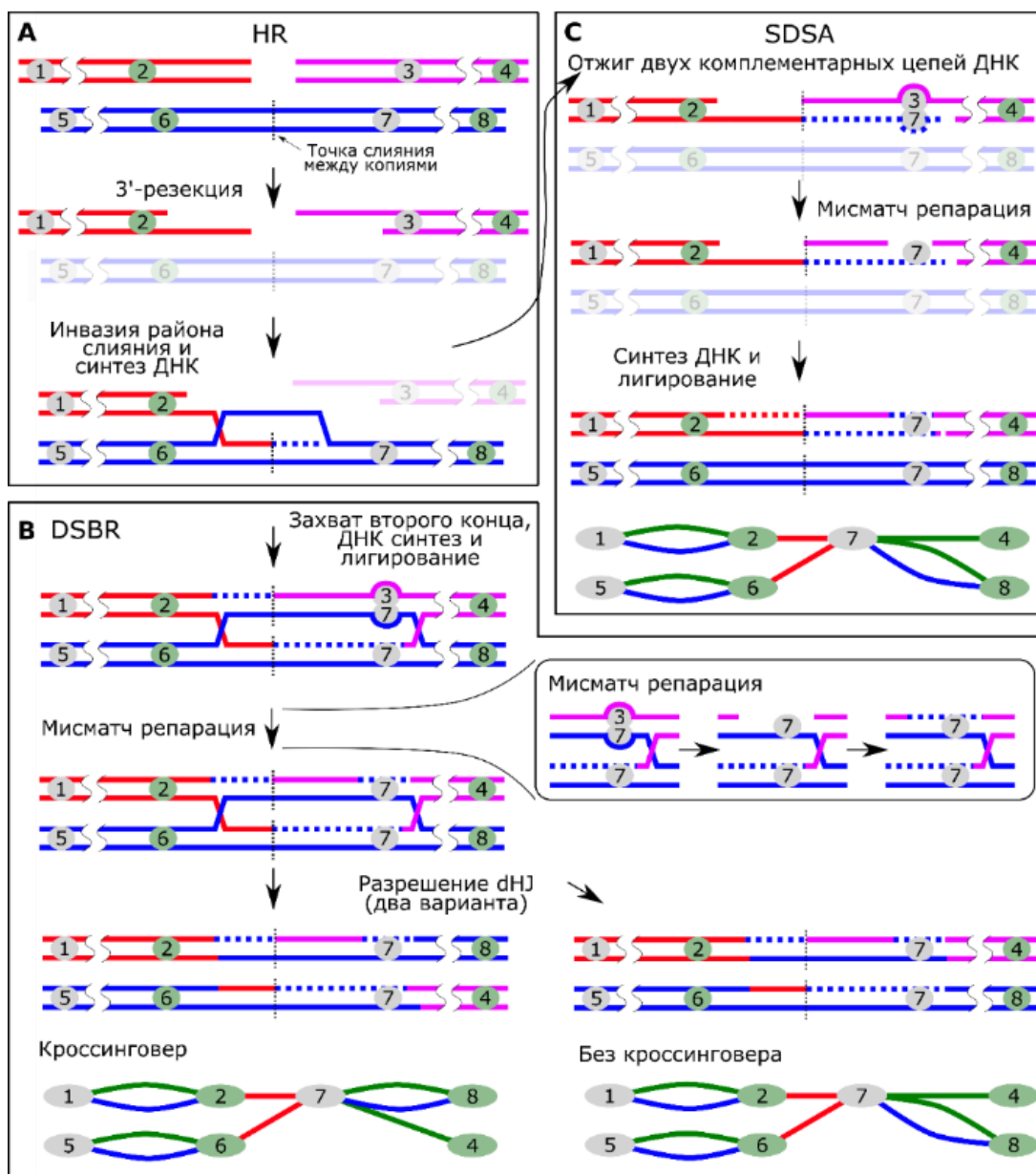
**Концы трансгенов активно процессируются HR.** На картах конкатемеров (Рис. 3) заметно большое количество *зеленых* связей, потерявших парную *синюю* связь. Это означает, что баркодная метка на одном или обоих концах инъецированной молекулы была изменена. Всего нами было проанализировано 1135 копий, среди которых только 222 (20%) сохранили свои исходные комбинации концевых баркодов (*сине-зеленые* связи) (Таблица 1). Можно заключить, что большинство молекул (80%) не лигируются напрямую путем негомологичного или микрогомологичного соединения концов (NHEJ/MMEJ), а процессируются факторами гомологичной рекомбинации (HR), что приводит к «переписыванию» баркодов. По нашим предположениям, это является следствием процессинга баркодированных концов по механизмам synthesis-dependent strand annealing (SDSA) и double-strand break repair (DSBR) (Рис. 4, см. описание).

Заметной чертой карт конкатемеров является наличие огромного числа развилок, когда из одного узла-баркода выходит несколько связей (Рис. 4) (Таблица 1, «развилки»). Это характерно в первую очередь для многокопийных конкатемеров из эмбрионов #2, #3, #4, #7, #8.

Карты для этих эмбрионов похожи на запутанные сети (см. файлы в статье Smirnov *et al.*, 2020), а не на линейные цепочки из малокопийных эмбрионов (Рис. 3). Появление развилок вызвано тем, что одно слияние с парой баркодов может копироваться несколькими независимыми атакующими молекулами. 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, когда за один клеточный цикл единственная молекула может быть донором несколько раз.

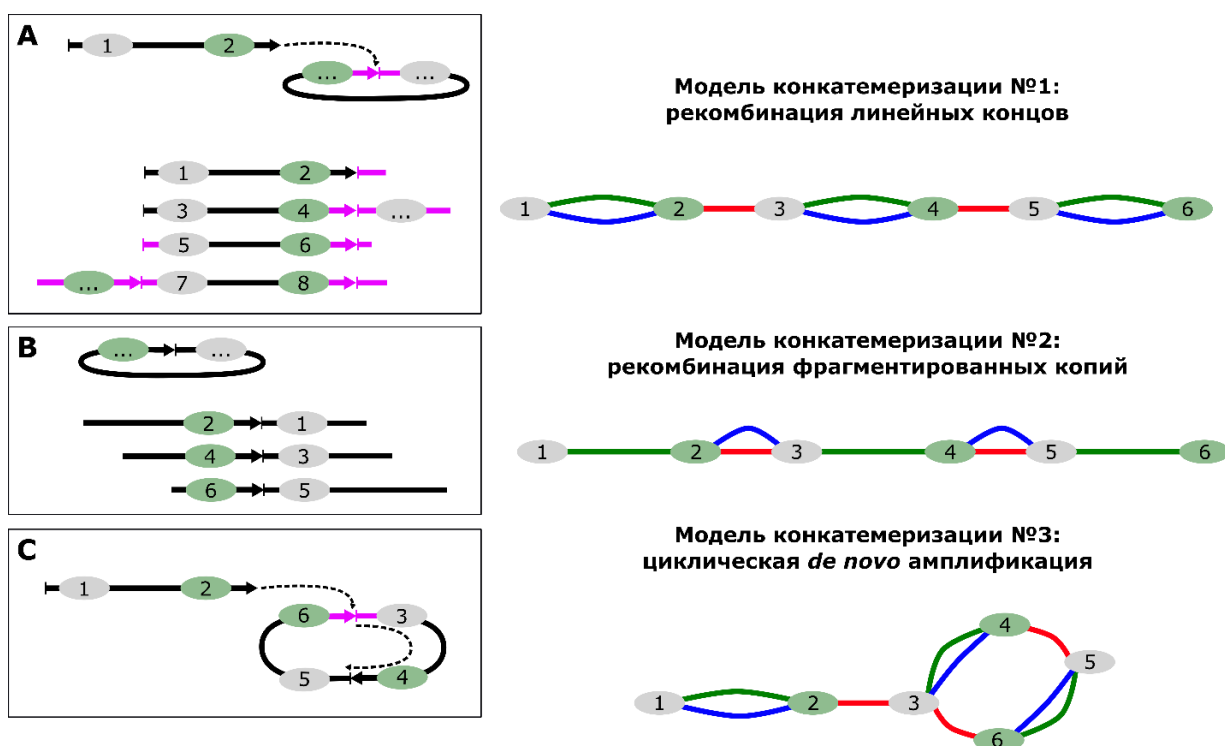
Наши результаты позволили однозначно ответить на основную загадку происхождения конкатемеров (Рис. 5). Изначально мы рассматривали три наиболее вероятных механизма конкатемеризации: рекомбинация за счет линейных концов (модель №1) (Рис. 5А); рекомбинация за счет перекрывающихся фрагментов (модель №2) (Рис. 5В); синтез путем *de novo* амплификации через циклическую копию (модель №3) (Рис. 5С). Данные баркодирования однозначно указывают на то, что конкатемеризация

происходит по модели №1. Об этом говорит преобладание рекомбинации на концах трансгенов (80% *зеленых*, рекомбинированных копий) (Таблица 1). 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. 5B), но таких связей очень мало (Таблица 1). Конкатемеризация по модели №3 показала бы низкое число уникальных баркодов относительно копиности и циклические паттерны баркодных связей (Рис. 5C), но этого не наблюдается.



**Рисунок 4.** Два основных механизма HR (SDSA/DSBR), которые проводят сборку трансгенов в тандем. (А) Общая стадия для двух путей: 3'-резекция двуцепочечных концов линейных молекул (красная и розовая) и поиск

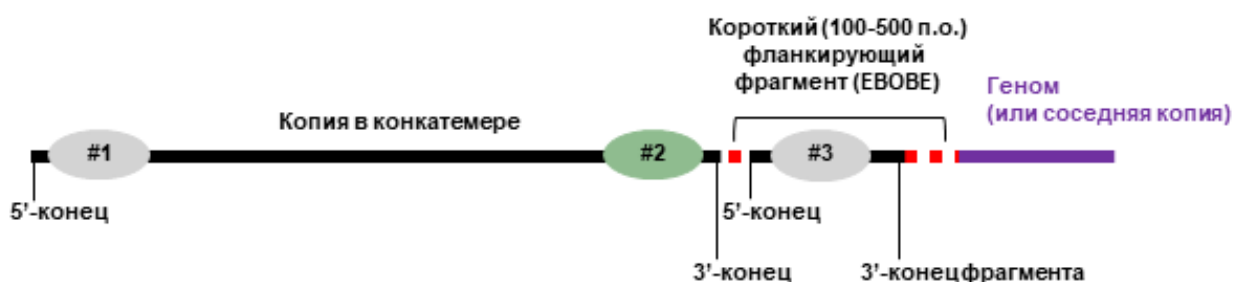
гомологичного района для инвазии. Этим районом обычно является участок слияния между двумя копиями (синяя молекула). (В) Этапы DSBR. После захвата второго резектированного конца формируется двойная холидеевская структура (dHJ), которая может быть разрешена двумя способами (с кроссинговером или без него). (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. После инвазии дуплекса и синтеза в D-петле на конце трансгена появляются несовпадающие баркодные участки, которые исправляются системой мисматч репарации (см. врезку). Таким образом, атакующий трансген всегда копирует баркод в результате репарации мисматча в новосинтезированной цепи.



**Рисунок 5.** Гипотетические модели конкатемеризации (А-С) и их визуализация в формате карт конкатемеров.

Мы также обнаружили и другой интересный признак активности HR - участки дубликации рядом с баркодами, которые иногда встречаются в конкатемерах (Рис. 6). 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. 4). Такой гетерологичный район на 3'-конце не дает копии вернуться для нового раунда инвазии, поэтому он присоединяется к конкатемеру через NHEJ/MMEJ, образуя

дублицированный участок (Рис. 6). 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). В имеющихся данных мы нашли EBOBE в эмбрионе #6 и на границах трансген-геном в эмбрионах #5 и #10. Мы также нашли 4 EBOBE участка в данных RasBio для эмбриона #8. Мы заметили похожие паттерны в ранее опубликованных статьях (Yan *et al.*, 2010), однако в тех случаях авторы объяснили эти паттерны случайной фрагментацией трансгенов или неполным копированием трансгена при встройке. Мы считаем, что EBOBE представляют собой результат взаимодействия терминированных производных HR с путями репарации NHEJ (или MMEJ). В дальнейшем важно будет изучить, насколько этот патологический путь репарации активен в зиготах и других типах клеток.

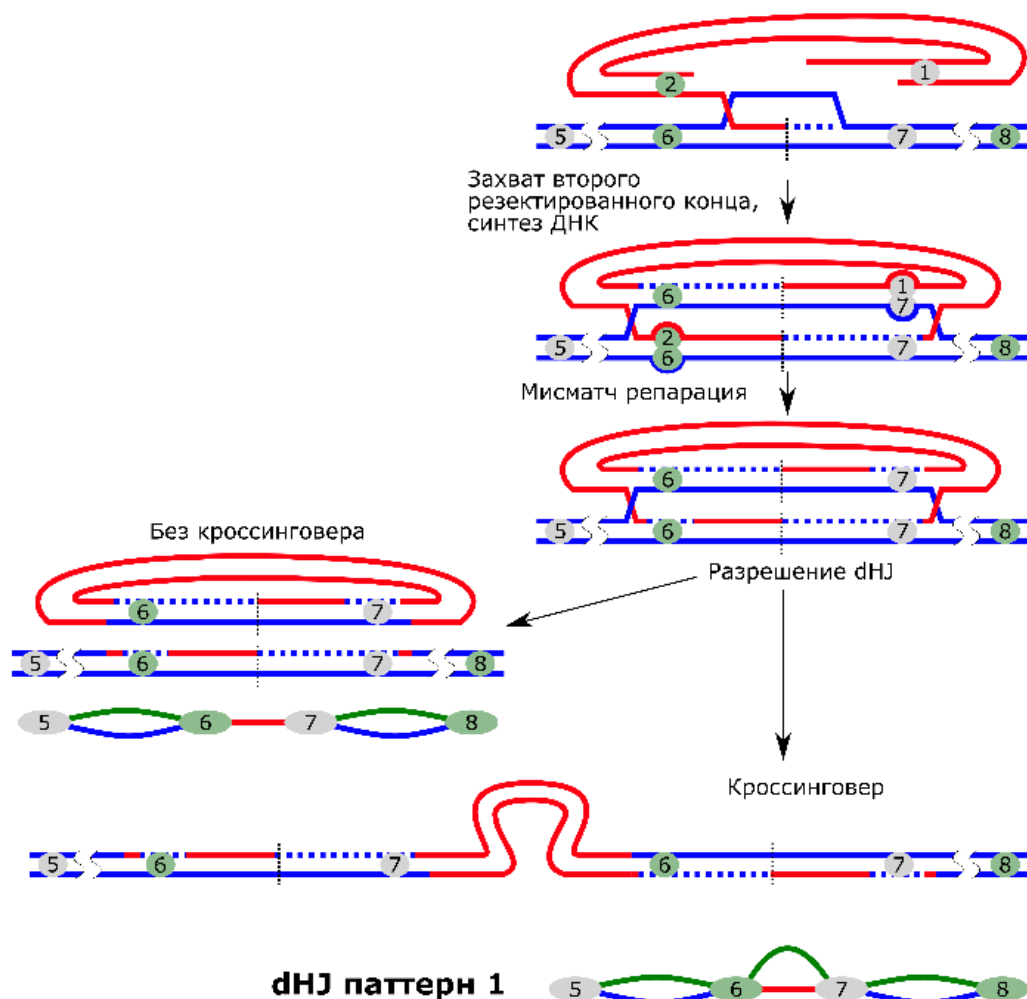


**Рисунок 6.** Схема паттерна EBOBE (elongation beyond original broken end), который формируется после удлинения конца трансгена в D-петле и его лигирования с другими молекулами через NHEJ/MMEJ.

**Индикаторы кроссинговера и dHJ среди паттернов связей.** В большинстве случаев продукты SDSA и DSBR имеют одинаковые паттерны связей на картах конкатемеров (Рис. 4), поэтому разделить их вклад в рекомбинацию невозможно. Однако косвенно на активность DSBR указывает появление продуктов митотического кроссинговера, которые образуются при разрешении двойных холидеевских структур (double Holliday Junction, dHJ) (Рис. 4В). Мы выделили три паттерна связей между баркодами, которые могут объясняться кроссинговером. Самый распространенный - dHJ паттерн 1 (Рис. 7): 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).

В норме частота митотического кроссинговера в эмбриональных стволовых клетках мыши составляет менее 2.5%. По нашим данным частота митотического кроссинговера в эмбрионах составила 5% (Таблица 1). Реальный уровень кроссинговера еще выше, так как многие сложные паттерны разрушаются из-за множественной рекомбинации, которая идет при конкатемеризации. Интересно, что разрешение двойной холидееской структуры (dHJ) с кроссинговером – нежелательная мера в клетке, так как приводит к перестройкам. Вне мейоза система репарации DSB избегает кроссинговера, благодаря геликазе Sgs1/BLM, устраняющей dHJ.



**Рисунок 7.** Разрешение dHJ может приводить к кроссинговеру и особым паттернам связей баркодов. На схеме показан кроссинговер между двумя атакующими концами трансгена (красные) и слиянием (синие) в конкатемере (dHJ паттерн 1).



**Роль NHEJ в конкатемеризации.** Чтобы оценить вклад NHEJ в соединение копий мы выбрали три эмбриона с высокой копийностью (#2, #3, #7) и просеквенировали ПЦР-продукты, соответствующие району слияния между копиями. Все сигнатуры NHEJ были привязаны к близлежащим баркодам. Всего было обнаружено 103 уникальных варианта слияний, представляющих разнообразные делеции и, в редких случаях, вставки или замены нуклеотидов (Smirnov *et al.*, 2020). Учитывая, что в данных для этих эмбрионов присутствует 1803 уникальных баркода (~900 уникальных молекул трансгенов), можно сказать, что в среднем 1 из 9 инъекцированных молекул участвует в лигировании через NHEJ. На самом деле, это значение должно быть гораздо выше, так как многие варианты слияний повторяют друг друга, а какие-то варианты не попадают в финальный конкатемер. В целом же, по совокупности данных о рекомбинации (*зеленые* копии) и сигнатур NHEJ, мы можем заключить, что более 10% инъекцированных трансгенных молекул предварительно обрабатываются NHEJ и лигируются друг с другом (или внутримолекулярно), а уже затем служат «затравкой» для формирующих конкатемер HR-процессированных копий. Таким образом, NHEJ важен на первом шаге формирования конкатемера, но он почти не присоединяет копии к растущему конкатемеру.

**Локализация встройек конкатемеров в геноме.** Для некоторых из эмбрионов встройка была картирована с помощью TAIL-ПЦР или PacBio. Всего было локализовано 11 границ, их строение мы изучили более детально и нашли участки микрогомологии почти в каждом эмбрионе. Похожие данные были получены и для трансгенных линий из архива нашей лаборатории (Serova *et al.*, 2012; Burkov *et al.*, 2013; Smirnov *et al.*, 2018). Проанализировав 8 трансген-геномных границ в 6 трансгенных линиях мышей, мы также обнаружили частое присутствие микрогомологии и высокую частоту интеграции в интроны генов (3 случая из 6). Примерно половина трансген-геномных границ несла признаки MMEJ, при этом средний размер микрогомологии составил около 3 нуклеотидов. Как показывают наши данные, MMEJ в зиготах работает гораздо активнее, чем в других типах клеток, где обычно служит «запасным» механизмом репарации DSB.

## **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 является перспективным инструментом для дальнейших исследований конкатемеризации в культурах клеток и эмбрионов.

## **ВЫВОДЫ**

1. Применение баркодированной библиотеки трансгенов с разнообразием 12.7 тыс. молекул позволяет оценивать процессы конкатемеризации в зиготе.
2. Негомологичное соединение концов участвует в лигировании как минимум 10% инъецированных копий, которые служат в качестве матриц для HR, но почти не встраиваются в конкатемер.
3. 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 гибриднему 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 РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. **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.
2. **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 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.
4. 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.