

Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук

На правах рукописи



Андрианова Мария Александровна

Характеристики точечного мутагенеза в раковых клетках человека

03.01.09 – математическая биология, биоинформатика

Диссертация на соискание ученой степени кандидата биологических наук

Научный руководитель:

доктор биологических наук

Георгий Александрович Базыкин

Москва – 2019

Оглавление

Введение.....	5
Актуальность темы исследования	5
Степень разработанности темы.....	6
Цели и задачи исследования.....	7
Научная новизна и практическая значимость.....	7
Основные положения, выносимые на защиту	8
Личный вклад автора в исследование.	9
Степень достоверности и апробация результатов.....	9
Структура и объем диссертации	10
Список публикаций по теме диссертации.....	10
Список сокращений.....	11
Глава 1. Обзор литературы.....	13
1.1 Мутационные подписи.....	14
1.2 Репликация как источник мутаций	16
1.3 Система репарации неспаренных нуклеотидов	18
1.4 Эксцизионная система репарации нуклеотидов, ассоциированная с транскрипцией.....	19
1.5 Распределение мутаций вдоль генома в раковых образцах	19
1.6 Асимметрия мутаций между цепями ДНК	20
1.7 Изменение мутационных процессов в ходе развития опухоли.....	21
Глава 2. Асимметричная работа системы репарации неспаренных нуклеотидов на лидирующей и отстающей цепях ДНК.....	25
2.1 Материалы и методы.....	25
2.1.1 Мутационные данные	25
2.1.2 Определение соматических мутаций в ультрамутабельных раках	26
2.1.3 Репликационная асимметрия	26
2.1.4 Данные секвенирования фрагментов Оказаки	27
2.1.5 Анализ вставок и выпадений	27
2.1.6 Обогащение подписью АРОВЕС	28
2.1.7 Модель для мутационных смещений между цепями	28
2.2 Результаты.....	29
2.2.1 Паттерны мутаций в раковых образцах с инактивированным экзонуклеазным доменом полимеразы эpsilon или дельта.....	29
2.2.2 Цепеспецифичные паттерны мутаций в раковых образцах с инактивированным экзонуклеазным доменом полимеразы эpsilon или дельта.....	31
2.2.3 Мутационные спектры в раковых образцах с инактивированной MMR	35

2.2.4 Мутационная асимметрия в раковых образцах с инактивированной MMR	38
2.2.5 Репликационная асимметрия в раковых образцах с инактивированной MMR соответствует асимметрии в раковых образцах с мутированной полимеразой дельта	42
2.3 Обсуждение результатов.....	44
Глава 3. Свойства раковых образцов с биаллельной инактивацией MMR и мутацией в гене полимеразы эpsilon	51
3.1 Материалы и методы.....	51
3.1.1 Оценка микросателлитной нестабильности	51
3.1.2 Описание семьи и характеристики опухолей.....	52
3.1.3 Экзомное и геномное секвенирование.....	53
3.1.4 Определение происхождения мутации с помощью генотипирования.....	54
3.1.5 ПЦР длинных фрагментов и вложенная ПЦР.....	54
3.1.6 Определение участков гомозиготности.....	55
3.1.7 Иммуногистохимия белков системы репарации неспаренных нуклеотидов	55
3.1.8 Определение цисспецифичности мутаций в раковой опухоли	55
3.1.9 Доступные данные из ранее опубликованных работ	55
3.1.10 Оценка доли мутаций под положительным отбором в генах-супрессорах	56
3.2 Результаты.....	56
3.2.1 Участки повышенной гомозиготности в исследуемых пациентах	56
3.2.2 Определение каузативного гомозиготного варианта в гене <i>PMS2</i>	57
3.2.3 Отсутствие белка PMS2 в опухолях и других клетках пациентов.....	58
3.2.4 Геномная и микросателлитная стабильность.....	59
3.2.5 Независимое происхождение стоп-кодона в гене <i>PMS2</i>	59
3.2.6 Ультрамутабельный фенотип опухолей с bMMRD и соматической мутацией в гене полимеразы эpsilon	60
3.2.7 Репликационная асимметрия опухолей с bMMRD и соматической мутацией в гене полимеразы эpsilon	62
3.2.8 Ультрамутабельные глиобластомы несут как минимум в 2 раза меньшую долю мутаций под положительным отбором, чем другие глиобластомы.....	64
3.3 Обсуждение результатов.....	67
Глава 4. Мутационные подписи базальноклеточной карциномы	68
4.1 Материалы и методы.....	68
4.1.1 Мутационные данные	68
4.1.2 Сбор образцов и секвенирование	68
4.1.3 Экспрессия в образцах меланомы и ВСС	69
4.1.4 Подсчет числа замен.....	69
4.2 Результаты и обсуждение	69

Глава 5. Изменение мутационных подписей в ходе развития раковой опухоли	75
5.1. Материалы и методы	75
5.1.1 Данные	75
5.1.2 Оценка доли раковых клеток, содержащих мутацию	76
5.1.3 Подписи в клональных и субклональных мутациях	77
5.1.4 Численное моделирование изменения интенсивности мутационных подписей	78
5.2 Результаты и обсуждение	78
5.2.1 Паттерны изменения активности основных мутационных процессов при развитии опухоли	78
5.2.2 Изменение распределения локальных скоростей мутирования при развитии рака	83
Основные результаты и выводы	86
Приложения	87
Благодарности	103
Список литературы	104

Введение

Актуальность темы исследования

Рак – это группа заболеваний, характеризующихся бесконтрольным ростом и делением клеток, которые могут проникать в близлежащие части организма и распространяться в другие органы. Раковые образования очень гетерогенны: даже в одном конкретном органе они отличаются между пациентами по фенотипам и генотипам, по скоростям развития и прогнозам. Считается, однако, что все они вызваны генетическими изменениями, приводящими к нарушению пролиферации и дифференцировки клеток.

Широкое разнообразие раковых опухолей, сложность их диагностики и лечения привели к тому, что в период активного развития технологий секвенирования были прочитаны геномные и экзомные последовательности огромного количества раковых образцов. Более того, объем таких данных продолжает увеличиваться с каждым годом. В настоящий момент молекулярная диагностика является неотъемлемой частью постановки диагноза для многих типов опухолей.

В результате секвенирования определяется нуклеотидная последовательность ДНК ракового образца. При сравнении ее с последовательностью здоровых клеток того же индивида можно определить мутации, характерные непосредственно для ракового образца. Полученные таким образом мутации можно исследовать с разных точек зрения. Например, можно изучать гены и функциональные участки ДНК, обогащенные мутациями в определенном типе рака, то есть заниматься поиском потенциальных генов и мутаций-драйверов развития опухоли данного типа. Это одна из необходимых прикладных задач молекулярной диагностики.

Известно, что большинство раковых образцов характеризуются достаточно высокими скоростями мутирования. Среди всех мутаций, наблюдаемых в раковом образце, в среднем лишь около 4 мутаций являются драйверами развития опухоли [1], то есть мутациями, которые находились под положительным отбором и привели к селективному преимуществу клетки – увеличению выживаемости и пролиферации. Остальные мутации являются «пассажирскими», и накапливались без прямого влияния положительного отбора [2]. Поскольку большинство из таких «пассажирских» мутаций не подвергались действию отбора, исследование их свойств является очень интересной задачей: они содержат информацию о мутационных процессах, происходивших в опухоли. Исследование спектров мутирования (то есть типов нуклеотидных замен и их контекстной зависимости), особенностей распределения мутаций вдоль генома и между цепями ДНК помогает лучше понять процесс возникновения и развития опухоли. Это дает представление о мутационных процессах, приводящих к повреждениям ДНК, а также о системах репарации, ответственных за исправление этих повреждений.

В данной работе проведен подробный анализ характеристик точечных мутаций в различных раковых образцах. Во второй главе диссертации с помощью анализа характеристик мутаций и их геномного распределения исследуются особенности молекулярных механизмов репликации и репарации в клетках человека, в том числе и в здоровых. В третьей главе аналогичные данные используются для уточнения причин и особенностей развития раковых опухолей в двух конкретных случаях детских раков мозга. В четвертой главе анализ распределения и свойств мутаций помогает подробнее разобраться в особенностях возникновения и репарации повреждений в отдельном типе рака, а именно в базальноклеточной карциноме. В пятой главе показано, что подробное исследование мутаций раковых образцов помогает восстановить историю развития опухоли, а именно узнать, какие мутационные процессы действовали на разных этапах развития опухоли, как это влияет на мутационный профиль и распределение мутаций по геному.

Степень разработанности темы

В настоящий момент имеется огромное количество работ с исследованием мутационных паттернов как в отдельных раковых образцах, так и в выборках образцов определенного типа рака. Известны определенные характерные черты мутационных паттернов разных типов рака; например, известно, что мутации в меланоме в основном представляют собой замены цитозина на тимин (C→T), вызванные воздействием ультрафиолета (УФ), а в раке легких преобладают мутации гуанина в тимин (G→T), вызванные воздействием табачного дыма. Одним из наиболее значимых достижений в данной теме является работа, в которой были выделены так называемые «мутационные подписи» [3; 4], то есть спектры мутаций, характерные для определенных мутационных процессов. Понимание того, какой мутационный процесс является источником определенной подписи – вопрос крайне интересный как с точки зрения фундаментальной, так и прикладной науки. Однако для многих подписей этиология до сих пор остается неизвестной (<https://cancer.sanger.ac.uk/cosmic/signatures>). Более того, каждый раковый образец уникален; набор подписей в нем тоже уникален, и может содержать как известные подписи, так и новые, еще не изученные. Наконец, неописанные ранее с этой точки зрения типы рака или подтипы интересны с точки зрения мутационных паттернов в них.

Известно, что мутации в раковых образцах очень неравномерно распределены вдоль генома. Частота мутаций в данном участке генома объясняется особенностями мутирования и репарации, а также эпигенетическими характеристиками этого участка. Принимая во внимание все известные эпигенетические разметки, можно объяснить около 90% изменчивости скорости мутирования вдоль генома в раковых образцах [5]. Также было показано, что мутации распределены неравномерно между цепями ДНК: скорость мутирования различается между

лидирующей и отстающей цепями при репликации, а также между матричной и нематричной цепями при транскрипции [6]. Не для всех этих неравномерностей хорошо понятна их природа, и это безусловно интересный вопрос, раскрывающий особенности функционирования клеточных механизмов.

Как было упомянуто ранее, большинство мутаций в раковых образцах являются «пассажирскими», то есть не являются драйверами развития или субклональной экспансии в опухолях. Однако таких «пассажирских» мутаций в раковых образцах тысячи, и они содержат важную информацию об истории развития опухоли. Подробный анализ таких мутаций показал, что интенсивность воздействия разных мутационных процессов изменяется в ходе развития опухолей [7]. Однако остается неизвестным, насколько эти изменения характерны для отдельных раковых образцов и как они влияют на распределение мутаций вдоль генома и между цепями ДНК на разных этапах развития опухоли.

Цели и задачи исследования

Целью данной работы являлось подробное изучение характеристик точечного мутагенеза в образцах раковых опухолей человека: мутационных спектров, распределения мутаций вдоль генома и между цепями ДНК.

Для достижения поставленной цели были решены следующие задачи:

- 1) описать мутационные спектры и асимметрию точечных мутаций между цепями ДНК в раковых образцах с инактивированной системой репарации неспаренных нуклеотидов и мутациями в экзонуклеазных доменах основных репликативных полимераз эпсилон и дельта;
- 2) описать мутационные спектры и асимметрию точечных мутаций в раковых образцах детской глиобластомы с ультрамутабельным фенотипом, а также проанализировать положительный отбор в этих образцах;
- 3) описать мутационные спектры и асимметрию точечных мутаций в раковых образцах самого частого рака кожи - базальноклеточной карциномы;
- 4) описать мутационные спектры, распределение мутаций вдоль генома, а также их асимметрию между цепями ДНК на разных этапах развития раковой опухоли.

Научная новизна и практическая значимость

В предыдущих работах было показано, что скорость накопления мутаций различается между лидирующей и отстающей цепями ДНК, в том числе и в раковых клетках человека. Это называется репликационной асимметрией. В частности, она была показана для раковых образцов

с инактивированной системой репарации неспаренных нуклеотидов. В настоящей работе впервые показано, что данная асимметрия вызвана тем, что полимераза дельта при репликации отстающей цепи вносит больше ошибок, чем полимераза эpsilon при репликации лидирующей цепи. Также впервые показано, что система репарации неспаренных нуклеотидов в клетках человека компенсирует неравномерность скоростей мутирования между лидирующей и отстающей цепями, удаляя больше неправильно спаренных оснований на отстающей цепи.

Благодаря использованию информации о репликационной асимметрии удалось показать, что ультрамутабельный фенотип в исследуемых образцах детских глиобластом с врожденной биаллельной инактивацией системы репарации неспаренных нуклеотидов вызван обилием мутаций на лидирующей цепи ДНК ввиду инактивации экзонуклеазного домена полимеразы эpsilon и системы репарации. Подобные анализы могут быть использованы и для других раковых образцов с подозрением на инактивацию системы репарации неправильно спаренных нуклеотидов или экзонуклеазных доменов репликативных полимераз для уточнения молекулярного механизма развития опухоли.

В работе впервые описаны особенности возникновения повреждений и их репарации в базальноклеточной карциноме, для чего использован подробный анализ спектров мутирования и транскрипционной асимметрии в данном типе рака, а также сравнение с другим хорошо изученным типом рака кожи – меланомой.

Более того, проанализированы спектры мутирования и распределение мутаций вдоль генома и между цепями ДНК для мутаций, возникших на разных стадиях развития раковой опухоли. Впервые показано, что изменение интенсивности мутационного процесса можно увидеть внутри одного конкретного ракового образца даже при секвенировании целого фрагмента опухоли. Также показано, что мутационный процесс, действующий в опухоли в конкретный момент времени, изменяет профиль мутаций и их распределение по геному согласно характерным особенностям этого мутационного процесса. Знания об изменении мутационного спектра и локальных скоростей мутирования в ходе развития раковой опухоли могут быть использованы для поиска субклональных драйверных мутаций.

Основные положения, выносимые на защиту

1. Полимераза дельта, реплицирующая отстающую цепь ДНК, вносит больше ошибок при репликации, чем полимераза эpsilon на лидирующей цепи. Система репарации неспаренных нуклеотидов в клетках человека компенсирует неравномерность скоростей мутирования между лидирующей и отстающей цепями, удаляя больше неправильно спаренных оснований на отстающей цепи.

2. Врожденная биаллельная инактивация системы репарации неспаренных нуклеотидов и последующая соматическая мутация в экзонуклеазном домене полимеразы эpsilon приводят к ультрамутабельному фенотипу опухоли ввиду избытка повреждений на лидирующей цепи ДНК, внесенных полимеразой эpsilon и не исправленных системой репарации неспаренных нуклеотидов.

3. Доля мутаций под положительным отбором в ультрамутабельных глиобlastомах как минимум в 2 раза меньше, чем в других глиобlastомах.

4. В базальноклеточной карциноме большая доля мутаций вызвана УФ, чем в меланоме. Эффективность репарации этих мутаций также выше в базальноклеточной карциноме. Мутации цитозина в аденин в базальноклеточной карциноме и в меланоме имеют разную природу: в базальноклеточной карциноме они преимущественно являются мутациями окисления гуанина, в то время как в меланоме – смесью мутаций окисления гуанина и мутаций цитозина под действием УФ.

5. В ходе развития отдельно взятой раковой опухоли наблюдается изменение в относительной активности разных мутационных процессов. Это приводит к изменению мутационных спектров, а также распределения мутаций вдоль генома и между цепями ДНК во времени.

Личный вклад автора в исследование.

Все результаты, включенные в диссертацию, получены лично соискателем или при его непосредственном участии, за исключением разделов 3.2.1-3.2.5, выполненных соавторами публикации. Обсуждение и интерпретация результатов осуществлялись автором совместно с научным руководителем и соавторами публикаций.

Степень достоверности и апробация результатов

По материалам диссертации опубликовано три статьи в рецензируемых журналах. Результаты работы были представлены на Московской международной конференции по биоинформатике (Bioinformatics'16) и Московской международной конференции по вычислительной молекулярной биологии (Moscow Conference on Computational Molecular Biology – MCCMB'17), на международной встрече Abscam'17 “Геномные перестройки и мутационные подписи в развитии и раке” (Бостон, США), на международной встрече Human Genome Meeting 2017 “От геномики к терапии” (Барселона, Испания), на 7-й немецко-российской неделе молодых исследователей 2017 (Москва, Россия), на конференции Gen-Y 2017 (Сочи, Россия).

Структура и объем диссертации

Диссертация изложена на 113 страницах машинописного текста и содержит следующие разделы: введение, обзор литературы, результаты и обсуждение в четырёх главах, основные результаты и выводы. В конце приведён список литературы. Материал включает 37 рисунков и 9 таблиц в основном тексте, 9 рисунков и 7 таблиц в приложении, а также список литературы, содержащий 128 ссылок.

Список публикаций по теме диссертации

По теме диссертации опубликовано три статьи в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (PubMed, WoS и Scopus):

1. Bonilla X, Parmentier L, King B, Bezrukov F, Kaya G, Zoete V, Seplyarskiy VB, Sharpe HJ, McKee T, Letourneau A, Ribaux PG, Popadin K, Basset-Seguín N, Chaabene RB, **Andrianova MA**, Guipponi M, Garieri M, Verdán C, Grosdemange K, Sumara O, Eilers M, Aifantis I, Michielin O, de Sauvage FJ, Antonarakis SE, Nikolaev SI. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma // *Nature Genetics*. 2016. Vol 48, №4. P. 398-406.

2. **Andrianova MA**, Chetan GK, Sibin MK, Mckee T, Merkler D, Narasinga RK, Ribaux P, Blouin JL, Makrythanasis P, Seplyarskiy VB, Antonarakis SE, Nikolaev SI. Germline PMS2 and somatic POLE exonuclease mutations cause hypermutability of the leading DNA strand in biallelic mismatch repair deficiency syndrome brain tumours // *Journal of Pathology*. 2017. Vol 243, № 3. P. 331-341.

3. **Andrianova MA**, Bazykin GA, Nikolaev SI, Seplyarskiy VB. Human mismatch repair system balances mutation rates between strands by removing more mismatches from the lagging strand // *Genome Research*. 2017. Vol 27, №8. P. 1336-1343.

Кроме того, результаты работы опубликованы в сборниках тезисов международных и российских конференций:

1. Andrianova MA, Bazykin GA, Bredikhin D, Seplyarskiy VB. Changes in mutational processes and mutations patterns during cancer development // *Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB' 17)*, Moscow, Russia, July 27-30, 2017. <http://mccmb.belozersky.msu.ru/2017/proceedings/abstracts/116.pdf>

2. Andrianova MA. Changes in mutational processes during cancer development // *The Seventh German-Russian Week of the Young Researcher "Computational biology and biomedicine"*, Moscow, Russia, September 11-14, 2017. <https://www.dwih-moskau.org/files/2018/11/7-WdjW.pdf>

Список сокращений

п.н. – пар нуклеотидов

ПЦР – полимеразная цепная реакция

УФ – ультрафиолет

BLCA – рак мочевого пузыря

bMMRD (biallelic mismatch repair deficiency) – врожденная биаллельная инактивация гена системы репарации неспаренных нуклеотидов

BRCA – рак груди

CCF (cancer cell fraction) – доля раковых клеток, несущих мутацию

CESC – плоскоклеточная карцинома шейки матки

COAD – аденокарцинома толстой кишки

fp (fork polarity) – направление вилки репликации

FPKM (fragments per kilobase per million mapped reads) – число фрагментов на 1000 нуклеотидов на миллион закартированных чтений

GBM – глиобластома

HNSC – рак головы и шеи

LIHC – рак печени

LUAD – аденокарцинома легкого

LUSC – плоскоклеточный рак легкого

MMR (mismatch repair) – система репарации неспаренных нуклеотидов

MSI (microsatellite instability) – фенотип микросателлитной нестабильности

MSS (microsatellite stability) – фенотип микросателлитной стабильности

OV – рак яичников

READ – рак прямой кишки

RPKM (reads per kilobase per million mapped reads) – число чтений на 1000 нуклеотидов на миллион закартированных чтений

RT (replication time) – время репликации

SKCM – меланома

STAD – рак желудка

vaf (variant allele frequency) – аллельная частота варианта

UCEC – карцинома эндометрия

Глава 1. Обзор литературы

Рак – это совокупность заболеваний, характеризующихся появлением бесконтрольно делящихся клеток. Причина появления состоит в нарушении пролиферации и дифференцировки клеток вследствие генетических изменений. Раковые заболевания очень трудно поддаются лечению. Одной из причин этого является высокая гетерогенность как между типами рака, так и между отдельными образцами внутри одного типа. С развитием технологий секвенирования появились глобальные проекты, которые генерируют огромные каталоги соматических мутаций десятков тысяч раковых образцов [8]. Сравнение их с данными секвенирования здоровых тканей позволило определять генетические изменения, произошедшие в раковых клетках. Это дало возможность изучать механизмы развития раковых опухолей. В частности, это позволило оценить скорость соматического мутагенеза в раковых клетках и показать, что в зависимости от типа рака образец может содержать от 1000 до 20000 соматических точечных мутаций (Рисунок 1.1) [9]. Детские раки мозга и лейкемии обычно имеют самые низкие скорости мутирования, в то время как раки, вызванные воздействием внешних мутагенов, таких как табачный дым или ультрафиолет (УФ), имеют самые высокие скорости мутирования. Однако есть раковые образцы, имеющие очень сильно повышенные скорости мутирования; такое может происходить, например, при инактивации систем репарации или экзонуклеазных доменов репликативных полимераз.

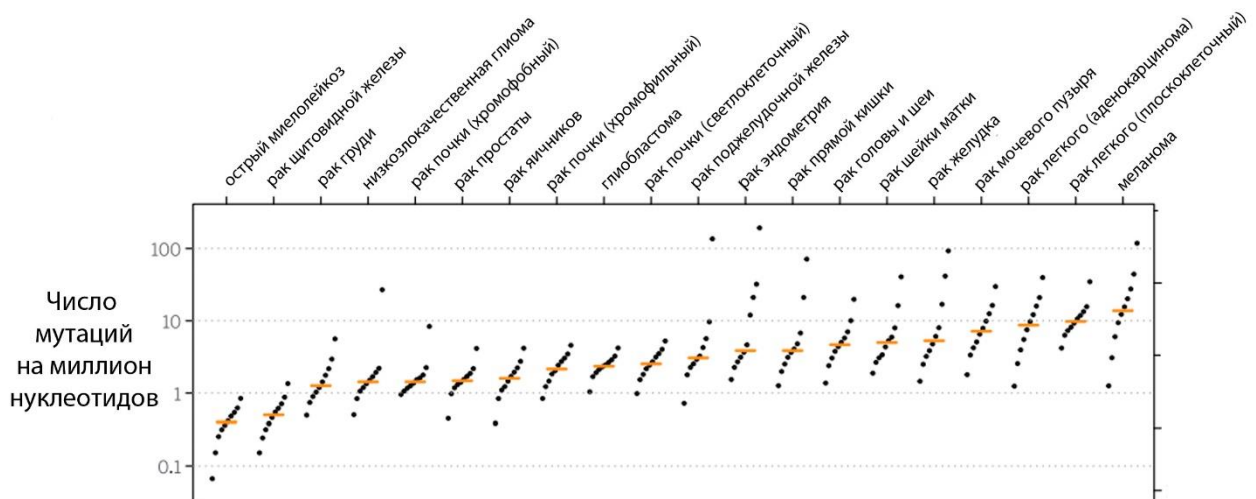


Рисунок 1.1. Скорость мутирования в 20 типах рака (по [9] с изменениями). Внутри каждого типа рака образцы разбиты на группы по 10 штук по числу мутаций в них. Точками обозначены медианные значения числа мутаций (точечных замен, а также вставок и выпадений) внутри десятка образцов, оранжевая линия показывает медианное значение для всех образцов данного типа рака вместе.

1.1 Мутационные подписи

Изучение мутаций в раковых образцах позволило узнать многое о мутационных процессах, приводящих к накоплению мутаций. Мутации, наблюдаемые при секвенировании ракового образца, являются кумулятивным результатом всех соматических мутационных механизмов, включая повреждения ДНК и процессы репарации, которые происходили в клеточной линии начиная с оплодотворенной яйцеклетки, из которой развился пациент, до раковой клетки [10]. Лишь малая часть мутаций в каждом раковом образце является «драйверами» онкогенеза, которые дают селективное преимущество роста одному из клонов (по последним оценкам число таких мутаций варьирует от 1 до 10 на образец в зависимости от типа рака [1]). Большинство же мутаций в раковых образцах считаются «пассажирскими»; паттерны их распределения слабо подвержены действию отбора [11]. Поэтому набор мутаций в раковом образце можно рассматривать как своеобразную археологическую запись, содержащую следы, или «подписи», мутационных процессов, действовавших в нём.

Хотя возникновение каждой мутации – вероятностный процесс, мутации различных типов не равновероятны. Известно, что некоторые мутагены оставляют очень специфические мутационные подписи, то есть вызывают повреждения, приводящие к определенным нуклеотидным заменам, часто в определенном нуклеотидном контексте. Например, типы мутаций в раках легких и кожи похожи на те, что возникают при обработке клеток канцерогенами табачного дыма или облучении ультрафиолетом соответственно [12–15]. В частности, в раках легких, ассоциированных с курением, преобладают трансверсии С:G→А:Т, в то время как для раков кожи, ассоциированных с УФ, характерны транзиции С:G→Т:А в дипиримидинах и двойные нуклеотидные замены СС:GG→ТТ:АА. Таким образом, известно, что сильное воздействие внешних мутагенов оставляет следы в виде определенных мутационных паттернов (подписей) в раковых геномах.

Эндогенные процессы также могут оставлять определенные мутационные подписи. Например, такие внутриклеточные процессы, как энергетический обмен и перекисное окисление липидов, являются источниками активных реагентов (например, активных форм кислорода), которые вызывают повреждения ДНК [16]. Известно, что эти эндогенные мутагены производят определенный подкласс мутаций и, таким образом, также влияют на набор мутаций, наблюдаемых в раковом образце [17]. Еще одним эндогенным мутагеном с известной подписью является белки семейства АРОВЕС. В нормальной клетке эти белки функционируют для защиты от ретровирусов, внося мутации на промежуточных стадиях их жизненного цикла, использующих ДНК. Недавно было показано, что в раковых клетках белки семейства АРОВЕС могут быть источником мутаций в собственной ДНК клетки [18–20]. Даже процесс репликации,

который является очень точным, может стать источником ошибочного спаривания оснований, приводящего в последствии к мутациям, которые тоже характеризуются определенными подписями.

Клетка запускает процессы репарации, чтобы сохранить целостность генома, исправляя повреждения, вызванные экзо- и эндогенными мутагенами [21]. Например, эксцизионные системы репарации неспаренных оснований (BER – base excision repair) и нуклеотидов (NER – nucleotide excision repair) работают с повреждениями ДНК, вызванными соответственно клеточными метаболитами и широким спектром повреждений, разрушающих спираль ДНК [22]. Процессы репарации в свою очередь имеют различную эффективность в исправлении мутационных подписей, оставленных мутагенами, и таким образом также влияют на финальный набор наблюдаемых мутаций.

Мутационные процессы имеют и другие характерные черты, помимо специфических типов нуклеотидных замен и локального контекста, в котором происходят повреждения. К таким особенностям можно отнести распределение мутаций вдоль генома и между цепями ДНК, исправление повреждений данного типа определенной системой репарации клетки, время действия мутационного процесса в ходе развития раковой опухоли.

Получается, что при секвенировании ракового образца мы имеем некий итоговый набор мутаций, который является результатом действия различных мутационных процессов (которые могли действовать в разные периоды времени и с разной интенсивностью), скорректированный действием различных систем репарации. Одним из крупных прорывов в исследовании ракового мутагенеза было восстановление «подписей» различных мутационных и репарационных процессов из итоговых наборов мутаций раковых опухолей [3; 4]. Использование компьютерной модели, в основе которой лежит математический метод неотрицательного матричного разложения, помог восстановить подписи мутационных процессов и интенсивность их воздействия в разных раковых образцах (Рисунок 1.2). Идея состоит в том, что, имея финальные наборы мутаций для большого числа раковых образцов, мы хотим восстановить подписи мутационных процессов и их интенсивность в каждом данном образце. В статье, описывающей данный метод, приводится наглядная аналогия данной математической задачи. Представим вечеринку, где в комнате находится некоторое количество людей, общающихся друг с другом, а в разных местах этой комнаты расставлены микрофоны, записывающие звук разговоров. Имея записи всего звука в комнате с разных микрофонов, мы хотим восстановить отдельные разговоры. Это возможно благодаря тому, что интенсивности звука разных разговоров отличаются из-за различий в расстоянии от конкретного разговора до конкретного микрофона. В данной аналогии имеющиеся записи аналогичны имеющимся итоговым наборам мутаций из

раковых образцов, отдельные разговоры аналогичны мутационным процессам и их подписям, а интенсивность мутационного процесса аналогична громкости данного разговора. Такая задача является примером задач, известных в математике как задачи слепого разделения сигнала. Одним из способов решения таких задач является неотрицательное матричное разложение, которое и было использовано в данной биологической задаче.

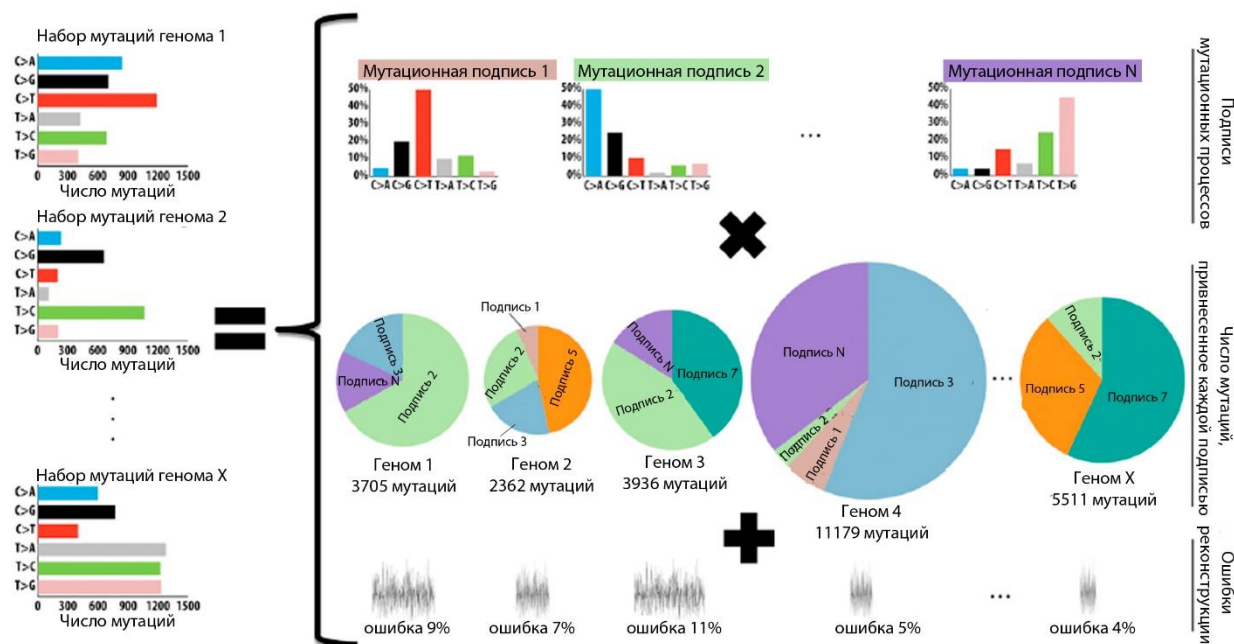


Рисунок 1.2. Использование наборов мутаций из раковых образцов для восстановления подписей мутационных процессов, а также числа мутаций, привнесённого каждым процессом в каждый геном. Извлеченные подписи и их вклады не позволяют восстановить исходные данные со 100% точностью, что приводит к наличию геном-специфических ошибок реконструкции (по [4] с изменениями).

1.2 Репликация как источник мутаций

Как было описано выше, одним из источников мутаций могут являться эндогенные процессы клетки, например, репликация. Репликация – это очень точный процесс. Его точность достигается тремя главными компонентами. Во-первых, избирательностью репликативных полимераз, которые вставляют правильный комплементарный нуклеотид с очень высокой вероятностью. Во-вторых, активностью экзонуклеазных доменов репликативных полимераз, которая позволяет заменять неправильно вставленное основание. В-третьих, исправлением ошибок, оставшихся после предыдущего шага, с помощью системы репарации неспаренных нуклеотидов (MMR – mismatch repair system) [23]. Исследования на дрожжах показали, что эффективность каждого из этих шагов зависит от типа неспаренного основания, и при этом MMR компенсирует неточность полимераз [24; 25]. Несмотря на все это, некоторое количество ошибок

все же остается не исправленным, и спустя еще один цикл репликации они становятся мутациями.

Известно, что две цепи ДНК реплицируются по-разному: одна цепь реплицируется непрерывно и называется лидирующей, а вторая – небольшими участками, называемыми фрагментами Оказаки, и называется отстающей [26]. Классическая модель вилки репликации эукариот [27] предполагает, что репликация лидирующей и отстающей цепей ДНК разделена между основными репликативными полимеразы: полимеразы эпсилон реплицирует лидирующую цепь, а полимеразы дельта и альфа – отстающую цепь (рисунок 1.3). Исключения составляют ориджины репликации и другие специфические участки, в которых полимеразы дельта может участвовать в репликации обеих цепей [28]. Классическую модель подтверждают многие эксперименты. В их числе эксперименты с мутантными полимеразы [29–31] или вставкой рибонуклеотидов [32], в которых соответствующие полимеразы наблюдались на соответствующих цепях ДНК; эксперименты, исследовавшие связывание полимераз с лидирующей и отстающей цепями ДНК [33]; биохимические эксперименты по сборке и стабилизации репликативного комплекса [34; 35]. Однако эта модель недавно была подвергнута сомнениям. Согласно альтернативной модели репликации полимеразы дельта отвечает за синтез обеих цепей ДНК, в то время как экзонуклеазная активность полимеразы эпсилон вовлечена в исправление ошибок, созданных полимеразой дельта во время репликации лидирующей цепи ДНК [36]. Однако эта новая модель очень противоречива и противоречит большинству имеющихся данных [37; 38]. В своем исследовании мы опирались на классическую модель, однако приводим и обсуждение новой модели.

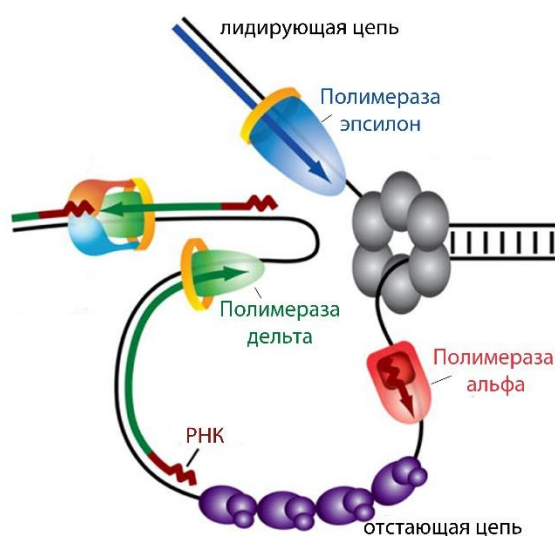


Рисунок 1.3. Схематическое изображение вилки репликации эукариот (по [31] с изменениями).

Вне зависимости от конкретной модели репликационной вилки тот факт, что две цепи реплицируются по-разному и ошибки на них исправляются независимо, может приводить к различию в числе и видам ошибок на разных цепях. Чуть подробнее об этом будет рассказано в пункте 1.4.

1.3 Система репарации неспаренных нуклеотидов

Система репарации неспаренных нуклеотидов (MMR) – одна из важнейших систем репарации в клетке. Действует она в основном корепликативно, то есть исправляет ошибки, возникшие в ходе репликации, практически одновременно с ней [39; 40]. В экспериментах по нокауту генов - основных участников данной системы в дрожжах было показано, что при таком нокаутировании скорость мутирования возрастает в десятки раз [25; 41]. Скорость мутирования также увеличена в разы в раковых образцах с инактивированной MMR [42]. Известно, что система MMR наиболее эффективно исправляет вставки и выпадения [43], что согласуется с тем фактом, что раковые образцы с инактивированной MMR содержат огромное количество вставок и выпадений [44], особенно в простых повторах. Такой фенотип микросателлитной нестабильности (MSI – microsatellite instability) используется для экспериментального подтверждения инактивации системы MMR.

Система MMR в эукариотах представлена в основном действием пути MutS α -MutL α . Этот путь инициируется, когда гетеродимер MutS α , состоящий из белков Msh2 and Msh6, связывается с неправильно спаренными нуклеотидами. MutS α отвечает за исправление основных ошибок репликации – точечных однонуклеотидных ошибок, а также вставок и выпадений. MutS α содержит два АТФазных активных сайта, необходимых для работы системы MMR. АТФ и связывание с неправильно спаренными нуклеотидами инициируют конформационные изменения в гетеродимере MutS α , такие, что он формирует кольцо, способное двигаться вдоль ДНК. Такая форма MutS α , активированная АТФ, дает возможность связывания с комплексом MutL α – гетеродимером, состоящим из белков Mlh1 and Pms2. Активированный MutL α вносит разрывы в цепи ДНК вокруг повреждения. Далее происходит удаление ошибки вместе с окружающим фрагментом ДНК, синтез нового фрагмента по образцу второй цепи ДНК и лигирование.

Инактивация системы MMR может быть вызвана вредными мутациями в генах данной системы. Однако чаще инактивация происходит из-за эпигенетических изменений, приводящих к изменениям в экспрессии генов системы MMR, например, метилирования промотора гена *MLH1* [45] или даже инактивации генов, не вовлеченных напрямую в систему MMR [46]. Инактивация MMR часто приводит к развитию рака, в частности рака кишки, рака эндометрия и

рака яичников. Врожденная инактивация MMR может приводить к развитию рака мозга еще в детском возрасте.

1.4 Эксцизионная система репарации нуклеотидов, ассоциированная с транскрипцией

Эксцизионная репарация нуклеотидов (NER) – это единственный механизм удаления крупных повреждений ДНК, вызванных различными генотоксичными агентами, включая УФ [47–49]. Эксцизионная система репарации состоит из 2 ветвей – глобальной (GG-NER – global genome NER) и ассоциированной с транскрипцией (TC-NER – transcription-coupled NER), которые отличаются прежде всего на этапе распознавания повреждения [50; 51]. В TC-NER распознавание происходит при остановке РНК полимеразы II во время транскрипции, что привлекает основные факторы данной системы репарации.

Основной механизм эксцизионной репарации нуклеотидов одинаков для GG-NER и TC-NER и хорошо изучен. Кратко говоря, он состоит из внесения разрезов с двух сторон от повреждения, в результате чего вырезается одноцепочечный фрагмент ДНК, содержащий повреждение, длиной от 24 до 32 нуклеотидов. Образовавшийся пробел заполняют ДНК полимеразы, после чего концы новосинтезированной ДНК лигируются [48; 49].

Мутации в генах NER могут приводить к фенотипам, которые варьируют от нормального развития до сильной предрасположенности к раку или дефектов развития нервной системы. Такой широкий спектр фенотипов объясняется разнообразием повреждений, устраняемых системой NER, наличием двух путей NER и множественными функциями различных факторов данной системы репарации.

1.5 Распределение мутаций вдоль генома в раковых образцах

Соматические мутации в раковых клетках распределены вдоль генома неравномерно. Эта гетерогенность объясняется особенностями мутационных процессов и неравномерностью эпигенетического ландшафта, т.е. различиями между участками генома по времени репликации, структуре хроматина и уровню экспрессии [5; 52; 53]. Было показано, что локальная скорость мутирования в раковых клетках хорошо коррелирует с различными эпигенетическими разметками. Важным фактором здесь является использование эпигенетических характеристик из ткани, соответствующей происхождению опухоли. В частности, было показано, что при использовании эпигенетических разметок соответствующей ткани (доступности и модификаций хроматина, времени репликации) можно объяснить до 86% дисперсии локальной скорости мутирования вдоль генома [5].

Мутации, вызванные разными мутационными процессами, по-разному зависят от характеристик генома. Так, большинство мутаций имеют сильную зависимость от времени репликации: участки поздней репликации имеют более высокие скорости мутирования в сравнении с участками ранней репликации [54–59]. Однако мутации, вызванные белками семейства АРОВЕС, напротив, не зависят от времени репликации [60]. Мутации, активно исправляемые системой TC-NER (например, вызванные УФ или воздействием табачного дыма), имеют сильную зависимость от уровня экспрессии генов: гены с высоким уровнем экспрессии имеют гораздо более низкие скорости мутирования [53].

1.6 Асимметрия мутаций между цепями ДНК

Одна из важных проблем при анализе мутаций – потеря информации о цепи, на которой изначально произошло повреждение. Например, рассмотрим мутационный процесс, который инициирует процесс окисления гуанина в паре С:G. В данных секвенирования ДНК мы видим мутацию только спустя много клеточных делений. В рассматриваемом примере после первого деления произойдет неправильное спаривание А:G, которое приведет к стабильной паре А:Т после еще одного цикла репликации. Поскольку примерно половина пар С:G ориентирована так, что С находится на референтной цепи, а половина так, что на нереферентной, то мы увидим примерно одинаковое количество мутаций G→Т и С→А. Повреждение цитозина в паре С:G приведет к аналогичному результату. Таким образом, мы не знаем ничего о том, какое именно основание было повреждено на самом деле.

Чтобы разобраться в том, какое повреждение происходит на самом деле, нужно как-то различать две цепи ДНК. В транскрибируемых участках генома можно сравнивать транскрибируемую и нетранскрибируемую цепи. Тогда мы можем рассматривать замену С:G→А:Т как мутацию С→А или G→Т в зависимости от того, С или G находится на транскрибируемой цепи. Если мы можем различать отстающую и лидирующую цепь ДНК, то мы можем также ориентировать мутации относительно направления репликации, то есть в данном примере различать мутации С→А и G→Т в зависимости от того, С или G находится на лидирующей цепи. Как при репликации, так и при транскрипции возникновение и исправление повреждений могут быть асимметричными (цепеспецифичными), и такой подход может помочь различить скорости и паттерны мутирования между цепями [6; 53; 61–67].

Асимметрия между цепями была достаточно хорошо изучена в контексте транскрипции; такая асимметрия называется транскрипционной. Как было описано выше, повреждение ДНК на транскрибируемой цепи может вызывать остановку РНК-полимеразы, что ведет к активации системы TC-NER, которая может исправить повреждение [51; 68–72]. Более высокий уровень

транскрипции связан с большей вероятностью исправления повреждения с помощью TC-NER, что приводит к обратной корреляции между уровнем экспрессии гена и скоростью мутирования [52; 53; 73]. Напротив, повреждение на нетранскрибируемой цепи не может остановить РНК-полимеразу и таким образом остается незамеченным для TC-NER. Более того, нетранскрибируемая цепь остается в одноцепочечном состоянии во время процесса транскрипции и поэтому более подвержена повреждениям [74]. В совокупности это приводит к различиям в скоростях и спектрах мутирования между транскрибируемой и нетранскрибируемой цепями ДНК. Изучение транскрипционной асимметрии дает дополнительную информацию о повреждениях и TC-NER, помимо той, что может быть получена из корреляции скорости мутирования с уровнем экспрессии, поскольку уровень экспрессии коррелирует с другими геномными факторами, например, хроматином.

Асимметрия между лидирующей и отстающей цепями ДНК (репликационная асимметрия) изучена меньше. Такая асимметрия изменяет направление на ориджинах репликации, где референтная цепь меняется с отстающей на лидирующую или наоборот. Репликационная асимметрия наблюдалась как локальные смещения в нуклеотидном составе в хромосомах бактерий [63; 75] и эукариот [67], связана с устойчивыми ориджинами репликации в дрожжах [76] и была экспериментально показана в дрожжах [64; 77]. Репликационная асимметрия также была показана в некоторых типах раковых образцов, а именно в раках с высокой активностью белков семейства APOBEC [6; 78], в раках с инактивированной системой репарации неспаренных нуклеотидов и в раках с инактивированным экзонуклеазным доменом полимеразы эпсилон [6; 61]. Природа этой асимметрии может быть различной; например, она может возникать вследствие того, что две цепи ДНК реплицируются разными полимеразы, которые могут различаться по своей точности и склонности к ошибкам определенного типа или в результате разной подверженности действию белков APOBEC. Остается еще много неизученных вопросов, связанных с репликационной асимметрией.

1.7 Изменение мутационных процессов в ходе развития опухоли

Большинство мутаций в раковых образцах являются «пассажирами»; однако таких мутаций тысячи, и, хотя они не являются причиной развития рака, они являются богатым источником информации об истории развития опухоли. Хотя эти мутации не находились под прямым действием положительного отбора, они являются результатом действия биологических (мутационных) процессов, которые происходили на всем протяжении развития опухоли. Итоговый набор мутаций, наблюдаемый в раковом образце, определяется силой и продолжительностью воздействия разных мутационных процессов (рисунок 1.4). Более того, раковая опухоль часто состоит из нескольких отдельных клеточных популяций (субклональных

популяций), которые по-разному подвергались воздействию мутационных процессов, что еще больше усложняет картину итогового набора мутаций, наблюдаемого в раковых образцах. Таким образом, мутации, наблюдаемые в опухоли после ее извлечения и секвенирования, являются сложным набором мутационных подписей.

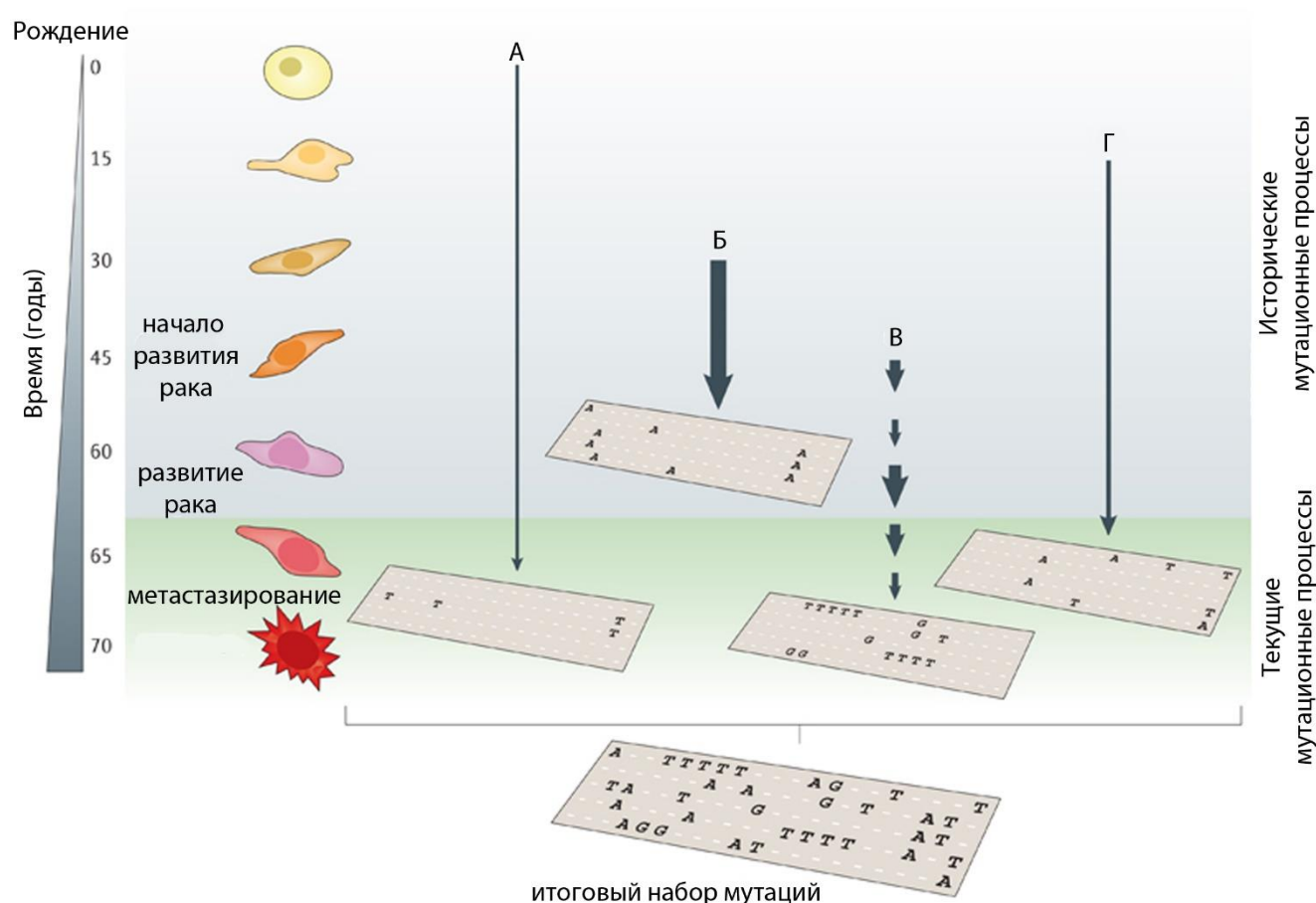


Рисунок 1.4. Мутационные процессы, результат которых наблюдается в мутациях при секвенировании раковой опухоли (по [79] с изменениями). Стрелками обозначена продолжительность и интенсивность мутационных процессов А-Г. Итоговый набор мутаций представляет собой сумму мутаций, привнесенных процессами А-Г, действовавших на протяжении всей жизни. Процесс А представляет собой деаминирование метилированных цитозинов, которое происходит на протяжении всей жизни. Процесс Б может быть, например, воздействием табачного дыма. Процесс В может представлять собой эпизодические всплески активности белков АРОВЕС. Процесс Г представляет результат инактивации системы исправления ошибок в ДНК.

Как было описано выше, с помощью математических моделей можно извлечь индивидуальные мутационные подписи отдельных мутационных процессов в каждом отдельном раковом образце. Число мутаций, привносимых конкретной подписью (мутационным процессом), может быть использовано для аппроксимации интенсивности воздействия данного мутационного процесса, которое может варьировать от образца к образцу. Таким образом,

мутационные подписи дают информацию не только о присутствии некоего мутационного процесса в истории развития опухоли, но и характеризуют интенсивность воздействия данного процесса.

Подробное изучение воздействия разных мутационных процессов в развитии опухоли очень важно. Оно помогает разделить процессы на те, что действовали на ранних этапах развития, и те, что действуют сейчас. Изучение процессов, которые действовали на ранних этапах, дает важное понимание того, как развивалась опухоль, а анализ таких процессов на большом числе образцов может предоставить важную информацию для профилактики раковых заболеваний. Процессы, которые происходят в опухоли в настоящее время, более важны для лечения существующей опухоли, поскольку они могут дать информацию о терапевтической чувствительности и потенциальных мишенях контроля заболевания.

Большое количество данных секвенирования раковых экзомов позволяет глубже изучить особенности развития опухолей разных типов рака именно с точки зрения изменения воздействия мутационных процессов во времени. Большинство раковых образцов на данный момент отсеквенировано не по одной клетке, а целыми фрагментами опухоли. Разные клетки раковой опухоли могут содержать разные мутаций в зависимости от того, на каком этапе возникла та или иная мутация. Если мутация возникла в клетке-предшественнице раковой или на первых делениях раковой клетки, то такая мутация будет присутствовать во всех или почти во всех клетках ракового образца. Если же мутация возникла на поздних стадиях развития рака, то она будет присутствовать в небольшом числе (подпопуляции) клеток. Такие мутации можно различить по данным секвенирования. Ранние мутации будут иметь высокие аллельные частоты (близкие к 0.5) при условии, что все отсеквенированные клетки являются раковыми, то есть отсеквенированный фрагмент опухоли не содержит примеси здоровых клеток. Поздние мутации будут иметь гораздо более низкие значения аллельных частот. Также стоит учитывать, что в раковых образцах достаточно часто встречаются крупные геномные перестройки и как следствие изменение копийности отдельных участков генома. Это, безусловно, тоже влияет на наблюдаемую аллельную частоту мутации. Так, чтения секвенирования двух разных фрагментов генома, возникших в результате соматической дупликации, могут картироваться на один и тот же фрагмент референтного генома. При этом мутация, возникшая в одной из копий фрагмента генома после его удвоения, будет иметь аллельную частоту вдвое ниже той, которая была бы у нее в однокопийном фрагменте. Для решения этих проблем можно оценивать примесь здоровых клеток в исследуемом образце и копийность отдельных участков генома.

Оценка времени возникновения мутации позволяет сравнивать мутационные процессы, оставившие подписи в ранних и поздних мутациях. Такой анализ раковых экзомов разных типов

рака показал, что интенсивность воздействия разных мутационных процессов меняется со временем в опухолях разных типов [7]. Было показано, что подпись 1¹, которая представляет собой мутации С→Т в СрG (подчеркнутый нуклеотид мутирует) и считается результатом спонтанного деаминирования цитозина, в 4 из 5 исследованных типах рака была более выражена среди ранних мутаций по сравнению с поздними. Это предполагает, что большинство ранних мутаций возникли благодаря не ракоспецифичным мутационным процессам. Интенсивность подписи 2, связанной с активностью белков АРОВЕС, возрастает в субклональных мутациях почти во всех исследованных типах рака. Для подписи 4 (связанной с воздействием табачного дыма) и подписи 7 (связанной с воздействием УФ) было показано, что доля привносимых ими мутаций уменьшается со временем развития рака, что говорит о возникновении и активности новых мутационных процессов с развитием опухоли.

Поиск драйверных мутаций требует построения точной мутационной нуль-модели, описывающей распределения мутаций в геноме без отбора. Анализ изменения мутационных процессов во времени важен не только с фундаментальной точки зрения, но и с практической - для поиска субклональных драйверных мутаций.

¹ Нумерация подписей приведена согласно базе данных COSMIC и статье [3]. Мутационные спектры подписей и их краткие характеристики можно посмотреть по ссылке https://cancer.sanger.ac.uk/cosmic/signatures_v2

Глава 2. Асимметричная работа системы репарации неспаренных нуклеотидов на лидирующей и отстающей цепях ДНК

В экспериментах с пекарскими дрожжами было показано, что ДНК полимеразы, реплицирующие лидирующую и отстающую цепи ДНК, отличаются по своей точности, то есть вероятности вставить неверный нуклеотид. MMR в свою очередь считается одной из систем, уравнивающих эти различия [64]. В ходе изучения раковых данных было показано, что в раках с инактивированной MMR наблюдается репликационная асимметрия [6], что также говорит о различиях в эффективности работы данной системы на разных цепях ДНК. Однако природа данной асимметрии оставалась неизвестной. В данной главе описан метод, с помощью которого мы установили, что в клетках человека MMR также компенсирует меньшую точность одной из полимераз.

2.1 Материалы и методы

2.1.1 Мутационные данные

В данной главе были использованы исключительно опубликованные ранее данные: (1) выявленные соматические мутации по данным полноэкзомного секвенирования 159 раковых образцов с микросателлитной нестабильностью (MSI – microsatellite instability) и 782 раковых образца с микросателлитной стабильностью (MSS – microsatellite stability) и по данным полногеномного секвенирования MSI (11 шт.) и MSS (27 шт.) раковых образцов из базы данных The Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>); (2) файлы в формате BAM с чтениями, выровненными на референтный геном, для ультрамутабельных раковых образцов с врожденной биаллельной инактивацией MMR (bMMRD - biallelic mismatch repair deficiency) и соматическими мутациями в генах полимеразы эpsilon или дельта [80]; (3) файлы в формате VCF для гипермутабельных раковых образцов с врожденной гомозиготной мутацией в гене *MSH2* (ген системы MMR) и соматической мутацией в гене полимеразы эpsilon [81]; (4) множественное выравнивание геномов человек-шимпанзе-орангутанг из геномного браузера UCSC (<https://genome.ucsc.edu/>) и (5) файлы в формате VCF, описывающие полиморфизм человека, из проекта 1000 геномов.

Скорость замен везде подсчитывалась как число замен определенного типа, разделенное на число сайтов соответствующего типа.

Для анализа межвидовых данных мы получили мутации в линии человека после его расхождения с шимпанзе методом наибольшей экономии, используя орангутанга в качестве аутгруппы.

2.1.2 Определение соматических мутаций в ультрамутабельных раках

Соматические мутации были определены с помощью программы MuTest (версия 1.1.4) [82] при параметрах по умолчанию. Затем мутации были профильтрованы против базы данных однонуклеотидных полиморфизмов.

2.1.3 Репликационная асимметрия

Для каждой мутации было определено, насколько вероятно, что референтная цепь ДНК в данной позиции генома реплицируется как лидирующая (отстающая) цепь. Для этого была использована величина, названная «направление вилки репликации» (fp – fork polarity). Определялась она как производная времени репликации (RT – replication time) в данной позиции по координате в геноме. Подробно метод описан в [78]. Для анализа было использовано время репликации лимфобластоидной клеточной линии [76]. Согласно значениям этой производной позиции генома были разделены на 10 равных категорий («бинов») (рисунок 2.1): бины с низкими и высокими значениями производной содержат позиции генома, в которых референтная цепь реплицируется преимущественно как отстающая и лидирующая цепь соответственно. Для каждого бина было подсчитано число замен и число соответствующих сайтов. Каждая замена была посчитана дважды: как замена на референтной цепи с соответствующим значением производной времени репликации и как комплементарная замена с производной с тем же значением, но противоположной по знаку. Поэтому все графики с репликационной асимметрией (рисунки 2.5, 2.8, 2.10, 2.11) симметричны относительно нуля. Доверительные интервалы были посчитаны с помощью пакета R «PropCIs». Для анализа отдельных участков позднего и раннего времени репликации геном был разделен на 2 категории согласно значению времени репликации (по 25% генома с самыми высокими и самыми низкими значениями RT).

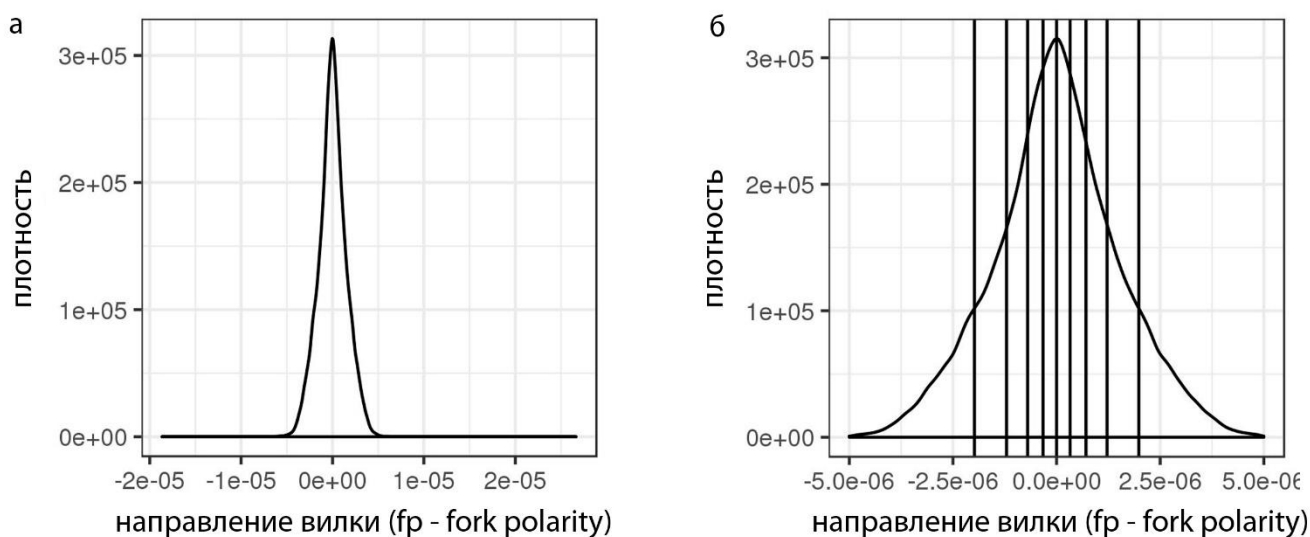


Рисунок 2.1. Распределение значений направления вилки репликации (fr) в геноме человека. Значения fr были усреднены внутри 10кб окон. **а**, распределение fr во всем геноме; **б**, увеличенный график распределения с вертикальными линиями, соответствующими квантилям, использованным для разделения на бины.

2.1.4 Данные секвенирования фрагментов Оказаки

Мы использовали опубликованные данные секвенирования фрагментов Оказаки в лимфобластоидной клеточной линии GM06990 [83]. Мы использовали значения RFD (RFD в данном методе подсчитывалось как разница между долей вилок репликации, направленных вправо, и долей вилок, направленных влево) из этих данных, чтобы определить позиции генома, где референтная цепь реплицируется преимущественно как лидирующая (низкие значения RFD) или отстающая (высокие значения RFD). Мы разделили позиции генома на 10 равных бинов согласно значениям RFD (рисунок 2.2).

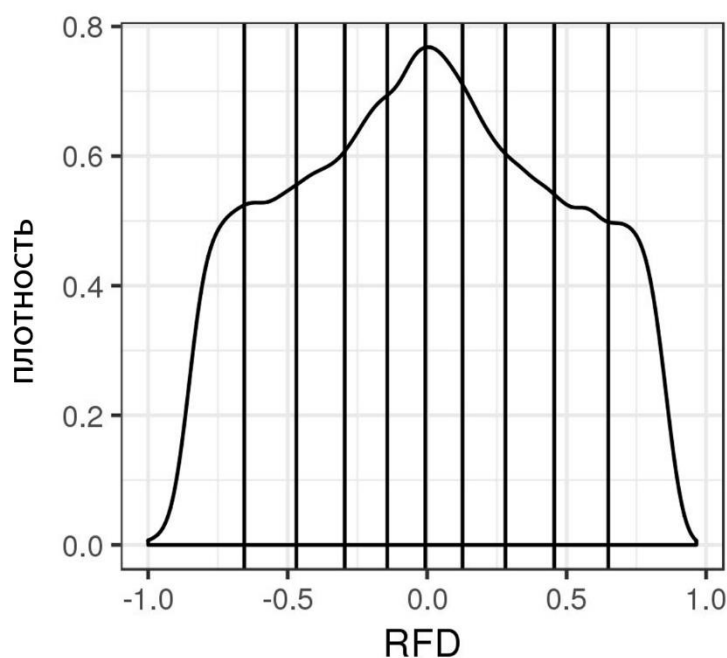


Рисунок 2.2. Распределение значений RFD для всего генома. Вертикальные линии соответствуют квантилям, использованным для разделения на бины.

2.1.5 Анализ вставок и выпадений

Мы использовали данные по однонуклеотидным выпадениям в поли(A) и поли(T) трактах в раковых геномах с микросателлитной нестабильностью (MSI) для трактов длиной от 6 до 8 идентичных нуклеотидов, поскольку для них нашлось достаточное для анализа количество выпадений.

2.1.6 Обогащение подписью АРОВЕС

Обогащение подписью белков АРОВЕС было рассчитано для каждого образца как отношение скоростей мутирования $C \rightarrow K$ ($K = G/T$) в контекстах $TrCpW$ и $VpCpW$ ($W = A/T, V = A/C/G$). Веса мутационных подписей были посчитаны с использованием пакета R «deconstructSigs»[84].

2.1.7 Модель для мутационных смещений между цепями

Мы посчитали отношения мутационных скоростей, используя следующую логику. Из рисунка 2.11 для каждого типа мутаций $A \rightarrow B$ мы получили отношение скорости данной мутации $r(A \rightarrow B)$ и скорости комплементарной мутации $r(A' \rightarrow B')$ на лидирующей цепи. Тогда

$$\frac{x_{MMR}}{(1-x_{MMR})} = \frac{x_\epsilon \alpha + (1-x_\delta)(1-\alpha)}{(1-x_\epsilon)\alpha + x_\delta(1-\alpha)}, \quad (1)$$

где α – это доля мутаций $A \rightarrow B$ и $A' \rightarrow B'$, привносимых полимеразой эpsilon на лидирующей цепи; $(1-\alpha)$ – это доля таких мутаций, привносимых полимеразой дельта на отстающей цепи; x_ϵ и $(1-x_\epsilon)$ – это доли мутаций $A \rightarrow B$ and $A' \rightarrow B'$ соответственно, привносимые полимеразой эpsilon; x_δ и $(1-x_\delta)$ – это доли мутаций $A \rightarrow B$ and $A' \rightarrow B'$ соответственно, привносимые полимеразой дельта; x_{MMR} и $(1-x_{MMR})$ – это доли мутаций $A \rightarrow B$ and $A' \rightarrow B'$ соответственно на лидирующей цепи в раках с инактивацией системы репарации неспаренных нуклеотидов.

Например, рассмотрим мутацию $C \rightarrow A/G \rightarrow T$. Из рисунка 2.11 находим, что значение $x_{MMR}/(1-x_{MMR})$ равно 0.65. Поскольку отношение $C \rightarrow A/G \rightarrow T$ на лидирующей цепи в раках с мутированной полимеразой эpsilon равно 1.96 (рисунок 2.11), доля мутаций $C \rightarrow A$, привносимых полимеразой эpsilon $x_\epsilon = 0.66$, а доля мутаций $G \rightarrow T$, привносимых полимеразой эpsilon $1-x_\epsilon = 0.34$. Аналогично, поскольку отношение $C \rightarrow A/G \rightarrow T$ на отстающей цепи в раках с мутированной полимеразой дельта равно 2.07, доля мутаций $C \rightarrow A$, привносимых полимеразой дельта $x_\delta = 0.67$, а доля мутаций $G \rightarrow T$, привносимых полимеразой дельта $1-x_\delta = 0.33$. Из уравнения (1) $(1-\alpha)/\alpha = 4.0$. Другими словами для данного типа мутаций полимеразы дельта производит на отстающей цепи ошибки, приводящие к данному типу мутаций, в ~4 раза чаще, чем полимеразы эpsilon производит их на лидирующей цепи. Поскольку в раках с работающей системой репарации неспаренных оснований не наблюдается никакой асимметрии, то смещение активности MMR должно быть строго противоположным, то есть MMR должна исправлять ошибки полимеразы дельта ~4 раза чаще, чем ошибки полимеразы эpsilon. Результаты для разных типов мутаций посчитаны аналогично (Таблица 2.6). Общая асимметрия системы репарации неспаренных оснований была посчитана как средняя асимметрия по 3 типам мутаций, наиболее эффективно исправляемым системой MMR, с учетом доли каждого типа мутаций в MSI раковых образцах.

2.2 Результаты

2.2.1 Паттерны мутаций в раковых образцах с инактивированным экзонуклеазным доменом полимеразы эpsilon или дельта

В раках с врожденной биаллельной инактивацией системы репарации неспаренных нуклеотидов (bMMRD) часто наблюдаются соматические мутации в экзонуклеазном домене репликативных полимераз. Поскольку в таком случае повреждения не только не исправляются, но и производятся с повышенной частотой, то это приводит к гипермутабельному фенотипу. В опухолях таких пациентов точность поврежденных полимераз снижена в 100-1000 раз и большинство наблюдаемых в опухоли мутаций сделаны именно этой полимеразой [80; 81; 85; 86]. Мы обнаружили, что мутационный спектр в таких раках отличается от спектра в раках с репликативными полимеразами дикого типа (Рисунок 2.3а). В частности, образцы с биаллельной инактивацией MMR и мутацией в экзонуклеазном домене полимеразы эpsilon сильно обогащены мутациями C→T. Для более детального изучения мы проанализировали контекстную зависимость мутаций в этих образцах. Оказалось, что данные мутации особенно часто появляются в определенных тринуклеотидных контекстах: мутация C→T наиболее часто происходит при наличии нуклеотида G в 3'-позиции, а особенно часто – если при этом в 5'-позиции находится нуклеотид G (GpCpG); мутация C→A наиболее часто происходит при наличии нуклеотида T в 3'-позиции (NpCpT, N = A/T/G/C) (рисунок 2.3б).

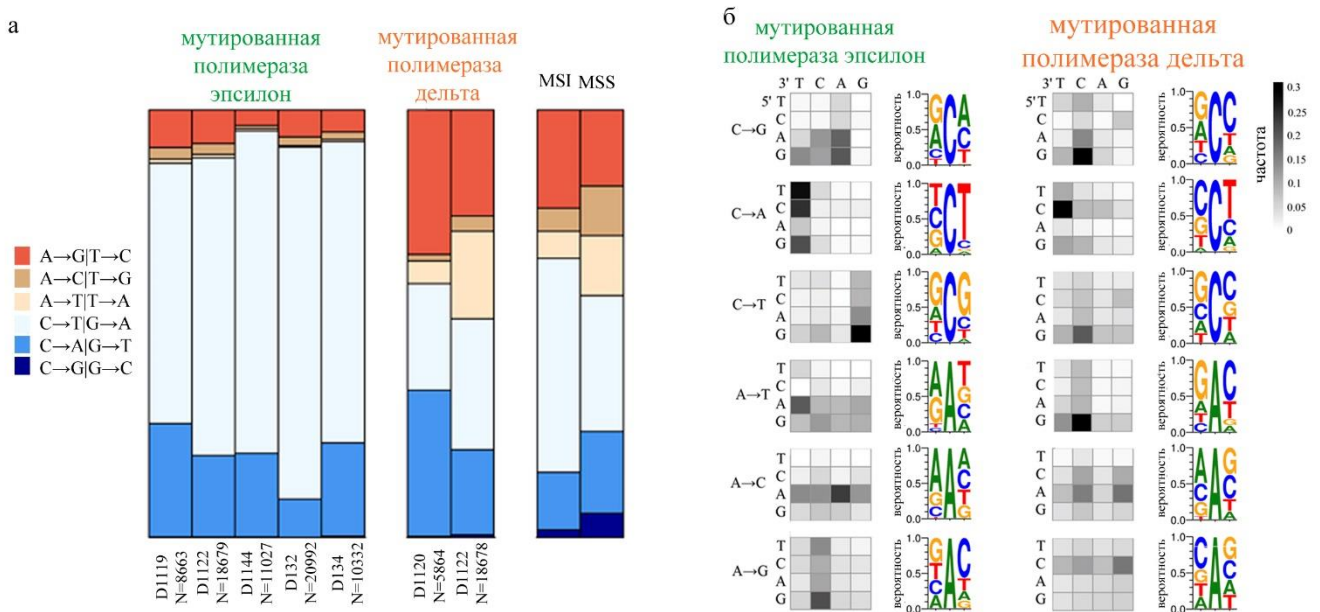


Рисунок 2.3. Мутационные спектры образцов с инактивированным экзонуклеазным доменом полимеразы эpsilon (мутированная полимеразы эpsilon) или дельта (мутированная полимеразы дельта). **а**, частоты мутаций в раковых образцах с bMMRD и последующей мутацией в полимеразе эpsilon или полимеразе дельта. Данные для 7 экзотов bMMRD раков (5 с мутированной полимеразой эpsilon и 2 с мутированной

полимеразой дельта). Относительные частоты однонуклеотидных замен показаны безотносительно цепи; данные из базы данных TCGA для MSI и MSS образцов с полимеразой эпсилон и дельта дикого типа показаны для сравнения. N - число мутаций, наблюдаемое в каждом образце. **б**, Контекстнозависимые частоты каждого типа мутаций, усредненные между пациентами внутри группы. Вертикальная ось соответствует 5' нуклеотиду, а горизонтальная - 3' нуклеотиду. Для каждого типа мутации цвет одной из 16 клеточек соответствует частоте данного типа мутации в данном трехнуклеотидном контексте. Лого содержат ту же информацию и приведены для наглядности.

Чтобы удостовериться, что полученные мутационные паттерны согласуются с имеющейся информацией о мутационных процессах и подписях в раковых образцах с инактивированной MMR и с инактивированным экзонуклеазным доменом полимеразы эпсилон, мы сравнили мутационный спектр bMMRD опухолей с мутированной полимеразой эпсилон с мутационными подписями базы данных COSMIC (Таблица 2.1, четвертый столбец).

Таблица 2.1. Косинусное сходство между подписями из базы данных COSMIC и мутационными спектрами образцов с мутацией в экзонуклеазном домене полимеразы эпсилон. Показаны подписи с косинусным сходством больше 0.7 хотя бы в одном сравнении. Жирным шрифтом выделены косинусные сходства больше 0.7. Нумерация подписей соответствует подписям из базы данных COSMIC (<https://cancer.sanger.ac.uk/cosmic/signatures>).

Подпись	MSS, мутированная полимераза эпсилон	MSI, мутированная полимераза эпсилон	bMMRD, мутированная полимераза эпсилон
Подпись 1	0.21	0.67	0.80
Подпись 6	0.17	0.87	0.84
Подпись 10	0.92	0.35	0.55
Подпись 14	0.33	0.97	0.74
Подпись 15	0.18	0.81	0.79
Подпись 20	0.16	0.73	0.41

Мутационный спектр оказался похож на подписи 6 и 15, которые в базе данных COSMIC приписаны инактивации MMR, а также на подпись 14 (косинусное сходство = 0.74) (Таблица 2.1). В базе данных COSMIC полимеразе эпсилон с инактивированным экзонуклеазным доменом приписана одна подпись – подпись 10. Этиология подписи 14 до этого была неизвестна, хотя она и наблюдалась в образцах с повышенной скоростью мутирования и репликационной асимметрией [87]. Чтобы это исследовать, мы использовали дополнительные данные – образцы

с мутированным экзонуклеазным доменом полимеразы эpsilon из базы данных The Cancer Genome Atlas (TCGA). Часть из этих образцов также имела инактивированную MMR, но соматическую, а не врожденную. Как и ожидалось, спектр bMMRD образцов с мутированным экзонуклеазным доменом полимеразы эpsilon больше похож на образцы из TCGA с соматической инактивацией MMR (косинусное сходство = 0.78), чем на образцы с активной MMR (косинусное сходство = 0.43) (рисунок 2.4). Мы обнаружили, что мутационный спектр образцов TCGA с мутированным экзонуклеазным доменом полимеразы эpsilon сильно зависит от статуса активности MMR: спектр образцов с рабочей MMR хорошо согласуется с подписью 10 (косинусное сходство = 0.92), а спектр образцов с инактивированной MMR согласуется с подписью 14 (косинусное сходство = 0.97). Таким образом, мы выявили, что подпись 14 является подписью полимеразы эpsilon с инактивированным экзонуклеазным доменом, не скорректированной MMR, а подпись 10 – это подпись полимеразы эpsilon с инактивированным экзонуклеазным доменом после действия системы репарации неспаренных нуклеотидов. Так нам удалось установить неизвестную ранее этиологию подписи 14. Чуть позже это также было подтверждено в одной из работ [88].

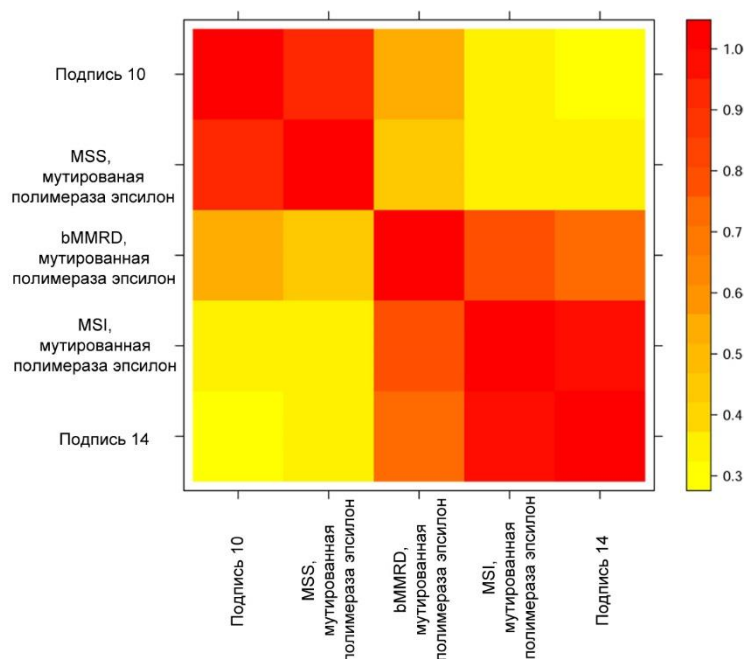


Рисунок 2.4. Косинусное сходство между мутационными спектрами образцов с мутированным экзонуклеазным доменом полимеразы эpsilon и подписями 10 и 14 из базы данных COSMIC.

2.2.2 Цепеспецифичные паттерны мутаций в раковых образцах с инактивированным экзонуклеазным доменом полимеразы эpsilon или дельта

Мы использовали данные по мутациям из образцов с bMMRD, чтобы проверить, с одинаковой ли частотой полимеразы, работающие в клетках человека, создают комплементарные

пары неспаренных нуклеотидов, то есть с одинаковой ли частотой, например, появляются пары C-dA и G-dT, а также определить предпочтительную пару в случае неравновесия. Мы использовали метод, позволяющий определить направление вилки репликации (f_r) с помощью производной времени репликации [6; 78; 89; 90]. Кратко говоря, значения f_r отражают отношение частот пассажей вилки репликации в прямом и обратном направлениях относительно референтной цепи. Мы определили, в каких участках генома референтная цепь реплицируется преимущественно как лидирующая ($f_r > 0$), а в каких – как отстающая ($f_r < 0$). Затем мы сравнили скорости комплементарных мутаций между лидирующей и отстающей цепями в раках с мутированной полимеразой ϵ и мутированной полимеразой δ (рисунок 2.5).

Мы обнаружили, что каждая полимеразы имеет предпочтение к одному типу неправильно спаренных оснований внутри комплементарной пары (Рисунок 2.5б). Этот эффект монотонно изменяется с изменением значений f_r . Самый большой эффект наблюдается в бинах с самыми крайними значениями f_r , которые соответствуют геномным участкам, где мы можем предсказать f_r с самой большой достоверностью [78]. Поэтому, значения именно в этих бинах мы используем в дальнейшем для наших сравнений и модели. Мы проверили, что при использовании всех данных, наблюдаемый эффект слабее, как и ожидается, но его направление сохраняется (Таблица 2.2). Направление асимметрии, наблюдаемое для полимеразы ϵ , для всех типов замен совпадает с ранее полученными экспериментальными данными [61].

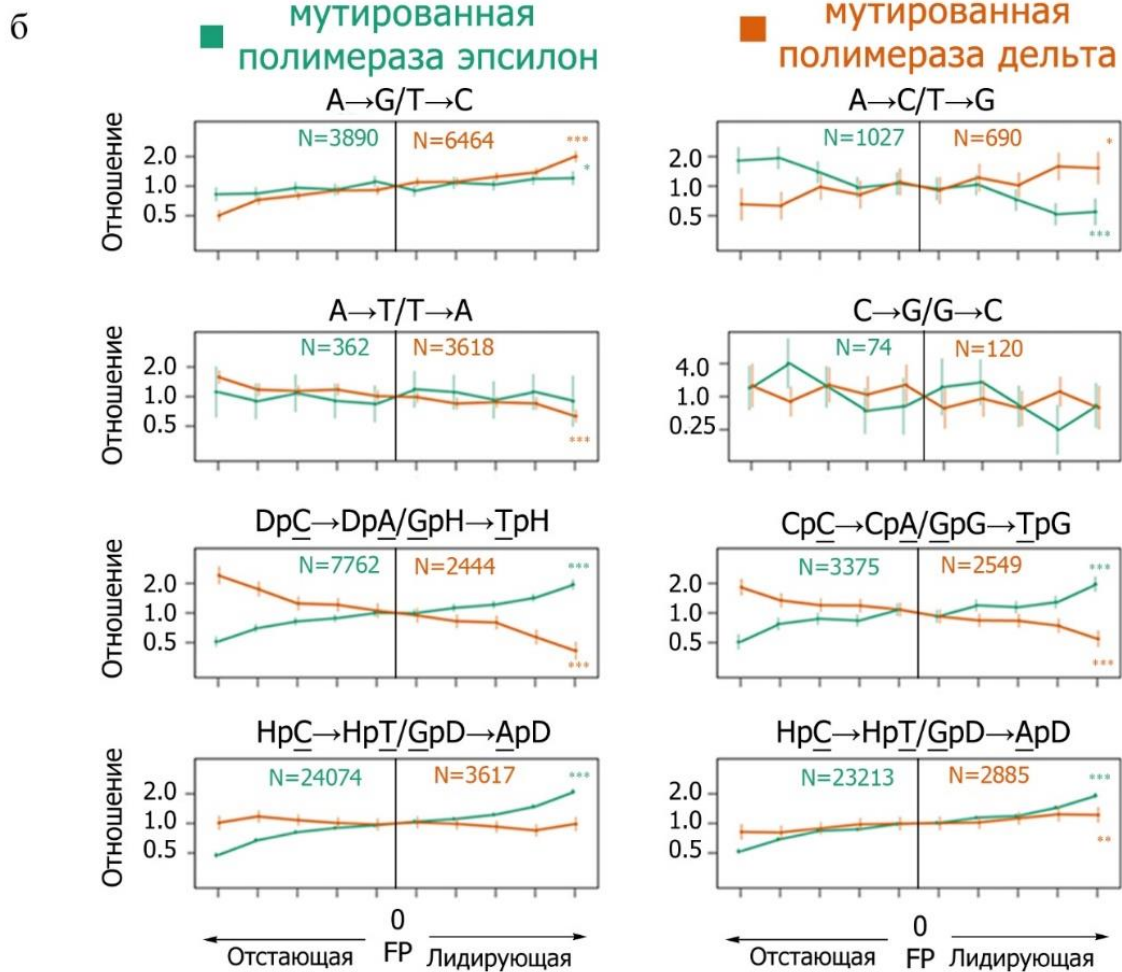
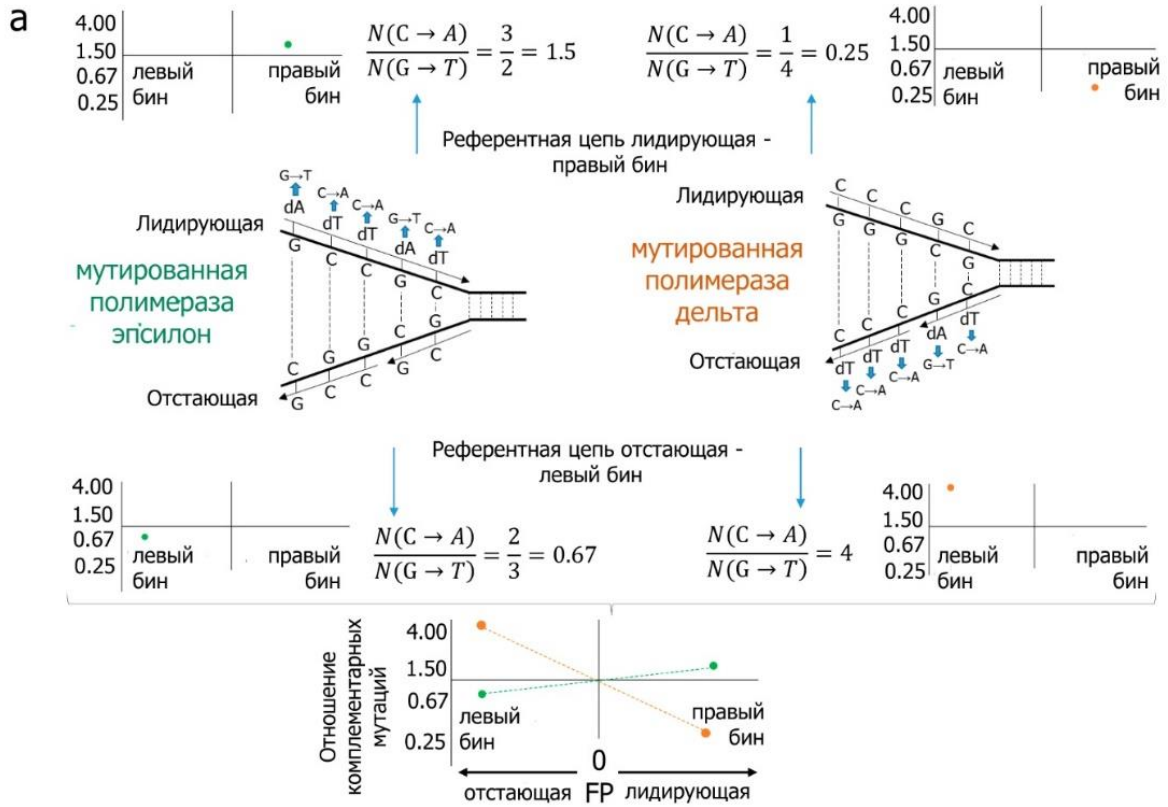


Рисунок 2.5. Репликационная асимметрия в bMMRD опухолях с мутациями в репликативных полимераз. **а**, схематическое изображение метода оценки отношения скоростей мутирования для комплементарных мутаций. В изображенном примере частота неправильно спаренных оснований C-dT, приводящих к мутациям C→A, в 1.5 раза выше, чем скорость мутаций G→T, когда мутирована полимеразы эпсилон, и в 4 раза выше, когда мутирована полимеразы дельта. Самый левый и самый правый бины соответствуют 20% генома, в которых мы с максимальной достоверностью можем определить, что референтная цепь преимущественно реплицируется как лидирующая или отстающая соответственно. **б**, отношение скоростей комплементарных мутаций (вертикальная ось) как функция вероятности референтной цепи реплицироваться как лидирующая или отстающая цепь (горизонтальная ось). Бины оси x соответствуют квантилям производной времени репликации. Вертикальные линии соответствуют 95% доверительному интервалу. Звездочками отмечена значимость отклонения от 1 в самом правом (или левом) бине. (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.001$. D соответствует нуклеотидам A, G или T. H соответствует нуклеотидам A, C или T. Вертикальная ось логарифмическая.

Таблица 2.2. Репликационная асимметрия в самом крайнем бине и с использованием всех данных для bMMRD раков (5 экзотов с мутированной полимеразой эпсилон и 2 экзота с мутированной полимеразой дельта). Значения в скобках соответствуют 95% доверительному интервалу. Звездочками показана значимость отклонения от 1. (*) $p < 0.05$; (**) $p < 0.001$; (***) $p < 0.001$. D соответствует нуклеотидам A, G или T. H соответствует нуклеотидам A, C или T.

Тип мутации	Мутированная полимеразы эпсилон		Мутированная полимеразы дельта	
	Крайний правый бин	Все данные	Крайний правый бин	Все данные
A→G/T→C	1.21 (1.03-1.42)*	1.07 (1.00-1.14)*	2.00 (1.75-2.28)***	1.30 (1.23-1.36)***
A→C/T→G	0.55 (0.40-0.75)***	0.74 (0.66-0.83)***	1.53 (1.05-2.23)*	1.20 (1.04-1.39)*
A→T/T→A	0.90 (0.49-1.63)	1.06 (0.87-1.29)	0.63 (0.54-0.74)***	0.84 (0.79-0.90)***
C→G/G→C	0.69 (0.27-1.76)	0.75 (0.49-1.15)	0.63 (0.25-1.57)	0.82 (0.58-1.15)
DpC→DpA/GpH→TpH	1.95 (1.73-2.20)***	1.27 (1.21-1.32)***	0.41 (0.34-0.51)***	0.71 (0.66-0.76)***
CpC→CpA/GpG→TpG	1.97 (1.65-2.34)***	1.23 (1.15-1.31)***	0.55 (0.45-0.66)***	0.78 (0.73-0.84)***
HpC→HpT/GpD→ApD	2.10 (1.96-2.25)***	1.30 (1.27-1.34)***	0.99 (0.84-1.16)	0.95 (0.89-1.01)
GpC→GpT/GpC→ApC	1.92 (1.80-2.06)***	1.27 (1.24-1.30)***	1.22 (1.02-1.46)*	1.11 (1.03-1.19)**

Более того, для двух пар мутаций (C→A/G→T и T→G/A→C) отклонения, ассоциированные с полимеразой эпсилон и полимеразой дельта, противоположны: полимеразы эпсилон чаще неправильно спаривает основания, приводящие к первой мутации из пары на лидирующей цепи, в то время как полимеразы дельта делает это на отстающей (рисунок 2.5б), что соответствует наблюдениям в дрожжах [91–93]. Похожие паттерны асимметрии наблюдались в

образцах с инактивированной MMR и мутированной полимеразой эpsilon в других доступных данных: полных геномах карциномы эндометрия (UCEC) и аденокарциномы толстой кишки (COAD), а также в другом независимом наборе данных для глиобластом с bMMRD (рисунки П2.1-П2.3). Эти наблюдения совпадают между образцами (Таблица П2.1), что говорит о том, что асимметрия специфична для мутационного процесса, а не для образца или типа рака.

2.2.3 Мутационные спектры в раковых образцах с инактивированной MMR

Для изучения свойств системы репарации неспаренных нуклеотидов мы сравнили раковые образцы COAD и UCEC с функциональным и нефункциональным путями MMR.

Для начала мы разделили имеющиеся раковые образцы на две группы по статусу системы репарации неспаренных нуклеотидов. Инактивация системы MMR ведет к большому количеству маленьких вставок и выпадений в простых повторах. Данный фенотип микросателлитной нестабильности (MSI) может быть использован как маркер для определения инактивации MMR. Мы использовали экспериментальные данные из базы данных TCGA, где раки были классифицированы по этому признаку как стабильные (MSS) и нестабильные (MSI).

Мы постарались подтвердить, что в группе MSI находятся образцы с инактивированной системой репарации. Поскольку инактивация MMR может быть вызвана мутациями в генах MMR, мы проверили образцы на наличие мутаций, приводящих к инактивации, основных генов, кодирующих белки системы MMR (стоп-кодонов, мутаций, приводящих к сдвигу рамки считывания и других вредных мутаций). Поскольку инактивация MMR часто вызвана эпигенетическими изменениями, которые изменяют экспрессию генов MMR или даже инактивацией генов, не вовлеченных напрямую в MMR, то не удивительно, что мы смогли найти такие мутации не во всех образцах (Таблица 2.3).

Таблица 2.3. Вредные (damaging) мутации в генах системы MMR в MSI образцах. Хр – хромосома, Реф – референтный аллель, Алт – альтернативный аллель.

Образец	Тип рака	Ген	Хр	Позиция начала	Позиция конца	Тип мутации	Реф	Алт	Идентификатор базы данных Cosmic
TCGA-AA-3516	COAD	MSH6	2	48027196	48027196	Frame_Shift_Del	A	-	-
		MSH6	2	48030640	48030640	Frame_Shift_Del	C	-	COSM330655
TCGA-AA-A01R	COAD	PMS2	7	6018236	6018236	Missense_Mutation	C	T	COSM179343
		PMS2	7	6038813	6038813	Nonsense_Mutation	G	A	COSM179354
TCGA-A5-A0GA	UCEC	MSH6	2	48030639	48030640	INS	-	C	COSM308681
		PMS2	7	6029581	6029581	Missense_Mutation	C	T	COSM1091091

TCGA-AP-A054	UCEC	MSH6	2	48030639	48030640	INS	-	C	COSM308681
--------------	------	------	---	----------	----------	-----	---	---	------------

В качестве альтернативного метода мы подтвердили инактивацию MMR в MSI образцах, проанализировав мутационный спектр образцов. Во всех MSI образцах мы наблюдали большой вклад мутационных подписей, ассоциированных с дефектной MMR системой (подписи 6, 15, 20 и 26) (Таблица П2.2).

Мы проанализировали мутационные спектры исследуемых образцов и заметили, что спектры образцов с инактивированной системой MMR двух разных типов рака (UCEC и COAD) несколько различаются (рисунок 2.6). Во многих MSI образцах UCEC мы наблюдали подпись 5 и другие мутационные подписи, не связанные с дефектностью MMR. Это говорит о том, что в MSI UCEC раках примешаны MMR-независимые мутационные процессы. Среди MSS образцов в UCEC, в отличие от COAD, наблюдается высокая доля мутаций $\text{TrCpN} \rightarrow \text{G}$ и $\text{TrCpN} \rightarrow \text{T}$ мутаций (соответствуют подписям 2 и 13); в остальном мутационные спектры MSS образцов двух типов рака были очень похожи. Подписи 2 и 13 ассоциированы с активностью белка APOBEC; мы исключили эти образцы (N=5шт.) из дальнейшего анализа, поскольку известно, что в образцах с высокой активностью APOBEC наблюдается репликационная асимметрия, связанная с активностью этого белка.

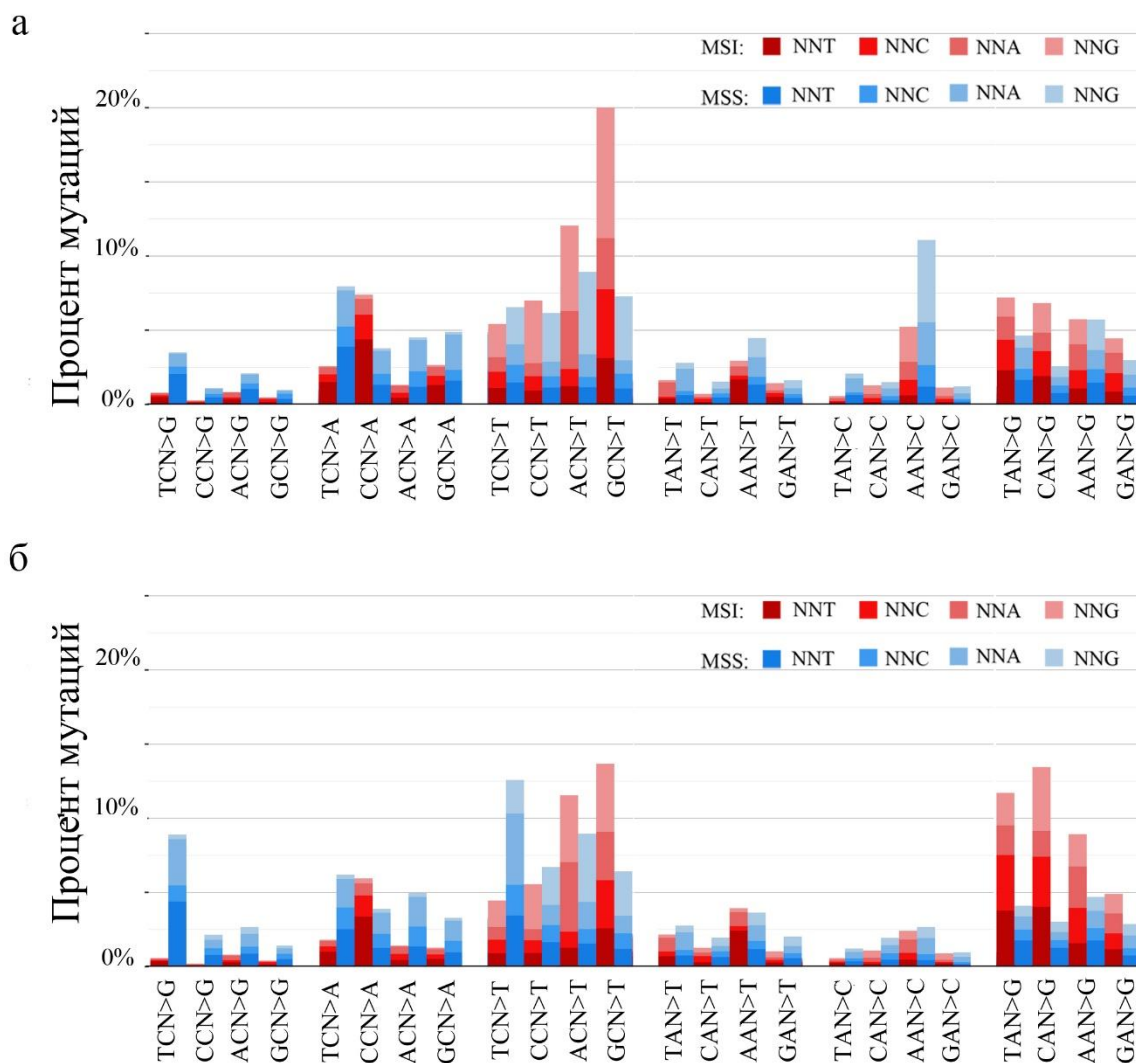


Рисунок 2.6. Мутационные спектры MSI и MSS образцов двух типов рака. **а**, относительные частоты 96 типов мутаций (все возможные типы мутаций в тринуклеотидном контексте) в образцах COAD (3 MSI образца и 8 MSS образцов); **б**, в образцах UCEC (7 MSI образцов и 19 MSS образцов).

Мы задались вопросом, какие мутации исправляются с помощью системы MMR в человеке наиболее эффективно. Для этого мы сравнили мутационные спектры образцов с активной и неактивной системой MMR (рисунок 2.7). Эффективность MMR для конкретного типа неправильно спаренных нуклеотидов может быть посчитана как отношение скоростей соответствующих мутаций в MMR-активных и MMR-инактивированных клетках. MMR исправляет все типы мутаций во всех контекстах (рисунок 2.7,б), с особо высокой эффективностью исправляет неправильно спаренные нуклеотиды, приводящие к следующим мутациям: C→T в контексте GpCpN, A→G во всех контекстах и C→A в контексте CpCpN, что согласуется с предыдущими исследованиями [94].

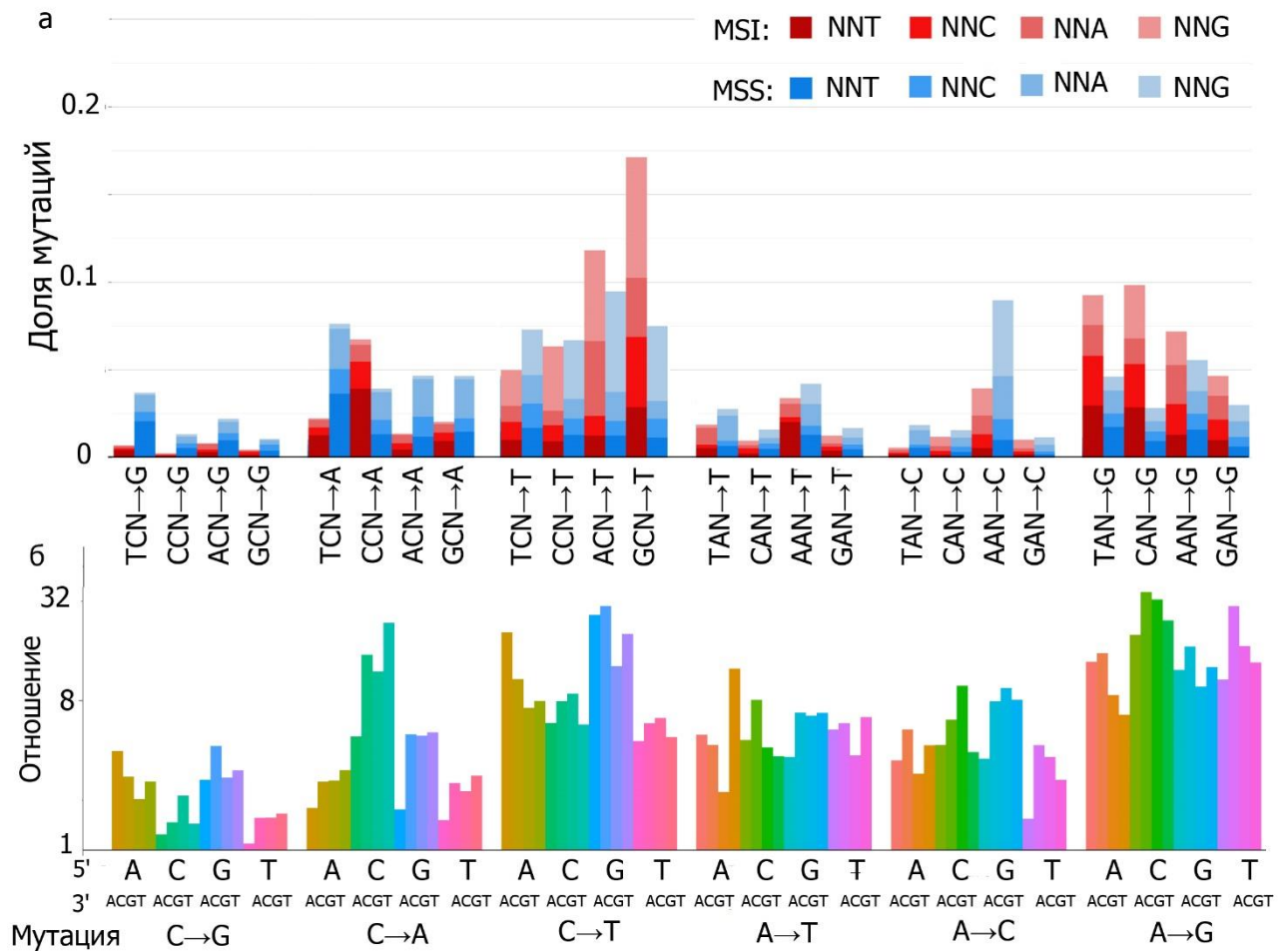


Рисунок 2.7. Сравнение мутационных спектров раковых образцов с активной и инактивированной системой репарации неспаренных нуклеотидов. Комплементарные мутации посчитаны вместе. Данные для COAD и UCEC посчитаны вместе. Данные представлены для 10 MSI образцов и 22 MSS образцов. **а**, относительные частоты 96 типов мутаций (все возможные типы мутаций в каждом тринуклеотидном контексте) в MSI и MSS образцах; **б**, отношение скоростей в MSI и MSS образцах для каждого типа замен. Вертикальная ось логарифмическая.

2.2.4 Мутационная асимметрия в раковых образцах с инактивированной MMR

По аналогии с анализом раковых образцов с мутированными полимеразы, мы сравнили скорости комплементарных мутаций между лидирующей и отстающей цепями для соматических мутаций в MSI и MSS раковых образцах (рисунок 2.8 – красная и синяя линии соответственно). Также для сравнения мы провели аналогичный анализ для врожденных мутаций, оцененных из полиморфизма человека (рисунок 2.8 – зеленая линия).

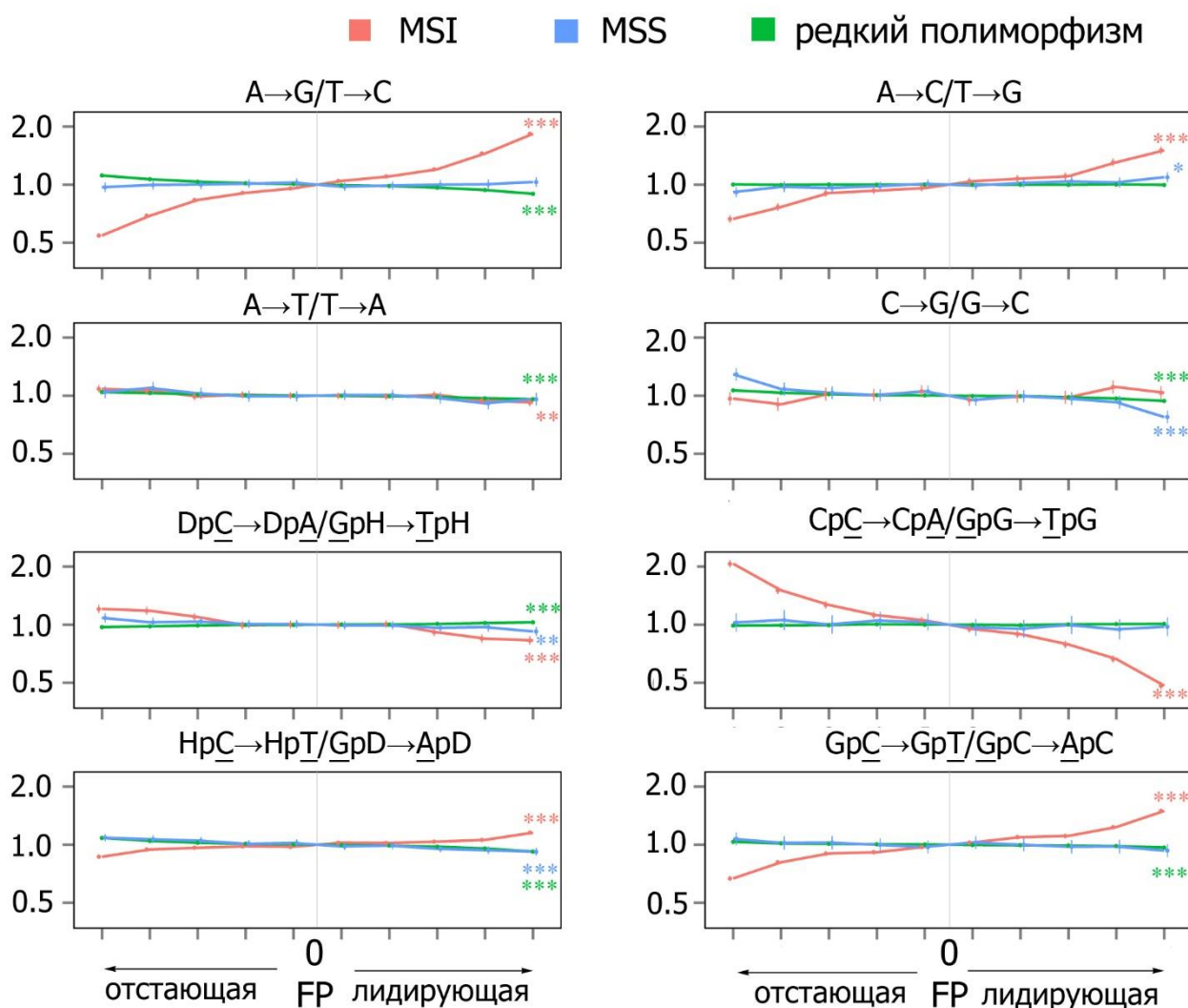


Рисунок 2.8. Репликационная асимметрия мутаций. Панели показывают отношение скоростей мутирования комплементарных мутаций в 10 раковых образцах с инактивированной MMR (3 образца COAD и 7 образцов UCEC) и 22 раковых образцах с активной MMR (8 образцов COAD и 14 образцов UCEC), а также для редкого полиморфизма человека. Используются полногеномные данные секвенирования. Оси и подписи аналогичны рисунку 2.5.

Как и в предыдущем случае, мы подсчитали асимметрию, не разбивая геном на бины по степени достоверности определения направления вилки репликации. Как и ожидалось, асимметрия снова оказалась слабее, но имеет то же направление (Таблица 2.4).

Таблица 2.4. Репликационная асимметрия в раковых образцах MSI в самом крайнем бине и с использованием всех данных для полных геномов из базы данных TCGA. Значения в скобках соответствуют 95% доверительному интервалу. Звездочки показывают значимость отклонения от 1: (*) $p < 0.05$; (**) $p < 0.001$; (***) $p < 0.001$. D соответствует нуклеотидам A, G или T. H соответствует нуклеотидам A, C или T.

Тип мутации	Крайний правый бин	Все данные
A→G/T→C	1.83*** (1.79-1.87)	1.28*** (1.27-1.29)
A→C/T→G	1.50*** (1.44-1.58)	1.18*** (1.15-1.20)
A→T/T→A	0.92*** (0.88-0.96)	0.97** (0.95-0.99)
C→G/G→C	1.04 (0.96-1.12)	1.01 (0.97-1.04)
DpC→DpA/GpH→TpH	0.83*** (0.79-0.87)	0.92*** (0.90-0.94)
CpC→CpA/GpG→TpG	0.48*** (0.46-0.51)	0.74*** (0.73-0.76)
HpC→HpT/GpD→ApD	1.15*** (1.13-1.18)	1.06*** (1.04-1.07)
GpC→GpT/GpC→ApC	1.50*** (1.45-1.54)	1.18*** (1.16-1.19)

В данных по редким полиморфным вариантам асимметрия между комплементарными мутациями значительно слабее, что соответствует предыдущим исследованиям [95]. Небольшая наблюдаемая асимметрия для некоторых типов мутаций скорее всего объясняется ошибками, не связанными с репликацией. Наиболее изученный пример – это мутация C→T, которая в избытке присутствует на отстающей цепи в клетках зародышевой линии. Вероятно, это объясняется активностью белков семейства APOBEC [96]. Таким образом, мутационная асимметрия, связанная с репликацией, в раковых образцах с активной MMR и в клетках зародышевых линий очень слаба. Напротив, в раковых образцах, где система MMR не функционирует, мы наблюдаем сильную асимметрию комплементарных мутаций между лидирующей и отстающей цепями ДНК.

Степень асимметрии сильно различается между типами мутаций. Очень сильная асимметрия (1.5-1.8) наблюдается для мутаций, которые соответствуют неправильно спаренным основаниям, эффективно исправляемым системой MMR (A→G, CpCpN→A и GpCpN→T). Асимметрия низкая или почти отсутствует для мутаций, которые почти не присутствуют в спектре MSI раковых образцов (A→T, C→G, DpCpN→A, HpCpN→T); исключение составляет мутация A→C, для которой наблюдаемая асимметрия составляет 1.5. Нам показалось удивительным, что для мутации A→C асимметрия была высокой, хотя эта мутация не связана с мутационной подписью инактивированной MMR. Мы попытались разобраться в этом и обнаружили, что скорость мутации A→C в MSI раках сильно повышена в определенных контекстах, а именно контекстах CpApG и GpApB (рисунок 2.9а). Оказалось, что асимметрия для данной мутации в данных контекстах также значительно выше, чем во всех других контекстах (рисунок 2.9б). Таким образом, мутация A→C в определенных контекстах, по-видимому, также является минорной подписью MMR.

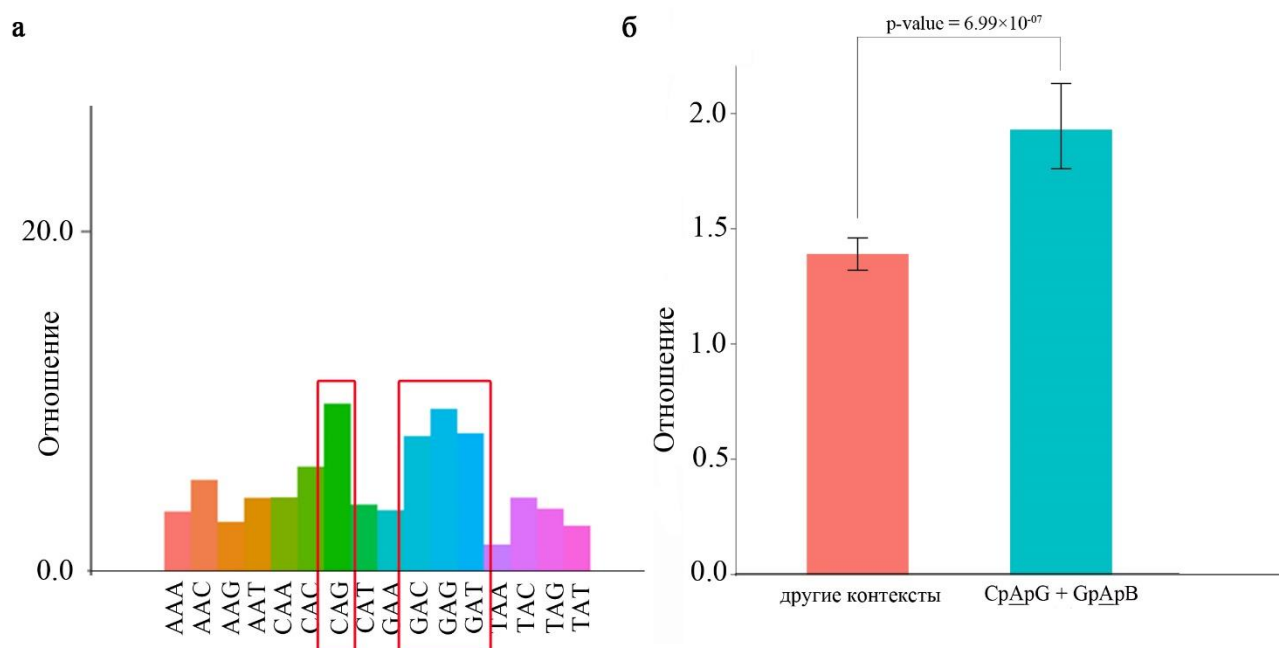


Рисунок 2.9. Мутация А→С является минорной подписью MMR. **а**, Отношение скоростей мутирования А→С во всех возможных тринуклеотидных контекстах в MSI и MSS раковых образцах. Красными рамками обозначены контексты с самыми большими значениями. **б**, Отношение скоростей мутирования А→С и Т→G на лидирующей цепи в разных контекстах в участках генома, где *fr* определено с высокой достоверностью.

Раздельный анализ образцов UCEC и COAD, а также индивидуальных образцов также показал наличие сильной асимметрии в MSI образцах (Таблицы П2.3 и П2.4), что подтверждает, что асимметрия определяется не типом рака, а активностью MMR. Наблюдаемые небольшие отличия между образцами, вероятно, свидетельствуют о примеси MMR-независимых мутационных процессов. Асимметрия имеет очень похожие значения, когда при анализе используются только межгенные интервалы (рисунок П2.4), что означает, что она не связана с транскрипцией. Мы также повторили наш анализ на независимом наборе данных - экзомных последовательностях MSI и MSS раковых образцов из TCGA, включающий большее число образцов (рисунок П2.5), разделив их по типам рака: COAD, STAD (рак желудка), UCEC. Значения асимметрии были схожими между типами рака (Таблица П2.3) и образцами (Таблица П2.4).

Среди всех однонуклеотидных ошибок система репарации неспаренных нуклеотидов наиболее эффективно исправляет однонуклеотидные вставки и выпадения. В клетках с инактивированной MMR вставки и выпадения в гомополимерных трактах и динуклеотидных tandemных повторах – это самые частые типы мутаций. В клетках человека, как и в дрожжах [43], с инактивированной MMR выпадения встречаются чаще, чем вставки, и их скорость возрастает

с длиной гомополимерного тракта. Мы изучили репликационную асимметрию наиболее частого типа выпадений – выпадений нуклеотидов А или Т в соответствующих гомополимерных трактах. Мы увидели небольшую асимметрию для данного типа мутаций (Рисунок 2.10), что подтверждает наше предположение о различии в активности MMR между цепями.

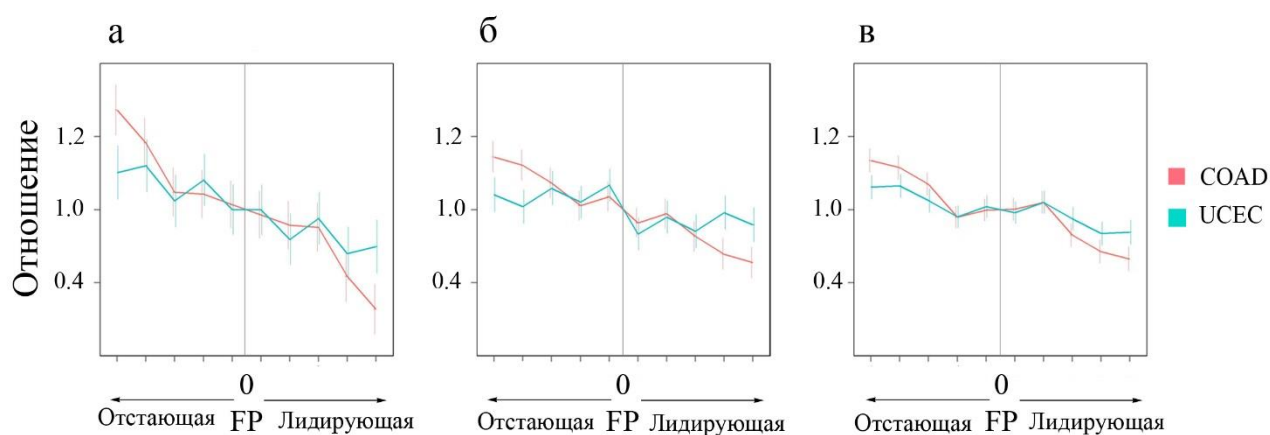


Рисунок 2.10. Репликационная асимметрия выпадений. Вертикальная ось соответствует отношению скоростей однонуклеотидных выпадений в полиА трактах и полиТ трактах в 10 геномах MSI раковых образцов (3 образца COAD и 7 образцов UCEC). Условные обозначения аналогичны рисунку 2.5. **а**, тракты длиной 10 нт; **б**, тракты длиной 7 нт; **в**, тракты длиной 8 нт.

Таким образом, асимметрия между цепями в раковых образцах с инактивированной MMR является устойчивой. В клетках с активной MMR мы не наблюдаем никакой асимметрии; значит, там она компенсируется активностью MMR. Это означает, что MMR исправляет больше ошибок на одной из двух цепей, таким образом уравнивая скорости мутирования между цепями.

2.2.5 Репликационная асимметрия в раковых образцах с инактивированной MMR соответствует асимметрии в раковых образцах с мутированной полимеразой дельта

Чтобы лучше изучить баланс между мутационным и репарационным смещениями, мы сравнили направление асимметрии в бинах с крайними значениями f_r в MSI образцах и в образцах с мутированными полимеразой дельта. Мы заметили, что асимметрия в раковых образцах с мутированной полимеразой дельта совпадает с асимметрией в MSI образцах с немутированными полимеразой дельта (рисунок 2.10), что дает основание предполагать, что полимеразой дельта вносит больше мутаций, даже когда она не поломана.

Мы повторили все наши анализы, используя другой способ определения направления вилки репликации – экспериментальные данные по секвенированию фрагментов Оказаки [83]. Результаты получились качественно очень похожими (рисунок П2.6).

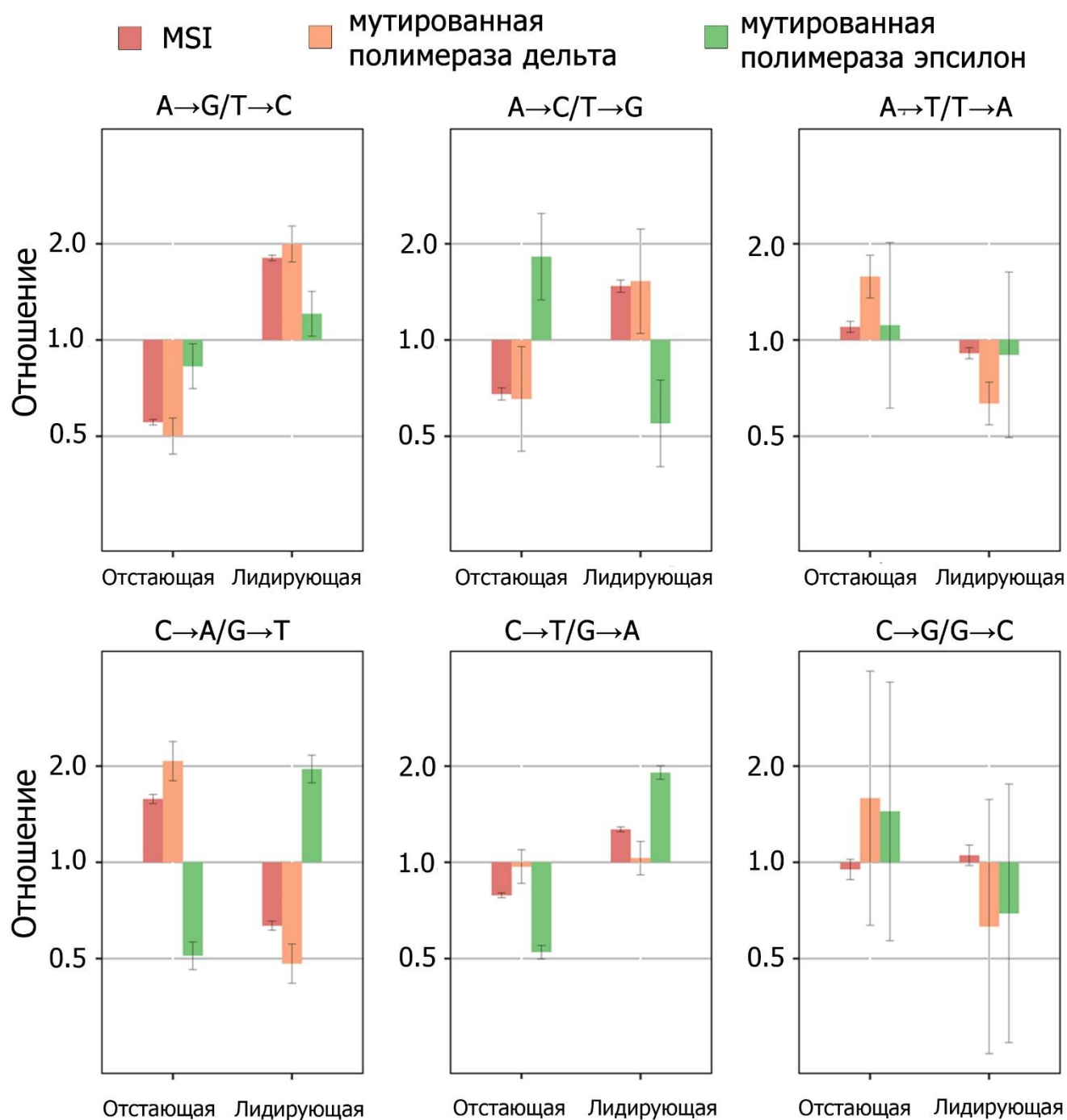


Рисунок 2.11. Репликационная асимметрия мутаций в раковых образцах с инактивированной системой MMR и в bMMRD образцах с мутированной полимеразой эпсилон или дельта. Асимметрия комплементарных мутаций по данным 20% генома, в которых направление вилки репликации было определено с высокой степенью достоверности (соответствует бинам 1 и 10 на рисунках 2.5 и 2.8). Вертикальные линии соответствуют 95% доверительным интервалам.

Более того, контексты мутаций в образцах с инактивированной MMR больше похожи на контексты мутаций в раковых образцах с мутированной полимеразой дельта, чем на контексты в раковых образцах с мутированной полимеразой эпсилон (Таблица П2.5), что также указывает на

то, что полимеразы дельта вносят больше ошибок при репликации в сравнении с полимеразой эпсилон.

Мы рассмотрели простейшую модель, используя значения асимметрии, наблюдаемые в раковых образцах с мутированной полимеразой эпсилон и в образцах с мутированной полимеразой дельта, чтобы оценить вклад этих полимераз в мутагенез (см. раздел 2.1.7). Модель предполагает, что (1) все мутации в раках с мутированной полимеразой эпсилон (мутированной полимеразой дельта) возникают в результате неправильно спаренных нуклеотидов, вставленных полимеразой при репликации лидирующей (отстающей) цепи ДНК; (2) каждая полимеразы независимо вносит определенную долю мутаций; (3) частоты этих мутаций не зависят от MMR системы в MSI раковых образцах; (4) направление вилки репликации в 20% генома, где репликационная асимметрия может быть посчитана с высоким уровнем достоверности, известна точно. Мы применили эту модель к 3 типам мутаций, частота которых сильно повышена в спектре MSI образцов, то есть эффективно исправляемых системой MMR (рисунок 2.7): A→G, C→A, C→T. Для этих типов мутаций мы обнаружили, что вклад полимеразы дельта в асимметрию, наблюдаемую в MSI раковых образцах, был в 1.9-4.4 раза выше, чем вклад полимеразы эпсилон (Таблица 2.5). С учетом частот соответствующих мутаций в общем мутационном спектре общий вклад полимеразы дельта в мутагенез примерно в 3 раза выше, чем полимеразы эпсилон (Таблица 2.5). Это означает, что основной двигатель асимметрии в MSI раковых образцах – это мутации, вносимые полимеразой дельта.

Таблица 2.5. Эффективность MMR на лидирующей и отстающей цепях ДНК. Результаты решения уравнения 1 и оценки суммарного вклада в асимметрию.

A→B	x_ϵ	x_δ	A'→B'	1- x_ϵ	1- x_δ	$x_{MMR}/(1-x_{MMR})$	1- α/α	Доля мутаций в MSI образцах	Вклад в общую асимметрию
A→G	0.547	0.334	T→C	0.453	0.666	1.81	4.27	0.33	1.41
C→A	0.662	0.675	G→T	0.338	0.325	0.63	4.39	0.12	0.53
C→T	0.656	0.492	G→A	0.344	0.508	1.27	1.89	0.38	0.72
Общая асимметрия									2.66

2.3 Обсуждение результатов

Асимметрия скоростей замен на лидирующей и отстающей цепях ДНК в клетках человека была показана ранее, однако ее причина оставалась неизвестной. Это было отчасти связано с тем, что сложно различить мутации, возникшие из неправильно спаренных оснований на комплементарных цепях. Например, избыток мутаций C→A на отстающей цепи и избыток

мутаций G→Т на лидирующей цепи будет выглядеть в данных секвенирования одинаково. В нашей работе мы попытались решить эту задачу, привязав наблюдаемые асимметрии к мутационным подписям полимераз. В частности, для мутации C→A/G→Т мы показали, что обе главные репликативные полимеразы демонстрируют избыток неправильно спаренных нуклеотидов C-dT (приводящих к мутациям C→A) в сравнении с неправильно спаренными нуклеотидами G-dA (приводящими к мутациям G→Т). Таким образом, наблюдаемая асимметрия скорее происходит из-за избытка мутаций C→A на отстающей цепи, чем мутаций G→Т на лидирующей.

Поскольку MMR – это прежде всего ко-репликативный процесс [39; 40], большинство мутаций в раковых образцах с инактивированной MMR являются ошибками репликации. Таким образом, асимметрии, наблюдаемые в раковых образцах с инактивированной MMR, выявляют асимметрии соответствующих полимераз без смещающего фактора MMR. Мы показали, что асимметрия, связанная с инактивацией MMR, совпадает с асимметрией, наблюдаемой в раковых образцах с мутированным экзонуклеазным доменом полимеразы дельта для разных типов мутаций. Это дает основание предполагать, что асимметрия в образцах с инактивированной MMR скорее всего связана с более высокой долей мутаций, вызванных полимеразой дельта во время репликации отстающей цепи.

В нашем анализе мы использовали ряд допущений. Во-первых, мы предполагали, что отношение скоростей мутирования для комплементарных мутаций, измеренное для мутантных полимераз, такое же, как для полимераз дикого типа. Это, по-видимому, правда, поскольку величина асимметрии в первую очередь отражает селективность вставки нуклеотида во время синтеза ДНК. Как минимум для мутаций C→A и G→Т уровень ошибок, посчитанный на лактазном опероне для полимеразы эпсилон дикого типа, подтверждает это [61]. Во-вторых, мы предполагали, что частоты, с которыми происходит неправильное спаривание нуклеотидов, не зависят от функциональности системы MMR. Это наиболее простое объяснение, и мы не имеем никаких данных, говорящих об обратном. В-третьих, наш подход позволяет предсказывать только предпочтительное направление вилки репликации; таким образом, величина асимметрии может быть недооценена, особенно если она очень сильная [96]. Эти допущения, однако вряд ли качественно влияют на наши результаты.

Наш анализ основан на классической модели вилки репликации, в которой каждая цепь реплицируется своей полимеразой. Недавно данная модель была подвергнута сомнениям. Согласно альтернативной модели репликации, полимеразы дельта отвечает за синтез обеих цепей ДНК, в то время как экзонуклеазная активность полимеразы эпсилон вовлечена в исправление ошибок, созданных полимеразой дельта во время репликации лидирующей цепи ДНК [36].

Однако эта новая модель противоречит большинству имеющихся данных и очень спорна [37; 38]. Многие эксперименты подтверждают классическую модель. В их числе эксперименты с мутантными полимеразми [29–31] или вставкой рибонуклеотидов [32], в которых соответствующие полимеразы наблюдались на соответствующих цепях ДНК; эксперименты, исследовавшие ассоциацию полимераз с лидирующей и отстающей цепями ДНК [33]; биохимические эксперименты по сборке и стабилизации репликативного комплекса [34; 35]. Более того, недавние исследования показали, что полимеразы эпсилон не исправляет ошибки, сделанные полимеразой дельта [97]. Новая модель также противоречит нашим данным, поскольку она не предполагает асимметрии противоположного направления, которую мы наблюдаем в образцах с мутированными экзонуклеазными доменами полимераз эпсилон и дельта. Более того, она не предсказывает асимметрии в раках с инактивированной MMR и мутированной полимеразой эпсилон, поскольку в этой модели обе цепи ДНК реплицируются полимеразой дельта. Напротив, наши данные хорошо согласуются с классической моделью вилки репликации.

Другое недавнее исследование предполагает, что полимеразы дельта могут реплицировать обе цепи ДНК около ориджинов репликации [28]. Мы считаем, что эти результаты не могут влиять на наши выводы. Даже если верно, что полимеразы дельта реплицируют обе цепи в ориджинах репликации, большая часть генома реплицируется по классической модели. Если бы большая часть генома реплицировалась только полимеразой дельта, мы не наблюдали бы противоположных паттернов асимметрии между полимеразми эпсилон и дельта.

Согласно классической модели репликации, вместе с полимеразой дельта в репликацию отстающей цепи ДНК также вовлечена полимеразы альфа, синтезирующая небольшие ДНК праймеры [98]. Считается, что перед сшиванием фрагментов Оказаки большая часть (или весь) праймер, синтезированный полимеразой альфа, удаляется и замещается ДНК, синтезированной полимеразой дельта. Недавнее исследование также показывает, что быстрая посадка ДНК-связывающих белков может мешать полимеразе дельта замещать праймеры, синтезированные полимеразой альфа, что может приводить к тому, что они остаются встроенными в новосинтезированную ДНК [99]. Поскольку полимеразы альфа действуют с более низкой точностью (ввиду отсутствия 3'-5' экзонуклеазной активности), это может увеличить скорость мутирования на отстающей цепи ДНК из-за неисправленных ошибок полимеразы альфа. Таким образом, полимеразы альфа могут вносить вклад в паттерны, наблюдаемые в раковых образцах с мутированной полимеразой дельта. К сожалению, у нас нет данных, с помощью которых мы могли бы различить вклад полимераз дельта и альфа. Однако у дрожжей с мутированной полимеразой альфа репликационная асимметрия была качественно похожа на то, что

наблюдается у дрожжей с мутированной полимеразой дельта [93], что показывает, что оба фермента вносят вклад в избыток мутаций на отстающей цепи ДНК. По-видимому, действие полимеразы альфа может влиять только на количественную оценку асимметрии разных типов мутаций.

Мутационный спектр отличается между MSI и MSS образцами с мутированными полимеразой эпсилон или дельта (рисунок 2.12). Однако уровень асимметрии очень похож между всеми образцами с мутированной полимеразой эпсилон или дельта (Таблица 2.6). Таким образом, как и в дрожжах [93], MMR, по-видимому, может изменить спектр мутирования, однако не может компенсировать очень сильную асимметрию, внесенную мутированными полимеразой эпсилон или дельта.

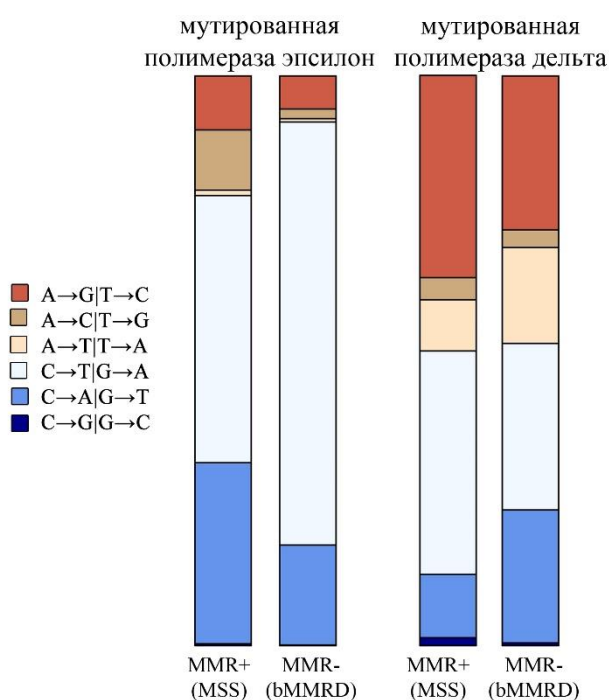


Рисунок 2.12. Частоты однонуклеотидных замен независимо от цепи ДНК в раковых образцах с активной (MMR+) и инактивированной (MMR-) системой репарации неспаренных нуклеотидов и мутированной полимеразой эпсилон или дельта.

Таблица 2.6. Отношение скоростей комплементарных мутаций на лидирующей цепи в следующих группах раковых образцов: MSS, MSI, MSS с мутированной полимеразой эпсилон, MSS с мутированной полимеразой дельта, bMMRD с мутированной полимеразой эпсилон и bMMRD с мутированной полимеразой дельта. Данные посчитаны по 20% генома, где направление вилки репликации определено с наибольшей степенью достоверности. Значения в скобках соответствуют 95% доверительному интервалу. N обозначает количество использованных для анализа образцов.

Тип мутации	MSS (n=782)	MSI (n=159)	MSS мутированная полимеразы эпсилон (n=27)	бMMRD мутированная полимеразы эпсилон (n=5)	MSS мутированная полимеразы дельта (n=1)	бMMRD мутированная полимеразы дельта (n=2)
A→G/T→C	1.03 (0.94-1.13)	1.92 (1.81-2.03)	0.95 (0.88-1.04)	1.20 (1.02-1.41)	2.39 (1.94-2.93)	1.98 (1.74-2.26)
C→T/G→A	0.89 (0.86-0.94)	1.22 (1.18-1.27)	1.93 (1.86-2.02)	2.01 (1.93-2.13)	1.08 (0.91-1.29)	1.09 (0.97-1.23)
A→T/T→A	1.10 (0.98-1.29)	0.73 (0.64-0.86)	0.62 (0.49-0.82)	0.89 (0.50-1.65)	0.44 (0.30-0.67)	0.63 (0.55-0.75)
T→G/A→C	0.87 (0.77-0.94)	0.72 (0.61-0.81)	2.27 (2.05-2.43)	1.84 (1.32-2.46)	0.54 (0.29-0.97)	0.66 (0.44-0.94)
C→A/G→T	0.92 (0.85-1.00)	0.72 (0.68-0.78)	2.14 (2.05-2.26)	1.97 (1.79-2.19)	0.51 (0.36-0.74)	0.48 (0.42-0.56)
C→G/G→C	0.86 (0.78-0.97)	0.98 (0.81-1.20)	0.74 (0.49-1.11)	0.69 (0.28-1.78)	0.43 (0.17-1.08)	0.63 (0.25-1.59)

Поскольку асимметрия, наблюдаемая в клетках с активной системой MMR, очень слабая, любая асимметрия, вносимая полимеразой, должна компенсироваться MMR. Поскольку больше неспаренных нуклеотидов вставляется на отстающей цепи ДНК, это также предполагает, что и MMR исправляет больше ошибок на отстающей цепи. Это может иметь два объяснения. Первое заключается в том, что эффективность MMR отличается между двумя цепями, так что она исправляет неправильно спаренные нуклеотиды на отстающей цепи с большей вероятностью, чем на лидирующей. Второе объяснение предполагает, что MMR имеет одинаковую эффективность на обеих цепях, сохраняя асимметрию корепликативных ошибок, но радикально уменьшая число всех типов таких ошибок. Поскольку при этом увеличивается доля симметричных нерепликативных ошибок, общая мутационная асимметрия снижается. Различить эти два объяснения можно, сравнив эффект инактивации MMR между клетками с мутированной полимеразой эпсилон и дельта. Если MMR более эффективна на отстающей цепи, то инактивация MMR должна приводить к более радикальному возрастанию скорости мутирования в клетках с мутированной полимеразой дельта по сравнению с клетками с мутированной полимеразой эпсилон. Действительно, в дрожжах инактивация MMR в 2 раза сильнее влияет на скорость мутирования в линиях с мутированной полимеразой дельта, что согласуется с большей эффективностью MMR на отстающей цепи. В человеке разрешить этот вопрос могут помочь данные по раковым образцам с активной MMR, но мутированными полимеразой. К сожалению, достаточное количество данных имеется только для образцов с мутированной полимеразой эпсилон, но не с мутированной полимеразой дельта. Однако, если MMR слабее влияет на неправильно спаренные основания на лидирующей цепи, то статус MMR должен меньше влиять на скорость мутирования в раках, где большинство мутаций приходят из неправильно спаренных оснований, производимых полимеразой эпсилон, по сравнению с раками, где обе полимеразы

вносят одинаковый вклад. В соответствии с этим ожиданием скорость мутирования отличается в MMR-активных и MMR-неактивных раках примерно в 9 раз, в то время, как это же отношение в раках с мутированной полимеразой эpsilon всего около 2 (рисунок 2.13). Таким образом, доступные данные позволяют предположить, что MMR не только удаляет больше неспаренных оснований на отстающей цепи, но и делает это более эффективно на отстающей цепи.

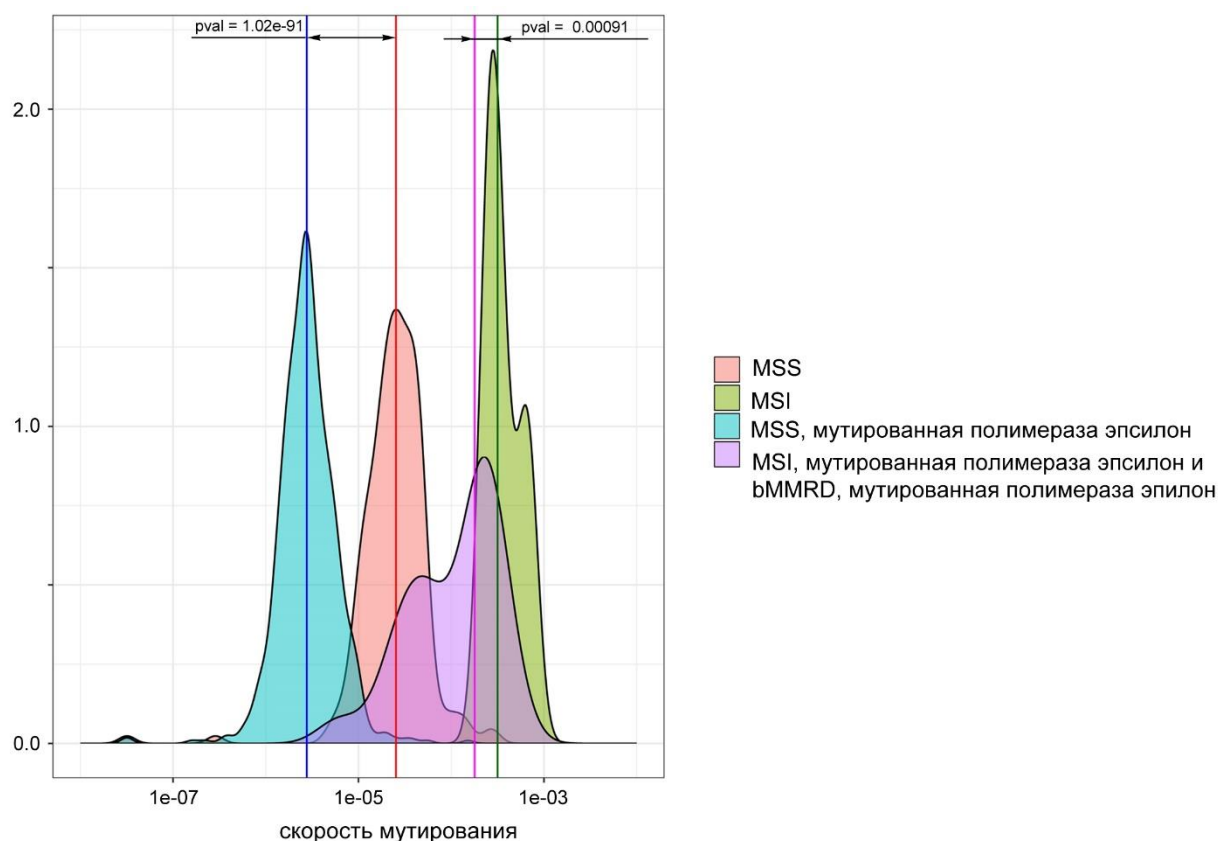


Рисунок 2.13. Распределение скоростей мутирования в образцах с активной и инактивированной MMR и мутациями в полимеразе эpsilon. Вертикальные линии соответствуют медианным значениям распределений (имеют соответствующий цвет). Логарифмическая ось x .

Таким образом, мы показали, что уровень ошибок при репликации отстающей цепи выше; что эта асимметрия появляется в основном за счет ошибок полимеразы дельта (и, возможно, альфа) на отстающей цепи; и что большее число и, по-видимому, большая доля этих неправильно спаренных оснований удаляется MMR на отстающей цепи (рисунок 2.14). Это согласуется с биохимическим свойством MMR предпочтительно удалять неправильно спаренные основания на цепи ДНК, содержащей разрывы [100]. Поскольку отстающая цепь реплицируется фрагментами Оказаки, их концы могут представлять сигнал дочерней цепи для MMR, привлекая его к этой цепи [64]. Наши наблюдения строго согласуются с экспериментами в дрожжах [93], показывая, что основные свойства MMR консервативны между дрожжами и людьми. Хотя наши результаты отражают свойства MMR в соматических клетках, они, вероятно, справедливы и для клеток

зародышевой линии. В отсутствие MMR асимметрия в числе и типах неправильно спаренных оснований, производимых при репликации лидирующей и отстающей цепей в клетках зародышевой линии, могли бы привести не только к увеличению зародышевой скорости мутирования, но также к смещению локального нуклеотидного контекста. С эволюционной точки зрения тот факт, что неравномерность появления ошибок компенсируется неравномерностью их исправлений, помогает уменьшить геномную скорость мутирования и предотвратить накопление локальных асимметрий в нуклеотидном составе.

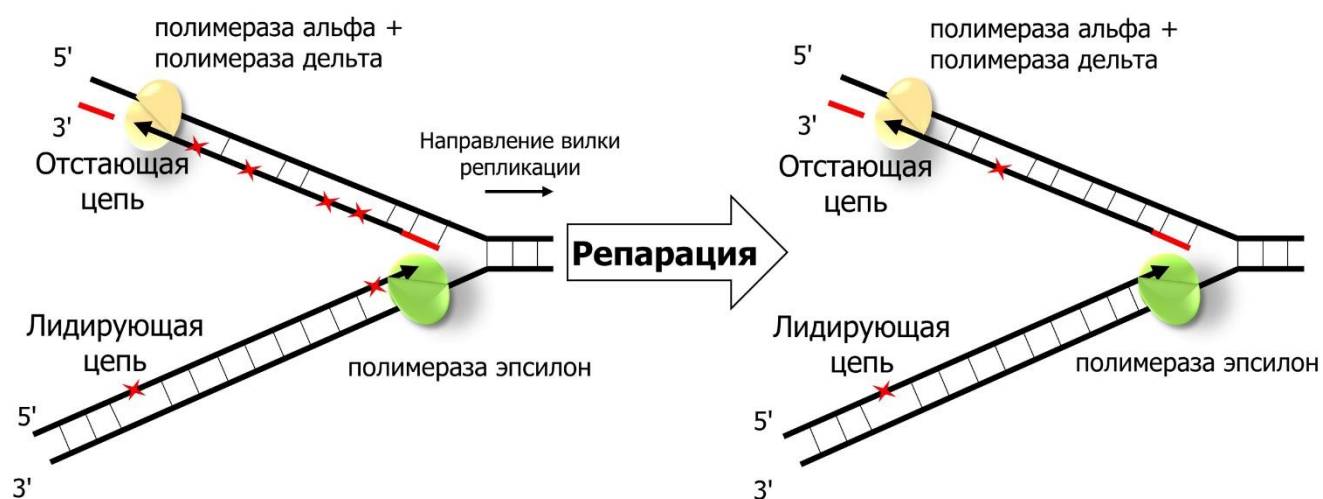


Рисунок 2.14. Схематическое изображение эффективности работы системы репарации неспаренных нуклеотидов на лидирующей и отстающей цепях при репликации ДНК. Неправильно спаренные основания (красные звездочки) чаще происходят при репликации отстающей цепи ДНК; система репарации неправильно спаренных нуклеотидов также исправляет большее число ошибок на отстающей цепи.

Глава 3. Свойства раковых образцов с биаллельной инактивацией MMR и мутацией в гене полимеразы эpsilon

Биаллельная инактивация системы репарации неспаренных нуклеотидов (bMMRD) в опухолях часто связана с соматическими мутациями в экзонуклеазных доменах ДНК-полимераз эpsilon и дельта и характеризуется особым мутационным профилем. В данной главе описывается молекулярный механизм ультрамутабельного фенотипа опухолей мозга высокой степени злокачественности в контексте bMMRD. Синдром bMMRD может быть генетической основой одной из форм синдрома Туретта (OMIM #276300) – состояния, которое характеризуется появлением рака в детском возрасте с высокой вероятностью опухолей мозга. Недавние генетические исследования показали, что глиомы пациентов с bMMRD часто имеют гипермутабельный фенотип и вследствие этого являются потенциально чувствительными к ингибиторам контрольных точек иммунного ответа [101], что подчеркивает важность ранней и точной молекулярной диагностики. Однако точная диагностика затруднена, поскольку некоторые гены, мутации в которых вызывают синдром bMMRD, расположены в участках сегментных дупликаций [102; 103], а полный список факторов генетического риска неизвестен [104]. В данной главе подробно изучен случай синдрома биаллельной инактивации системы репарации неспаренных нуклеотидов, вызванной врожденной биаллельной мутацией p.R802* в гене *PMS2*, у двух пациентов с детской глиобластомой в консангвиальной индийской семье. Геномное секвенирование опухолей в обоих случаях выявило соматическую инактивацию экзонуклеазного домена полимеразы эpsilon и ультрагипермутабельный фенотип лидирующей цепи ДНК, синтезируемой данной полимеразой.

3.1 Материалы и методы

3.1.1 Оценка микросателлитной нестабильности²

Для оценки микросателлитной нестабильности из биопсий была выделена ДНК, после чего мононуклеотидные микросателлитные маркеры BAT25, BAT26 и CAT25, а также динуклеотидные маркеры D2S123, D5S346 и D17S250 были амплифицированы с помощью полимеразной цепной реакции (ПЦР).

Смешивание компонентов ПЦР проводилось следующим образом: 1.0 мкл ДНК с содержанием 100 нг/мкл, 0.8 мкл каждого праймера (15мкмоль/л), 2.5 мкл 10× буфера и 0.25 мкл ДНК-полимеразы «Accuprime II Taq» (Thermo Fisher Scientific, Waltham, MA, USA) в общем

² Работа, методы которой описаны в разделах 3.1.1 – 3.1.7, а результаты в разделах 3.2.1-3.2.5, не выполнялась диссертантом. Описания приведены здесь для полноты.

объеме 25 мкл. Программа цикла ПЦР была следующей: исходная денатурация при 94°C в течение 1 мин, затем 4 цикла последовательных этапов при 94°C, 55°C и 72°C по 30 сек, затем 31 цикл последовательных этапов при 94°C, 54°C и 72°C по 30 сек и затем финальная элонгация при 72°C в течение 8 мин.

Аmplифицированные продукты были денатурированы при 95°C в формамидном буфере HiDye (Thermo Fisher Scientific) в течение 3 минут и проанализированы на капиллярном секвенаторе (ABI3500; Thermo Fisher Scientific) совместно со стандартом определения размеров GeneScan (GS500 LIZ; Thermo Fisher Scientific). Результаты секвенирования анализировались с помощью программного обеспечения Genemarker (Softgenetics, State college, PA, USA).

3.1.2 Описание семьи и характеристики опухолей

Три случая детской высокозлокачественной глиомы (диагностирована в 12, 12 и 15 лет) были диагностированы в большой индийской консангвиальной семье (рисунок 3.1, образцы 32, 33 и 34).

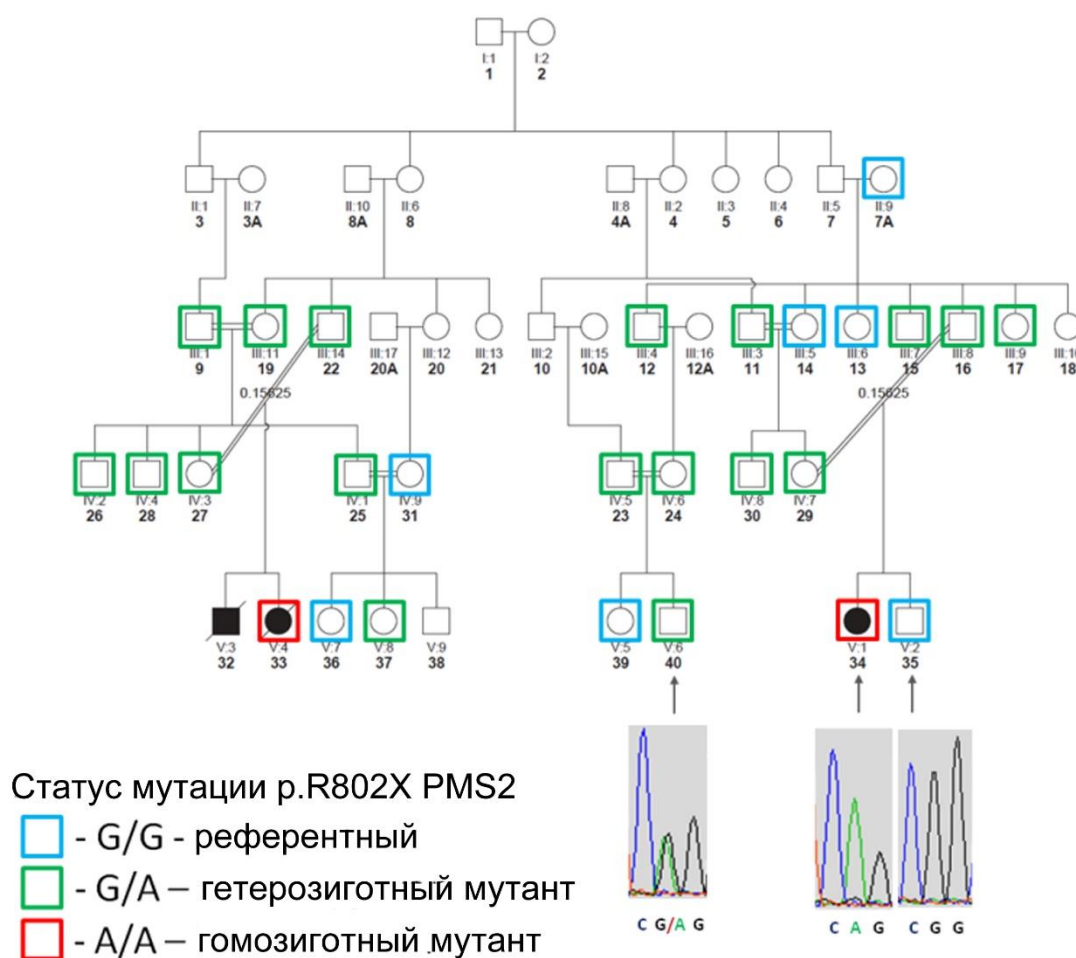


Рисунок 3.1. Родословная изучаемой консангвиальной индийской семьи. Статус мутации (p.Arg802X) в гене *PMS2* доступных ДНК образцов обозначен цветами квадратных рамок и описан на рисунке. Также для примера представлены хроматограммы секвенирования по Сэнгеру для двух гомозиготных и одного гетерозиготного образца по данной мутации.

Во всех случаях был поставлен диагноз астроцитомы или высокозлокачественная глиома. Родословная предполагает аутосомное рецессивное наследование. ДНК была доступна для 27 членов семьи (обозначены квадратными рамками разных цветов на рисунке 3.1), включая 2 из 3 больных индивидуумов (пациенты 33 и 34).

В пациенте 33 с помощью гематоксилин-эозинового окрашивания биопсии была диагностирована глиобластома IV степени злокачественности согласно международной классификации раковых опухолей (WHO, World Health Organisation). В пациенте 33 только один из 6 маркеров (мономерный повтор CAT25) показал микросателлитную нестабильность (рисунок ПЗ.1). Мы присвоили данному образцу статус образца с низкой микросателлитной нестабильностью (MSI-low, <30% нестабильных маркеров) [105].

В пациенте 34 с помощью гематоксилин-эозинового окрашивания биопсии была диагностирована анапластическая астроцитома III степени злокачественности согласно WHO. В опухоли пациента 34 ни один из 6 маркеров не показал микросателлитной нестабильности (рисунок ПЗ.1). Образцу был присвоен статус образца с микросателлитной стабильностью (MSS).

3.1.3 Экзомное и геномное секвенирование

Для каждого парафинового блока первые 5 срезов были удалены, следующие 10-20 были использованы для выделения ДНК с помощью набора «DNA FFPE Tissue kit» (QIAmp kit; Qiagen, Hilden, Germany) согласно инструкции фирмы-производителя. Качество всей ДНК было проверено с помощью гель-электрофореза или с помощью автоматической платформы для электрофореза «TapeStation genomic DNA ScreenTape» (Agilent, Santa Clara, CA, USA). ДНК из крови была выделена с помощью жидкофазной автоматизированной системы (Autopure LS; Qiagen). Экзомное обогащение и секвенирование библиотек было сделано с помощью реагентов Agilent SureSelect или с помощью KAPA Hyper Prep kit (KAPA Biosystems, Wilmington, CA, USA), с 50–200 нг исходной ДНК; обогащение было проведено с помощью Agilent SureSelect Human All Exon V5 по протоколу фирмы-производителя.

Библиотеки были секвенированы на платформе HiSeq2000 (Illumina, San Diego, CA, USA), с получением парноконцевых чтений длиной 100 пар нуклеотидов (п.н.). Медианное покрытие для экзотов составило 142× (разброс 134×-169×). Для геномного секвенирования покрытие было следующим: 26× для опухоли пациента 33, 27× для опухоли пациента 34, 23× для крови пациента

33 и 26× для крови пациента 34. Однако качество геномного секвенирования образца 33 оказалось недостаточно высоким, поэтому для последующего анализа был использован только геном пациента 34.

Чтения были откартированы на референтный геном с помощью программы BWA [106], а затем обработаны с помощью программы Genome Analysis Toolkit (GATK version 3.3.0) [107]. Качество секвенирования было проверено с помощью программы Picard tools. Для определения соматических однонуклеотидных вариантов была использована программа MuTest [108] (MuTest version 1.1.4) со стандартными параметрами, для определения соматических вставок и делеций - программа GATK HaplotypeCaller с опцией `-comp`. Полиморфизмы, обнаруженные в проектах 1000 геномов и The Exome Aggregation Consortium (EXAC), были исключены. Варианты рассматривались как потенциально драйверные, если они присутствовали в базе данных COSMIC v71 или были мутациями потери функции в известных опухолевых генах-супрессорах.

Соматические изменения копийности участков ДНК были определены из экзомных/геномных данных, учитывая аллельное покрытие, процент чтений врожденных гетерозиготных вариантов и аллельную частоту в образце. Границы и типы участков измененной копийности были определены с помощью алгоритма HMM [109] с последующей ручной проверкой.

3.1.4 Определение происхождения мутации с помощью генотипирования

Возможность общего происхождения мутации R802X, обнаруженной в индийской семье в данном исследовании и в 5 пакистанских семьях [102], была оценена с помощью 3 коротких нуклеотидных полиморфных повторов, расположенных около сайта мутаций: последовательность (CA)_n расположенная примерно 205 кб с 5' стороны от мутации (205 tel), последовательность (TA)_n, расположенная примерно 335 кб с 3' конца от мутации (335 cen), и последовательность (CA)_n, расположенная примерно по середине между двумя остальными маркерами и менее чем в 100 кб с 3' стороны от мутации (D7S481)

Размер ДНК фрагментов был оценен с помощью платформы «Fragment Analyzer INFINITY» (Advanced Analytical, Ankeny, IA, USA).

3.1.5 ПЦР длинных фрагментов и вложенная ПЦР

Для анализа длинных фрагментов ДНК смешивание компонентов ПЦР проводилось по следующему протоколу: 1.0 мкл ДНК концентрации 100 нг/мкл, 2 мкл каждого праймера (20мкмоль/л), 5 мкл 10× буфера, 8 мкл дНТФ и 0.5 мкл полимеразы Takara общим объемом 50мкл. Цикл программы ПЦР: денатурация при 94°C в течение 1 мин, 30 циклов при 98°C по 10 сек, 60°C в течение 12 мин и элонгация при 72°C в течение 10 мин.

Для вложенной ПЦР компоненты смешивались следующим образом: 1.0 мкл Takara PCR, 10 мкл Jumpstart mix и 1 мкл каждого праймера (10мкмоль/л) общим объемом 20мкл. Программа ПЦР: денатурация при 94°C в течение 3 мин, 35 циклов 94°C по 30 сек, 52°C по 30 сек, 72°C 1 мин и элонгация при 72°C в течение 5 мин. Праймеры для ПЦР длинных фрагментов и вложенной ПЦР приведены в таблице ПЗ.1.

3.1.6 Определение участков гомозиготности

Участки гомозиготности были определены согласно синонимичным однонуклеотидным полиморфизмам высокого качества экзомного секвенирования опухолей пациентов 33 и 34. В частности, участком гомозиготности назывался участок, содержащий как минимум 10 последовательных гомозиготных однонуклеотидных полиморфизмов на протяжении менее 1Мб.

3.1.7 Иммуногистохимия белков системы репарации неспаренных нуклеотидов

Иммуногистохимический анализ был произведен с помощью оборудования Ventana Benchmark Ultra Immunostainer (Ventana Medical Systems, Tucson, AZ, USA). Были использованы следующие антитела: анти-MLH1 клон M1 (Ventana Medical Systems), анти-PMS2 клон EPR3947 (Cell Marque, Rocklin, CA, USA), анти-PMS2 клон G219-1129 (Cell Marque) и анти-MSH6 клон 44 (Ventana Medical Systems). Демаскирование антигена было проведено с помощью буфера CC1 (Ventana Medical Systems) в течение 48 мин (MLH1 и MSH2) и 56 мин (MLH6 и PMS2). Проявление цвета осуществлялось с помощью набора «Optiview kit» (Ventana Medical Systems) в течение 8 мин для всех антител.

3.1.8 Определение цепеспецифичности мутаций в раковой опухоли

Репликационная асимметрия мутаций определялась аналогично разделу 2.1.3. А именно, для аппроксимации вероятности того, что референтная цепь реплицируется как лидирующая (отстающая) в данной позиции генома, была использована производная времени репликации в позиции мутации. Геном был разделен на 10 равных бинов по значению производной, где низкие значения производной соответствуют склонности данной позиции референтного генома быть реплицированной как отстающая цепь, а высокие значения – как лидирующая цепь. Для каждого бина было подсчитано число замен и число соответствующих сайтов. Каждая замена была посчитана дважды аналогично анализу раздела 2.1.3, поэтому графики асимметрии (рисунок 3.7) симметричны относительно нуля.

3.1.9 Доступные данные из ранее опубликованных работ

В данной части работы были переанализированы данные из 2 источников [80; 81]. Мы выбрали 17 образцов, в которых были проаннотированы врожденные мутации в генах системы MMR, и разделили их на 3 группы: (1) 7 bMMRD образцов без мутаций в репликативных

полимеразах эпсилон или дельта; (2) 8 bMMRD образцов с мутациями в гене полимеразы эпсилон; (3) 2 bMMRD образца с мутациями в гене полимеразы дельта.

3.1.10 Оценка доли мутаций под положительным отбором в генах-супрессорах

Чтобы оценить долю мутаций под положительным отбором в опухолевых генах-супрессорах, мы разработали статистический тест, включающий следующие шаги: (1) подсчет общего числа мутаций в 103 опухолевых генах-супрессорах из базы данных COSMIC (<http://cancer.sanger.ac.uk/census/>) и 300 случайно выбранных генах не из этого списка; (2) случайное перераспределение мутаций (1000 итераций) в этих генах с учетом тринуклеотидного мутационного спектра образца, определенного из полноэкзомных данных; (3) оценка доли потенциально вредных (“probably damaging”) мутаций (с использованием аннотации программы PolyPhen-2 [110]) на симулированных и настоящих данных; (4) оценка доли мутаций под положительным отбором согласно избытку потенциально вредных мутаций в настоящих данных по сравнению с симулированными. Доля мутаций под положительным отбором была посчитана из уравнения:

$$x = (F_{\text{наблюдаемая}} - F_{\text{симулированная}}) / F_{\text{наблюдаемая}} \quad (2),$$

где x – доля потенциальных драйверов под положительным отбором среди потенциально вредных мутаций, $F_{\text{наблюдаемая}}$ – наблюдаемая доля потенциально вредных мутаций в данном наборе генов, $F_{\text{симулированная}}$ – среднее распределения доли потенциально вредных мутаций в симулированных данных для того же набора генов.

Такой тест был проведен отдельно для набора генов супрессоров и для случайных генов в качестве контроля.

Мы также провели оценку силы нашего теста, основываясь на симулированных данных, чтобы определить минимальную долю потенциальных драйверов под положительным отбором, при которой наш тест покажет значимое (< 0.05) отклонение от симулированных данных. Оказалось, что наш тест будет показывать значимое обогащение в потенциальных драйверах только при следующих долях драйверов под положительным отбором: $>18\%$ для образца опухоли пациента 33; 23% для образца опухоли пациента 34; и 2% для всех мутабельных генов генома.

3.2 Результаты

3.2.1 Участки повышенной гомозиготности в исследуемых пациентах

Родословная изучаемой семьи предполагает рецессивное наследование детской глиобластомы (рисунок 3.1). Определение участков гомозиготности выявило, что 512 Мб

(17.68%) и 493Мб (17.03%) исследуемых геномов пациентов 33 и 34 являются гомозиготными, из них 38.2Мб (1.32%) составляют перекрывающиеся гомозиготные участки и 28.9Мб (1%) - перекрывающиеся гомозиготные участки с одинаковым аллелем (рисунок 3.2).

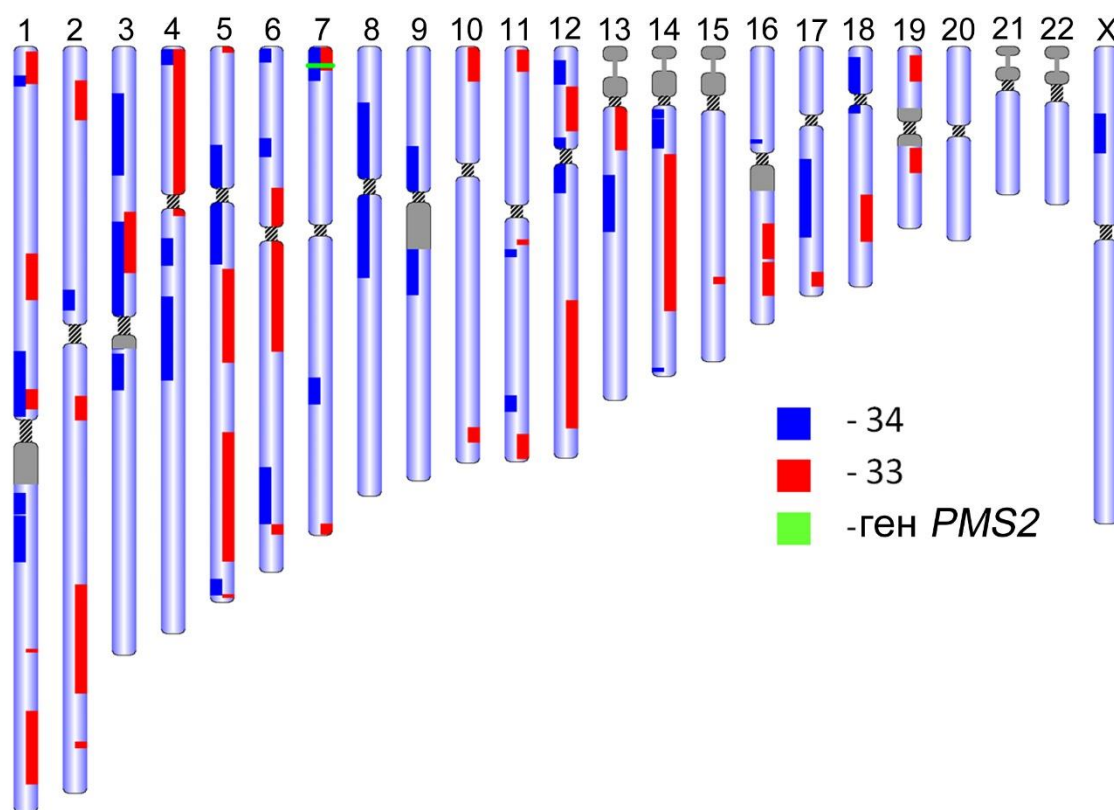


Рисунок 3.2. Участки гомозиготности в пациентах 33 (синий) и 34 (красный). Сверху подписаны названия хромосом.

3.2.2 Определение каузативного гомозиготного варианта в гене *PMS2*

Поиск мутаций с аутосомным рецессивным наследованием с помощью программы VariantMaster [111] не нашел никаких кандидатных вариантов. Однако при исследовании нуклеотидных замен в участках общей гомозиготности была обнаружена замена, приводящая к стоп-кодону в экзоне 14 гена *PMS2* (OMIM#600259), а именно NM_000535.5:c.2404C>T:p.(Arg802*) (rs63751466; не найдено в ExAC). Данная замена была проаннотирована как гетерозигота в геномах обоих пациентов и в геномах их родителей со следующими частотами закартированных чтений: пациент 33 – 40%; пациент 34 – 43%, индивидуум 27 – 17%, 22 – 18%, 29 – 23% и 16 – 38%. Исследование геномного участка с мутацией показало, что экзоны 11-15 гена *PMS2* перекрываются с 96-килобазной дупликацией гомологии 99%, расположенной в коротком плече хромосомы 7, включающей часть гена *PMS2* в качестве непротранскрибированного псевдогена *PMS2CL* (NR_002217, chr7:6 749 757 – 6 793 493, hg19).

Это наблюдение привело к гипотезе о том, что определение замен могло быть неправильным из-за ошибочного картирования чтений последних экзонов гена *PMS2* (в результате ошибочного картирования туда чтений, соответствующих гену *PMS2CL*) и соответственно неверно определенным аллельным частотам в образцах. Если данная гипотеза верна, то наблюдаемые аллельные частоты в образцах должны быть вдвое меньше ожидаемых для недуплицированного эуплоидного генома, то есть 50% для гомозиготного варианта и 25% для гетерозиготного. Чтобы проверить эту гипотезу, был проведен ПЦР длинных фрагментов таким образом, что один из праймеров находился за пределами сегментной дупликации (рисунок 3.3 и таблица ПЗ.1).

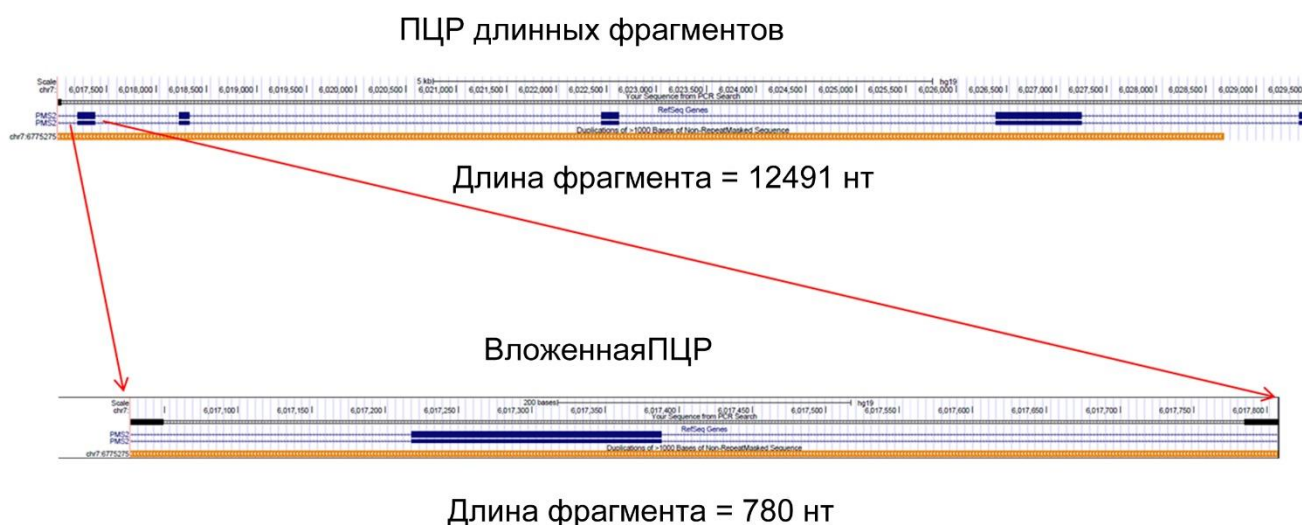


Рисунок 3.3. ПЦР длинных фрагментов и вложенная ПЦР для определения мутации p.R802X в гене *PMS2*, расположенном в участке сегментной дупликации (отмечен оранжевым).

Вложенная ПЦР с последующим секвенированием по Сэнгеру подтвердила гетерозиготность мутации p.(Arg802X) в гене *PMS2* для всех 4 родителей и гомозиготность для пациента 34. ПЦР для пациента 33 не получился, вероятнее всего из-за фрагментированности ДНК. ПЦР длинных фрагментов в гене *PMS2* для всех остальных доступных членов семьи не показал больше ни одного случая гомозиготности для мутации p.(Arg802*).

3.2.3 Отсутствие белка *PMS2* в опухолях и других клетках пациентов

Иммуногистохимия с антителами против 4 белков системы MMR (*MLH1*, *MSH2*, *MSH6* и *PMS2*) показала полное отсутствие белка *PMS2* в обеих опухолях, что подтверждает гомозиготную мутацию потери функции в гене *PMS2* (рисунок 3.4). Более того, в этих же образцах белок *PMS2* отсутствует и в других (не раковых) клетках пациентов, что согласуется с тем, что мутация является гомозиготной и врожденной.

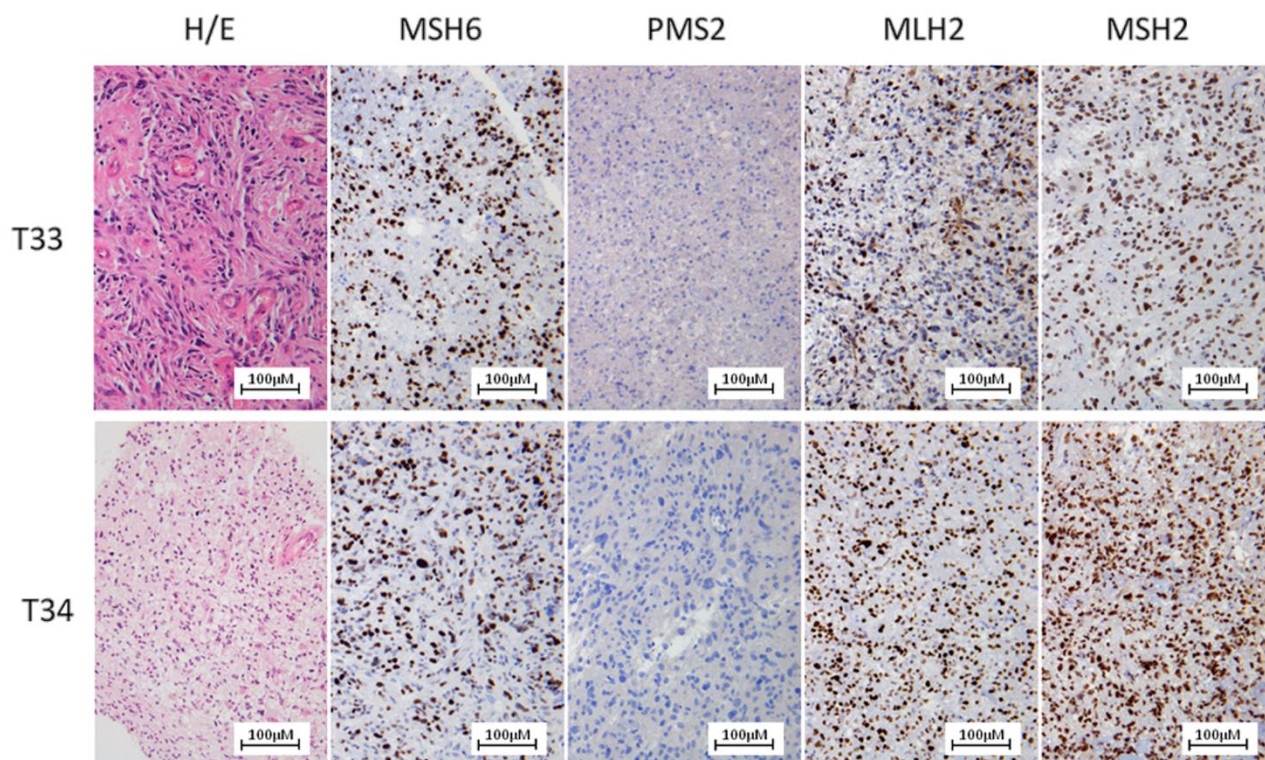


Рисунок 3.4. Иммуногистохимия для белка PMS2 и других белков системы MMR для 2 исследуемых образцов пациентов 33 и 34. Шкала 100 мкм.

3.2.4 Геномная и микросателлитная стабильность

Тест на микросателлитную стабильность выявил микросателлитную стабильность в опухоли пациента 33 (MSS) и низкую нестабильность в опухоли пациента 34 (MSI-low). Более того, было обнаружено только по 1 случаю соматического изменения копийности участков ДНК в каждой опухоли: участок потери гетерозиготности на хромосоме 9p, перекрывающий ген *CDKN2A* в опухоли пациента 33 и участок потери гетерозиготности на хромосоме 8p в опухоли пациента 34. Эти результаты согласуются с предыдущими наблюдениями о том, что ультрамутабельные опухоли мозга с врожденной bMMRD характеризуются микросателлитной стабильностью и отсутствием участков изменения копийности [80].

3.2.5 Независимое происхождение стоп-кодона в гене *PMS2*

Мутация p.(Arg802*) была ранее обнаружена в пяти пакистанских семьях [102]; более того, гаплотип мутации был общим для всех пациентов тех семей, что предполагает эффект основателя. С использованием 3 микросателлитных маркеров, определяющих пакистанский мутантный гаплотип, было показано, что вариант p.(Arg802*), найденный в изучаемой индийской семье, находится в контексте другого гаплотипа и с 5', и с 3' стороны. Эти данные предполагают независимое происхождение варианта p.(Arg802X) в гене *PMS2* в индийской семье (Таблица 3.1).

Таблица 3.1. Микросателлитные маркеры, показывающие независимое происхождение мутации R802X в гене *PMS2* в изучаемой индийской семье и в пакистанских семьях*.

	Маркер		
	205tel	D7S481	335cen
Праймеры	cataatgtccagactttgatta gggtcatcttgcttctcta	ctcattctcaccccatatt cccactgtctaaaaaaaaaat	atccaaaatcacaccactgc ctgtgggaagaacgaagtgt
Пациент 33	301/301	191/191	146/146
Пациент 34	301/301	191/191	146/146
Пакистанские семьи (01,02,04,05)*	299/299	193/193	149/149

*- Данные для пакистанских семей взяты из [102]

3.2.6 Ультрамутабельный фенотип опухолей с bMMRD и соматической мутацией в гене полимеразы эpsilon

Чтобы реконструировать этапы развития рака в пациентах 33 и 34, мы провели полноэкзомное и полногеномное секвенирование опухолей пациентов 33 и 34, а также ДНК их крови. Мы выявили ультрамутабельный фенотип обеих опухолей, а именно обнаружили 11 902 и 6185 аутомных мутаций на экзом соответственно, и 297 793 мутации на геном (с маскированными повторами) опухоли пациента 34. Таким образом, частоты мутаций в экзомах изучаемых пациентов составляют 237/Мб и 123/Мб соответственно, а частота мутаций в геноме опухоли пациента 34 - 198/Мб. Опухоли обоих пациентов несут мутации в экзонуклеазном домене полимеразы эpsilon: NM_006231.3:c.1307C>A:p.(Pro436His) в опухоли пациента 33 и c.1270C>G:p.(Leu424Val) в опухоли пациента 34. Мы попытались установить, насколько рано в ходе развития опухоли появились эти мутации. Для этого мы провели анализ аллельных частот данных мутаций и сравнили их с аллельными частотами всех мутаций в образце. Поскольку при секвенировании ракового образца происходит секвенирование большого числа раковых клеток, то аллельная частота мутации отражает, в каком количестве клеток присутствует данная мутация. Для гетерозиготных мутаций, возникших рано в ходе развития опухоли, мы ожидаем, что они будут присутствовать во многих клетках опухоли, поэтому их частота будет близка к 0.5; мутации же, возникшие позже в ходе развития опухоли, будут присутствовать только в небольшом количестве клеток, и их ожидаемые аллельные частоты будут гораздо ниже. Анализ аллельных частот исследуемых мутаций в экзонуклеазных доменах показал, что в обоих пациентах мутации в полимеразе эpsilon возникли в ходе развития опухоли рано (рисунок 3.5).

Распределение аллельных частот

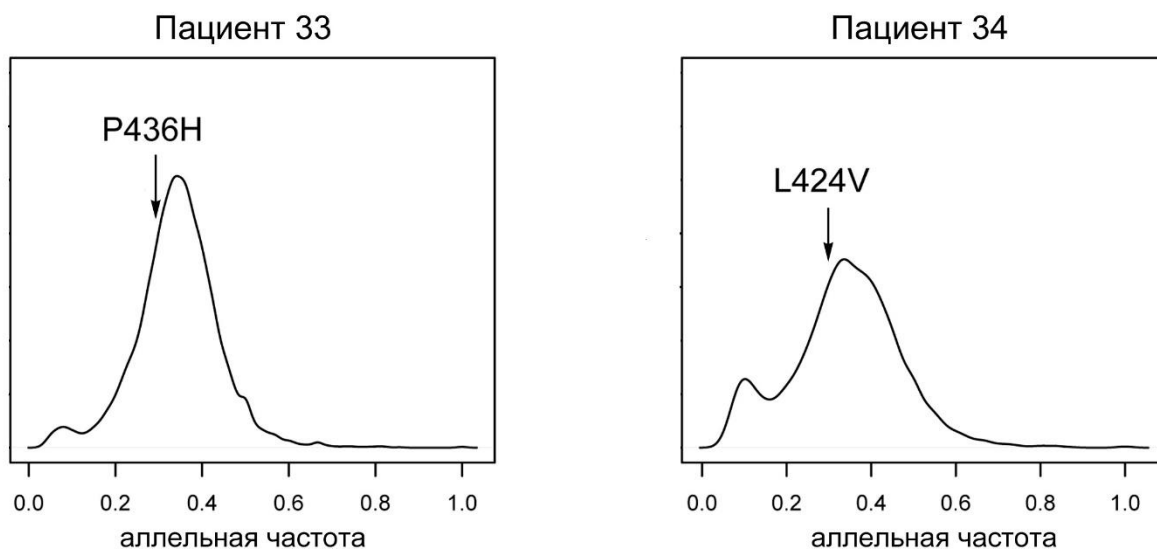


Рисунок 3.5. Распределение аллельных частот всех экзомных мутаций в опухолях пациентов 33 и 34 (число чтений, несущих мутацию, по отношению к общему числу чтений, покрывающих позицию мутации). Стрелочками обозначены аллельные частоты мутаций, инактивирующих функции экзонуклеазного домена в гене полимеразы эpsilon.

Эти мутации были и ранее обнаружены в других типах рака как соматические или мутации зародышевых линий. Также было показано, что они приводят к инактивации экзонуклеазной активности полимеразы эpsilon [61; 112]. Эти результаты совместимы с предыдущими наблюдениями, что ультрамутабельный фенотип в опухолях с bMMRD связан с мутациями в одной из репликативных полимераз. Скорость мутирования в таких опухолях увеличена в 230 раз ($> 250/\text{Mб}$) [80]. Мы исследовали мутационные свойства ультрамутабельных опухолей с bMMRD, проанализировав опухоли пациентов 33 и 34 и сравнив это с результатами, полученными при анализе опубликованных ранее данных [80; 81].

Спектр мутирования обоих образцов обогащен мутациями C→T (особенно в контексте GpCpG) и C→A (особенно в контексте NpCpT) (рисунок 3.6). Такой же паттерн наблюдался в описанных ранее образцах с bMMRD и инактивацией экзонуклеазного домена полимеразы эpsilon (рисунок 3.6).

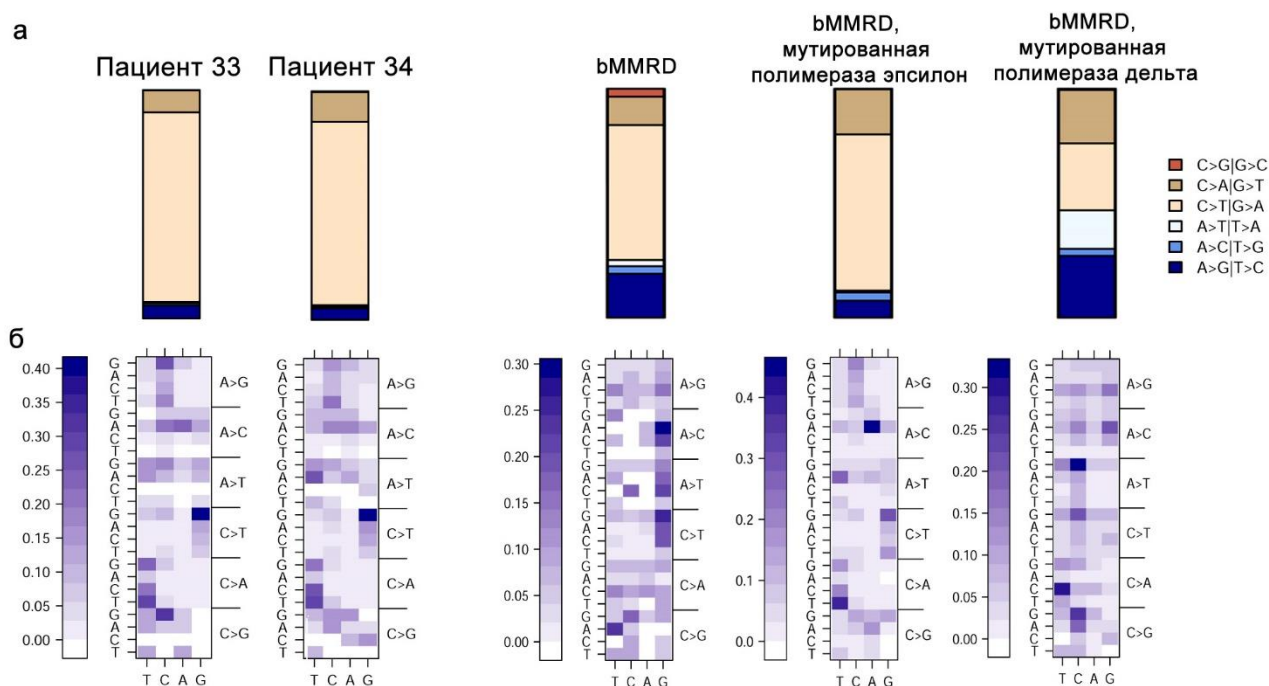


Рисунок 3.6. Мутационные спектры bMMRD образцов. **а**, относительные частоты однонуклеотидных замен в образцах 33 и 34, а также в образцах bMMRD и bMMRD с мутированными полимеразами из ранее опубликованных данных [80; 81] независимо от цепи ДНК; **б**, контекст-зависимые частоты для каждого типа мутации в том же наборе образцов. Вертикальная ось - 5' контекст, горизонтальная - 3' контекст. Для каждого типа однонуклеотидной замены цвет одной из 16 клеток показывает частоту данной мутации в определенном контексте.

3.2.7 Репликационная асимметрия опухолей с bMMRD и соматической мутацией в гене полимеразы эpsilon

Как было описано выше, классическая модель вилки репликации эукариот предполагает, что полимеразы эpsilon и дельта реплицируют лидирующую и отстающую цепи ДНК, соответственно. Для комплементарной пары мутаций $C \rightarrow A$ и $G \rightarrow T$ в экспериментах было показано, что мутированная полимеразы эpsilon вставляет неправильно спаренные нуклеотиды, приводящие к мутациям $C \rightarrow A$, чаще, чем те, что приводят к мутациям $G \rightarrow T$ [61; 113]. В главе 2.2.2 мы экстраполировали этот паттерн на клетки человека и показали, что в клетках человека мутированные полимеразы эpsilon и дельта также вставляют неправильно спаренные нуклеотиды, приводящие к мутации $C \rightarrow A$ чаще, чем те, что приводят к комплементарной мутации $G \rightarrow T$. Также для пары комплементарных мутаций $C \rightarrow T$ и $G \rightarrow A$ мы показали, что полимеразы эpsilon чаще вставляет неправильно спаренные нуклеотиды, приводящие к мутации $C \rightarrow T$. Чтобы проверить, что большинство мутаций $C \rightarrow T$ и $C \rightarrow A$ в исследуемых образцах глиобластом предпочтительно находится на лидирующей цепи ДНК из-за потери экзонуклеазной активности полимеразы эpsilon, мы оценили асимметрию между лидирующей и отстающей

цепями для исследуемых образцов аналогично анализам раздела 2.2.2 (см. раздел 3.1.8). В участках генома, где референтная цепь преимущественно синтезируется как лидирующая, мы наблюдали приблизительно в два раза больше мутаций C→A по сравнению с комплементарными G→T. Похожий результат был также получен для второй пары комплементарных мутаций C→T и G→A. Таким образом, неправильно спаренные нуклеотиды, наиболее часто вставляемые полимеразой эpsilon, происходят преимущественно на лидирующей цепи, с приблизительно двукратным избытком как в экзонах, так и в геноме. Мы сравнили асимметрию, полученную для исследуемых образцов с асимметрией для bMMRD образцов с мутированными полимеразой эpsilon и дельта. Мы увидели, что направление асимметрии согласуется с направлением для других образцов, имеющих мутацию в экзонуклеазном домене полимеразы эpsilon, и противоположно направлению, наблюдаемому в образцах с мутацией в экзонуклеазном домене полимеразы дельта (рисунок 3.7).

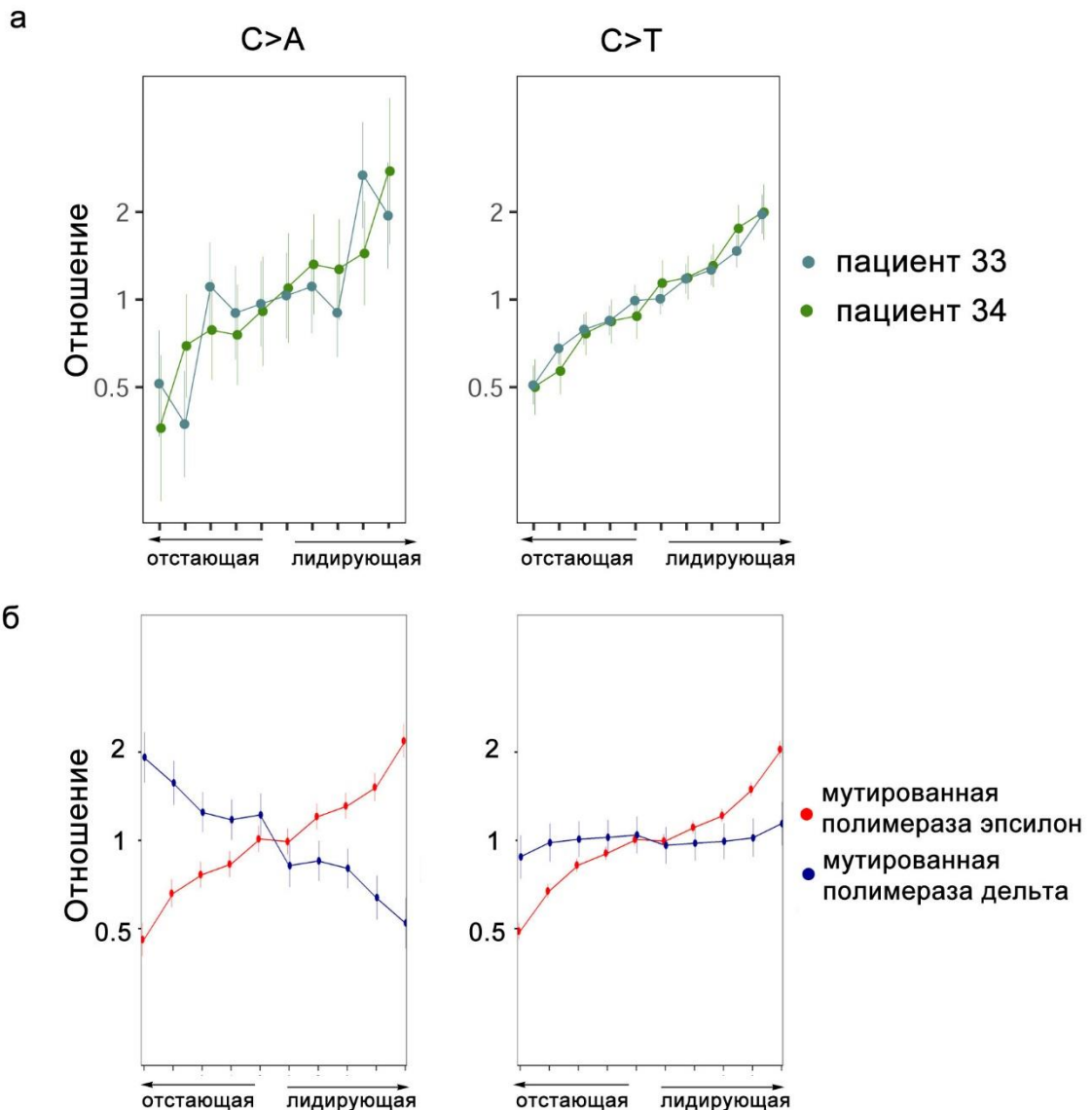


Рисунок 3.7. Отношение скоростей комплементарных мутаций **а**, в наших образцах и **б**, в опубликованных ранее данных по образцам с bMMRD и мутациями в экзонуклеазных доменах полимераза эпсилон и дельта. Горизонтальная ось: 10 бинов по значениям fp (fork polarity). Вертикальная ось: отношение скоростей комплементарных мутаций данного типа. Вертикальные линии соответствуют 95% доверительному интервалу. Вертикальная ось логарифмическая.

3.2.8 Ультрамутабельные глиобластомы несут как минимум в 2 раза меньшую долю мутаций под положительным отбором, чем другие глиобластомы

Вследствие ультрамутабельного фенотипа обе опухоли накопили высокое число потенциально драйверных мутаций в генах, кодирующих фосфатидилинозитол 3-киназу, WNT, и в других раковых путях (PARP1, ARID2, CREBBP и TP53 в опухоли пациента 33 и PIK3CA, APC и NF1 в опухоли пациента 34) (Таблица П3.2). Эти события характеризуются высокой рекуррентностью (>50) и высокими аллельными частотами (рисунок П3.2 и П3.3), что означает, что они могут быть потенциальными клональными драйверами. Однако, поскольку большинство мутаций в этих опухолях является клональными, это доказательство недостаточно для того, чтобы заключить, что они действительно являются драйверами. Более того, ультрамутабельные опухоли могут накапливать большое число потенциально вредных мутаций в драйверных генах случайно.

Чтобы предсказать наличие потенциальных драйверов в опухолевых генах-супрессорах, мы разработали статистический тест, который оценивает долю мутаций, участвующих в образовании опухоли (см. раздел 3.1.10). Этот тест основан на сравнении ожидаемой и наблюдаемой доли вредных мутаций в генах-супрессорах. Коротко говоря, мы подсчитывали количество потенциально вредных мутаций, наблюдаемое в заданном наборе генов, а затем сравнивали его с количеством потенциально вредных мутаций в этом же наборе генов при условии, что мутации мы распределяем по генам случайно, сохраняя только общую скорость мутирования и учитывая общий спектр мутирования. Случайное распределение мутаций мы производили 1000 раз. Сначала мы провели данный тест на интересующем нас наборе генов – опухолевых генах-супрессорах. Для контроля мы провели аналогичный тест для случайного набора генов не из этого списка.

Мы применили этот тест к раковым образцам пациентов 33 и 34, а также к доступным опубликованным данным ультрамутабельных и неультрамутабельных глиобластом. В образцах опухолей пациентов 33 и 34 и в образцах других ультрамутабельных глиобластом мы не обнаружили значимого избытка потенциально вредных мутаций в генах-супрессорах; однако обогащение было очень сильным в немутабельных образцах (46% потенциально вредных мутаций в генах-супрессорах находятся под положительным отбором) (рисунок 3.8г). Отсутствие

эффекта в ультрамутабельных опухолях может быть связано со слабым эффектом и недостаточной силой нашего теста. Мы оценили, что для наших ультрамутабельных образцов тест покажет значимое обогащение только если доля мутаций под положительным отбором будет выше 23% (см. раздел 3.1.10). Это означает, что доля мутаций под положительным отбором в ультрамутабельных раках ниже этого значения, то есть как минимум в 2 раза меньше, чем это значение для немутабельных глиобластом. Однако, стоит заметить, что, хотя доля мутаций под положительным отбором в генах-супрессорах ниже в ультрамутабельных опухолях, чем в немутабельных, абсолютные значения числа драйверных событий тем не менее могут быть выше в связи с очень высокими скоростями мутирования.

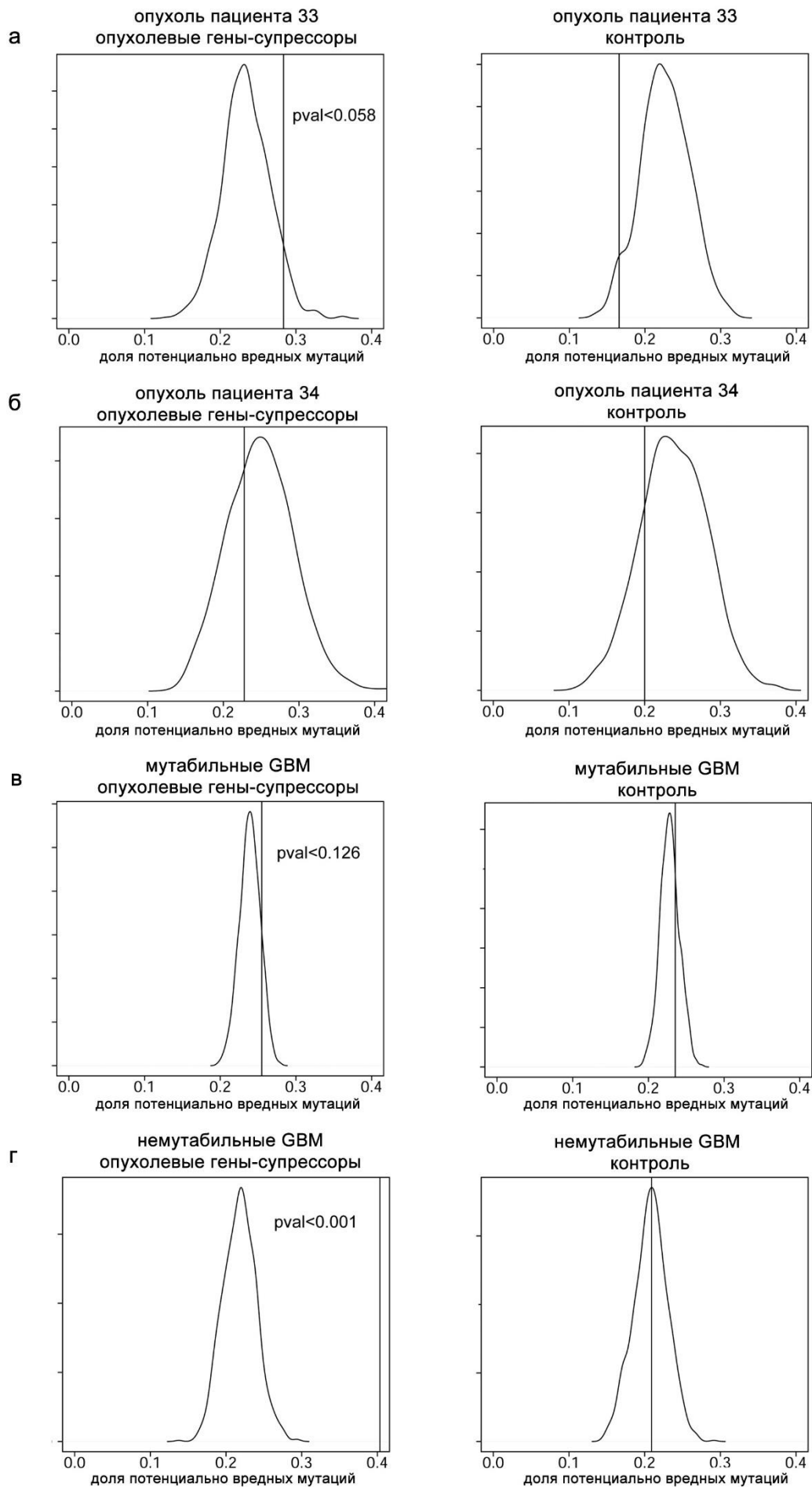


Рисунок 3.8 Наблюдаемые и ожидаемые доли потенциально вредных мутаций среди всех мутаций в генах-супрессорах опухолей (левая панель) и контрольном наборе генов (правая панель) исследуемых пациентов 33 (а) и 34 (б), других ультрамутабельных (в) и немутабельных (г) образцов глиобластом. Распределения соответствуют ожидаемым долям, полученным путем 1000 случайных «разбрасываний» мутаций по генам. Вертикальные линии соответствуют наблюдаемой доле.

3.3 Обсуждение результатов

Мутация p.(Arg802*) в гене *PMS2*, обнаруженная в данном исследовании, была ранее обнаружена в 5 семьях пакистанского происхождения с предполагаемым эффектом основателя. В исследуемых пациентах 33 и 34 эта мутация имеет независимое происхождение.

После подробного генетического анализа экзомов (и одного генома) исследуемых пациентов мы предполагаем, что гипермутабельный фенотип в описанной семье объясняется следующим: (1) накоплением большого числа ошибок, произведенных полимеразой эpsilon при репликации лидирующей цепи ДНК из-за инактивации экзонуклеазной активности этой полимеразы соматической мутацией, и (2) неактивной системой MMR из-за врожденной гомозиготной мутации зародышевой линии в гене *PMS2*.

Мы показали, что исследуемые образцы опухолей содержат большое количество потенциально вредных мутаций в генах-супрессорах раковых опухолей. Однако, несмотря на то, что эти мутации имеют высокую рекуррентность и являются клональными, они могли быть получены случайно в силу очень высоких скоростей мутирования. Мы разработали статистический тест, который помог нам показать, что доля мутаций под положительным отбором в генах-супрессорах в исследуемых нами образцах не превышала 23%, что примерно в 2 раза меньше, чем доля мутаций под положительным отбором в немутабельных глиобластомах. Однако стоит помнить, что это ничего не говорит об абсолютном числе мутаций под положительным отбором, поскольку из-за высоких скоростей мутирования абсолютное число может быть и больше по сравнению с немутабельными опухолями.

В исследуемой семье эволюция опухолей была быстрой и агрессивной, что схоже с остальными описанными ранее случаями bMMRD [80; 113]. Опухоли с мутацией в экзонуклеазном домене полимеразы эpsilon потенциально могут лечиться ингибиторами контрольных точек иммунного ответа, такими как ниволумаб [101], что объясняет важность точного молекулярного диагноза соматических мутаций и мутаций зародышевой линии в таких семьях.

Глава 4. Мутационные подписи базальноклеточной карциномы

Базальноклеточная карцинома – тип рака кожи, который включает около 90% всех случаев рака кожи [114]. Базальноклеточная карцинома происходит из стволовых клеток в интерфолликулярном эпидермисе или фолликуле волоса вследствие мутагенного воздействия ультрафиолета [115]. Среди раков кожи с точки зрения паттернов мутагенеза хорошо изучена только меланома. В этой главе приводится наш анализ мутационных паттернов базальноклеточной карциномы и их сравнение с данными по меланоме.

4.1 Материалы и методы

4.1.1 Мутационные данные

В данной части работы были проанализированы мутационные данные, полученные для 68 экзомов базальноклеточной карциномы и 371 экзома меланомы. Данные по меланоме были взяты из базы данных TCGA [116]. Данные по базальноклеточной карциноме были получены коллегами в рамках данного проекта.

4.1.2 Сбор образцов и секвенирование³

Были получены свежие образцы опухолей базальноклеточной карциномы, а также образцы крови из пациентов. ДНК из опухолей была выделена с помощью набора «DNeasy Blood and Tissue kit» (Qiagen) согласно протоколу фирмы-производителя, ДНК из крови была очищена с помощью прибора «Qiagen Autopure LS». Качество ДНК всех образцов было проверено на геле-электрофорезе или с помощью автоматической платформы для электрофореза «TapeStation Genomic DNA ScreenTape» (Agilent). Экзомное обогащение и подготовка библиотек были проведены с использованием набора «SureSelect Human All Exon V5» (Agilent) согласно протоколу фирмы-производителя. Библиотеки были секвенированы на платформе HiSeq 2000 (Illumina) с получением парноконцевых чтений длиной 100 п.н. Медианное покрытие образцов составило 100× (76–204×). Чтения были картированы на референтный геном с помощью программы BWA [106] и затем обработаны с помощью программы Genome Analysis Toolkit (GATK; версия 3.3.0) [107]. Качество секвенирования и обогащения было проверено с помощью метрик программы Picard. Для определения соматических мутаций была использована программа MuTect (версия 1.1.4) [82] со стандартными параметрами. Известные из базы данных 1000 геномов полиморфизмы были отфильтрованы.

³ Работа по сбору образцов, а также ДНК- и РНК-секвенированию не является частью диссертационной работы, данные приводятся для полноты картины. Мутационный анализ проведен совместно с коллегами при непосредственном участии диссертанта.

4.1.3 Экспрессия в образцах меланомы и базальноклеточной карциномы

Секвенирование РНК было проведено для 61 образца опухолей базальноклеточной карциномы. Для приготовления библиотек было использовано от 500 нг до 2 мкг общей ДНК. Библиотеки были приготовлены с использованием набора «TruSeq RNA sample prep kit» (RS-122-2001, Illumina) по инструкции производителя. Библиотеки были секвенированы на платформе «HiSeq 2000» (Illumina) с использованием парноконцевых чтений длиной 100 п.н. Чтения были картированы на референтный геном (сборка 38) с помощью программы TopHat. Для подсчета покрытия каждого экзона были использованы уникально картированные чтения (аннотация GENCODE v21). Для нормализации на длину экзона и размер библиотеки для каждого экзона было посчитано значение FPKM (fragments per kilobase per million mapped reads, число фрагментов на тысячу п.н. на миллион картированных чтений), поскольку секвенирование РНК проводилось парноконцевыми чтениями. Для меланомы данные по экспрессии генов были взяты из базы данных TCGA [116]. Данные были получены для 471 образца меланомы. Значения экспрессии были указаны в RPKM (reads per kilobase per million mapped reads, число чтений на 1000 п.н. на миллион закартированных чтений). Тот факт, что значения экспрессии выражаются разными мерами для двух типов рака, не влияет на проведенный анализ, поскольку экспрессия нигде не сравнивается напрямую между типами рака; она была использована лишь для того, чтобы разделить гены на категории по уровню экспрессии внутри каждого типа рака отдельно. Для обоих типов рака экспрессия была нормирована внутри каждого образца на экспрессию гена *TBP* и затем усреднена по всем образцам. Для анализа были взяты гены, имеющие ненулевую экспрессию в обоих типах рака.

4.1.4 Подсчет числа замен

Скорости мутирования везде подсчитывались как число замен определенного типа мутаций, разделенное на число сайтов соответствующего типа.

4.2 Результаты и обсуждение

Мутационный профиль базальноклеточной карциномы соответствует воздействию ультрафиолета как основного мутагенного фактора: в среднем, 90% однонуклеотидных вариантов – это замены C→T в контексте YpC или CpY (где Y=C или T), и 5% всех мутаций – это динуклеотидные замены (рисунок 4.1).

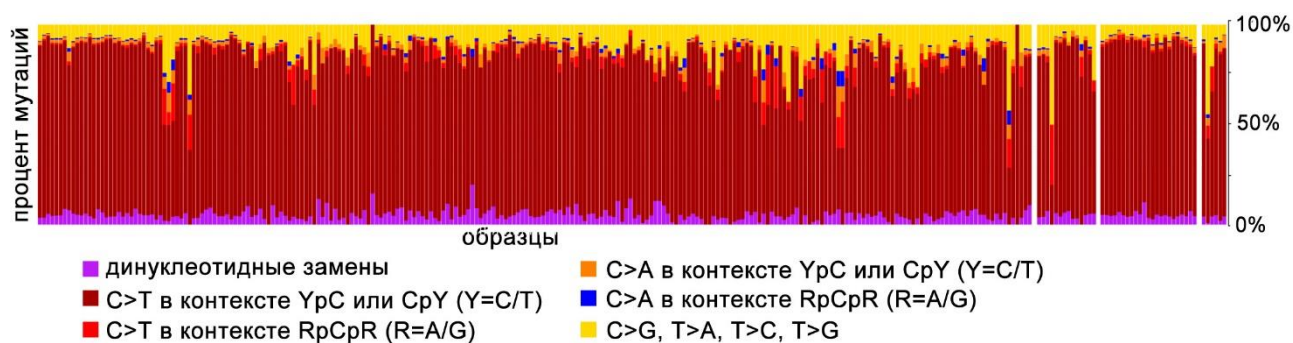


Рисунок 4.1. Мутационный спектр изученных образцов базальноклеточной карциномы.

Чтобы лучше понять мутационные процессы, действующие в базальноклеточной карциноме, мы сравнили ее мутационный профиль с мутационным профилем меланомы. В частности, мы исследовали мутагенез, вызванный УФ и окислительным стрессом, а также эффективность действия TC-NER.

Для начала мы сравнили, насколько отличается вклад мутаций, вызванных УФ, в этих двух типах рака кожи. Известно, что УФ вызывает мутации C→T (особенно часто в тринуклеотидном контексте TrCpC), а также динуклеотидные мутации CC→TT [3; 117; 118]. Оказалось, что доля всех мутаций, вызванных УФ, в базальноклеточной карциноме выше, чем в меланоме (90% против 85% мутаций, $p < 2.2 \times 10^{-16}$). Мы также оценили доли мутаций C→T в тринуклеотидном контексте TrCpC (очень специфическом контексте УФ) среди всех мутаций цитозина, а также долю мутации CC→TT среди всех двойных мутаций из CC. В базальноклеточной карциноме оба типа мутаций оказались более частыми в сравнении с меланомой. Мутации C→T в контексте TrCpC происходили на 2% чаще, чем в меланоме ($p < 2.2 \times 10^{-16}$, тест χ^2), а мутации CC→TT – на 15% чаще ($p < 2.2 \times 10^{-16}$, тест χ^2) (Таблица 4.1).

Таблица 4.1. Доли мутаций, вызванных УФ, в базальноклеточной карциноме и в меланоме. R=A/G.

	Тип мутации	Контекст	Базальноклеточная карцинома	Меланома
УФ	C→T	TrCpC	0.25	0.23
	CC→TT		0.96	0.81
УФ	C→A	CpC	0.35	0.53
Окисление гуанина	C→A	RpCpR	0.21	0.10

Затем мы оценили эффективность работы TC-NER. Известно, что повреждения ДНК, вызванные УФ, исправляются только с помощью NER [47–49]. Действие этой системы репарации основано на распознавании повреждения ДНК РНК-полимеразой в ходе процесса транскрипции,

поэтому данная система репарации эффективно исправляет повреждения ДНК, находящиеся на транскрибируемой цепи ДНК. В результате этого скорости мутирования на нетранскрибируемой и транскрибируемой цепях ДНК могут сильно различаться при эффективном действии системы TC-NER. Таким образом, определить эффективность работы данной системы можно, сравнив отношения скоростей мутирования на транскрибируемой и нетранскрибируемой цепях, то есть сравнив значения транскрипционной асимметрии. Мы сконцентрировались на анализе транскрипционной асимметрии мутаций, вызванных УФ, в частности мутации C→T в контексте TrCpC. Отношение скоростей мутирования на транскрибируемой и нетранскрибируемой цепях ДНК в базальноклеточной карциноме оказалось ниже, чем в меланоме (0.59 против 0.68, $p < 2.2 \times 10^{-16}$, тест χ^2), что показывает, что TC-NER более эффективна в базальноклеточной карциноме, чем в меланоме.

Таблица 4.2. Отношение скоростей мутирования на транскрибируемой и нетранскрибируемой цепях для УФ мутаций и не УФ мутаций в меланоме и базальноклеточной карциноме. $R=A/G$.

	Тип мутации	Контекст	Базальноклеточная карцинома	Меланома
УФ	C→T	Tr <u>C</u> pC	0.59	0.68
	CC→TT		0.57	0.59
УФ	C→A	Cp <u>C</u>	1.09	0.89
окисление гуанина	C→A	Rp <u>C</u> pR	1.26	1.27

Более того, известно, что скорость мутирования зависит от уровня экспрессии генов, поскольку в высокоэкспрессируемых генах TC-NER исправляет больше повреждений. Действительно, оказалось, что при рассмотрении отдельно группы высокоэкспрессируемых генов, эффект TC-NER оказался в базальноклеточной карциноме еще более сильным, чем в меланоме (0.21 против 0.29, $p < 2.2 \times 10^{-16}$, тест χ^2) (рисунок 4.2).

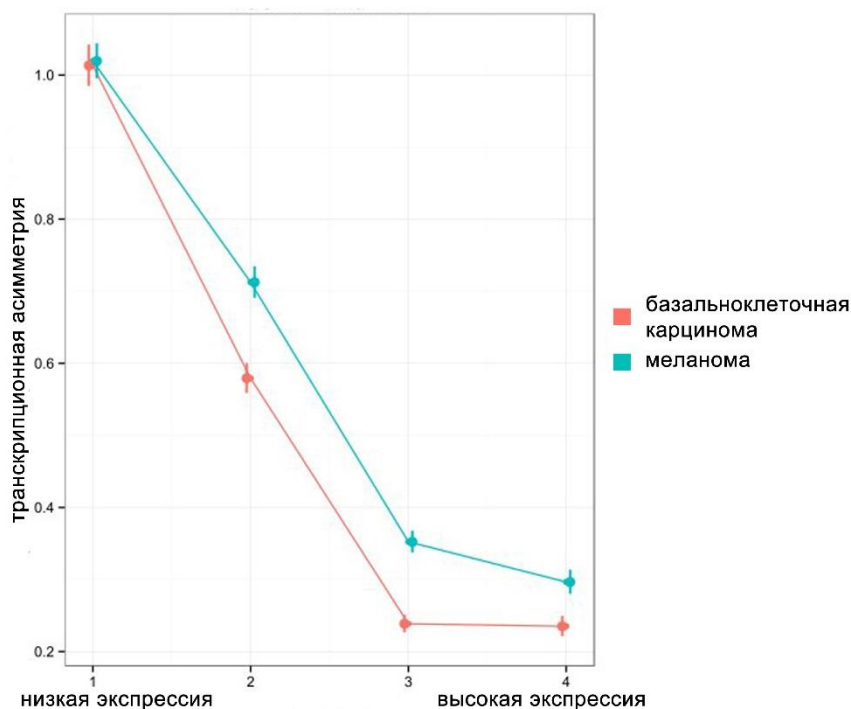


Рисунок 4.2. Транскрипционная асимметрия (отношение скоростей мутирования на транскрибируемой и нетранскрибируемой цепях ДНК) мутации $\text{TrCpC} \rightarrow \text{T}$ в базальноклеточной карциноме и меланоме в генах с разным уровнем экспрессии. Экспрессия генов специфична для каждого типа рака. 95% доверительные интервалы обозначены вертикальными линиями.

Затем мы подробно изучили мутации $\text{C} \rightarrow \text{A}$, которые могут происходить в результате либо окислительного стресса [119], либо воздействия УФ [120]. Мы выбрали для анализа мутации $\text{C} \rightarrow \text{A}$ в контексте RpCpR ($\text{R}=\text{A}$ или G), который слабо подвержен воздействию УФ, поскольку цитозины в этих контекстах не образуют фотодимеры при облучении УФ [121]. Доля мутаций $\text{C} \rightarrow \text{A}$ в контексте RpCpR была в два раза выше в базальноклеточной карциноме, чем в меланоме (21% в ВСС против 10% в меланоме) (таблица 4.1). В этом контексте мутации $\text{C} \rightarrow \text{A}$ могут быть обогащены в результате окисления гуанина. Мы проанализировали отношение скоростей мутирования $\text{C} \rightarrow \text{A}$ в контексте RpCpR на транскрибируемой и нетранскрибируемой цепях ДНК. Мутации $\text{C} \rightarrow \text{A}$ наблюдались в избытке на транскрибируемой цепи как в базальноклеточной карциноме, так и в меланоме (исследуемое отношение равно 1.26 и 1.27, в обоих типах рака $p < 0.001$, тест χ^2). Это указывает на то, что на самом деле мутация представляет собой повреждение гуанина, то есть мы наблюдаем избыток мутаций $\text{G} \rightarrow \text{T}$ на нетранскрибируемой цепи в результате действия TC-NER на транскрибируемой (таблица 4.2).

Затем мы выбрали мутации $\text{C} \rightarrow \text{A}$ в контексте CpC - контексте УФ мутагенеза, где эти мутации встречаются наиболее часто (рисунок 4.3). В каждой Т-клеточной лимфоме, для которой

была показана подпись УФ [122; 123], мутации С→А происходят специфично именно в СрС контексте.

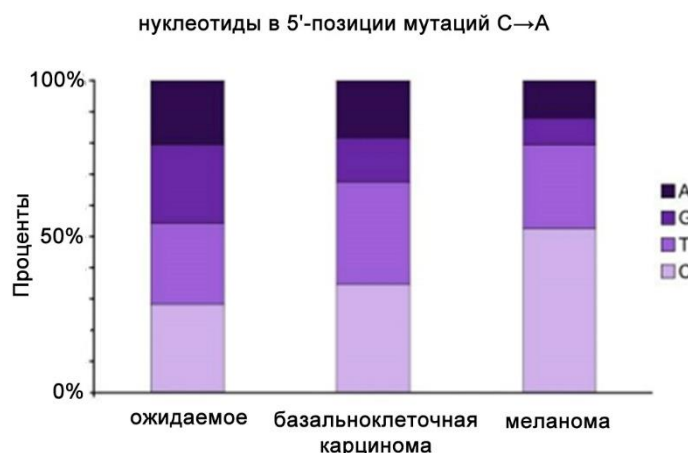


Рисунок 4.3. Доля разных нуклеотидов в 5' позиции мутаций С→А в меланоме и базальноклеточной карциноме.

Доля мутаций С→А в контексте СрС (обогащенном УФ мутациями) была в 1.6 раз выше в меланоме, чем в ВСС (53% в меланоме против 35% в ВСС) (таблица 4.1). Мы не обнаружили статистически значимой транскрипционной асимметрии для мутаций С→А в базальноклеточной карциноме (1.09 $p = 0.10$, тест χ^2), в то время как в меланоме был значимый избыток мутаций С→А на нетранскрибируемой цепи (0.89 , $p < 0.001$, тест χ^2) (таблица 4.2). Принимая во внимание эффект ТС-NER, возможно, что С→А мутации в контексте СрС в базальноклеточной карциноме представляют собой смесь мутаций гуанина и цитозина, в то время как в меланоме мутации цитозина являются более частыми.

Эти данные показывают, что механизм мутаций С→А отличается в базальноклеточной карциноме и в меланоме. По-видимому, в базальноклеточной карциноме мутации С→А в основном происходят в результате окисления гуанина, в то время как в меланоме вклад от мутирования цитозинов под действием УФ и окисления гуанинов примерно одинаков.

Таким образом, подробный мутационный анализ образцов базальноклеточной карциномы и сравнение с данными по меланоме привел нас к следующим выводам:

1) Мутации, которые являются характерной подписью УФ, составляют большую долю от общего числа мутаций в базальноклеточной карциноме, чем в меланоме. Эффективность

репарации этих мутаций с помощью TC-NER в базальноклеточной карциноме также выше, чем в меланоме.

2) Мутации C→A в базальноклеточной карциноме и в меланоме, по-видимому, имеют разную природу. Доля таких мутаций в контексте, характерном для УФ, выше в меланоме, в то время как в контексте, не характерном для УФ, выше в базальноклеточной карциноме. Анализ транскрипционной асимметрии показал, что в меланоме есть избыток мутаций C→A в ультрафиолетовом контексте на нетранскрибируемой цепи, в то время как избыток G→T мутаций в не УФ контексте на нетранскрибируемой цепи есть как в меланоме, так и в базальноклеточной карциноме. Таким образом, это приводит нас к заключению о том, что мутации C→A в базальноклеточной карциноме являются на самом деле мутациями окисления гуанина, в то время как в меланоме это смесь мутаций окисления гуанина и мутаций цитозина под действием УФ.

Глава 5. Изменение мутационных подписей в ходе развития раковой опухоли

Мутации, наблюдаемые в опухоли, происходят не одновременно, а накапливаются в течение длительного времени вследствие действия различных мутационных процессов, некоторые из которых действовали еще до превращения клетки в опухолевую. В связи с этим паттерны мутирования также могут изменяться со временем. Эпизодическое воздействие специфических мутационных процессов может приносить мутации определенного типа или с определенными свойствами. Так, например, действие белка APOBEC на ранних или поздних стадиях развития рака вносит мутации, распределение которых слабо зависит от времени репликации, в то время как остальные мутации, накопленные в другие периоды времени, имеют сильную зависимость от времени репликации.

В недавних исследованиях мутации, наблюдаемые в опухоли, начали разделять по времени их происхождения. Даже секвенирование фрагмента опухоли, а не отдельных клеток, позволяет разделить мутации, наблюдаемые в образце, согласно времени их появления [3; 7]. Мутации, накопленные в клетках-предшественниках рака и на ранних стадиях развития рака, присутствуют почти во всех клетках ракового образца и называются клональными мутациями. Мутации, появившиеся на поздних стадиях развития опухоли, присутствуют только в части раковых клеток и называются субклональными. Доля раковых клеток, содержащая мутацию (CCF, cancer cell fraction), может быть использована для аппроксимации времени появления мутации. Предыдущие исследования фокусировались в основном на сравнении тринуклеотидных спектров клональных и субклональных мутаций. Однако для определения клональных и субклональных драйверов важна оценка локальных скоростей мутирования [124], и изменение распределения мутаций вдоль генома в ходе развития раковой опухоли представляет большой интерес.

5.1. Материалы и методы

5.1.1 Данные

Мы получили данные по мутациям и файлы в формате bam из базы данных TCGA. С использованием файлов в формате bam мы определили для каждой мутации ее аллельную частоту (vaf, variant allele frequency).

Мы также загрузили данные Level1 чипов Affymetrix из базы данных TCGA. Данные Affymetrix SNP6 для пар опухоль-норма были нормализованы и предварительно обработаны с использованием пакета PennCNV [125] с настройками по умолчанию. Изменение копийности в

опухолях, ploидность и примесь нормальных клеток были определены с помощью пакета ASCAT. В качестве контроля были использованы образцы здоровых тканей. Координаты соответствуют сборке hg19. Половые хромосомы были исключены из анализа.

5.1.2 Оценка доли раковых клеток, содержащих мутацию

Для каждой мутации аллельная частота (vaf) зависит от локальной копийности опухоли (CPN_{mut}), чистоты ракового образца ($purity, p$), локальной копийности здоровой клетки (CPN_{norm}), а также доли раковых клеток, несущих мутацию. Ожидаемая аллельная частота может быть посчитана следующим образом:

$$vaf = \frac{p \times CCF}{CPN_{norm} \times (1-p) + CPN_{mut} \times p} \quad (2)$$

Рассмотрим простейший пример (рисунок 5.1). В данном примере образец содержит 3 раковые клетки и 2 здоровые, следовательно, чистота образца составляет $3/5$. Здоровые клетки имеют копийность 2, раковые – 4. Одна из трех раковых клеток несет мутацию, то есть $CCF = 1/3$. Ожидаемая аллельная частота в данном случае составляет $1/16$.

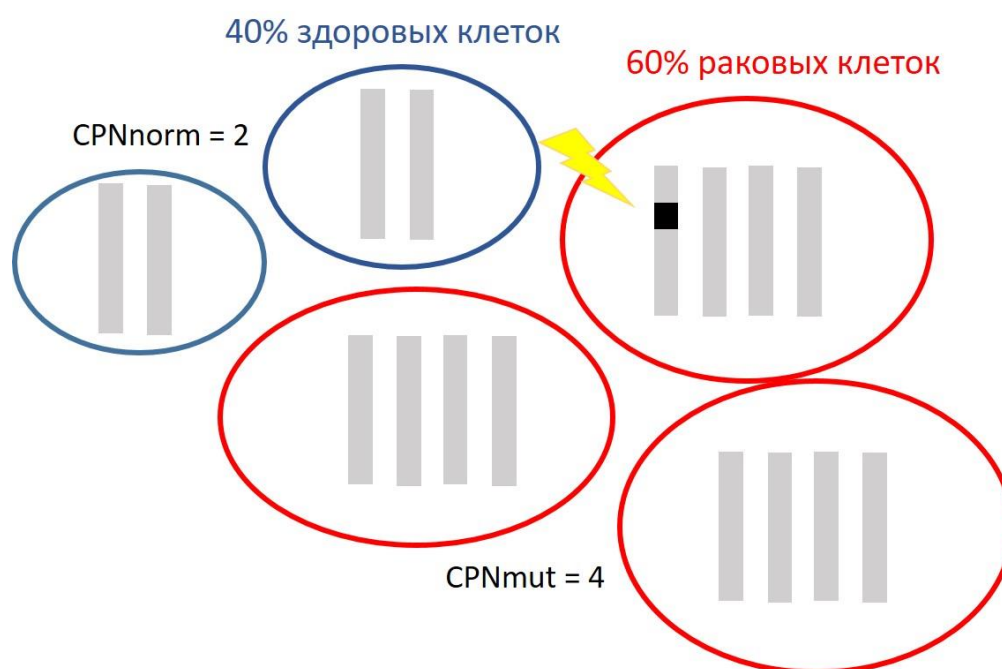


Рисунок 5.1. Схематическое изображение клеток секвенированного ракового образца. Пример, изображающий образец, содержащий 3 раковые клетки (красные овалы) и 2 здоровые (синие овалы). Серым обозначены хромосомы внутри клеток. Все раковые клетки имеют полногеномную дупликацию. Черным обозначена мутация в раковой клетке, представленная только в 1 клетке из 3.

Используя уравнение 2, мы подсчитали CCF для каждой мутации. Чистоту и копийность мутации мы получили с помощью пакета ASCAT; аллельную частоту для каждой мутации мы получили, разделив число чтений, несущих мутацию, на общее число чтений, покрывающих

данную позицию; копияность здоровой клетки везде приняли равной 2. Далее мы разделили мутации на три категории согласно их CCF: «ранние клональные» ($CCF > 1.5$), «поздние клональные» ($0.7 < CCF < 1.5$) и «субклональные» ($CCF < 0.7$). На рисунке 5.2 представлено схематичное изображение алгоритма разделения мутаций на клональные и субклональные.

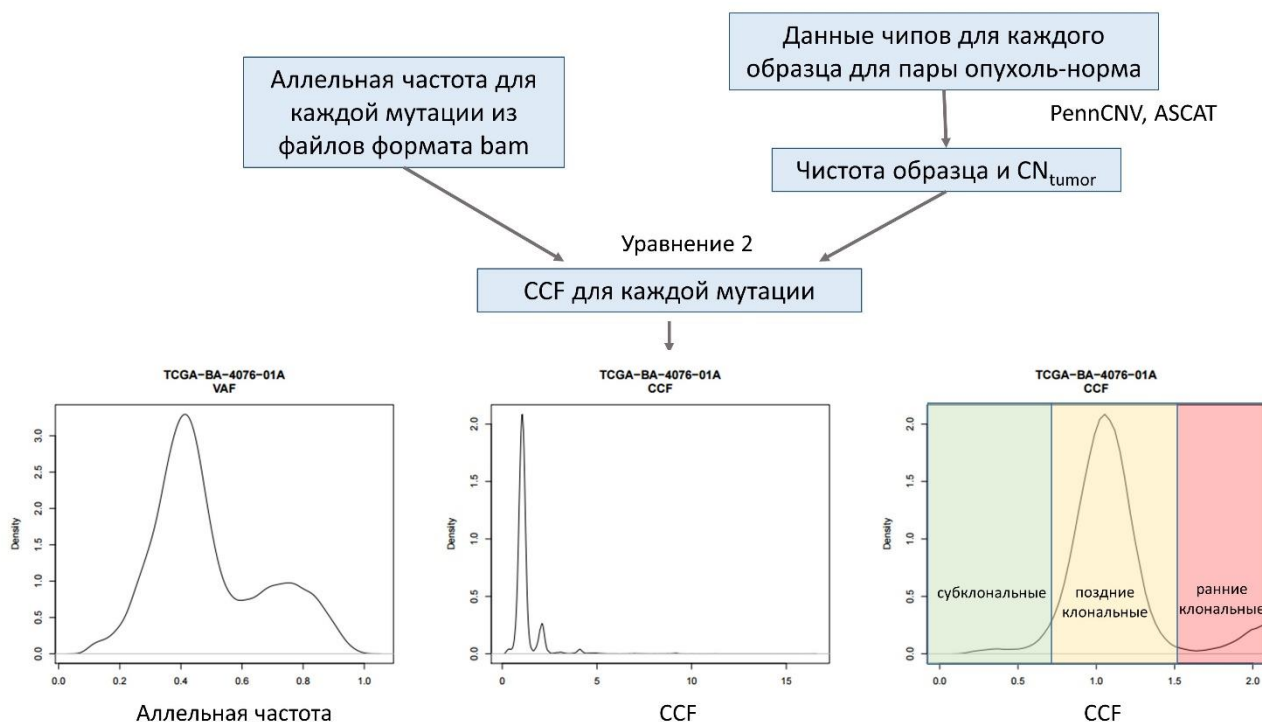


Рисунок 5.2. Схема оценки CCF для каждой мутации из имеющихся данных и классификация мутаций на три категории: «ранние клональные», «поздние клональные», «субклональные».

5.1.3 Подписи в клональных и субклональных мутациях

Для каждого образца мы разделили все мутации на бины согласно их CCF (рисунок 5.2). Число бинов определялось индивидуально для каждого образца так, чтобы каждый бин содержал 1000 мутаций. Для каждого образца в каждом бине мы определили доли всех мутационных процессов (доли соответствующих мутационных подписей) с использованием пакета «deconstructSigs». Затем мы оценили изменение доли мутаций всех доминантных подписей от бинов с высокими CCF к бином с низкими CCF. Поскольку не все мутационные подписи являются ортогональными векторами, изменения в доле некоторых мутационных подписей могут быть вызваны изменением доли других похожих подписей. Поскольку у нас в каждом бине всего 1000 мутаций, это может быть причиной ошибок.

Мы исключили из анализа образцы, которые имели менее 50 субклональных мутаций. Затем мы оценили долю подписей отдельно для ранних и поздних клональных и субклональных мутаций. Для каждого образца мы исключили подписи, имеющие вес меньше 0.2 во всех категориях мутаций. Для каждого типа рака мы выбрали подписи, которые присутствовали (с

долей > 0.2) более, чем в 1 образце, и изучили, изменяется ли доля таких подписей с клональностью и схожи ли эти тенденции между образцами. Для каждой подписи внутри каждого образца мы построили линейную зависимость доли мутаций заданной подписи от CCF отдельно для ранних и поздних мутаций. К ранним мы отнесли ранние клональные мутации, к поздним – поздние клональные и субклональные мутации. В каждом случае мы оценили значимость линейной регрессии. Значимой считалась регрессия с $p < 0.05$. Почти во всех исследуемых образцах регрессия для ранних клональных мутаций была не значимой, и доля заданной подписи оставалась постоянной и не зависела от CCF. Это можно объяснить тем, что значения $CCF > 1$ не отражают долю клеток, несущих мутацию; все значения $CCF > 1$ говорят лишь о том, что мутация произошла до изменения копийности. Мы считали, что доля подписи растет (падает) с уменьшением CCF, если мы наблюдали значимый рост (падение) в зависимости доли подписи от CCF для группы поздних мутаций.

5.1.4 Численное моделирование изменения интенсивности мутационных подписей

Чтобы оценить нашу способность улавливать изменения интенсивностей подписей, мы использовали численное моделирование. Мы симулировали 1000 мутаций в каждой временной точке, комбинируя подписи 1, 2, 3, 5, 8 и 13 (https://cancer.sanger.ac.uk/cosmic/signatures_v2). Подписи мы выбрали произвольно, включив в выборку как подписи с очень характерным мутационным спектром (например, подписи 1, 2 и 13), так и с более гладкими спектрами (например, подписи 3, 5 и 8). Мы изменяли активность одной подписи от 0.1 до 0.6 с шагом 0.1. Доли мутаций, приносимых другими подписями, были постоянными. Затем мы оценили вклад подписей с использованием пакета «deconstructSigs», как и для реальных данных. Мы использовали значения R-квадрат как меру оценки качества нашего предсказания. Подписи 2 и 13 были предсказаны очень хорошо, что может быть объяснено их очень специфическим мутационным спектром. Подписи 1, 3 и 8 были определены достаточно хорошо; эффективность предсказания варьировалась в зависимости от доли подписи. Подпись 5 предсказывалась очень плохо. Это говорит о том, что эффективность нашего метода зависит от особенностей мутационного спектра изучаемой подписи.

5.2 Результаты и обсуждение

5.2.1 Паттерны изменения активности основных мутационных процессов при развитии опухоли

Мы проанализировали 825 раковых геномов с высоким покрытием из базы данных TCGA. Мы использовали CCF для оценки времени появления мутаций. Исследование полных геномов, а не экзотов, в отличие от предыдущих исследований, дало нам возможность посмотреть на изменения мутационных паттернов внутри индивидуальных раковых образцов. Мы оценили

интенсивность мутационных процессов с известными мутационными подписями, измерив доли мутаций, привнесенных этими подписями среди клональных и среди субклональных мутаций. В качестве статистической оценки для каждой опухоли мы построили зависимость доли мутаций, привнесённых данной подписью, среди всех мутаций от CCF (рисунок 5.3). Интересный вывод состоит в том, что в исследованных образцах изменение доли подписи происходит плавно и непрерывно с изменением CCF; этого нельзя было утверждать ранее, без исследования индивидуальных раковых образцов.

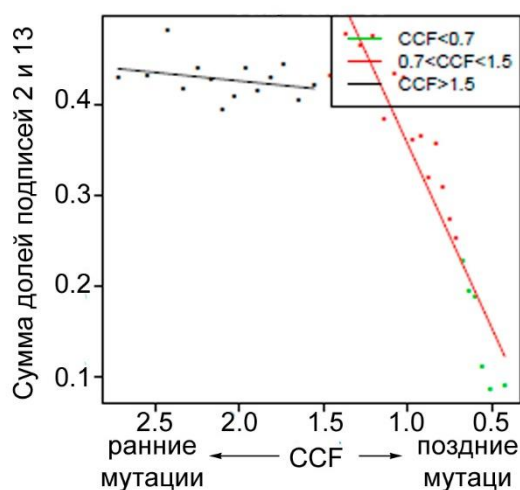


Рисунок 5.3. Пример зависимости доли подписи белка АРОВЕС (сумма долей подписей 2 и 13) от CCF для одного образца аденокарциномы легкого (идентификатор базы данных TCGA TCGA-49-6742). Черные точки соответствуют ранним клональным мутациям, красные – поздним клональным, зеленые – субклональным мутациям. Черная линия соответствует линейной регрессии доли подписей от CCF для ранних клональных мутаций ($p = 0.289$). Красная линия соответствует линейной регрессии доли подписей от CCF для поздних клональных и субклональных мутаций вместе ($p = 2 \times 10^{-9}$).

Для каждой подписи мы сгруппировали образцы по типам рака и выявили преимущественное направление изменения интенсивности каждой подписи внутри каждого типа рака, где она присутствует, подсчитав число образцов, в которых доля подписи растет или падает в поздних мутациях по сравнению с ранними (рисунок 5.4).

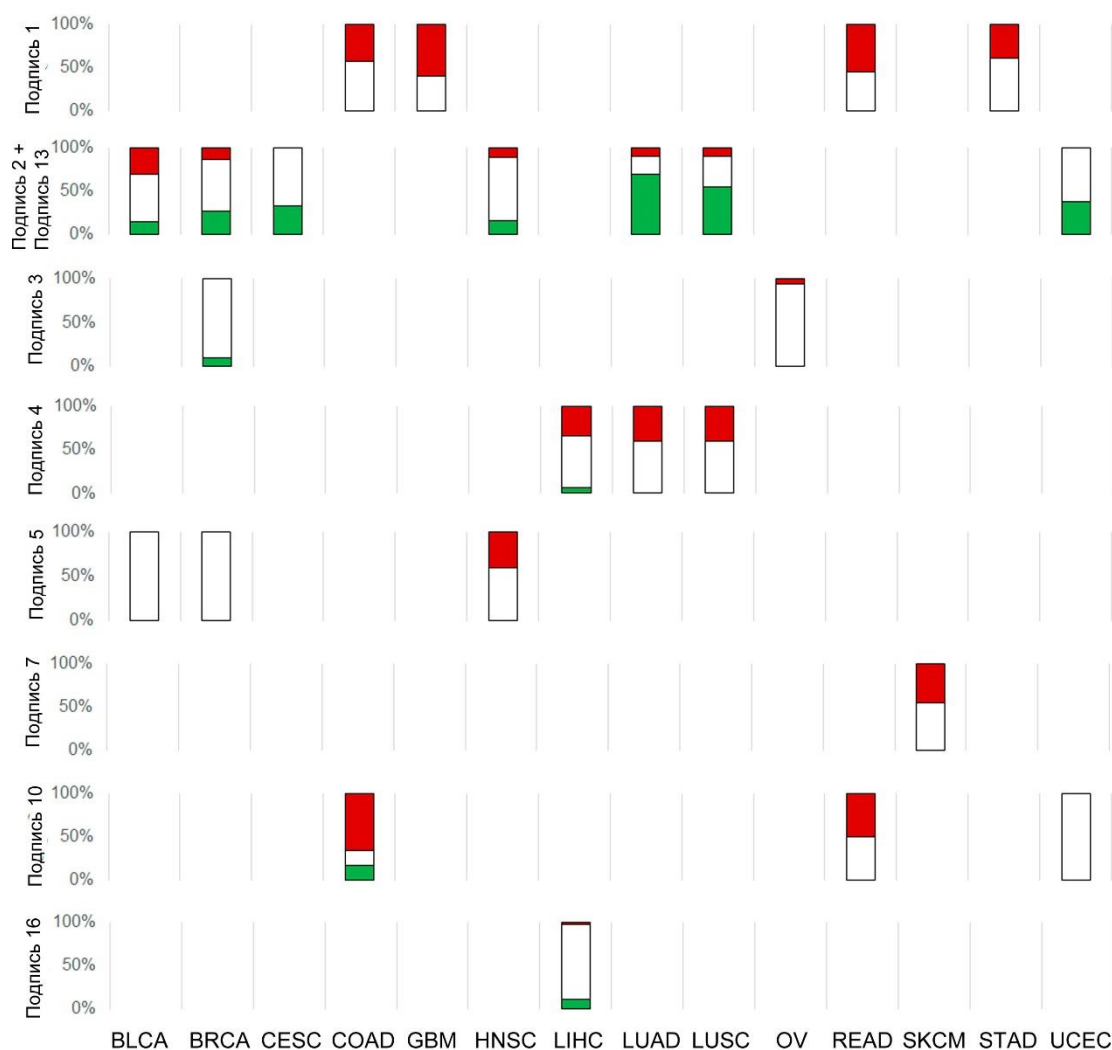


Рисунок 5.4. Изменение мутационных процессов в ходе развития раковых опухолей. По оси x – типы рака. BLCA – рак мочевого пузыря, BRCA – рак груди, CESC – плоскоклеточная карцинома шейки матки, COAD – аденокарцинома толстой кишки, GBM – глиобластома, HNSC – рак головы и шеи, LIHC – рак печени, LUAD – аденокарцинома легкого, LUSC – плоскоклеточный рак легкого, OV – рак яичников, READ – рак прямой кишки, SKCM – меланома, STAD – рак желудка, UCEC – карцинома эндометрия. По оси y – доля образцов со значимым возрастанием (зеленый) или убыванием (красный) доли мутаций, привнесенных данной мутационной подписью, в субклональных мутациях по сравнению с клональными. Белым обозначены образцы без значимой зависимости доли мутаций от CCF

Мы обнаружили, что вклад подписи УФ (подпись 7) и табачного дыма (подпись 4) постепенно падает в ходе развития опухоли (рисунок 5.4), что согласуется с предыдущими исследованиями [7; 126; 127]. Этот результат, по-видимому, может иметь два объяснения одновременно. Во-первых, меньшая подверженность опухоли соответствующим мутагенам на поздних стадиях развития рака уменьшает долю соответствующей мутационной подписи. Большинство образцов меланомы были получены из метастазов, которые меньше подвержены воздействию солнца, а большинство пациентов с раком легких бросают курить после постановки

диагноза. Во-вторых, появление специфических мутагенных процессов на поздних стадиях развития опухоли уменьшает долю мутационных подписей, наблюдаемых на ранних стадиях. На самом деле для рака легких падение доли мутаций, вызванных курением, хорошо коррелирует с ростом доли мутаций, вызванных действием белков семейства APOBEC (рисунок 5.6).

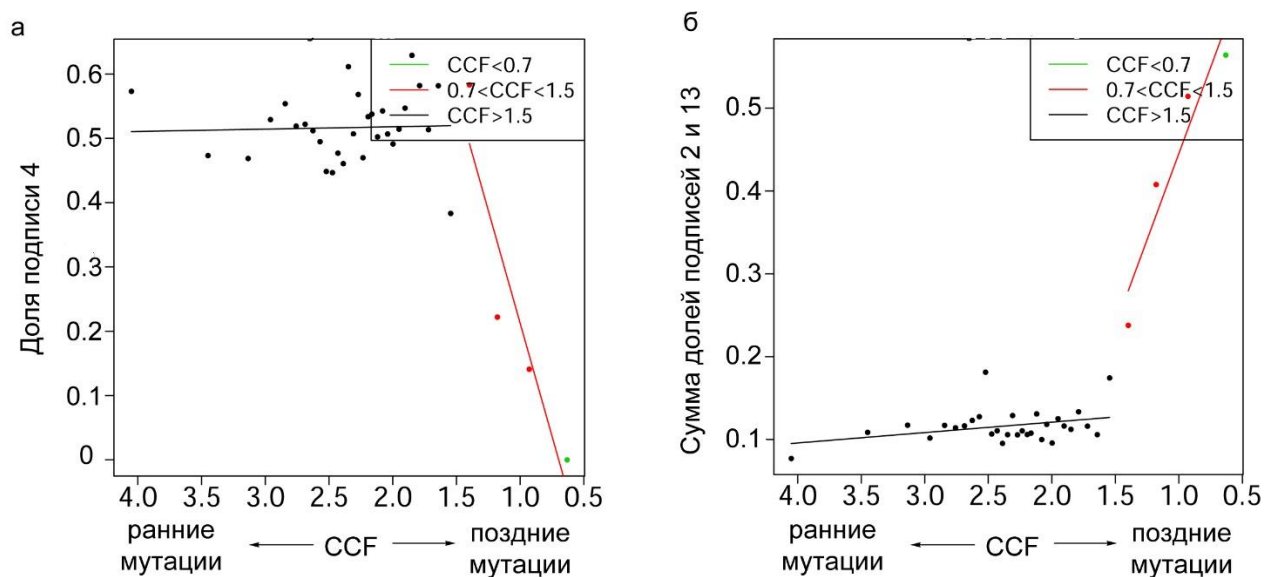


Рисунок 5.5. Изменение долей подписи (а) табачного дыма (подписи 4) и (б) белков APOBEC (подписи 2 и 13) в одном и том же образце рака легких.

Мутации, вызванные белками APOBEC (подпись 2 и подпись 13), показывают неоднозначность среди образцов, которая, по-видимому, специфична для типов рака (рисунок 5.4, 5.7 а, б). В плоскоклеточной карциноме шейки матки (CESC), карциноме эндометрия (UCEC) и раках легких (LUAD и LUSC) мутации, вызванные действием белков семейства APOBEC, имеют тенденцию накапливаться на поздних стадиях развития опухоли, в то время как в раке мочевого пузыря (BLCA) они часто являются ранними событиями. В раке груди (BRCA) и в раке головы и шеи (HNSC) направления изменения активности белка APOBEC разнятся между образцами. Однако внутри опухоли доля мутаций, вызванных активностью белка APOBEC, обычно показывает монотонное поведение.

Подпись спонтанного деаминарования цитозина, ассоциированная с возрастом индивида (подпись 1), была более активно представлена в ранних мутациях (рисунок 5.4). Эта подпись характерна для всех клеток, в том числе и здоровых. Наблюдаемая тенденция скорее всего объясняется тем, что поздние мутации обогащены мутациями, специфичными для раковых процессов, за счет чего уменьшается доля подписи, характерной для здоровых клеток. Подпись полимеразы эpsilon с мутацией в экзонуклеазном домене (подпись 10) также была чаще представлена среди клональных мутаций. Это показывает, что инактивация экзонуклеазного

домена полимеразы эpsilon происходит достаточно рано в ходе развития опухоли, а поздние мутации обогащены дополнительными раковыми процессами, возникшими при развитии опухоли. Например, падение доли подписи 10 хорошо коррелирует с ростом доли подписи 28, происхождение которой неизвестно. Для остальных подписей мы не наблюдали стабильных трендов. Мы предположили, что возможность наблюдать изменения в доле той или иной подписи может зависеть от самой подписи, а именно – от ее мутационного спектра. Изучение спектров разных подписей сразу дает понять, что некоторые подписи имеют очень характерный спектр, то есть приносят очень конкретный тип мутаций с определенным нуклеотидным контекстом (например, подпись 1, представляющая собой мутации C→T в контексте CpG). Другие подписи имеют более гладкие мутационные спектры, то есть вносят достаточно большое разнообразие мутаций и имеют низкую контекстную специфичность (например, подпись 5). Кажется очевидным, что выделять долю подписи с четким контекстно-зависимым мутационным спектром можно точнее, чем долю подписи с большим разнообразием мутаций. Мы, однако, решили подкрепить это предположение симуляцией (см. раздел 5.1.4). Как и предполагалось, оказалось, что чем более специфична подпись, тем точнее определяется ее доля среди всех мутаций. В частности, R-квадрат для подписей 2 и 13 был близок к 1, а для подписи 5 составляет около 0.7 (рисунок 5.6).

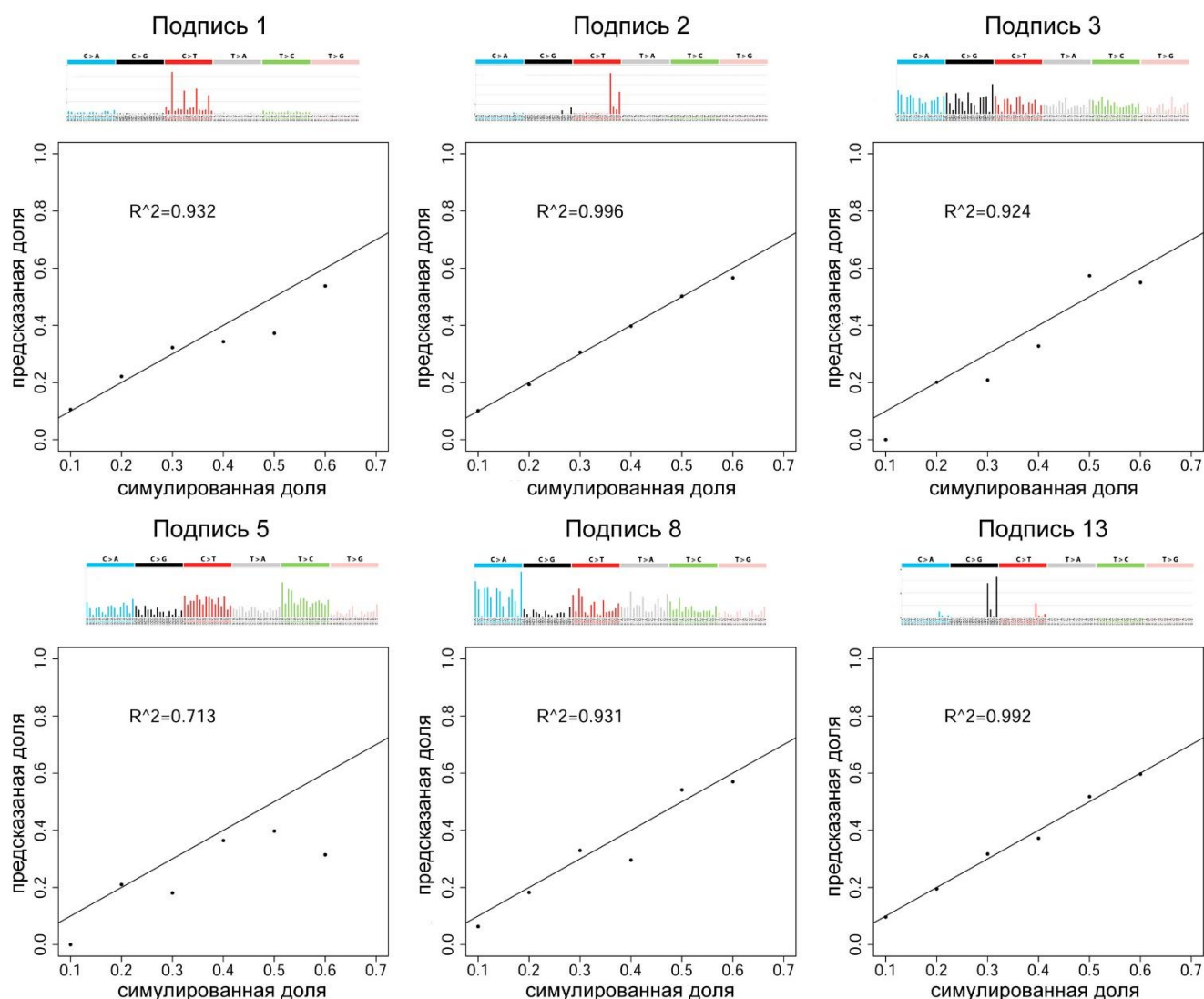


Рисунок 5.6. Симулированная и предсказанная доли разных подписей. Мутационные спектры подписей приведены для наглядности (взяты из https://cancer.sanger.ac.uk/cosmic/signatures_v2 с изменениями). Сплошные линии соответствуют совпадению предсказания и симуляции.

4.2.2 Изменение распределения локальных скоростей мутирования при развитии рака

Наблюдаемые изменения в составе мутационных процессов, действующих в ходе развития опухоли, приводят не только к изменению мутационного спектра, но и к изменению в распределении локальных скоростей мутирования вдоль генома.

Известно, что повреждения, вызванные УФ и табачным дымом, эффективно исправляются TC-NER, что приводит к наличию транскрипционной асимметрии. Мы наблюдали, что доля мутаций, вызванных этими мутагенами выше среди ранних мутаций по сравнению с поздними. Мы подсчитали транскрипционную асимметрию (отношение скоростей мутирования на нетранскрибируемой и транскрибируемой цепях) в ранних и поздних мутациях для мутаций, характерных для УФ и табачного дыма: мутации C→T в меланоме и мутации G→A в раках

легких. Мы увидели, что, как и ожидалось, значения транскрипционной асимметрии для ранних мутаций значительно выше, чем для поздних и в меланоме, и в раках легких (рисунок 5.7б).

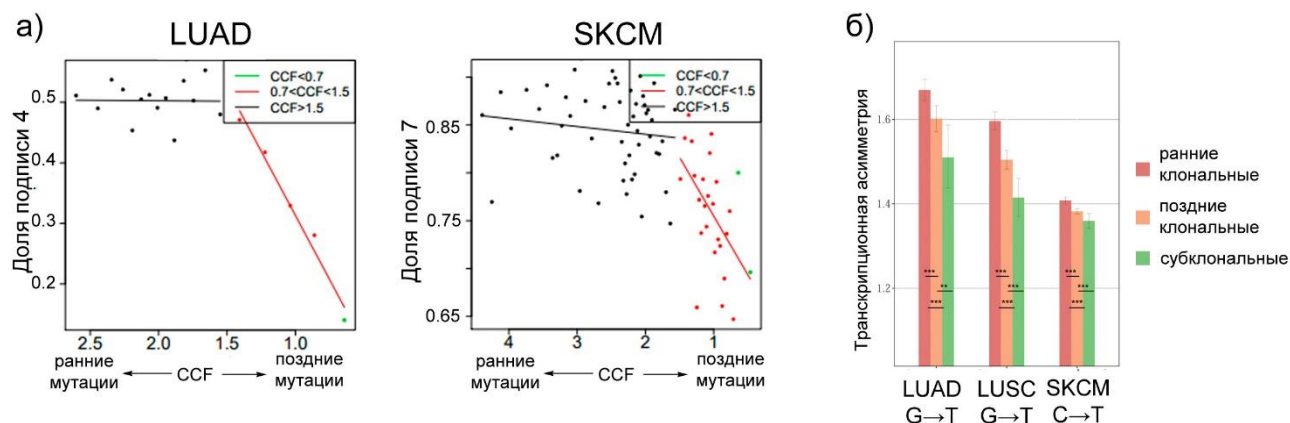


Рисунок 5.7. Изменения в спектре и распределении вдоль генома мутаций в ходе развития раковой опухоли. **а**, примеры изменения доли мутаций табачного дыма и УФ в образцах рака легких и меланомы соответственно; **б**, транскрипционная асимметрия (отношение скоростей мутирования на нетранскрибируемой и транскрибируемой цепях ДНК) ранних и поздних клональных и субклональных мутаций в раках легких (мутация G→T в LUAD и LUSC) и меланоме (мутация C→T).

Мутации, вызванные белками семейства АРОВЕС, практически не зависят от времени репликации, в отличие от мутаций, вызванных большинством других мутационных процессов [60]. Мутации, вызванные АРОВЕС, преобладают среди ранних или поздних мутаций в зависимости от ракового образца. Это предоставляет уникальную возможность показать, что действие одного и того же мутационного процесса может изменять распределение локальных скоростей мутирования по-разному в зависимости от типа опухоли. Мы разделили все образцы со значимым изменением доли подписи АРОВЕС на 2 группы в зависимости от направления изменения этой доли и подсчитали отношение скоростей мутирования в участках ранней и поздней репликации в обеих группах. В наших данных, в образцах рака с большой долей мутаций, вызванных АРОВЕС, среди ранних мутаций ранние мутации слабее зависят от времени репликации, чем поздние мутации (рисунок 5.8г). Ровно противоположный паттерн наблюдался для опухолей, где мутации, вызванные АРОВЕС, преобладают среди поздних мутаций: там поздние мутации слабее зависят от времени репликации (рисунок 5.8в). Мы изучили влияние времени репликации как на все мутации вместе, так и отдельно на мутации, специфичные и не специфичные для АРОВЕС. Интересным образом оказалось, что активность белков семейства АРОВЕС изменяет распределение вдоль генома не только мутаций специфичных для них (TrCpW→K), но и каким-то образом также влияет на распределение остальных мутаций (VpCpW→K). Мы также оценили изменение репликационной асимметрии в ходе развития опухолей с подписью белка АРОВЕС, поскольку известно, что активность этого белка повышена

на отстающей цепи ДНК [78; 128]. Мы наблюдали, что при увеличении активности АРОВЕС среди поздних мутаций ожидаемо увеличивается и репликационная асимметрия (рисунок 5.8в), однако результаты оказались статистически незначимыми.

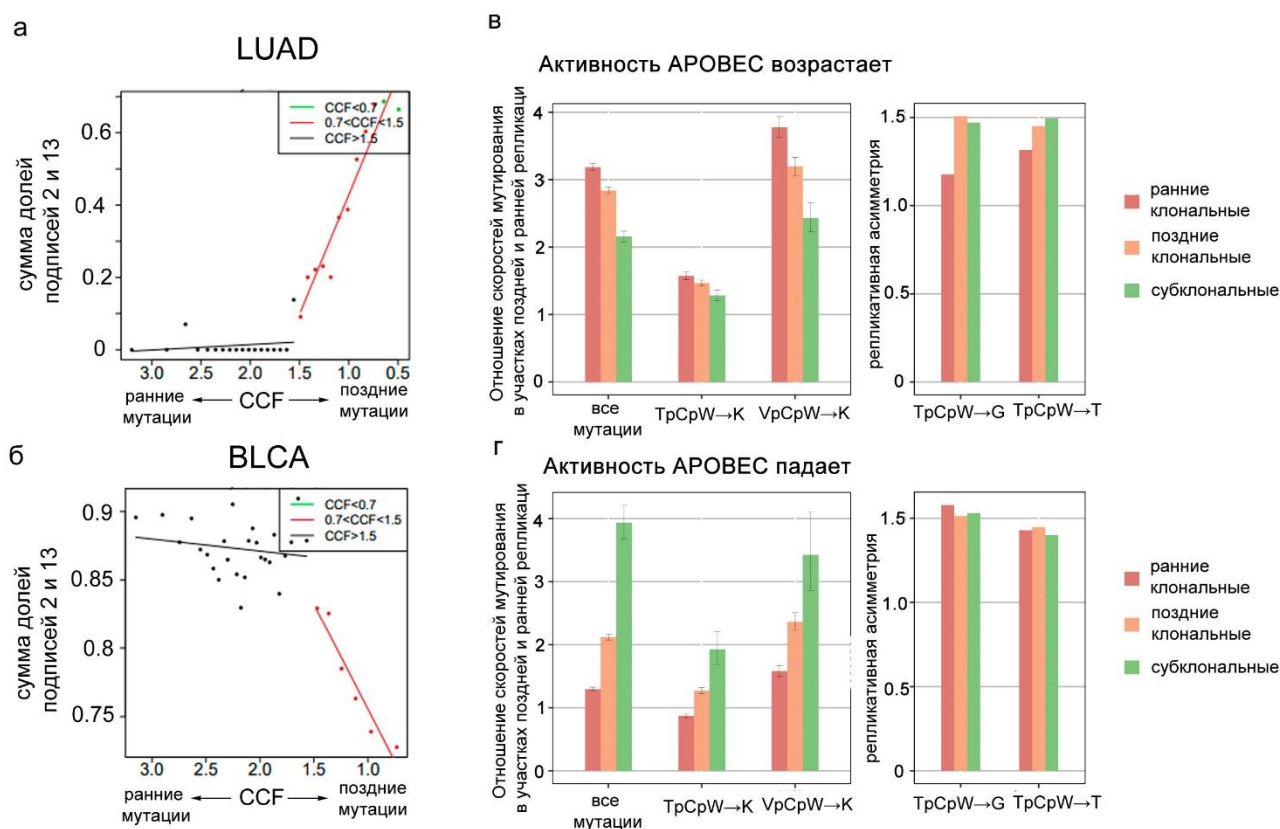


Рисунок 5.8. Изменения в спектре и распределении вдоль генома мутаций в ходе развития раковой опухоли. **а-б**, примеры изменения доли мутаций, вызванных АРОВЕС, в образцах рака легких (**а**) и рака мочевого пузыря (**б**); **в-г**, отношение скоростей мутирования в участках поздней и ранней репликации и репликационная асимметрия (отношение скоростей мутирования на отстающей и лидирующей цепях) ранних и поздних клональных и субклональных мутаций в раке легких и раке мочевого пузыря соответственно. W = A/T, V = A/C/G, K = G/T.

Таким образом, действие разных мутационных процессов в ходе развития раковой опухоли приводит не только к тому, что изменяется характерный мутационный спектр опухоли в разные отрезки времени, но и к тому, что изменяется распределение мутаций вдоль генома и между цепями ДНК. Более того, изменение распределения мутаций по геному не всегда затрагивает только мутации, характерные для определенного процесса; иногда может изменяться распределение и других мутаций. Подробное изучение изменения локальных скоростей мутирования вдоль генома в ходе развития рака важно и с прикладной точки зрения, поскольку помогает правильно оценивать избыточность мутаций при поиске субклональных драйверов.

Основные результаты и выводы

1. Полимераза дельта вызывает большее число неправильно спаренных нуклеотидов при репликации отстающей цепи ДНК, чем полимеразы эпсилон при репликации лидирующей цепи. Это приводит к асимметрии в скоростях мутирования между лидирующей и отстающей цепями ДНК. Система репарации неправильно спаренных нуклеотидов компенсирует данную асимметрию за счет того, что удаляет примерно в 3 раза больше повреждений на отстающей цепи ДНК.

2. В двух исследованных случаях детских глиобластом ультрамутабельный фенотип опухолей объясняется: 1) накоплением большого числа повреждений на лидирующей цепи ДНК в связи с инактивацией экзонуклеазного домена полимеразы эпсилон соматической мутацией; 2) невозможностью репарации данных ошибок из-за инактивации системы репарации неспаренных нуклеотидов в связи с врожденной биаллельной мутацией в гене *PMS2*. Мутационный спектр и направление репликационной асимметрии в данных образцах совпадает с направлением в других изученных ранее образцах с bMMRD и инактивацией экзонуклеазного домена полимеразы эпсилон.

3. Доля мутаций под положительным отбором в ультрамутабельных глиобластомах как минимум в 2 раза ниже, чем в немутабельных глиобластомах.

4. Доля мутаций, характерных для УФ, в базальноклеточной карциноме выше, чем в меланоме. Эффективность репарации таких мутаций с помощью TC-NER также выше в базальноклеточной карциноме. Мутации C→A в базальноклеточной карциноме и в меланоме, по-видимому, имеют разную природу: в базальноклеточной карциноме они на самом деле являются мутациями окисления гуанина, в то время как в меланоме представляют собой смесь мутаций окисления гуанина и мутаций цитозина под действием УФ.

5. Доли мутаций, вызванных определённым мутационным процессом, может изменяться в ходе развития раковой опухоли. В исследованных случаях изменения происходят непрерывно и плавно. Направление изменения зависит от мутационного процесса, а также от типа рака. Изменение активности мутационных процессов влияет не только на особенности мутационного спектра, но и на распределение мутаций вдоль генома и между цепями ДНК.

Приложения

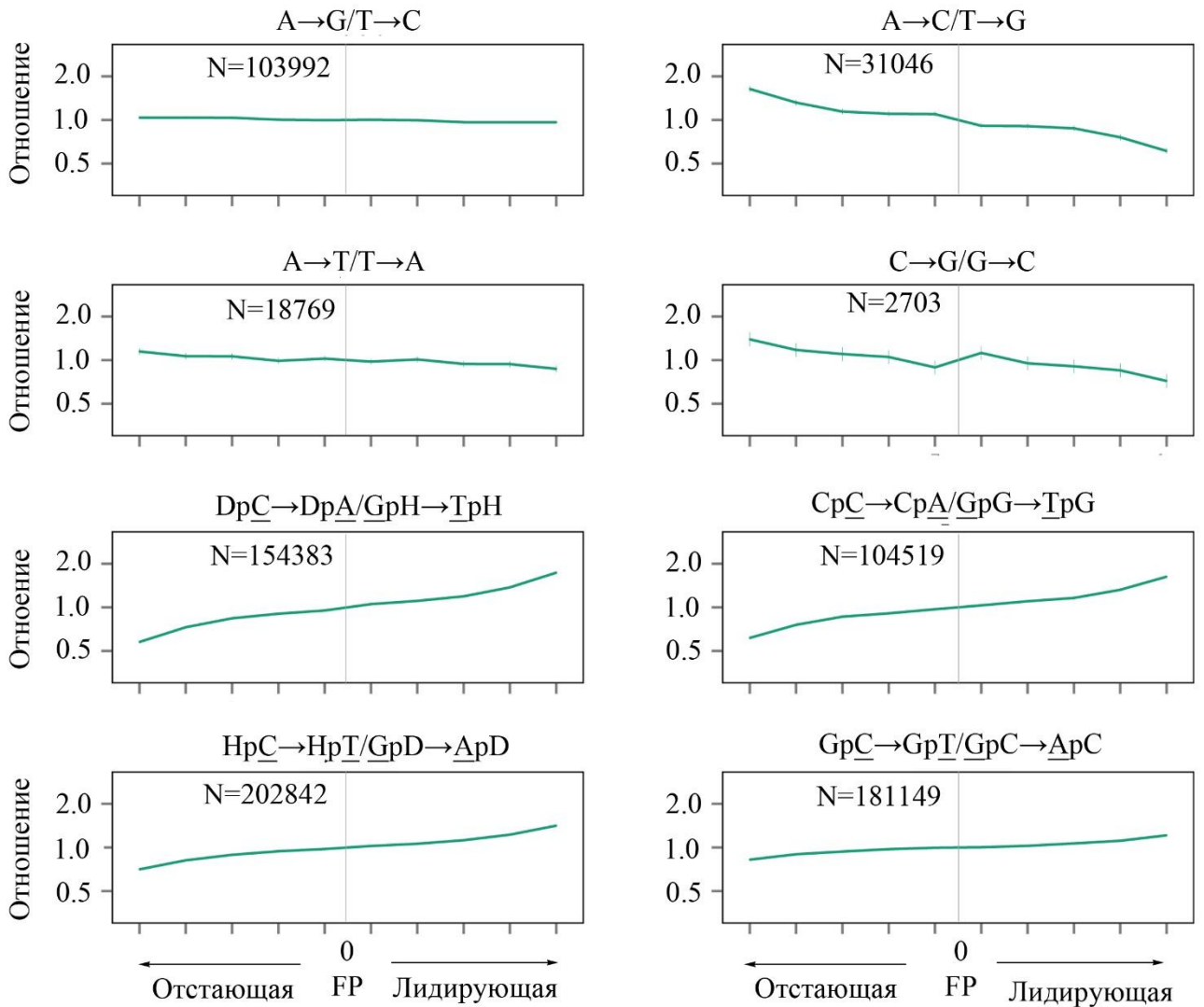


Рисунок П2.1. Репликационная асимметрия в полных геномах карциномы матки с инактивированной MMR и мутацией в экзонуклеазном домене полимеразы эpsilon. Данные по 1 геному из базы данных TCGA. Условные обозначения как на рисунке 2.5.

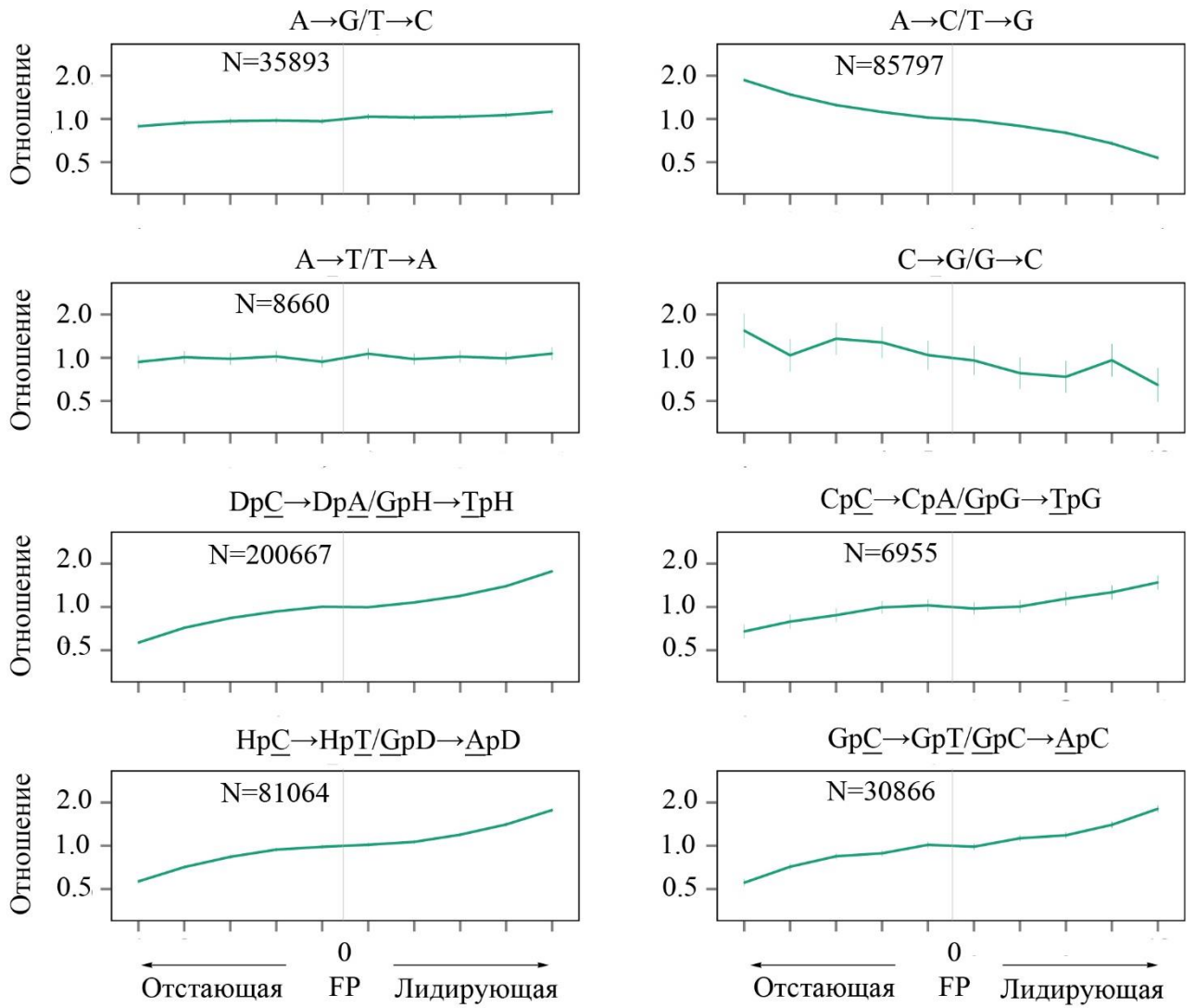


Рисунок П2.2. Репликационная асимметрия в полных геномах аденокарцином толстой кишки с мутацией в экзонуклеазном домене полимеразы эpsilon. Данные для 2 геномов из базы данных TCGA. Условные обозначения как на рисунке 2.5.

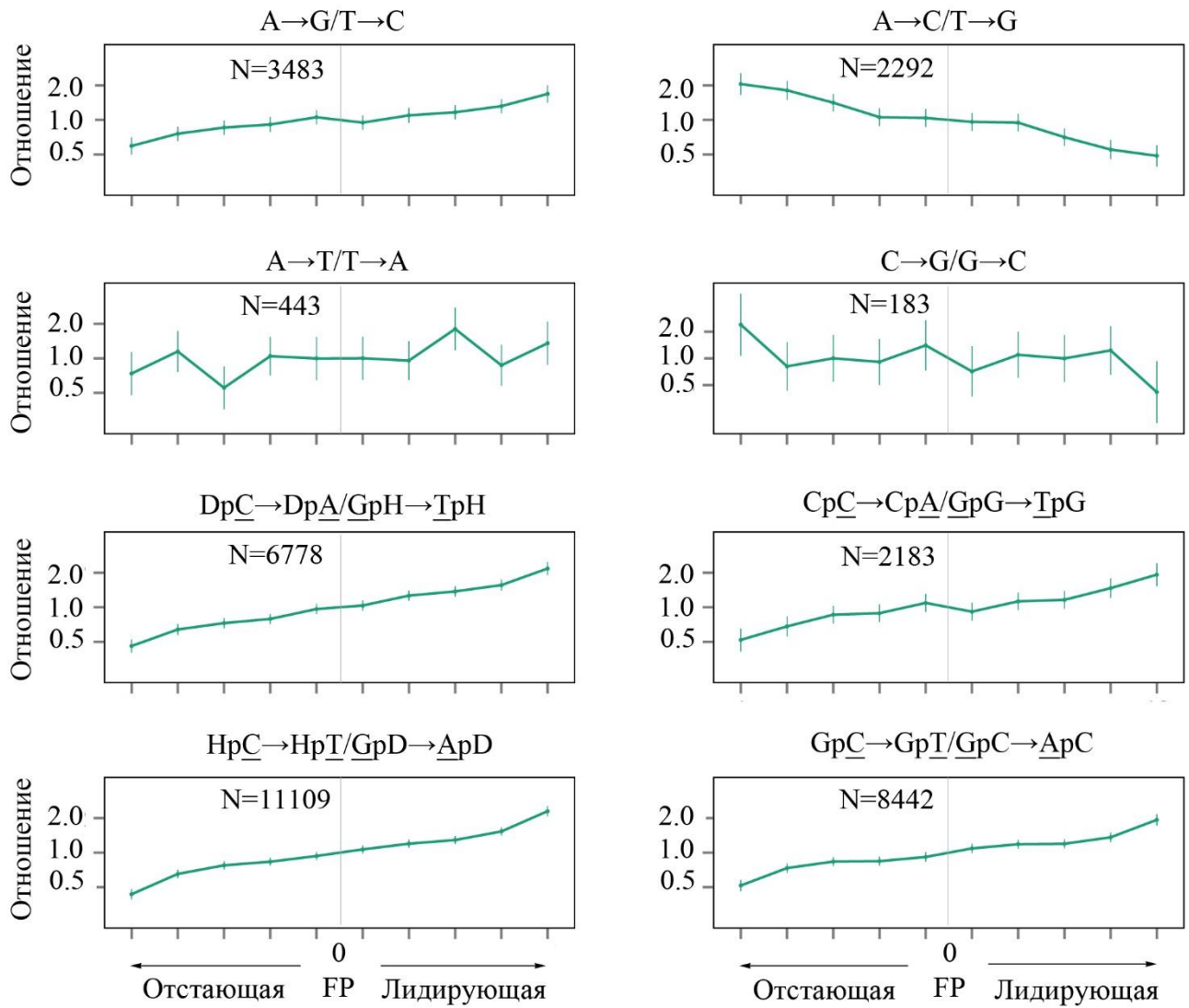


Рисунок П2.3. Репликационная асимметрия в экзонах глиобластом с мутацией в экзонуклеазном домене полимеразы эпсилон и гомозиготной врожденной мутацией в гене *MSH6* (ген системы MMR). Данные для 3 экзонов из [81]. Условные обозначения как на рисунке 2.5.

Таблица П2.1. Отношение скоростей комплементарных мутаций на лидирующей цепи в участках, где наиболее точно определяется направление вилки репликации, для отдельных образцов раков с bMMRD и инактивирующей мутацией в репликативных полимеразах. Используются данные по мутациям для экзомов. Значения в скобках соответствуют 95% доверительному интервалу. Для мутаций C→G было недостаточно данных для оценки асимметрии для отдельных образцов.

Образец	Мутированная полимераза	A→G	A→C	A→T	DpC→DpA	CpC→CpA	HpC→HpT	GpC→GpT
D1119	эпсилон	1.22 (0.86-1.74)	0.59 (0.29-1.20)	0.33 (0.08-1.43)	2.12 (1.57-2.86)	1.68 (1.14-2.48)	2.44 (1.88-3.18)	2.03 (1.62-2.53)
D1121	эпсилон	1.33 (0.86-2.05)	0.62 (0.29-1.34)	0.99 (0.27-3.61)	2.28 (1.63-3.19)	3.74 (2.00-6.99)	2.38 (1.84-3.08)	1.83 (1.46-2.30)
D1144	эпсилон	1.20 (0.82-1.77)	0.82 (0.36-1.86)	4.94 (1.22-20.06)	1.70 (1.38-2.10)	1.90 (1.46-2.48)	2.24 (1.89-2.65)	1.80 (1.61-2.00)
D132	эпсилон	1.20 (0.91-1.57)	0.47 (0.28-0.8)	0.28 (0.07-1.20)	2.29 (1.75-3.00)	2.48 (1.32-4.66)	2.06 (1.75-2.42)	2.11 (1.87-2.38)
D134	эпсилон	1.12 (0.73-1.71)	0.44 (0.21-0.96)	0.66 (0.13-3.30)	1.75 (1.32-2.30)	1.68 (1.14-2.48)	1.83 (1.50-2.24)	1.86 (1.52-2.27)
60001	эпсилон	2.19 (1.69-2.82)	0.51 (0.4-0.66)	2.37 (1.15-4.88)	2.34 (1.90-2.87)	2.20 (1.30-3.70)	2.34 (2.05-2.68)	1.99 (1.62-2.43)
60003	эпсилон	0.99 (0.69-1.42)	0.47 (0.26-0.84)	1.29 (0.64-2.62)	2.34 (1.82-3.02)	1.95 (1.36-2.8)	2.18 (1.75-2.73)	1.83 (1.51-2.21)
60004	эпсилон	1.65 (1.18-2.31)	0.41 (0.24-0.69)	0.58 (0.23-1.42)	1.84 (1.45-2.34)	1.78 (1.24-2.55)	2.28 (1.80-2.88)	2.01 (1.64-2.46)
D1120	дельта	2.03 (1.60-2.58)	1.98 (0.54-7.21)	1.25 (0.69-2.26)	0.44 (0.29-0.66)	0.65 (0.49-0.85)	0.49 (0.33-0.73)	0.89 (0.58-1.36)
D1122	дельта	1.99 (1.70-2.32)	1.5 (1.01-2.21)	0.60 (0.51-0.71)	0.41 (0.32-0.52)	0.47 (0.36-0.62)	0.92 (0.72-1.16)	1.31 (1.07-1.60)

Таблица П2.2. Доля мутационных подписей, ассоциированных с инактивацией системы MMR, в раковых образцах COAD и UCEC с функциональным и нефункциональным путями MMR.

Образец	Тип рака	Статус системы MMR	Подписи			
			6	15	20	26
TCGA-AA-3516	COAD	MSI	0.22	0.13	-	0.07
TCGA-AA-A01R	COAD	MSI	0.4	-	-	0.07
TCGA-D5-6540	COAD	MSI	0.38	0.15	-	-
TCGA-A5-A0G9	UCEC	MSI	0.32	-	0.1	-
TCGA-A5-A0GA	UCEC	MSI	0.32	-	0.22	-
TCGA-AP-A054	UCEC	MSI	-	-	0.07	0.59
TCGA-AP-A0LD	UCEC	MSI	0.18	-	0.26	0.12
TCGA-AP-A0LE	UCEC	MSI	0.3	-	0.26	-
TCGA-AX-A05S	UCEC	MSI	0.45	-	0.1	-
TCGA-A5-A0GG	UCEC	MSI	0.27	-	0.28	-
TCGA-A6-2681	COAD	MSS	-	-	-	-
TCGA-A6-2683	COAD	MSS	-	-	-	-
TCGA-AA-3666	COAD	MSS	-	-	-	-
TCGA-AA-3956	COAD	MSS	-	-	-	-
TCGA-AA-A01S	COAD	MSS	-	-	-	-
TCGA-AA-A02O	COAD	MSS	-	-	-	-
TCGA-AA-A02Y	COAD	MSS	-	-	-	-
TCGA-AA-A03F	COAD	MSS	-	-	-	-
TCGA-A5-A0GE	UCEC	MSS	-	-	-	-
TCGA-AP-A052	UCEC	MSS	-	-	-	-
TCGA-AP-A053	UCEC	MSS	-	-	-	-
TCGA-AP-A05A	UCEC	MSS	-	-	-	-
TCGA-AP-A05D	UCEC	MSS	-	-	-	-
TCGA-AP-A0L8	UCEC	MSS	-	-	-	-
TCGA-AP-A0L9	UCEC	MSS	-	-	-	-
TCGA-AP-A0LH	UCEC	MSS	-	-	-	-
TCGA-AP-A0LI	UCEC	MSS	-	-	-	-
TCGA-AP-A0LL	UCEC	MSS	-	-	-	-
TCGA-AP-A0LO	UCEC	MSS	-	-	0.08	-
TCGA-B5-A0JN	UCEC	MSS	-	-	-	-
TCGA-BK-A0CC	UCEC	MSS	-	-	-	-
TCGA-BS-A0TC	UCEC	MSS	-	-	-	-
TCGA-BS-A0TD	UCEC	MSS	-	-	-	-
TCGA-BS-A0TG	UCEC	MSS	-	-	-	-
TCGA-D1-A16G	UCEC	MSS	-	-	-	-
TCGA-AX-A1CI	UCEC	MSS	-	-	-	-
TCGA-EY-A1GW	UCEC	MSS	-	-	-	-

Таблица П2.3. Отношение скоростей комплементарных мутаций на лидирующей цепи в участках с достоверно определенным направлением вилки репликации в MSI и MSS образцах разных типов рака. Звездочками отмечены недостаточно надежные значения, посчитанные по маленькому числу мутаций (<100 шт.). Значения в скобках соответствуют 95% доверительному интервалу.

Полноэкзомные данные						
Тип рака	COAD		UCEC		STAD	
Тип мутации	MSI	MSS	MSI	MSS	MSI	MSS
A→C/T→G	*1.60 (1.11-2.30)	*1.08 (0.79-1.46)	*0.89 (0.63-1.28)	*0.88 (0.55-1.41)	1.53 (1.28-1.83)	1.21 (1.08-1.36)
A→G/T→C	1.55 (1.33-1.82)	0.90 (0.75-1.09)	1.96 (1.74-2.21)	1.18 (0.89-1.55)	2.01 (1.87-2.16)	1.07 (0.95-1.21)
A→T/T→A	*0.94 (0.66-1.34)	*0.90 (0.67-1.22)	*0.89 (0.65-1.22)	*1.28 (0.82-2.01)	0.64 (0.52-0.77)	1.17 (0.97-1.40)
C→A/G→T	0.74 (0.61-0.88)	0.89 (0.76-1.04)	0.68 (0.58-0.79)	0.90 (0.74-1.09)	0.73 (0.66-0.80)	0.92 (0.81-1.04)
C→T/G→A	1.18 (1.09-1.27)	0.94 (0.87-1.02)	1.25 (1.16-1.34)	0.88 (0.78-0.99)	1.21 (1.17-1.26)	0.84 (0.78-0.90)
C→G/G→C	*0.94 (0.60-1.47)	0.98 (0.78-1.24)	*0.79 (0.52-1.19)	0.75 (0.59-0.96)	1.07 (0.83-1.39)	0.91 (0.78-1.07)
Полногеномные данные						
Тип рака	COAD		UCEC			
Тип мутации	MSI	MSS	MSI	MSS		
A→C/T→G	1.43 (1.35-1.52)	1.12 (1.05-1.20)	1.53 (1.43-1.64)	0.94 (0.80-1.11)		
A→G/T→C	1.54 (1.49-1.59)	1.03 (0.97-1.05)	1.99 (1.94-2.04)	1.03 (0.92-1.15)		
A→T/T→A	0.73 (0.69-0.78)	0.94 (0.87-1.03)	1.06 (1.01-1.12)	1.01 (0.87-1.16)		
C→A/G→T	0.62 (0.60-0.65)	0.95 (0.90-1.00)	0.65 (0.62-0.68)	0.88 (0.80-0.97)		
C→T/G→A	1.24 (1.22-1.27)	0.94 (0.90-0.99)	1.30 (1.27-1.34)	0.87 (0.81-0.94)		
C→G/G→C	0.92 (0.83-1.01)	0.72 (0.65-0.79)	1.24 (1.11-1.38)	0.93 (0.80-1.06)		

Таблица П2.4. Отношение скоростей комплементарных мутаций на лидирующей цепи в участках с достоверно определенным направлением вилки репликации в геномах индивидуальных образцов MSI раков. Значения в скобках соответствуют 95% доверительному интервалу.

Образец	A→G	A→C	A→T	C→G	C→A	CpC→CpA	C→T	GpC→GpT
TCGA-A5-A0G9	1.80 (1.58-2.06)	1.16 (0.86-1.56)	0.93 (0.78-1.12)	1.07 (0.74-1.55)	0.92 (0.76-1.12)	0.55 (0.43-0.69)	0.99 (0.84-1.16)	1.39 (1.20-1.60)
TCGA-A5-A0GA	1.63 (1.45-1.83)	1.15 (0.82-1.62)	0.78 (0.61-0.99)	1.29 (0.89-1.88)	0.79 (0.62-1.02)	0.54 (0.45-0.66)	1.12 (0.96-1.31)	1.68 (1.48-1.92)
TCGA-A5-A0GG	1.91 (1.65-2.20)	1.34 (0.89-2.03)	0.91 (0.66-1.26)	1.43 (0.97-2.12)	0.80 (0.59-1.09)	0.46 (0.35-0.59)	1.02 (0.84-1.23)	1.56 (1.32-1.83)
TCGA-AP-A054	2.34 (2.25-2.44)	1.88 (1.69-2.10)	1.47 (1.35-1.60)	1.35 (1.07-1.70)	0.96 (0.81-1.14)	0.42 (0.35-0.50)	1.11 (0.97-1.25)	1.73 (1.55-1.93)
TCGA-AP-A0LD	2.10 (1.92-2.30)	1.81 (1.44-2.28)	0.89 (0.73-1.08)	1.52 (1.05-2.21)	0.80 (0.64-1.01)	0.46 (0.38-0.56)	1.12 (0.97-1.30)	1.65 (1.45-1.88)
TCGA-AP-A0LE	1.67 (1.53-1.83)	1.26 (0.97-1.63)	0.87 (0.72-1.06)	1.11 (0.80-1.55)	0.82 (0.67-1.00)	0.40 (0.34-0.47)	1.18 (1.04-1.35)	1.73 (1.56-1.92)
TCGA-AX-A05S	1.73 (1.54-1.95)	1.41 (1.09-1.83)	0.88 (0.72-1.07)	1.04 (0.74-1.46)	0.89 (0.69-1.14)	0.61 (0.50-0.75)	1.03 (0.90-1.17)	1.57 (1.40-1.76)
TCGA-AA-3516	1.36 (1.29-1.42)	1.35 (1.24-1.46)	0.73 (0.67-0.81)	0.82 (0.72-0.94)	0.79 (0.72-0.88)	0.50 (0.45-0.55)	0.90 (0.84-0.97)	1.32 (1.25-1.39)
TCGA-AA-A01R	1.70 (1.60-1.80)	1.64 (1.47-1.83)	0.82 (0.74-0.91)	1.05 (0.86-1.28)	0.83 (0.73-0.94)	0.47 (0.42-0.52)	1.07 (1.00-1.15)	1.53 (1.44-1.63)
TCGA-D5-6540	1.71 (1.59-1.84)	1.38 (1.20-1.58)	0.61 (0.54-0.70)	1.07 (0.85-1.35)	0.81 (0.73-0.90)	0.50 (0.45-0.56)	1.00 (0.92-1.09)	1.47 (1.38-1.57)

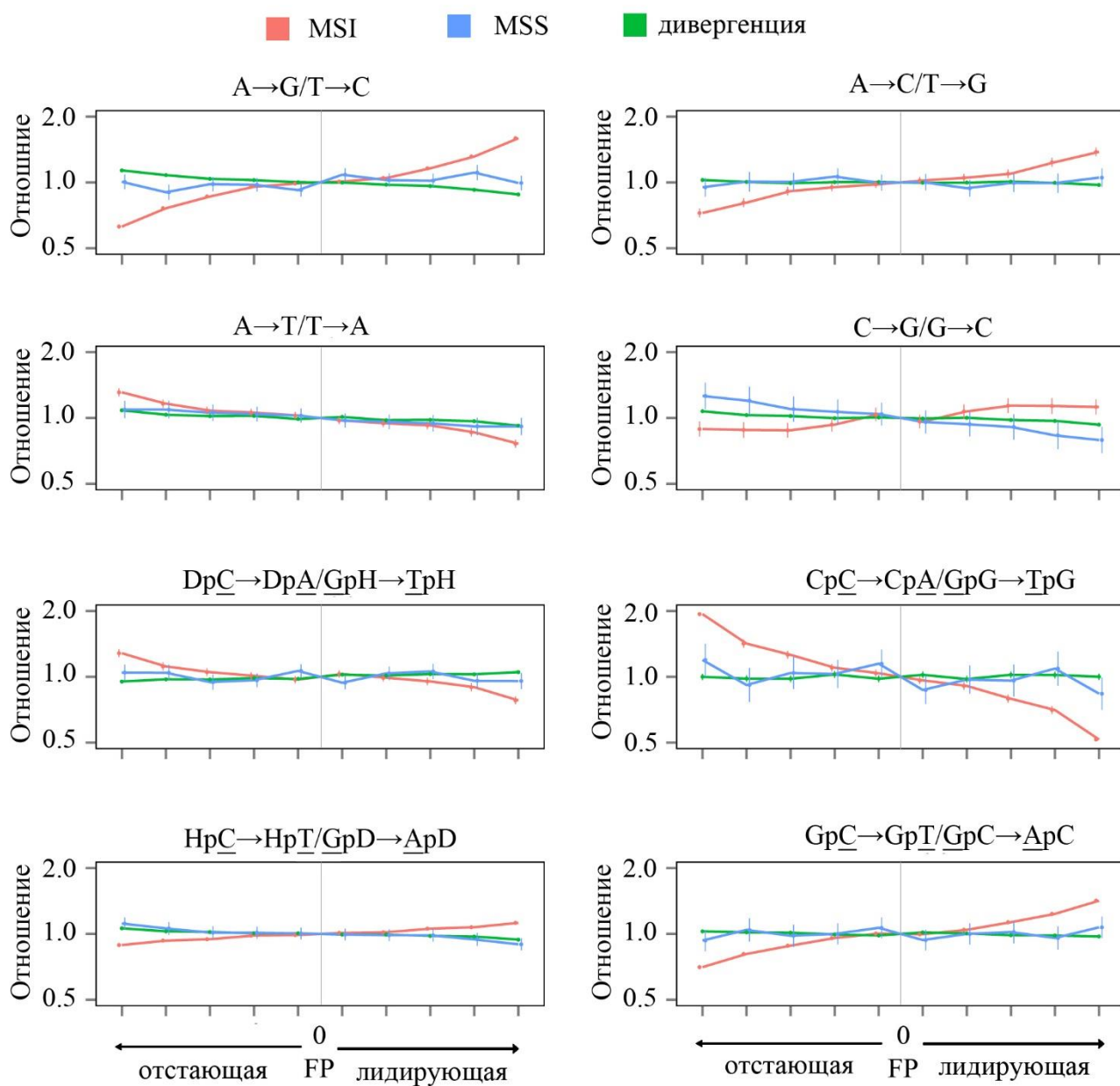


Рисунок П2.4. Репликационная асимметрия в межгенных интервалах. Данные и условные обозначения как на рисунке 2.5.

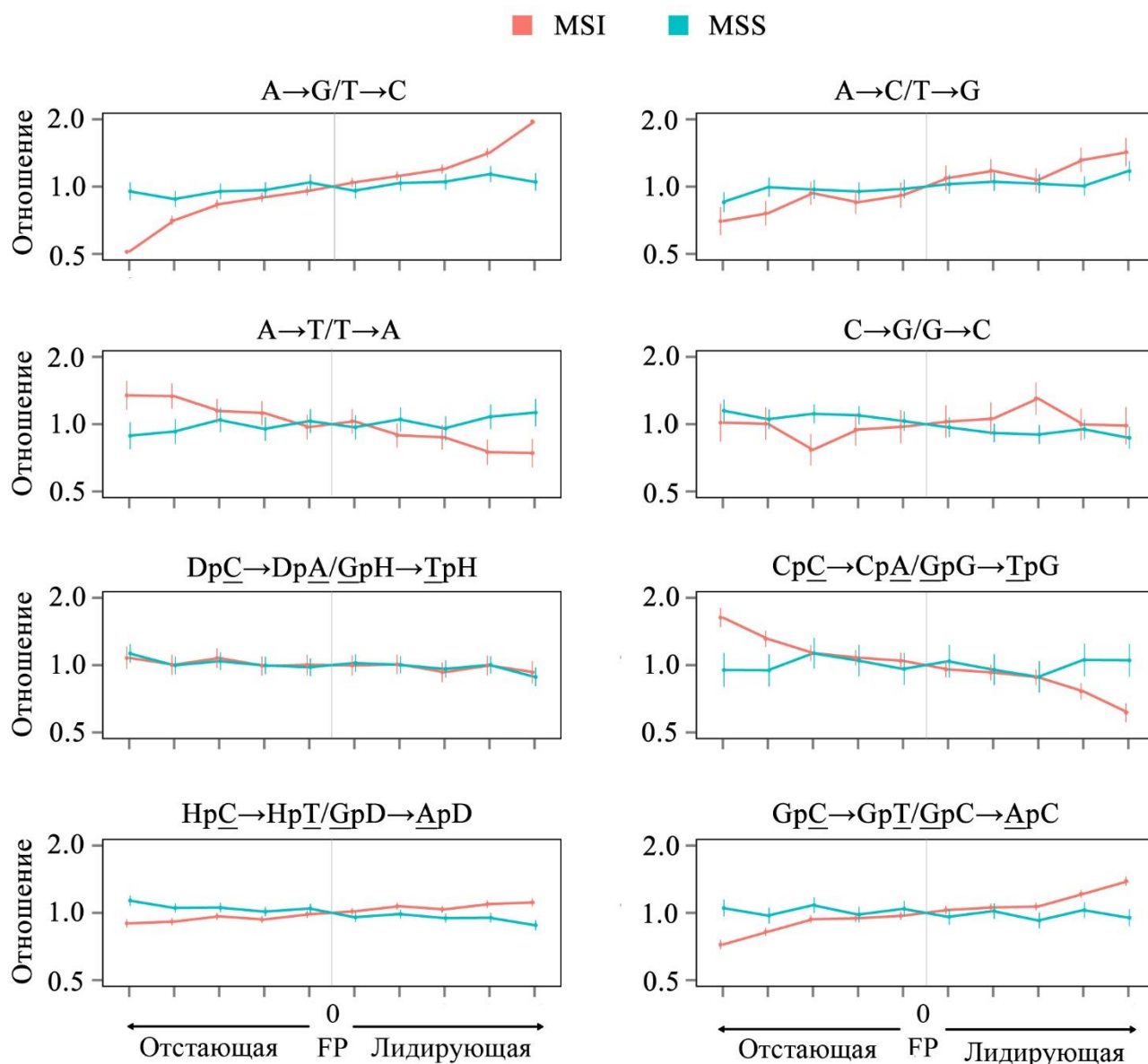


Рисунок П2.5. Репликационная асимметрия в MSI и MSS образцах по экзомным данным. Используются данные для 159 MSI и 782 MSS образцов с диким типом полимераз эпсилон и дельта. Условные обозначения как на рисунке 2.5.

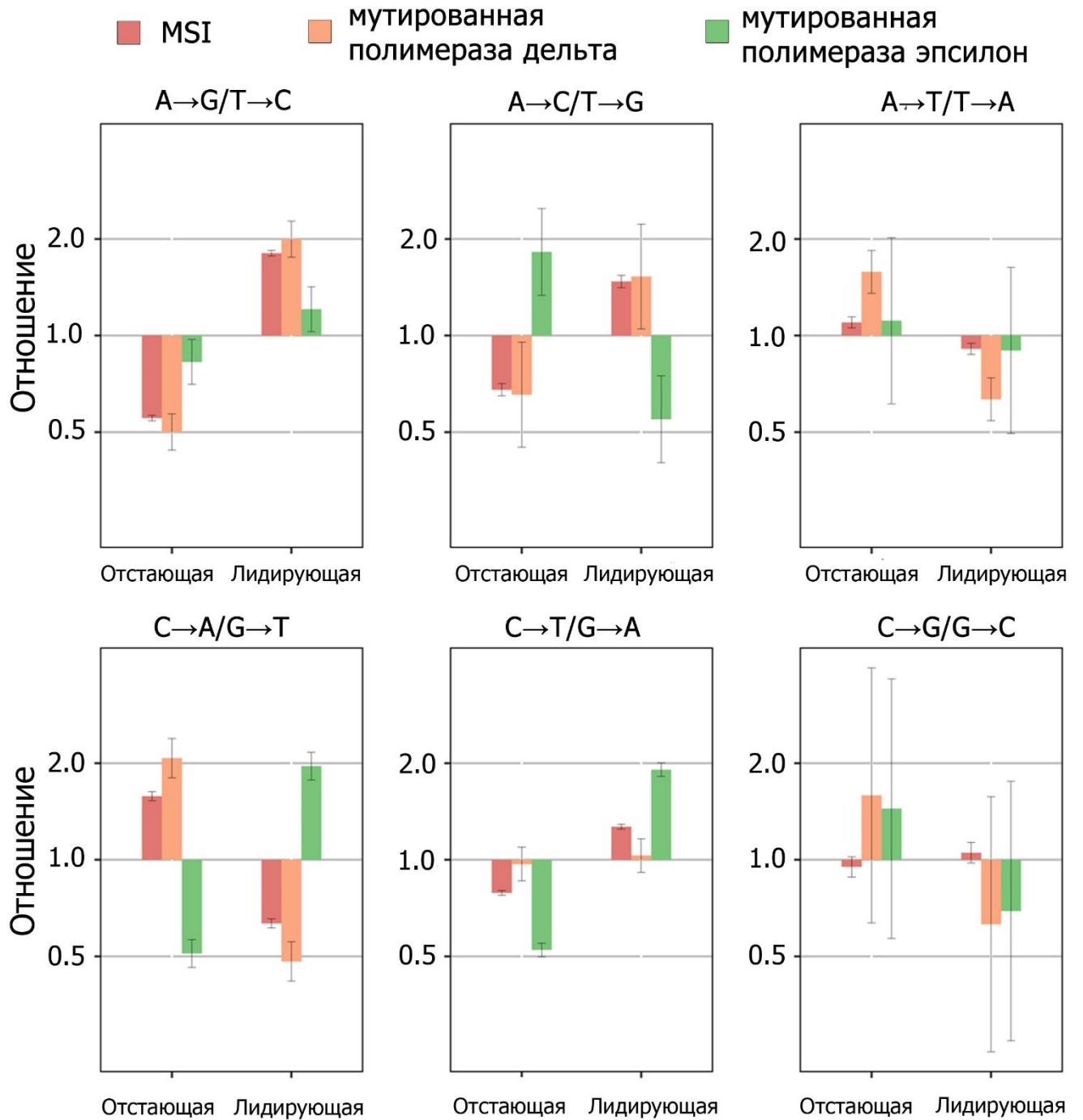


Рисунок П2.6. Репликационная асимметрия с использованием данных секвенирования фрагментов Оказаки. Данные и условные обозначения аналогичны рисунку 2.11.

Таблица П2.5. Корреляция тринуклеотидных мутационных спектров в MSI образцах с полимеразы дикого типа и в bMMRD образцах с мутированными полимеразы.

Тип мутации	Коэффициент корреляции	
	MSI образцы против bMMRD образцов с мутированной полимеразой эpsilon	MSI образцы против bMMRD образцов с мутированной полимеразой delta
C→A	0.59	0.97
A→G	0.19	0.64
A→C	0.44	0.89
C→T	0.82	0.35
C→G	0.48	0.16
A→T	0.85	-0.02

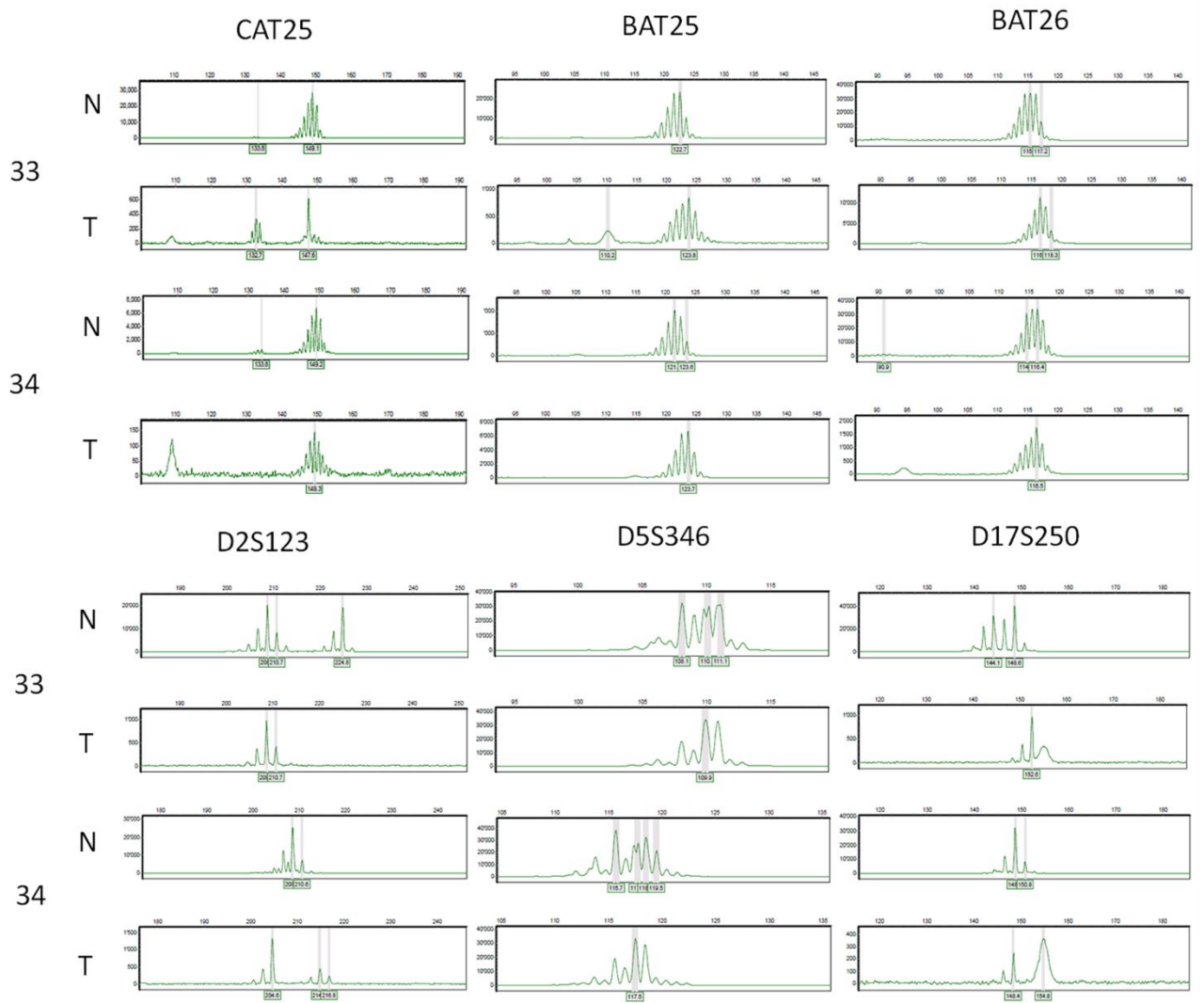


Рисунок П3.1. Профиль геномной микросателлитной нестабильности маркеров для опухоли (Т) и здоровой ткани (N) для двух исследуемых образцов.

Таблица ПЗ.1 Последовательности праймеров для ПЦР длинных фрагментов и вложенной ПЦР 14 экзона гена *PMS2*.

ПЦР длинных фрагментов	CAATAATTATGAGTTCAAGGTCACAGA
	AGACCTCTTTGATAGGAATGTTTGATA
Вложенная ПЦР	TCCAAAAAGCATTTTGTGAGTT
	GAGTTCAAGGTCACAGAGAACG

Таблица ПЗ.2. Список потенциально драйверных мутаций в исследуемых образцах. Хр – хромосома, Реф – референтный аллель, Алт – альтернативный аллель.

Имя гена	Хр	Позиция начала	Позиция конца	Тип мутации	Реф	Алт	Пациент	Замена в белке
<i>POLE</i>	12	133250213	133250213	Missense_Mutation	G	T	33	p.P436H
<i>TP53</i>	17	7578389	7578389	Missense_Mutation	G	A	33	p.R181C
<i>TP53</i>	17	7578407	7578407	Missense_Mutation	G	A	33	p.R175C
<i>PIK3CA</i>	3	178916876	178916876	Missense_Mutation	G	A	33	p.R88Q
<i>PIK3CA</i>	3	178952018	178952018	Missense_Mutation	A	G	33	p.T1025A
<i>SETD2</i>	3	47164385	47164385	Nonsense_Mutation	C	A	33	p.E581*
<i>PBRM1</i>	3	52696293	52696293	Splice_Site	C	T	33	
<i>PARP1</i>	1	226567812	226567812	Nonsense_Mutation	G	A	34	p.R452*
<i>POLE</i>	12	133250250	133250250	Missense_Mutation	G	C	34	p.L424V
<i>ARID2</i>	12	46245723	46245723	Nonsense_Mutation	C	T	34	p.R1273*
<i>CREBBP</i>	16	3795286	3795286	Nonsense_Mutation	C	T	34	p.W1302*
<i>TP53</i>	17	7577094	7577094	Missense_Mutation	G	A	34	p.R282W
<i>TP53</i>	17	7578242	7578242	Missense_Mutation	C	T	34	p.V203M
<i>BAP1</i>	3	52443880	52443880	Nonsense_Mutation	C	T	34	p.W5*
<i>APC</i>	5	112174631	112174631	Nonsense_Mutation	C	T	34	p.R1114*
<i>NF1</i>	17	29683524	29683525	Frame_Shift_Ins	-	A	34	p.K2555fs

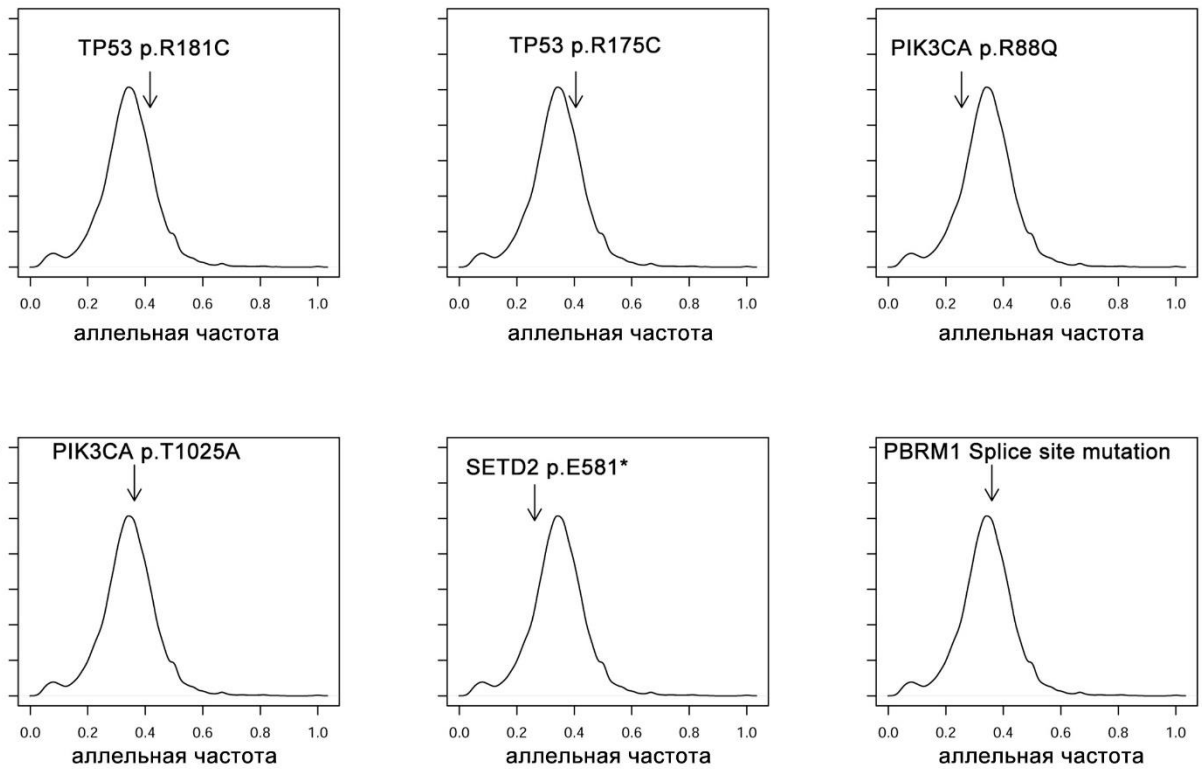


Рисунок ПЗ.2. Распределение аллельных частот потенциальных драйверных мутаций (обозначены стрелками) относительно частот всех мутаций в экзоме образца опухоли пациента 33.

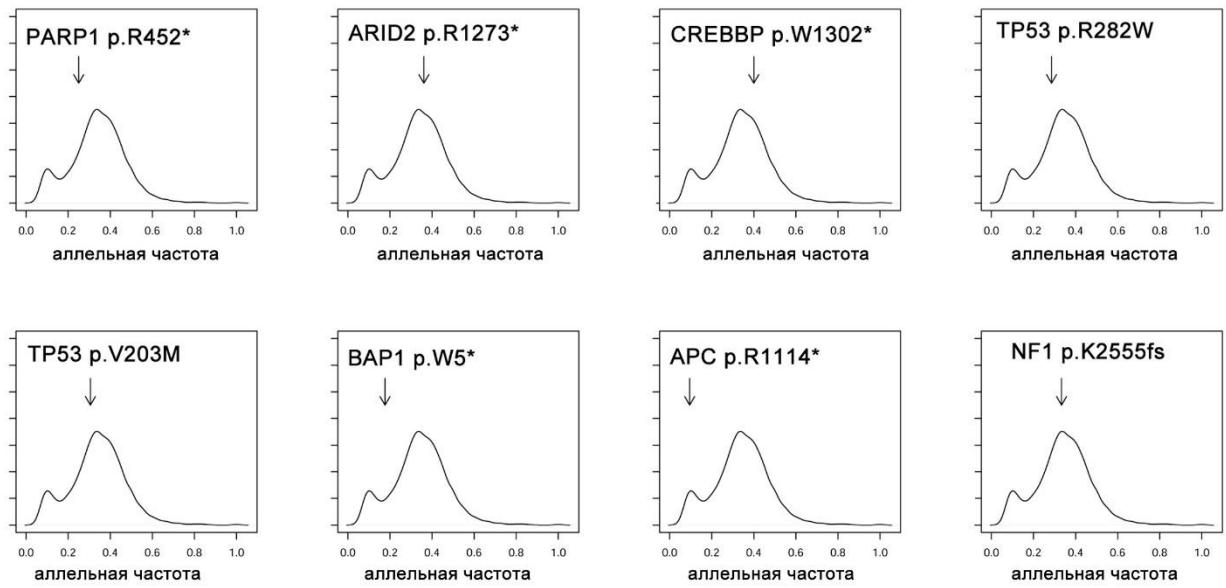


Рисунок ПЗ.3. Распределение аллельных частот потенциальных драйверных мутаций (обозначены стрелками) относительно частот всех мутаций в экзOME образца опухоли пациента 34.

Благодарности

Я благодарна своему научному руководителю Г.А. Базыкину за помощь и поддержку во всех проектах, за огромную помощь в редактировании статей, за создание приятной атмосферы в лаборатории. Отдельно хочу поблагодарить В.Б. Сеплярского за то, что научил меня науке о мутациях, помогал с постановкой задач, за продуктивные обсуждения и моральную поддержку. Хочу сказать большое спасибо С.И. Николаеву за помощь в получении данных, за содержательные обсуждения и помощь в написании статей. Я признательна всем остальным своим соавторам, а также коллегам в лаборатории. Я благодарна за помощь и поддержку при написании текста своим друзьям и семье.

Список литературы

1. Universal Patterns of Selection in Cancer and Somatic Tissues / I. Martincorena [et al.] // *Cell*. – 2017. – Vol. 171. – № 5. – P. 1029-1041.e21.
2. Stratton M.R. The cancer genome / M.R. Stratton, P.J. Campbell, P.A. Futreal // *Nature*. – 2009. – Vol. 458. – № 7239. – P. 719-724.
3. Signatures of mutational processes in human cancer / L.B. Alexandrov [et al.] // *Nature*. – 2013. – Vol. 500. – № 7463. – P. 415-421.
4. Deciphering signatures of mutational processes operative in human cancer / L.B. Alexandrov [et al.] // *Cell Reports*. – 2013. – Vol. 3. – № 1. – P. 246-259.
5. Cell-of-origin chromatin organization shapes the mutational landscape of cancer / P. Polak [et al.] // *Nature*. – 2015. – Vol. 518. – № 7539. – P. 360-364.
6. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair / N.J. Haradhvala [et al.] // *Cell*. – 2016. – Vol. 164. – № 3. – P. 538-549.
7. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution / N. McGranahan [et al.] // *Science Translational Medicine*. – 2015. – Vol. 7. – № 283. – P. 283ra54-283ra54.
8. International network of cancer genome projects / International Cancer Genome Consortium [et al.] // *Nature*. – 2010. – Vol. 464. – № 7291. – P. 993-998.
9. Martincorena I. Somatic mutation in cancer and normal cells / I. Martincorena, P.J. Campbell // *Science*. – 2015. – Vol. 349. – № 6255. – P. 1483-1489.
10. Stratton M.R. Exploring the genomes of cancer cells: progress and promise / M.R. Stratton // *Science (New York, N.Y.)*. – 2011. – Vol. 331. – Exploring the genomes of cancer cells. – № 6024. – P. 1553-1558.
11. Rubin A.F. Mutation patterns in cancer genomes / A.F. Rubin, P. Green // *Proceedings of the National Academy of Sciences of the United States of America*. – 2009. – Vol. 106. – № 51. – P. 21766-21770.
12. Hainaut P. Patterns of p53 G→T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke / P. Hainaut, G.P. Pfeifer // *Carcinogenesis*. – 2001. – Vol. 22. – № 3. – P. 367-374.
13. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers / G.P. Pfeifer [et al.] // *Oncogene*. – 2002. – Vol. 21. – № 48. – P. 7435-7451.
14. Pfeifer G.P. Mutations induced by ultraviolet light / G.P. Pfeifer, Y.-H. You, A. Besaratinia // *Mutation Research*. – 2005. – Vol. 571. – № 1-2. – P. 19-31.

15. The genome as a record of environmental exposure / S. Nik-Zainal [et al.] // *Mutagenesis*. – 2015. – Vol. 30. – № 6. – P. 763-770.
16. Pluskota-Karwatka D. Modifications of nucleosides by endogenous mutagens–DNA adducts arising from cellular processes / D. Pluskota-Karwatka // *Bioorganic Chemistry*. – 2008. – Vol. 36. – № 4. – P. 198-213.
17. Ames B.N. Endogenous mutagens and the causes of aging and cancer / B.N. Ames, L.S. Gold // *Mutation Research*. – 1991. – Vol. 250. – № 1-2. – P. 3-16.
18. Mutational Processes Molding the Genomes of 21 Breast Cancers / S. Nik-Zainal [et al.] // *Cell*. – 2012. – Vol. 149. – № 5. – P. 979-993.
19. Burns M.B. Evidence for APOBEC3B mutagenesis in multiple human cancers / M.B. Burns, N.A. Temiz, R.S. Harris // *Nature Genetics*. – 2013. – Vol. 45. – № 9. – P. 977-983.
20. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers / S.A. Roberts [et al.] // *Nature Genetics*. – 2013. – Vol. 45. – № 9. – P. 970-976.
21. Berwick M. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review / M. Berwick, P. Vineis // *Journal of the National Cancer Institute*. – 2000. – Vol. 92. – Markers of DNA repair and susceptibility to cancer in humans. – № 11. – P. 874-897.
22. Fuss J.O. DNA repair: dynamic defenders against cancer and aging / J.O. Fuss, P.K. Cooper // *PLoS biology*. – 2006. – Vol. 4. – DNA repair. – № 6. – P. e203.
23. Kunkel T.A. Evolving views of DNA replication (in)fidelity / T.A. Kunkel // *Cold Spring Harbor Symposia on Quantitative Biology*. – 2009. – Vol. 74. – P. 91-101.
24. Kunkel T.A. Balancing eukaryotic replication asymmetry with replication fidelity / T.A. Kunkel // *Current Opinion in Chemical Biology*. – 2011. – Vol. 15. – № 5. – P. 620-626.
25. Quantifying the contributions of base selectivity, proofreading and mismatch repair to nuclear DNA replication in *Saccharomyces cerevisiae*. / J.A. St Charles [et al.] // *DNA repair*. – 2015. – Vol. 31. – P. 41-51.
26. OKAZAKI T. Days weaving the lagging strand synthesis of DNA — A personal recollection of the discovery of Okazaki fragments and studies on discontinuous replication mechanism — / T. OKAZAKI // *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*. – 2017. – Vol. 93. – № 5. – P. 322-338.
27. Genome-wide model for the normal eukaryotic DNA replication fork. / A.A. Larrea [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2010. – Vol. 107. – № 41. – P. 17674-17679.
28. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication / J.T.P. Yeeles [et al.] // *Molecular Cell*. – 2017. – Vol. 65. – № 1. – P. 105-116.

29. Yeast DNA polymerase epsilon participates in leading-strand DNA replication / Z.F. Pursell [et al.] // *Science (New York, N.Y.)*. – 2007. – Vol. 317. – № 5834. – P. 127-130.
30. Kunkel T.A. Dividing the workload at a eukaryotic replication fork / T.A. Kunkel, P.M. Burgers // *Trends in Cell Biology*. – 2008. – Vol. 18. – № 11. – P. 521-527.
31. Division of labor at the eukaryotic replication fork. / S.A. Nick McElhinny [et al.] // *Molecular cell*. – 2008. – Vol. 30. – № 2. – P. 137-144.
32. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation / A.R. Clausen [et al.] // *Nature Structural & Molecular Biology*. – 2015. – Vol. 22. – № 3. – P. 185-191.
33. Strand-specific analysis shows protein binding at replication forks and PCNA unloading from lagging strands when forks stall / C. Yu [et al.] // *Molecular Cell*. – 2014. – Vol. 56. – № 4. – P. 551-563.
34. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation / R.E. Georgescu [et al.] // *eLife*. – 2015. – Vol. 4.
35. CMG helicase and DNA polymerase form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication / L.D. Langston [et al.] // *Proceedings of the National Academy of Sciences*. – 2014. – Vol. 111. – № 43. – P. 15390-15395.
36. A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands / R.E. Johnson [et al.] // *Molecular Cell*. – 2015. – Vol. 59. – № 2. – P. 163-175.
37. Burgers P.M.J. Who Is Leading the Replication Fork, Pol ϵ or Pol δ ? / P.M.J. Burgers, D. Gordenin, T.A. Kunkel // *Molecular Cell*. – 2016. – Vol. 61. – № 4. – P. 492-493.
38. Lujan S.A. Eukaryotic genome instability in light of asymmetric DNA replication / S.A. Lujan, J.S. Williams, T.A. Kunkel // *Critical Reviews in Biochemistry and Molecular Biology*. – 2016. – Vol. 51. – № 1. – P. 43-52.
39. Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication / H. Hombauer [et al.] // *Science (New York, N.Y.)*. – 2011. – Vol. 334. – № 6063. – P. 1713-1716.
40. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair / Y. Liao [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2015. – Vol. 112. – № 50. – P. E6898-6906.
41. Differing patterns of genetic instability in mice deficient in the mismatch repair genes Pms2, Mlh1, Msh2, Msh3 and Msh6 / D.C. Hegan [et al.] // *Carcinogenesis*. – 2006. – Vol. 27. – № 12. – P. 2402-2408.
42. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer / The Cancer Genome Atlas Network // *Nature*. – 2012. – Vol. 487. – № 7407. – P. 330-337.

43. Lujan S.A. Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. / S.A. Lujan, A.B. Clark, T.A. Kunkel // *Nucleic acids research*. – 2015. – Vol. 43. – № 8. – P. 4067-4074.
44. Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks / H. Zhao [et al.] // *eLife*. – 2014. – Vol. 3. – P. e02725.
45. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers / S.B. Simpkins [et al.] // *Human Molecular Genetics*. – 1999. – Vol. 8. – № 4. – P. 661-666.
46. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSa / F. Li [et al.] // *Cell*. – 2013. – Vol. 153. – № 3. – P. 590-600.
47. Sancar A. DNA Excision Repair / A. Sancar // *Annual Review of Biochemistry*. – 1996. – Vol. 65. – № 1. – P. 43-81.
48. Reardon J.T. Nucleotide excision repair / J.T. Reardon, A. Sancar // *Progress in Nucleic Acid Research and Molecular Biology*. – 2005. – Vol. 79. – P. 183-235.
49. Wood R.D. Nucleotide excision repair in mammalian cells / R.D. Wood // *The Journal of Biological Chemistry*. – 1997. – Vol. 272. – № 38. – P. 23465-23468.
50. Mellon I. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene / I. Mellon // *Cell*. – 1987. – Vol. 51. – № 2. – P. 241-249.
51. Hanawalt P.C. Transcription-coupled DNA repair: two decades of progress and surprises / P.C. Hanawalt, G. Spivak // *Nature Reviews. Molecular Cell Biology*. – 2008. – Vol. 9. – Transcription-coupled DNA repair. – № 12. – P. 958-970.
52. Mutational heterogeneity in cancer and the search for new cancer-associated genes / M.S. Lawrence [et al.] // *Nature*. – 2013. – Vol. 499. – № 7457. – P. 214-218.
53. A comprehensive catalogue of somatic mutations from a human cancer genome / E.D. Pleasance [et al.] // *Nature*. – 2010. – Vol. 463. – № 7278. – P. 191-196.
54. Donley N. DNA replication timing, genome stability and cancer / N. Donley, M.J. Thayer // *Seminars in Cancer Biology*. – 2013. – Vol. 23. – № 2. – P. 80-89.
55. Differential Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation / A. Koren [et al.] // *The American Journal of Human Genetics*. – 2012. – Vol. 91. – № 6. – P. 1033-1040.
56. Liu L. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. / L. Liu, S. De, F. Michor // *Nature communications*. – 2013. – Vol. 4. – P. 1502.

57. Schuster-Böckler B. Chromatin organization is a major influence on regional mutation rates in human cancer cells / B. Schuster-Böckler, B. Lehner // *Nature*. – 2012. – Vol. 488. – № 7412. – P. 504-507.
58. Sima J. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution / J. Sima, D.M. Gilbert // *Current Opinion in Genetics & Development*. – 2014. – Vol. 25. – Complex correlations. – P. 93-100.
59. Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair / P. Polak [et al.] // *Nature Biotechnology*. – 2014. – Vol. 32. – № 1. – P. 71-75.
60. APOBEC-induced cancer mutations are uniquely enriched in early replicating, gene dense, and active chromatin regions / M.D. Kazanov [et al.] // *Cell reports*. – 2015. – Vol. 13. – № 6. – P. 1103-1109.
61. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. / E. Shinbrot [et al.] // *Genome research*. – 2014. – Vol. 24. – № 11. – P. 1740-1750.
62. Transcription-associated mutational asymmetry in mammalian evolution / NISC Comparative Sequencing Program [et al.] // *Nature Genetics*. – 2003. – Vol. 33. – № 4. – P. 514-517.
63. Lobry J.R. Asymmetric substitution patterns in the two DNA strands of bacteria / J.R. Lobry // *Molecular Biology and Evolution*. – 1996. – Vol. 13. – № 5. – P. 660-665.
64. Mismatch repair balances leading and lagging strand DNA replication fidelity. / S.A. Lujan [et al.] // *PLoS genetics*. – 2012. – Vol. 8. – № 10. – P. e1003016.
65. Polak P. Transcription induces strand-specific mutations at the 5' end of human genes / P. Polak, P.F. Arndt // *Genome Research*. – 2008. – Vol. 18. – № 8. – P. 1216-1223.
66. Polak P. The evolution of transcription-associated biases of mutations across vertebrates / P. Polak, R. Querfurth, P.F. Arndt // *BMC Evolutionary Biology*. – 2010. – Vol. 10. – № 1. – P. 187.
67. Replication-associated strand asymmetries in mammalian genomes: Toward detection of replication origins / M. Touchon [et al.] // *Proceedings of the National Academy of Sciences*. – 2005. – Vol. 102. – Replication-associated strand asymmetries in mammalian genomes. – № 28. – P. 9836-9841.
68. Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template / B.A. Donahue [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 1994. – Vol. 91. – № 18. – P. 8502-8506.
69. Fousteri M. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects / M. Fousteri, L.H. Mullenders // *Cell Research*. – 2008. – Vol. 18. – Transcription-coupled nucleotide excision repair in mammalian cells. – № 1. – P. 73-84.

70. Jiang G. Recruitment of DNA damage checkpoint proteins to damage in transcribed and nontranscribed sequences / G. Jiang, A. Sancar // *Molecular and Cellular Biology*. – 2006. – Vol. 26. – № 1. – P. 39-49.
71. Transcription-coupled repair deficiency and mutations in human mismatch repair genes. / I. Mellon [et al.] // *Science (New York, N.Y.)*. – 1996. – Vol. 272. – № 5261. – P. 557-560.
72. Spivak G. The complex choreography of transcription-coupled repair / G. Spivak, A.K. Ganesan // *DNA repair*. – 2014. – Vol. 19. – P. 64-70.
73. Initial genome sequencing and analysis of multiple myeloma / M.A. Chapman [et al.] // *Nature*. – 2011. – Vol. 471. – № 7339. – P. 467-472.
74. Jinks-Robertson S. Transcription-Associated Mutagenesis / S. Jinks-Robertson, A.S. Bhagwat // *Annual Review of Genetics*. – 2014. – Vol. 48. – № 1. – P. 341-359.
75. McLean M.J. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes / M.J. McLean, K.H. Wolfe, K.M. Devine // *Journal of Molecular Evolution*. – 1998. – Vol. 47. – № 6. – P. 691-696.
76. Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase / A. Koren [et al.] // *PLoS genetics*. – 2010. – Vol. 6. – № 8. – P. e1001068.
77. Pavlov Y.I. Yeast origins establish a strand bias for replicational mutagenesis / Y.I. Pavlov, C.S. Newlon, T.A. Kunkel // *Molecular Cell*. – 2002. – Vol. 10. – № 1. – P. 207-213.
78. APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication / V.B. Seplyarskiy [et al.] // *Genome Research*. – 2016. – Vol. 26. – № 2. – P. 174-182.
79. Helleday T. Mechanisms underlying mutational signatures in human cancers / T. Helleday, S. Eshtad, S. Nik-Zainal // *Nature Reviews Genetics*. – 2014. – Vol. 15. – № 9. – P. 585-598.
80. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermuted cancers. / A. Shlien [et al.] // *Nature genetics*. – 2015. – Vol. 47. – № 3. – P. 257-262.
81. Somatic *POLE* mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis / E.Z. Erson-Omay [et al.] // *Neuro-Oncology*. – 2015. – Vol. 17. – № 10. – P. 1356-1364.
82. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples / K. Cibulskis [et al.] // *Nature Biotechnology*. – 2013. – Vol. 31. – № 3. – P. 213-219.
83. Replication landscape of the human genome / N. Petryk [et al.] // *Nature Communications*. – 2016. – Vol. 7. – P. 10208.

84. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution / R. Rosenthal [et al.] // *Genome Biology*. – 2016. – Vol. 17. – deconstructSigs. – № 1.
85. Korona D.A. The high fidelity and unique error signature of human DNA polymerase epsilon / D.A. Korona, K.G. Lecompte, Z.F. Pursell // *Nucleic Acids Research*. – 2011. – Vol. 39. – № 5. – P. 1763-1773.
86. Henninger E.E. DNA polymerase ϵ and its roles in genome stability / E.E. Henninger, Z.F. Pursell // *IUBMB life*. – 2014. – Vol. 66. – № 5. – P. 339-351.
87. Widespread impact of DNA replication on mutational mechanisms in cancer / Tomkova, M. [et al.]. – 2017.
88. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair / N.J. Haradhvala [et al.] // *Nature Communications*. – 2018. – Vol. 9. – № 1.
89. Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines / A. Baker [et al.] // *PLoS computational biology*. – 2012. – Vol. 8. – № 4. – P. e1002443.
90. The topography of mutational processes in breast cancer genomes / S. Morganella [et al.] // *Nature Communications*. – 2016. – Vol. 7. – P. 11383.
91. Unique error signature of the four-subunit yeast DNA polymerase epsilon. / P.V. Shcherbakova [et al.] // *The Journal of biological chemistry*. – 2003. – Vol. 278. – № 44. – P. 43770-43780.
92. *Saccharomyces cerevisiae* DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. / J.M. Fortune [et al.] // *The Journal of biological chemistry*. – 2005. – Vol. 280. – № 33. – P. 29980-29987.
93. Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. / S.A. Lujan [et al.] // *Genome research*. – 2014. – Vol. 24. – № 11. – P. 1751-1764.
94. Supek F. Differential DNA mismatch repair underlies mutation rate variation across the human genome. / F. Supek, B. Lehner // *Nature*. – 2015. – Vol. 521. – № 7550. – P. 81-84.
95. Replication-associated mutational asymmetry in the human genome / C.-L. Chen [et al.] // *Molecular Biology and Evolution*. – 2011. – Vol. 28. – № 8. – P. 2327-2337.
96. Seplyarskiy V.B. APOBEC3A/B-induced mutagenesis is responsible for 20% of heritable mutations in the TpCpW context / V.B. Seplyarskiy, M.A. Andrianova, G.A. Bazykin // *Genome Research*. – 2017. – Vol. 27. – № 2. – P. 175-184.
97. Replicative DNA Polymerase δ but Not ϵ Proofreads Errors in Cis and in Trans / C.L. Flood [et al.] // *PLOS Genetics*. – 2015. – Vol. 11. – № 3. – P. e1005049.
98. Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta / Y.I. Pavlov [et al.] // *Current biology: CB*. – 2006. – Vol. 16. – № 2. – P. 202-207.

99. Lagging-strand replication shapes the mutational landscape of the genome / M.A.M. Reijns [et al.] // *Nature*. – 2015. – Vol. 518. – № 7540. – P. 502-506.
100. PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. / A. Pluciennik [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2010. – Vol. 107. – № 37. – P. 16066-16071.
101. Immune Checkpoint Inhibition for Hypermutant Glioblastoma Multiforme Resulting From Germline Biallelic Mismatch Repair Deficiency / E. Bouffet [et al.] // *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. – 2016. – Vol. 34. – № 19. – P. 2206-2211.
102. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome / M. De Vos [et al.] // *American Journal of Human Genetics*. – 2004. – Vol. 74. – № 5. – P. 954-964.
103. Extensive gene conversion at the PMS2 DNA mismatch repair locus / B.E. Hayward [et al.] // *Human Mutation*. – 2007. – Vol. 28. – № 5. – P. 424-430.
104. Genome-wide high-density SNP linkage search for glioma susceptibility loci: results from the Gliogene Consortium / S. Shete [et al.] // *Cancer Research*. – 2011. – Vol. 71. – Genome-wide high-density SNP linkage search for glioma susceptibility loci. – № 24. – P. 7568-7575.
105. Guidelines on Genetic Evaluation and Management of Lynch Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer / F.M. Giardiello [et al.] // *Diseases of the Colon & Rectum*. – 2014. – Vol. 57. – Guidelines on Genetic Evaluation and Management of Lynch Syndrome. – № 8. – P. 1025-1048.
106. Li H. Fast and accurate short read alignment with Burrows-Wheeler transform / H. Li, R. Durbin // *Bioinformatics (Oxford, England)*. – 2009. – Vol. 25. – № 14. – P. 1754-1760.
107. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data / A. McKenna [et al.] // *Genome Research*. – 2010. – Vol. 20. – The Genome Analysis Toolkit. – № 9. – P. 1297-1303.
108. The genetic landscape of high-risk neuroblastoma / T.J. Pugh [et al.] // *Nature Genetics*. – 2013. – Vol. 45. – № 3. – P. 279-284.
109. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma / X. Bonilla [et al.] // *Nature Genetics*. – 2016. – Vol. 48. – № 4. – P. 398-406.
110. A method and server for predicting damaging missense mutations / I.A. Adzhubei [et al.] // *Nature Methods*. – 2010. – Vol. 7. – № 4. – P. 248-249.
111. Simultaneous identification and prioritization of variants in familial, de novo, and somatic genetic disorders with VariantMaster / F.A. Santoni [et al.] // *Genome Research*. – 2014. – Vol. 24. – № 2. – P. 349-355.

112. A panoply of errors: polymerase proofreading domain mutations in cancer / E. Rayner [et al.] // *Nature Reviews. Cancer.* – 2016. – Vol. 16. – A panoply of errors. – № 2. – P. 71-81.
113. Bebenek K. Analyzing fidelity of DNA polymerases / K. Bebenek, T.A. Kunkel // *Methods in Enzymology.* – 1995. – Vol. 262. – P. 217-232.
114. The molecular genetics underlying basal cell carcinoma pathogenesis and links to targeted therapeutics / J.K. Iwasaki [et al.] // *Journal of the American Academy of Dermatology.* – 2012. – Vol. 66. – № 5. – P. e167-e178.
115. Identification of the cell lineage at the origin of basal cell carcinoma / K.K. Youssef [et al.] // *Nature Cell Biology.* – 2010. – Vol. 12. – № 3. – P. 299-305.
116. Genomic Classification of Cutaneous Melanoma / R. Akbani [et al.] // *Cell.* – 2015. – Vol. 161. – № 7. – P. 1681-1696.
117. Characteristics of UV-induced mutation spectra in human XP-D/ERCC2 gene-mutated xeroderma pigmentosum and trichothiodystrophy cells / C. Marionnet [et al.] // *Journal of Molecular Biology.* – 1995. – Vol. 252. – № 5. – P. 550-562.
118. Drobetsky E.A. UV-induced G:C-->A:T transitions at the APRT locus of Chinese hamster ovary cells cluster at frequently damaged 5'-TCC-3' sequences / E.A. Drobetsky, E. Sage // *Mutation Research.* – 1993. – Vol. 289. – UV-induced G. – № 2. – P. 131-138.
119. Riboflavin activated by ultraviolet A1 irradiation induces oxidative DNA damage-mediated mutations inhibited by vitamin C / A. Besaratinia [et al.] // *Proceedings of the National Academy of Sciences of the United States of America.* – 2007. – Vol. 104. – № 14. – P. 5953-5958.
120. Brash D.E. UV signature mutations / D.E. Brash // *Photochemistry and Photobiology.* – 2015. – Vol. 91. – № 1. – P. 15-26.
121. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution / J. Hu [et al.] // *Genes & Development.* – 2015. – Vol. 29. – № 9. – P. 948-960.
122. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome / A.C. da Silva Almeida [et al.] // *Nature Genetics.* – 2015. – Vol. 47. – № 12. – P. 1465-1470.
123. Spectrum of p53 Gene Mutations Suggests a Possible Role for Ultraviolet Radiation in the Pathogenesis of Advanced Cutaneous Lymphomas / J.M. McGregor [et al.] // *Journal of Investigative Dermatology.* – 1999. – Vol. 112. – № 3. – P. 317-321.
124. Discovery and characterization of coding and non-coding driver mutations in more than 2,500 whole cancer genomes / E. Rheinbay [et al.]. – *Genomics*, 2017.
125. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data / K. Wang [et al.] // *Genome Research.* – 2007. – Vol. 17. – PennCNV. – № 11. – P. 1665-1674.

126. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution / E.C. de Bruin [et al.] // *Science*. – 2014. – Vol. 346. – № 6206. – P. 251-256.
127. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing / J. Zhang [et al.] // *Science*. – 2014. – Vol. 346. – № 6206. – P. 256-259.
128. APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication / J.I. Hoopes [et al.] // *Cell Reports*. – 2016. – Vol. 14. – № 6. – P. 1273-1282.