

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
НАУКИ «ИНСТИТУТ МОЛЕКУЛЯРНОЙ И КЛЕТОЧНОЙ БИОЛОГИИ
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК»

На правах рукописи

РОМАНОВ СТАНИСЛАВ ЕВГЕНЬЕВИЧ

**Роль белков CP190 и CG9879 в регуляции генов дифференцировки
сперматоцитов *Drosophila melanogaster***

1.5.7. - генетика

Диссертация на соискание ученой степени кандидата биологических наук

Научный руководитель

к.б.н. Лактионов П.П.

Новосибирск – 2022

ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ	5
ВВЕДЕНИЕ.....	6
Актуальность темы исследования	6
Степень разработанности темы исследования	7
Цели и задачи исследования	9
Научная новизна.....	10
Теоретическая и научно-практическая ценность.....	11
Основные положения, выносимые на защиту.....	12
Публикации по теме работы	12
Апробация работы.....	13
Вклад автора	14
Структура и объем работы	14
Благодарности.....	14
1. ОБЗОР ЛИТЕРАТУРЫ.....	15
1.1. Сперматогенез дрозофилы как модель эпигенетической регуляции дифференцировки.....	15
1.1.1. Стадии дифференцировки клеток в ходе сперматогенеза	15
1.1.2. Переход клеток зародышевого пути от пролиферации к терминальной дифференцировке	18
1.1.3. Семенник-специфичные гены	18
1.1.4. Гены задержки мейоза.....	20
1.1.5. Роль системы репрессии <i>Polysomb</i> в регуляции семенник- специфичных генов.....	24
1.1.6. <i>tTAF</i> и <i>tMAC</i> запускают регуляторный каскад активации генов в сперматоцитах	26
1.1.7. Гомология <i>tMAC</i> и <i>MMB/dREAM</i> указывает на роль инсуляторного белка <i>CP190</i> в регуляции семенник-специфичных генов.....	27

1.2. Роль инсуляторного белка CP190 в регуляции генов	30
1.2.1 Инсуляторы – энхансер-блокирующие и барьерные элементы генома	30
1.2.2. CP190 опосредует физические контакты между инсуляторами.....	32
1.2.3. CP190 организует барьеры хроматиновых доменов	34
1.2.4. Роль CP190 на границах доменов репрессии <i>Polysorb</i>	37
1.2.5. Активность инсуляторов динамична	37
1.3. Участие ТВР и его паралога в генетической регуляции	38
1.3.1. <i>Tbp</i>	39
1.3.2. <i>Trf</i>	39
1.3.3. <i>Trf2</i>	40
1.3.4. <i>CG9879</i> и <i>CG15398</i>	42
1.4. Заключение	43
2. МАТЕРИАЛЫ И МЕТОДЫ	44
2.1. Генетические конструкции	44
2.2. Молекулярное клонирование	45
2.3. Линии мух.....	48
2.4. Трансформация генома дрозофил	49
2.5. Генетические скрещивания	50
2.6. Экстракция геномной ДНК.....	52
2.7. Полимеразная Цепная Реакция	53
2.8. Секвенирование коротких нуклеотидных последовательностей	53
2.9. Выделение РНК и обратная транскрипция	53
2.10. Количественная ПЦР	53
2.11. Микроскопия	54
2.12. DamID-seq.....	55
2.13. Анализ транскриптома методом RNA-seq.	55
2.14. Поиск доменов обогащения H3K27me3 при помощи скрытой марковской модели (НММ).....	57
2.15. Доступность данных.....	57

3. РЕЗУЛЬТАТЫ.....	58
3.1. Генетическая система для условного спасения мутантов <i>Cp190</i>	58
3.2. Система условного спасения значительно снижает уровень эктопической экспрессии <i>Cp190</i> в клетках мужского зародышевого пути имаго.....	62
3.3. Мутация <i>Cp190</i> влияет на экспрессию генов дифференцировки в клетках мужского зародышевого пути	64
3.4. Мутация <i>Cp190</i> нарушает экспрессию генов-мишеней tMAC	71
3.5. Дифференцировка клеток сопровождается динамичным связыванием CP190 с хроматином	75
3.6. Активная транскрипция и SetDB1, но не связывание CP190, маркирует границы доменов H3K27me3	80
3.7. Роль гомолога ТВР CG9879 в регуляции генов дифференцировки мужского зародышевого пути.....	85
4. ОБСУЖДЕНИЕ	93
4.1. Исследование функции генов в системе условного спасения	94
4.2. Роль белка CP190 в регуляции генов дифференцировки сперматоцитов	96
4.3. Динамика связывания CP190 с хроматином в процессе дифференцировки клеток мужского зародышевого пути.....	98
4.4. SetDB1, но не CP190, может участвовать в поддержании доменной организации хромосом.....	100
4.5. Избыточная функция ТВР-подобных белков поддерживает стабильность генетической системы сперматогенеза	103
4.6. Заключение	105
ВЫВОДЫ.....	106
СПИСОК ЛИТЕРАТУРЫ.....	107

СПИСОК СОКРАЩЕНИЙ

DamID-seq - DNA adenine methyltransferase identification with sequencing

MMB/dREAM – комплекс MybMuvB/drosophila Rbf, E2F And Mip

PRC1/2 – Polycomb Repressive Complex 1/2

RNA-seq – RNA sequencing

TBP – TATA box binding protein, ТАТА бокс связывающий белок

TFIID – general transcription factor IID, базальный фактор транскрипции II D

tMAC – testis Meiotic Arrest Complex, семенник-специфичный комплекс задержки мейоза

TRF1/2 – TATA box binding protein-related factor 1/2

tTAF – testis TBP-Associated Factors, семенник-специфичные TBP-ассоциированные факторы

TSS – Transcription Start Site, сайт начала транскрипции

TES – Transcription End Site, сайт завершения транскрипции

кДНК – кодирующая ДНК

кПЦР – количественная ПЦР

мРНК – матричная РНК

ОТ-кПЦР – количественная ПЦР продуктов обратной транскрипции

ТАД – топологически ассоциированные домены

ВВЕДЕНИЕ

Актуальность темы исследования

Специализация клеток в ходе развития многоклеточных организмов сопровождается активацией тканеспецифичных генов. Ответственные за это эпигенетические программы сложны и многообразны, а их изучение является одной из важнейших задач генетики развития. Удобной моделью для изучения механизмов, управляющих специализацией клеток, является сперматогенез *D. melanogaster*, поскольку на каждой стадии дифференцировки клетки мужского зародышевого пути приобретают контрастные морфологические признаки, и все клеточные типы в семенниках легко идентифицируются [1]. Отличительной особенностью сперматогенеза дрозофилы является взрывная транскрипционная активность сперматоцитов. За короткий промежуток времени в этих клетках активируется более тысячи генов, которые необходимы для завершения мейоза и яркой морфологической трансформации гаплоидных клеток в специализированные сперматозоиды [2,3]. Интересно, что большинство этих генов больше нигде в организме не функционирует, то есть являются семенник-специфичными [3]. Нарушение масштабной активации семенник-специфичных генов в сперматоцитах приводит к остановке развития мужских половых клеток в профазе мейоза I [2,3]. Благодаря такому контрастному фенотипическому эффекту путем мутационного анализа были обнаружены ключевые гены, ответственные за запуск программы терминальной дифференцировки сперматоцитов [3]. Они кодируют транскрипционные факторы, активность которых в организме самцов ограничена недифференцированными клетками мужского зародышевого пути [4,5]. Однако молекулярные механизмы масштабной активации генов в сперматоцитах во многом остаются неясными. Изучение этих механизмов имеет важное фундаментальное значение, поскольку улучшит понимание того, как клетки контролируют

экспрессию генов в процессе специализации и какие процессы лежат в основе передачи наследственной информации между поколениями.

Степень разработанности темы исследования

К настоящему моменту обнаружено тринадцать генов, чья функция необходима для запуска генетической программы терминальной дифференцировки сперматоцитов [3,6]. Эти гены были названы генами задержки мейоза, и их продукты образуют два белковых комплекса, регулирующих транскрипцию [7]. Первый представляет из себя комплекс из семенник-специфичных ТВР-ассоциированных факторов tTAF, паралогичный базальному фактору транскрипции TFIID [8,9]. Второй – это мультисубъединичный семенник-специфичный комплекс задержки мейоза tMAC, чей паралог MMB/dREAM регулирует гены дифференцировки и клеточного цикла в соматических клетках [4,10].

Механизмы масштабной активации генов в сперматоцитах были изучены в нашей лаборатории при помощи высокопроизводительных методов геномного анализа. Исследование транскриптомов в семенниках у мутантов по генам задержки мейоза позволило выявить гены с tTAF и tMAC-зависимой активностью [11,12]. Кроме того, при помощи метода DamID-seq в клетках мужского зародышевого пути были прокартированы сайты связывания tTAF и tMAC. Это позволило доказать, что оба фактора служат транскрипционными активаторами. В то же время оказалось, что сотни семенник-специфичных генов, чья активация зависит от двух факторов, не связываются ими напрямую, то есть являются их непрямыми мишенями [11,12]. Это открывает вопрос о том, какие дополнительные механизмы ответственны за индукцию семенник-специфичных генов в сперматоцитах помимо транскрипционных активаторов tTAF и tMAC.

Было обнаружено, что tMAC и tTAF зачастую связываются с хроматином вдали от промоторов [11,12]. Это может указывать на то, что активация генов под действием двух комплексов зависит от дистанционных

регуляторных взаимодействий. Известно, что динамика экспрессии генов в ходе дифференцировки связана с изменением пространственной архитектуры хроматина [13]. В силу этого изучение функции архитектурных белков, регулирующих дистальные взаимодействия между участками хромосом, важно для понимания связи между организацией и функцией генома в клетках мужского зародышевого пути. У дрозофилы обнаружено более дюжины архитектурных белков, необходимых для поддержания трехмерной организации хроматина [14]. Примечательно, что паралог tMAC в соматических клетках, комплекс MMB/dREAM, привлекается к мишеням благодаря связыванию с белком CP190, который регулирует промотор-энхансерные взаимодействия в хромосомах дрозофилы [14–16]. В силу этого можно предположить, что регуляция генов под действием tMAC зависит от дистальных взаимодействий, организованных благодаря CP190. Стоит отметить, что до настоящего времени геномные исследования, изучавшие функции CP190, проводились либо на культурах клеток, либо в эмбрионах дрозофилы. По этой причине сохраняется множество вопросов о механизмах действия этого белка. Например, было показано, что CP190 находится на границах хромосомных доменов, обогащенных меткой H3K27me3, и препятствует распространению этой метки за пределы доменов [17,18]. Однако остается неясным, является ли поддержание доменной организации хроматина прямой функцией CP190 или он необходим только на ранних этапах эмбриогенеза, когда впервые устанавливаются домены H3K27me3 [18]. Изучение CP190 в клетках взрослого организма позволит не только выявить его роль в эпигенетической регуляции дифференцировки, но также сделает возможным проверить гипотезу о функции этого белка в поддержании доменной организации хромосом.

Большое число не прямых мишеней tTAF и tMAC может объясняться существованием транскрипционных факторов, чья экспрессия в сперматоцитах индуцируется под действием этих комплексов. Поиск ассоциированных с регуляцией транскрипции генов-мишеней tMAC

позволил выявить, что только ген *CG9879* теряет активность у мутантов с задержкой мейоза и кодирует транскрипционный фактор [12]. Эти данные указывают на то, что белок *CG9879* может являться вторичным регулятором в генетическом каскаде, приводящем к массовой активации генов в сперматоцитах. Ген *CG9879* является одним из паралогов *Tbr*, который кодирует ТАТА-бокс связывающий белок ТВР, принимающий участие в инициации транскрипции множества белок-кодирующих генов дрозофилы. Исследование роли белка *CG9879* в масштабной активации семенник-специфичных генов представляется крайне полезным в силу возрастающего числа исследований, посвященных роли ТВР и его паралогов в дифференцировке клеток [19].

Цели и задачи исследования

Целью данной работы было изучение роли белков *CP190* и *CG9879* в регуляции активности генов в сперматогенезе *Drosophila melanogaster*. Для достижения этой цели были сформулированы следующие задачи:

1. Разработать генетическую систему для специфичной инактивации белка *CP190* в клетках мужского зародышевого пути дрозофилы.
2. Установить сайты связывания *CP190* на хромосомах клеток мужского зародышевого пути на разных этапах дифференцировки. Определить влияние *CP190* на экспрессию генов в семенниках и исследовать динамику его связывания с регуляторными элементами генома, ассоциированными с tTAF и tMAC.
3. Исследовать связывание *CP190* с хромосомами на границах доменов H3K27me3 в терминально дифференцированных клетках дрозофилы.
4. Исследовать связывание белка *CG9879* с генами-мишенями транскрипционных активаторов tTAF и tMAC в клетках мужского зародышевого пути.

5. Получить мутантов с потерей функции гена *CG9879* и определить влияние этой мутации на экспрессию генов в семенниках дрозофилы.

Научная новизна

Мутанты по гену *Cp190* погибают, не доживая до стадии имаго, что существенно усложняет изучение его функций в клетках взрослого организма. Чтобы преодолеть эту проблему и исследовать регуляторные эффекты CP190 в развитии взрослых мух, впервые разработали систему условного спасения мутантов *Cp190*. Эта система включает «спасающую» конструкцию, в которой при помощи Cre/loxP-опосредованной рекомбинации кодирующая последовательность гена *Cp190* эффективно удаляется в выбранной популяции клеток.

При помощи разработанной системы была впервые показана связь между CP190 и регуляцией генов дифференцировки в сперматоцитах под действием tMAC. Было обнаружено, что инактивация *Cp190* в клетках мужского зародышевого пути приводит преимущественно к нарушению экспрессии tMAC-зависимых генов дифференцировки сперматоцитов. С использованием метода DamID-seq также впервые прокартированы сайты связывания этого белка с хроматином в клетках мужского зародышевого пути на разных этапах дифференцировки и обнаружено tMAC-зависимое взаимодействие CP190 с мишенями двух транскрипционных активаторов в сперматоцитах. Полученные результаты указывают на то, что основной функцией CP190 в ходе сперматогенеза является координация взаимодействий между генами дифференцировки и специфичными для них транскрипционными активаторами.

С использованием полученных данных мы смогли проверить гипотезу об участии CP190 в поддержании границ доменов H3K27me3. Мы не обнаружили корреляции между CP190 и расположением барьеров триметилирования H3K27 в дифференцированных клетках *in vivo*. Неожиданно, мы выявили четкую связь между гистон-метилтрансферазой

SetDB1 и пограничными элементами доменов репрессии Polycomb, что указывает на роль этого фактора в поддержании доменной организации хроматина.

Мы также впервые исследовали функцию белка CG9879 в контексте регуляции экспрессии генов. Были прокартированы сайты связывания этого белка на хромосомах клеток мужского зародышевого пути и обнаружена колокализация между CG9879 и tTAF. Это указывает на возможное участие исследованного белка в инициации транскрипции в составе семенник-специфичного варианта TFIIID. Однако мы показали, что удаление гена *CG9879* не приводит к нарушению морфологии семенников или снижению фертильности, а также вызывает только незначительные изменения в транскриптоме. Мы предполагаем, что функция *CG9879* компенсируется другими паралогами *Tbr*.

Теоретическая и научно-практическая ценность

Исследования последних лет демонстрируют, что специализация клеток и дифференциальная экспрессия генов в многоклеточном организме возможна благодаря динамичной архитектуре ядра и изменчивому эпигенетическому ландшафту [20]. Однако, хотя связь между инсуляторными белками и регуляцией генов на уровне архитектуры хроматина хорошо описана, полногеномные исследования инсуляторных белков у дрозофилы до настоящего времени проводилось главным образом на изолированных клеточных культурах вне контекста дифференцировки [21–24]. Обнаружение динамики связывания CP190 с хроматином клеток в ходе сперматогенеза расширяет существующие представления о механизмах действия инсуляторных белков *in vivo*. Полученные данные могут быть использованы для исследования динамичной архитектуры хроматина в развивающихся мужских половых клетках и позволят охарактеризовать механизмы масштабной активации семенник-специфичных генов в сперматоцитах на уровне трехмерной организации ядра.

Преимуществом разработанной нами системы условного спасения является то, что одновременно с утратой CP190 в клетках запускается экспрессия репортерного белка GFP, что позволяет изолировать их методами сортировки. Это дает возможность исследовать функцию инсульторного белка CP190 высокопроизводительными геномными методами специфично в развивающихся мужских половых клетках. Принцип работы данной системы универсален, что позволяет использовать ее для исследования других белков в разных типах клеток *in vivo*.

Основные положения, выносимые на защиту

1. Белок CP190 *Drosophila melanogaster* изменяет свой паттерн связывания с хромосомами в процессе дифференцировки клеток мужского зародышевого пути и в сперматоцитах локализуется в сайтах связывания семенник-специфичных транскрипционных факторов tMAC и tTAF, участвуя в регуляции tMAC-зависимых генов дифференцировки.
2. Белок CP190 не обязателен для поддержания функции барьеров на границах доменов H3K27me3 в терминально-дифференцированных клетках дрозофилы.
3. Семенник-специфичный транскрипционный фактор CG9879 регулирует в сперматоцитах *D. melanogaster* гены дифференцировки вместе с факторами tTAF и tMAC, однако его потеря слабо влияет на экспрессию генов в семенниках и не приводит к стерильности самцов.

Публикации по теме работы

По материалам диссертации опубликовано 3 научные работы в изданиях, индексируемых в научных базах данных «Scopus», «Web of Science» и РИНЦ:

1. Laktionov P.P., Maksimov D.A., **Romanov S.E.**, Antoshina P.A., Posukh O.V., White-Cooper H., Koryakov D.E., Belyakin S.N. Genome-wide analysis of gene regulation mechanisms during *Drosophila* spermatogenesis //

Epigenetics and Chromatin – 2018. – V. 11. – 14. DOI: 10.1186/s13072-018-0183-3. IF=5.465.

2. Kalashnikova D.A.*, Maksimov D.A.*, **Romanov S.E.***, Laktionov P.P., Koryakov D.E. SetDB1 and Su(var)3-9 play non-overlapping roles in somatic cell chromosomes of *Drosophila melanogaster* // J Cell Sci – 2021. №134(2). – jcs253096. DOI: 10.1242/jcs.253096. IF=5.235.

* - равный вклад.

3. **Романов С.Е.**, Шлома В.В., Коряков Д.Е., Белякин С.Н., Лактионов П.П. Инсуляторный белок CP190 регулирует экспрессию генов дифференцировки сперматоцитов в клетках мужского зародышевого пути *Drosophila melanogaster* // Молекулярная биология. IF=1.540. *Принята к печати.*

Апробация работы

Результаты работы были представлены на международных конференциях в «Хромосома-2018» и «Хромосомы и митоз» (Новосибирск, Россия) в форме стендовых сообщений, а также вошли в состав сборника «Методы редактирования генов и геномов»:

1. **Romanov S.E.**, Posukh O.V., Maksimov D.A., Belyakin S.N., Laktionov P.P. (2018). *In vivo* DamID mapping uncovered dynamic CP190 chromatin binding landscape in course of *Drosophila* spermatogenesis // International conference Chromosome – Новосибирск: Хромосома, 2018. – С. 68-69.

2. **Romanov S.E.**, Laktionov P.P., Belyakin S.N. (2019) The role of CP190 in genetic regulation of *Drosophila* spermatogenesis // International mini-conference Chromosomes and Mitosis – Новосибирск: НИГУ, 2019. – С. 30.

3. **Романов С.Е.**, Лактионов П.П. Получение направленных делеций в геноме *Drosophila melanogaster* с помощью системы CRISPR/Cas9 // Методы редактирования генов и геномов / отв. ред. Закиян С.М., Медведев

С.П., Дементьева Е.В., Власов В.В. – Новосибирск: Издательство СО РАН, 2020. – С. 133-148.

Вклад автора

Основные результаты получены автором самостоятельно. Работа по математической обработке данных выполнена совместно с к.б.н. П.П. Лактионовым и к.б.н. Д.А. Максимовым. Иммуноокрашивание и анализ препаратов на конфокальном микроскопе проводил к.б.н. В.В. Шлома. (ИМКБ СО РАН) Гибридизация РНК *in situ* была проведена Dr. Н. White-Cooper (Университет Кардиффа, Великобритания).

Структура и объем работы

Диссертация состоит из введения, 4 глав, выводов и списка литературы. Работа изложена на 130 страницах, содержит 15 рисунков и 5 таблиц.

Благодарности

Работа выполнена в лаборатории геномики ИМКБ СО РАН. Автор выражает глубокую признательность С.Н. Белякину и П.П. Лактионову за поддержку и помощь при проведении всех этапов исследования. Также автор благодарит В.В. Шлому за помощь в проведении экспериментов по иммуноокрашиванию фиксированных препаратов. Автор особенно благодарен Д.Е. Корякову из лаборатории молекулярной цитогенетики ИМКБ СО РАН за ценные советы и поддержку на протяжении всей работы.

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Сперматогенез дрозофилы как модель эпигенетической регуляции дифференцировки

Специализация клеток многоклеточного организма сопровождается активацией одних генов и отключением других. Дифференциальная экспрессия генов отражается не только на морфологии и функции клеток, но и на организации хроматина. Сперматогенез *D. melanogaster* является удобной моделью для изучения генной регуляции в процессе клеточной дифференцировки и выгодно отличается рядом свойств. Во-первых, внутри семенника можно легко отличить все стадии этого процесса: деление стволовых предшественников, пролиферацию слабодифференцированных клеток и терминальную дифференцировку [1]. Во-вторых, начало терминальной дифференцировки сопровождается масштабными перестройками хроматина с последующим запуском специфической генетической программы, которая в конечном итоге приводит к образованию зрелых гамет [3]. В-третьих, известны ключевые факторы, запускающие терминальную дифференцировку клеток мужского зародышевого пути [4,5]. Основные особенности сперматогенеза дрозофилы, а также молекулярные механизмы регуляции генов в ходе терминальной дифференцировки будут рассмотрены далее.

1.1.1. Стадии дифференцировки клеток в ходе сперматогенеза

Зрелые семенники дрозофилы имеют вид закрученной в спираль замкнутой с одного конца трубки из мышечных и пигментных клеток, заполненной клетками мужского зародышевого пути [1]. Сперматогенез *Drosophila melanogaster* начинается на апикальном конце семенника в так

называемом зародышевом пролиферативном центре, который состоит из трех типов клеток [25]. Ядро пролиферативного центра образовано 10-12 плотно сомкнутыми покоящимися клетками, формирующими структуру в форме конуса, которую называют хаб. Вокруг хаба располагается 5-9 стволовых клеток зародышевого пути, к каждой из которых прилегает по две соматические стволовые клетки [1]. Клетки пролиферативного центра имеют разное происхождение. Стволовые клетки зародышевого пути являются потомками примордиальных зародышевых клеток, которые формируются на заднем конце зародыша после девятого деления дробления. Клетки хаба и соматические стволовые клетки формируются из соматических клеток в висцеральной мезодерме эмбриона [26,27].

Дифференцировка мужских половых клеток у *D. melanogaster* начинается с асимметричного деления стволовой клетки зародышевого пути, в результате которого одна из дочерних клеток остается стволовой и сохраняет контакт с хабом, а вторая вытесняется из пролиферативного центра и становится гониобластом (**Рисунок 1**) [28]. Одновременно с этим асимметрично делятся две окружающие соматические стволовые клетки, каждая из которых дает начало одной цистой клетке. Две цистные клетки образуют капсулу вокруг гониобласта – цисту, больше не делятся и только увеличиваются в размерах [1]. Новые цисты вытесняют более зрелые к дистальному концу семенника, поэтому стадии дифференцировки клеток мужского зародышевого пути последовательно сменяют друг друга вдоль семенника.

Гониобласт претерпевает четыре раунда митотических делений, в результате которых образуется 16 сперматогониев (**Рисунок 1**). Цитоплазму сперматогониев соединяют кольцевые каналы, образовавшиеся в результате неполного цитокинеза. По завершению последнего митотического деления 16 сперматогониев быстро проходят предмейотическую S-фазу клеточного цикла и становятся сперматоцитами первого порядка [1].

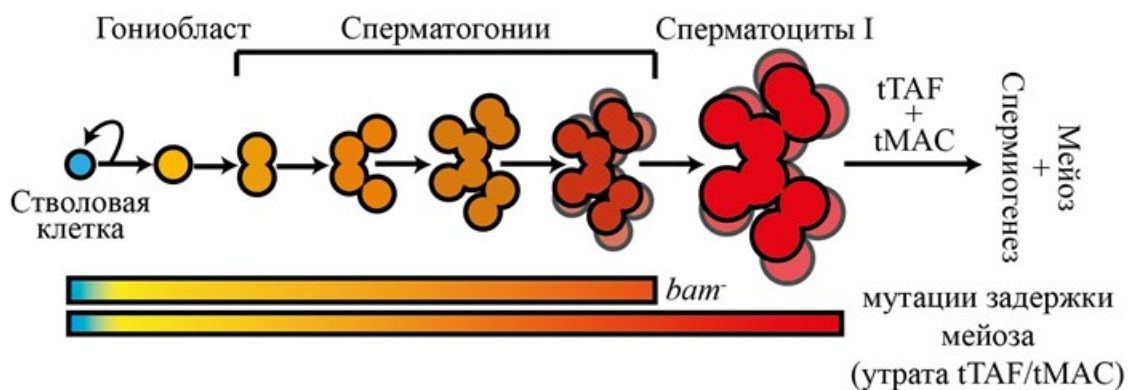


Рисунок 1. Стадии дифференцировки клеток мужского зародышевого пути. Клетки на разных этапах сперматогенеза указаны цветами от синего до красного. Мутации в гене *bam* и генах задержки мейоза нарушают дифференцировку клеток мужского зародышевого пути на ключевых этапах дифференцировки. Прямоугольниками внизу показаны типы клеток, которые присутствуют в семенниках у мутантов.

Все последующие этапы сперматогенеза принято объединять в стадию терминальной дифференцировки (**Рисунок 1**). Начинается терминальная дифференцировка с того, что сперматоциты вступают в продолжительную предмейотическую G2-фазу, или фазу роста, которая длится около 90 часов. В это время размер сперматоцитов увеличивается в 25 раз [29]. В фазе роста в клетках происходит активная транскрипция и накопление мРНК, однако со вступлением в мейоз транскрипция в сперматоцитах останавливается. После завершения мейотических делений синтез мРНК возобновляется только для 24 генов [30]. Таким образом, в сперматоцитах первого порядка происходит накопление подавляющего большинства транскриптов, необходимых для прохождения последующих этапов сперматогенеза.

В результате двух делений мейоза из 16 сперматоцитов формируется 64 гаплоидных сперматиды [1]. Вследствие неполного цитокинеза цитоплазма всех этих клеток по-прежнему соединена посредством кольцевых каналов. В дальнейшем сперматиды вступают в стадию спермиогенеза, в ходе которой эти клетки удлиняются, приобретают жгутик, отделяются друг от друга и превращаются в зрелые сперматозоиды. Наконец, сперматозоиды покидают цисту и скапливаются в семенном мешке, где хранятся до копуляции [1,31].

1.1.2. Переход клеток зародышевого пути от пролиферации к терминальной дифференцировке

Каждый гониобласт внутри цисты проходит четыре раунда митотических делений (**Рисунок 1**) [1]. На стадии 16-клеточной цисты клетки мужского зародышевого пути становятся сперматоцитами и вступают в длительную предмейотическую G2-фазу клеточного цикла. Ключевую роль в регуляции числа делений на стадии пролиферации играет продукт гена *bag of marbles (bam)* [32]. У гомозиготных мутантов *bam* пролиферация сперматогониев не останавливается на стадии 16 клеток, и формирования сперматоцитов не происходит [32,33]. В клетках зародышевого пути белок Bam впервые обнаруживается на стадии четырех сперматогониев, в дальнейшем его концентрация в этих клетках растет и достигает своего максимума в интерфазе восьмиклеточной цисты, а на стадии 16 сперматогониев этот белок исчезает из цитоплазмы [34]. У самцов с пониженной экспрессией *bam* сперматогонии вступают в предмейотическую интерфазу на стадии 32 и более клеток [34]. Напротив, повышенная доза этого гена приводит к формированию сперматоцитов уже на стадии 8 клеток. Преждевременная активация *bam* приводит к накоплению в семенниках одноклеточных цист, а также к дегенерации клеток зародышевого пути [34,35]. В то же время, накопление белка Bam в сперматоцитах приводит к нарушениям терминальных этапов дифференцировки мужских гамет и стерильности. Таким образом, концентрация белка Bam в клетках определяет момент перехода сперматогоний в сперматоциты.

1.1.3. Семенник-специфичные гены

Тканеспецифичные гены – это группа генов, которые экспрессируются и выполняют свою функцию преимущественно в одном типе клеток или тканей [36,37]. У дрозофилы около полутора тысяч генов экспрессируются

исключительно в мужских гонадах, то есть являются семенник-специфичными [11,38–40]. Интересно, что в других органах число тканеспецифичных генов не превышает нескольких сотен [40]. Экспрессия большинства семенник-специфичных генов запускается в сперматоцитах на стадии роста, что позволяет охарактеризовать их как гены дифференцировки сперматоцитов [3].

В соматических и недифференцированных клетках мужского зародышевого пути дрозофилы семенник-специфичные гены зачастую располагаются в областях транскрипционно неактивного хроматина [38,39]. В связи с этим предполагается, что активация семенник-специфичных генов должна сопровождаться масштабными перестройками хроматина [41].

Характерной особенностью семенник-специфичных генов является их склонность располагаться в кластерах – группах близкорасположенных генов со схожим уровнем транскрипции [38]. Нахождение генов в кластерах можно объяснить тем, что они произошли в результате дупликации. Однако ранее было показано, что лишь небольшая часть кластеров семенник-специфичных генов содержит паралогичные гены [38]. С другой стороны, близкое расположение коэкспрессирующихся генов может говорить о существовании общего механизма регуляции их транскрипции. В пользу этой гипотезы говорит ряд наблюдений. Так, в соматических клетках кластеры семенник-специфичных генов расположены в областях неактивного хроматина [38,42], в частности, в слюнных железах семенник-специфичные гены широко представлены в областях интеркалярного гетерохроматина [39]. Также показано, что репрессированное состояние некоторых кластеров семенник-специфичных генов может быть связано с их перемещением на периферию ядра за счет взаимодействия с ядерной ламиной [43]. Однако до настоящего времени общего регуляторного элемента в кластерах найти не удалось. Более того, разделение кластера генов с помощью протяженных инверсий не приводит к изменению уровня их экспрессии в организме дрозофилы [44].

1.1.4. Гены задержки мейоза

Ключевую роль в активации генетической программы терминальной дифференцировки мужских гамет дрозофилы играют так называемые гены задержки мейоза [45–47]. Мутации в этих генах приводят к невозможности клеток вступить в мейоз и остановке развития мужских гамет на стадии сперматоцитов первого порядка (**Рисунок 1**). В результате у мутантов по генам задержки мейоза семенники лишены сперматид и сперматозоидов и заполнены крупными шаровидными сперматоцитами.

По цитологическим особенностям и регуляторным эффектам гены задержки мейоза разделяют на два класса: *aly* и *can* [45]. У мутантов по генам *aly*-класса не происходит характерной для первого деления мейоза конденсации хромосом [7], и в семенниках нарушается экспрессия почти 1000 генов, среди которых гены клеточного цикла и спермиогенеза [12,45]. У мутантов по генам *can*-класса хромосомы в сперматоцитах находятся в частично конденсированном состоянии, характерном для профазы первого деления мейоза [7]. У таких мух в семенниках инактивируется порядка 600 генов спермиогенеза, причем те же гены инактивируются и у мутантов *aly*-класса.

В состав *can*-класса входят гены *can* (*cannonball*), *mia* (*meiosis I arrest*), *sa* (*spermatocyte arrest*), *rye* (*ryan express*), *nht* (*no hitter*) а также *mip40* (*Myb interacting protein 40*) [4,7]. Гены *nht*, *can*, *mia*, *sa* и *rye* экспрессируются семенник-специфично и являются паралогами генов TBP-ассоциированных факторов TAF4, TAF5, TAF6, TAF8 и TAF12, соответственно [8,46]. TBP-ассоциированные факторы (TAF, TBP Associated Factors) являются компонентами базального фактора транскрипции TFIID, который участвует в инициации транскрипции РНК-полимеразы II. На основе установленной гомологии продукты генов *can*, *mia*, *sa*, *rye* и *nht* были названы семенник-специфичные TBP-ассоциированные факторы, tTAF (*testis TBP associated factors*).

В недавних исследованиях было показано, что в двугибридной дрожжевой системе белки Can, Mia, Sa, Rye и Nht взаимодействуют с транскрипционными факторами tBRD-1/2/3 [9]. Гены *tbrd-1*, *tbrd-2* и *tbrd-3* экспрессируются исключительно в семенниках. Продукты этих генов характеризуются наличием бромодомена, обладающего сродством к ацетилированным лизинам в гистонах – маркерам активного хроматина. Мутации в генах *tbrd-1* и *tbrd-2* вызывают нарушения мейоза и приводят к стерильности самцов [48,49]. В семенниках дрозофилы эти гены начинают экспрессироваться в клетках зародышевого пути в фазе роста сперматоцитов. На основе этих фактов был сделан вывод о существовании единого комплекса, включающего белки tTAF, а также tBRD-1/2/3 [9]. В исследованиях нашей лаборатории было показано, что белок Can из этого комплекса связывается с промоторами генов, которые инактивируются у мутантов *can*-класса. Это позволяет охарактеризовать комплекс белков tTAF как семенник-специфичный активатор транскрипции, паралогичный TFIID [11].

В состав генов *aly*-класса входят семенник-специфичные гены *aly* (*always early*), *comr* (*cookie monster*), *tomb* (*tombola*), *topi* (*matotopetli*) и *achi/vis* (*achintya* и *vismay*) [7,47,50–52]. Ген *comr* является консервативным внутри рода *Drosophila*, но его гомологи не обнаружены у более отдаленных видов насекомых [47]. Для белка Comr предсказано наличие ДНК-связывающего мотива «спираль-поворот-спираль». Семенник-специфичный ген *topi* консервативен у насекомых, его продукт содержит мотив цинковых пальцев [51]. Гены *achi* и *vis* являются дублированными копиями друг друга и располагаются в одном локусе, их продукты практически идентичны друг другу и гомологичны транскрипционному фактору TGIF человека [50]. Гены *aly* и *tomb* являются семенник-специфичными паралогами генов *mir130* и *mir120* соответственно [4]. В клетках дрозофилы белки Mir130 и Mir120 входят в состав комплекса MMB/dREAM [53]. Этот комплекс крайне консервативен, его гомологи обнаружены как у животных, так и у растений

[54,55]. В соматических клетках дрозофилы он участвует в регуляции широкого круга генов, ответственных, например, за контроль клеточной пролиферации, дифференцировки и апоптоза [54,56].

Ген *mir40*, который принадлежит *cap*-классу генов задержки мейоза, кодирует одну из субъединиц комплекса MMB/dREAM и в организме дрозофилы экспрессируется повсеместно [4]. С помощью коиммунопреципитации удалось показать, что белок Mir40, а также еще один компонент MMB/dREAM, Caf1, образуют комплекс с продуктами генов *aly*, *comr*, *tomb* и *topi* [4]. Обнаруженный комплекс получил название testis Meiotic Arrest Complex (tMAC). У мутантов по компонентам tMAC в сперматоцитах инактивируется большое количество семенник-специфичных генов. Вместе с тем, в ряде исследований установлено, что компоненты комплекса tMAC локализуются в областях эухроматина [52,57,58], а в работе нашей лаборатории было показано участие tMAC в прямой активации семенник-специфичных генов [12]. Таким образом, tMAC является семенник-специфичным паралогом MMB/dREAM и служит активатором транскрипции в сперматоцитах.

С помощью метода коиммунопреципитации в семенниках дрозофилы был также обнаружен комплекс, образованный белками Achi/Vis, Aly и Comr, но не включающий Mir40, Topi или Tomb [58]. Поскольку транскрипционные факторы Achi и Vis проявляют свойства сайт-специфичных ДНК-связывающих белков, предполагается, что они привлекают белки Aly и Comr к генам-мишеням.

Еще один ген задержки мейоза – *wuc* (*wake up call*). Подавление экспрессии *wuc* в семенниках с помощью РНК-интерференции приводило к остановке развития сперматоцитов в фазе роста, но вызывало только незначительные изменения в экспрессии генов в семенниках [6]. Этот ген является паралогом гена *dLin52*, который кодирует компонент комплекса MMB/dREAM. В двугибридной дрожжевой системе Wuc физически взаимодействовал с белками Aly, Comr и Topi [6]. Можно предположить, что

Achi/Vis и Wuc входят в состав комплекса tMAC, но в отличие от белков Topi, Tomb, Aly, Comr и Mir40 теряются в ходе очистки продукта иммунопреципитации. С другой стороны, Achi/Vis и Wuc могут быть компонентами одного или нескольких схожих по составу комплексов, лишенных Mir40.

Помимо tTAF и tMAC в активации семенник-специфичных генов принимают участие другие регуляторные факторы. Во-первых, было установлено, что мутация в гене *dany* вызывает задержку мейоза в семенниках дрозофилы [59]. Продукт этого гена, гомологичный специфичным для нервной ткани транскрипционным репрессорам Danr и Dan, содержит ДНК-связывающий мотив «спираль-поворот-спираль». Считается, что Danу служит активатором семенник-специфичных генов в семенниках. Во-вторых, было показано, что подавление экспрессии в семенниках гена *med22*, кодирующего один из компонентов медиаторного комплекса транскрипции, также приводит к задержке мейоза [60]. Медиатор – это один из ключевых компонентов транскрипционной машины, который необходим для взаимодействия транскрипционных факторов с комплексом инициации транскрипции. Более того, в экспериментальных условиях Med22 физически взаимодействовал с компонентом комплекса tMAC - белком Topi, что может говорить о связи компонентов tMAC с медиаторным комплексом. В-третьих, обнаружено, что к задержке мейоза приводят мутации в генах *p8* и *p52*, кодирующих регуляторные субъединицы базального фактора транскрипции TFIIN [61].

Важную роль в активации семенник-специфичных генов в сперматоцитах первого порядка играют также факторы ремоделирования хроматина. Обнаружено, что удаление С-концевого домена у белка NURF301 – большой субъединицы комплекса ремоделирования хроматина NURF – приводит к формированию жизнеспособных самцов с задержкой мейоза [62]. Удаление же других районов белка NURF301 летально для мух. Интересно,

что по экспериментальным данным NURF301 взаимодействует с белком Caf1 [63,64], который входит в состав комплексов MMB/dREAM и tMAC.

Примечательно, что продукты генов задержки мейоза зачастую напрямую или через посредников взаимодействуют друг с другом в различных экспериментальных системах. Это может говорить о том, что в основе масштабной активации генов в сперматоцитах лежит цепочка последовательных взаимодействий между комплексами tMAC и tTAF, факторами ремоделирования хроматина, медиатором, базальными факторами транскрипции и, в конечном счете, РНК-полимеразой.

1.1.5. Роль системы репрессии Polycomb в регуляции семенник-специфичных генов

Факторы, необходимые для активации семенник-специфичных генов, известны и активно изучаются. Однако вопрос о том, каким образом семенник-специфичные гены поддерживаются в неактивном состоянии в соматических клетках, и предшественниках сперматоцитов остается открытым.

В работе Chen et al. (2005) было показано, что в сперматоцитах дикого типа белки tTAF накапливаются в ядрышках, где с ними колокализуются компоненты репрессорного комплекса Polycomb 1 (Polycomb Repression Complex 1, PRC1). Перемещение комплекса PRC1 в ядрышко совпадало с активацией генов *cap*-класса в ранних сперматоцитах, тогда как у мутантов по этим генам его ядрышковая локализация нарушалась. Мишенями PRC1 служат области хроматина, обогащенные меткой H3K27me3, которую устанавливает комплекс PRC2 (Polycomb Repression Complex 2), однако в ядрышках такой модификации хроматина не удалось обнаружить. Эти наблюдения говорили о том, что привлечение PRC1 в ядрышко происходит в присутствии tTAF независимо от PRC2.

Позже эти же авторы с помощью иммунопреципитации хроматина показали, что PRC1 связывается с тремя семенник-специфичными генами-мишенями tTAF (*dj*, *Mst87f*, *fzo*) у мутантов по генам *can*-класса, тогда как в диком типе связывания не происходит [41]. Кроме того, в отсутствие tTAF промоторы этих генов оказались обогащены H3K27me3. На основе этих наблюдений исследователи сделали вывод о том, что репрессия семенник-специфичных генов в малодифференцированных клетках мужского зародышевого пути обеспечивается комплексами Polycomb, и активаторные белки tTAF конкурируют с ними за связывание с промоторными областями генов-мишеней [41,65]. Кроме того, авторы предположили, что tTAF опосредует транспортировку комплекса PRC1 в ядрышко растущего сперматоцита. Стоит отметить, что в своей работе авторы не обсуждают функциональное значение такого перераспределения белков.

Дальнейшие исследования роли Polycomb в репрессии семенник-специфичных генов давали весьма противоречивые результаты: если в одних работах обогащение семенник-специфичных генов *dj*, *Mst87f* и *fzo* модификацией H3K27me3 было подтверждено [41,65], то в других работах обогащение было опровергнуто [66]. Полученные в нашей лаборатории полногеномные данные по связыванию белка Pc (комплекс PRC1) с хроматином в клетках мужского зародышевого пути не подтверждают гипотезу Chen: хотя Pc связывается с некоторыми прямыми мишенями Can и Comr в сперматогониях дрозофилы (в том числе *fzo* и *dj*), Pc не может считаться главным репрессором семенник-специфичных генов в недифференцированных мужских половых клетках (П.П. Лактионов, неопубликованные данные). Также было показано, что многие семенник-специфичные гены становятся мишенями Pc после формирования комплексов tMAC и tTAF. Согласно одной из современных гипотез, комплекс PRC1 по умолчанию связывается с генами до начала их транскрипции и в присутствии активаторов транскрипции удаляется из генов, но устанавливает сигналы репрессии, если активаторы отсутствуют [68]. В

соответствии с этой гипотезой связывание Рс с семенник-специфичными генами может происходить в момент переключения этих генов из неактивного состояния в активное и не связано напрямую с репрессией транскрипции.

1.1.6. tTAF и tMAC запускают регуляторный каскад активации генов в сперматоцитах

Механизмы масштабной активации генов в сперматоцитах были изучены в нашей лаборатории при помощи высокопроизводительных методов геномного анализа. Для этого было исследовано связывание компонентов tMAC и tTAF с хромосомами и изучено их влияние на транскрипцию. В частности, в геноме клеток мужского зародышевого пути были картированы сайты связывания Mir40 (tMAC и MMB/dREAM), Comr (tMAC) и Can (tTAF) [11,12]. Mir40 демонстрировал свойства и активатора, и репрессора, причем в сперматогониях его функция сводилась к подавлению активности генов [11]. Двойная функция Mir40 хорошо согласуется с тем, что он входит в состав активатора tMAC и репрессора MMB/dREAM. Напротив, распределение Comr и Can позволило охарактеризовать два белка как активаторы транскрипции семенник-специфичных генов, что согласуется с их ролью в регуляции транскрипции в составе tMAC и tTAF, соответственно.

В то же время было обнаружено, что в процессе масштабной индукции генов в сперматоцитах только часть генов подвергается прямой активации под действием tTAF или tMAC [11]. Например, среди 1043 генов, которые были инактивированы в семенниках у мутантов *comr*, только 243 содержали сайт связывания Comr в промоторах. В свою очередь, из 630 генов, чья активация происходила под действием Can, только 151 содержали сайт связывания этого фактора вблизи TSS. Таким образом, роль tMAC и tTAF

может заключаться в запуске специфичной генетической программы, которая разворачивается с помощью других регуляторов транскрипции.

Интересно, что среди семенник-специфичных генов, кодирующих транскрипционные факторы, только *CG9879* являлся прямой мишенью белка *Comr* [12]. Эти данные указывают на то, что белок *CG9879* может являться вторичным регулятором в генетическом каскаде, приводящем к масштабной активации генов в сперматоцитах. Функции белка *CG9879* не изучены, и его принадлежность к транскрипционным факторам предсказана на основе гомологии. Ген *CG9879* является одним из паралогов *Tbr*, который кодирует ТАТА-бокс связывающий белок ТВР. Этот фактор принимает участие в транскрипции множества белок-кодирующих генов дрозофилы и является одним из ключевых компонентов транскрипционной машины [69]. Обнаруженная связь между геном-паралогом *Tbr* и регуляцией транскрипции в сперматоцитах представляется крайне интересным наблюдением, поскольку предполагается, что в сперматоцитах семенник-специфичные белки tTAF функционируют аналогично компонентам базального фактора транскрипции TFIID (**Рисунок 2А**) [3,5,9,49]. Однако остается неясным, какие факторы могут выполнять роль ТВР в комплексе с tTAF. В то же время известно, что некоторые паралоги *Tbr* участвуют в инициации транскрипции тканеспецифичных генов [70]. Исследование функции *CG9879* в сперматогенезе могло бы расширить представления о роли паралогов ТВР в специализации клеток.

1.1.7. Гомология tMAC и MMB/dREAM указывает на роль инсуляторного белка CP190 в регуляции семенник-специфичных генов

Белок *Comr*, который входит в состав tMAC, связывается с тысячами сайтов в геноме сперматоцитов [11,12]. Вместе с тем, 43% сайтов связывания *Comr* обнаруживаются вдали от промоторов в межгенных областях. Можно предположить, что активация генов под действием tMAC может зависеть от

дистальных взаимодействий в геноме. Пространственная архитектура хромосом играет важную роль в регуляции генной экспрессии и зависит от функции архитектурных белков [13]. Однако у дрозофилы обнаружено более дюжины таких факторов, что затрудняет поиск белков-кандидатов, чья функция необходима для масштабной активации генов в сперматоцитах [14].

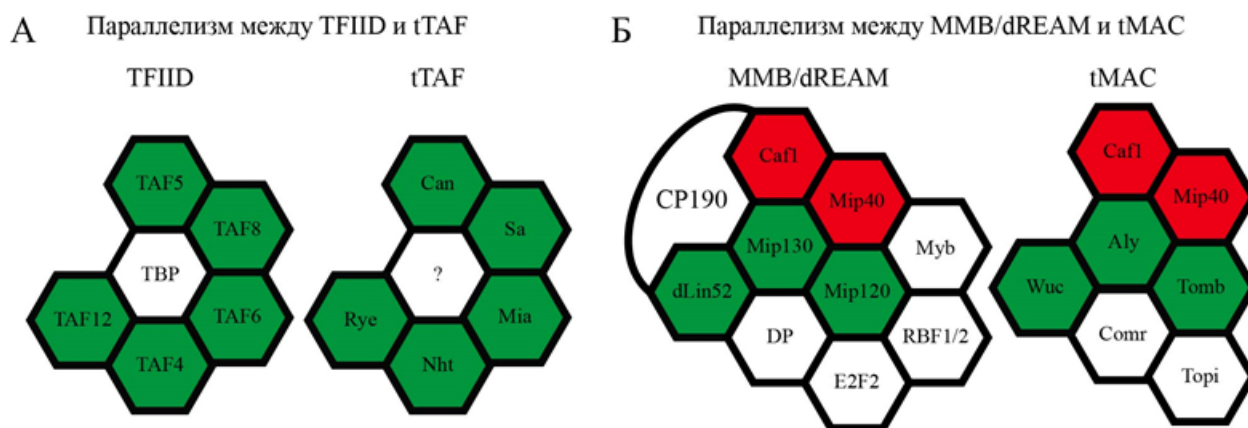


Рисунок 2. Сравнение tTAF и tMAC с паралогичными транскрипционными факторами. Гомологичные субъединицы показаны зеленым цветом. Красным цветом показаны повторяющиеся субъединицы. А) Сравнение tTAF и TFIIID позволяет предположить взаимодействие tTAF с одним из паралогов TBP. Б) Сравнение tMAC и MMB/dREAM позволяет предположить взаимодействие между tMAC и инсуляторным белком CP190.

В то же время, гомология tMAC и комплекса MMB/dREAM может указывать на возможные механизмы действия семенник-специфичного транскрипционного активатора. Комплекс MMB/dREAM (Myb-MuvB/*Drosophila* RBF, E2F, and Myb-interacting proteins) состоит из множества белков, включая транскрипционные факторы Myb, E2F2, DP, RBF1/2, Mip40, Mip120, Mip130, L(3)MBT, Lin52 гистоновый шаперон Caf1 (p55) и деацетилазу RPD3 (**Рисунок 2Б**) [53,71]. Функцией этого комплекса в соматических клетках является репрессия генов клеточного цикла, репликации, митоза и дифференцировки [54].

Особенностью MMB/dREAM является присутствие в его составе как репрессорных, так и активаторных субъединиц. Например, если E2F2 служит репрессором транскрипции, то белки Myb и Lin52 необходимы для активации некоторых генов, ассоциированных с MMB/dREAM, а Mip120 и

Mir130 задействованы как в репрессии, так и в активации [72,73]. Сочетание активаторных и репрессорных компонентов приводит к функциональному антагонизму между субъединицами MMB/dREAM. Например, потеря Myb сопровождается нарушениями конденсации и расхождения митотических хромосом в делящихся клетках, и поэтому летальна для зародышей дрозофилы. Однако двойные мутанты *tub mir130*, *tub mir120* или *tub mir40* жизнеспособны [4,74]. Подобным образом, мутация *lin-52* летальна для зародышей из-за остановки митоза на ранних стадиях развития, тогда как мутанты *lin-52 mir40* или *lin-52 mir130* доживают до стадии имаго [73]. Предполагается, что функция активаторных субъединиц Myb и Lin52 заключается в подавлении репрессорного воздействия компонентов E2F2, Mir40, Mir130, Mir120 [73,75,76]. Стоит отметить, что функциональный антагонизм обнаружен и в исследованиях семенник-специфичных генов *wis* (паралог *lin-52*) и *aly* (паралог *mir130*), кодирующих компоненты комплекса tMAC [6]. Это наблюдение подчеркивает степень гомологии между tMAC и MMB/dREAM и может указывать на схожий для двух комплексов механизм регуляции транскрипции.

Было показано, что комплекс MMB/dREAM регулирует активность генов через взаимодействие с инсуляторным белком CP190 [15,77]. Показано, что он необходим для связывания MMB/dREAM с хроматином и взаимодействует с компонентами этого комплекса [15,77]. Примечательно, что в клетках дрозофилы он регулирует дистальные промотор-энхансерные взаимодействия. Участки генома, ассоциированные с CP190 представляют собой барьеры, препятствующие физическому контакту между промоторами и дистальными регуляторными элементами [14–16]. С помощью репортерного анализа было показано, что участки генома, связанные с MMB/dREAM и CP190 блокируют промотор-энхансерные взаимодействия, но в отсутствие MMB/dREAM энхансер-блокирующая функция этих геномных элементов исчезает [77]. Поэтому предполагается, что MMB/dREAM кооперирует с CP190 и контролирует экспрессию генов через

регуляцию дистальных взаимодействий. Мы предполагаем, что тканеспецифичный активатор транскрипции tMAC может регулировать гены сходным образом. В силу этого важно проверить, зависит ли активация генов под действием tMAC от функции белка CP190 (**Рисунок 2Б**).

1.2. Роль инсуляторного белка CP190 в регуляции генов

Инсуляторные белки играют важную роль в регуляции генов. Взаимодействуя с инсуляторными элементами генома эти белки контролируют промотор-энхансерные взаимодействия и устанавливают границы между доменами хромосом. CP190 опосредует физические контакты между инсуляторами дрозофилы и считается одним из ключевых архитектурных белков *Drosophila melanogaster*. Ниже будет рассмотрена роль CP190 в регуляции транскрипции.

1.2.1 Инсуляторы – энхансер-блокирующие и барьерные элементы генома

Геномы эукариот организованы в ряд структурно и функционально независимых друг от друга доменов хроматина [78–80]. Границы доменов определяются благодаря пространственным взаимодействиям между специализированными *цис*-регуляторными элементами генома, называемыми инсуляторы. Впервые инсуляторы были обнаружены в геноме *Drosophila melanogaster*, а затем у дрожжей и позвоночных, включая человека [81–83]. С использованием репортерного анализа было выявлено два ключевых свойства инсуляторных элементов генома. Во-первых, инсуляторы способны блокировать промотор-энхансерные взаимодействия [81,84]. Такой эффект возникает, если в репортерной конструкции инсуляторы помещены между промотором и энхансером. Во-вторых, инсуляторы являются барьерными элементами и предотвращают распространение репрессорных эффектов гетерохроматина [81,85]. Это выражается в том, что трансгенные

конструкции, фланкированные инсуляторами, сохраняют активность после встройки в гетерохроматин. Некоторые инсуляторные элементы действуют главным образом как барьеры для гетерохроматина, другие могут обладать одновременно энхансер-блокирующей и барьерной активностью [86].

В геноме дрозофилы обнаружено несколько классических инсуляторных элементов. Среди них *scs* и *scs'* в локусе теплового шока *hsp70*, инсулятор в 5'-нетранслируемой области ретротранспозона *gypsy*, инсулятор *SF1* в кластере генов *Antennapedia* и серия элементов *Mcp*, *Fab-7* и *Fab-8* в регуляторной области гомеозисного гена *Abd-B* [84,87–91]. Более дюжины специализированных белков связываются с инсуляторами и отвечают за их функционирование у дрозофилы. Большинство белков напрямую связываются с ДНК. Например, за активность *scs* и *scs'* ответственны ДНК-связывающие белки Zest-white-5 (Zw5) и BEAF32; с инсуляторным элементом в ретротранспозоне *gypsy* связывается белок Su(Hw), с *Fab-7* и *SF1* связывается фактор GAGA (GAF), а с *Fab-8* связывается dCTCF [91–95]. Некоторые инсуляторные белки не способны связываться с ДНК напрямую и рекрутируются к инсуляторам ДНК-связывающими белками посредством белок-белковых взаимодействий. Например, инсуляторный белок Mod(mdg4)2.2 связывается с белком Su(Hw) благодаря своему С-терминальному домену и необходим для функционирования инсулятора *gypsy* [96]. Белок CP190 ассоциирован с большинством известных инсуляторов дрозофилы и рекрутируется к своим мишеням через взаимодействие с dCTCF, BEAF32 и Su(Hw) [16,22,97,98]. Поскольку утрата CP190 приводит к потере инсуляторами энхансер-блокирующей и барьерной функции, этот фактор считается одним из ключевых инсуляторных белков дрозофилы [17,99].

CP190 (Centrosomal Protein 190 kDa) длиной 1096 аминокислот был впервые обнаружен в эмбрионах дрозофилы на стадии дробления синцитиальной бластодермы как компонент centrosом с молекулярной массой 190 кДа [100]. После целлюляризации эмбрионов этот белок

обнаруживался в интерфазных ядрах, а в ядрах слюнных желез связывался с политенными хромосомами преимущественно на границах дисков и междисков [101,102]. Строение этого белка достаточно хорошо описано. N-терминальный домен ВТВ/POZ необходим для димеризации CP190 [16]. Рядом с ним располагается аспартат-обогащенный (D-Rich) домен, который обеспечивает взаимодействие CP190 с другими инсуляторными белками [23]. С-терминальный глутамат-обогащенный (E-Rich) домен регулирует связывание CP190 с мишенями. В середине аминокислотной последовательности располагается домен CENT, необходимый для ассоциации CP190 с микротрубочками и centrosомой [102]. Хотя изначально CP190 был обнаружен в цитоскелете, утрата этого белка оказывает лишь незначительный эффект на организацию микротрубочек и клеточное деление [103]. Кроме того, CENT домен оказался необязательным для жизнеспособности мух [23]. Тем не менее, потеря функции гена *Cp190* приводит к гибели мух в куколке, что указывает на важную роль этого белка в развитии дрозофилы [23,103].

1.2.2. CP190 опосредует физические контакты между инсуляторами

Способность формировать физические контакты друг с другом является характерной особенностью инсуляторов. Например, с помощью метода захвата конформации хромосом 3C было показано, что инсуляторы *scs* и *scs'*, связанные с Zw5 и BEAF32, физически взаимодействуют друг с другом [104]. В локусе *Bithorax* инсуляторы *Fab-7* и *Mcp*, ассоциированные с GAF, и элемент *Fab-8*, связанный с dCTCF и CP190, контактируют друг с другом с образованием петли хроматина [105,106].

С помощью метода иммунопреципитации (ChIP) исследователи в деталях изучили связывание инсуляторных белков с хромосомами в культивируемых клетках дрозофилы [18,22,107,108]. Инсуляторные белки связываются с тысячами сайтов в разных комбинациях друг с другом, что

позволило выделить среди инсуляторов от двух до шестнадцати классов регуляторных элементов, причем CP190 входит в большинство из них [18,22]. На геномном уровне CP190 колокализуется с белками dCTCF, Su(Hw), BEAF32 и GAF. Например, в клетках S2 больше половины сайтов Su(Hw) и dCTCF и 80% сайтов связывания GAF ассоциированы с CP190 [107]. Хотя CP190 не обладает ДНК-связывающей активностью, существенная доля его сайтов не колокализуется с известными инсуляторными белками, что указывает на существование неизвестных инсуляторных белков. Анализ партнеров CP190 по белок-белковым взаимодействиям позволяет обнаруживать новые ДНК-связывающие факторы, такие как Pita и ZIPIC, необходимые для функционирования инсуляторных элементов генома [109,110].

Механизм взаимодействия инсуляторов дрозофилы является предметом активных исследований. Согласно основной модели, ДНК-связывающие инсуляторные белки привлекают CP190, и он образует мост между инсуляторами благодаря образованию гомодимеров через свой домен VTB/POZ [16,111,112]. Этот механизм был продемонстрирован в эксперименте *in vitro*, где фрагменты ДНК, связанные с BEAF32, взаимодействовали друг с другом после добавления CP190 [16]. В экспериментах *in vivo* было показано, что CP190 необходим для образования петель хроматина в локусе *Abd-B* в клетках S2 [113]. Кроме того, в клетках Kc167, подвергавшихся воздействию экдизона, CP190 стабилизировал взаимодействия между инсуляторными элементами в экдизон-зависимом локусе *Eip75B*, и ограничивал активирующий эффект экдизона узким кругом экдизон-зависимых генов. Эта модель также подтверждается исследованием трехмерной организации хромосом, где было показано, что сайты связывания dCTCF, BEAF32 и CP190 взаимодействуют друг с другом с высокой частотой [80].

Хотя инсуляторы характеризуются способностью блокировать промотор-энхансерные взаимодействия, дистальные контакты между

инсуляторами, образованные при помощи CP190, могут приводить к сближению промоторов с дистальными регуляторными элементами [114–116]. Это проявляется в том, что в системах репортерного анализа встройка двух идентичных инсуляторов между энхансером и промотором восстанавливает способность энхансеров усиливать транскрипцию [117]. Другим примером служит энхансер *T1*, который запускает экспрессию гомеозисного гена *Scr* в каудальных головных сегментах дрозофилы благодаря петлевому взаимодействию промежуточных CP190-зависимых инсуляторов *SF1* и *SF2* [99,118]. В экспериментах *in vivo* показано, что петли ДНК между инсуляторами могут приводить к взаимодействию промотора и энхансера, разделенных расстоянием 140 кб [119–121]. Помимо энхансеров дистальные контакты потенциально могут приводить к репрессии генов. Возможный механизм был продемонстрирован в S2 клетках, у которых петли хроматина, образованные белком CP190 на границах доменов репрессии Polycomb, позволяют перескакивать меткам H3K27me3 из районов гетерохроматина в эухроматин через основание петель [122]. Стоит отметить, что роль CP190 в поддержании дистальных промотор-энхансерных взаимодействий не доказана и главной функцией этого белка считается блокирование промотор-энхансерных взаимодействий [99].

1.2.3. CP190 организует барьеры хроматиновых доменов

Исследования последних лет показывают, что хромосомы эукариот разделены на тысячи доменов, которые служат элементарными единицами генетической регуляции, а инсуляторные белки устанавливают границы этих доменов [80,123]. Визуально это разделение отражается в характерной исчерченности политенных хромосом, причем инсуляторные белки располагаются на границах дисков/междисков [124]. Систематический анализ белкового состава хромосом позволяет выявить доменную организацию на уровне структуры хроматина [125]. Каждый домен

характеризуется специфичной комбинацией эпигенетических модификаций, составом хроматиновых белков, скоординированной экспрессией генов, а также четкими границами. Примером служат районы обогащения гистоновой модификации H3K27me3 [17,18]. Эта метка ассоциирована с репрессией генов, осуществляемой белками группы Polycomb, и играет ключевую роль в регуляции сложного паттерна экспрессии генов [126]. Локус гомеозисных генов *Bithorax* служит классическим объектом для изучения регуляторной роли инсуляторов дрозофилы. Расположение границ домена H3K27me3 в данном локусе зависит от парасегмента тела и в точности совпадает с сайтами связывания инсуляторных белков [127]. В исследованиях *in vivo* показано, что инсуляторные белки dCTCF и CP190 необходимы для установки барьеров и обеспечивают правильное распределение H3K27me3 в локусе *Bithorax* [98,107,128].

Механизм действия инсуляторных белков в барьерных элементах H3K27me3 до конца не ясен. Одно из возможных объяснений заключается в способности инсуляторных белков регулировать процессы конденсации/деконденсации хроматина. В экспериментах на культурах клеток было показано, что привлечение CP190 к искусственным сайтам на хромосомах приводит к деконденсации хроматина в соответствующих областях ядра [129], в то же время нативные сайты CP190 обеднены нуклеосомами [107]. Это означает, что CP190 привлекает к местам своей локализации комплексы ремоделирования хроматина. В самом деле, для CP190 показано взаимодействие с ремоделирующим фактором NURF [77,130]. Активные промоторы, обогащенные CP190, характеризуются низкой плотностью нуклеосом, причем потеря этого белка приводит к обогащению нуклеосомами [17]. Помимо факторов ремоделирования, CP190 привлекает различные транскрипционные регуляторы и эпигенетические модификаторы, например, гистон-ацетилтрансферазу Gcn5, что приводит к активации гетерохроматина [131–133]. Таким образом, ассоциация CP190 с

активными генами и границами хромосомных доменов может быть следствием вмешательства CP190 в структуру хроматина.

Применение метода высокопроизводительного анализа конформации хромосом (Hi-C) позволило выявить, что хромосомы дрозофилы разделены на домены ~100 kb [79]. Взаимодействия между участками ДНК внутри этих доменов происходят с высокой частотой, но между участками в разных доменах контакты затруднены. Выявленные элементы ядерной архитектуры были названы Топологически Ассоциированными Доменами (ТАД). Сравнение Hi-C профилей между хромосомами эмбрионов, различных клеточных линий и политенизированных тканей дрозофилы позволило выявить высокую стабильность ТАД в ходе дифференцировки клеток [80,134,135]. Кроме того, ТАД зачастую совпадают с доменами репрессии Polycomb, ассоциированными с меткой H3K27me3 [80,136]. Как и ожидалось, инсуляторы и инсуляторные белки оказались вовлечены в разделение генома на ТАД. Классическим примером является ассоциированный с dCTCF и CP190 инсуляторный элемент *Fub*, который разбивает кластер генов *Bithorax* на два топологически ассоциированных домена. Сравнение профилей Hi-C с геномным распределением инсуляторных белков позволило обнаружить колокализацию CP190 и BEAF32 с границами ТАД [80,134]. В то же время не ясна причинно-следственная связь между инсуляторными белками и границами топологических доменов. Дело в том, что оба инсуляторных белка склонны связываться с промоторами активных генов, а активная транскрипция как раз наблюдается в участках хромосом между топологическими доменами [107,134]. Таким образом, высокая представленность CP190 и BEAF32 на границах топологических доменов может отражать их склонность к взаимодействию с активными генами.

1.2.4. Роль CP190 на границах доменов репрессии *Polyscomb*

Установка барьерных элементов может быть одной из основных функций CP190. В самом деле, как уже упоминалось, CP190 обеспечивают правильные границы распределения H3K27me3 в локусе *Bithorax* [98,107,128]. В клетках S2 в барьерные элементы, связанные с CP190 и dCTCF, обнаружены на границах доменов репрессии *Polyscomb* в четверти случаев [17], а в работе [18] в половине. Однако до сих пор неизвестно, необходим ли CP190 для поддержания барьерной функции элементов, граничащих с доменами репрессии *Polyscomb*. С одной стороны, на примере нескольких доменов H3K27me3 было продемонстрировано, что нокдаун *Cp190* приводит к распространению метки за пределы барьеров [17]. Напротив, полногеномный анализ показал, что нокдаун генов, кодирующих инсуляторные белки приводит к экспансии H3K27me3 только через 19% барьерных элементов [18]. При этом распространение метки происходило только в локусах с низкой транскрипционной активностью. Таким образом, несмотря на очевидную роль CP190 в поддержании состояния хроматина в отдельных локусах, в масштабе генома эта роль может оказаться весьма незначительной. Поэтому предполагается, что функция CP190 может быть наиболее важной, когда репрессия *Polyscomb* впервые устанавливается у эмбрионов [18]. Исследование CP190 в терминально-дифференцированных клетках может показать, сохраняется ли локализация CP190 вблизи границ доменов H3K27me3 в ходе дифференцировки клеток в организме. Это позволит определить, необходимо ли присутствие этого инсуляторного белка для поддержания состояния хроматина на границах доменов.

1.2.5. Активность инсуляторов динамична

Активность инсуляторных белков может регулироваться путем посттрансляционных модификаций [137,138]. Например, присоединение

полипептида SUMO к Mod(mdg4)2.2 и CP190 подавляет функцию инсуляторов, а нарушение механизмов сумоилирования приводит к усилению их энхансер-блокирующей активности. Также было обнаружено, что связывание CP190 с хроматином изменяется под действием теплового шока и экдизонового гормона [23,24]. Можно предположить, что контроль активности инсуляторов является инструментом эпигенетической регуляции. В таком случае связывание инсуляторных белков с хроматином должно быть тканеспецифичным. В пользу этого, сравнение клеток нейрального происхождения Kc167 и клеток гемопоэтической линии Mbn2 выявило, что некоторые сайты связывания инсуляторных белков появляются только в одном клеточном типе [108]. Например, для белка Su(Hw) тканеспецифичными являются 18% сайтов в клетках Kc167 и 5% сайтов в Mbn2. В свою очередь, для белка CP190 доля таких сайтов достигает 15%.

Можно ожидать, что дифференцировка клеток должна сопровождаться динамичным связыванием инсуляторных белков с хроматином. Однако стоит отметить, что исследование молекулярных механизмов действия инсуляторных белков до настоящего времени проводилось на клеточных культурах. По этой причине остается неясным, как динамическая активность инсуляторных белков связана с дифференциальной активностью генов, которая имеет место в процессе дифференцировки.

1.3. Участие TBP и его паралогов в генетической регуляции

CG9879 является членом небольшого семейства генов-паралогов *Tbp*, куда у дрозофилы входят также *Trf*, *Trf2* и *CG15398*. Известно, что некоторые паралоги *Tbp* участвуют в инициации транскрипции тканеспецифичных генов [70]. В связи с этим считается, что изменение структуры преинициаторного комплекса транскрипции может служить одним из механизмов транскрипционной регуляции [19,139].

1.3.1. *Tbp*

Активация транскрипции промотор-кодирующих генов у эукариот с РНК-полимеразой II происходит в коровой части промоторов. Первым базальным фактором транскрипции, который связывается с коровыми промоторами, является TFIID, который состоит из белка TBP и нескольких TBP-ассоциированных факторов TAF [140]. Считается, что TBP инициирует связывание TFIID с промоторами через взаимодействие со специфической последовательностью ДНК под названием ТАТА-бокс [141]. Консенсусный мотив ТАТА-бокса (TATAWAAR), располагается приблизительно на 30 нуклеотидов выше TSS и является крайне консервативной последовательностью, которая обнаружена даже у архей [142]. Хотя ТАТА-бокс был первым регуляторным элементом, обнаруженным в промоторах, и достаточно хорошо изучен, только 20% генов дрозофилы содержат эту последовательность [143,144]. Белок TBP содержит два основных домена. N-концевой домен относительно изменчив по последовательности и размеру между разными видами, но консервативен у позвоночных. С-концевой коровый домен крайне консервативен и сохраняет 80% гомологии между человеком и дрожжами [145]. Он состоит из двух прямых псевдоповторов и складывается в седлообразную структуру с выпуклой и вогнутой поверхностью. Выпуклая поверхность связывается с малой бороздкой ДНК ТАТА-бокса.

1.3.2. *Trf*

Белок Trf (TBP-related factor) обнаружен у двукрылых насекомых, но отсутствует у других животных [146,147]. Он имеет 63% гомологии с С-концевым коровым доменом TBP. В культурах клеток он связывается с ТАТА-боксом и ТС-богатыми кластерами ДНК и замещает TBP в процессе транскрипции некоторых ассоциированных с РНК-полимеразой II

промоторов [146,148]. Кроме того, Trf регулирует транскрипцию с некоторых промоторов, ассоциированных с РНК-полимеразой III [147,149,150]. Интересно отметить, что в сумме функции TBP и Trf у дрозофилы сходны с функциями TBP у организмов, у которых отсутствуют Trf, таких как дрожжи *Saccharomyces cerevisiae* [151]. Следовательно, появление Trf в эволюции насекомых привело к разделению первоначальных функций ТАТА-бокс связывающего белка между TBP и Trf.

1.3.3. *Trf2*

Trf2 - это еще один ген-паралог *Tbp*, который был обнаружен у множества модельных организмов, включая млекопитающих и насекомых [152–157]. У дрозофилы этот ген кодирует два белковых продукта: короткую изоформу длиной 632 аминокислотных остатка (Trf2S) и длинную изоформу из 1715 аминокислот (Trf2L), которая отличается от первой наличием N-концевого домена с мотивами спиральной катушки [158]. Синтез короткой изоформы, вероятно, происходит в результате внутренней инициации трансляции по механизму IRES [158]. Обе изоформы демонстрируют сходство с С-концевым коровым доменом TBP. Примечательно, что белок Trf2L был идентифицирован только у дрозофилы, тогда как короткая изоформа Trf2 высоко консервативна в эволюции [70,159].

Коровый домен Trf2 дрозофилы на 40% идентичен TBP [156]. Однако несмотря на гомологию между Trf2 и TBP, остатки фенилаланина, взаимодействующие с ТАТА-боксом в TBP, подверглись мутации в TRF2, из-за чего последний не может связываться с этим мотивом [153,156,159]. Таким образом, разнообразие факторов, связанных с TBP, добавляет еще один уровень сложности в регуляцию транскрипции.

Trf2 связывается с политенными хромосомами в сайтах, отличных от TBP [156,160]. Это позволяет предположить, что Trf2 регулирует подмножество генов, отличное от TBP. Это подтверждается тем, что Trf2

регулирует ген гистона H1 без TATA-бокса, тогда как TBP регулирует гены коровых гистонов, где TATA-бокс присутствует [160]. Более того, экспрессия гистона H1 и коровых гистонов разобщены во времени относительно S-фазы клеточного цикла [161]. Это подчеркивает, что разделение функции TBP и TRF2 необходимо для тонкой регуляции генов.

Помимо гистоновых генов Trf2 связывается с некоторыми генами рибосомных белков [160,162,163]. Примечательно, что эти гены не взаимодействуют с TBP, не распознаются каноническим комплексом TFIID и содержат в промоторной области специфический TCT-мотив. Кроме того, для Trf2 дрозофилы показано участие в регуляции генов с DPE-мотивом в промоторе (Downstream Promoter Element), но без TATA-бокса [163,164]. Это свидетельствует о том, что Trf2 регулирует специализированные транскрипционные пути независимо от TBP.

Гипоморфные мутации в гене *Trf2* приводят к дефектам развития, которые аналогичны таковым при нарушениях в системе экдизоновой регуляции, подразумевая участие этого фактора в контроле метаморфоза [165]. Оказалось, что Trf2 дрозофилы необходим для своевременной активации и надлежащего уровня экспрессии экдизон-зависимых генов, которые необходимы для вступления в метаморфоз [165]. Кроме того, было показано, что Trf2 дрозофилы необходим для дифференцировки половых клеток как самцов, так и самок [158]. Интересно, что у мышей Trf2 может формировать тканеспецифичные комплексы преинициации, которые регулируют транскрипцию определенных подмножеств генов, необходимых для дифференцировки половых клеток [166]. Это снова подтверждает гипотезу о том, что Trf2 регулирует транскрипцию специфических регуляторных путей и доказывает, что функции паралога *Tbp* важны для специализации клеток.

1.3.4. *CG9879* и *CG15398*

Гены *CG9879* и *CG15398* кодируют полипептиды длиной 331 и 305 аминокислотных остатков соответственно. Оба гена были отнесены к семейству TBP-подобных белков на основе автоматической аннотации [167]. Продукты двух генов обладают 40% сходством и демонстрируют гомологию с TBP и Trf2 [168]. Филогенетический анализ при помощи BLAST показывает, что *CG9879* и *CG15398* более тесно связаны друг с другом, чем с другими TBP-подобными белками. Два гена специфичны для видов *Drosophila* и у других животных не обнаруживаются. Так как эти гены имеют сходства в последовательности некодирующей ДНК, было предположено, что они возникли путем дупликации общего предшественника у предкового вида дрозофил [168]. В настоящее время функции этих генов не изучены. Мы обнаружили что среди семенник-специфичных генов, кодирующих транскрипционные факторы, только *CG9879* являлся прямой мишенью белка Comr [12]. Эти данные указывают на то, что белок *CG9879* может являться вторичным регулятором в генетическом каскаде, приводящем к масштабной активации генов в сперматоцитах.

1.4. Заключение

Транскрипционные факторы tTAF и tMAC разворачивают в сперматоцитах сложную генетическую программу, в результате чего в клетках активируется семенник-специфичные гены, необходимые для всех последующих стадий сперматогенеза. При этом многие гены активируются двумя комплексами не напрямую. Мы предположили, что белки CP190 и CG9879 являются наиболее вероятными претендентами на роль факторов, ответственных за непрямую регуляцию генов двумя транскрипционными активаторами. Исследование функции этих белков на модели сперматогенеза позволит раскрыть новые аспекты генетической и эпигенетической регуляции у дрозофилы. В частности, исследование связывания CP190 с хроматином в терминально-дифференцированных клетках может дать ключ к пониманию роли инсуляторных белков дрозофилы в поддержании доменной организации хромосом. В свою очередь, исследование CG9879 в контексте регуляции семенник-специфичных генов могло бы расширить представления о функции паралога TBP в специализации клеток.

2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Генетические конструкции

Все конструкции были клонированы в векторы на основе плазмиды *pUAST-attB* (GenBank EF362409), несущей сайт *attB* для *phiC31*-опосредованной интеграции, а также репортерный ген *mini-white*, обеспечивающий красную окраску глаз у мух. Конструкции для DamID белков CG9879 и CP190 были заклонированы в плазмиды *pUAST-attB-hsp70-Stop-Dam-CG9879* (GenBank KY930504) и *pUAST-attB-hsp70-Stop-Dam-CP190* (GenBank MT832326) соответственно. Конструкция для условного спасения мутантов *Cp190* была заклонирована в *CP190-mCD8-GFP* (GenBank ON783212).

pUAST-attB-hsp70-Stop-Dam-CG9879. Кодированная последовательность белка CG9879 была амплифицирована из кДНК-клона RT01008 (Bloomington *Drosophila* Stock Center) по праймерам CG9879L (5'-ctgcggccgcacatggattccgtgatcaaaag-3') и CG9879R (5'-gaggtacctaagagtaaagcaacatgccc-3') и встроена в вектор *pUAST-hsp70>Stop>Dam* (GenBank JN993988) [169] по сайтам рестрикции NotI и KpnI.

pUAST-attB-hsp70-Stop-Dam-CP190. Кодированная последовательность белка CP190 была амплифицирована из вектора *pUC18-lexA-CP190* (GenBank AQA27273) по праймерам 5'-ctgcggccgcaaatgggtgaagtcaagtccgt-3' и 5'-gaggtaccttatagctcctccttcgccg-3', в плазмиду *pUAST-hsp70>Stop>Dam* по сайтам KpnI и NotI.

CP190-mCD8-GFP. Химерный терминатор *HIS3-SV40* [170,171] был амплифицирован из плазмиды *p-attB-min.hsp70P-FRT-STOP#1-FRT-DamMyc[closed]* (Addgene #71810, предоставлена Пиндюриным А.В., ИМКБ

СО РАН) по праймерам 5'-ccgctagcaggtcggggacaccaaatat-3' и 5'-ttacgcgtgtcagatccagacatgataaga-3' и клонирован по сайту EcoRV в плазмиду *pUC18-2loxP* (предоставлена Белякиным С.Н., ИМКБ СО РАН) между двумя *loxP* сайтами, расположенными в одной ориентации. Далее флоксированный терминатор был амплифицирован с праймерами 5'-ccgcggccgcgtcgatcgacggtatcgccg-3' и 5'-aagcggccgcgggaattcgattgatcccc-3' и клонирован по сайту рестрикции NotI в плазмиду *pUAST-Ubi-mCD8GFP* (предоставлена Белякиным С.Н., ИМКБ СО РАН) в одной ориентации с промотором *Ubi63E*. Вслед за терминатором по KpnI сайту в идентичной ориентации была клонирована кодирующая последовательность гена *Cp190*, полученная путем амплификации из *pUC18-lexA-CP190* (GenBank KX357893) с праймерами 5'-gaggtaccttatagctcctcctcgccg-3' и 5'-aaggtaccatgggtgaagtcaagtcctga-3'.

Конструкции для CRISPR/Cas9-опосредованного удаления *CG9879* были получены следующим образом. Последовательности для направляющих химерных РНК были получены при помощи программы CRISPRdirect [172]. Пары комплементарных олигонуклеотидов L1: 5'-cttcgacgatggtgacaggtgtct-3', L2: 5'-aaacagacacctgtcaccatcgctc-3' R1: 5'-cttcgtgccagtgggtggcccgag-3' и R2: 5'-aaacctcgggccaaccactggcac-3' были подвергнуты отжигу и встроены в конструкцию *pU6-BbsI-chiRNA* (Addgene, #45946) по сайту BbsI [173]. В результате были получены конструкции *pU6-BbsI-chiRNA-left* (кодирует направляющую химерную РНК перед геном *CG9879*) и *pU6-BbsI-chiRNA-right* (кодирует направляющую химерную РНК вслед за геном *CG9879*).

2.2. Молекулярное клонирование

Амплификация ДНК-встроек. Амплификацию ДНК для клонирования производили с помощью Phusion-полимеразы (Thermo Scientific, F-530XL) в соответствии с рекомендациями производителя.

Расщепление ДНК эндонуклеазами рестрикции. Для проведения специфического гидролиза ДНК применяли эндонуклеазы рестрикции FastDigest (Fermentas) согласно протоколам производителя.

Разделение фрагментов рестрикции и ПЦР-продуктов методом электрофореза. Электрофорез ДНК проводили при постоянном напряжении 10 В/см в 1% агарозном геле, приготовленном на основе ТАЕ-буфера (40 мМ Трис-ацетат, 1 мМ ЭДТА) с добавлением бромистого этидия. На фореz образцы наносили после смешивания с буфером для нанесения (50% глицерин, 0.05% бромфеноловый синий, 0.05% ксиленцианол). Результаты детектировали с использованием системы гель-документации G:Box Chemi XT4 (Syngene).

Лигирование фрагментов ДНК. Лигирование фрагментов ДНК проводили с использованием Quick ligation kit (NEB, M2200S) согласно методике производителя.

Дефосфорилирование концевых участков ДНК. Удаление 5'-фосфатных групп в молекулах линеаризованной плазмидной ДНК проводили при помощи термолабильной щелочной фосфатазы (СибЭнзим, E365) согласно инструкции производителя.

Фосфорилирование концевых групп ДНК. Перед лигированием с плазмидой встройки, наработанные методом ПЦР или отжигом олигонуклеотидов, фосфорилировались с использованием Т4-Полинуклеотидкиназы (СибЭнзим, E311) согласно методике производителя.

Отжиг олигонуклеотидов для клонирования. Смесь комплементарных олигонуклеотидов после фосфорилирования нагревали до 95 °С 5 минут и охлаждали во льду в течение 20 минут.

Выделение ДНК из бактериальных суспензий, реакционных смесей и агарозного геля. Очистку ДНК из реакционных смесей, выделение

плазмидной ДНК в ходе клонирования, элюцию ДНК из агарозного геля проводили с помощью спин-микроколонок BioSilica (Биосилика, Россия) по протоколу производителя.

Трансформация бактерий методом электропорации. В пробирку с 50 мкл суспензии электрокомпетентных бактерий добавляли 1-10 мкл ДНК. Далее переносили суспензию бактерий в охлажденную 1 мм кювету (BIO-RAD, #1652089), и проводили электропорацию в режиме 5 мс / 1800 В на приборе BIORAD GENEPulser XCell. После электропорации немедленно ресуспендировали бактерий в 1 мл среды LB и переносили полученную суспензию в 1,5 мл центрифужную пробирку. Далее бактерий осаждали центрифугированием, осадок ресуспендировали в 100 мкл среды LB и высевали на чашки Петри с агаризованной средой (LB+1.5% бактериальный агар) с добавлением селективного антибиотика. Чашки Петри инкубировали при 37⁰С в течение 10-14 часов.

Получение электрокомпетентных бактерий. Клетки *E. coli* штамма *XL1-Blue MRF'* (Stratagene, #200230) выращивали в 5 мл среды LB (1% NaCl, 1% триптон, 0.5% дрожжевой экстракт) с добавлением тетрациклина (25 µg/ml) при 37⁰С в течение 8 часов. Затем переносили 1 мл полученной бактериальной суспензии в 500 мл среды LB и инкубировали при 23⁰С. По достижении оптической плотности 0,5 на длине волны 600 нм бактериальную суспензию охлаждали на льду в течение 30 минут. Далее клетки осаждали центрифугированием 15 минут при 1500 g и температуре 4⁰С. Бактериальный осадок дважды промывали в 250 мл автоклавированной сверхчистой воды и один раз в 20 мл охлажденного 10% глицерина и собирали центрифугированием в течение 15 минут при 1500 g. Очищенные клетки ресуспендировали в 2 мл охлажденного 10% глицерина и фасовали по 50 мкл и хранили при -70⁰С. Эффективность трансформации электрокомпетентных клеток составляла ~10⁸ КОЕ/µг.

Выделение апирогенной плазмидной ДНК. Очистку плазмидной ДНК для трансформации эмбрионов дрозофилы производили при помощи колонок Qiagen Plasmid midi kit (Qiagen, #12143) согласно протоколу производителя.

2.3. Линии мух

В Таблице 1 перечислены исходные линии мух, которые были использованы в работе. Стоковые линии содержались при температуре 18 градусов на стандартном корме.

Таблица 1. Список исходных линий, использованных в работе

Линия	Примечание
$y^l w^*$; $P\{y[+t7.7]=CaryP\}attP40; +; M\{vas-int.B\}ZH-102D$ (сокращенно $attP40$)	BDSC, #36304 и #2364
$y^l w^*$; $Mi\{y[+mDint2]=MIC\}MI04214$ (сокращенно $MI04214$)	BDSC, #37285
$y^l M\{RFP[3xP3.PB]GFP[E.3xP3]=vas-Cas9, U6-tracrRNA\}ZH-2Aw^{1118}$ (сокращенно $vas-Cas9$)	BDSC, #51326
$y^l w^{67}; If/CyO; can^l/TM6B, Tb^l$	[46]
$y^l w^{67}; If/CyO; bam^{\Delta 86}/TM6B, Tb^l$	[174]
$y^l w^{67}; mip40^{EY16520}/CyO$	[4]
$y^l w^{67}; If/CyO; Cp190^2, e^l/TM6B, Tb^l$	[103]
$y^l w^{67}; If/CyO; Cp190^3, e^l/TM6B, Tb^l$	[23]
$y^l w^{67}; hsp70-stop-Dam-Comr$	Конструкция $hsp70>Stop>DamComr$ интегрирована в сайт $attP40$ [12]
$y^l w^{67}; hsp70-stop-Dam$	Конструкция $hsp70>Stop>Dam$ интегрирована в сайт $attP40$ [12]
$y^l w^{67}; nanos-Cre$	Конструкция $nanos-Cre$ интегрирована в сайт $attP40$ [169]
$y^l w^{67}; Ubi-stop-mCD8-GFP$	Конструкция $Ubi>Stop>mCD8-GFP$ интегрирована в сайт $attP40$ [169]
$y^l w^{67}; If/CyO$	
$y^l w^{67}$	

2.4. Трансформация генома дрозофил

Интеграция конструкций в геном. Плазмиды *pUAST-attB-hsp70-Stop-Dam-CG9879*, *pUAST-attB-hsp70-Stop-Dam-CP190* и *CP190-mCD8-GFP* были очищены от пирогенов и инъецированы в пребластодерму эмбрионов линии *attP40* для ϕ C31-опосредованной трансформации [175,176]. Перед инъекцией эмбрионов собирали в течение часа после откладки при температуре 25⁰С, вручную дехорионизировали, прикрепляли к предметному стеклу с помощью двустороннего скотча, подсушивали 4-7 минут и покрывали тонким слоем минерального масла Halocarbon 200 (Halocarbon, #25073). Далее производили инъекцию раствора апирогенной плазмидной ДНК с концентрацией 0,1-0,25 мкг/мкл в полярную плазму зародышей. После инъекции эмбрионов содержали во влажной камере. Выживших личинок первого возраста переносили в стакан со стандартным кормом. Внедрение конструкции в геном мух происходило за счет экспрессии в полярных клетках интегразы ϕ C31, которая осуществляет рекомбинацию между сайтом *attP* на второй хромосоме и *attB* в векторной конструкции [175,176]. Используемый нами метод позволял на каждые 300 проколотых эмбрионов получать 30-50 фертильных имаго. Вылетевших имаго индивидуально скрещивали с мухами *y¹w⁶⁷;If/Cy*. Потомков анализировали на предмет экспрессии репортерного гена *mw* и выводили в гомозиготные линии. Список полученных трансгенных линий приведен в **Таблице 2**.

Редактирование генома с помощью CRISPR/Cas9. Для CRISPR/Cas9-опосредованного удаления *CG9879* аналогичным образом проводили инъекцию эквимолярной смеси плазмид *pU6-BbsI-chiRNA-left* и *pU6-BbsI-chiRNA-right*, в полярную плазму эмбрионов *y¹,M{vas-Cas9,U6-tracrRNA}ZH-2A,w¹¹¹⁸;MI04214*. Вылетевших имаго индивидуально скрещивали с мухами *y¹w⁶⁷;If/CyO*. Среди потомков отбирали мух, утративших маркер *yellow⁺* (серые крылья), и выводили в линию. Скрининг делеции *CG9879* проводили

при помощи праймеров для клонирования CG9879-L и CG9879-R, которые не давали продукта в случае удаления целевого гена. Было получено две линии с подтвержденной делецией CG9879 (Таблица 2). Для секвенирования делеции нарабатывали ДНК из локуса CG9879^{Δ1} при помощи праймеров 5'-ggactttggcatcggaatga-3' и 5'-aagacagtggcctccaatctg-3'. Секвенирование ПЦР-продукта проводили при помощи праймера 5'-ctggctagtgggtccctctg-3'.

Таблица 2. Список трансгенных линий, полученных в работе

Генотип	Метод получения
<i>y¹w⁶⁷;hsp70-stop-Dam-CG9879</i>	Интеграция <i>pUAST-attB-hsp70-Stop-Dam-CG9879</i> в сайт <i>attP40</i>
<i>y¹w⁶⁷;hsp70-stop-Dam-CP190</i>	Интеграция <i>pUAST-attB-hsp70-Stop-Dam-CP190</i> в сайт <i>attP40</i>
<i>y¹w⁶⁷;CP190-mCD8-GFP</i>	Интеграция <i>CP190-mCD8-GFP</i> в сайт <i>attP40</i>
<i>y¹w⁶⁷;CG9879^{Δ1}</i>	CRISPR/Cas9-опосредованное удаление CG9879
<i>y¹w⁶⁷;CG9879^{Δ2}</i>	

2.5. Генетические скрещивания

Стоковые линии мух, полученные путем скрещивания исходных и трансгенных линий перечислены в Таблице 3.

Таблица 3. Список стоковых линий, полученных в исследовании

Эксперимент	Линия
DamID-seq CP190 в клетках мужского зародышевого пути мутантов <i>bam</i>	<i>y¹w⁶⁷;hsp70-stop-Dam-CP190;bam^{delta86}</i> <i>y¹w⁶⁷;hsp70-stop-Dam;bam^{delta86}</i> <i>y¹w⁶⁷;nanos-Cre;bam^{delta86}</i>
DamID-seq CP190 и Comr в клетках мужского зародышевого пути мутантов <i>mip40</i>	<i>y¹w⁶⁷;hsp70-stop-Dam-Comr,mip40^{EY16520}</i> <i>y¹w⁶⁷/Y;hsp70-stop-Dam-CP190,mip40^{EY16520}</i> <i>y¹w⁶⁷/Y;hsp70-stop-Dam,mip40^{EY16520}</i> <i>y¹w⁶⁷/Y;nanos-Cre,mip40^{EY16520}</i>
DamID-seq CP190 в клетках мужского зародышевого пути мутантов <i>can</i>	<i>y¹w⁶⁷;hsp70-stop-Dam-CP190;can¹</i> <i>y¹w⁶⁷;hsp70-stop-Dam;can¹</i> <i>y¹w⁶⁷;nanos-Cre;can¹</i>
Условное спасение мутантов <i>Cp190</i>	<i>y¹w⁶⁷;CP190-mCD8-GFP;Cp190²,e¹/TM6</i>

	$y^l w^{67}; nanos-Cre; Cp190^3, e^l / TM6$
DamID-seq CP190 в личиночных слюнных железах	$y^l w^{67}; hsp70-Dam-CP190$ $y^l w^{67}; hsp70-Dam$

Для выведения мух $y^l w^{67}; hsp70-Dam-CP190$ и $y^l w^{67}; hsp70-Dam$ самок $y^l w^{67}; hsp70-stop-Dam-CP190$ и $y^l w^{67}; hsp70-stop-Dam$ скрещивали с самцами $y^l w^{67}; nanos-Cre$. Потомков F1 индивидуально скрещивали с мухами $y^l w^{67}; If/CyO$, и с помощью ПЦР идентифицировали мух, несущих Dam-содержавшую конструкцию с удаленной стоп-кассетой и сводили в линию

Для проведения Cre/loxP-опосредованной рекомбинации в семенниках скрещивали самок, несущих спасающую конструкцию, с самцами, несущими конструкцию *nanos-Cre*.

Генотипы мух, использованных в экспериментах RNA-seq и DamID-seq перечислены в **Таблице 4**.

Таблица 4. Список генотипов, использованных в экспериментах DamID-seq и RNA-seq

Эксперимент	Генотип
DamID-seq CP190 и CG9879 в клетках мужского зародышевого пути дикого типа	$y^l w^{67}/Y; hsp70-stop-Dam-CP190/nanos-Cre$ $y^l w^{67}/Y; hsp70-stop-Dam-CG9879/nanos-Cre$ $y^l w^{67}/Y; hsp70-stop-Dam /nanos-Cre$
DamID-seq CP190 в клетках мужского зародышевого пути мутантов <i>bam</i>	$y^l w^{67}/Y; hsp70-stop-Dam-CP190/nanos-Cre; bam^{\Delta 86}$ $y^l w^{67}/Y; hsp70-stop-Dam /nanos-Cre; bam^{\Delta 86}$
DamID-seq CP190 и Comr в клетках мужского зародышевого пути мутантов <i>mip40</i>	$y^l w^{67}/Y; hsp70-stop-Dam-Comr, mip40^{EY16520}/nanos-Cre, mip40^{EY16520}$ $y^l w^{67}/Y; hsp70-stop-Dam-CP190, mip40^{EY16520}/nanos-Cre, mip40^{EY16520}$ $y^l w^{67}/Y; hsp70-stop-Dam, mip40^{EY16520}/nanos-Cre, mip40^{EY16520}$
DamID-seq CP190 в клетках мужского зародышевого пути мутантов <i>can</i>	$y^l w^{67}/Y; hsp70-stop-Dam-CP190/nanos-Cre; can^l$ $y^l w^{67}/Y; hsp70-stop-Dam /nanos-Cre; can^l$
DamID-seq CP190 в личиночных слюнных железах	$y^l w^{67}/Y; hsp70-Dam-CP190$ $y^l w^{67}/Y; hsp70-Dam$

RNA-seq в семенниках мутантов	$y^1w^{67}/Y;CG9879^{A1}$
с делецией <i>CG9879</i>	y^1w^{67}/Y
RNA-seq в семенниках со спасением мутации <i>Cp190</i>	$y^1w^{67}/Y;CP190-mCD8-GFP/nanos-Cre;Cp190^2,e^1/Cp190^3,e^1$ $y^1w^{67}/Y;CP190-mCD8-GFP/nanos-Cre$ $y^1w^{67}/Y;Ubi-stop-mCD8-GFP/nanos-Cre$

2.6. Экстракция геномной ДНК

Образцы тканей дрозофилы собирали в 50 мкл охлажденного PBS (137 mM NaCl, 2,7 mM KCl, 10 mM NaH_2PO_4 , 1,8 mM KH_2PO_4) в 1,5 мл микроцентрифужных пробирках и хранили до выделения при -70°C . Затем гомогенизировали образцы в 700 мкл лизирующего буфера (состав: 0,1 M NaCl, 0,2 M сахароза, 0,1 M Tris HCl (pH = 9,1), 0,05 M EDTA (pH = 8,0), 0,5% SDS). К гомогенату добавляли диспазу II (Sigma, 42613-33-2) до концентрации 100 мкг/мл и инкубировали 12-16 часов при температуре 56°C . Далее экстрагировали нуклеиновые кислоты из лизата добавлением равного объема смеси фенол-хлороформа (в соотношении 1:1) с последующим центрифугированием 15 минут при 13000 g. Затем водную фазу инкубировали с 100 мкг РНКазы А (Биолот, Р-15-01) 30 минут при 37°C . ДНК экстрагировали фенол-хлороформом, затем смешивали с равным объемом хлороформа и центрифугировали 15 минут при 13000 g. Затем ДНК осаждали добавлением 1,5 объемов изопропилового спирта и последующим центрифугированием 20 минут при 13000g. Осадок ДНК промывали 1 мл 70% этанола и центрифугировали 5 минут при 13000g, супернатант удаляли, осадок высушивали на воздухе в течение 5 минут и растворяли в необходимом объеме автоклавированной сверхчистой воды.

2.7. Полимеразная Цепная Реакция

Для скрининга на наличие специфических коротких последовательностей в образцах ДНК использовали ПЦР с Taq-полимеразой (предоставлена Барановым К. О., ИМКБ СО РАН).

2.8. Секвенирование коротких нуклеотидных последовательностей

Секвенирующую реакцию проводили при помощи набора реактивов BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, #4337455) по рекомендациям производителя.

Очистку секвенирующей реакции проводили с помощью набора BigDyeX Terminator Purification Kit (Applied Biosystems, #4376486) в соответствии с протоколом производителя. Анализ последовательности проводили при помощи капиллярного секвенатора ABI 3500 Genetic Analyser (Applied Biosystems).

2.9. Выделение РНК и обратная транскрипция

Для анализа экспрессии генов в целых семенниках диссектировали 25 пар семенников и выделяли РНК с использованием реагента TRIZOL (Invitrogen, #15596018). Для наработки кДНК использовали 1 мкг тотальной РНК, обратную транскрипцию проводили с помощью Superscript II Reverse Transcriptase (Invitrogen, #18080093) с использованием олигонуклеотидов dT₂₀ в качестве затравки.

2.10. Количественная ПЦР

Количественную ПЦР проводили с использованием реакционной смеси БиоМастер HS-qPCR SYBR Blue (Биолабмикс, МНС030).

Последовательности использованных праймеров указаны в **Таблице 5**. Эффективность удаления спасающей кассеты оценивали при помощи метода $\Delta\Delta C_t$, экспрессию CP190 измеряли методом стандартной кривой. Для нормировки использовали ген *Actin42A*. Для построения стандартной кривой использовали серийные разведения образцов геномной ДНК личиночных имагинальных дисков самок $y^1w^{67};CP190-mCD8-GFP$.

Таблица 5. Список праймеров для количественной ПЦР

	Последовательность	Мишень
Actin42A	5'-acgcaaagagctgctagatg-3'	Кодирующая последовательность гена <i>Actin42A</i>
	5'-cctcgctccgttactgaccta-3'	
CP190_CDS	5'-acgcaaagagctgctagatg-3'	Четвертый экзон гена <i>Cp190</i>
	5'-cctcgctccgttactgaccta-3'	
CP190_3UTR	5'-cctcccatcccaagtatgtt-3'	3'-НТО область <i>Cp190</i> в пятом экзоне
	5'-tgtgaagaacggacactca-3'	
Res+	5'-ggatgagaatggggttagtgc-3'	Промежуточная область между кодирующей последовательностью <i>Cp190</i> и терминатором HIS3 в спасающей кассете <i>CP190-mCD8-GFP</i>
	5'-gtagatggatcgatggcaaca-3'	

2.11. Микроскопия

Гибридизацию РНК в семенниках *in situ* и подбор зондов проводили по методу, описанному в [177]. Для анализа флуоресценции GFP *in vivo* семенники переносили на предметное стекло в каплю PBS, накрывали покровным стеклом и сразу анализировали на флуоресцентном микроскопе. Для иммуноокрашивания семенники фиксировали стандартным методом при помощи 2% формальдегида в PBST [178]. В работе использовали первичные крысиные антитела anti-CP190 (любезно предоставлены А.К. Головниным) [179] и вторичные антитела Goat anti-Rat IgG (H+L) Cross-Adsorbed A568 (Thermo Fisher, A-11077). Препараты анализировали на конфокальном микроскопе Carl Zeiss LSM 710.

2.12. DamID-seq

По 25 пар семенников помещали в охлажденный буфер PBS. Затем по описанной выше методике из семенников выделяли геномную ДНК. Обработку образцов производили по протоколу, описанному в работе [180] с модификациями, указанными в [181]. Пробоподготовку образцов проводили с помощью Illumina DNA TruSeq Nano Kit (Illumina, #20015964) по протоколу производителя без селекции фрагментов по длине. Полученные библиотеки геномных фрагментов секвенировали на платформе Illumina MiSeq в режиме парных прочтений 2x75 п.н. Анализ секвенированных библиотек для получения геномных профилей проводили по алгоритму, описанному в работе [181]. Два независимых биологических повтора было получено для каждого образца. Данные по связыванию белков Can, Comr, Mir40, Pc и SetDB1 с хромосомами были получены ранее при помощи метода DamID-seq (GEO GSE97182, GSE155579) [11,182].

2.13. Анализ транскриптома методом RNA-seq.

Тотальную РНК выделяли из 25 пар семенников с использованием реагента TRIZOL. Пробоподготовку библиотек проводили с использованием набора TruSeq RNA Sample Preparation v2 Kit (Illumina, RS-121-2002). Секвенирование библиотек осуществляли на платформе Illumina MiSeq в режиме парных прочтений 2x75 п.н. Данные по экспрессии генов в целых семенниках мутантов *can*, *comr*, *mir40*, *bat* были получены ранее и доступны в базе данных Gene Expression Omnibus (GSE97129) [11]. Данные по экспрессии генов в личиночных слюнных железах были взяты из базы данных GEO по идентификатору GSE33017 [183].

Делеция CG9879. Эффект делеции *CG9879* исследовали у трехдневных самцов $y^1w^{67}/Y;CG9879^{41}$, а в качестве контроля использовали семенники y^1w^{67} (Таблица 4). Было исследовано по два биологических повтора на

образец. Данные высокопроизводительного секвенирования были проанализированы с помощью набора инструментов Galaxy [184]: прочтения были выравнены на сборку генома *D. melanogaster* BDGP R5/dm3 (<https://genome-euro.ucsc.edu/>) с использованием TopHat (-r 200 –mate-std-dev 50) [185]. Анализ дифференциальной экспрессии транскриптов был проведен при помощи программы Cuffdiff с использованием геометрической нормализации, объединенной оценкой дисперсии и граничным значением FDR=0.05 [186].

Условное спасение мутантов Cp190. Эффект конструкций для условного спасения на экспрессию генов в семенниках исследовали у трехнедельных самцов $y^1w^{67}/Y;CP190-mCD8-GFP/nanos-Cre;Cp190^2,e^1/Cp190^3,e^1$ и $y^1w^{67}/Y;CP190-mCD8-GFP/nanos-Cre$ (Таблица 4). В качестве контроля использовали семенники самцов $y^1w^{67};Ubi>stop>mCD8-GFP/nanos-Cre$, с репортерной конструкцией $Ubi>stop>mCD8-GFP$, интегрированной в сайт *attP40* во второй хромосоме [169]. Для каждого генотипа было получено от двух до трех биологических повторов. Геномное выравнивание осуществляли инструментом HiSat2 [187] (сборка генома dm6, Release 6 plus ISO1 MT, Aug. 2014) со стандартными параметрами. Количество прочтений для каждого гена было суммировано при помощи FeatureCounts [188] с использованием геномной аннотации UCSC refGene (<https://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/genes/dm6.refGene.gtf.gz>). Анализ дифференциальной экспрессии генов проводился при помощи алгоритма DESeq2 [189] с параметрами “ashr” [190] и “ihv” [191]. Данные по экспрессии генов в тканях дрозофилы были получены из базы modENCODE [192] и кластеризованы при помощи программы Cluster 3.0 [193]. Анализ обогащения тканеспецифичными генами проводили при помощи пакета TissueEnrich [37]. Данные секвенирования транскриптома единичных клеток семенников были получены из базы Fly Cell Atlas и проанализированы с использованием программного пакета Seurat [194,195].

2.14. Поиск доменов обогащения H3K27me3 при помощи скрытой марковской модели (HMM)

Данные эксперимента ChIP-chip для H3K27me3 в политенных хромосомах личиночных слюнных желез были получены из базы данных GEO (GSE33017) [183]. Эти данные были использованы для расчета распределения сигналов $\log_2(\text{IP}/\text{Input})$ вдоль хромосом. Профиль распределения для каждой хромосомы был проанализирован как независимый дискретный временной ряд. Для этого распределение было сглажено с помощью алгоритма скользящего среднего, а затем с помощью пакета `dermixS4` [196] была определена наиболее вероятная последовательность из трех скрытых состояний. Точки перехода между состояниями разбили хромосомы на фрагменты. Если фрагмент имел среднее значение $\log_2(\text{IP}/\text{Input})$ больше нуля, то его считали доменом H3K27me3. Если домены располагались вплотную, то их объединяли в один. Далее координаты границ доменов H3K27me3 были скорректированы. Для этого с помощью функций `computeMatrix` и `plotHeatmap` из набора программ `deepTools2` [197] была построена тепловая карта обогащения H3K27me3 вблизи первичных границ доменов. Затем каждая граница была сдвинута так, чтобы переходы между областями обогащения/обеднения H3K27me3 лежали на одной линии тепловой карты.

2.15. Доступность данных

Данные RNA-seq в семенниках со спасением *Cp190* были загружены в базу данных BioProject (PRJNA847720). Результаты RNA-seq мутантов с делецией *CG9879* доступны по идентификатору GSE97129. Данные по связыванию *CG9879* в клетках мужского зародышевого пути доступны по идентификатору GSE97182. DamID-seq *CP190* в семенниках дикого типа доступен по идентификатору GSE155579.

3. РЕЗУЛЬТАТЫ

Целью данной работы является изучение роли белков CP190 и CG9879 в регуляции активности генов в сперматогенезе *Drosophila melanogaster*. Белок CP190 необходим для жизнеспособности мух, и мутанты по гену, кодирующему этот белок, не доживают до стадии имаго. Поэтому сначала мы разработали генетическую систему, которая позволила нам добиться утраты этого белка специфично в клетках мужского зародышевого пути. С использованием этой системы мы исследовали изменение активности генов в семенниках у мух с нарушением функции CP190 в половых клетках. Далее с помощью DamID-seq мы исследовали динамику связывания этого белка с хроматином клеток в процессе сперматогенеза. С помощью полученных данных мы смогли изучить, сохраняется ли CP190 на границах доменов репрессии Polycomb в терминально дифференцированных клетках дрозофилы. Далее мы исследовали роль CG9879, гомологичного TBP, в регуляции генов в семенниках. Для этого с помощью DamID-seq мы прокартировали сайты связывания этого белка на хромосомах клеток мужского зародышевого пути и сопоставили с распределением tMAC и tTAF. Чтобы изучить функцию CG9879 в генетической регуляции мы получили мух с делецией соответствующего гена и проанализировали изменение транскрипции генов в семенниках у мутантов.

3.1. Генетическая система для условного спасения мутантов *Cp190*

Мутанты по гену, кодирующему инсультный белок CP190, не доживают до взрослой стадии [23,103], поэтому исследование функции этого гена в тканях имаго сопряжено с экспериментальными трудностями. Чтобы преодолеть эту проблему и установить регуляторную функцию CP190 в

клетках зародышевого пути взрослых самцов, мы разработали генетическую систему для условного спасения мутантов *Cp190* (**Рисунок 3А**).

Система состоит из двух генетических конструкций. Первая – спасающая конструкция *CP190-mCD8-GFP* – содержит кодирующую последовательность гена *Cp190* под контролем конститутивного промотора гена *Ubi63E*, который успешно использовали ранее для спасения летальной мутации *Cp190* [198,199]. Вслед за *Cp190* помещен гибридный терминатор транскрипции *HIS3-SV40*, и кодирующая последовательность гена химерного репортерного белка *mCD8-GFP* [171,200]. Кодирующая последовательность *Cp190* и терминатор транскрипции фланкированы *loxP*-сайтами узнавания рекомбиназы Cre и образуют спасающую генную кассету (**Рисунок 3А**). Вторая конструкция содержит ген рекомбиназы Cre под управлением промотора гена *nanos*, что обеспечивает наработку белка исключительно в клетках зародышевого пути (**Рисунок 3А**) [11,169]. При наличии обеих конструкций в организме мутантов *Cp190* спасающая кассета будет специфически удалена в клетках зародышевого пути рекомбиназой Cre. В соматических клетках спасающая кассета не вырезается, обеспечивая жизнеспособность организма. При удалении кассеты в клетках зародышевого пути кодирующая последовательность гена химерного репортерного белка *mCD8-GFP* сближается с промотором гена *Ubi63E*, что позволяет получить *Cp190*-дефицитные клетки, помеченные GFP (**Рисунок 3Б**).

С использованием системы *phiC31*-опосредованного трансгеноза генетическая конструкция *CP190-mCD8-GFP* была встроена в сайт *attP40* во второй хромосоме [176]. Трансгенные мухи были жизнеспособны и фертильны как с одной, так и с двумя копиями конструкции. При этом одной копии оказалось достаточно для спасения летального фенотипа компаунд-гетерозигот *Cp190²/Cp190³*, у которых не образуется функциональный белок CP190 [23,103]. Одновременно с этим, экспрессия одной или двух копий спасающей конструкции не вызывала аномалий развития как в комбинации с *Cp190²/Cp190³*, так и на фоне дикого типа. Это свидетельствует об

умеренном уровне эктопической экспрессии CP190, поскольку сильное повышение экспрессии этого гена вызывает серьезные дефекты развития [198].

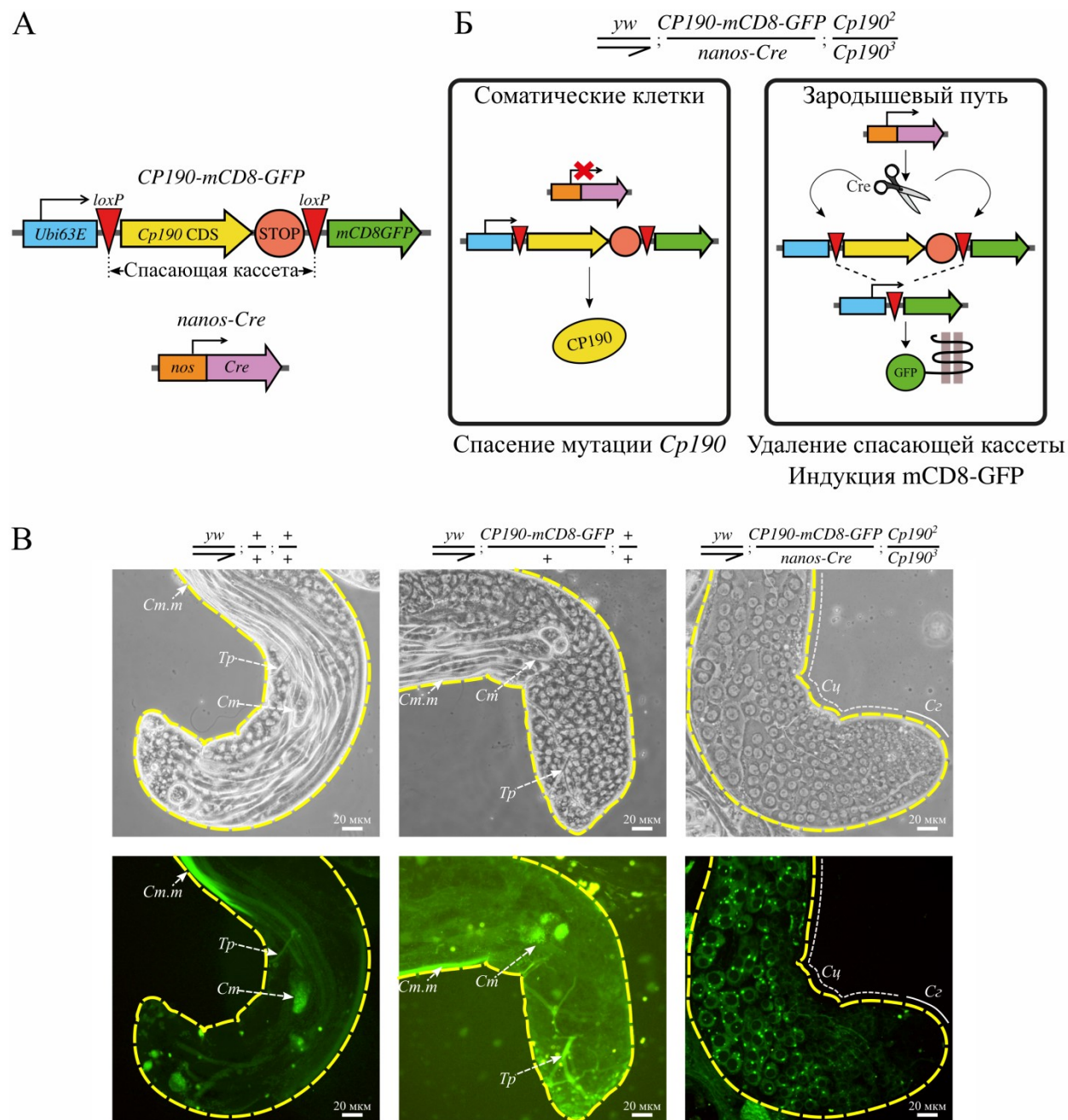


Рисунок 3. Генетическая система для условного спасения мутантов с потерей функции *Cp190*. А) Схема спасающей конструкций *Cp190-mCD8-GFP* и конструкции *nanos-Cre* для экспрессии рекомбиназы Cre в клетках зародышевого пути дрозиды. Б) Принципиальная схема условного спасения мутантов *Cp190²/Cp190³* (см. пояснения в тексте). В) Флуоресценция GFP в семенниках без спасающей конструкции, с одной дозой спасающей конструкции и у мутантов с условным спасением. Желтым

пунктиром обозначены границы семенников. Условные обозначения: *St.m* – хвосты сперматид; *St* – головы сперматид; *Tr* – трахеи; *C2* – область семенника, содержащая сперматогонии (непрерывная линия); *C4* – область семенника, содержащая сперматоциты (пунктирная линия). Масштаб – 20 мкм.

Серией генетических скрещиваний были получены самцы, несущие конструкции *CP190-mCD8-GFP* и *nanos-Cre* на фоне мутации *Cp190* (y^1w^{67} ; *CP190-mCD8-GFP/nanos-Cre*; *Cp190²/Cp190³*). В клетках зародышевого пути таких самцов, начиная со стадии ранних сперматоцитов, детектировалась активация репортерного белка mCD8-GFP, что свидетельствовало об удалении спасающей генной кассеты (**Рисунок 3В**). Такие самцы обладали нормальной жизнеспособностью и не демонстрировали видимых нарушений сперматогенеза. GFP-негативные сперматоциты и GFP-позитивные соматические клетки оболочки визуально не были обнаружены, что указывает на высокую эффективность и специфичность удаления спасающей конструкции.

В контрольных образцах семенников y^1w^{67} , а также семенников самцов, несущих спасающую конструкцию в отсутствие Cre-рекомбиназы, имела место только фоновая автофлуоресценция, характерная для дыхательных трубочек и фрагментов сперматид (**Рисунок 3В**). Это подтверждает надежное блокирование экспрессии репортерного белка во всех типах клеток семенника.

Дополнительная проверка эффективности удаления спасающей кассеты из генома клеток зародышевого пути при помощи количественной ПЦР показала, что спасающая конструкция сохраняется в 49,6% клеток целого семенника (**Рисунок 4А**). Это значение соответствует доле соматических клеток в семеннике (50.4%), оцененной исходя из данных секвенирования транскриптома единичных клеток семенника [194]. Этот результат свидетельствует о том, что спасающая кассета удаляется в подавляющем большинстве половых клеток.

3.2. Система условного спасения значительно снижает уровень эктопической экспрессии *Cp190* в клетках мужского зародышевого пути имаго

Мы проанализировали уровень экспрессии мРНК эндогенного и трансгенного *Cp190* методом количественной ПЦР, для чего были подобраны праймеры на кодирующую область гена *Cp190* и 3'-нетранслируемую область эндогенного *Cp190* (**Рисунок 4Б**). Использование первой пары праймеров позволяет измерить уровень экспрессии и эндогенного, и трансгенного *Cp190*, а второй пары – только эндогенной мРНК *Cp190*. Было установлено, что удаление спасающей кассеты под воздействием *nanos-Cre* приводило к снижению экспрессии кодирующей последовательности *Cp190* в целых семенниках в 3.8 раза (**Рисунок 4В**). Вместе с тем, уровень экспрессии кодирующей последовательности *Cp190* оставался в 2.4 раз выше по сравнению с контролем дикого типа, что, по всей видимости, объясняется повышенным относительно эндогенного уровнем экспрессии трансгенного *Cp190* в соматических клетках и сохранением спасающей кассеты в единичных сперматоцитах.

Получение мутантных по *Cp190* клеток зародышевого пути было подтверждено иммуноокрашиванием (**Рисунок 4Г**). В трехнедельных семенниках $y^1w^{67}; CP190-mCD8-GFP/nanos-Cre; Cp190^2/Cp190^3$ белок CP190 был обнаружен в ядрах клеток на апикальном конце семенника, где располагаются недифференцированные сперматогонии, а также в ядрах соматических клеток оболочки. Также сигнал наблюдался в единичных ядрах сперматоцитов. В семенниках дикого типа стабильный уровень экспрессии CP190 наблюдался как в соматических, так и в половых клетках. Таким образом, разработанная система обеспечивает спасение мутантного фенотипа *Cp190* в соматических клетках, а также достаточно эффективное удаление спасающей кассеты в клетках мужского зародышевого пути, что на фоне

мутации позволяет воспроизвести мутантный фенотип *Cp190* изолированно в половых клетках взрослых самцов.

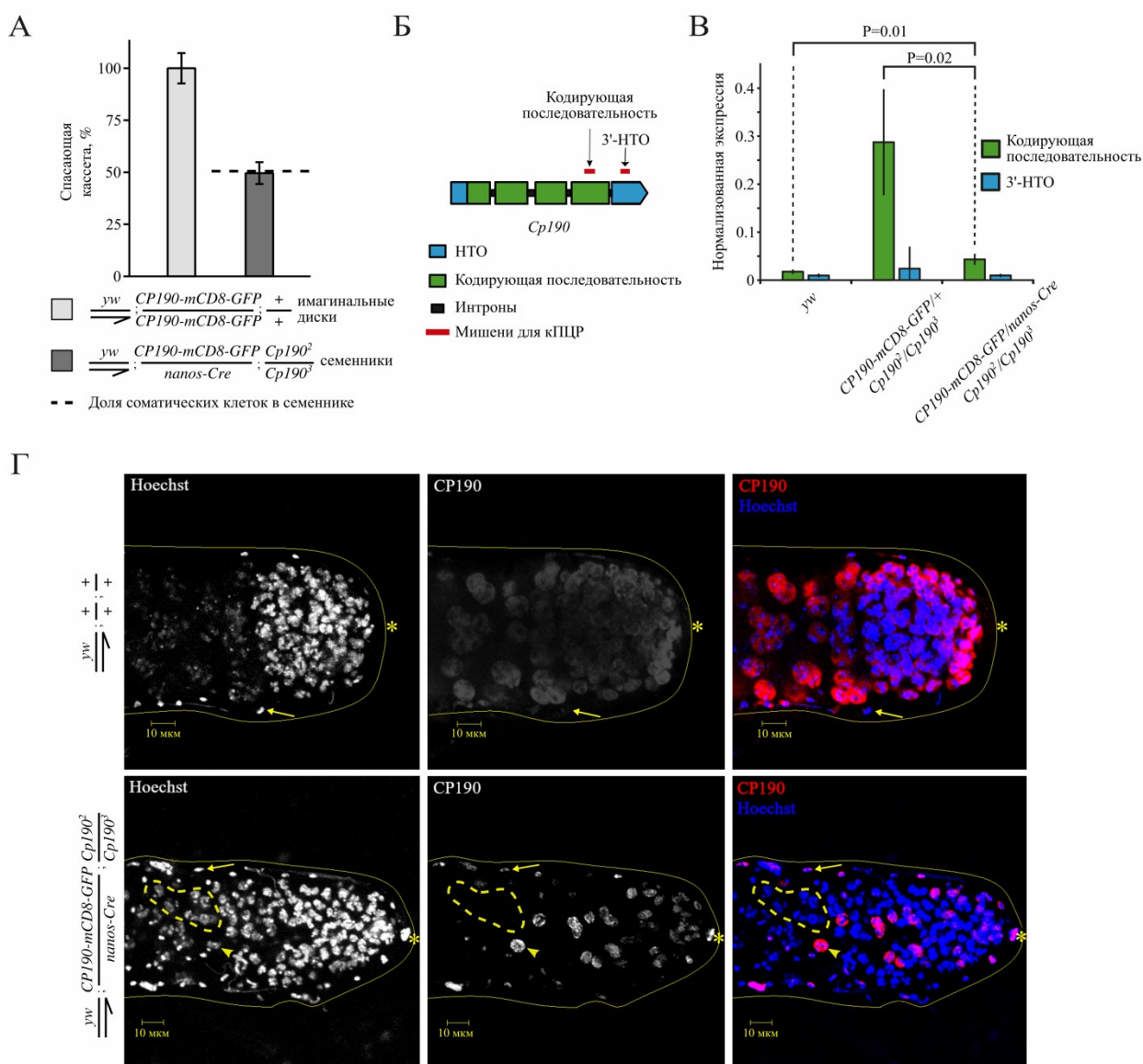


Рисунок 4. Cre-опосредованная рекомбинация приводит к удалению спасающей кассеты и утрате CP190 в клетках мужского зародышевого пути. А) Количество копий спасающей кассеты в семенниках с условным спасением сравнивали с имажинальными дисками личинок, несущих двойную дозу конструкции *CP190-mCD8-GFP* (диаграмма слева). Число копий спасающей кассеты нормировали на дозу конструкции. Пунктирной линией обозначена доля соматических клеток в семенниках дрозофилы по данным секвенирования транскриптома единичных клеток [194]. **Б)** Расположение мишеней для ОТ-кПЦР в гене *Cp190*. Соответствующие праймеры использованы для измерения экспрессии эндогенного и трансгенного *Cp190*. **В)** Нормализованная экспрессия *Cp190* в семенниках трехдневных имаго с конструкциями *CP190-mCD8-GFP* и *nanos-*

Cre. Экспрессию измеряли при помощи праймеров к локусам, обозначенным на Диаграмме Б. Экспрессию нормализовали на значение для гена домашнего хозяйства *Actin42A*. Величина P-value получена с помощью теста Стьюдента с двусторонним распределением. Г) Конфокальные изображения семенников с иммуноокрашиванием антителами анти-CP190. В верхнем ряду – семенник *uw*. В нижнем ряду – семенник с условным спасением. Окраску ДНК проводили при помощи Hoechst. Непрерывной желтой линией подчеркнуты границы семенников. Звездочкой показан апикальный конец семенника. Длинной стрелкой показаны соматические клетки оболочки. Пунктирной линией показано скопление CP190-негативных сперматоцитов. Короткой стрелкой показан CP190-позитивный сперматоцит. Масштаб – 10 мкм.

3.3. Мутация *Cp190* влияет на экспрессию генов дифференцировки в клетках мужского зародышевого пути

Для оценки влияния CP190 на экспрессию генов в семенниках имаго мы провели высокопроизводительное секвенирование транскриптома (RNA-seq) и последующий анализ дифференциальной экспрессии. Важно отметить, что активность спасающей кассеты контролируется промотором гена *Ubi63E*, из-за чего уровень транскрипции *Cp190* превышает эндогенный, что может оказывать влияние на экспрессию генов. Чтобы аккуратно учесть этот эффект в нашем эксперименте, мы оценили, как повышенная экспрессия *Cp190* в соматических клетках влияет на транскриптом целых семенников. В эксперименте мы анализировали трехнедельных самцов, так как в этом возрасте семенники $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;+/+$ и $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;Cp190^2/Cp190^3$ демонстрировали эквивалентный уровень эктопической экспрессии *Cp190* (Рисунок 5).

Мы измерили экспрессию генов в семенниках самцов $y^1w^{67};CP190-mCD8-GFP/nanos-Cre;+/+$, в которых функционирует эндогенный белок CP190, а спасающая кассета активна только в соматических клетках. Для сравнения использовали экспрессию генов в семенниках трехнедельных самцов контрольной линии. Присутствие спасающей конструкции привело к умеренным изменениям транскрипции: 12 генов изменяли экспрессию более

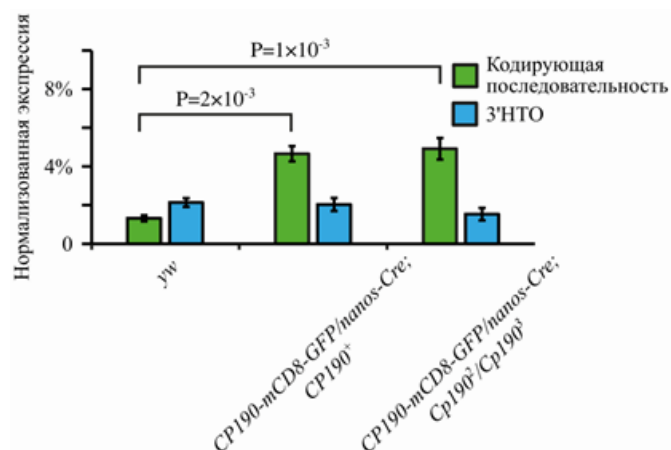


Рисунок 5. Нормализованная экспрессия фрагментов *Cp190* в семенниках трехнедельных имаго с конструкциями *CP190-mCD8-GFP* и *nanos-Cre*. Экспрессию нормализовали на значение для гена домашнего хозяйства *Actin42A*. Значение P-value получено с помощью теста Стьюдента с двусторонним распределением. Условные обозначения расшифрованы на **Рисунке 4Б**.

чем в 8 раз (**Рисунок 6А**). При анализе же семенников с условным спасением мутации *Cp190* в соматических клетках ($y^1w^{67}; CP190-mCD8-GFP/nanos-Cre; Cp190^2/Cp190^3$) было обнаружено 89 генов, чья экспрессия изменялась более чем в 8 раз в сравнении с контролем (**Рисунок 6А**).

В дальнейшем мы считали гены, чей уровень активности достоверно изменился в 2 и более раза, дифференциально экспрессирующимися. В семенниках с условным спасением мутации *Cp190* таких генов было 226 (154 повысили экспрессию, 72 – понизили), при этом транскрипция 43 изменялась и на фоне повышенной экспрессии *Cp190* (34 повысили экспрессию, 9 – понизили) (**Рисунок 6Б**). Эти 43 гена были исключены из дальнейшего анализа, чтобы выделить специфический эффект мутации в гене *Cp190* на экспрессию генов в клетках зародышевого пути.

После этой поправки было обнаружено 120 генов, чья экспрессия усиливалась на фоне мутации *Cp190* (Набор №1), и 63 гена, уровень экспрессии которых снижался (Набор №2) (**Рисунок 6Б**). Чтобы охарактеризовать эти выборки, мы проанализировали профили экспрессии входящих в них генов в разных тканях с использованием транскриптомных данных базы modENCODE

[192]. Оказалось, что среди генов Набора №1 можно выделить три основных кластера, представленных генами, активными в мужской репродуктивной, нервной и пищеварительной системах (**Рисунок 6В**). Гены же Набора №2 экспрессируются повсеместно, и лишь четырнадцать из них можно причислить к семенник-специфичным. Анализ паттернов экспрессии генов в различных тканях дрозофилы с использованием программы TissueEnrich [37] подтвердил обогащение семенник-специфичных генов в Наборе №1, но не в Наборе №2 (**Рисунок 6Г**). Таким образом, белок CP190 в семенниках может участвовать как в поддержании экспрессии повсеместно активных генов, так и в репрессии (или модуляции) активности тканеспецифичных генов, в том числе и генов, участвующих в сперматогенезе.

Вместе с тем, необходимо учитывать, что клеточный состав семенника достаточно разнообразен и представлен несколькими типами соматических клеток, а также клетками зародышевого пути, находящихся на разных стадиях дифференцировки. Различные клеточные типы демонстрируют характерные отличия в профиле экспрессии, что подтверждается данными секвенирования транскриптома единичных клеток семенников [194]. Охарактеризованные профили экспрессии отдельных клеточных популяций позволяют выделить специфичные гены-маркеры, а также установить, в каких клетках активен тот или иной ген в норме.

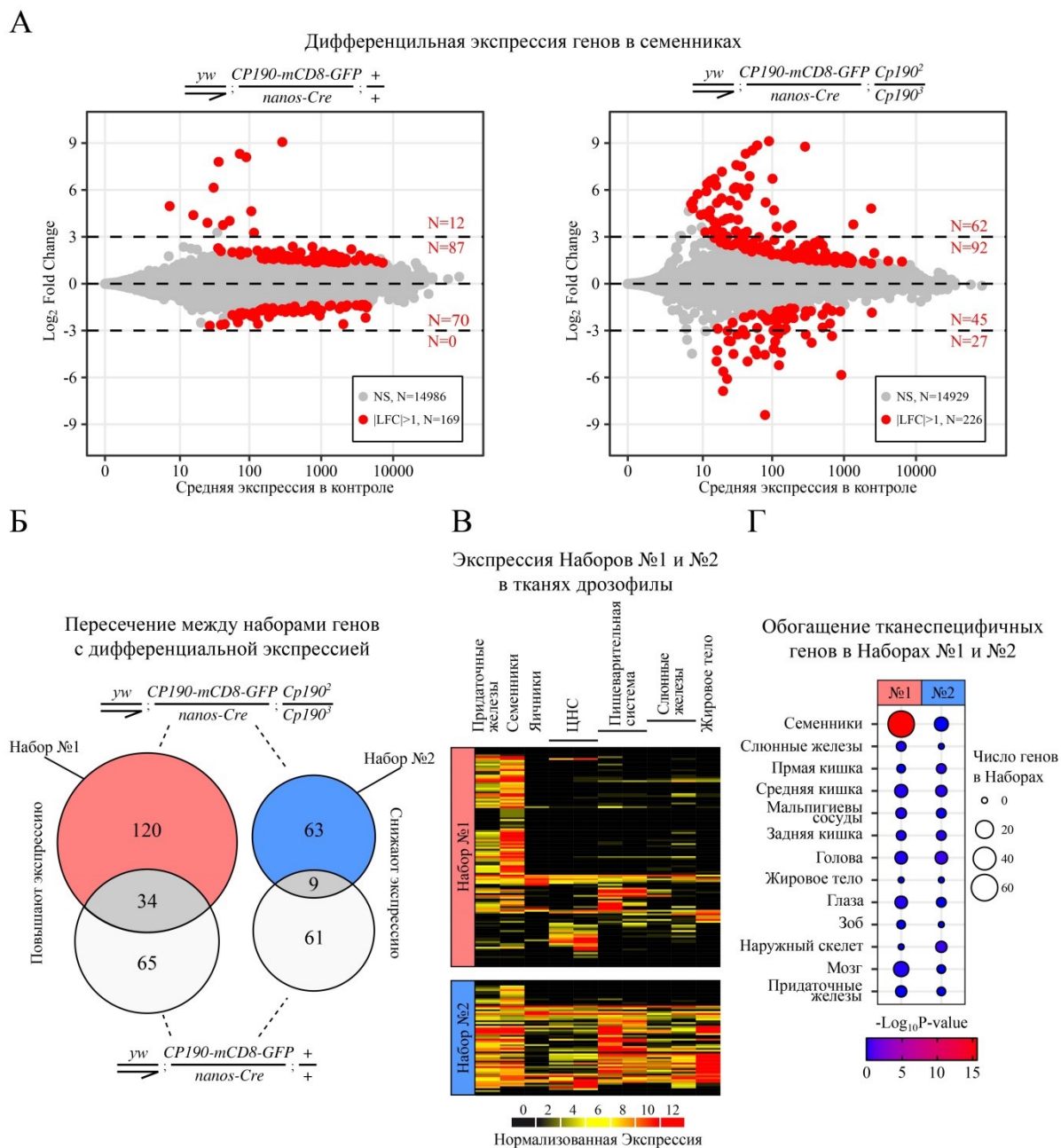


Рисунок 6. Утрата CP190 в клетках мужского зародышевого пути сопровождается активацией семенник-специфичных генов и инактивацией генов домашнего хозяйства. А) Дифференциальная экспрессия генов в семенниках с повышенной экспрессией CP190 в соматических клетках (диаграмма слева) или в семенниках с условным спасения (диаграмма справа) по результатам RNA-seq. По вертикальной оси отложен логарифмированный диапазон изменения экспрессии (Log₂FoldChange, LFC) трансгенных семенников по сравнению с контролем. По горизонтальной оси отложена нормализованная средняя экспрессия в семенниках *yw*. Гены с достоверным изменением экспрессии (|LFC|>1, P<0.05) помечены красным цветом. Значения LFC для генов с недостоверным изменением экспрессии (серые точки, N.S.) были усечены при помощи эмпирического байесовского подхода для наглядности. Пунктирными

линиями указаны значения $LFC \pm 3$ и 0. Красным текстом показано число красных точек со значением LFC в интервале между ближайшими пунктирными линиями. **Б)** Пересечение между множествами генов, достоверно повысившими и понизившими экспрессию в семенниках с условным спасением или семенниках с повышенной экспрессией CP190 в соматических клетках (красные точки на **Диаграмме А**). Гены, изменившие экспрессию только на фоне условного спасения, обозначены как Набор №1 и Набор №2. **В)** Нормализованная экспрессия генов из Наборов №1 и №2 в разных тканях дрозофилы по данным modENCODE [192]. **Г)** Анализ тканеспецифичной экспрессии для генов с CP190-зависимой экспрессией в клетках мужского зародышевого пути.

Для исследования, в каких конкретно клетках семенника в норме активны *Sp190*-зависимые гены, мы использовали гены-маркеры различных клеточных популяций семенников дрозофилы из базы данных DRscDB [194,201]. В Наборе №1 были обнаружены лишь единичные гены-маркеры соматических клеток, тогда как маркеры сперматоцитов в сумме составляли 10% (**Рисунок 7А**). В Наборе №2 22% генов являлись маркерами жировых и пигментных клеток, маркеры же клеток зародышевого пути были единичны (**Рисунок 7А**). Кластеризация выборки генов Набора №1 по уровню экспрессии в разных клеточных популяциях семенника с использованием базы данных транскриптома единичных клеток Fly Cell Atlas [194] позволила установить, что транскрипция 66% генов набора характерна для сперматоцитов (**Рисунок 7Б**). Не менее 60% транскриптов таких генов обеспечивается клетками зародышевого пути, что позволяет отнести большинство генов Набора №1 к генам дифференцировки сперматоцитов (**Рисунок 7Б**). Напротив, лишь пятую часть генов Набора №2 можно идентифицировать как гены дифференцировки сперматоцитов, а их общее количество в 5.7 раз меньше, чем в Наборе №1. Все вместе это позволяет предположить, что CP190 оказывает преимущественно репрессорное воздействие по отношению к генам дифференцировки сперматоцитов.

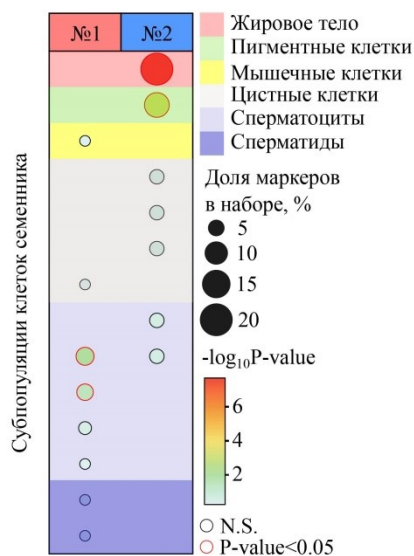
Основная часть генов Набора №2 в норме преимущественно активна в соматических клетках семенников (**Рисунок 7Б**). Вместе с тем, оставался открытым вопрос, изменяется ли уровень экспрессии этих генов в клетках

зародышевого пути на фоне деплеции *Cp190*. С использованием данных о транскриптомах единичных клеток [194], которые позволяют оценить, какой вклад в уровень экспрессии генов вносит отдельная клеточная популяция семенников, мы смоделировали два профиля экспрессии, соответствующих образцам целых семенников, в одном из которых исследуемые гены были инактивированы в клетках зародышевого пути, а в другом – в соматических клетках. На следующем шаге мы сопоставили реальные изменения экспрессии генов Набора №2 с моделями инактивации генов в отдельных клеточных популяциях. Для генов, активных преимущественно в соматических клетках, инактивация только в соматических или только в половых клетках не позволяет смоделировать такое же сильное падение транскрипции, какое мы наблюдали в семенниках с условным спасением ($P < 5 \times 10^{-8}$ в тесте Вилкоксона) (**Рисунок 7В**). Напротив, для генов дифференцировки сперматоцитов оценка на основе гипотезы об инактивации в зародышевом пути не отличается от реальных значений ($P = 0.23$ в тесте Вилкоксона) (**Рисунок 7В**). Такое наблюдение может указывать на то, что падение транскрипции генов Набора №2, характерных для соматических клеток, происходит как в клетках зародышевого пути, так и в соматических клетках.

Интересно отметить, что схожие регуляторные эффекты *CP190* были обнаружены и для культур соматических клеток дрозофилы. С использованием полученных ранее данных транскриптомного анализа, выполненного на клетках Kc167 с РНК-интерференцией *Cp190*, мы установили, что 820 генов повышали экспрессию, тогда как снижали только 260 (**Рисунок 8А**) [21]. Аналогичным образом, активирующиеся гены были представлены преимущественно генами, специфичными для семенников, нервной ткани и пищеварительной системы (**Рисунок 8Б, В**). Гены же, чья экспрессия снижалась на фоне нокдауна *Cp190*, демонстрировали неспецифичный паттерн экспрессии (**Рисунок 8Б, В**).

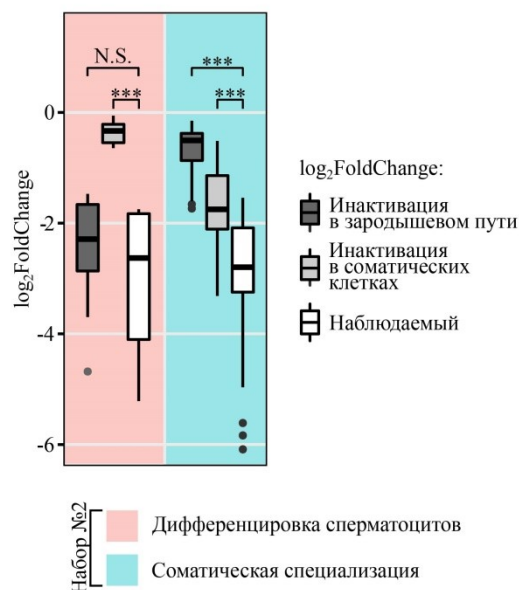
А

Обогащение Наборов №1 и №2
генами-маркерами клеток семенников



В

Оценка изменения экспрессии
генов Набор №2



Б

Экспрессия генов из Наборов №1 и №2 в популяциях клеток семенника

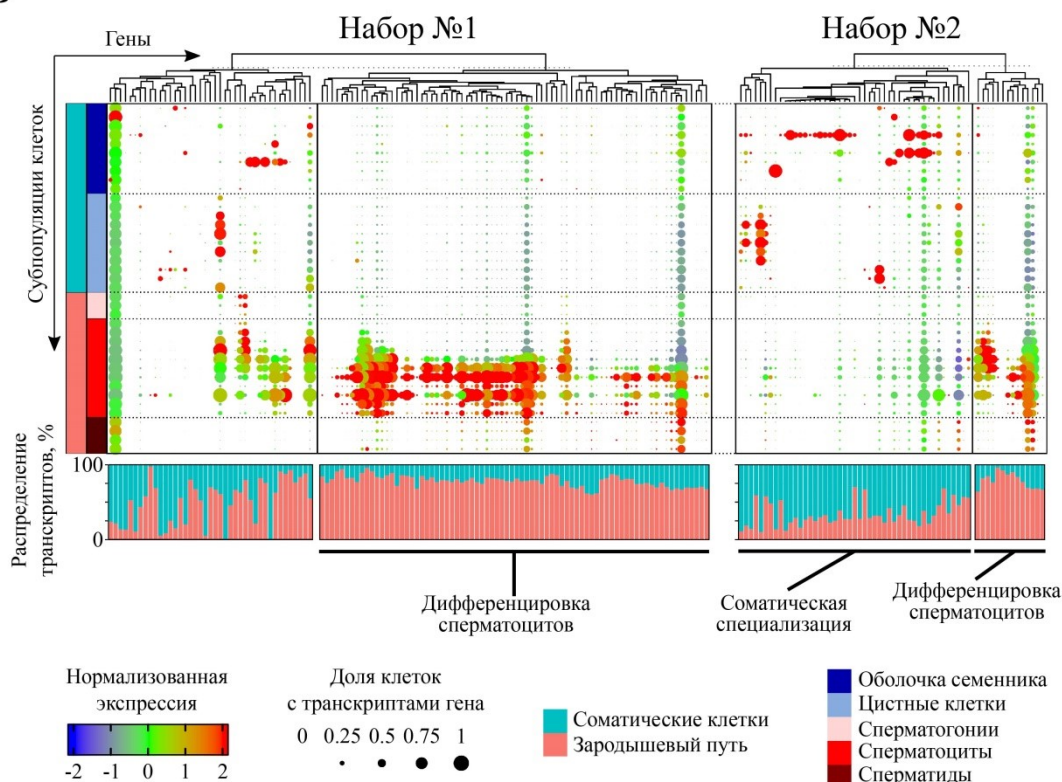


Рисунок 7. Утрата CP190 в клетках мужского зародышевого пути приводит к нарушению активности генов дифференцировки сперматогоний. А) Обогащение генов-маркеров клеточных типов в Наборе №1 и №2 по данным DRscDB [201]. Клеточные типы были аннотированы в эксперименте по секвенированию транскриптома единичных клеток [194] и в некоторых случаях включают несколько субпопуляций с собственным

набором генов-маркеров. Принадлежность субпопуляций клеточным типам обозначена цветными прямоугольниками. Размер кругов характеризует представленность маркеров во всем наборе. Достоверность обогащения вычислена в точном тесте Фишера. Маркеры с достоверным обогащением обведены красным цветом. Б) Экспрессия генов из Наборов №1 и №2 в клеточных популяциях семенника по данным Fly Cell Atlas [194]. Принадлежность популяций клеточным типам и анатомическим структурам в семеннике закодирована цветными прямоугольниками. Цвет точек характеризует уровень экспрессии гена в сравнении со средней экспрессией в клетках семенника. Размер точек описывает долю клеток в популяции, в которых транскрипты данного гена были обнаружены. На столбчатой диаграмме показан вклад клеток зародышевого пути и соматических клеток в общую транскрипцию каждого гена в целом семеннике. При помощи иерархической кластеризации по уровням экспрессии в популяциях клеток гены в каждом наборе были разделены на два кластера. В Наборе №1 отчетливо выделяются гены с высокой активностью в сперматоцитах (гены дифференцировки сперматоцитов). В Наборе №2 кластеры можно охарактеризовать как гены соматической специализации и гены дифференцировки сперматоцитов. Д) Сравнение оценок изменения экспрессии для двух кластеров из Набора 2. Оценки $\text{Log}_2\text{FoldChange}$ были получены отдельно для генов дифференцировки сперматоцитов (красный цвет) и генов соматической специализации (синий цвет) из предположения, что в семенниках с условным спасением инактивация произошла только в клетках зародышевого пути (темно-серые прямоугольники) или только в соматических клетках (светло-серые прямоугольники). Наблюдаемые значения $\text{Log}_2\text{FoldChange}$ показаны белыми прямоугольниками. Условные обозначения: *** – $P < 5 \times 10^{-8}$ в тесте Вилкоксона, N.S. – $P > 0.05$.

3.4. Мутация *Sp190* нарушает экспрессию генов-мишеней tMAC

Активация генов дифференцировки происходит скоординировано в фазе роста сперматоцитов первого порядка, и описано несколько мутаций, которые нарушают этот процесс [3]. Так, мутация гена *bam* вызывает остановку сперматогенеза на стадии сперматогониев, в результате чего семенники оказываются обогащены слабо дифференцированными клетками зародышевого пути [32]. Гены *mir40*, *comr* и *cap* являются генами задержки мейоза, кодируют транскрипционные факторы, а их мутации приводят к остановке сперматогенеза на стадии первичных сперматоцитов [3].

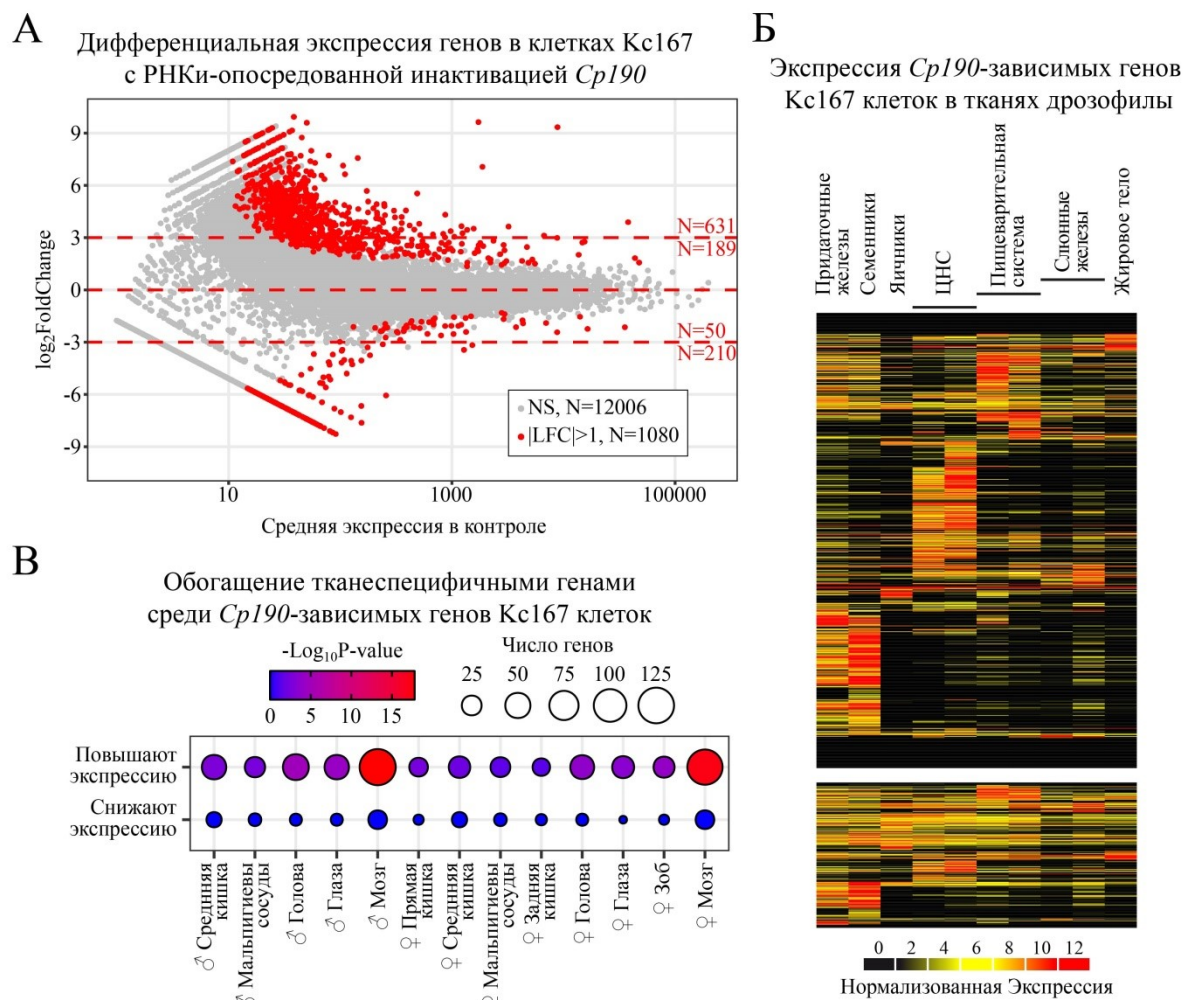


Рисунок 8. CP190 подавляет активность тканеспецифичных генов и способствует активации генов домашнего хозяйства в клетках Kc167. А) Эффект инактивации *Cp190* в клетках Kc167 согласно данным эксперимента [21]. Анализ дифференциальной экспрессии проводили тем же способом, что и для **Рисунка 6А**. Б) Экспрессия CP190-зависимых генов (красные точки на диаграмме А) в тканях дрозофилы согласно данным modENCODE [192]. В) Анализ тканеспецифичной экспрессии для CP190-зависимых генов Kc167-клеток.

Нарушения в работе Mir40 и Comr, которые входят в состав семенник-специфичного комплекса задержки мейоза tMAC, вызывают драматичные изменения экспрессии генов в семенниках [3,11]. Нарушения же в работе белков из группы семенник-специфичных факторов транскрипции tTAF, включая Can, вызывают более мягкие изменения в транскриптом семенников. Вместе tTAF и tMAC обеспечивают активацию программы

дифференцировки сперматоцитов, причем только часть генов регулируется этими факторами напрямую [11].

Учитывая выявленную связь CP190 с генами дифференцировки, мы предположили функциональное взаимодействие между CP190, tTAF и tMAC. Ранее нами были получены профили экспрессии генов в семенниках мутантов *bat*, *mir40*, *comr* и *cap*, а также прокартированы сайты связывания Mir40, Comr и Cap с хромосомами в клетках зародышевого пути трехдневных самцов [11]. С помощью этих данных мы проанализировали экспрессию генов, которые активируются в семенниках с условным спасением (Набор №1) в семенниках дикого типа и мутантов *bat*, *mir40*, *comr* и *cap* (**Рисунок 9А**).

Гены дифференцировки сперматоцитов из Набора №1 (**Рисунок 9Б**) значительно снижают активность у мутантов *bat*, *mir40* и *comr*, но не у мутантов *cap*. Остальные гены из Набора №1 почти не реагируют на мутации *mir40*, *comr* и *cap*, и только мутация *bat* приводит к незначительному падению их экспрессии, что можно объяснить присутствием среди них генов с высокой активностью в сперматоцитах (**Рисунок 9Б**). Таким образом, CP190 ограничивает активность tMAC-зависимых генов дифференцировки сперматоцитов.

Профили распределения Comr, Mir40 и Cap на хромосомах показывают, что эти белки чаще ожидаемого связываются с генами из Набора №1, причем наиболее интенсивное связывание наблюдается для транскрипционного фактора Comr (**Рисунок 9Б**). Таким образом, мутация *Cp190* затрагивает активность генов, непосредственно регулируемых комплексом tMAC.

Для экспрессии генов Набора №2 характерны более мягкие отличия между мутантами *bat*, *mir40*, *comr* и *cap*, и диким типом (**Рисунок 9В**), хотя и гены дифференцировки сперматоцитов, и гены, характерные для соматических клеток, снижают экспрессию как у мутантов с нарушениями в tMAC, так и у мутантов *bat*. Гены из Набора №2 демонстрируют некоторую

связь с транскрипционным фактором *Comr* (Рисунок 9Г). Это указывает на то, что их экспрессия может напрямую зависеть от взаимодействия между tMAC и CP190.

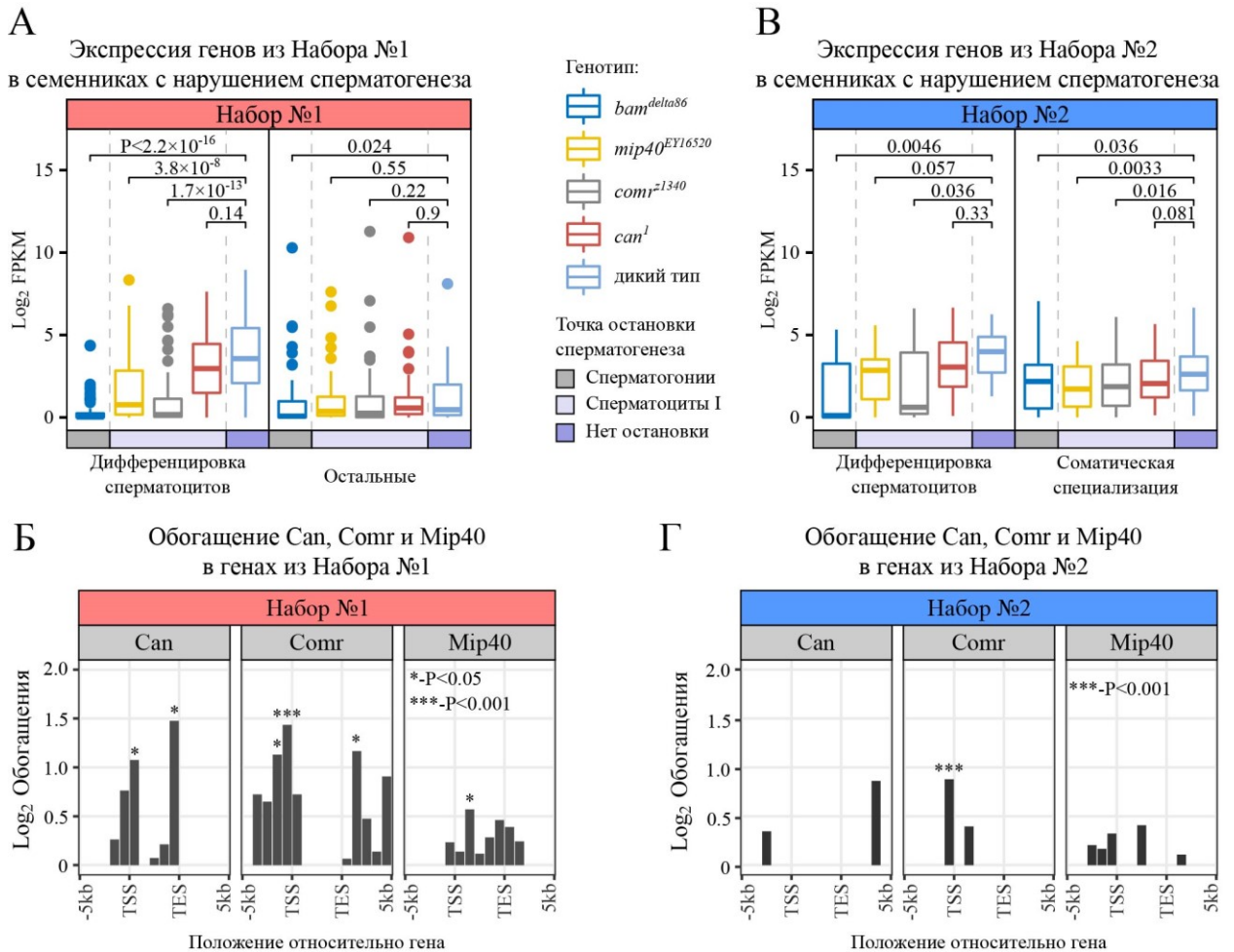


Рисунок 9. Деплеция CP190 в клетках мужского зародышевого пути приводит к нарушению активности tMAC-зависимых генов дифференцировки сперматоцитов. А) Экспрессия генов из двух кластеров Набора №1 (см. Рисунок 7В) в семенниках дикого типа и мутантов с нарушением дифференцировки клеток мужского зародышевого пути. Цветными прямоугольниками показано, на каких стадиях сперматогенеза останавливается дифференцировка клеток в семенниках каждого генотипа. Гены *mir40* и *comr* кодируют компоненты семенник-специфичного транскрипционного активатора tMAC. Ген *can* кодирует один из белков tTAF. Достоверность отличий между мутантами и диким типом получена в тесте Вилкоксона. Б) Обогащение Набора №1 генами, вокруг которых в хромосомах клеток зародышевого пути самцов в разных участках связываются транскрипционные факторы *Can*, *Comr* и *Mir40* [11]. Достоверность обогащения измерена с помощью точного теста Фишера. Условные обозначения: TSS – сайт начала транскрипции; TES – сайт

завершения транскрипции. В) То же, что в А, для генов из двух кластеров Набора №2. Г) То же, что в Б, для генов из Набора №2.

3.5. Дифференцировка клеток сопровождается динамичным связыванием CP190 с хроматином

Известно, что CP190 регулирует активность множества генов через взаимодействие с рядом транскрипционных факторов. Поэтому мы хотели выяснить, связывается ли CP190 с регуляторными элементами генома, ассоциированными с tMAC. С другой стороны, дифференцировка сперматоцитов дрозофилы сопровождается масштабными изменениями в структуре хроматина и активности генов [3]. Однако неизвестно, отражаются ли перестройки хроматина в клетках на связывании инсуляторных белков с хромосомами. С целью понять это мы использовали метод тканеспецифичного DamID-seq [12], чтобы прокартировать CP190 в клетках зародышевого пути в семенниках дикого типа, а также у мутантов *cap*, *tip40* и *bat*. Как уже упоминалось, мутации *cap* и *tip40* приводят к задержке мейоза из-за нарушений функции tTAF и tMAC, в результате чего семенники переполняются недифференцированными сперматоцитами. У мутантов *bat* дифференцировка клеток мужского зародышевого пути останавливается до формирования сперматоцитов, и семенники переполняются пролиферирующими сперматогониями (**Рисунок 1**). Таким образом, DamID-seq CP190 клеток мужского зародышевого пути мутантов позволил нам охарактеризовать связывание CP190 с хромосомами на стадии сперматогоний и недифференцированных сперматоцитов. Чтобы оценить общие закономерности распределения CP190, мы также прокартировали этот белок в терминально-дифференцированных соматических клетках личиночных слюнных желез. В этой ткани tTAF и tMAC не функционируют, а гены дифференцировки сперматоцитов подвержены репрессии.

Полученные профили очень похожи визуально, что продемонстрировано на **Рисунке 10А**, тем не менее, отличия в распределении CP190 имеют место. Чтобы охарактеризовать различия, мы в первую очередь проанализировали распределение GATC-фрагментов, достоверно связанных с CP190, по функциональным районам генома (**Рисунок 10Б**). Во всех случаях CP190 демонстрирует тенденцию к ассоциации с промоторными областями генов, так как от 38% до 57% всех фрагментов локализованы вблизи TSS (-1000:+300 п.н.). Это достоверно выше случайно-ожидаемого значения в 23% ($P < 0.05$), поэтому мы сконцентрировались на генах с которыми CP190 связывается в промоторной области.

Далее мы сравнили наборы генов, ассоциированные с CP190 в клетках мужского зародышевого пути и слюнных железах (**Рисунок 10В**). Оказалось, что только 47% генов, для которых мы обнаружили связывание CP190 в зародышевом пути, ассоциированы с ним на политенных хромосомах. В то же время, только 1275 генов оказались связаны с CP190 во всех профилях, полученных для клеток мужского зародышевого пути, что составляет 30% от общего числа генов, ассоциированных с CP190 на профиле дикого типа (4136 генов) (**Рисунок 10В**). Таким образом, набор генов, ассоциированных с CP190 в промоторной области, значительно отличается между типами клеток. Это говорит о том, что в ходе дифференцировки связывание CP190 с хромосомами меняется более значительно, чем было известно ранее. Так, при сравнении клеток Kc167 и Mbn профили связывания CP190 пересекаются не менее чем на 85% [108].

В клеточных культурах CP190 связывается с промоторами активных генов [17], поэтому мы предположили, что причина обнаруженных нами различий заключается в разной активности генов, с которыми связан этот белок в конкретных клетках. Чтобы проверить это, мы проанализировали доступные транскриптомные данные, полученные для личиночных слюнных желез, семенников дикого типа, а также мутантов *cap*, *mir40* и *bat* [11,202].

В каждом случае мы разделили все гены с ненулевой транскрипцией по квартилям распределения величины FPKM, которая отражает уровень

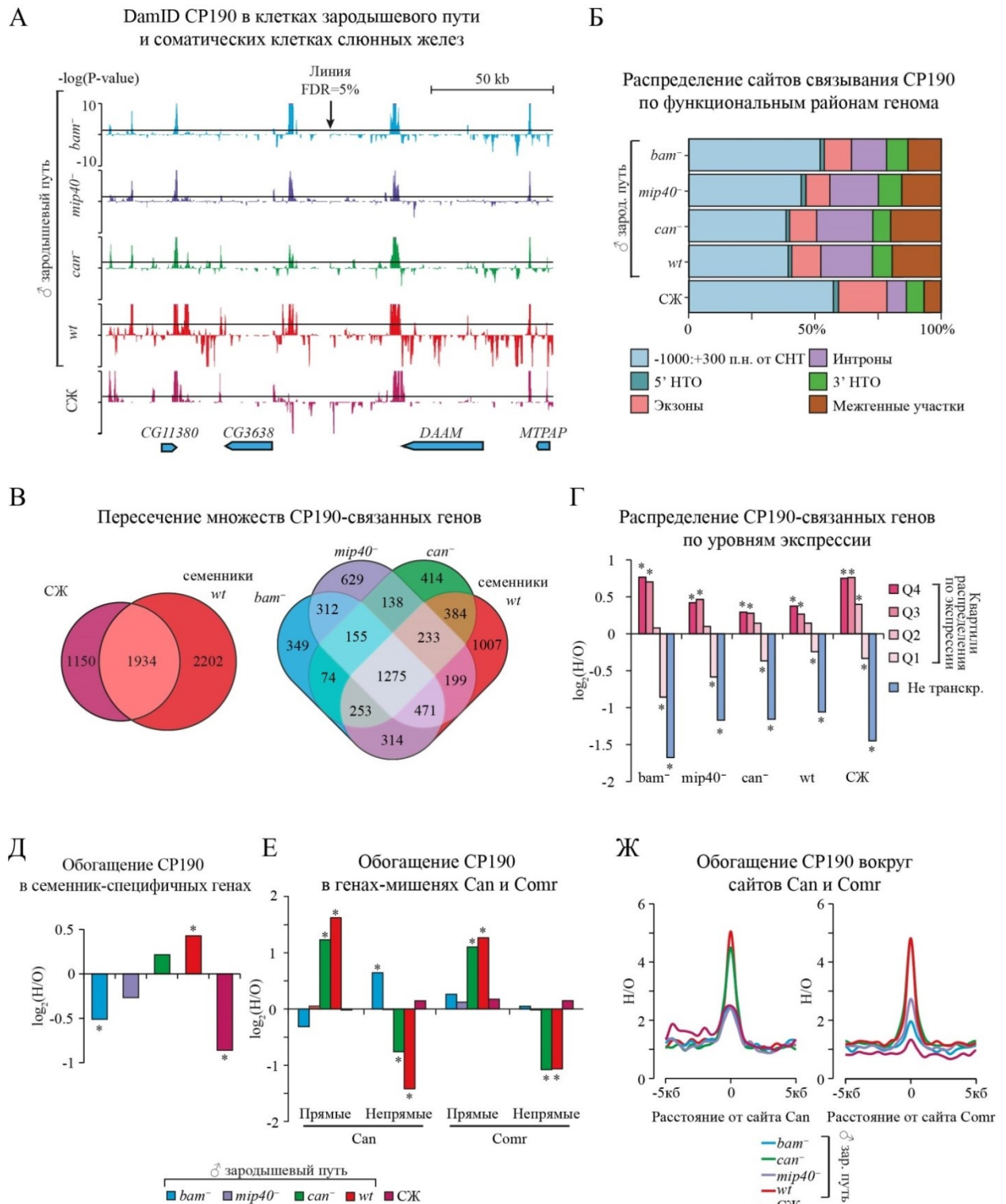


Рисунок 10. В ходе дифференцировки клеток мужского зародышевого пути CP190 перераспределяется к генам дифференцировки сперматоцитов. А) DamID-профили связывания белка CP190 в клетках мужского зародышевого пути дикого типа (*wt*) и мутантов

по генам *cap*, *mir40* и *bat*, а также в политенных хромосомах личиночных слюнных желез (СЖ). По вертикальной оси отложена достоверность связывания CP190 с GATC-фрагментами генома. Горизонтальной линией показан уровень достоверности, при котором вероятность ложноположительного открытия (FDR) составляет 5%. Пики на профиле, превышающие эту линию, считаются достоверными. **Б)** Достоверные сайты связывания CP190 были распределены по категориям в зависимости от положения в функциональных участках генов. **В)** Наборы генов, ассоциированные с CP190 в промоторной области, были пересечены на диаграмме Эйлера. **Г)** Гены с недетектируемой экспрессией помещались в категорию «не транскрибируются». Остальные гены были разделены на квартили по уровню транскрипции. В ряду Q1, Q2, Q3, Q4 экспрессия увеличивается. Наблюдаемое число генов из каждой категории (Н) сравнивали с ожидаемым (О) при помощи теста Хи-квадрат. **Д)** Наблюдаемое (Н) число семенник-специфичных генов, связанных с CP190 в каждом профиле, было сравнено с ожидаемым (О) при помощи теста Хи-квадрат. **Е)** Гены-мишени *Cap* и *Comr* были определены как гены, которые инактивируются в семенниках у мутантов *cap* и *comr*. Гены, которые одновременно с этим связываются с *Cap* или *Comr* согласно DamID, названы прямыми мишенями. Для диаграммы наблюдаемое число генов из каждой группы (Н) сравнивали с ожидаемым числом (О) среди всех генов, ассоциированных с CP190 согласно DamID. Достоверность обогащения измеряли в тесте Хи-квадрат. **Ж)** Сайты связывания *Cap* и *Comr* в хромосомах клеток мужского зародышевого пути были определены с помощью тканеспецифичного DamID. Область генома вокруг каждого сайта была разделена на участки, и общее число сайтов связывания CP190 в каждом участке было просуммировано. Наблюдаемое число сайтов связывания CP190 (Н) сравнивали с ожидаемым (О), полученным исходя из случайного распределения CP190. Условные обозначения: * – P-value<0.001.

экспрессии гена. В этом анализе гены с самой сильной экспрессией попадали в квартиль Q4, с самой низкой экспрессией – в Q1. Это позволило обнаружить, что в каждом типе клеток CP190 действительно связывается с наиболее активными генами (**Рисунок 10Д**). Таким образом, наблюдаемое нами динамичное связывание CP190 с хромосомами вызвано дифференциальной активностью генов.

Далее мы проанализировали, как меняется связывание CP190 с группой семенник-специфичных генов (**Рисунок 10Е**). Мы снова обнаружили достоверное обогащение семенник-специфичными генами только в семенниках дикого типа, где эти гены активно транскрибируются. Таким

образом, в ходе дифференцировки клеток мужского зародышевого пути CP190 перераспределяется в промоторы семенник-специфичных генов.

Ранее мы показали, что CP190 необходим для поддержания правильного уровня активности генов дифференцировки сперматоцитов, ассоциированных с tMAC (См. §3.4). Поэтому мы проверили, с какой долей прямых и непрямых мишеней tTAF и tMAC белок CP190 ассоциирован в разных профилях (**Рисунок 10Ж**). Показанное ранее обогащение CP190 в промоторах прямых мишеней и обеднение в непрямых мишенях *Can* (tTAF) и *Comr* (tMAC) выявляется только в клетках зародышевого пути дикого типа и мутантов *can*. Таким образом, в ходе дифференцировки сперматоцитов CP190 перераспределяется к промоторам генов, активируемых транскрипционными факторами tTAF и tMAC, и белок Mir40 важен для перераспределения. Мы предположили, что показанная ранее колокализация CP190 с сайтами связывания *Can* и *Comr* также будет наблюдаться лишь в клетках мужского зародышевого пути дикого типа и мутантов *can* (**Рисунок 10З-И**). Как и ожидалось, колокализация CP190 с белками *Can* и *Comr* пропадает в семенниках у мутантов *bat* и *mir40* и не наблюдается в слюнных железах.

Белок Mir40 входит в состав и tMAC и dREAM, и эффект мутации в гене *mir40* фенотипически схож с мутациями, затрагивающими tTAF [4]. Поэтому не ясно, приводит ли мутация *mir40* к потере функции tMAC. Чтобы убедиться, что связывание tMAC с хроматином нарушается в отсутствие белка Mir40, мы прокартировали сайты связывания белка *Comr* в клетках мужского зародышевого пути у мутантов *mir40* с помощью DamID. В результате мы получили профиль, в котором только 63 GATC-фрагмента в геноме дрозофилы оказались достоверно связаны с *Comr*. Участок полученного профиля показан на **Рисунке 11**. В клетках мужского зародышевого пути дикого типа DamID профиль *Comr* содержит более пяти тысяч достоверных сайтов связывания. Мы проверили, совпадают ли сайты связывания *Comr* у мутантов *mir40*, с пиками этого же белка в диком типе.

Мы установили, что из 63 сайтов связывания в мутантах только 8 сохраняются в диком типе. Мы дополнительно проанализировали выборку из 200 наиболее достоверных сайтов связывания Comr у мутантов *mir40*, при этом количество пересечений увеличилось только до 19. Таким образом, в отсутствие функционального белка Mir40 транскрипционный фактор Comr не способен стабильно связываться с хромосомами клеток мужского зародышевого пути. В связи с этим мы приходим к выводу, что роль белка Mir40 в привлечении CP190 к семенник-специфичным генам в сперматоцитах опосредована комплексом tMAC.

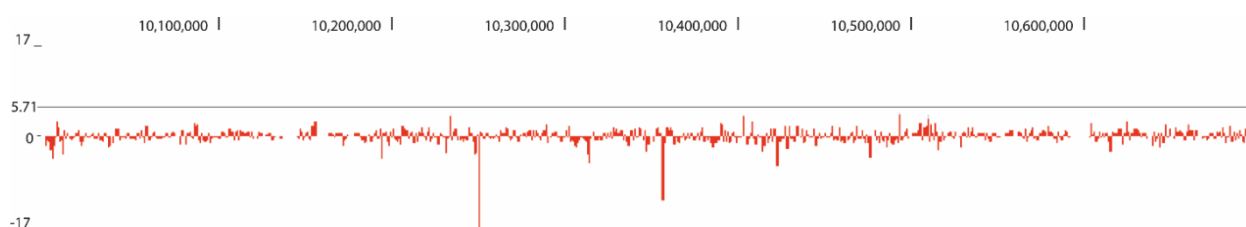


Рисунок 11. DamID-профиль связывания белка Comr в клетках мужского зародышевого пути мутантов *mir40*. Приведен участок хромосомы 2L.

3.6. Активная транскрипция и SetDB1, но не связывание CP190, маркирует границы доменов H3K27me3

CP190 является одним из ключевых инсуляторных белков дрозофилы, и неоднократно высказывалось предположение, что он может устанавливать барьеры на границах доменов H3K27me3 [17,18]. Эта эпигенетическая модификация появляется благодаря действию репрессоров семейства Polycomb и связана с инактивацией генов в факультативном гетерохроматине. С использованием полученных нами данных мы хотели определить, участвует ли CP190 в поддержании границ доменов H3K27me3 в терминально-дифференцированных клетках дрозофилы. Мы проверили связь между CP190 и доменной организацией хроматина в политенных хромосомах личиночных слюнных желез. Для этого мы объединили полученный нами

DamID профиль CP190 в политенных хромосомах с опубликованными ранее экспериментами.

Мы использовали данные по распределению метки H3K27me3 в политенных хромосомах личиночных слюнных желез, полученные методом ChIP-chip [183], и воспользовались алгоритмом скрытых марковских моделей, чтобы определить точные границы доменов репрессии Polycomb. В результате мы выявили 387 доменов. Чтобы определить, есть ли связь между CP190 и границами доменов, мы проанализировали обогащение CP190 вокруг границ. Мы не выявили никакой связи между этим инсультаторным белком и границами (**Рисунок 12А**). Поэтому мы провели поиск факторов, которые могут служить более надежными маркерами границ и, к нашему удивлению, гистон-метилтрансфераза SetDB1 демонстрировала явную предрасположенность к связыванию на границах доменов репрессии Polycomb.

При более детальном рассмотрении, DamID-профиль белка CP190 содержал пики как на границах доменов, так и внутри. Например, обширный домен репрессии Polycomb в кластере гомеозисных генов *Antennapedia* содержал несколько сайтов связывания CP190, причем этот белок локализовался в инсультаторах I класса, которые были предсказаны ранее [203] (**Рисунок 12Б**). Напротив, SetDB1 локализовался на границах домена, и только один пик находился в локальном «провале» метилирования внутри этого домена. Мы также обратили внимание, что сайты связывания SetDB1 на границе доменов или в «провалах» метилирования совпадают с промоторами высокоактивных генов (**Рисунок 12Б,В**). Напротив, те домены, которые не были окружены пиками SetDB1, зачастую фланкировались генами с низкой активностью. Например, в локусе *dar1-Awh*, только домен *Awh* фланкирован пиками CP190, SetDB1 и генами с высокой транскрипционной активностью (**Рисунок 12В**).

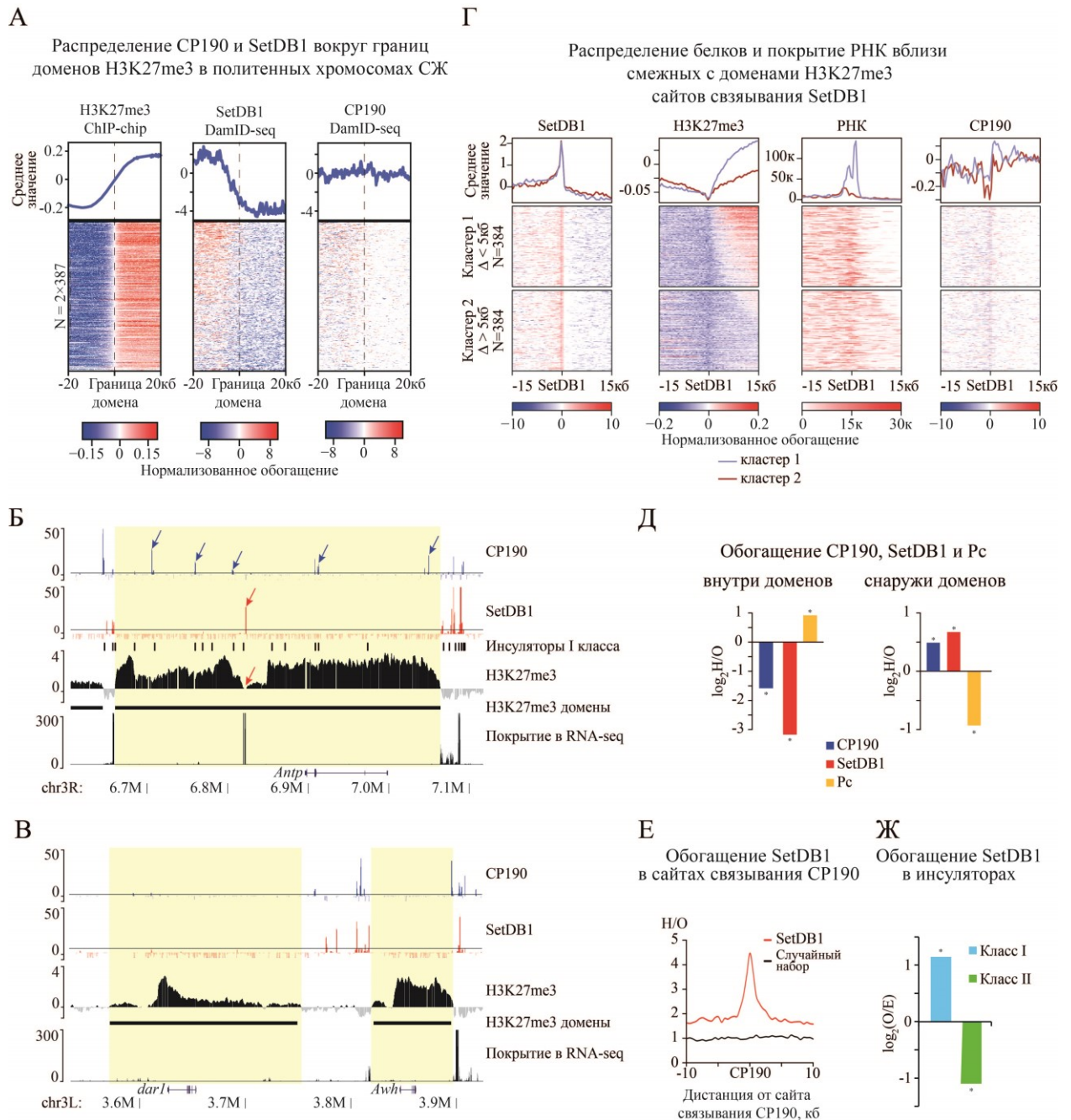


Рисунок 12. CP190 не демонстрирует связи с границами доменов H3K27me3. А) SetDB1, но не CP190 обогащен на границах доменов H3K27me3 в политенных хромосомах личиночных слюнных желез. Положения границ определяли с помощью скрытой марковской модели. На тепловой карте показано обогащение локусов вокруг установленных границ доменов заданным белком. Границы доменов являются точкой отсчета, координаты нуклеотидов вне домена отсчитывается со знаком минус, а внутри домена со знаком плюс. Б) Распределение CP190, SetDB1 и прочтений RNA-seq в локусе гомеозисных генов *Antp*. CP190 обнаруживается как на границах, так и внутри доменов репрессии Polycomb, тогда как SetDB1 маркирует области с высокой транскрипцией на границах доменов. Домен K27me3 помечен желтым прямоугольником. Расположение инсуляторных элементов I класса (связываются с CP190/BEAF-32/dCTCF) было предсказано

в работе [22]. Синими стрелками показаны сайты связывания CP190 внутри домена. Красными стрелками показан сайт связывания SetDB1 внутри проаннотированного домена, где наблюдается «провал» в уровне метилирования и высокий уровень транскрипции. **В)** Распределение CP190, SetDB1 и прочтений RNA-seq в локусе *dar1-Awh*. Из двух доменов H3K27me3 только один ограничен сайтами связывания CP190, SetDB1 и активной транскрипцией. **Г)** Границы доменов H3K27me3 были отсортированы по дистанции до ближайшего сайта связывания SetDB1, и разделены на две группы по дистанции от SetDB1 до границы домена (Δ). Для каждого пограничного сайта SetDB1 было определено обогащение SetDB1, CP190, H3K27me3, и покрытие прочтениями в RNA-seq. Сайт связывания SetDB1 взят за точку отсчета и ориентированы таким образом, чтобы домены H3K27me3 находились справа. **Д)** Сайты связывания CP190 и SetDB1 обеднены внутри доменов H3K27me3. Для сравнения показано обогащение Polysomb. Для диаграмм сравнивали наблюдаемое число сайтов каждого белка (Н) внутри или вне доменов со случайно-ожидаемым (О). **Е)** SetDB1 колокализуется с CP190 в политенных хромосомах слюнных желез. Для анализа районы вокруг всех сайтов связывания CP190 были разделены на фрагменты 500 п.н. и общее число пиков SetDB1 в равноудаленных фрагментах суммировалось (Н), а также сравнивалось со случайно-ожидаемым (О). Для сравнения показано обогащение для случайного набора геномных локусов. **Ж)** Обогащение SetDB1 в инсуляторах первого класса (связываются с CP190) и второго класса (Su(Hw)-связанные). Сайты SetDB1 считались связанными с инсулятором, если находились на расстоянии не более 500 п.н. от них. Расположение инсуляторов было предсказано в [204]. Условные обозначения: * - P-value<0.001 в тесте Хи-квадрат.

Мы предположили, что высокая транскрипция SetDB1-связанных генов служит лучшим маркером границ доменов H3K27me3 и, чтобы убедиться в этом, нашли ближайшие к границам сайты связывания SetDB1, после чего проанализировали распределение метилирования, CP190 и транскрипционной активности вокруг этих сайтов (**Рисунок 12Г**). Оказалось, что половина всех границ располагается на расстоянии менее 5 кб от SetDB1, и уровень транскрипции вблизи таких сайтов многократно превышает аналогичный показатель в удаленных от H3K27me3 сайтах SetDB1. В свою очередь, CP190 демонстрировал лишь небольшое обогащение вблизи сайтов SetDB1, расположенных рядом с доменами.

Таким образом, SetDB1 не является обязательным фактором, образующим границы доменов H3K27me3, но в комбинации с активной транскрипцией он присутствует примерно на половине из них. При этом, именно на границе можно видеть и максимум транскрипции, и максимум связывания SetDB1, и оба эти параметра быстро убывают по мере удаления от домена H3K27me3, чего нельзя сказать о CP190, который не имеет определенного тренда распределения.

Сравнение профилей SetDB1 и CP190 позволило установить общие закономерности в распределении двух белков на хромосомах. Так, нахождение SetDB1 вне доменов H3K27me3 аналогично распределению инсуляторного белка CP190 (**Рисунок 12Д**). Анализ колокализации пиков показал, что SetDB1 значительно чаще ожидаемого совпадает с CP190 (**Рисунок 12Е**). Кроме этого, мы оценили, насколько сайты связывания SetDB1 совпадают с инсуляторами двух разных классов, найденных в геноме дрозофилы [22]. Оказалось, что в ближайшей окрестности инсуляторов класса I (CP190/BEAF-32/dCTCF) SetDB1 связывается в 2.2 раза чаще ожидаемого, а возле инсуляторов класса II (Su(Hw)) – в 2.1 раза реже ожидаемого (**Рисунок 12Ж**). Это указывает на возможную связь сайтов SetDB1 с разделением хроматина на физические домены. С другой стороны, SetDB1 потенциально может оказаться еще одним примером эпигенетического регулятора, чья активность связана с инсуляторными белками.

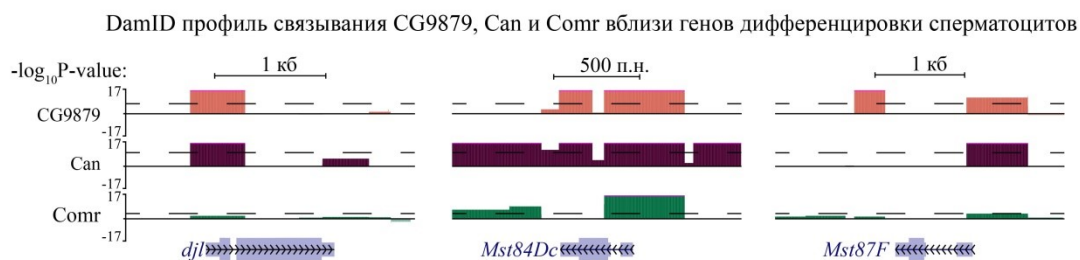
Таким образом, в политенных хромосомах слюнных желез CP190 плохо коррелирует с границами доменов H3K27me3. Учитывая, что в культурах клеток до половины пограничных элементов таких доменов ассоциированы с CP190, это может означать, что в ходе дифференцировки CP190 утрачивает связь с границами доменов репрессии Polycomb. Наоборот, SetDB1 вместе с активной транскрипцией служат более надежными маркерами пограничных элементов в этих доменах, что указывает на его роль в поддержании доменной организации хроматина в клетках слюнных желез.

3.7. Роль гомолога ТВР CG9879 в регуляции генов дифференцировки мужского зародышевого пути

Транскрипционные факторы tTAF и tMAC разворачивают в сперматоцитах сложную генетическую программу, результатом которой является активация более тысячи семенник-специфичных генов, необходимых для всех последующих стадий сперматогенеза. Вместе с тем, большое число этих генов не связываются с двумя транскрипционными активаторами напрямую [11,12]. Большое число не прямых мишеней tTAF и tMAC может объясняться существованием транскрипционных факторов, чья экспрессия в сперматоцитах индуцируется под действием этих комплексов. Поиск среди мишеней tMAC генов, ассоциированных с регуляцией транскрипции, позволил выявить, что только CG9879, гомологичный *Tbr*, теряет активность у мутантов с задержкой мейоза и при этом регулируется с помощью tMAC напрямую [12]. Мы предположили, что CG9879 служит промежуточным звеном в генетическом каскаде, контролирующем семенник-специфичные гены в сперматоцитах, и исследовали его функцию в сперматогенезе.

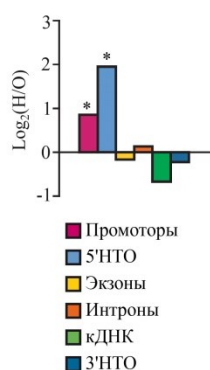
Мы использовали тканеспецифичный DamID-seq, чтобы прокартировать сайты связывания CG9879 на хромосомах в клетках мужского зародышевого пути. Полученный профиль продемонстрировал, что этот белок имеет тенденцию связываться с 5'-нетранслируемыми и промоторными областями семенник-специфичных генов (биномиальный тест, $P < 0.001$, **Рисунок 13А-Б**). Кроме того, CG9879 обнаруживается преимущественно в проксимальной области промоторов (**Рисунок 13А,В**), что является характерной особенностью ТВР и других базальных факторов транскрипции [205].

А



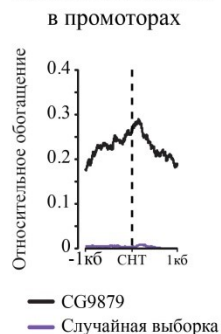
Б

Обогащение CG9879
в функциональных
районах генома



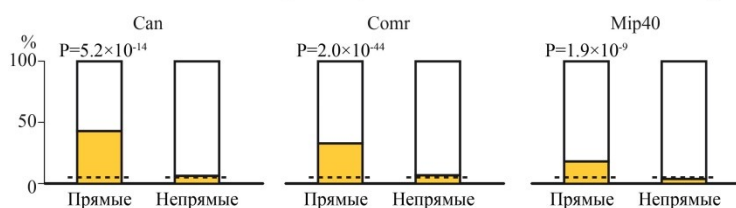
В

Распределение сайтов
связывания CG9879
в промоторах



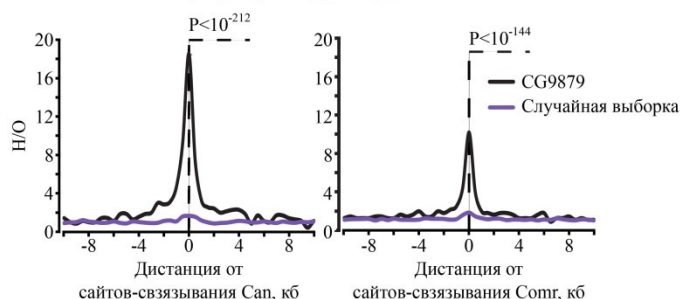
Г

Обогащение CG9879 в промоторах генов-мишеней Can, Comr и Mir40



Д

Обогащение CG9879 вблизи сайтов
связывания Can и Comr



Е

Топ 3 мотивов внутри сайтов связывания CG9879

Мотив	Лого	P-value
1. AAATWATH		9.5×10^{-118}
2. AAMAHAA		1.4×10^{-94}
3. ANAAAT		1.9×10^{-77}

Мотив связывания в
CG9879-ассоциированных промоторах

Мотив	Лого	P-value
1. DAAATAWA		1.4×10^{-20}

Мотив связывания ТВР



Рисунок 13. CG9879 колокализуется с белками Can и Comr в клетках мужского зародышевого пути. А) CG9879 колокализуется с Can, Comr в промоторах генов дифференцировки сперматозоидов. На диаграмме показаны фрагменты DamID-профилей белков CG9879, Can, Comr вблизи генов дифференцировки сперматозоидов *djl*, *Mst84Dc*, *Mst87F*. По вертикальной оси отложена достоверность связывания белка ($-\log_{10}P$ -value). Пунктирной линией показано значение достоверности, при которой вероятность ложного обнаружения (FDR) равна 5%. Б) CG9879 связывается с регуляторными районами семенник-специфичных генов. На диаграмме показано обогащение сайтов связывания CG9879 в промоторах, нетранслируемых областях (НТО), кодирующих областях (кДНК), экзонах и интронах семенник-специфичных генов. По вертикальной оси отложен логарифм отношения наблюдаемой (Н) частоты сайтов в данных районах генов к ожидаемой (О). Звездочкой показаны области генов с достоверным

обогащением ($P\text{-value} < 0.001$ в биномиальном тесте). **В)** CG9879 связывается с промоторами преимущественно в проксимальной области. На диаграмме изображено распределение пиков CG9879 вблизи сайтов начала транскрипции (СНТ) генов. Фиолетовым цветом показано распределение, посчитанное для случайного набора геномных сайтов. **Г)** Частота сайтов связывания CG9879 в промоторах генов мишеней транскрипционных факторов Can, Comr и Mir40. Гены разделены на прямые мишени (содержат сайт связывания соответствующего транскрипционного фактора) и не прямые мишени (не содержат сайт связывания). Достоверность обогащения ($P\text{-value}$) посчитано с помощью теста Хи-квадрат. **Д)** CG9879 колокализуется с транскрипционными факторами Can и Comr. На диаграммах показано отношение наблюдаемой (Н) и ожидаемой (О) частоты пиков CG9879 (черная кривая) и случайного набора геномных сайтов (фиолетовая линия) на заданном удалении от сайтов связывания Can и Comr. Достоверность обогащения CG9879 в сайтах связывания посчитана с помощью биномиального теста. **Е)** Сайты связывания CG9879 содержат АТ-богатые мотивы. Поиск мотивов проводился с помощью программы DREME [206]. На диаграмме показано три мотива с наиболее достоверным обогащением. Для сравнения показан мотив связывания TBP [207].

Далее мы исследовали нашу гипотезу о том, что CG9879 служит промежуточным фактором, ответственным за регуляцию не прямых мишеней tTAF и tMAC. Однако это оказалось не так, поскольку гены, которые не являлись прямыми мишенями Comr, Can и Mir40, не были обогащены пиками CG9879 (**Рисунок 13Г**). Наоборот, 43% прямых мишеней гена Can имели пик CG9879 в пределах 1 кб вокруг TSS. Прямые гены-мишени Comr и Mir40 имели пики CG9879 вблизи TSS в 25% и 18% случаев, соответственно (**Рисунок 13А,Г**). Изучив обогащение CG9879 вокруг сайтов связывания Can и Comr, мы подтвердили колокализацию этого гомолога TBP с компонентами семенник-специфичных транскрипционных активаторами с сильным перевесом в сторону tTAF (**Рисунок 13Д**).

Активация транскрипции большинства генов эукариот начинается с взаимодействия между ТАТА-бокс связывающим белком TBP и группой TBP-ассоциированных факторов (TAF) с образованием базального фактора транскрипции TFIID [69]. Считается, что семенник-специфичные белки tTAF функционируют аналогичным образом, и образуют семенник-специфичную

версию TFIID [3,5,49]. В то же время не было известно, какие ТВР-подобные белки могут кооперировать с tTAF в сперматоцитах, и CG9879 может быть одним из кандидатов. Поиск коротких консенсусных последовательностей ДНК в сайтах связывания CG9879, а также в промоторах связанных с этим белком генов позволил обнаружить обогащение АТ-богатыми мотивами, схожими с ТАТА-боксом (**Рисунок 13Е**). Таким образом, наши результаты позволяют предположить, что CG9879 может функционировать как паралог ТВР в комплексе с tTAF.

Чтобы понять, как CG9879 влияет на экспрессию генов в семенниках, мы удалили кодирующий его ген с помощью системы CRISPR/Cas9 (**Рисунок 14А**). Удивительно, но мутанты с гомозиготной делецией не демонстрировали никаких морфологических дефектов или нарушений фертильности (**Рисунок 14Б**). Более того, анализ транскриптома показал, что только 28 генов значительно снижают экспрессию в семенниках у мутантов (**Рисунок 14В**), но ни один из них не был связан с CG9879 (данные не показаны). Учитывая наши данные о специфическом связывании CG9879 с прямыми мишенями tTAF, такой слабый эффект на транскрипцию, вероятно, объясняется избыточностью CG9879 в присутствии других паралогов *Tbr*, которые могут полностью заменить его функцию.

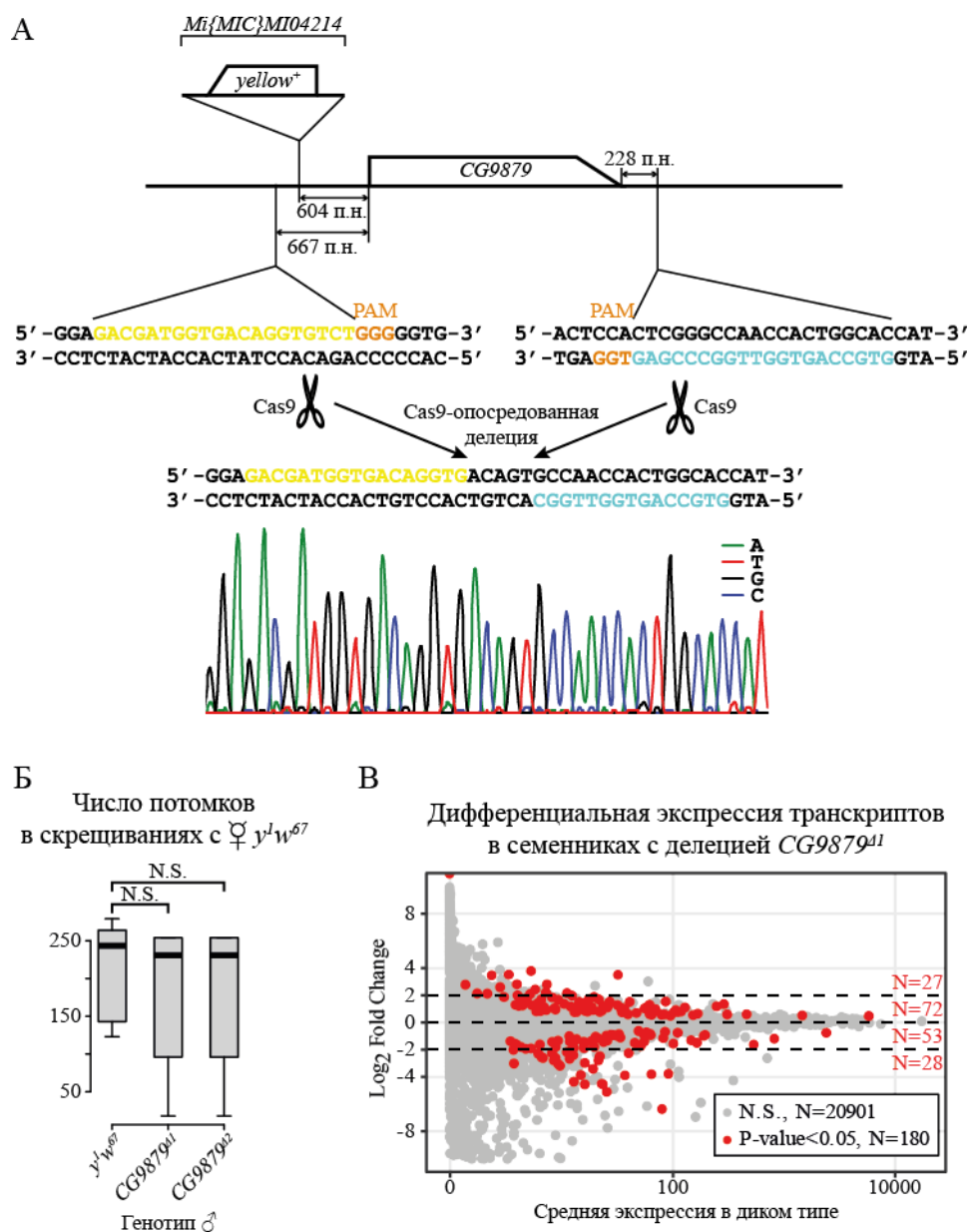


Рисунок 14. Делеция *CG9879* не приводит к нарушениям фертильности или значительным нарушениям транскрипции в семенниках. А) Схема CRISPR/Cas9-опосредованного удаления гена *CG9879*. Для получения делеции была выбрана трансгенная линия мух, несущая искусственный транспозон *MI04214* с репортерным геном *yellow*, встроенный на 604 п.н. выше гена *CG9879*. Были подобраны последовательности для двух направляющих РНК (желтый и голубой цвет), фланкирующие точку встройки транспозона и целевой ген. Нуклеаза Cas9 вносит разрывы на расстоянии 3-4 п.н. перед РАМ-сайтами (оранжевый цвет). Селекцию мутантов проводили по утрате маркера *yellow* (серые крылья). В результате трансгенеза было получено две гомозиготные линии мутантов. Последовательность ДНК и фрагмент секвенограммы в локусе, фланкирующем точку слияния разрывов у мутантов *CG9879^{Δ1}*, показаны внизу. Б) Делеция *CG9879* не приводит к нарушению фертильности. На диаграмме показано число потомков, полученных в серии скрещиваний

гомозиготных самцов *CG9879^{Δ1}* и *CG9879^{Δ2}* с самками *y¹w⁶⁷*. В качестве контроля использовали скрещивание с самцами *y¹w⁶⁷*. Достоверность отличий с контролем измеряли в тесте Манна-Уитни. **В)** Делеция *CG9879* приводит к незначительным изменениям транскрипции. На диаграмме показана дифференциальная экспрессия транскриптов в семенниках с делецией *CG9879^{Δ1}*. По горизонтальной оси отложен средний уровень экспрессии каждого транскрипта в семенниках дикого типа. По вертикальной оси отложено логарифм изменения экспрессии транскриптов ($\log_2\text{FoldChange}$). Транскрипты с достоверным изменением показаны красными точками. Пунктирными линиями показаны граничные значения $\log_2\text{FoldChange}$ -2, 0 и 2. Число красных точек в интервалах $\log_2\text{FoldChange}$ между граничными значениями показано красным текстом. Только 28 генов достоверно снижают экспрессию у мутантов с делецией более чем в 4 раза.

Чтобы показать такую возможность, мы исследовали экспрессию других генов из семейства гомологов *Tbp* в семенниках. Помимо *CG9879* в него входят *Trf*, *Trf2*, и *CG15398*. Гены *Tbp* и *CG15398* демонстрируют повышенную экспрессию в семенниках *bam*, обогащенных сперматогониями, и почти не изменяют активность в семенниках мутантов *cap*, *comr*, *mir40*, обогащенных недифференцированными сперматоцитами (**Рисунок 15А**). Это позволяет говорить о высокой активности этих генов в сперматогониях, что подтверждается данными по секвенированию транскриптома единичных клеток семенника (**Рисунок 15Б**) [194]. В то же время, гены *Trf* и *Trf2*, демонстрируют существенную активность во всех генотипах (**Рисунок 15А**). Анализ транскриптома единичных клеток демонстрирует, что они на высоком уровне транскрибируются в сперматоцитах параллельно с генами задержки мейоза (**Рисунок 15Б**). Таким образом, *Trf* и *Trf2* функционируют независимо от генов задержки мейоза, что было подтверждено при помощи *in situ* гибридизации РНК (**Рисунок 15В**). Напротив, *CG9879* обладает высокой активностью в семенниках дикого типа, но почти полностью инактивируется у мутантов (**Рисунок 15А**), а также демонстрирует паттерн экспрессии, характерный для генов дифференцировки сперматоцитов *fzo* и *gdl* (**Рисунок 15В**). Таким образом, *Trf2* и *Trf* активны в сперматоцитах и потенциально могут замещать утраченную функцию *CG9879* у мутантов.

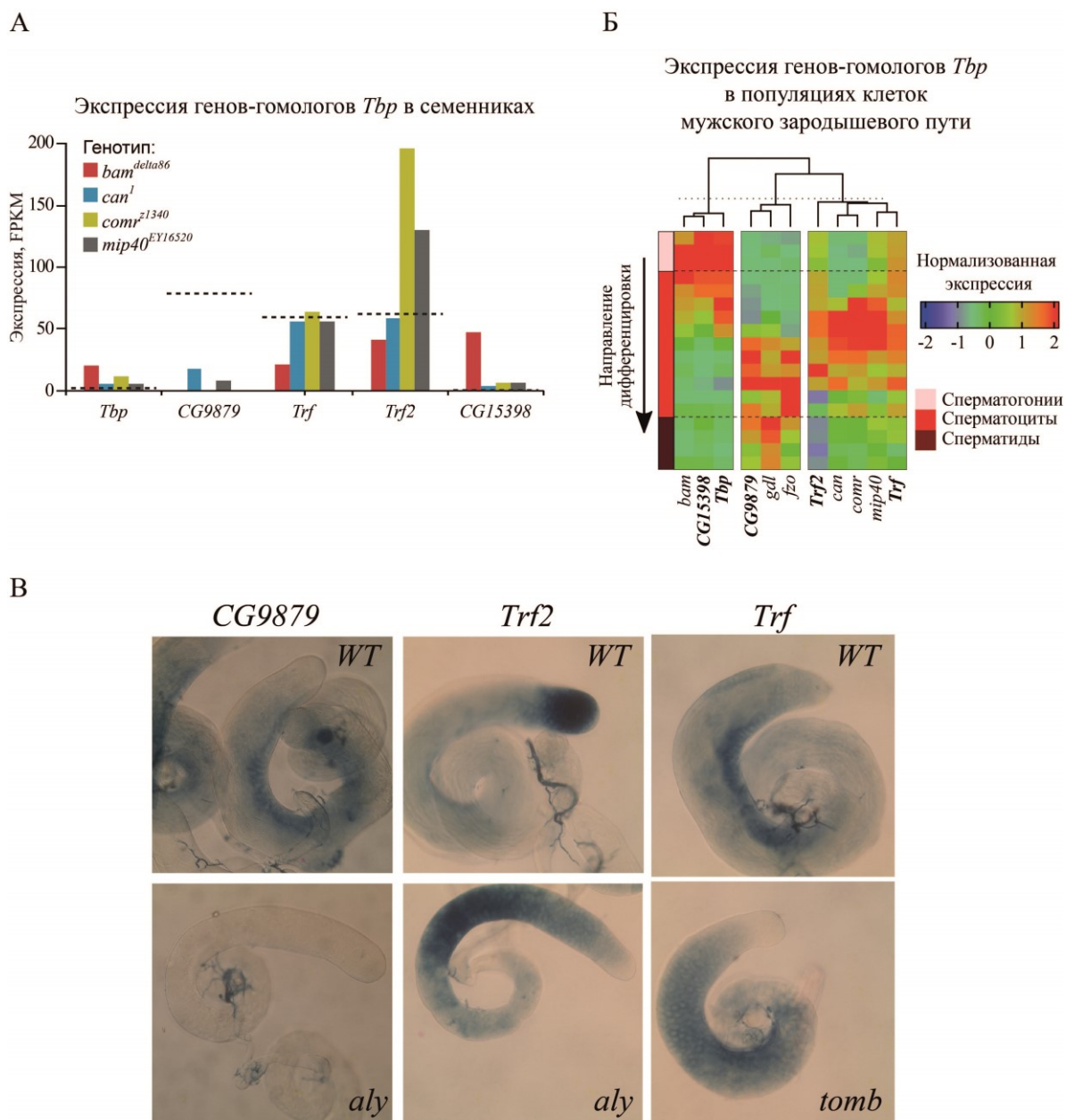


Рисунок 15. Экспрессия генов-паралогов *Tbp* в семенниках дикого типа и мутантов с задержкой мейоза. А) Экспрессия генов-паралогов *Tbp* в семенниках дикого типа и мутантов *bam*, *mip40*, *comr*, *can*. *CG9879* является единственным семенник-специфичным паралогом *Tbp*. Гены *Trf* и *Trf2* демонстрируют высокую экспрессию в семенниках дикого типа и у мутантов с нарушением сперматогенеза. Гены *Tbp* и *CG15398* активны преимущественно в сперматогониях. Пунктирные линии обозначают уровни экспрессии генов в семенниках дикого типа в единицах FPKM. **Б)** Паттерн экспрессии генов-паралогов *Tbp*, а также гена-маркера сперматогоний (*bam*), генов задержки мейоза (*can*, *comr*, *mip40*) и генов дифференцировки-сперматоцитов (*fzo*, *gdl*) в клетках мужского зародышевого пути по данным транскриптома единичных клеток семенника [194]. Тепловой картой показан нормализованный уровень экспрессии каждого гена. Клеточные типы включают несколько субпопуляций клеток с разным уровнем специализации.

Принадлежность клеток к сперматогониям, сперматоцитам и сперматидам обозначена цветными прямоугольниками. С помощью иерархической кластеризации гены были разделены на три группы по уровню экспрессии в клетках мужского зародышевого пути. Гены *Tbp* и *CG15398* экспрессируются преимущественно в сперматогониях. Гены *Trf* и *Trf2* экспрессируются в сперматоцитах на ранних этапах дифференцировки параллельно *cap*, *comr* и *mir40*. Ген *CG9879* демонстрирует паттерн экспрессии, характерный для генов дифференцировки сперматоцитов. **В)** Окраска транскриптов *CG9879* (левый столбец), *Trf2* (средний столбец) и *Trf* (правый столбец) в семенниках при помощи *in situ* гибридизации РНК. Верхний ряд – семенники дикого типа. Нижний ряд – семенники с задержкой мейоза.

4. ОБСУЖДЕНИЕ

Целью данной работы являлось изучение роли белков CP190 и CG9879 в регуляции активности генов в сперматогенезе *Drosophila melanogaster*. Интерес к CP190 обусловлен тем, что он является ключевым инсультаторным белком дрозофилы, и исследование его функции в сперматогенезе важно для понимания роли архитектурных белков в генетических программах развития. Чтобы исследовать функцию CP190 в клетках мужского зародышевого пути, мы разработали систему условного спасения мутантов *Cp190*, которая позволила добиться тканеспецифичной утраты этого белка начиная со стадии сперматоцитов. Мы продемонстрировали, что этот белок необходим для регуляции нескольких десятков генов дифференцировки сперматоцитов, чья активность в семенниках зависит от tMAC, и обнаружили tMAC-зависимый паттерн связывания CP190 с семенник-специфичными генами. Объединив полученные нами данные с опубликованными ранее результатами, нам удалось показать, что поддержание границ доменов H3K27me3 не может считаться основной функцией CP190, что являлось одним из нерешенных вопросов относительно функций этого белка.

Белок CG9879 является потенциальным участником вторичного регуляторного каскада масштабной активации генов в сперматоцитах. Поэтому мы использовали геномные методы, чтобы выявить роль этого белка в регуляции сперматогенеза. Однако, хотя мы выявили явную связь между распределением CG9879 в геноме и белками Can (tTAF) и Comr (tMAC), полученная нами полноразмерная делеция *CG9879* гена не привела к нарушению морфологии семенников или фертильности, а также вызывала только незначительные изменения в транскриптом семенников. Таким образом, *CG9879* не может считаться ключевым регулятором масштабной активации генов в сперматоцитах.

4.1. Исследование функции генов в системе условного спасения

Исследование функций генов, мутации которых вызывают гибель на ранних стадиях развития дрозофилы, является технически сложной задачей в клетках имаго. Одним из путей решения является опосредованная РНК-интерференцией (РНКи) инактивация, которая позволяет тканеспецифично снижать уровень экспрессии таких генов [208]. Широкий выбор GAL4-драйверов и линий мух, несущих конструкции для РНКи-инактивации, дают экспериментальные возможности для проведения подобных исследований [208–210].

Инструменты инактивации генов на основе РНКи демонстрируют различную эффективность в тканях мух. Так, многие трансгенные линии, доступные для заказа, не позволяют проводить инактивацию в клетках зародышевого пути самок без модификации генетических конструкций для экспрессии дцРНК [211,212]. Использование аналогичных линий позволяет добиться сравнительно эффективной инактивации генов в клетках мужского зародышевого пути дрозофилы [6,208]. Однако проведение подобных экспериментов требует содержания мух на повышенной температуре, что, согласно нашему опыту, может вызывать неконтролируемые нарушения сперматогенеза, включая задержку мейоза.

Для исследования функций генов с летальными мутациями также может быть использован подход на основе митотической рекомбинации ДНК, индуцированной системой FLP/FRT [213]. В результате применения такой методологии возможно получение мозаичных популяций клеток, в том числе несущих гомозиготную мутацию исследуемого гена [213]. В комбинации с системой UAS/GAL4 были разработаны модификации метода, в ходе которых происходит одновременная активация флуоресцентных маркеров в мутантных клетках либо клетках дикого типа, или специфическая элиминация клеток, не несущих мутацию [214–216]. Однако, важно отметить, что возникновение гомозиготных по мутации клеточных клонов

возможно только в ходе митотических делений, что существенно ограничивает применение таких методов.

Альтернативным подходом является условный нокаут, успешно применяемый в работах на дрозофиле, при котором происходит удаление спасающей конструкции на фоне мутации исследуемого гена под воздействием различных систем рекомбинации [217–220]. Наиболее близким к предложенному нами подходу является метод MARGE, удаление спасающей конструкции в котором происходит под воздействием тканеспецифичной FLP [217]. Однако в рамках данного подхода происходит мечение клеточной популяции, несущей спасающую конструкцию, что затрудняет идентификацию мутантных клеток.

Нами была разработана методология для условного спасения мутации генов в соматических клетках, что позволило исследовать регуляторные эффекты *Crp190* в клетках зародышевого пути самцов. Разработанная система основана на использовании тканеспецифичной системы Cre/loxP-рекомбинации, успешно применяемой в клетках зародышевого пути дрозофилы для активации трансгенов в специфических клеточных популяциях [11,169]. В данном же случае аналогичный подход использован для удаления спасающей генной кассеты.

При помощи разработанной системы мы добились значительного снижения уровня экспрессии *Crp190* в клетках зародышевого пути. Около 50% клеток семенника теряли спасающую кассету, что соответствует числу клеток зародышевого пути по результатам анализа транскриптома единичных клеток, и согласуется с отсутствием GFP-негативных сперматоцитов в семенниках с условным спасением. Высокую эффективность предложенной нами системы подтверждают и результаты иммуноокрашивания, согласно которым экспрессия CP190 происходит только в единичных сперматоцитах. Вместе с тем, система условного спасения не вызывает отклонений в развитии имаго и нарушений фертильности, что характерно для мух с оверэкспрессией CP190 [198]. Мы

также не обнаружили в семенниках с условным спасением видимых дефектов клеточного деления. Это согласуется с наблюдениями, сделанными в личиночных семенниках нуль-мутантов *Cp190*, где утрата этого белка не влияла на деление клеток [103].

Важно отметить, что принцип работы системы условного спасения универсален, и ее можно использовать для исследования других факторов, что требует только заменить кодирующую последовательность ДНК в спасающей конструкции и внедрить ее в геном мутантов. Кроме того, можно использовать другие типы кодирующих рекомбиназу Cre конструкций, чтобы добиться удаления спасающей кассеты в выбранном типе клеток, что значительно расширяет возможности системы. Таким образом, разработанная система является удобной альтернативой другим методам, позволяющим добиться инактивации генов в клетках *in vivo*. С использованием данной системы в клетках с утраченной функцией исследуемого фактора, запускается экспрессия репортерного белка GFP, что позволяет изолировать их методами сортировки. Это дает возможность использовать высокопроизводительные геномные методы, чтобы изучать роль регуляторов транскрипции в интересующей популяции клеток взрослого организма.

4.2. Роль белка CP190 в регуляции генов дифференцировки сперматоцитов

В этой работе мы провели подробное исследование эффекта белка CP190 на активность генов в клетках зародышевого пути *D. melanogaster*. Как оказалось, основное влияние мутация *Cp190* оказывает на субпопуляцию генов дифференцировки сперматоцитов, которые регулируются транскрипционным комплексом tMAC. Интересно, что в отсутствии белка CP190 активность этих генов усиливается в 2 и более раза, что позволяет сделать вывод, что этот белок в норме модулирует их активность, ограничивая уровень их экспрессии.

В целом, для инсуляторного белка CP190 в исследованиях на культурах клеток дрозофилы показаны умеренные изменения транскриптома на фоне его деплеции [18,21,107]. Нами также были выявлены лишь 183 гена, экспрессия которых зависит от CP190 в сперматоцитах. Из них 120 повышали экспрессию на фоне мутации *Sp190*, а 63 – снижали, что говорит в пользу превалирующей репрессорной функции CP190. В работах на культурах клеток также было показано, что CP190, как и другие инсуляторные белки, в большей степени склонен подавлять транскрипцию, несмотря на одновременную активаторную функцию [21,221].

Нами было обнаружено противоположное влияние CP190 на активность тканеспецифичных генов и генов домашнего хозяйства, что указывает на несколько возможных механизмов, с помощью которых CP190 контролирует их активность. Можно предположить, что для генов, активация которых зависит от CP190, он способен выступать в качестве транскрипционного фактора. Это косвенно подтверждается тем, что для активации CP190-связанных генов требуется кооперативное действие CP190 и фактора M1BP – регулятора генов домашнего хозяйства [21]. Причем гены, снижавшие экспрессию в клетках с РНКи-опосредованным нокдауном, демонстрировали обогащение CP190 в промоторах, в то время как репрессорный эффект CP190, по всей вероятности, реализуется не напрямую [21].

Важно отметить, что активация генетической программы дифференцировки клеток мужского зародышевого пути сопровождается значительной перестройкой архитектуры хроматина [222]. Так, на стадии сперматоцитов уменьшается количество дистальных контактов, а промоторы активных генов дифференцировки оказываются в изолированном окружении [222]. Для CP190, наряду с белком Chro, было показано участие в формировании и поддержании топологически ассоциированных доменов (ТАД) [221]. Причем было обнаружено, что нарушение экспрессии генов в результате удаления CP190, коррелирует с изменением границ окружающего их ТАД [221]. Более того, было обнаружено, что ТАД, содержащие гены

домашнего хозяйства, имеют тенденцию к более высокой стабильности [221]. Это позволяет предположить, что наблюдаемое дифференциальное влияние CP190 на экспрессию тканеспецифичных генов и генов домашнего хозяйства может быть обусловлено различиями в механизмах регуляции на уровне структуры хроматина, в результате чего гены домашнего хозяйства активируются под действием CP190, тогда как тканеспецифичные гены подавляются.

Учитывая отсутствие влияния на дифференцировку клеток мужского зародышевого пути, нормальную фертильность мутантных самцов, а также умеренный эффект на изменение паттерна экспрессии генов, инсуляторный белок CP190 вряд ли может рассматриваться в качестве ключевого регуляторного фактора сперматогенеза. Для сравнения, мутация *comr* (tMAC) затрагивает экспрессию более 2500 генов и вызывает задержку мейоза [12]. Однако, с учетом ключевой роли CP190 в регуляции архитектуры ядра, можно предположить, что недостаточность CP190 приводит к сбоям в коммуникации специфичных транскрипционных факторов и их регуляторных элементов, а также общим изменениям структуры хроматина, что вызывает нарушения в механизмах тонкой регуляции экспрессии. Чтобы проверить это предположение, мы планируем дополнить полученные в данном исследовании результаты анализом архитектуры ядра в развивающихся мужских половых клетках. Это позволит охарактеризовать механизмы масштабной активации семенник-специфичных генов в сперматоцитах на уровне трехмерной организации ядра.

4.3. Динамика связывания CP190 с хроматином в процессе дифференцировки клеток мужского зародышевого пути

До настоящего момента анализ связывания CP190 с хромосомами дрозофилы на полногеномном уровне производили в клеточных культурах. Было продемонстрировано, что расположение сайтов связывания

инсуляторных белков дрозофилы в основном одинаково для разных типов клеток [107]. Полученные нами данные демонстрируют крайне вариабельный паттерн связывания CP190 с хроматином. Обнаруженная нами динамика CP190 не была описана ранее и требует внимательного анализа. Мы приходим к выводу, что значительная вариабельность может объясняться несколькими причинами.

Во-первых, данные на клетках в культуре могут не отражать события, происходящие в организме, должным образом. Например, известно довольно много примеров, когда инсуляторные белки обеспечивают специфическую для разного типа клеток организма конформацию хроматина в локусах, кодирующих тканеспецифичные гены [113,223,224]. В семенниках *D. melanogaster* экспрессируется 50% из примерно 14000 всех генов плодовой мушки [3]. При этом 1600 генов являются семенник-специфичными [38–40]. Такое обилие тканеспецифичных генов является отличительной особенностью семенников – в других тканях экспрессируется не более нескольких сотен специфичных генов [40]. Возможно, вариабельное связывание CP190 является уникальной чертой клеток мужского зародышевого пути.

Во-вторых, механизмы регуляции инсуляторных белков в ходе дифференцировки клеток недостаточно изучены. Хотя показано, что реакция на тепловой шок и экдизон могут изменять связывание CP190 с хроматином, эти данные получены на клеточных культурах [24]. Первые доказательства тканеспецифичной регуляции инсуляторов были получены для центральной нервной системы дрозофилы, где был обнаружен тканеспецифичный РНК-связывающий белок *Shep*, который играл важную роль в ремоделировании нейронов через регуляцию экспрессии генов. Было показано, что *Shep* взаимодействует с инсуляторными белками *Su(Hw)* и CP190 и подавляет барьерную активность инсуляторов *gypsy* в ЦНС [225]. Проведенный нами анализ показал роль *Mir40* в связывании CP190 с хроматином. Поскольку в отсутствие *Mir40* белок *Comr* не способен стабильно связываться с

хромосомами, мы приходим к выводу, что эффект Mir40 на связывание CP190 локусами, ассоциированными с tTAF и tMAC, обусловлен потерей функции tMAC.

Интересно, что нарушение связывания CP190 с хромосомами обнаруживается в клетках, у которых нарушена функция MMB/dREAM [15,77]. Это снова указывает на схожий механизм действия его семенник-специфичного паралога tMAC. Вместе с тем, MMB/dREAM также необходим для энхансер-блокирующей активности некоторых дистальных регуляторных элементов генома, которые связаны с компонентом этого комплекса E2F2, и CP190 [15]. Это позволяет предположить, что механизм действия MMB/dREAM в некоторых случаях связан с регуляцией пространственных взаимодействий в геноме. В нашей работе мы обнаружили значительное пересечение между CP190 и белком Comr из комплекса tMAC. В силу этого заманчиво предположить, что tMAC также может контролировать экспрессию через регуляцию дистальных промотор-энхансерных взаимодействий. В то же время может оказаться крайне трудной задачей исследовать этот вопрос при помощи репортерного анализа, так как барьерную активность ассоциированных с tMAC регуляторных элементов нужно будет проверять только в клетках мужского зародышевого пути, где этот комплекс функционирует. Однако анализ пространственной архитектуры хроматина в клетках мужского зародышевого пути может помочь проверить эту гипотезу.

4.4. SetDB1, но не CP190, может участвовать в поддержании доменной организации хромосом

CP190 является одним из ключевых архитектурных белков дрозофилы который важен для барьерной функции инсуляторов, и неоднократно высказывалось предположение, что он может устанавливать барьеры на границах доменов H3K27me3 [98,107,128]. Эти выводы подкреплялись

примерами в единичных локусах генома. В то же время, исследование на полногеномном уровне в культурах клеток S2 давали противоречивые результаты [17,18]. Хотя колокализация CP190 с границами доменов была очевидной, подавление экспрессии этого белка в клетках приводило лишь к незначительным эффектам, и экспансия H3K27me3 за пределы границ происходила в считанных случаях [18]. Это может означать, что CP190 не важен для поддержания функции барьеров в дифференцированных клетках. С использованием полученных нами данных мы хотели проверить гипотезу о том, что CP190 поддерживает границы доменов H3K27me3 и исследовали колокализацию этого белка с границами доменов репрессии Polycomb в политекных хромосомах личиночных слюнных желез. Однако мы не обнаружили обогащения CP190 на границах. Напротив, он зачастую находился внутри доменов. Это означает, что поддержание доменной организации районов репрессии Polycomb не является функцией CP190.

К нашему удивлению, очень высокую корреляцию с положением пограничных элементов триметилирования H3K27 демонстрировал SetDB1. При этом половина всех пограничных элементов располагалась на расстоянии менее 5 кб от этого белка. Примечательно, что связывание этого белка на границах доменов совпадало с высокой транскрипционной активностью рядом лежащих генов. Обнаруженная нами ассоциация между SetDB1 и границами доменов репрессии Polycomb не была описана ранее и требует внимательного анализа.

Изначально SetDB1 был описан как метилтрансфераза, ответственная за метилирование H3K9 в хромосоме 4 и усиление мозаичного фенотипа у мух, несущих репортерные конструкции с геном *white* [226]. Другие H3K9-специфичные метилтрансферазы Su(var)3-9 и G9a являются основными факторами, вовлеченными в реессию генов через формирование конститутивного гетерохроматина. Установлено, что SetDB1 играет ключевую роль в инактивации мобильных элементов, что необходимо для инициации репрессивного состояния гетерохроматиновых доменов [227]. В

нашей работе показано, что SetDB1 ассоциирован с высоко активными генами на границах доменов репрессии Polycomb. Таким образом, в данных локусах этот белок не функционирует как репрессор и его роль не заключается в метилировании H3K9. Однако это не должно быть противоречием, поскольку несмотря на то, что изначальной функцией SetDB1 считалось метилирование H3K9, его связь с этой меткой не является однозначной. Исследования на мышах показали, что 45,7% сайтов связывания SetDB1 не имели близлежащих пиков сигнала H3K9me3 [228]. Кроме того, на различных модельных объектах продемонстрировано, что гистонметилтрансферазы модифицируют негистоновые субстраты [229–231]. В частности, сообщалось, что SetDB1 метилирует белок Tat ВИЧ [232], а также p53 человека [233] и киназу АКТ [234,235].

В нашей работе SetDB1 был обнаружен примерно на половине границ доменов H3K27me3, что свидетельствует о его участии в разделении хроматина на функциональные домены. Предполагаемая роль SetDB1 в формировании пограничных элементов согласуется с результатами предыдущих исследований. Например, некоторые районы, обогащенные H3K9me2/3, как правило, не связаны с SetDB1, а у мутантов по гену *egg*, кодирующему этот белок у дрозофилы, метилирование H3K9 в таких регионах не уменьшается, а увеличивается, причем домены становятся несколько шире [182]. Приблизительно 40% мишеней Su(var)3-9 не связаны с SetDB1, однако мутация *egg* приводит к исчезновению Su(var)3-9 из таких регионов. Кроме того, ранее в нашей лаборатории было описано небольшое подмножество генов, которые не связаны с Su(var)3-9 у животных дикого типа, но становятся мишенями Su(var)3-9 у мутантов *egg* [236]. Эти факты можно объяснить не прямым, а косвенным влиянием мутации *egg* на метилирование гистонов или связывание белков, например, через слияние или расщепление доменов. В результате часть геномных локусов может оказаться в новом эпигенетическом окружении.

Наша находка о связи SetDB1 с CP190 и пограничными элементами доменов H3K27me3 хорошо согласуется с полученными для млекопитающих данными функции SetDB1 в локальной регуляции архитектуры хроматина. Так в нейронах мыши подмножество очень крупных ТАД распадается после удаления SetDB1 [237]. Делеция или оверэкспрессия *Setdb1* в нейронах мыши и человека изменяет высокоуровневую архитектуру хроматина [238]. В эмбриональных стволовых клетках SetDB1 колокализуется с инсуляторным белком CTCF и когезином в более чем 10000 тысячах сайтах в эухроматине и его утрата приводит к нарушению транскрипции из-за переключения генов между геномными компартментами [239]. Таким образом, обнаруженная нами связь между SetDB1 и CP190 отражает высоко консервативную функцию SetDB1, которая может заключаться в регуляции активности барьерных элементов генома.

4.5. Избыточная функция TBP-подобных белков поддерживает стабильность генетической системы сперматогенеза

Активация генов дифференцировки в ходе сперматогенеза представляет из себя эволюционно консервативный механизм, и участие семенник-специфичных транскрипционных активаторов, паралогичных компонентам базальных факторов транскрипции, показано у млекопитающих. Например, белки TAF7L и TAF4B паралогичные компонентам TFIID, а также паралог TFIIA под названием ALF, играют ключевую роль в активации генов в гаплоидных клетках семенников у мышей [166,240,241]. Белок TRF2, гомолог Trf2 дрозофилы, работает совместно с ALF в регуляции подмножества постмейотических генов, направляющих спермиогенез [242]. Более того, у мышей предполагается существование разных преинициаторных комплексов транскрипции, ассоциированных либо с TRF2, либо с TBP и TRF2 одновременно [242]. Таким образом, существует высокая вероятность, что факторы tTAF дрозофилы образуют транскрипционный

активатор, паралогичный TFIID, однако остается неясным какие из белковых гомологов *Tbp* образуют ядро комплекса tTAF [3,5,65].

В данной работе мы выяснили, что белок CG9879 может играть роль ТВР в комплексе с tTAF. Об этом говорит то, что CG9879 имеет тенденцию к колокализации с белком Cap в сперматоцитах. Однако делеция гена *CG9879* привела к очень слабым изменениям в транскриптоме семенников и не оказала заметного влияния на сперматогенез. Мы предполагаем, что функция этого гена может быть компенсирована другими ТВР-подобными белками, функционирующими в сперматоцитах, например Trf2. В пользу роли Trf2 в сперматогенезе говорит то, что мутации с потерей функции *Trf2* приводят к нарушению дифференцировки половых клеток у самцов и у самок [158]. В то же время, согласно данным DamID только около половины генов, которые связываются с Cap напрямую, также ассоциированы с CG9879. Можно предположить, что CG9879 участвует в регуляции ограниченного круга генов дифференцировки сперматоцитов, тогда как остальные ТВР-подобные белки, включая Trf2, вовлечены в этот процесс в большей степени. Такая избыточность может способствовать поддержанию стабильности этой важной генетической системы.

4.6. Заключение

Специализация клеток в ходе развития многоклеточных организмов вызывается активацией тканеспецифичных генов. Особенностью сперматогенеза является то, что тканеспецифичные гены здесь активируются взрывообразно на стадии сперматоцитов первого порядка. Ключевую роль в такой масштабной активации играют транскрипционные активаторы tTAF и tMAC. Однако эти факторы напрямую регулируют лишь некоторую часть семенник-специфичных генов. В данной работе мы обратили внимание на белки, чья функция может вносить вклад в непрямую регуляцию экспрессии генов в сперматоцитах. Мы продемонстрировали, что CP190 принимает участие в контроле транскрипции семенник-специфичных генов и его ассоциация с регуляторными областями генома сильно зависит от функции tMAC. Мы также исследовали функцию белка CG9879, который, как предполагалось, являлся наиболее вероятным вторичным посредником между tMAC, tTAF и их непрямыми мишенями. Однако, хотя мы выявили явную связь между распределением CG9879 в геноме и белками Cap (tTAF) и Comr (tMAC), его утрата не вызывала только значительных изменений в транскриптоме семенников, что, по-видимому, связано с пересекающейся функцией паралогов *Tbr*.

Объединив полученные данные с опубликованными ранее результатами, нам удалось показать, что поддержание границ доменов H3K27me3 не может считаться основной функцией CP190, что являлось одним из нерешенных вопросов относительно функции этого белка. Полученные результаты расширяют представление о роли исследованных нами факторов в регуляции клеточной дифференцировки и в будущем позволят построить исчерпывающую модель масштабной активации генов в сперматогенезе дрозофилы.

ВЫВОДЫ

1. Разработана генетическая система условного спасения мутантов *Cp190*. Она позволяет добиться утраты белка CP190 в клетках мужского зародышевого пути, начиная со стадии ранних сперматоцитов, но сохраняет его в соматических клетках.

2. Установлено, что связывание CP190 с хромосомами динамично. По мере дифференцировки клеток мужского зародышевого пути белок CP190 перераспределяется к сайтам связывания tMAC и tTAF и в сперматоцитах регулирует активность tMAC-зависимых генов дифференцировки.

3. Установлено, что в политенных хромосомах личиночных слюнных желез связывание CP190 не коррелирует с границами доменов H3K27me3.

4. Установлено, что белок CG9879 является транскрипционным фактором – семенник-специфичным паралогом TBP. В клетках мужского зародышевого пути белок CG9879 связывается с АТ-богатыми мотивами в промоторах прямых мишеней tTAF и tMAC.

5. С использованием системы CRISPR/Cas9 получена полноразмерная делеция гена *CG9879*. Установлено, что она вызывает лишь незначительные изменения в экспрессии генов в семенниках и не приводит к нарушению фертильности самцов.

СПИСОК ЛИТЕРАТУРЫ

1. Fuller M. Spermatogenesis // The development of *Drosophila melanogaster* / ed. Bate M., Martinez A. – NY: Cold Spring Harbor Laboratory Press, 1993. – V. 1. – P. 71–147.
2. Fabian L., Brill J. A. *Drosophila* spermiogenesis: Big things come from little packages // Spermatogenesis. – 2012. – V. 2. – P. 197–212.
3. White-Cooper H. Molecular mechanisms of gene regulation during *Drosophila* spermatogenesis // Reproduction. – 2010. – V. 139. – P. 11–21.
4. Beall E.L. et al. Discovery of tMAC: A *Drosophila* testis-specific meiotic arrest complex paralogous to Myb-Muv B // Genes and Development. – 2007. – V. 21. – P. 904–919.
5. Metcalf C.E., Wassarman D. a. Nucleolar colocalization of TAF1 and testis-specific TAFs during *Drosophila* spermatogenesis // Developmental Dynamics. – 2007. – V. 236. – № 10. – P. 2836–2843.
6. Doggett K. et al. Wake-up-call, a lin-52 paralogue, and Always early, a lin-9 homologue physically interact, but have opposing functions in regulating testis-specific gene expression // Developmental Biology. – 2011. – V. 355. – № 2. – P. 381–393.
7. Lin T.Y. et al. Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. // Development. – 1996. – V. 122. – P. 1331–1341.
8. Hiller M. et al. Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. // Development. – 2004. – V. 131. – P. 5297–5308.
9. Theofel I. et al. tBRD-1 Selectively Controls Gene Activity in the *Drosophila* Testis and Interacts with Two New Members of the Bromodomain and Extra-Terminal (BET) Family // PLoS ONE. – 2014. – V. 9. – № 9. – P. e108267.
10. Beall E.L. et al. Role for a *Drosophila* Myb-containing protein

complex in site-specific DNA replication. // *Nature*. – 2002. – V. 420. – P. 833–837.

11. Laktionov P.P. et al. Genome-wide analysis of gene regulation mechanisms during *Drosophila* spermatogenesis // *Epigenetics and Chromatin*. – 2018. – V. 11. – № 1. – P. 14.

12. Laktionov P.P. et al. Transcription factor Comr acts as a direct activator in the genetic program controlling spermatogenesis in *D. melanogaster* // *Molecular Biology*. – 2014. – V. 48. – № 1. – P. 130–140.

13. Gómez-Díaz E., Corces V.G. Architectural proteins: regulators of 3D genome organization in cell fate // *Trends in Cell Biology*. – 2014. – V. 24. – № 11. – P. 703–711.

14. Cubeñas-Potts C., Corces V.G. Architectural proteins, transcription, and the three-dimensional organization of the genome // *FEBS Letters*. – 2015. – V. 589. – № 20. – P. 2923–2930.

15. Korenjak M. et al. dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. // *Nucleic Acids Research*. – 2014. – V. 42. – № 14. – P. 8939–8953.

16. Vogelmann J. et al. Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the *Drosophila* genome. // *PLoS Genetics*. – 2014. – V. 10. – № 8. – P. e1004544.

17. Bartkuhn M. et al. Active promoters and insulators are marked by the centrosomal protein 190 // *EMBO Journal*. – 2009. – V. 28. – № 7. – P. 877–888.

18. Schwartz Y.B. et al. Nature and function of insulator protein binding sites in the *Drosophila* genome // *Genome Research*. – 2012. – V. 22. – № 11. – P. 2188–2198.

19. Mishal R., Luna-Arias J.P. Role of the TATA-box binding protein (TBP) and associated family members in transcription regulation // *Gene*. – 2022. – V. 833. – P. 146581.

20. Winick-Ng W. et al. Cell-type specialization is encoded by specific chromatin topologies // *Nature*. – 2021. – V. – 599. – № 7886. – P. 684–691.

21. Bag I. et al. M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity // *Nature Communications*. – 2021. – V. 12. – № 1. – P. 4170.
22. Nègre N. et al. A comprehensive map of insulator elements for the *Drosophila* genome // *PLoS Genetics*. – 2010. – V. 6. – № 1. – P. e1000814.
23. Oliver D. et al. The chromosomal association/dissociation of the chromatin insulator protein CP190 of *Drosophila melanogaster* is mediated by the BTB/POZ domain and two acidic regions. // *BMC Cell Biology*. – 2010. – V. 11. – P. 101.
24. Wood A.M. et al. Regulation of Chromatin Organization and Inducible Gene Expression by a *Drosophila* Insulator // *Molecular Cell*. – 2011. – V. 44. – № 1. – P. 29–38.
25. Hardy R.W. et al. The germinal proliferation center in the testis of *Drosophila melanogaster*. // *Journal of ultrastructure research*. – 1979. – V. 69. – № 2. – P. 180–190.
26. Sonnenblick B.P. Germ Cell Movements and Sex Differentiation of the Gonads in the *Drosophila* Embryo. // *Proceedings of the National Academy of Sciences of the United States of America*. – 1941. – V. 27. – № 10. – P. 484–489.
27. Dinardo S. et al. lines and bowl affect the specification of cyst stem cells and niche cells in the *Drosophila* testis. // *Development*. – 2011. – V. 138. – № 9. – P. 1687–1696.
28. Gönczy P., DiNardo S. The germ line regulates somatic cyst cell proliferation and fate during *Drosophila* spermatogenesis. // *Development*. – 1996. – V. 122. – P. 2437–2447.
29. Olivieri G., Olivieri A. Autoradiographic study of nucleic acid synthesis during spermatogenesis in *Drosophila melanogaster* // *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. – 1965. – V. 2. – № 4. – P. 366–380.
30. Barreau C. et al. Post-meiotic transcription in *Drosophila* testes. // *Development*. – 2008. – V. 135. – P. 1897–1902.

31. Tokuyasu K.T., Peacock W.J., Hardy R.W. Dynamics of spermiogenesis in *Drosophila melanogaster* - II. Coiling process // Zeitschrift für Zellforschung und Mikroskopische Anatomie. – 1972. – V. 127. – № 4. – P. 492–525.
32. Gönczy P., Matunis E., DiNardo S. bag-of-marbles and benign gonial cell neoplasm act in the germline to restrict proliferation during *Drosophila* spermatogenesis. // Development. – 1997. – V. 124. – № 21. – P. 4361–4371.
33. McKearin D.M., Spradling a C. bag-of-marbles: a *Drosophila* gene required to initiate both male and female gametogenesis // Genes and Development. – 1990. – V. 4. – № 12b. – P. 2242–2251.
34. Insko M.L. et al. Accumulation of a differentiation regulator specifies transit amplifying division number in an adult stem cell lineage. // Proceedings of the National Academy of Sciences of the United States of America. – 2009. – V. 106. – № 52. – P. 22311–22316.
35. Schulz C. et al. A misexpression screen reveals effects of bag-of-marbles and TGF β class signaling on the *Drosophila* male germ-line stem cell lineage // Genetics. – 2004. – V. 167. – № 2. – P. 707–723.
36. Xiao S.J. et al. TiSGeD: A database for tissue-specific genes // Bioinformatics. – 2010. – V. 26. – № 9. – P. 1273–1275.
37. Jain A., Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis // Bioinformatics. – 2019. – V. 35. – № 11. – P. 1966–1967.
38. Boutanaev A.M. et al. Large clusters of co-expressed genes in the *Drosophila* genome // Nature. – 2002. – V. 420. – P. 666–669.
39. Belyakin S.N. et al. Genomic analysis of *Drosophila* chromosome underreplication reveals a link between replication control and transcriptional territories. // Proceedings of the National Academy of Sciences of the United States of America. – 2005. – V. 102. – № 23. – P. 8269–8274.
40. Chintapalli V.R., Wang J., Dow J. a T. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. // Nature genetics. – 2007. – V. 39. – № 6. – P. 715–720.

41. Chen X. et al. Sequential changes at differentiation gene promoters as they become active in a stem cell lineage. // *Development*. – 2011. – V. 138. – P. 2441–2450.
42. Kalmykova A.I. et al. Regulated chromatin domain comprising cluster of co-expressed genes in *Drosophila melanogaster* // *Nucleic Acids Research*. – 2005. – V. 33. – № 5. – P. 1435–1444.
43. Shevelyov Y.Y. et al. The B-type lamin is required for somatic repression of testis-specific gene clusters. // *Proceedings of the National Academy of Sciences of the United States of America*. – 2009. – V. 106. – № 9. – P. 3282–3287.
44. Meadows L. a. et al. Neighbourhood continuity is not required for correct testis gene expression in *Drosophila* // *PLoS Biology*. – 2010. – V. 8. – № 11.
45. White-Cooper H. et al. Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. // *Development*. – 1998. – V. 125. – P. 125–134.
46. Hiller M. a. et al. Developmental regulation of transcription by a tissue-specific TAF homolog // *Genes and Development*. – 2001. – V. 15. – P. 1021–1030.
47. Jiang J., White-Cooper H. Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of aly and a novel meiotic arrest gene cookie monster. // *Development*. – 2003. – V. 130. – P. 563–573.
48. Leser K. et al. The bromodomain-containing protein tBRD-1 is specifically expressed in spermatocytes and is essential for male fertility // *Open biology*. – 2012. – V. 1. – P. 597–606.
49. Kimura S. et al. Two bromodomain proteins functionally interact to recapitulate an essential BRDT-like function in *Drosophila* spermatocytes // *Open biology*. – 2015. – V. 5. – № 2. – P. 140–145.
50. Ayyar S. et al. *Drosophila* TGIF is essential for developmentally

regulated transcription in spermatogenesis. // *Development*. – 2003. – V. 130. – P. 2841–2852.

51. Perezgasga L. et al. Regulation of transcription of meiotic cell cycle and terminal differentiation genes by the testis-specific Zn-finger protein matotopetli. // *Development*, – 2004. – V. 131. – № 8. – P. 1691–1702.

52. Jiang J. et al. Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the *Drosophila* male germline and physically interacts with always early. // *Development*. – 2007. – V. 134. – P. 1549–1559.

53. Lewis P.W. et al. Identification of a *Drosophila* Myb-E2F2/RBF transcriptional repressor complex // *Genes and Development*. – 2004. – V. 18. – № 23. – P. 2929–2940.

54. Sadasivam S., DeCaprio J. a. The DREAM complex: master coordinator of cell cycle-dependent gene expression. // *Nature reviews. Cancer*. – 2013. – V. 13. – № 8. – P. 585–595.

55. Kobayashi K. et al. Transcriptional repression by MYB 3 R proteins regulates plant organ growth // *The EMBO Journal*. – 2015. V 34. – №15. – P. 1992–2007.

56. Rovani M.K. et al. The dREAM/Myb-MuvB complex and Grim are key regulators of the programmed death of neural precursor cells at the *Drosophila* posterior wing margin // *Developmental Biology*. – 2012. – V. 37. – № 1. – P. 88–102.

57. Kiger A.A., White-Cooper H., Fuller M.T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. // *Nature*. – 2000. – V. 407. – № 6805. – P. 750–754.

58. Wang Z., Mann R.S. Requirement for two nearly identical TGIF-related homeobox genes in *Drosophila* spermatogenesis. // *Development*. – 2003. – V. 130. – P. 2853–2865.

59. Trost M. et al. *Drosophila* dany is essential for transcriptional control and nuclear architecture in spermatocytes // *Development*. – 2016. – V. 143. – № 14. – P. 2664–2676.

60. Lu C., Fuller M.T. Recruitment of Mediator Complex by Cell Type and Stage-Specific Factors Required for Tissue-Specific TAF Dependent Gene Activation in an Adult Stem Cell Lineage // PLoS Genetics. – 2015. – V. 11. – № 12. – P. 1–24.
61. Cruz-Becerra G. et al. Analysis of *Drosophila* p8 and p52 mutants reveals distinct roles for the maintenance of TFIIH stability and male germ cell differentiation // Open biology. – 2016. – V. 6. – № 10. – P. 160222.
62. Kwon S.Y. et al. Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities // PLoS Genetics. – 2009. – V. 5. – № 7.
63. Tsukiyama T., Wu C. Purification and properties of an ATP-dependent nucleosome remodeling factor // Cell. – 1995. – V. 83. № 6. – P. 1011–1020.
64. Tsukiyama T. et al. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor // Cell. – 1995. – V. 83. – № 6. – P. 1021–1026.
65. Chen X. et al. Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. // Science. – 2005. – V. 310. – № 2005. – P. 869–872.
66. El-Sharnouby S., Redhouse J., White R. a H. Genome-Wide and Cell-Specific Epigenetic Analysis Challenges the Role of Polycomb in *Drosophila* Spermatogenesis // PLoS Genetics. – 2013. – V. 9. – № 10. – e1003842.
67. Laktionov P.P. et al. Genome-wide profiling of gene expression and transcription factors binding reveals new insights into the mechanisms of gene regulation during *Drosophila* spermatogenesis. // Epigenetics and Chromatin – 2018. – V. 11. – 14.
68. Klose R.J. et al. Chromatin Sampling-An Emerging Perspective on Targeting Polycomb Repressor Proteins // PLoS Genetics. – 2013. – V. 9. – № 8.
69. Louder R.K. et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. // Nature. – 2016. – V. 531. – № 7596. – P. 604–609.

70. Akhtar W., Veenstra G.J.C. TBP-related factors: A paradigm of diversity in transcription initiation // *Cell and Bioscience*. – 2011. – V. 1. – № 1. – P. 23.
71. Korenjak M. et al. Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes // *Cell*. – 2004. – V. 119. – № 2. – P. 181–193.
72. Georlette D. et al. Genomic profiling and expression studies reveal both positive and negative activities for the *Drosophila* Myb MuvB/dREAM complex in proliferating cells. // *Genes and development*. – 2007. – V. 21. – № 22. – P. 2880–2896.
73. Lewis P.W. et al. *Drosophila* Lin-52 Acts in Opposition to Repressive Components of the Myb-MuvB/dREAM Complex // *Molecular and Cellular Biology*. – 2012. – V. 32. – № 16. – P. 3218–3227.
74. Beall E.L. et al. Dm-myb mutant lethality in *Drosophila* is dependent upon mip130: Positive and negative regulation of DNA replication // *Genes and Development*. – 2004. – V. 18. – № 14. – P. 1667–1680.
75. Sim C.K. et al. Epigenetic regulation of olfactory receptor gene expression by the Myb-MuvB/dREAM complex // *Genes and Development*. – 2012. – V. 26. – № 22. – P. 2483–2498.
76. DeBruhl H., Wen H., Lipsick J.S. The Complex Containing *Drosophila* Myb and RB/E2F2 Regulates Cytokinesis in a Histone H2Av-Dependent Manner // *Molecular and Cellular Biology*. – 2013. – V. 33. – № 9. – P. 1809–1818.
77. Bohla D. et al. A Functional Insulator Screen Identifies NURF and dREAM Components to Be Required for Enhancer-Blocking // *PLoS ONE*. – 2014. – V. 9. – № 9. – P. e107765.
78. Dixon J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions // *Nature*. – 2012. – V. 485. – № 7398. – P. 376–380.
79. Sexton T. et al. Three-Dimensional Folding and Functional

Organization Principles of the *Drosophila* Genome // Cell. – 2012. – V. 148. – № 3. – P. 458–472.

80. Hou C. et al. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. // Molecular Cell. – 2012. – V. 48. – № 3. – P. 471–484.

81. Geyer P.K., Corces V.G. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. // Genes and Development. – 1992. – V. 6. – № 10. – P. 1865–1873.

82. Li Q., Stamatoyannopoulos G. Hypersensitive site 5 of the human beta locus control region functions as a chromatin insulator // Blood. – 1994. – V. 84. – № 5. – P. 1399–1401.

83. Donze D. et al. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae* // Genes and development. – 1999. – V. 13. – № 6. – P. 698–708.

84. Kellum R., Schedl P. A group of scs elements function as domain boundaries in an enhancer-blocking assay // Molecular and cellular biology. – 1992. – V. 12. – № 5. – P. 2424–2431.

85. Kellum R., Schedl P. A position-effect assay for boundaries of higher order chromosomal domains // Cell. – 1991. – V. 64. – № 5. – P. 941–950.

86. West A.G., Gaszner M., Felsenfeld G. Insulators: many functions, many mechanisms // Genes and Development. – 2002. – V. 16. – № 3. – P. 271–288.

87. Udvardy A., Maine E., Schedl P. The 87A7 chromomere // Journal of Molecular Biology. – 1985. – V. 185. – № 2. – P. 341–358.

88. Gdula D.A., Gerasimova T.I., Corces V.G. Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila* // Proceedings of the National Academy of Sciences of the United States of America. – 1996. – V. 93. – № 18. – P. 9378–9383.

89. Belozarov V.E. et al. A novel boundary element may facilitate independent gene regulation in the Antennapedia complex of *Drosophila* // The

EMBO journal. – 2003. – V. 22. – № 12. – P. 3113–3121.

90. Hagstrom K., Muller M., Schedl P. Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. // Genes and Development. – 1996. – V. 10. – № 24. – P. 3202–3215.

91. Moon H. et al. CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator // EMBO reports. – 2005. – V. 6. – № 2. – P. 165–170.

92. Gaszner M., Vazquez J., Schedl P. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction // Genes and development. – 1999. – V. 13. – № 16. – P. 2098–2107.

93. Hart C.M., Zhao K., Laemmli U.K. The scs' boundary element: characterization of boundary element-associated factors // Molecular and cellular biology. – 1997. – V. 17. – № 2. – P. 999–1009.

94. Gerasimova T.I. et al. A drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation // Cell. – 1995. – V. 82. – № 4. – P. 587–597.

95. Schweinsberg S. et al. The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila* bithorax complex requires GAGA-factor-binding sites // Genetics. – 2004. – V. 168. – № 3. – P. 1371–1384.

96. Ghosh D., Gerasimova T.I., Corces V.G. Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function // The EMBO journal. – 2001. – V. 20. – № 10. – P. 2518–2527.

97. Pai C.-Y.Y. et al. The Centrosomal Protein CP190 Is a Component of the gypsy Chromatin Insulator. // Molecular Cell. – 2004. – V. 16. – № 5. – P. 737–748.

98. Mohan M. et al. The *Drosophila* insulator proteins CTCF and CP190 link enhancer blocking to body patterning. // The EMBO Journal. – 2007. – V. 26. – № 19. – P. 4203–4214.

99. Kaushal A. et al. Essential role of Cp190 in physical and regulatory

boundary formation // Science Advances. – 2022. – V. 8. – № 19. – P. eabl8834.

100. Whitfield W.G. et al. Cloning of a gene encoding an antigen associated with the centrosome in *Drosophila*. // Journal of cell science. – 1988. – V. 89. – P. 467–480.

101. Whitfield W.G.F. et al. The 190 kDa centrosome-associated protein of *Drosophila melanogaster* contains four zinc finger motifs and binds to specific sites on polytene chromosomes. // Journal of Cell Science. – 1995. – V. 108. – № 11. – P. 3377–3387.

102. Oegema K. et al. The cell cycle-dependent localization of the CP190 centrosomal protein is determined by the coordinate action of two separable domains // Genes and Development. – 2011. – V. 6. – № 5. – P. 1261–1273.

103. Butcher R.D.J. et al. The *Drosophila* centrosome-associated protein CP190 is essential for viability but not for cell division // Journal of Cell Science. – 2004. – V. 117. – № 7. – P. 1191–1199.

104. Blanton J., Gaszner M., Schedl P. Protein:protein interactions and the pairing of boundary elements in vivo // Genes and development. – 2003. – V. 17. – № 5. – P. 664–675.

105. Kyrchanova O. et al. Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the *Drosophila* Abd-B gene // Molecular and cellular biology. – 2008. – V. 28. – № 12. – P. 4188–4195.

106. Bantignies F. et al. Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in *Drosophila* // Cell. – 2011. – V. 144. – № 2. – P. 214–226.

107. Bartkuhn M. et al. Active promoters and insulators are marked by the centrosomal protein 190 // The EMBO Journal. – 2009. – V. 28. – № 7. – P. 877–888.

108. Bushey A.M., Ramos E., Corces V.G. Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions // Genes and Development. – 2009. – V. 23. – № 11. – P. 1338–1350.

109. Maksimenko O. et al. Two new insulator proteins, Pita and ZIPIC,

target CP190 to chromatin. // *Genome Research*. – 2015. – V. 25. – № 1. – P. 89–99.

110. Cuartero S. et al. Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. // *The EMBO Journal*. – 2014. – V. 33. – № 6. – P. 637–647.

111. Bonchuk A. et al. *Drosophila* BTB/POZ Domains of ‘ttk Group’ Can Form Multimers and Selectively Interact with Each Other. // *Journal of Molecular Biology*. – 2011. – V. 412. – № 3. – P. 423–436.

112. Golovnin A. et al. Integrity of the Mod(mdg4)-67.2 BTB domain is critical to insulator function in *Drosophila melanogaster*. // *Molecular and Cellular Biology*. – 2007. – V. 27. – № 3. – P. 963–974.

113. Moshkovich N. et al. RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function // *Genes and Development*. – 2011. – V. 25. – № 16. – P. 1686–1701.

114. Liang J. et al. Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing // *Molecular Cell*. – 2014. – V. 53. – № 4. – P. 672–681.

115. Comet I. et al. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber // *Proceedings of the National Academy of Sciences of the United States of America*. – 2011. – V. 108. – № 6. – P. 2294–2299.

116. Kyrchanova O. et al. Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements // *Nucleic Acids Research*. – 2008. – V. 36. – № 22. – P. 7019–7028.

117. Chetverina D. et al. Red flag on the white reporter: a versatile insulator abuts the white gene in *Drosophila* and is omnipresent in mini-white constructs // *Nucleic Acids Research*. – 2008. – V. 36. – № 3. – P. 929–937.

118. Yokoshi M., Segawa K., Fukaya T. Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication // *Molecular Cell*. – 2020. – V. 78. – № 2. – P. 224–235.e5.

119. Chen H. et al. Dynamic interplay between enhancer–promoter topology and gene activity // *Nature Genetics*. – 2018. – V. 50. – № 9. – P. 1296–1303.
120. Fujioka M. et al. Determinants of Chromosome Architecture: Insulator Pairing in cis and in trans // *PLOS Genetics*. – 2016. – V. 12. – № 2. – P. e1005889.
121. Lim B. et al. Visualization of Transvection in Living *Drosophila* Embryos // *Molecular Cell*. – 2018. – V. 70. – № 2. – P. 287-296.e6.
122. Heurteau A. et al. Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes // *Genome Biology*. – 2020. – V. 21. – № 1. – P. 193.
123. Ho J.W. et al. Comparative analysis of metazoan chromatin organization. // *Nature*. – 2014. – V. 512. – № 7515. – P. 449–452.
124. Gerasimova T.I. et al. Coordinated control of dCTCF and gypsy chromatin insulators in *Drosophila*. // *Molecular Cell*. – 2007. – V. 28. – № 5. – P. 761–772.
125. Filion G.J. et al. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in *Drosophila* Cells // *Cell*. – 2010. – V. 143. – P. 212–224.
126. Steffen P.A., Ringrose L. What are memories made of? How polycomb and trithorax proteins mediate epigenetic memory // *Nature Reviews Molecular Cell Biology*. – 2014. – V. 15. – № 5. – P. 340–356.
127. Bowman S.K. et al. H3K27 modifications define segmental regulatory domains in the *Drosophila* bithorax complex. // *eLife*. – 2014. – V. 3. – P. e02833.
128. Savitsky M. et al. Distinct Roles of Chromatin Insulator Proteins in Control of the *Drosophila* Bithorax Complex. // *Genetics*. – 2016. – V. 202. – № 2. – P. 601–617.
129. Ahanger S.H. et al. Ectopically tethered CP190 induces large-scale chromatin decondensation. // *Scientific reports*. – 2014. – V. 4. – P. 3917.
130. Kwon S.Y. et al. Genome-Wide Mapping Targets of the Metazoan

Chromatin Remodeling Factor NURF Reveals Nucleosome Remodeling at Enhancers, Core Promoters and Gene Insulators // PLoS Genetics. – 2016. – V. 12. – № 4. – P. 1–26.

131. Ali T. et al. Chromatin binding of Gcn5 in *Drosophila* is largely mediated by CP190 // Nucleic Acids Research. – 2017. – V. 45. – № 5. – P. 2384–2395.

132. Korenjak M. et al. DREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes // Nucleic Acids Research. – 2014. – V. 42. – № 14. – P. 8939–8953.

133. Gerland T.A. et al. The *Drosophila* speciation factor HMR localizes to genomic insulator sites // PLOS ONE. – 2017. – V. 12. – № 2. – P. e0171798.

134. Ulianov S.V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains // Genome Research. – 2016. – V. 26. – № 1. – P. 70–84.

135. Stadler M.R., Haines J.E., Eisen M.B. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo // Elife. – 2017. – V. 17. – № 6. – P. e29550.

136. Eagen K.P., Hartl T.A., Kornberg R.D. Stable Chromosome Condensation Revealed by Chromosome Conformation Capture // Cell. – 2015. – V. 163. – № 4. – P. 934–946.

137. Handu M. et al. SUMO-Enriched Proteome for *Drosophila* Innate Immune Response. // G3 : genes - genomes - genetics. – 2015. – V. 5. – № 10. – P. 2137–2154.

138. Jox T. et al. *Drosophila* CP190- and dCTCF-mediated enhancer blocking is augmented by SUMOylation. // Epigenetics and chromatin. – 2017. – V. 10. – № 1. – P. 32.

139. Ravarani C.N.J. et al. Molecular determinants underlying functional innovations of TBP and their impact on transcription initiation // Nature Communications. – 2020. – V. 11. – № 1. – P. 2384.

140. Dynlacht B.D., Hoey T., Tjian R. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation // *Cell*. – 1991. – V. 66. – № 3. – P. 563–576.
141. Chalkley G.E. DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the Initiator // *The EMBO Journal*. – 1999. – V. 18. – № 17. – P. 4835–4845.
142. Reeve J.N. Archaeal chromatin and transcription // *Molecular Microbiology*. – 2003. – V. 48. – № 3. – P. 587–598.
143. Gershenzon N.I., Trifonov E.N., Ioshikhes I.P. The features of *Drosophila* core promoters revealed by statistical analysis // *BMC Genomics*. – 2006. – V. 7. – № 1. – P. 161.
144. Zehavi Y. et al. Core Promoter Functions in the Regulation of Gene Expression of *Drosophila* Dorsal Target Genes // *Journal of Biological Chemistry*. – 2014. – V. 289. – № 17. – P. 11993–12004.
145. Burley S.K. The TATA box binding protein // *Current Opinion in Structural Biology*. – 1996. – V. 6. – № 1. – P. 69–75.
146. Hansen S.K. et al. Transcription properties of a cell type-specific TATA-binding protein, TRF // *Cell*. – 1997. – V. 91. – № 1. – P. 71–83.
147. Takada S. et al. A TRF1:BRF Complex Directs *Drosophila* RNA Polymerase III Transcription // *Cell*. – 2000. – V. 101. – № 5. – P. 459–469.
148. Holmes M.C., Tjian R. Promoter-Selective Properties of the TBP-Related Factor TRF1 // *Science*. – 2000. – V. 288. – № 5467. – P. 867–870.
149. Isogai Y. et al. Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription // *EMBO Journal*. – 2007. – V. 26. – № 1. – P. 79–89.
150. Verma N. et al. Differential Utilization of TATA Box-binding Protein (TBP) and TBP-related Factor 1 (TRF1) at Different Classes of RNA Polymerase III Promoters // *Journal of Biological Chemistry*. – 2013. – V. 288. – № 38. – P. 27564–27570.
151. Vo Ngoc L., Kassavetis G.A., Kadonaga J.T. The RNA Polymerase II

Core Promoter in *Drosophila* // Genetics. 2019. – V. 212. – № 1. – P. 13–24.

152. Dantonel J.C. et al. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. Elegans* // Molecular Cell. – 2000. – V. 6. – № 3. – P. 715–722.

153. Moore P.A. et al. A Human TATA Binding Protein-Related Protein with Altered DNA Binding Specificity Inhibits Transcription from Multiple Promoters and Activators // Molecular and Cellular Biology. – 1999. – V. 19. – № 11. – P. 7610–7620.

154. Ohbayashi T. et al. Isolation of cDNA, chromosome mapping, and expression of the human TBP-like protein // Biochemical and Biophysical Research Communications. – 1999. – V. 255. – № 1. – P. 137–142.

155. Ohbayashi T., Makino Y., Tamura T.A. Identification of a mouse TBP-like protein (TLP) distantly related to the *Drosophila* TBP-related factor // Nucleic Acids Research. – 1999. – V. 27. – № 3. – P. 750–755.

156. Rabenstein M.D. et al. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family // Proceedings of the National Academy of Sciences of the United States of America. – 1999. – V. 96. – № 9. – P. 4791–4796.

157. Teichmann M. et al. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with HTFIIa in HeLa cells // Proceedings of the National Academy of Sciences of the United States of America. – 1999. – V. 96. – № 24. – P. 13720–13725.

158. Kopytova D.V. et al. Two Isoforms of *Drosophila* TRF2 Are Involved in Embryonic Development, Premeiotic Chromatin Condensation, and Proper Differentiation of Germ Cells of Both Sexes // Molecular and Cellular Biology. – 2006. – V. 26. – № 20. – P. 7492–7505.

159. Duttke S.H.C. et al. TRF2 and the evolution of the bilateria // Genes and Development. – 2014. – V. 28. – № 19. – P. 2071–2076.

160. Isogai Y. et al. Transcription of histone gene cluster by differential core-promoter factors // Genes and Development. – 2007. – V. 21. – № 22. – P.

2936–2949.

161. Guglielmi B., LaRoche N., Tjian R. Gene-specific transcriptional mechanisms at the histone gene cluster revealed by single-cell imaging // *Molecular Cell*. – 2013. – V. 51. – № 4. – P. 480–492.

162. Parry T.J. et al. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery // *Genes and Development*. – 2010. – V. 24. – № 18. – P. 2013–2018.

163. Wang Y.L. et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes // *Genes and Development*. – 2014. – V. 28. – № 14. – P. 1550–1555.

164. Kedmi A. et al. *Drosophila* TRF2 is a preferential core promoter regulator // *Genes and Development*. – 2014. – V. 28. – № 19. – P. 2163–2174.

165. Bashirullah A. et al. dTrf2 is required for transcriptional and developmental responses to ecdysone during *Drosophila* metamorphosis // *Developmental Dynamics*. – 2007. – V. 236. – № 11. – P. 3173–3179.

166. Zhou H. et al. Taf7l cooperates with Trf2 to regulate spermiogenesis // *Proceedings of the National Academy of Sciences of the United States of America*. – 2013. – V. 110. – № 42. – P. 16886–16891.

167. Biswas M. et al. Applications of InterPro in protein annotation and genome analysis. // *Briefings in bioinformatics*. – 2002. – V. 3. – № 3. – P. 285–295.

168. Kurshakova M.M. et al. TRF4, the novel TBP-related protein of *Drosophila melanogaster*, is concentrated at the endoplasmic reticulum and copurifies with proteins participating in the processes associated with endoplasmic reticulum // *Journal of Cellular Biochemistry*. – 2019. – V. 120. – № 5. – P. 7927–7939.

169. Лактионов П.П. et al. Генетическая система для мечения соматических и герминальных клеточных линий в гонадах *Drosophila melanogaster* // *Цитология*. – 2013. – Т. 55. – С. 185–189.

170. Lakso M. et al. Targeted oncogene activation by site-specific

recombination in transgenic mice. // Proceedings of the National Academy of Sciences. – 1992. – V. 89. – № 14. – P. 6232–6236.

171. Pindyurin A. – V. et al. Inducible DamID systems for genomic mapping of chromatin proteins in *Drosophila* // Nucleic Acids Research. – 2016. – V. 44. – № 12. – P. 5646–5657.

172. Naito Y. et al. CRISPRdirect: Software for designing CRISPR/Cas guide RNA with reduced off-target sites // Bioinformatics. – 2015. – V. 31. – № 7. – P. 1120–1123.

173. Gratz S.J. et al. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease // Genetics. – 2013. – V. 194. – № August. – P. 1029–1035.

174. Shivdasani A.A., Ingham P.W. Regulation of Stem Cell Maintenance and Transit Amplifying Cell Proliferation by TGF- β Signaling in *Drosophila* Spermatogenesis // Current Biology. – 2003. – V. 13. – № 23. – P. 2065–2072.

175. Bischof J. et al. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. // Proceedings of the National Academy of Sciences of the United States of America. – 2007. – V. 104. – P. 3312–3317.

176. Markstein M. et al. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. // Nature genetics. – 2008. – V. 40. – № 4. – P. 476–483.

177. Morris C.A., Benson E., White-Cooper H. Determination of gene expression patterns using in situ hybridization to *Drosophila* testes // Nature Protocols. – 2009. – V. 4. – № 12. – P. 1807–1819.

178. Solovei I., Cremer M. 3D-FISH on Cultured Cells Combined with Immunostaining // Methods in molecular biology (Clifton, N.J.). – 2010. – V. 659. – P. 117–126.

179. Golovnin A., Volkov I., Georgiev P. SUMO conjugation is required for the assembly of *Drosophila* Su(Hw) and Mod(mdg4) into insulator bodies that facilitate insulator complex formation // Journal of Cell Science. – 2012. – V. 125.

– № 8. – P. 2064–2074.

180. Greil F., Moorman C., van Steensel B. DamID: Mapping of In Vivo Protein-Genome Interactions Using Tethered DNA Adenine Methyltransferase // *Methods in Enzymology*. – 2006. – V. 410. – № 06. – P. 342–359.

181. Maksimov D. a., Laktionov P.P., Belyakin S.N. Data analysis algorithm for DamID-seq profiling of chromatin proteins in *Drosophila melanogaster* // *Chromosome Research*. – 2016. № October. – P. 1–14.

182. Kalashnikova D.A. et al. SetDB1 and Su(var)3-9 play non-overlapping roles in somatic cell chromosomes of *Drosophila melanogaster* // *Journal of Cell Science*. – 2021. – V. 134. – № 2.

183. Sher N. et al. Developmental control of gene copy number by repression of replication initiation and fork progression // *Genome Research*. – 2012. – V. 22. – № 1. – P. 64–75.

184. Giardine B. et al. Galaxy: A platform for interactive large-scale genome analysis // *Genome Research*. – 2005. – V. 15. – № 10. – P. 1451–1455.

185. Kim D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. // *Genome biology*. – 2013. – V. 14. – № 4. – P. R36.

186. Trapnell C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation // *Nature Biotechnology*. – 2010. – V. 28. – № 5. – P. 511–515.

187. Kim D. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype // *Nature Biotechnology*. – 2019. – V. 37. – № 8. – P. 907–915.

188. Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features // *Bioinformatics*. – 2014. – V. 30. – № 7. – P. 923–930.

189. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 // *Genome Biology*. – 2014. – V. 15. – № 12. – P. 550.

190. Stephens M. False discovery rates: a new deal // *Biostatistics*. – 2016. – V. 18. – № 2. – P. 275–294.
191. Ignatiadis N. et al. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing // *Nature Methods*. – 2016. – V. 13. – № 7. – P. 577–580.
192. Brown J.B. et al. Diversity and dynamics of the *Drosophila* transcriptome // *Nature*. – 2014. – V. 512. – № 7515. – P. 393–399.
193. de Hoon M.J.L. et al. Open source clustering software // *Bioinformatics*. – 2004. – V. 20. – № 9. – P. 1453–1454.
194. Li H. et al. Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly // *Science*. – 2022. – V. 375. – № 6584. – P. eabk2432.
195. Hao Y. et al. Integrated analysis of multimodal single-cell data // *Cell*. – 2021. – V. 184. – № 13. – P. 3573–3587.e29.
196. Visser I., Speekenbrink M. depmixS4: An R package for hidden markov models // *Journal of Statistical Software*. 2010. – V. 36. – № 7. – P. 1–21.
197. Ramírez F. et al. deepTools2: a next generation web server for deep-sequencing data analysis // *Nucleic Acids Research*. – 2016. – V. 44. – № W1. – P. W160–W165.
198. Akbari O.S. et al. An Entry/Gateway cloning system for general expression of genes with molecular tags in *Drosophila melanogaster*. // *BMC Cell Biology*. – 2009. – V. 10. – № 1. – P. 8.
199. Evans C.J. et al. G-TRACE: rapid Gal4-based cell lineage analysis in *Drosophila* // *Nature Methods*. – 2009. – V. 6. – № 8. – P. 603–605.
200. Lee T., Luo L. Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development // *Trends in Neurosciences*. – 2001. – V. 24. – № 5. – P. 251–254.
201. Hu Y. et al. DRscDB: A single-cell RNA-seq resource for data mining and data comparison across species // *Computational and Structural Biotechnology Journal*. – 2021. – V. 19. – P. 2018–2026.
202. Celniker S.E. et al. Unlocking the secrets of the genome // *Nature*. –

2009. – V. 459. – № 7249. – P. 927–930.

203. modENCODE Consortium et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. // Science. – 2010. – V. 330. – № 6012. – P. 1787–1797.

204. Nègre N. et al. A comprehensive map of insulator elements for the *Drosophila* genome. // PLoS Genetics. – 2010. – V. 6. – № 1. – P. e1000814.

205. Petrenko N., Struhl K. Comparison of transcriptional initiation by RNA polymerase II across eukaryotic species // eLife. – 2021. – V. 10. – P. e67964

206. Bailey T.L. DREME: motif discovery in transcription factor ChIP-seq data // Bioinformatics. – 2011. – V. 27. – № 12. – P. 1653–1659.

207. Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences // Journal of Molecular Biology. – 1990. – V. 212. – № 4. – P. 563–578.

208. White-Cooper H. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the *Drosophila* testis // Spermatogenesis. – 2012. – V. 2. – № 1. – P. 11–22.

209. Dietzl G. et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila* // Nature. – 2007. – V. 448. – № 7150. – P. 151–156.

210. Perkins L.A. et al. The Transgenic RNAi Project at Harvard Medical School: Resources and Validation // Genetics. – 2015. – V. 201. – № 3. – P. 843–852.

211. Rørth P. Gal4 in the *Drosophila* female germline // Mechanisms of Development. – 1998. – V. 78. – № 1–2. – P. 113–118.

212. DeLuca S.Z., Spradling A.C. Efficient Expression of Genes in the *Drosophila* Germline Using a UAS Promoter Free of Interference by Hsp70 piRNAs // Genetics. – 2018. – V. 209. – № 2. – P. 381–387.

213. Xu T., Rubin G.M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues // Development. – 1993. – V. 117. – № 4. – P. 1223–1237.

214. Lee T., Luo L. Mosaic Analysis with a Repressible Cell Marker for

Studies of Gene Function in Neuronal Morphogenesis // *Neuron*. – 1999. – V. 22. – № 3. – P. 451–461.

215. Stowers R.S., Schwarz T.L. A Genetic Method for Generating *Drosophila* Eyes Composed Exclusively of Mitotic Clones of a Single Genotype // *Genetics*. – 1999. – V. 152. – № 4. – P. 1631–1639.

216. Yu H.-H. et al. Twin-spot MARCM to reveal the developmental origin and identity of neurons // *Nature Neuroscience*. – 2009. – V. 12. – № 7. – P. 947–953.

217. Zhou Q., Neal S.J., Pignoni F. Mutant analysis by rescue gene excision: New tools for mosaic studies in *Drosophila* // *Genesis*. – 2016. – V. 54. – № 11. – P. 589–592.

218. Ou H., Lei T. A novel strategy for conditional gene knockout based on Φ C31 integrase and Gal4/UAS system in *Drosophila* // *IUBMB Life*. – 2013. – V. 65. – № 2. – P. 144–148.

219. Frickenhaus M. et al. Highly efficient cell-type-specific gene inactivation reveals a key function for the *Drosophila* FUS homolog cabeza in neurons // *Scientific Reports*. – 2015. – V. 5. – № 1. – P. 9107.

220. Jin M. et al. Conditional knockout of retinal determination genes in differentiating cells in *Drosophila* // *The FEBS Journal*. – 2016. – V. 283. – № 15. – P. 2754–2766.

221. Chathoth K.T. et al. The role of insulators and transcription in 3D chromatin organization of flies // *Genome Research*. – 2022. – V. 32. – № 4. – P. 682–698.

222. Ilyin A.A. et al. Comparison of genome architecture at two stages of male germline cell differentiation in *Drosophila* // *Nucleic Acids Research*. – 2022. – V. 50. – № 6. – P. 3203–3225.

223. Splinter E. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus // *Genes and Development*. – 2006. – V. 20. – № 17. – P. 2349–2354.

224. Li H.-B. et al. Insulators, not Polycomb response elements, are

required for long-range interactions between Polycomb targets in *Drosophila melanogaster* // Molecular and cellular biology. – 2011. – V. 31. – № 4. – P. 616–625.

225. Matzat L.H. et al. Tissue-specific regulation of chromatin insulator function. // PLoS Genetics. – 2012. – V. 8. – № 11. – P. e1003069.

226. Seum C. et al. *Drosophila* SETDB1 Is Required for Chromosome 4 Silencing // PLoS Genetics. – 2007. – V. 3. – № 5. – P. e76.

227. Sienski G. et al. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery // Genes and Development. – 2015. – V. 29. – № 21. – P. 2258–2271.

228. Fei Q. et al. SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells // Genome Research. – 2015. – V. 25. – № 9. – P. 1325–1335.

229. Lanouette S. et al. The functional diversity of protein lysine methylation // Molecular Systems Biology. – 2014. – V. 10. – № 4. – P. 724.

230. Moore K.E., Gozani O. An unexpected journey: Lysine methylation across the proteome // Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. – 2014. – V. 1839. – № 12. – P. 1395–1403.

231. Zhang X., Wen H., Shi X. Lysine methylation: beyond histones // Acta Biochimica et Biophysica Sinica. – 2012. – V. 44. – № 1. – P. 14–27.

232. Van Duyne R. et al. Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR // Retrovirology. – 2008. – V. 5. – № 1. – P. 40.

233. Fei Q. et al. Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53 // Nature Communications. – 2015. – V. 6. – № 1. – P. 8651.

234. Guo J. et al. AKT methylation by SETDB1 promotes AKT kinase activity and oncogenic functions // Nature Cell Biology. – 2019. – V. 21. – № 2. – P. 226–237.

235. Wang G. et al. SETDB1-mediated methylation of Akt promotes its

K63-linked ubiquitination and activation leading to tumorigenesis // *Nature Cell Biology*. – 2019. – V. 21. – № 2. – P. 214–225.

236. Maksimov D.A., Koryakov D.E. Binding of SU(VAR)3-9 Partially Depends on SETDB1 in the Chromosomes of *Drosophila melanogaster* // *Cells*. – 2019. – V. 8. – № 9. – P. 1030.

237. Jiang Y. et al. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain // *Nature Genetics*. – 2017. – V. 49. – № 8. – P. 1239–1250.

238. Bharadwaj R. et al. Conserved Higher-Order Chromatin Regulates NMDA Receptor Gene Expression and Cognition // *Neuron*. – 2014. – V. 84. – № 5. – P. 997–1008.

239. Warriar T. et al. SETDB1 acts as a topological accessory to Cohesin via an H3K9me3-independent, genomic shunt for regulating cell fates // *Nucleic Acids Research*. – 2022. – V. 50. – № 13. – P. 7326-7349

240. Falender A.E. et al. Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID // *Genes and Development*. – 2005. – V. 19. – № 7. – P. 794–803.

241. Ozer J., Moore P.A., Lieberman P.M. A Testis-specific Transcription Factor IIA (TFIIA τ) Stimulates TATA-binding Protein-DNA Binding and Transcription Activation // *Journal of Biological Chemistry*. – 2000. – V. 275. – № 1. – P. 122–128.

242. Martianov I. et al. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression // *Scientific Reports*. – 2016. – V. 6. – № 1. – P. 32069.