

На правах рукописи

ЛАКТИОНОВ ПЕТР ПАВЛОВИЧ

**РОЛЬ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ COOKIE MONSTER И
CANNONBALL В РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНОВ В
СПЕРМАТОГЕНЕЗЕ
*DROSOPHILA MELANOGASTER***

Молекулярная генетика – 03.01.07

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Новосибирск – 2013

Работа выполнена в лаборатории геномики Федерального государственного бюджетного учреждения науки Институт молекулярной и клеточной биологии СО РАН, г. Новосибирск, Россия.

Научный руководитель: **Белякин Степан Николаевич**
кандидат биологических наук,
зав. лабораторией геномики,
Федеральное государственное бюджетное
учреждение науки Институт молекулярной и
клеточной биологии СО РАН, г. Новосибирск

Официальные оппоненты: **Омельянчук Леонид Владимирович**,
доктор биологических наук,
зав. лабораторией генетики клеточного цикла,
Федеральное государственное бюджетное
учреждение науки Институт молекулярной и
клеточной биологии СО РАН, г. Новосибирск

Медведев Сергей Петрович,
кандидат биологических наук,
старший научный сотрудник,
Федеральное государственное бюджетное
учреждение науки Институт цитологии и генетики
СО РАН, г. Новосибирск

Ведущее учреждение: Федеральное государственное бюджетное
учреждение науки Институт молекулярной
генетики Российской академии наук, г. Москва

Защита диссертации состоится «___» _____ 2013 г. на утреннем заседании
диссертационного совета Д 003.074.01 на базе Федерального государственного
бюджетного учреждения науки Институт молекулярной и клеточной биологии СО
РАН по адресу:

630090, г. Новосибирск, проспект ак. Лаврентьева, 8/2, т./ф. (383)363-90-45, факс
(383) 363-90-78, e-mail: kokoza@mcb.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИМКБ СО РАН.

Автореферат разослан «___» _____ 2013 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук

Е.Б. Кокоза

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Одним из главных направлений функциональной геномики является изучение регуляторных систем, управляющих генетическими и эпигенетическими программами развития организма. Сперматогенез у *Drosophila melanogaster* является удобной моделью для изучения регуляции экспрессии генов в процессе клеточной дифференцировки. В семеннике дрозофилы можно легко различить все стадии развития клеток зародышевого пути – образование клеток из стволовых предшественников, пролиферацию и терминальную дифференцировку. Терминальная дифференцировка клеток мужского зародышевого пути у *D. melanogaster* сопровождается перестройкой хроматина, за которой следует активация генов, необходимых для сперматогенеза. Активация генов дифференцировки контролируется несколькими специализированными транскрипционными факторами, которые кодируются так называемыми генами задержки мейоза и которые можно разделить на два класса - *aly* и *can*. Белковые продукты генов *aly* класса образуют комплекс tMAC (testis-specific meiotic arrest complex), а гены *can* класса являются семенник-специфичными паралогами (tTAF, testis-specific TBP-associated factors) генов TAF у *D. melanogaster*. Мутации генов задержки мейоза приводят к инактивации множества генов в семенниках, неспособности клеток зародышевого пути вступить в мейоз и остановке гаметогенеза на стадии сперматоцитов. Было показано, что на фоне мутаций по генам задержки мейоза промоторные области генов дифференцировки остаются связанными белком-репрессором Polycomb, что по всей вероятности вызывает подавление экспрессии. В работе Chen et al. [2011] была сформулирована гипотеза, согласно которой активация генов в семенниках зависит от синхронизированного действия белков tMAC и tTAF. Это предположение основано на наблюдении о том, что в отсутствие компонентов tMAC нарушается связывание tTAF с ДНК. Однако существующие данные описывают конечный эффект мутаций, что не дает представления о механизме действия и конкретной роли этих транскрипционных факторов. В первую очередь остаются невыясненными гены-мишени белков, кодируемых генами задержки мейоза. Более того, до сих пор не определено, являются ли эти белки прямыми активаторами или подавляют активность генов-репрессоров в клетках зародышевого пути. Также остается открытым вопрос о том, каким образом эти транскрипционные факторы участвуют в эпигенетической регуляции активности генов. Ответы на эти вопросы позволят приблизиться к пониманию принципов работы регуляторных систем, управляющих генетическими программами дифференцировки клеток зародышевого пути *D. melanogaster*.

В представленной работе для исследования роли транскрипционных факторов, кодируемых генами задержки мейоза, на первом этапе были установлены гены-мишени транскрипционного фактора Cookie monster (Comr), кодируемого геном, относящимся к *aly* классу. Картирование Comr проводили методом тканеспецифичного DamID с последующим анализом методом микрочипов в масштабе генома. Анализ данных о локализации сайтов связывания Comr и экспрессии генов в семенниках мутантов по генам *cookie monster* (*aly* класс) и *cannonball* (*can* класс) позволил установить роль Comr и Cannonball (Can) в регуляции экспрессии генов при дифференцировке клеток мужского зародышевого пути и сформулировать гипотезу относительно механизмов активации генов, необходимых для дифференцировки мужских гамет у *D. melanogaster*.

Цели и задачи исследования. Целью данной работы являлось изучение роли транскрипционных факторов Comr и Cap в регуляции активности генов в процессе сперматогенеза у *D. melanogaster*. Для достижения этой цели были сформулированы следующие задачи:

1. Создать генетическую систему для картирования сайтов связывания транскрипционных факторов методом DamID в клетках мужского зародышевого пути у *D. melanogaster*.
2. Определить профиль связывания транскрипционного фактора Comr с ДНК герминальных клеток самцов *D. melanogaster*.
3. Определить эффект связывания Comr на экспрессию генов в сперматоцитах *D. melanogaster*.
4. Определить регуляторный эффект белка Cap на гены, связанные с белком Comr

Научная новизна. Разработан подход, позволяющий картировать сайты связывания хроматиновых белков с ДНК в отдельных клеточных популяциях в организме *D. melanogaster* методом DamID. Этим методом впервые были локализованы участки связывания транскрипционного фактора Comr в геноме и выявлены его гены-мишени в герминальных клетках самца *D. melanogaster*. Показано, что белок Comr является прямым активатором части своих генов-мишеней в сперматоцитах. Выявлен возможный вторичный регуляторный фактор, экспрессия которого контролируется Comr. Предложена модель, описывающая этапы активации семенник-специфичных генов в сперматоцитах *D. melanogaster* и роль белков tMAC и tTAF в этом процессе.

Научно-практическая ценность. Результаты исследования вносят вклад в современные представления о регуляторных системах, управляющих генетическими и эпигенетическими программами развития клеток. Полученные данные содержат принципиально новую информацию о механизме действия белков комплексов tMAC и tTAF, что значительно расширяет представления о механизмах регуляции генов в семенниках *D. melanogaster* процессе клеточной дифференцировки. Разработанная генетическая система для проведения тканеспецифичного DamID может быть использована для исследования транскрипционных факторов и иных ассоциированных с ДНК белков в клетках зародышевого пути у *D. melanogaster*. Полученная система универсальна и может быть легко адаптирована для исследования тканеспецифичных транскрипционных факторов в других органах *D. melanogaster*.

Основные положения диссертации, выносимые на защиту. Проведен анализ профиля связывания белка Comr в герминальных клетках самца *D. melanogaster*. Белок Comr является прямым активатором генов-мишеней в сперматоцитах. Установлено наличие неизвестных ранее вторичных генов-регуляторов транскрипции, находящихся под контролем белков Comr и Cap.

Личный вклад автора. Основные результаты получены автором самостоятельно. Работа по математической обработке данных выполнена в сотрудничестве с Белякиным С.Н. и Максимовым Д.А. Данные по транскрипции генов в семенниках предоставлены Dr. H. White-Cooper (Университет Кардиффа, Великобритания).

Апробация работы. Результаты работы представлены в виде стендовых сообщений: на международной конференции по системной биологии дрозофилы (Пултуск, Польша, 2012), на международной конференции «Хромосома-2012»

(Новосибирск, Россия, 2012), на международной конференции «Высокопроизводительное секвенирование в геномике» (Новосибирск, Россия, 2013); а также в устных докладах: на международной конференции по молекулярным механизмам регуляции сперматогенеза «ETW-16» (Эльба, Италия, 2010), на международной конференции «Хромосома-2012» (Новосибирск, Россия, 2012).

Публикации. Результаты исследования представлены в 3 статьях и 5 тезисах конференций.

Структура и объем работы. Диссертация состоит из введения, обзора литературы, описания материалов и методов, результатов, обсуждения, а также выводов и списка цитируемой литературы, в которые входит 167 ссылок. Работа изложена на 112 страницах печатного текста, содержит 1 таблицу и 12 рисунков.

МАТЕРИАЛЫ И МЕТОДЫ

Линии мух. Для проведения сайт-специфичной интеграции генетических конструкций использовали линии мух, несущие сайты интеграции attP40 и attP18, полученные из Bloomington Stock Center (США), описанные в базе данных Flybase (<http://flybase.org>) и работе Markstein et al. [2008]. Также использовались линии *Oregon R*, y^1w^{67} (фонд лаборатории), $uw; can^3 red e/TM6C, uw; cn comr^{z2-1340} bw/CyO$ (Предоставлены Dr. Н. White-Cooper, Университет Кардиффа, Великобритания).

Генетические конструкции. Все этапы молекулярного клонирования проводили по стандартным протоколам или в соответствии с рекомендациями фирм-производителей (БиоСилика, Qiagen, Fermentas, Promega). В ходе работы были созданы следующие конструкции: *nanos-cre*, *Ubipro>stop>mCD8-GFP*, *hsp70>stop>dam*, *hsp70>stop>dam-comr*, полные нуклеотидные последовательности которых можно найти в базе данных GenBank (www.ncbi.nlm.nih.gov/genbank/) под номерами KC845567, KC845568, JN993988 и KC845569, соответственно.

Трансформация генома дрозофилы. Введение генетических конструкций проводили при помощи системы attP/attB phiC31-опосредованной трансформации дрозофилы, которая описана в Bischof et al., [2007]. Трансформацию проводили по методике, описанной в Spradling, Rubin, [1982]. Конструкции *nanos-cre* и *Ubipro>stop>mCD8-GFP* были встроены в сайт интеграции attP40, а конструкции *hsp70>stop>dam-comr* и *hsp70>stop>dam* – в attP18.

Проведение DamID. По 200 пар семенников самцов генотипов $y^1, w^{67}, hsp70>stop>dam-comr/Y; nanos-cre/+$ (№1), $y^1, w^{67}, hsp70>stop>dam/Y; nanos-cre/+$ (№2) и $y^1, w^{67}, hsp70>stop>dam/y^1, w^{67}, hsp70>stop>dam; +/+$ (№3) собирали диссекцией в охлажденном буфере PBS. Экспериментальный образец ДНК был выделен из семенников самцов генотипа №1, образец, выделенный из семенников самцов генотипа №2, использовали для контроля фонового метилирования, ДНК семенников самцов генотипа №3 использовали в качестве отрицательного контроля. Процедуру DamID проводили согласно методике, описанной в Greil et al. [2006] с последующим анализом полученных результатов с использованием микрочипов *D. melanogaster* Gene Expression 4x72K Arrays (Nimblegen, Roche). Гибридизацию микрочипов проводили на базе университета г. Торонто (Canadian Drosophila Microarray Centre, <http://www.flyarrays.com/>). Эксперимент был повторен дважды на независимых биологических образцах. Полученные данные были обработаны с использованием программного

обеспечения ArrayStar (<http://www.dnastar.com>). После проведения коррекции фона и RMA-нормализации для всех точек был посчитан логарифм отношения интенсивностей сигнала в опытном и контрольном образце ($\text{Log}_2(\text{Dam-Comr}/\text{Dam})$). Результаты были отфильтрованы на воспроизводимость по методу, описанному в Yang et al. [2002]. Данные двух независимых воспроизведений эксперимента были усреднены и сопоставлены с геномными координатами. Интенсивности сигналов от точек, расположенных внутри каждого гена, были усреднены. Экспериментальные данные DamID доступны в базе данных GEO (<http://www.ncbi.nlm.nih.gov/geo/>) под номером GSE49799.

Для определения Comr-связанных, Comr-несвязанных и промежуточных геномных областей мы использовали Скрытую Марковскую модель с тремя состояниями (Hidden Markow Model, HMM). Геном *D. melanogaster* разбили на отрезки длиной 200 п.н., как описано в Kharchenko et al. [2011]. Для определения исходных параметров HMM геном разделили на участки в 5 т.п.н., перекрывающиеся на 1 т.п.н., для каждого из которых посчитали усредненную интенсивность сигнала ($\text{Log}_2(\text{Dam-Comr}/\text{Dam})$). Из общего числа выбрали по 2,5% участков, демонстрирующих самую низкую интенсивность сигнала, самую высокую и среднюю по геному, которые были использованы для определения Comr-несвязанных, Comr-связанных и индифферентных районов. Для оптимизации параметров HMM и предсказания по алгоритму Витерби использовали библиотеку «mhsmm» (O'Connell and Hojsgaard, 2011) в оболочке R (<http://r-project.org>).

Анализ экспрессии генов в семенниках самцов, мутантных по генам *comr* и *can* проводили методом микрочипов, как описано в Doggett et al. [2011]. Выделение, обработку РНК и гибридизацию кДНК проводили в Glasgow Affymetrix microarray service (<http://www.gla.ac.uk/faculties/fbls/functionalgenomicsfacility/>). Полученные данные были нормализованы с использованием метода Affymetrix t100. Данные для каждого гена представлены в виде логарифма по основанию 2 от отношения интенсивности сигнала в образце семенников мутантов *comr* и *can* к интенсивности в образце семенников линии дикого типа ($\text{Log}_2(\text{comr}/\text{wt})$, $\text{Log}_2(\text{can}/\text{wt})$). Экспериментальные данные доступны в базе данных GEO (<http://www.ncbi.nlm.nih.gov/geo/>) под номером GSE49563.

РЕЗУЛЬТАТЫ

Определение сайтов связывания белка Comr с помощью тканеспецифичного DamID. Для определения профиля связывания белка Comr с хромосомами в клетках мужского зародышевого пути нами был использован метод DamID, который позволяет картировать сайты связывания хромосомных белков без применения фиксации материала и является альтернативой методу хроматин-иммунопреципитации. Метод основан на использовании химерного полипептида, состоящего из исследуемого белка, соединенного с ДНК-аденин-метилтрансферазой (Dam) *E. coli*, способной метилировать аденин в последовательностях GATC. Экспрессия такого полипептида осуществляется при помощи генетической конструкции, введенной в клетку. Химерный полипептид направляется к сайтам связывания исследуемого белка на хромосомах, а присоединенная к нему Dam метилирует последовательности GATC в окрестности этого сайта. В геноме дрозофилы GATC последовательности встречаются в среднем каждые 200-300 п.н., что определяет высокую разрешающую способность метода. Метилированные фрагменты ДНК избирательно выделяют,

амплифицируют и используют в экспериментах с микрочипами для определения геномного расположения фрагментов из обогащенных фракций. Для коррекции фонового метилирования, через все стадии эксперимента проводится контрольный образец, экспрессирующий только Dam.

До настоящего времени, использование метода DamID для определения профиля связывания исследуемого белка в определенных клеточных линиях или тканях организма не представлялось возможным. Это обусловлено тем, что для корректной работы DamID требуется крайне низкий уровень экспрессии химерного белка – в противном случае резко возрастает фоновое метилирование и снижается специфичность метода. В предыдущих исследованиях эмпирически было показано, что поддержание низкой, но достаточной для проведения экспериментов, активности химерного белка обеспечивается базальной активностью промотора белка теплового шока *hsp70* при 25⁰С, которая, однако, проявляется во всех клетках организма. Это является препятствием для картирования тканеспецифичных транскрипционных факторов, поскольку химерный белок образуется во всех клетках организма, что может приводить к неконтролируемому искажению картины связывания и не позволить адекватно интерпретировать полученные результаты. Исследуемый транскрипционный фактор *Comr in vivo* экспрессируется исключительно в клетках мужского зародышевого пути, поэтому необходимо было разработать систему, которая бы позволила проводить DamID только в этих клетках. Поставленная задача была решена в два этапа: в ходе первого этапа была создана генетическая система, позволяющая ограничить экспрессию трансгенов популяцией клеток зародышевого пути и, таким образом, избежать неспецифического сигнала из других клеточных типов. В ходе второго этапа мы адаптировали полученную генетическую систему для проведения DamID в герминальных клетках *D. melanogaster*.

Для создания генетической системы для запуска эктопической экспрессии трансгенов исключительно в клетках зародышевого пути сконструировали и ввели в геном дрозофилы две конструкции. Первая конструкция содержала ген рекомбиназы, фланкированный регуляторными областями гена *nanos*, белковый продукт которого является специфичным маркером стволовых герминальных клеток в организме *D. melanogaster* (Рис. 1А). Во второй конструкции под контролем промотора гена *Ubi-p63E* был размещен ген репортерного белка mCD8-GFP. Промотор и репортерный ген были разделены стоп-кассетой, состоящей из терминатора транскрипции гена *hsp70AB*, фланкированного *loxP* сайтами узнавания рекомбиназы Cre (Рис. 1Б). Промотор гена *Ubi-p63E* позволяет активировать репортерный ген во всех клетках организма, однако экспрессия блокируется стоп-кассетой. Под действием рекомбиназы Cre происходит удаление стоп-касеты и запускается экспрессия репортерного гена.

Чтобы убедиться в том, что под контролем регуляторных областей *nanos* рекомбиназа Cre способна экспрессироваться и удалять стоп-кассету только в клетках зародышевого пути, мы скрестили линии мух, несущие полученные генетические конструкции. У самцов F₁ были выделены гонады, в которых детектировали флуоресценцию GFP. Было обнаружено, что в семенниках самцов генотипа $y^1w^{67}/Y; Ubipro>stop>mCD8-GFP/nanos-cre; +/-$ репортерный белок нарабатывался исключительно в клетках зародышевого пути и не экспрессировался в соматических клетках (Рис. 1Г). Флуоресцентный сигнал обнаруживался на стадии сперматогониев (Рис. 1Г, стрелка) и достигал максимума на стадии

первичных сперматоцитов. На рисунке отчетливо видны отдельные цисты, содержащие по 16 сперматоцитов (Рис. 1Г, штриховая линия), экспрессирующих mCD8-GFP. В семенниках самцов y^1w^{67}/Y ; $Ubi\text{pro}>stop>mCD8-GFP/Ubi\text{pro}>stop>mCD8-GFP$; +/+, используемых в качестве отрицательного контроля, и у которых между промотором и репортерным геном присутствует стоп-кассета, маркерный белок не активировался (Рис. 1Д). Таким образом было установлено, что конструкция *nanos-cre* обеспечивает наработку рекомбиназы и эффективное удаление стоп-касеты исключительно в клетках зародышевого пути.

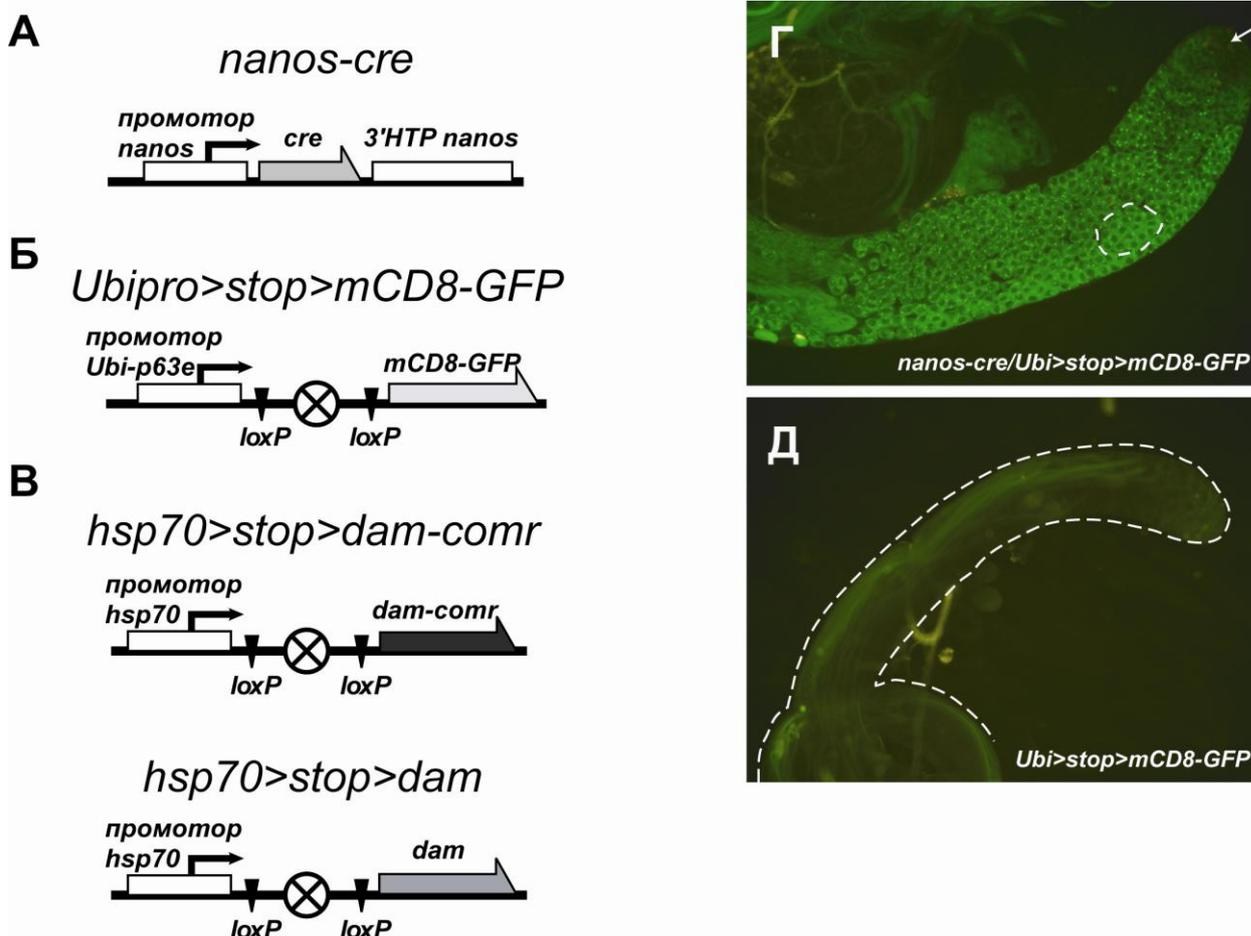
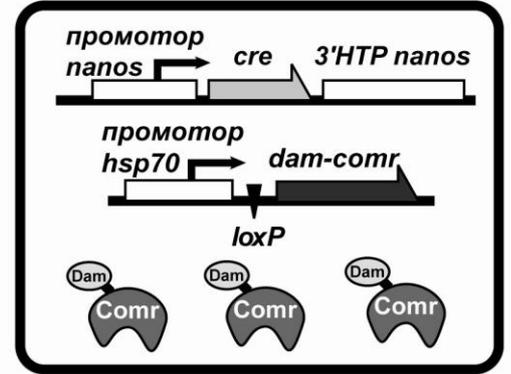
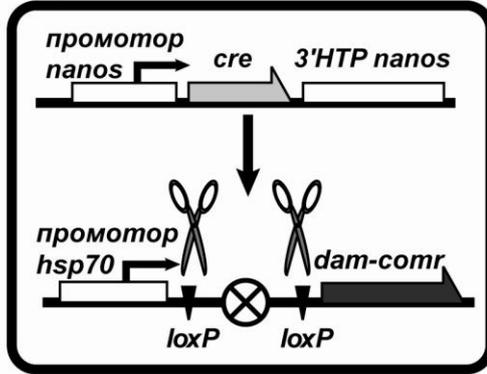


Рис. 1. Генетические конструкции для специфического мечения герминальных клеток самца и проведения тканеспецифичного DamID. (А) Конструкция *nanos-cre*; (Б) Конструкция $Ubi\text{pro}>stop>mCD8-GFP$; (В) Конструкции для DamID. Конструкции $hsp70>stop>dam-comr$ и $hsp70>stop>dam$. (Г) Семенник самца генотипа y^1w^{67}/Y ; $Ubi\text{pro}>stop>mCD8-GFP/nanos-cre$; +/+; (Д) Семенник самца генотипа y^1w^{67}/Y ; $Ubi\text{pro}>stop>mCD8-GFP/Ubi\text{pro}>stop>mCD8-GFP$; +/+ (см. объяснения в тексте).

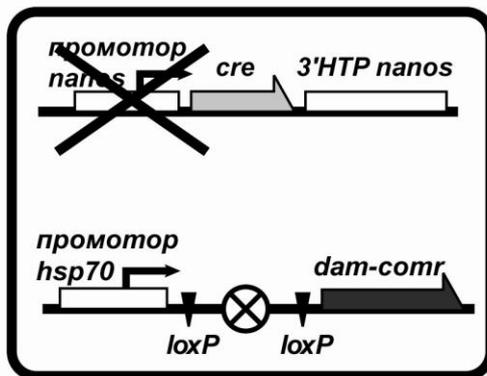
Для проведения тканеспецифичного DamID (Рис. 2) в геном дрозофилы ввели две генетические конструкции, одна из которых – описанная выше конструкция *nanos-cre* (Рис. 1А). Вторая конструкция содержала минимальный промотор гена *hsp70*, стоп-касету и последовательность, кодирующую либо химерный белок Dam-Comr ($hsp70>stop>dam-comr$), либо отдельно Dam ($hsp70>stop>dam$) (Рис. 1В). Конструкция $hsp70>stop>dam$ была использована в экспериментах для контроля фонового метилирования. Для проведения DamID

скрещивали самок линии мух, несущих конструкции для DamID с самцами линии $y^1, w^{67}/Y; nanos-cre/nanos-cre$.

А герминальные клетки



соматические клетки



Б

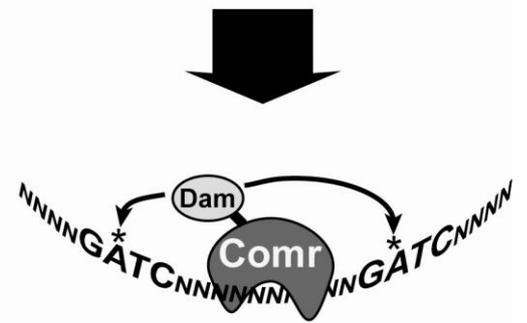


Рис. 2. Схема генетической системы для проведения DamID в герминальных клетках *D. melanogaster*. **(А)** В герминальных клетках нарабатывается рекомбиназа Cre (изображена в виде ножниц), которая удаляет терминатор транскрипции из конструкции *hsp70>stop>dam-comr*, что приводит к ее активации и наработке химерного белка Dam-Comr. В соматических клетках семенника конструкция *nanos-cre* неактивна и система не запускается. **(Б)** Химерный белок Dam-Comr направляется к генам-мишеням белка Comr, а Dam метилирует окружающие последовательности GATC в геномной ДНК.

У потомков от скрещиваний в клетках зародышевого пути происходило удаление терминатора и запускалась наработка малых количеств химерного белка (либо Dam) за счет базальной активности промотора *hsp70* (Рис. 2А). В результате метилирование ДНК происходило избирательно в клетках зародышевого пути (Рис. 2Б). С помощью этого подхода нами были получены данные о связывании транскрипционным фактором Comr для 11161 из 13653 всех аннотированных в базе данных Flybase генов *D. melanogaster* (<http://flybase.org>). После нормировки и усреднения данных оказалось, что 3120 генов проявили более чем двукратное превышение интенсивности сигнала в образцах Dam-Comr по сравнению с контрольными образцами Dam ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) > 1$). Эта выборка была значительно обогащена семенник-специфичными генами, а именно содержала 596 генов, что приблизительно в два раза выше, чем случайно ожидаемое значение (χ^2 , $P=3.9\text{E-}19$). В то же время представленность специфичных для яичников генов в

выборке *Comr*-связанных генов была более чем в два раза снижена по сравнению со случайно ожидаемым (75 генов, χ^2 , $P=2.5E-18$). Используемые здесь и далее в работе списки семенник-специфичных генов (1653 гена) и генов, специфически активирующихся в яичниках (907 гена), опубликованы в работе Belyakin et al. [2010].

Активирующая роль белка *Comr* в сперматоцитах. Чтобы выявить гены, экспрессия которых зависит от наличия функционального белка *Comr*, был проведен анализ представленности транскриптов в семенниках мутантов *comr*^{z2-1340} по сравнению с семенниками контрольной линии. Профиль экспрессии генов в семенниках линии *comr*^{z2-1340} получен с использованием микрочипов Affymetrix в лаборатории Dr Н. White-Cooper (*Drosophila* Genome 1.0 array; <http://affymetrix.com>). В результате этого анализа была охарактеризована экспрессия 12414 генов. Из них 2706 генов снижали экспрессию у мутантов более чем в 2 раза, а 978 генов – более чем в 8 раз. Чтобы оценить, на какие гены оказывает влияние мутация *comr*, был использован упомянутый выше список семенник-специфичных генов. В качестве контрольной группы использовали гены, специфически активные в яичниках дрозофилы. На Рис. 3А показаны распределения для этих двух категорий генов в зависимости от их экспрессии у мутантов *comr*. Третье распределение относится к генам, не попадающим ни в одну из этих категорий. Видно, что гены, активные только в яичниках, распределены симметрично относительно значения $\text{Log}_2(\text{comr}/\text{wt})=0$. Однако распределение семенник-специфичных генов сильно сдвинуто в сторону отрицательных значений $\text{Log}_2(\text{comr}/\text{wt})$, что свидетельствует об уменьшении их экспрессии у мутантов *comr*. На Рис. 3Б видно, что доля семенник-специфичных генов возрастает со снижением экспрессии у мутантов *comr*, в то время как доля генов, экспрессирующихся в яичниках, быстро падает. Таким образом, мутация по гену *comr* приводит к снижению уровня экспрессии множества генов, причем доля семенник-специфичных генов возрастает в группе, проявляющей более сильное снижение экспрессии.

Для того чтобы определить, является ли *Comr* прямым активатором генов в сперматоцитах *D. melanogaster*, или осуществляется иной механизм, мы сопоставили полученные данные по экспрессии с результатами DamID белка *Comr*. Мы сосредоточились на генах, которые существенно снижают экспрессию у мутантов *comr* и проанализировали их связывание с *Comr* согласно полученным данным DamID. Среди 978 генов, снижающих уровень экспрессии на фоне мутации более чем в 8 раз, данные DamID о связывании белком *Comr* имелись для 881 гена, из которых 700 оказались семенник-специфичными. Из этих 881 гена только 375 (42.5%) демонстрировали повышенный уровень связывания с транскрипционным фактором *Comr* ($\text{Log}_2(\text{Dam-Comr}/\text{Dam})>1$), причем 306 из них оказались семенник-специфичными.

С одной стороны, это говорит о том, что 375 генов проявляют связывание с белком *Comr*, присутствие которого необходимо для их активации, из чего можно сделать вывод, что он является их прямым активатором. С другой стороны, полученные нами результаты говорят о том, что далеко не все семенник-специфичные гены, экспрессия которых падает у мутантов *comr*, проявляют связывание с белком *Comr*. Это указывает на наличие вторичных регуляторов, экспрессия которых в свою очередь зависит от *Comr*.

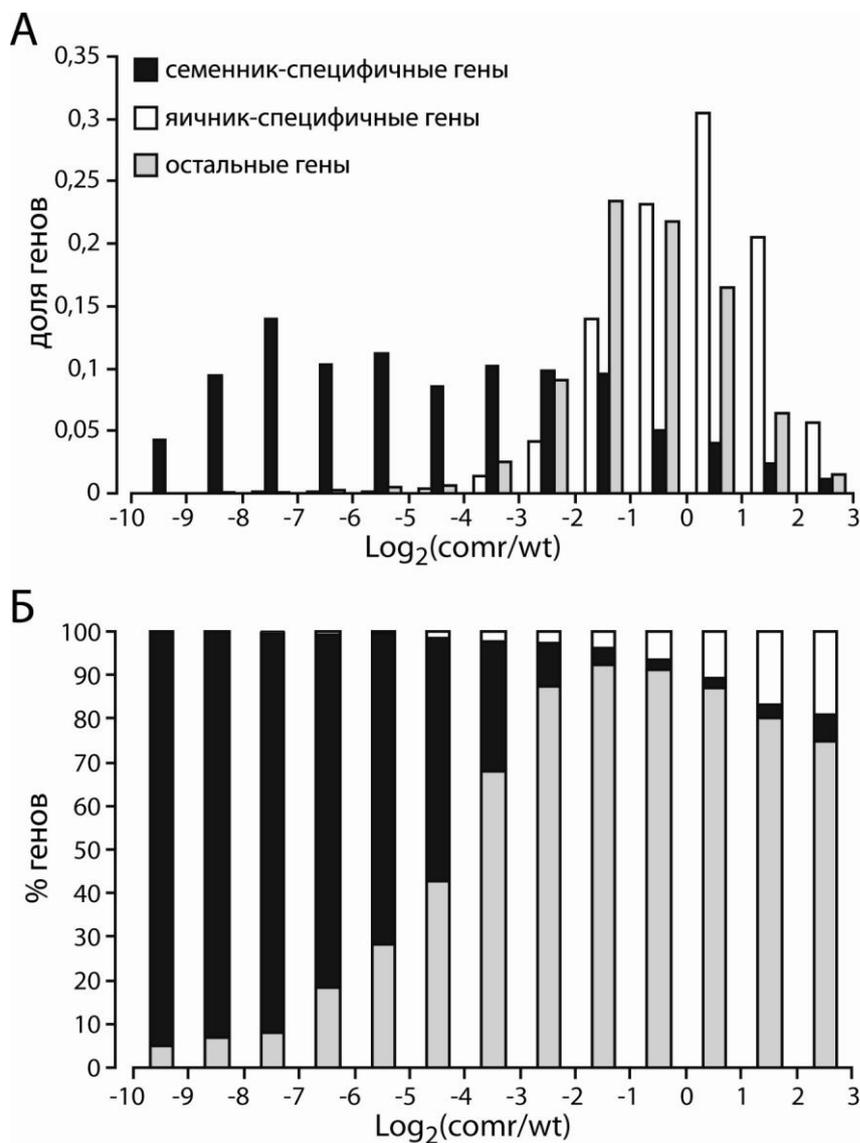


Рис. 3. Влияние мутации *comr* на экспрессию генов в семенниках *D. melanogaster*. **(А)** – Мутация *comr* снижает экспрессию семенник-специфичных генов (показаны черным). Для сравнения показаны гены, специфически активируемые в яичниках (показаны белым) и остальные гены *D. melanogaster* (показаны серым). **(Б)** – Доля семенник-специфичных генов увеличивается среди генов, снижающих экспрессию у мутантов *comr*. Гены, специфически активные в яичниках, демонстрируют обратную тенденцию.

Идентификация вторичных Comr-зависимых регуляторов транскрипции. Около 6% от всего числа генов *D. melanogaster* не связывались транскрипционным фактором Comr ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) < 0$), но демонстрировали пониженный уровень экспрессии в семенниках *comr*-мутантов ($\text{Log}_2(\text{comr}/\text{wt}) < -1$). Среди семенник-специфичных генов их доля составляла уже около 19%. Эти гены зависят от действия Comr, но не являются его прямыми мишенями. Регуляция таких генов может осуществляться вторичными посредниками, то есть семенник-специфичными транскрипционными факторами, которые сами регулируются белком Comr. В первую очередь мы обратили внимание на гены задержки мейоза. Оказалось, что уровень их экспрессии не снижается в семенниках мутантов *comr*

по сравнению с семенниками самцов дикого типа, так что, по-видимому, не они играют роль вторичных регуляторов в активации *Comr*-зависимых генов.

Поиск генов, кодирующих семенник-специфичные транскрипционные факторы, был проведен с использованием ресурса GoMiner, описанного в Zeeberg et al. [2003]. Среди 1653 семенник-специфичных генов было обнаружено 26 генов, имеющих отношение к регуляции транскрипции, включая гены *aly* и *can* классов. Сопоставление данных об экспрессии этих генов у мутантов *comr* и о связывании с ними белка *Comr* выявило, что ген *CG9879* проявил самый высокий уровень связывания в экспериментах DamID, а уровень его экспрессии был снижен в 64 раза в семенниках самцов, несущих мутацию гена *comr*. Ген *CG9879* не исследован, но согласно аннотации в базе данных FlyBase (www.flybase.org), кодируемый им белок имеет гомологию с ТАТА-связывающим белком (ТАТА-binding protein, TBP). Эти данные указывают на то, что ген *CG9879* может являться вторичным регулятором в генетическом каскаде, приводящем к массовой активации генов в сперматоцитах.

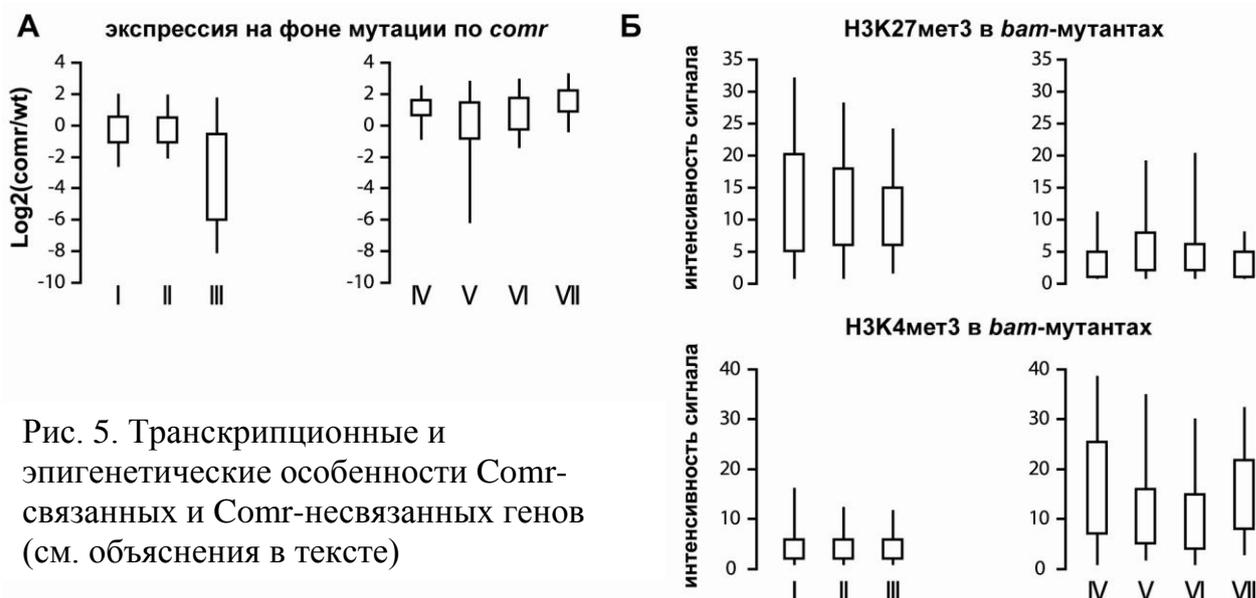
Связывание *Comr* не всегда приводит к изменению экспрессии генов.

Среди генов, проявляющих повышенное связывание с белком *Comr*, помимо семенник-специфичных были обнаружены гены, активные в яичниках, но не в семенниках. Это может означать, что связывание белка *Comr* не обязательно приводит к активации генов. Для исследования этого вопроса мы проанализировали выборку генов, значительно обогащенных по связыванию с белком *Comr* ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) > 2$). Для 749 из 1029 таких генов имелись экспериментальные данные об экспрессии в семенниках самцов *comr*^{z2-1340}. Из этих генов только у 320 (43%) генов наблюдалось более чем двукратное снижение уровня экспрессии на фоне мутации по *comr*, а около половины *Comr*-связанных генов не изменяли уровень экспрессии на фоне мутации по *comr*. Из этого следует, что приблизительно в половине случаев связывание *Comr* не влияет на активность генов.

Для исследования состава выборки *Comr*-связанных генов мы провели анализ их экспрессии при помощи базы данных FlyAtlas, которая содержит профили экспрессии всех генов *D. melanogaster* в различных органах. Группа *Comr*-связанных ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) > 2$) состояла из 1029 генов, в качестве контроля использовали выборку *Comr*-несвязанных ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) < -2$) генов (698 генов). К профилям экспрессии этих генов мы применили метод кластеризации k-means clustering и выделили подгруппы генов, обладающих схожими профилями экспрессии. Среди *Comr*-связанных генов было выделено три таких подгруппы. Подгруппа III состояла из 323 семенник-специфичных генов, уровень их экспрессии снижался в семенниках самцов *comr*^{z2-1340} по сравнению с семенниками дикого типа (Рис. 5А). Подгруппа I содержала специфичные для нервной ткани гены, уровень экспрессии которых не изменялся в семенниках на фоне мутации *comr* (Рис. 5А). Не изменялся уровень экспрессии и у генов подгруппы II, представители которой неактивны в семенниках и не имеют четко выраженной специфичной экспрессии в других органах (Рис. 5А).

В составе выборки *Comr*-несвязанных генов ($\text{Log}_2(\text{Dam-Comr}/\text{Dam}) < -2$) было обнаружено четыре подгруппы генов (IV-VII, Рис. 6А). Подгруппа IV была аналогична подгруппе II *Comr*-связанных генов. Подгруппа V состояла из семенник-специфичных генов. Гены подгруппы VI активируются в нервной ткани. Подгруппа VII состояла из генов, специфичных для яичников. Уровень

транскрипции генов подгруппы V снижался в семенниках самцов, мутантных по гену *comr*. По всей вероятности гены, входящие в состав этой подгруппы, регулируются *Comr*-зависимыми вторичными регуляторами транскрипции. Уровень экспрессии генов остальных подгрупп несколько повышался в семенниках мутантных по *comr* самцов, что может быть связано с изменением клеточного состава в мутантных семенниках (Рис. 5А).



В результате мы установили, что *Comr*-связанные гены обладают различными профилями экспрессии. И по характеру их активности в различных органах можно выделить две относительно гомогенные группы, одна из которых состоит из семенник-специфичных генов (III), а другая – из генов, активирующихся в нервной ткани (I). Причем только гены группы III регулируются белком *Comr*. Мы задались вопросом, какие общие свойства могут быть причиной связывания *Comr* с генами, имеющими настолько различающиеся паттерны экспрессии? Ранее было показано, что в клетках слюнных желез личинок, которые являются соматическими, семенник-специфичные гены у *D. melanogaster* зачастую содержатся в составе доменов транскрипционно неактивного хроматина, ассоциированных с белком SUUR (Belyakin et al., 2005). Сопоставление данных о связывании генов с белком *Comr* в семенниках с данными о связывании белка SUUR в слюнных железах выявило существенную положительную корреляцию (Максимов и др., 2013). Это свидетельствует о том, что гены, проявляющие в семенниках связывание с *Comr*, в слюнных железах имеют тенденцию располагаться в репрессированных доменах хроматина, что может иметь отношение к их регуляции в развитии.

Гены-мишени *Comr* располагаются в районах неактивного хроматина в недифференцированных клетках мужского зародышевого пути. Белок *Comr* впервые обнаруживается на стадии первичных сперматоцитов, в которых он необходим для активации нескольких сотен семенник-специфичных генов. На предыдущей стадии сперматогониев эти гены неактивны, причем в подавлении их экспрессии участвует белок Polycomb.

В работе Gan et al. [2010] методом ChIP-seq были установлены профили распределения гистоновых модификаций H3K27me3 и H3K4me3 в хроматине

семенников самцов, мутантных по гену *bam*. Мутация по гену *bam* блокирует переход от стадии сперматогониев к сперматоцитам. Это приводит к тому, что семенники заполняются цистами сперматогониев и не содержат клеток, находящихся на последующих стадиях развития. Поэтому семенники *bam*-мутантных самцов могут быть использованы в качестве модели для изучения структуры хроматина в недифференцированных мужских клетках зародышевого пути.

Мы воспользовались данными (Gan et al., 2010), чтобы проверить, не является ли состояние хроматина объединяющим свойством для генов-мишеней белка Comr. Для этого были просуммированы интенсивности сигналов ChIP-seq для H3K27me3 и H3K4me3 в области ± 200 п.н. вокруг сайтов инициации транскрипции для каждого гена генов *D. melanogaster*. Триметилирование гистона H3 по остатку лизина 27 (H3K27me3) наблюдается в транскрипционно неактивных районах хроматина, обогащенных белками группы Polycomb. Напротив, метилирование гистона H3 по остатку лизина 4 (H3K4me3) характерно для генов, проявляющих повышенный уровень транскрипции и повышенное содержание белков группы Trithorax.

На Рис. 5Б видно, что по сравнению с Comr-несвязанными генами, промоторные области всех Comr-связанных обогащены меткой неактивного хроматина H3K27me3. Кроме того, уровни маркера активного хроматина H3K4me3, были значительно ниже в группе Comr-связанных генов, чем в группе Comr-несвязанных. Причем эти наблюдения справедливы для всех подгрупп Comr-связанных и Comr-несвязанных генов. Таким образом, в недифференцированных клетках мужского зародышевого пути промоторные области Comr-связанных генов обогащены репрессивной гистоновой меткой. Это свойство наблюдается у всех Comr-связанных генов, вне зависимости от того, экспрессируются ли они в семенниках или нет, и может свидетельствовать о том, что связывание Comr определяется состоянием хроматина.

Транскрипционный фактор Comr связывается с протяженными участками хромосом. Ранее было показано, что репрессированный хроматин в геноме дрозофилы занимает протяженные области хромосом, зачастую достигающие нескольких сотен т.п.н. Поскольку наш анализ показал обогащение Comr-связанных генов гистоновой модификацией, характерной для репрессированных участков генома, мы решили проверить, не располагаются ли эти гены группами в геноме. В результате визуального анализа полногеномного профиля связывания белка Comr при помощи программного обеспечения UCSC genome browser (<http://genome.ucsc.edu/>) на хромосомах были также обнаружены протяженные непрерывные участки связывания белка Comr. Для объективного определения Comr-связанных, Comr-несвязанных и промежуточных геномных областей мы проанализировали наши данные с использованием Скрытой Марковской модели (Hidden Markow Model, НММ). В результате мы обнаружили 725 геномных областей, обогащенных белком Comr, и 837 областей, с которыми Comr не связывался. Остальные геномные области принадлежали к промежуточному классу. Размеры Comr-связанных и Comr-несвязанных областей значительно отличались: размеры 70% Comr-связанных областей превышали 20 т.п.н., а 30% - 100 т.п.н. Что касается Comr-несвязанных областей, то их размеры были меньше - протяженность только 26% из них превышала 20 т.п.н., и ни одна не достигала 100 т.п.н.

Области связывания Comr содержали 3036 генов, причем лишь 693 из них были семенник-специфичными. Кроме того, 350 областей связывания Comr не содержали семенник-специфичных генов. Протяженность 192 таких областей превышала 20 т.п.н., а 37 из них – 100 т.п.н. В составе этих 350 областей связывания Comr, располагались 638 генов, причем лишь 124 из них снижали уровень экспрессии в два и более раз на фоне мутации по *comr*. Из этого наблюдения следует, что многие такие области не содержали генов, регулируемых Comr. Промоторные области таких генов были обогащены репрессивной гистоновой меткой H3K27me3 и обеднены активной меткой H3K4me3.

Полученные данные свидетельствуют о том, что Comr-связанные области образуют протяженные домены, и что Comr-связанные гены образуют кластеры в их составе. Кроме того, области связывания Comr обогащены H3K27me3, из чего мы предполагаем, что Comr, как часть tMAC-комплекса, может участвовать в процессах перестройки хроматина, причем связывание Comr определяется состоянием хроматина, а именно наличием репрессивной метки H3K27me3.

Активация генов в семенниках *D. melanogaster* транскрипционными факторами tMAC и tTAF. Важно отметить, что в отсутствие tTAF действия только белков комплекса tMAC не достаточно для активации генов в сперматоцитах дрозофилы. Согласно существующим данным, регуляторное влияние белков tTAF и tMAC в конечном итоге направлено на пересекающиеся группы генов, однако до сих пор остается не изученным вопрос о возможной совместной регуляции генов белками tTAF и tMAC. Для исследования механизма совместной регуляции генов-мишеней белками, кодируемыми генами *aly* (tMAC) и *cap* классов (tTAF), мы сопоставили данные по экспрессии генов в семенниках самцов, мутантных по генам *cap* и *comr*. Мутация гена *cap* (*cap*³) вызывала нарушения в транскрипции значительно меньшего числа генов, чем было обнаружено в семенниках *comr*-мутантов. Так, лишь 309 генов снижали уровень экспрессии в 8 и более раз. Среди них, 277 (90%) также снижали уровень экспрессии в 8 и более раз и в семенниках *comr*-мутантов. Из 779 генов, снижающих экспрессию в 4 и более раз в семенниках *cap*-мутантов, 702 гена (88%) экспрессировались на таком же или более низком уровне в семенниках *comr*^{z2-1340} самцов. Это означает, что активность большинства *cap*-зависимых генов также зависит и от *comr*. Как уже было упомянуто, белок Comr напрямую связывался приблизительно с 40% генов, снижающих экспрессию в 8 и более раз в семенниках самцов *comr*^{z2-1340}. Среди 277 генов, снижающих экспрессию не менее чем в 8 раз в семенниках как самцов *cap*³, так и *comr*^{z2-1340}, с белком Comr были связаны только 112 генов, что также составляет 40%. Оставшиеся 60% генов в этой группе, по-видимому, регулируются при участии вторичных Comr-зависимых регуляторов, к которым может принадлежать CG9879. Необходимо отметить, что активность гена CG9879 снижается в 8 раз в семенниках самцов, несущих мутацию по *cap*, то есть его активность регулируется как tMAC, так и tTAF.

ОБСУЖДЕНИЕ

Для понимания роли отдельного транскрипционного фактора в генетической регуляторной сети необходимо обладать тремя типами данных. Во-первых, требуется определить, когда и где он работает в организме. Во-вторых, необходимо установить участки генома, с которыми связывается изучаемый транскрипционный фактор. И в-третьих, надо оценить, какой эффект связывание оказывает на

активность генов-мишеней. Обладая такими данными можно реконструировать часть генной сети, контролируемой исследуемым транскрипционным фактором.

Гены *aly* и *can* классов обладают семенник-специфичным профилем экспрессии и активны только в течение G₂-фазы первого деления мейоза сперматоцитов. Следовательно, в организме дрозофилы действие транскрипционных факторов, кодируемых генами задержки мейоза, ограничено одной стадией развития клеток зародышевого пути. Значительно менее определенным является механизм действия этих транскрипционных факторов.

В проведенных ранее работах исследователи предполагали активирующую роль белков tMAC, к которым принадлежит и Comr, опираясь на тот факт, что они ассоциированы с эухроматином, а у мутантов *aly* класса снижается уровень экспрессии многих генов. Однако такие наблюдения не являются однозначным доказательством активирующей роли продуктов генов *aly* класса, поскольку на основании имеющихся данных невозможно было исключить возможность активации генов-мишеней посредством подавления неизвестных генных репрессоров. Также, невозможно было исключить и существование tMAC-зависимых вторичных регуляторов, входящих в состав активирующего каскада в сперматоцитах. Более того, все предположения относительно механизма действия белков tMAC строились на основании исследования небольшого числа генов, для которых даже не было показано прямого физического взаимодействия с tMAC комплексом.

Проведенное нами картирование сайтов связывания белка Comr в масштабе генома позволило выявить более трех тысяч генов, вблизи которых детектировалось значительное связывание Comr. Проведенный сравнительный анализ связывания Comr с хромосомами и данных об экспрессии генов в семенниках мутантов *comr* предполагает, что белок Comr действительно необходим для активации генов, с которыми он непосредственно связан. Поскольку многие гены, связанные с Comr, требуют для активации функции tTAF, можно предположить, что, по крайней мере в ряде случаев, связывание Comr необходимо, но не достаточно для их активации.

Интеграция данных о связывании белка Comr с хромосомами и об эффекте мутации на генную активность позволила предположить наличие вторичных генов-регуляторов, экспрессия которых зависит от белка Comr. Действительно, в списке генов-мишеней белка Comr обнаружился ген *CG9879*, кодирующий семенник-специфичный ТАТА-связывающий белок, экспрессия которого значительно падала как на фоне мутации *comr*, так и *can*. Этот ген не описан в литературе, и полученные нами данные делают его кандидатом на роль вторичного регулятора в каскаде активации генов в сперматоцитах, действующего после генов *aly* и *can* классов (Рис. 6). Выяснение роли *CG9879* требует отдельного исследования, однако сам факт того, что в генетическом каскаде, запускаемом генами задержки мейоза, имеются ранее не охарактеризованные вторичные регуляторы, по нашему мнению весьма существенен.

В предыдущих исследованиях было показано, что многие гены, активные исключительно в семенниках *D. melanogaster*, в соматических клетках располагаются в неактивных хроматиновых доменах. Используя опубликованные паттерны гистоновых модификаций H3K27me3 и H3K4me3 в семенниках мутантов *bam*, мы показали, что в сперматогониях гены, которые после дифференцировки

сперматоцитов будут связываться с белком *Comr*, обогащены репрессивной модификацией H3K27me3.

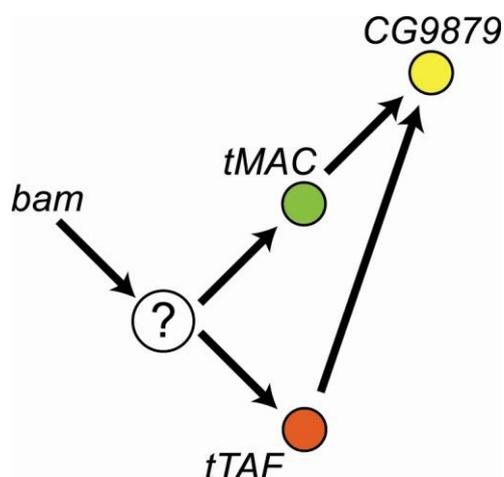


Рис. 6. Предполагаемый генетический каскад, регулирующий активность генов в сперматоцитах. В результате действия неизвестного механизма в начале стадии сперматоцитов запускается экспрессия генов *tMAC* и *tTAF*, которые активируют свои гены-мишени, среди которых содержатся вторичные регуляторы. Для активации гена *CG9879*, который может отвечать за активацию *tMAC*-несвязанных генов, требуются как *tMAC*, так и белки *tTAF*.

Полученные нами результаты улучшают понимание механизма массовой активации генов в сперматоцитах *D. melanogaster*. Для запуска экспрессии генов в сперматоцитах требуется активность обоих классов генов задержки мейоза. Мы убедились в том, что большинство генов, чья активность зависит от действия *tTAF*, зависели также и от *tMAC*. В проведенных ранее исследованиях предполагалось, что белки *tTAF* являются промотор-специфичными и не способны связываться с ДНК в отсутствие *tMAC*-компонентов. Принимая во внимание полученные данные, можно предположить, что *Comr*-связанные гены, активность которых снижается как на фоне мутации *comr*, так и *can*, напрямую регулируются и *tMAC*, и *tTAF* белками. Кроме того, ранее было установлено, что совместное действие генов *aly* и *can* классов необходимо для удаления H3K27me3, белка Polycomb и загрузки на промоторы РНК-полимеразы II, что в конечном итоге приводит к активации генов дифференцировки в сперматоцитах. Опираясь на полученные нами и другими исследователями данные, можно предположить, что на начальном этапе активации белок *Comr* в составе *tMAC* связывается с участками транскрипционно неактивного хроматина, обогащенного H3K27me3, что приводит к изменению его структуры. В пользу предположения о способности *tMAC* изменять свойства хроматина свидетельствует тот факт, что в составе этого комплекса содержится гистоновый шаперон Caf1, являющийся также компонентом других комплексов, способных модулировать структуру хроматина, таких как NURF и PRC2. По нашему предположению в результате перестройки хроматина становятся доступными сайты связывания промотор-специфичных транскрипционных факторов, таких как семенник-специфичные *tTAF*. Специфичные транскрипционные факторы активируют свои гены-мишени, среди которых содержатся вторичные регуляторы транскрипции, продолжающие дальнейший активирующий каскад в сперматоцитах. Предложенная модель может являться

основой для дальнейших исследований регуляции генов в процессе сперматогенеза *D. melanogaster* и может быть полезна для понимания механизмов генетической регуляции клеточной дифференцировки.

ВЫВОДЫ

1. Создана система для проведения картирования сайтов связывания транскрипционных факторов методом DamID в клетках мужского зародышевого пути у *D. melanogaster*.
2. Определен профиль связывания транскрипционного фактора Cookie Monster. Установлено, что Cookie Monster связывается с протяженными участками хромосом, содержащими более трех тысяч генов, промоторные области которых обогащены репрессивной гистоновой модификацией H3K27me3 и обеднены меткой активного хроматина H3K4me3 в недифференцированных герминальных клетках.
3. Установлено, что активирующее действие белка Cookie Monster осуществляется за счет прямого воздействия на гены-мишени, а также путем активации вторичных регуляторных генов. Кандидатом на роль такого регулятора является ген *CG9879*.
4. Установлено, что Cannonball участвует как в активации части генов, напрямую активируемых Cookie Monster, так и ряда генов, контролируемых Cookie Monster через вторичные регуляторы.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ РАБОТЫ

Статьи:

1. **Лактионов П.П.**, Андреева Е.Н., Шлома В.В., Максимов Д.А., Белякин С.Н. Генетическая система для мечения соматических и герминальных клеточных линий в гонадах *Drosophila melanogaster* // Цитология. 2013. Т. 55. С. 185-189.
2. Максимов Д.А., **Лактионов П.П.**, Белякин С.Н. Доменная регуляция экспрессии генов в районах интеркалярного гетерохроматина *Drosophila melanogaster* // Цитология. 2013. Т. 55. С. 190-193.
3. **Лактионов П.П.**, Н. White-Cooper, Максимов Д.А., Белякин С.Н. Транскрипционный фактор Comg играет роль прямого активатора в генетической программе сперматогенеза у *D. melanogaster*// Молекулярная биология. *Принята к публикации.*

Тезисы конференций:

1. **Laktionov P.P.**, Belyakin S.N., Maksimov D.A., Shloma V.V., Zhimulev I.F. Identification of transcriptional regulators for testis-specific gene clusters in *Drosophila melanogaster*// 16th European Workshop on Molecular and Cellular Endocrinology of the Testis, 2010, Elba, Italy. P. 76.
2. **Лактионов П.П.**, Максимов Д.А., Белякин С.Н. Роль транскрипционного фактора Cookie monster в регуляции генной экспрессии в процессе сперматогенеза *Drosophila melanogaster*// Международная конференция «Хромосома-2012», 2012, Новосибирск, Россия. С. 126.

3. Белякин С.Н., Максимов Д.А., **Лактионов П.П.**, Коряков Д.Е. Регуляция репрессированных районов генома в развитии дрозофилы// Международная конференция «Хромосома-2012», 2012, Новосибирск, Россия. С. 44.
4. **Laktionov P.P.**, Maksimov D.A., White-Cooper H., Belyakin S.N. Male meiosis arrest factor Cookie monster binds the extensive chromosome regions and counteracts the repression of testis-specific genes// ESF-EMBO Symposium on Systems Biology of Drosophila Development, Pultusk (Poland), May 22-25, 2012.
5. **Лактионов П.П.**, Максимов Д.А., Белякин С.Н. Метод тканеспецифичного DamID для полногеномного картирования сайтов связывания транскрипционных факторов в клетках мужского зародышевого пути *D.melanogaster*// Международная конференция «Высокопроизводительное секвенирование в геномике», 2013, Новосибирск, Россия. С17.