

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт фундаментальной биологии и биотехнологии
Базовая кафедра защиты и современных технологий мониторинга лесов

УТВЕРЖДАЮ
Заведующий кафедрой
_____ И.Е.Ямских
подпись инициалы, фамилия
« _____ » _____ 20 __ г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Определение гипер- и гипо- метилированных участков ДНК и
фенотипических последствий эпигенетических изменений с помощью
методов иммунопреципитации

06.04.01 – Биология

06.04.01.06 – Геномика и биоинформатика

Научный руководитель	_____	<u>доцент, PhD</u>	<u>Татарина Т.В.</u>
	подпись, дата	должность, ученая степень	инициалы, фамилия
Выпускник	_____		<u>Поплаухин А.А.</u>
	подпись, дата		инициалы, фамилия
Рецензент	_____	<u>Профессор,</u>	<u>Орлов Ю.Л.</u>
	подпись, дата	<u>Д. б. н.</u>	инициалы, фамилия
		должность, ученая степень	

Красноярск 2019

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Обзор литературы.....	6
1.1 История эпигенетики	6
1.2 Метилирование ДНК	13
1.3 Масличная пальма.....	16
2 Материалы и методы	17
2.1 Исходные данные.....	18
2.2 Проверка качества и очистка ридов	20
2.3 Получение выравниваний	20
2.4 Обработка данных пакетом MEDIPS	21
2.4.1 Анализ насыщения ридами	22
2.4.2 Покрытие ридами контекста CpG	23
2.4.3 Получение профилей метилирования.....	23
2.5 Определение обогащенных регионов набором программ MACS2	24
2.6 Сравнение профилей метилирования MEDIPS и MACS2 с профилем бисульфитного секвенирования	24
3 Результаты.....	24
3.1 Результаты выравнивания.....	Ошибка! Закладка не определена.
3.7 Сравнение профилей метилирования MEDIPS.....	Ошибка! Закладка не определена.
ЗАКЛЮЧЕНИЕ	26
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	27
ПРИЛОЖЕНИЕ А	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ Б.....	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ В	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ Г	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ Д	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ Е.....	Ошибка! Закладка не определена.
ПРИЛОЖЕНИЕ Ж	Ошибка! Закладка не определена.

ВВЕДЕНИЕ

Актуальность проблемы. В течение последних 15 лет понимание работы механизмов наследственности и изменчивости вышло на новый уровень благодаря молекулярным методам исследования, а в частности разработке методов высокопроизводительного секвенирования последовательности ДНК. Это позволило собрать обширную базу данных по геномам многих организмов и изучить структурно-функциональные особенности реализации генетического материала. Глубокие исследования геномов привели к подтверждению того, что экспрессия генов, взаимодействие генов и вообще генная регуляция не обуславливается одной лишь последовательностью ДНК, а кодируется каким-то иным образом. Было установлено, что метилирование ДНК, а также модификации гистонов, конформация ДНК - все эти факторы влияют на реализацию генетического материала в клетке, отвечают за дифференциацию клеток, влияют на рост и развитие, а также отвечают за некоторые заболевания. Область знаний предметом которой являются изменения фенотипа без изменения генотипа называется эпигенетика. А когда речь идет об изучении не отдельных генов, а целых геномов, то речь идет об эпигеномике.

В первую очередь, эпигенетические модификации изучались на млекопитающих, однако сейчас поле эпигенетики расширяется и все для большего числа организмов получают информацию по их эпигенетическому ландшафту. Многие явления вносят вклад в эпигенетику организма, но первым открытым и наиболее изученным является метилирование ДНК. Это эпигенетическая модификация заключается в присоединении метильной группы к цитозину в составе контекста CpG. Метилирование в промоторной зоне гена приводит к подавлению экспрессии гена. Существует несколько методов для детекции метилирования, но методом с самым большим разрешением является бисульфитное секвенирование. Этот метод позволяет узнать конкретные метилированные основания, то есть его разрешение это 1

нуклеотид. Однако, бисульфитное секвенирование, несмотря на снижение цены на секвенирование, до сих пор остается дорогим методом для эпигеномных экспериментов, так как в отличие от геномного секвенирования, требует многих повторений, в зависимости от условий и, соответственно, изменений в эпигенетическом ландшафте организма. Также существуют различные другие методы детекции метилирования, которые дают меньшее разрешение, но стоят дешевле. Одним из таких методов является MBD-seq, аффинный метод, который использует белок, связывающийся с метилированными основаниями ДНК. MBD-seq был разработан и активно использовался для млекопитающих.

Объектом данного исследования является масличная пальма. Масличная пальма является важным источником жирных кислот в виде масла. Из плодов масличной пальмы получают два вида масла: пальмовое масло из мякоти, которое используется в пищевой промышленности, а также пальмоядровое масло, которое получают из ядер плодов и используют для производства глицерина и в косметической промышленности. Масличную пальму размножают клонами из каллуса. Это позволяет добиться быстрого приготовления посевного материала и размножения. Однако, возникает проблема того, что 5% от пальм, полученных таким способом не фертильны, и, соответственно не дают плодов. Генетический состав у этих пальм одинаков, а вот эпигенетический ландшафт очень сильно отличается и определяет будет ли пальма фертильна. Как уже было сказано, проверка каждой пальмы бисульфитным секвенированием экономически невыгодна, поэтому было предложено оценивать метилирование ДНК масличной пальмы с помощью метода MBD-seq. Этот метод применялся в масштабных исследованиях на животных, но для масличной пальмы не тестировался.

Цели и задачи исследования. Целью данного исследования является проверка эффективности метода MBD-seq для эпигенома пальмы путем сравнения профилей метилирования биологических и технических реплик между собой и с данными бисульфитного секвенирования.

Для достижения данной цели были поставлены следующие задачи:

- проверить качество исходных данных;
- очистить и отфильтровать данные низкого качества;
- оценить уровни метилирования в пакете MEDIPS и MACS2;
- сравнить полученные уровни метилирования с данными бисульфитного метилирования.

Научная новизна. Впервые проверено качество MBD-seq для эпигенома масличной пальмы. Впервые проведено сравнение MBD-seq и бисульфитного секвенирования для масличной пальмы.

Практическая значимость. Данное исследование имеет важное значение для лесного хозяйства и лесной эпигеномики. В результате работы будут получены данные, обуславливающие возможность или невозможность использования более дешевого метода детекции эпигенетических модификаций, что облегчит выведение новых масличных пальм из каллуса и увеличит объемы его производства.

Благодарности. Автор выражает искреннюю признательность сотрудникам и аспирантам кафедры базовой кафедры защиты и современных технологий мониторинга лесов СФУ за ценные указания, советы и замечания по ходу работы. Особую благодарность хотелось бы выразить своему научному руководителю Татариновой Татьяне Валерьевне PhD, которая курировала и направляла меня для успешного достижения целей исследования.

Магистерская диссертация выполнена в лаборатории лесной геномики СФУ под руководством К.В. Крутовского и базовой кафедры защиты и современных технологий мониторинга лесов (зав. каф. д.б.н. И. Е. Ямских)

1 Обзор литературы

1.1 История эпигенетики

Современная эпигенетика это междисциплинарная область знания, которая находится на стыке генетики, молекулярной биологии и биологии развития. Термин “эпигенетика” был впервые предложен Конрадом Уоддингтоном в 1942 г., как изучение причинных механизмов реализации генотипа в фенотип [1]. Термин является производным от слов “генетика” и “эпигенез”. Генетика – раздел биологии, наука о закономерностях наследственности и изменчивости живых организмов; эпигенез - учение об эмбриональном развитии организмов, в ходе которого происходят последовательные новообразования органов. В течение следующих пятидесяти лет, по мере накопления знаний о строении клеточного ядра и механизмах передачи наследственной информации это достаточно размытое определение дополнялось и уточнялось. Д. Нэнни (D.L.Nanney) рассматривал эпигенетику как область знания, объясняющую, как и почему клетки (организмы) с идентичным генотипом могут различаться по наследуемому фенотипу. Д. Готчлинг и А. Риггс описывали эпигенетику как науку о наследуемых изменениях, которые не связаны с мутациями собственно в ДНК [2]. С. Эллис определил эпигенетику как наследуемую «клеточную память», связанную со структурными изменениями хроматина. Несмотря на очень общую формулировку данного определения, оно очень верно, потому что многое в онтогенезе в конечном итоге определяется именно модуляциями структуры хроматина. Генетик Робин Холлидей рассматривал эпигенетику как «изучение контроля за активностью генов во времени и пространстве в процессе развития сложных организмов». Именно Холлидей один из первых указал на возможную биохимическую природу наследуемых эпигенетических сигналов (например, метилирование ДНК). Эпигенетику рассматривают как совокупность знаний об изменениях в экспрессии генов в результате модуляции организации хроматина без изменения последовательности ДНК. Б.Ф. Ванюшин добавляет, что

«эпигенетика — это наука о наследуемых свойствах организмов, которые не связаны с изменением собственно нуклеотидной последовательности ДНК, и они могут быть не прямо, а косвенно закодированы в геноме» [3].

Эпигенетика, хоть ещё и в неоформленном виде, разрабатывалась на стыке биологии развития и генетики с конца XIX в., когда естествоиспытатели уже руководствовались научным методом и имели в своем распоряжении методы химии и микроскопии. Тогда исследователи задавались вопросом о том, как может развиваться сложный организм из одной клетки, как объяснить сложность структурно-функциональной организации организма.

Однако этот вопрос, только в более общем виде задавали себе ещё натурфилософы античности. Они не руководствовались научным методом познания, так как его еще не было, но уже тогда философы могли наблюдать, что из маленького семени вырастает сложное растение, а также имели представление о репродукции животных. Объяснения данным явлениям философы выводили из своих философских систем. Таким образом еще в античности возникло два учения: преформизм и эпигенез.

Оба учения были разработаны греческими натурфилософами и возрождены в 19 веке. Так, древнегреческий философ Алкмеон Кротонский (VI – V век до н.э.) предположил, что оплодотворение есть слияние мужского и женского «семени», Эмпедокл (490 - 430 л. до н.э.) добавил идею смешения отцовских и материнских свойств, а Анаксагор (500 - 428 л. до н.э.) – идею «семян всех вещей», т.е. мельчайших частиц, несущих свойства наблюдаемых предметов (в т.ч. живых). «Семена» организуются в наблюдаемые вещи под действием либо «мирового ума», либо (в организме) – материнского тепла. Стоицизм, основанный Зеноном в Греции в III веке до н.э., развил идею «семян» Анаксагора в концепцию «семенных логосов», которые сами мыслились в качестве единиц организованности. Средневековые схоласты Григорий Нисский (335 – 394 г.) и Аврелий Августин (354 – 430 г.) толковали творение мира как творение «семенных логосов», в которых изначально заключены потенции всех появившихся и еще обреченных появиться в

будущем объектов. Античный преформатизм был вместе с тем и неким подобием теории предустановленной эволюции [4].

Эпигенез в свою очередь разрабатывался в противовес преформатизму Аристотелем, и представлял собой возврат к древнейшей идее прорастания семени: самка дает материю, а самец – форму зародыша (ее содержит мужское семя). Он утверждал, что дифференциация и развитие это продолжительные, сложные биологические процессы, где каждый процесс зависит от предыдущего, что в итоге и дает организм [5].

Эпигенез развивал Гален в книге «О естественных способностях». В. Гарвей в произведении «Исследования о зарождении животных» 1651 года, где обобщил многолетние исследования эмбрионального развития позвоночных и беспозвоночных, сформулировал теорию эпигенеза. Н. Тербервилль в работе «Наблюдения за образованием, составом и разложением животных и растительных веществ» (1749) также поддерживал идею эпигенеза, как и другие ученые 18-19 веков. Однако, микроскопические исследования (Я. Сваммердам, 1669, М. Мальпиги, Ш. де Бонне 1762 г., А. Галлей и Л. Спалланцани) привели к возрождению преформатизма. Преформированный зародыш, известный как гомункул, «видели» либо в спермии, либо в яйцеклетке, который после оплодотворения начинал увеличиваться в размере. Преформатизм плохо объяснял сходство потомства с обоими родителями (что намного лучше объяснил Аристотель в произведении «О поколении животных»), феномены уродства и регенерации органов, поэтому с 1780-х гг. вновь стал популярен эпигенез благодаря диссертации К.Ф. Вольфа «Теория поколений» в 1759 г., описавшей развитие органов у зародыша и положившей начало современной эмбриологии [6]. Но в 1890-е годы А. Вейсман опубликовал теорию зародышевой плазмы, идея которой состояла в «непрерывности зародышевой плазмы — своего рода семейной реликвии, передаваемой от родителей к детям в полном объеме и претерпевающей изменения только в результате смешения, неизбежно вызываемого половым

размножением». Данная публикация фактически вернула науку на уровень преформатизма 17 века, а эволюцию представила, как бессмыслицу.

Согласно интерпретации Конрада Уоддингтона в произведении «Введение в современную генетику» 1939-го года, преформатизм и эпигенез являются дополняющими друг друга теориями. Он связал основополагающие элементы развития (гены) с эмбриологией для формирования термина «эпигенетика». В своей работе он пишет: «...оплодотворенная яйцеклетка содержит компоненты, которые имеют определенные свойства, что позволяет реализовать лишь определенное ограниченное количество реакций; в случае если это верно, то можно сказать, что развитие происходит на основе «преформированных» качеств оплодотворенного яйца. Однако столь же очевидно, что взаимодействие этих компонентов приводит к появлению новых тканей и органов, которые не были присутствовали в зародыше изначально, и такое развитие должно рассматриваться как «эпигенетические» ...». В дальнейшем Уоддингтон уточнил, что генотип и фенотип связаны комплексной совокупностью генетических и негенетических процессов развития, которые он назвал «эпигенотипом», и предложил обновить систему $\text{генотип} + \text{окружающая среда} = \text{фенотип}$ на $\text{генотип} + \text{эпигенотип} + \text{окружающая среда} = \text{специфичный фенотип}$ [7].

В 1957 году Уоддингтон в качестве поясняющей метафоры сформулировал концепцию «эпигенетического ландшафта», в котором клетка «двигается по каналу», чтобы достичь ожидаемого фенотипа за счет «скатывания» вниз по «волнистой поверхности», форма которой определяется взаимодействием между различными генов. С точки зрения развития, когда клетка запрограммирована на достижение определенной стадии дифференцировки, программа не может быть изменена для достижения иной конечной формы.

Концепция эпигенетики в большей степени относилась к дифференциации клетки как к «процессу, через который генотип реализует фенотип». Но только в 1958 году второй появился важный подход к

эпигенетике, что помогло приблизиться к современному пониманию термина. В своем эссе «Системы эпигенетического контроля» Дэвид Нэнни утверждал, что очевидно существование двух клеточных регуляторных систем. Одна относится к механизмам транскрипции ДНК (генетическая), и дополняющая система, имеющая иные механизмы действия, отвечает за определение какую информацию реализовывать и в каких клетках (эпигенетика) [8]. Нэнни знал, что трудно различить свойства клетки из-за изменений в генетическом материале от изменений, вызванных другими механизмами, и как объяснить то, как эффекты этих «вспомогательных» процессов могут сохраняться в фенотипе на протяжении многих поколений (1958). Нэнни привел аргументы, которые казались «сумасшедшими» для научного сообщества того времени. Так, например, наблюдения Мюллера (1930) в его исследованиях транслокаций, инерции и делений в хромосомах *Drosophila melanogaster* предполагает, что при отсутствии любых других изменений, изменение позиции гена в геноме может менять экспрессию гена (фенотипически проявляется как: непрерывно выщепляющий рецессивные формы (eversporting), пятнистость и эффект мозаики). Только в 1951 году А. Ханна (Aloha Hannah) подтвердила, что эффект варьирования, наблюдаемый Мюллером, был связан с тем, что гены, расположенные в эухроматине, были перемещены в область гетерохроматина. В то же время Барбара Мак-Клинтон была заинтересована данным феноменом в исследовании кукурузы (1945-1956). Она открыла, что перемещения локуса Ds и Ac вдоль генома (транспозиция) стали причиной фенотипического изменения цвета кукурузы по типу эффекта мозаики, в зависимости от новой локации транспозона за счет модификации активности близлежащих генов (или регуляции активности) (1950).

Аналогичным образом были заданы первые вопросы касательно фенотипических различий между клетками, которые якобы несут один и тот же генетический материал, например, различия в морфологии и пигментации среди колоний, возникающих в результате роста чистой культуры некоторых бактерий. Это явление указывало на то, что экспрессия определенных

признаков определялась не только наличием определенных генов в клетке. Явления, такие как вариация серотипа и бактериальной фазы были вызваны окружающей средой и отнесены с эпигенетическим [10]. Фазовая изменчивость в настоящее время известна как наследственный, но обратимый механизм регуляции генов, который вызывает фенотипическую гетерогенность за счет стимулирования синтеза белка в отдельных клетках клональной популяции [3]. Фенотипически это проявляется в появлении изменений в морфологии и цвете микробной колонии и в выражении пили, фимбрий и других белков, непосредственно связанных с патогенностью [11]. Другие примеры, описанные Нэнни, включают явления, в которых системы регуляции обратной связи клеток могут вызывать эпигенетические события. Однако, не было экспериментальных подтверждений, что организмы с идентичным генотипом способны образовать разный фенотип.

К тому времени, когда Уоддингтон ввел термин «эпигенетика» (1939) [1], теория наследования хромосом была продемонстрирована Томасом Морганом в *D. melanogaster* X-сцепленных генах (1911) [12,13], и А.Х. Стёртевант создал первую генетическую карту, основанную на частотах рекомбинации в этом организме (1913) [14]. Точно так же изменение в определении эпигенетики, предложенное Нэнни, возникло в то время, когда уже было известно, что ДНК — это молекула, несущая генетическую информацию об организмах (1952) [15,16] и структурные детали. нуклеиновые кислоты были доступны (1953) [17,18]. Несколько важных открытий в эпигенетике еще не были достигнуты, включая инактивацию X-хромосомы у млекопитающих (1961) [19], цитоплазматическое наследование в *Paramecium* (1965) [20], характеристику структуры нуклеосом (1974) [21] и открытие прионов (самореплицирующихся белков, способных передавать информацию по аналогии с нуклеиновыми кислотами) (1982) [22,23]. Тем не менее, исследования, которые в конечном итоге позволили представить концепции эпигенетики, представленные сегодня, были выполнены в основном на эукариотах и относятся к локализации и организации генетического материала в клетках и ковалентной модификации

ДНК и гистонов. Работы Аллфрея и соавт. (1964) [24] об ацетилировании и метилировании гистонов, и работы Риггса (1975) [25], Холлидея и Пью (1975) [26] о метилировании ДНК дали начало новому этапу изучения эпигенетики, что привело к выяснению его основных молекулярных механизмов: а также формировало современный взгляд на эпигенетику [27]. Однако новые достижения усложнили нахождение определения, охватывающего все явления, которые изменяют фенотип и не зависят от последовательности ДНК в организмах, особенно когда речь идет о наследовании.

В настоящее время эпигенетика обычно определяется как изучение митотически и/или мейотически унаследованных изменений функций, которые не могут быть объяснены изменениями в последовательности ДНК [9]. Несмотря на то, что в отношении этого определения был достигнут относительный консенсус, остается вопрос о том, должна ли эпигенетика охватывать понятие наследуемости между клеточными делениями. Не все эпигенетические модификации наследуются, и некоторые из них могут быть только временными. Кроме того, несмотря на наличие доказательств, подтверждающих наследственность некоторых эпигенетических изменений, в основном у растений и грибов, многие исследователи считают их плохо изученными и неопределенными, главным образом в случае с млекопитающими [28]. Может показаться, что самым практичным решением будет устранить требование наследственности, включенное в ранее представленное определение. Были предприняты новые попытки переопределить термин эпигенетика. Например, [28, с. 398] представил эпигенетику как «структурную адаптацию хромосомных областей для определения, сигнализации или стабилизации измененных состояний активности». Динс и Маггерт (2015) [29, с. 892] предположили, что эпигенетика - это «изучение явлений и механизмов, которые вызывают хромосомные наследственные изменения в экспрессии генов, которые не зависят от изменений в последовательности ДНК». Эти концепции пытаются решить проблему наследуемости, но они игнорируют или недооценивают такие

явления, как цепи обратной связи, структурное или трехмерное наследование конформации (инфузории и прионы), цитоплазматическая память, временные модификации хроматина и др., в связи с, нехваткой подробных знаний или доказательств. Эпигенетика - это не только изучение метилирования ДНК или посттрансляционных модификаций гистонов в эукариотах. В этом отношении Манн (2014) [30] предложил использовать дополнительный термин, когда наследуется эпигенетическое состояние (мемигенетика), освобождая термин эпигенетика от требования наследуемости.

В соответствии с идеей Манна, эпигенетику можно рассматривать как исследование как временных, так и постоянных (наследуемых) изменений в экспрессии генов, не вызванных изменениями в последовательности ДНК. Эпигенетика активно участвует в регуляции генов клетки и обеспечивает разнообразие генома, используя механизмы, которые вызывают временные изменения в качестве постоянных модификаций в функции гена, без изменений в последовательности ДНК [31,32]. Эти механизмы участвуют в дифференцировке клеток, их развитии и, следовательно, в проявлении фенотипа.

1.2 Метилирование ДНК

Инактивация X-хромосомы у мыши обеспечила раннюю модель такого рода эпигенетического механизма, который не включал перестройку ДНК [19]. Инактивированная X-хромосома была четко выбрана случайным образом, а затем клонально унаследована в соматических клетках, и не было никаких признаков изменений в самой последовательности ДНК. Риггс (1975) и Холлидей и Пью (1975) [25,26] предположили, что метилирование ДНК может выступать в качестве эпигенетической метки, частично для объяснения такого рода инактивации. Ключевыми элементами в этой модели были идеи о том, что сайты метилирования ДНК были палиндромными, и что разные ферменты были ответственны за метилирование немодифицированной ДНК и ДНК, уже

метилированной на одной цепи. Предполагалось, что первое событие метилирования ДНК будет намного сложнее, чем второе; после модификации первой цепи комплементарная цепь быстро модифицируется в том же палиндромном участке. Метилирующая метка ДНК, присутствующая на родительской нити, будет скопирована на дочернюю нить после репликации, что приведет к достоверной передаче метилированного состояния следующему поколению.

Вскоре после этого Бёрд воспользовался тем фактом, что основной целью метилирования у животных являются последовательности CpG, чтобы ввести использование чувствительных к метилированию рестрикционных ферментов в качестве способа обнаружения состояния метилирования ДНК. Последующие исследования [35] затем показали, что эндогенные сайты CpG были либо полностью неметилированы, либо полностью метилированы. Предсказания модели, таким образом, были подтверждены, установив механизм эпигенетической передачи метки метилирования через полуконсервативное распространение характера метилирования. Также был большой интерес к возможным механизмам, с помощью которых метильная группа в метилированных остатках цитозина (или сама 5 мС) удаляется, что происходит на ранних стадиях развития и в зародышевой линии.

Очевидно, степень, в которой метилирование ДНК может быть эпигенетической меткой, сохраняемой через зародышевую линию, определяется тем, какие сайты переживают события деметилирования. Хотя большая часть метилирования ДНК, наблюдаемая у позвоночных, связана с повторяющимися и ретровирусными последовательностями и может служить для поддержания этих последовательностей в постоянно неактивном состоянии, эта модификация так же может обеспечивать основу для эпигенетической передачи статуса активности гена.

Это наиболее четко продемонстрировано в импринтированных локусах [36], таких как локус мыши или человека *Igf2/H19*, в которых один аллель отмечен метилированием ДНК, который, в свою очередь, контролирует

экспрессию обоих генов [37]. В то же время было ясно, что это не может являться единственным механизмом эпигенетической передачи информации.

На данный момент известно, что метилирование ДНК это ненаследственная симметричная эпигенетическая метка, которая в высших эукариотах, практически исключительно встречается на пятом углероде остатка цитозина [38].

Главной целью метилирования являются динуклеотиды CpG, однако в растениях так же выделяют тринуклеотиды CHG и CHH. Гетерохроматиновые участки генома (центромеры, транспозоны, теломеры, повторяющиеся элементы) сильно метилированы. CpG островки, которые представляют собой области длиной не менее 550 п.н. с содержанием G + C более 50% и наблюдаемым отношением CpG к ожидаемому $> 0,65$, как правило, остаются свободными от метилирования. CpG островки обнаружены примерно в 70% промоторных областей человеческих генов.

В млекопитающих метилирование цитозина в ДНК обеспечивается действием фермента ДНК метилтрансферазы (DNMT), который катализирует перенос метильной группы с S-аденозил-1-метионина (SAM) кофактора на остаток цитозина. DNMT1, DNMT3a и DNMT3b, вероятно, являются наиболее важными белками в поддержании и создании эпигенетического импринта. DNMT1 связан с хроматином в фазе клеточного цикла, и, следовательно, он связан с репликацией ДНК и сохранении метилированного состояния. DNMT3a и DNMT3b являются ферментами, ответственными за установление тканеспецифического метилирования цитозина *de novo* во время эмбрионального развития; такое метилирование является решающим событием в дифференцировании и развитии клеток [38,39]. Метилирование цитозина ДНК участвует в таких явлениях, как инактивация X-хромосомы у самок млекопитающих и установление родительского эпигенетического импринта, при котором один из двух наследуемых родительских аллелей гена избирательно инактивируется [34].

Есть три основных механизма через которые предполагается, что метилирование ДНК приводит к репрессии транскрипции: путем ингибирования связывания фактора транскрипции с регуляторными последовательностями, путем привлечения метил-СpG-связывающих белков (МВР), которые блокируют элементы ответа, избегая связывания с агентами транскрипции, и путем образования комплексов между МВР и корепрессорами, конденсирующими хроматин [40]. Из-за своей значимости метилирование цитозина ДНК многими считается пятым, «забытым» основанием ДНК. Метилирование может быть активно обращено ферментативными методами (под действием диоксигеназ) или пассивно обратимо, когда механизмы метилирования выходят из строя после репликации ДНК [39]. Метилцитозин диоксигеназное семейство белков ТЕТ (Ten-Eleven-Translocation) (TET1, TET2 и TET3) катализирует последовательное окисление (процессивное) 5-метилцитозина (5mC) до 5-гидроксиметилцитозина (5hmC), 5-формилцитозина (5fC) и 5-карбоксилцитозина (5caC) у эукариот. Из этих продуктов было определено, что 5hmC является наиболее распространенным, и вероятно, 5hmC способствует пассивному деметилированию ДНК путем зависимой от репликации потери 5mC, когда цитозин в одной цепи ДНК окисляется, а в комплементарной цепи метилируется (сайт гемиметилирования). Ферменты ТЕТ могут быть частью пути, вовлеченного в замену 5fC и 5caC немодифицированным цитозином (эксцизионная репарация комплексом ТЕТ/тимин-ДНК-гликозилаза/основание), что приводит к деметилированию ДНК. Пока еще недостаточно доказательств, подтверждающих гипотезу о том, что 5hmC, 5fC и 5caC могут быть новыми эпигенетическими метками [41].

1.3 Масличная пальма

Масличная пальма (*Elaeis guineensis*) это растение из семейства пальмовых, также известное как Африканская пальма или Элеис гвинейский. Происходит из западной и юго-западной Африки, но также произрастает как

культурное растение на Мадагаскаре, Шри Ланке, в Малайзии, Индонезии, Центральной Америке, а также на нескольких островах Индийского и Тихого океанов.

Археологические находки говорят о том, что человечество использовало масло масличной пальмы уже 3000 лет назад до н.э. в Египте, но ботаническое описание данного растения первым составил французский натуралист Мишель Адансон.

Взрослые растения представляют собой неветвящиеся деревья высотой до 20 м. Листья перистые и достигают 3-5 метров в длину. Молодая пальма производит около 30 листьев в год. Взрослые пальмы производят около 20 листьев в год. Цветки растут в плотных кластерах, каждый отдельный цветок небольшой, с тремя чашелистиками и тремя лепестками.

Созревание плода пальмы после опыления занимает 5-6 месяцев. Плод красноватый, размером с большую сливу, и растут они в плотных мутовках. Каждый плод состоит из сочного, маслянистого внешнего слоя - перикарпа, и одного семени, также богатого маслом. При созревании каждая мутовка плодов весит от 5 до 30 кг, в зависимости от возраста пальмового дерева.

2 Материалы и методы

Биоинформатический анализ проводился для файлов, предоставленных малазийским советом по пальмовому маслу (Malaysian Palm Oil Board). Обработка данных включала в себя следующие стадии: проверка качества исходных данных, фильтрация данных, повторная проверка качества, выравнивание ридов на референсный геном, нормализация данных по покрытию и плотности CpG островков, сравнение реплик между собой, сравнение покрытия обогащенных регионов (enriched regions) с данными бисульфитного секвенирования.

2.1 Исходные данные

Образцы ДНК масличной пальмы были получены из тканей центрального листа (пальмовая ветвь F0), растущего из апикальной меристемы растения. Образцы были взяты у пяти клональных пальм (ID 753, 758, 764, 767, 768) на станции малазийского совета по пальмовому маслу в области Кератонг, штате Паханг, Малайзия. В последующем анализе данные образцы называются биологическими репликами.

Образцы ДНК пальмы Q38 были получены ранее на United Plantations BHD., в Пераке, в Малайзии. Так как эти образцы из одной пальмы, далее они называются техническими репликами. Всего было 6 технических реплик (ID Q38_1, Q38_2, Q38_3, Q38_5, Q38_9, Q38_10).

Сначала, геномная ДНК была разрушена до фрагментов средней длины 300 п.н. с помощью ультразвука Covaris, Inc. со следующими параметрами: 3 мкг геномной ДНК в 130 мкг пробирке microTUBE (Covaris, Inc), 50 peak power, 20% duty factor, 200 cycles/burst, 18°C температура водяной бани в течение 5 минут 50 секунд. Малые фрагменты ДНК были удалены с использованием Agencourt® AMPure® XP beads (Beckman Coulter) очисткой, согласно инструкциям производителя. Подсчет фрагментированной ДНК проводился с помощью dsDNA High Sensitivity assay (Invitrogen), а распределение размеров было проверено с помощью High Sensitivity DNA Kit (Agilent Technologies) на биоанализаторе Agilent 2100 (Agilent Technologies).

Иммунопреципитация метилированной ДНК шести технических реплик и пяти биологических реплик проводилась набором MethylCap Kit (Diagenode, Inc.), согласно инструкциям производителя. После шага иммунопреципитации, связанные метилированные молекулы ДНК были восстановлены в High Elution буфере с концентрацией соли 0.8 М (Aberg, Xie et al. 2015). Иммунопреципитированные (MBD-IP) и контрольные (input) продукты были очищены с использованием QIAquick PCR колонок очистки (Qiagen).

Так как в экспериментах с иммунопреципитацией ДНК на выходе получается небольшое количество ДНК, создание библиотек для секвенирования было произведено с помощью набора TruSeq[®] ChIP kit (Illumina), который подходит для экспериментов с низким содержанием ДНК. Библиотеки были подготовлены с 10 нг ДНК. Восстановление концов, добавление А-баз и лигирование адаптеров были проведены согласно протоколу изготовителя. Библиотеки были отобраны по размеру, затем провели 18 циклов ПЦР амплификации. Распределение размеров каждой библиотеки было оценено с помощью набора High Sensitivity DNA Kit (Agilent Technologies) на биоанализаторе Agilent 2100 (Agilent Technologies), в то время как количество амплифицированных библиотек было оценено qPCR с набором KAPA Library Quantification kit (Kapa Biosystems) и Qubit dsDNA High Sensitivity assay (Invitrogen). Генерация кластеров и секвенирование были проведены с использованием наборов TruSeq PE Cluster Kit v3-cBot-HS (Illumina) и TruSeq SBS Kit v3-HS (Illumina) соответственно, согласно инструкциям производителя. Каждая библиотека был секвенирована с помощью парноконцевых прочтений (ридов), с длиной рида 100 п.н. на секвенаторе HiSeq2000 (Illumina).

В качестве референсного генома для выравнивания ридов использовали сборку генома версии p5 пальмы AVROS pisifera, которая состоит из 40,360 геномных скаффолдов (N50 длина: 1,045,414 нуклеотидов, самый длинный скаффолд - 22,100,610 нуклеотидов, самый короткий - 1992 нуклеотида). Для генерации профилей метилирования использовали соответствующую аннотацию генома.

Данные метилирования бисульфитного секвенирования для сравнения результатов MBD-seq были предоставлены малазийским коллегами для пальмы Q38.

2.2 Проверка качества и очистка ридов

Работа с данными проходила на вычислительном кластере СФУ кафедры высокопроизводительных вычислений института космических и информационных технологий. В ходе работы были написаны несколько программ на скриптовых языках программирования bash и R. Эти скрипты с комментариями находятся в приложениях.

Данные с секвенатора были представлены в формате fastq, по несколько файлов для прямых и обратных прочтений. Сначала файлы прямых и обратных прочтений были конкатенированы соответственно. Затем каждый файл оценивался программой FastQC v0.11.5 [42]. По результатам оценки качества ридов были подобраны параметры для программы обрезки и фильтрации ридов по качеству - Trimmomatic v0.33 [43]. Позиции, у которых показатель качества Phred+33 был меньше тридцати в скользящем окне размером 4 нуклеотида, были обрезаны на концах ридов. После фильтрации ридов снова выполнили проверку качества программой FastQC, чтобы убедиться в улучшении качества.

2.3 Получение выравниваний

Референсный геном был представлен в виде fasta файла. Для выравнивания ридов на референсный геном использовали программу bowtie2 версии 2.3.4.2 [44]. Прежде всего сгенерировали bowtie2 индекс соответствующей командой. Для выравнивания выставили пресет настроек --very-sensitive, минимальную длину фрагмента - 50, максимальную длину фрагмента - 1000.

Результатом выравнивания bowtie2 является файл формата SAM. Для удобства работы и дополнительной очистки файлы формата SAM были переведены в формат BAM, а также из них были удалены все невыровненные риды и риды качеством ниже 10. Качество выравнивания 10 указывает на то, что вероятность того, что рид выровнен случайно равна 1/10. То есть остались

только выровненные риды с вероятностью правильного выравнивания 90% и выше.

Программой Picard markduplicates в bam файлах были отмечены риды дубликаты, которые являются артефактами ПЦР, появляющимися из-за слишком большого числа циклов амплификации, или оптическими дубликатами - артефактами метода секвенирования в секвенаторах фирмы Illumina. Такие риды происходят из одного кластера амплификации, который неправильно воспринимается оптическими сенсорами секвенатора, как множественные кластеры.

Программой qualimap v.2.2.1 [45] была проведена проверка качества выравниваний, были вычислены для каждого bam файла: средняя длина ридов, процент ридов-дубликатов, среднее покрытие, среднее качество выравнивания, общее число ошибок (инсерций, делеций) при выравнивании.

Затем программой samtools index для всех bam файлов был создан индекс. Это необходимо для получения матрицы корреляций файлов выравнивания с помощью пакета программ deepTools 2.0. Сначала программой multiBamSummary был создан сводный файл для всех файлов выравнивания. Затем к нему применили программу plotCorrelation, для того чтобы построить тепловую карту корреляции bam файлов между собой. Для этого файлы были разбиты на интервалы длиной 100 п.н. Сравнение проводили по этим интервалам. Применяли корреляцию Спирмена, так как она менее чувствительна к выбросам и не предполагает нормального распределения данных.

2.4 Обработка данных пакетом MEDIPS

Пакет MEDIPS статистического языка R был разработан для анализа данных полученных в экспериментах с использованием иммунопреципитации метилированной ДНК (MeDIP) с последующим секвенированием (MeDIP-seq). Тем не менее, он предоставляет необходимый функционал для анализа других

типов количественных данных секвенирования (ChIP-seq, MBD-seq, CMS-seq и др.), включая расчет дифференциального покрытия между группами образцов, а также анализ насыщения (saturation analysis) и корреляционный анализ.

2.4.1 Анализ насыщения ридами

Анализ насыщения проводится для определения достаточно ли выровненных ридов для получения насыщенного и воспроизводимого профиля референсного генома. Только в случае того, если коротких ридов достаточно, то геномный профиль покрытия будет воспроизводимым на другом независимом наборе коротких ридов. Анализ насыщения неспецифичен для данных MeDIP-seq и поэтому может быть использован для нашего эксперимента.

При анализе насыщения геном, покрытый ридами, делится на два различных случайных набора (A и B). Оба набора A и B затем снова делятся на случайные сабсеты, и число сабсетов определяется параметром `nit` (в нашем эксперименте 50). Для каждого набора A и B, итеративно выбирается возрастающее число сабсетов и рассчитывается покрытие короткими ридами в окнах, размер которых устанавливается заранее (100 в нашем эксперименте). На каждой итерации покрытия текущих сабсетов A и B сравниваются при помощи корреляции Пирсона. Так как число ридов растет с каждой итерацией в каждом наборе, то предполагается, что получающиеся геномные покрытия становятся более похожими и эта зависимость выражается в увеличении корреляции.

Важной особенностью анализа насыщения является то, что он может быть сделан только на двух независимых наборах коротких ридов. Таким образом, насыщение может быть рассчитано только для половины доступных коротких ридов. Так как нам необходимо оценить воспроизводимость всех данных в каждом образце, за анализом исходных данных следует анализ оценочных (estimated) данных. Для получения оценочных данных исходный набор данных умножается на два, учитывая каждый регион дважды. То есть

умножается не покрытие, а сам регион. После анализа насыщения исходных данных проводится анализ насыщения оценочных данных. В данном исследовании использовалась непараметрическая ранговая корреляция, которая не требует нормального распределения и не зависит от выбросов. Результаты анализа насыщения представлены в таблице 5. Программа для анализа насыщения представлена в приложении Е.

2.4.2 Покрытие ридами контекста CpG

Проверка покрытия ридами контекста CpG при правильном ходе эксперимента должна показать, что в иммунопреципитированных образцах число ридов не покрывающих CpG контекст должно быть меньше, чем в контрольных образцах. Также во время проверки покрытия вычисляется глубина покрытия на CpG контекстах. Выходные данные представляются в виде круговых диаграмм. Для лучшей читаемости и экономии места результаты данной проверки собраны в таблицы 7,8.

2.4.3 Получение профилей метилирования

Для большей эффективности работы был написан скрипт для параллельного считывания подготовленных bam файлов. Параметры запуска с комментариями находятся в приложении А. На вход подавались иммунопреципитированные файлы и контрольные файлы, согласно руководству MEDIPS. Геном был разделен на окна по 100 нуклеотидов. Была применена нормализация, для каждого окна был рассчитан показатель rms - relative methylation score. Согласно автору пакета MEDIPS, нормализация rms учитывает региональную плотность паттернов CpG и лучше коррелирует с результатами бисульфитного секвенирования.

2.5 Определение обогащенных регионов набором программ MACS2

Сначала качество обогащения каждого файла выравнивания было проверено программой phantompeakqualtools.

Также был проверен подход получения пиков обогащения (peak calling) с помощью MACS2. Набор программ изначально разработан для ChIP-seq экспериментов и не учитывает влияния плотности CpG островков на обогащение ДНК. Однако, MACS2 использовался другими авторами в MBD-seq экспериментах, поэтому было решено проверить его работу с этими данными. Программа запускалась с опцией broadPeaks для получения широких регионов. Полный набор параметров описан в приложении А.

2.6 Сравнение профилей метилирования MEDIPS и MACS2 с профилем бисульфитного секвенирования

Для сравнения профилей метилирования, полученных после анализа пакетом MEDIPS, были созданы файлы формата bed, включающие координаты окон, в которых relative methylation score был больше 0.

Сравнение геномных интервалов проводилось набором программ bedtools, в частности bedtools intersect. В результате вычислили процент контекстов CpG, который пересекается с профилем бисульфитного секвенирования.

Таким же образом проверили пересечение полученных пиков программой MACS2 и метилированными контекстами CpG в профиле бисульфитного секвенирования.

3 Результаты

Обработка данных с подобранными параметрами заняла 5 суток с задействованием 40 процессорных ядер и более чем 150 гигабайт оперативной

памяти на вычислительном комплексе СФУ. Этот анализ с таким объемом данных можно провести только на мощном сервере, и не представляется возможным проводить на персональном компьютере.

Изъято 12 страниц.

ЗАКЛЮЧЕНИЕ

В ходе работы был оценен метод детекции метилирования ДНК MBD-seq для масличной пальмы. Было произведено выравнивание коротких прочтений на референсный геном, проверка качества прочтений и выравнивания. Средствами пакета MEDIPS была также проведена проверка качества, а именно были проверены насыщение ридов и покрытие ридов контекста CpG.

Профили метилирования были получены пакетом MEDIPS, а гиперметилированные участки программой MACS2. Получили корреляции между профилями MEDIPS и рассчитали IDR для пиков MACS2.

В данных наблюдается несогласованность, особенно в технических репликах. Анализ насыщения показал, что ридов недостаточно для воспроизводимости эксперимента. Результаты MACS2 и MEDIPS слабо соотносятся.

Можно сделать вывод, что MBD-seq именно в том виде, в котором он был проведен не подходит для масличной пальмы. Требуется дополнительный эксперимент, в котором можно проверить разные концентрации соли при элюции преципитированных образцов. Также нужно учитывать, что метод MBD-seq был разработан и активно использовался на млекопитающих и могут существовать факторы, затрудняющие его использование на масличной пальме.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Waddington C. H. Canalization of development and the inheritance of acquired characters //Nature. – 1942. – Т. 150. – №. 3811. – С. 563.
2. Nanney D. L. Epigenetic control systems //Proceedings of the National Academy of Sciences of the United States of America. – 1958. – Т. 44. – №. 7. – С. 712.
3. Ванюшин Б. Ф. Эпигенетика сегодня и завтра //Вавиловский журнал генетики и селекции. – 2015. – Т. 17. – №. 4/2. – С. 805-832.
4. Rieppel O. Atomism, epigenesis, preformation and pre-existence: a clarification of terms and consequences //Biological journal of the Linnean Society. – 1986. – Т. 28. – №. 4. – С. 331-341.
5. Шишкин М. А. Индивидуальное развитие и уроки эволюционизма //Онтогенез. – 2006. – Т. 37. – №. 3. – С. 179-198.
6. Van der Woude MW. Some types of bacterial phase variation are epigenetic. // Microbe. –2008& – 3:21. —26.
7. Riggs AD. Epigenetic Mechanisms of Gene Regulation: Overview of epigenetic mechanisms // Riggs AD, Porter TN; под редакцией Russo VEA, Martienssen R, Riggs AD. – New York: Cold Spring Harbor Laboratory Press, 1996. – с. 29–45.
8. Cohen G.N, Monod J. Bacterial permeases // Bacterial Rev. – 1957. – Sep;21(3). – с. 169–194.
9. 7. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon // Proc Natl Acad Sci USA. – 1957. – 43(7). – с. 553–566.
10. Satory D, Gordon AJ, Halliday JA, Herman C. Epigenetic switches: can infidelity govern fate in microbes? // Curr Opin Microbiol. – 2011. – 14:212–217. Epub 2011 Apr 19.
11. Cao Y, Lu HM, Liang J. Probability landscape of heritable and robust epigenetic state of lysogeny in phage lambda // Proc Natl Acad Sci USA. – 2010. – 107. – с. 18445–18450. Epub 2010 Oct 13.

12. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins // *J Mol Biol.* – 1961. – 3. – c. 318–356.
13. Hollick JB, Dorweiler JE, Chandler VL. Paramutation and related allelic interactions // *Trends Genet.* – 1997. – 13. – c. 302–308. Epub 1997 Aug 1.
14. Geoghegan JL, Spencer HG. The evolutionary potential of paramutation: a population-epigenetic model // *Theor Popul Biol.* – 2013. – 88. –c. 9–19.
15. Morgan TH. An attempt to analyze the constitution of the chromosomes on the basis of sex-limited inheritance in *Drosophila* // *J Exp Zool.* – 1911. – 11. –c. 365–413.
16. Sturtevant AH. A third group of linked genes in *Drosophila ampelophila* // *Science.* – 1913. – 37. – c. 990–992.
17. Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III // *J Exper Med.* – 1944. – 79. – c. 137–158.
18. Hershey AD, Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage // *J Gen Physiol.* – 1952. – 36. – c. 39–56.
19. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid // *Nature.* – 1953. – 171. – c. 737–738.
20. Franklin RE, Gosling RG. Molecular configuration in sodium thymonucleate // *Nature.* – 1953. –171. –c. 740–741.
21. Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome // *Am J Hum Genet.* – 1962. – 14. – c. 135–148.
22. Beisson J, SonnebornTM. Cytoplasmic inheritance of the organization of the cell cortex in *Paramecium Aurelia* // *Proc Natl Acad Sci USA.* – 1965. – 53. – c.275–282.
23. Kornberg RD, Thomas JO. Chromatin structure; oligomers of the histones // *Science.* – 1974. – 184. – c. 865–868.
24. Bolton DC, McKinley MP, Prusiner SB. Identification of a protein that purifies with the scrapie prion // *Science.* – 1982. – 218. – c. 1309–1311.

25. Prusiner SB, Bolton DC, Groth DF, Bowman KA, Cochran SP, McKinley MP. Further purification and characterization of scrapie prions // *Biochemistry*. – 1982. – 21. – c. 6942–6950.
26. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis // *Proc Natl Acad Sci USA*. – 1964. – 51. – c. 786–794.
27. Riggs AD. X inactivation, differentiation, and DNA methylation // *Cytogene Cell Gene*. – 1975. – 14. – c. 9–25.
28. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development // *Science*. – 1975 – 187. – c. 226–232.
29. Felsenfeld G. A brief history of epigenetics // *Cold Spring Harb Perspect Biol*. – 2014. – 6. [DOI:10.1101/cshperspect.a018200]
30. Bird A. Perceptions of epigenetics // *Nature*. – 2007. – 447. – c. 396–398.
31. Deans C, Maggert KA. What do you mean, ‘epigenetic’? // *Genetics*. – 2015. – 199. – c. 887–896.
32. Mann JR. Epigenetics and memigenetics // *Cell Mol Life Sci*. – 2014. – 71. – c. 1117–1122.
33. Luger K. Encyclopedia of life sciences: Nucleosomes: structure and function // под редакцией Finazzi-Agró A. – Wiley- Blackwell. – 2001. с. 1–8.
34. Battistini F, Hunter CA, Gardiner EJ, Packer MJ. Structural mechanics of DNA wrapping in the nucleosome // *J Mol Biol*. – 2010. – 396. – c. 264–279.
35. Szyf M. The early life environment and the epigenome // *BBA*. 2009. – 1790. – c. 878–885.
36. Choudhuri S. From Waddington’s epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research // *Toxicol Mech Method*. – 2011. – 21. – c. 252–274.
37. Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis* // *J Mol Biol*. – 1978. – 118(1). – c. 27–47.

38. Cattanach BM, Kirk M. Differential activity of maternally and paternally derived chromosome regions in mice // *Nature*. – 1985. – 315(6019). – c. 496-498.
39. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene // *Nature*. – 2000. – 405. – c. 482–485.
40. Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases // *Chem- biochem*. 2011. – 12. – c. 206–222.
41. Chen ZX, Riggs AD. DNA methylation and demethylation in mammals // *J Biol Chem*. – 2011. – 286. – c. 18347–18353.
42. Andrews S. et al. FastQC: a quality control tool for high throughput sequence data. – 2010.
43. Bolger A. M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // *Bioinformatics*. – 2014. – T. 30. – №. 15. – C. 2114-2120.
44. Langmead B., Salzberg S. L. Fast gapped-read alignment with Bowtie 2 // *Nature methods*. – 2012. – T. 9. – №. 4. – C. 357.
45. García-Alcalde F. et al. Qualimap: evaluating next-generation sequencing alignment data // *Bioinformatics*. – 2012. – T. 28. – №. 20. – C. 2678-2679.
46. Lienhard M. et al. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments // *Bioinformatics*. – 2013. – T. 30. – №. 2. – C. 284-286.
47. Landt S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia // *Genome research*. – 2012. – T. 22. – №. 9. – C. 1813-1831.

ПРИЛОЖЕНИЯ

Изъято 10 страниц.

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт фундаментальной биологии и биотехнологии
Базовая кафедра защиты и современных технологий мониторинга лесов

УТВЕРЖДАЮ

Заведующий кафедрой

 И.Е.Ямских

подпись инициалы, фамилия

« 25 » июль 2019г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Определение гипер- и гипо- метилированных участков ДНК и
фенотипических последствий эпигенетических изменений с помощью
методов иммунопреципитации

06.04.01 – Биология

06.04.01.06 – Геномика и биоинформатика

Научный руководитель


подпись, дата

доцент, PhD
должность, ученая
степень

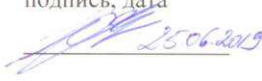
Татаринова Т.В.
инициалы, фамилия

Выпускник


подпись, дата

Поплаухин А.А.
инициалы, фамилия

Рецензент


подпись, дата

Профессор.
д. б. н.
должность, ученая
степень

Орлов Ю.Л.
инициалы, фамилия

Красноярск 2019