

## ОБЗОР ОТЗЫВОВ,

поступивших на диссертацию и автореферат диссертации Колмыкова Семёна Константиновича,

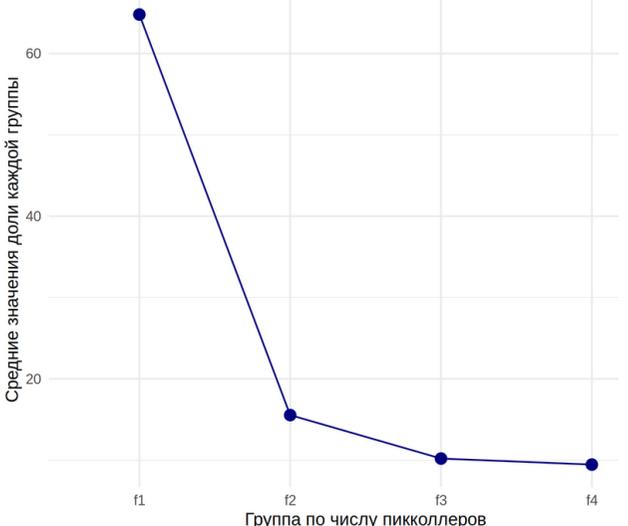
представленную на соискание ученой степени кандидата биологических наук

по научной специальности 1.5.8. Математическая биология, биоинформатика,

тема: Разработка методов контроля качества и построения карты геномных районов связывания транскрипционных факторов на основе сравнительного анализа ChIP-seq экспериментов

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
от Рогаева Евгения Ивановича, доктора биологических наук, профессора, академика РАН, научного руководителя научного центра генетики и наук о жизни автономной некоммерческой образовательной организации высшего образования «Научно-технологический университет «Сириус», федеральная территория «Сириус»	1. Одним из замечаний, которое хотелось бы отметить, является недостаточное освещение автором в литературном обзоре роли различных молекулярных механизмов в конкретных нарушениях морфологии сперматозоидов. Кроме того, в разделе, посвященном полученным результатам, не раскрыт вопрос биологической интерпретации выявленных однонуклеотидных вариантов. Хотелось бы увидеть более подробное обсуждение потенциальных механизмов, с помощью которых данные геномные варианты могут приводить к нарушениям морфологии сперматозоидов. Это позволило бы глубже понять биологические последствия генетической изменчивости в контексте патогенеза сперматогенеза.	С замечанием согласен.
	2. В работе представлено решение для ChIP-seq экспериментов. Интересно узнать, рассматривалась ли возможность адаптации предложенных алгоритмов для других типов данных, например, для DNase-seq или ATAC-seq,	Разработанные методы и подходы могут быть применены к любым позиционным методам NGS, в результате обработки которых являются наборы геномных районов. В программный модуль METARA, в частности, добавлена

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>которые также имеют проблемы с шумом и вариабельностью?</p>	<p>возможность проводить сравнительный анализ по имеющимся в БД GTRD DNase-seq данным. В данном случае объединение будет происходить на уровне клеточных типов. Однако отдельного исследования получаемых мета-кластеров открытого хроматина не проводилось.</p>
	<p>3. Описанный в диссертационной работе метод приоритизации районов связывания ТФ позволяет выделять достоверные районы связывания. Использовались ли дополнительные низкопроизводительные методы картирования районов связывания ТФ для валидации полученных результатов? Поскольку хранящиеся в базе данных GTRD ChIP-seq эксперименты получены из разных клеточных типов в разных условиях, интересно, как на результаты работы данного алгоритма влияет биологическая гетерогенность используемых ChIP-seq данных?</p>	<p>Такой работы не проводилось, однако отбор и использование результатов применения низкопроизводительных методов идентификации взаимодействия ТФ с ДНК таких как: SELEX и EMSA, полученных из анализа литературных данных, послужило бы дополнительной валидацией разработанных подходов приоритизации районов связывания ТФ.</p>
<p>от Афонникова Дмитрия Аркадьевича, доктора биологических наук, доцента, ведущего научного сотрудника, заведующего лабораторией эволюционной биоинформатики и теоретической генетики Федерального государственного</p>	<p>1. Мера FPCM предложенная автором для оценки доли ложноположительных предсказаний РСТФ основана на предположении, что «неизвестное число подлинных РСТФ является случайной величиной с распределением Пуассона», однако какого-либо обоснования этого не приведено. На чем оно основано и можно ли каким-то образом проверить такое предположение?</p>	<p>На рисунке ниже представлены средние доли групп (F1-F4) в более чем 11 тысяч ChIP-seq экспериментов.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
<p>бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН), г. Новосибирск</p>		 <p>На основе данного графика было сделано предположение (на основе визуальной схожести функции вероятности), что число истинных районов связывания в каждой из групп (F1-F4) соответствует распределению Пуассона. В качестве дополнительной проверки интересно было бы добавить еще несколько дополнительных алгоритмов идентификации пиков в ChIP-seq экспериментах, чтобы получить больше информации о наблюдаемом распределении.</p> <p>Следует обратить внимание, что разработанная характеристика базируется на оценке отклонения</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		от распределения Пуассона. Предполагается, что в группе F1 содержится большое количество истинных событий связывания. Я согласен с замечанием, что данное распределение не самым лучшим образом описывает исследуемый объект, однако все еще позволяет детектировать отклонения в доле F1 РСТФ. Более оправданным было бы использование биномиального, или даже отрицательного биномиального распределения, что будет сделано в последующих работах.
	2. На рисунке 3.4.4 приведен несколько неожиданный результат: в области РСТФ для белка ТВР наблюдается наименьшая среди всех факторов доля пиков с мотивом связывания ТФ. Почему такое происходит, ведь мотив ТВР, ТАТА-бокс, известен как классический и наиболее изученный ранее мотив, а связывание его с ТВР это сигнал к старту транскрипции для большинства генов? Чем такой результат можно объяснить?	В отличие от других организмов, например, дрожжей, в геноме человека большинство промоторов не содержат ТАТА-боксы, что также подтверждается ChIP-seq данными. Данные результаты согласуются с литературными данными (Venters et al., 2013). Таким образом, ТВР способен связываться с альтернативными участками в промоторных областях генов, либо участвует в процессе инициации транскрипции опосредованно через другие ТФы.  Venters B. J., Pugh B. F. Genomic organization of human transcription initiation complexes //Nature. – 2013. – Т. 502. – №. 7469. – С. 53-58.
	3. К замечаниям следует отнести небрежность в оформлении текста диссертации: (1) Встречается три разных рисунка с одним номером 3.1.2, два рисунка с номером 3.1.1; (2) на некоторых	С замечанием согласен.

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>рисунках присутствуют надписи на английском; (3) не для всех сокращений приведена расшифровка в специальном списке, например ФАФ; (4) в некоторых местах показатель степени значения p-value не приведен в верхнем регистре; (5) на рис. 3.1.2 стр. 74 пропущена панель E, на рис. 3.3.2 не подписаны и не описаны панели; неудачные фразы – кальки с английского (орфаны, корзина и пр.), опечатки.</p>	
	<p>4. Полагаю, что в формулировке цели работы напрасно отсутствует фраза о практическом применении комплекса разработанных методов для исследования связи между однонуклеотидными вариациями и нарушением структуры сперматозоидов у мужчин, проживающих на территории Российской Федерации.</p>	<p>Данная часть работы нужна для демонстрации применения разрабатываемых в диссертационной работе методов и подходов для интерпретации однонуклеотидных геномных вариантов, поэтому не была вынесена в цель.</p>
<p>от Головина Андрея Викторовича, доктора химических наук, профессора научного центра генетики и наук о жизни направления «Вычислительная биология» автономной некоммерческой образовательной организации высшего образования «Научно-технологический университет «Сириус», федеральная территория «Сириус».</p>	<p>1. В обзоре литературы автор описывает сложность предсказания вариантов влияния SNV на уровень экспрессии, для читателя было бы интересно узнать более детально влияние SNV на связывание факторов транскрипции в том числе и структурные аспекты этого процесса.</p>	<p>С замечанием согласен.</p>
	<p>2. В разделе 3.5 при обсуждении интерпретации SNV ассоциированных с нарушениями сперматогенеза, автор приводит лаконичную таблицу с названиями генов и детальное описание кандидатных генов текстом, было бы значительно удобнее сделать общее описание влияния той или иной SNV на клеточные процессы.</p>	<p>С замечанием согласен. Было бы полезно ознакомить читателя, например, с принадлежностью генов, изменяющих экспрессию в зависимости от присутствия рассматриваемых аллелей в генотипе. Данный анализ будет проведен в продолжении данной работы.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	3. Неприятной деталью оформления мне показалось неуместное использование светло серого фона в главе 3.5 для отображения графиков типа Violin Plot, белый фон является более уместным для графиков. Надо отметить, что оформление графиков в других главах мне показалось гармоничным.	С замечанием согласен.
от Кулаковского Ивана Владимировича, доктора биологических наук, кандидата физико-математических наук, ведущего научного сотрудника группы регуляции биосинтеза белка Федерального государственного бюджетного учреждения науки «Институт белка Российской академии наук», г. Пущино.	1. Автор вольно и взаимозаменяемо использует термины SNV (однонуклеотидный вариант) и SNP (однонуклеотидный полиморфизм), хотя последний акроним соответствует более узкому термину, который отражает не любые варианты, а лишь достаточно распространенные в популяции.	С замечанием согласен. В тексте диссертации следовало бы использовать только термин SNV, поскольку читателя не приводятся данные о частоте данных однонуклеотидных геномных вариантов
	2. В обзоре литературы присутствует раздел, посвященный морфологии сперматозоидов. Не вполне понятно насколько этот раздел информативен, учитывая, что апробация авторских методов на задаче функциональной аннотации однонуклеотидных вариантов не достигает стадии интерпретации и проверки конкретного молекулярного механизма и связи с физиологией процесса сперматогенеза.	С замечанием согласен
	3. С точки зрения оформления работы, часть рисунков представляет собой снимки экрана («скриншоты») программы BioUML. Такие технические материалы не следовало включать в работу вовсе, либо расположить в приложениях. Также, работа не избежала использования жаргона, например «пикколлеры» на рисунке 3.1.2.	С замечанием согласен. Термин пикколлер был введен на 23 странице диссертации, как наиболее емкий термин для обозначения алгоритмов идентификации пиков. Однако он продолжает оставаться жаргонизмом.

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
от Миронова Андрея Александровича, доктора биологических наук, кандидата физико-математических наук, профессора факультета биоинженерии и биоинформатики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М.В.Ломоносова», г. Москва.	1. Таблица 1.1.1., посвященная экспериментальным методам, связанным с NGS, либо избыточна, поскольку содержит ссылки на методы, не связанные непосредственно с темой работы, либо неполна, поскольку не указаны, например, методы определения РНК-хроматинового интерактома. В обзоре не упомянут протокол ChIP-ехо. Имеющий непосредственное отношение к теме исследования.	С замечанием согласен.
	2. Раздел про строение и аномалию сперматозоидов смотрится странно. Следовало бы дать короткое введение – зачем этот раздел в этом месте. Более того я бы рекомендовал этот раздел сделать последним разделом введения.	С замечанием согласен. Схожие замечания, касающиеся этой главы лит. обзора, встречаются и в других отзывах на диссертацию. Следовало бы расширить данную главу
	3. Стр. 46 «размер свободных от хроматина участков ДНК» – неудачное выражение, поскольку ДНК является непосредственным компонентом хроматина.	С замечанием согласен
	4. В разделе 3.1 проводится сравнение разных методов поиска пиков. Насколько я понял сравнение предсказаний производилось на бинах (рис. 3.1.1) и для каждого бина вводятся признак – сколькими программами поиска пиков в бине обнаружен сигнал. Здесь требуется более аккуратное описание – размер бина, что делается, если пик перекрывает границу бина и пр.	При разбиении найденных РСТФ на четыре группы не было этапа разбиения генома на корзины фиксированной длины. Для того, чтобы отнести РСТФ к конкретной группе выполнялась операция пересечения границ всех РСТФ, найденных четырьмя алгоритмами идентификации пиков. На основании числа пиков (кол-ва алгоритмов), попавших в это пересечение, делался вывод о принадлежности к РСТФ к определенной группе.
	5. Поскольку обнаружение сигнала программами поиска пиков опирается на достаточно произвольно выбранные порог, то возникает	Согласен с замечанием – дополнительная иллюстрация, демонстрирующая относительную воспроизводимость методов, позволила бы

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>вопрос – может при удачном выборе порога в каждой программе можно добиться консистентности? В любом случае была бы поучительна диаграмма Венна, показывающая число пиков (бинов) для каждого метода. На такой диаграмме наглядно можно было бы увидеть сравнение методов.</p>	<p>читателю лучше познакомиться с выбранными алгоритмами идентификации пиков. Рекомендованные разработчиками данных алгоритмов пороговые значения, или использование одинаковых значений p-value или FDR не приводят к одинаковым объемам наборов РСТФ. Вероятно, при варьировании пороговых значений значимостей можно добиться большей консистентности данных алгоритмов. Можно ожидать, что при уменьшении пороговых значений значимости, помимо уменьшения общего числа РСТФ снизится и их вариабельность. В этом случае разница в результатах между алгоритмами идентификации пиков будет обуславливаться различиями в статистических моделях и подходах, лежащих в основе этих методов.</p> <p>Также, в работе Bailey с соавторами рекомендуется использовать динамические пороговые значения, основанные на анализе воспроизводимости результатов обработки ChIP-seq экспериментов.</p> <p>Bailey T. et al. Practical guidelines for the comprehensive analysis of ChIP-seq data // PLoS computational biology. – 2013. – Т. 9. – №. 11. – С. e1003326.</p>
от Орлова Юрия Львовича, доктора биологических наук, профессора РАН, профессора	1. Раздел по интерпретации однонуклеотидных геномных вариаций, ассоциированных с нарушениями сперматогенеза представлен	С замечанием согласен.

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
<p>кафедры информационных и интернет-технологий Федерального государственного автономного образовательного учреждения высшего образования «Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации» (Сеченовский Университет), г. Москва.</p>	<p>слишком кратко. Стоило бы его расширить как важную практическую часть работы.</p>	
	<p>2. Следует избегать сокращений в заголовках, выводах, заключительных фразах, тем более в смеси сокращений на русском и английском. Например «ФАФ METARA»</p>	<p>С замечанием согласен.</p>
	<p>3. Аббревиатура РСТФ для районов связывания транскрипционных факторов не является общепринятой. Чаще пишут ССТФ – «сайты» не «районы» в литературе и на русском, и на английском языке.</p>	<p>В работе используются обе аббревиатуры. Это было сделано, чтобы подчеркнуть их различия. Под РСТФ понимаются, как правило, достаточно протяженные геномные районы (150-200 п.н.), соответствующие пикам в ChIP-seq экспериментам, тогда как под ССТФ в работе понимались небольшие геномные районы (10-20 п.н.), соответствующие сайтам связывания ТФ с ДНК.</p>
<p>4. Замечание по нумерации рисунков - работа содержит 35 рисунков со сложной, вводящей в заблуждение тройной нумерацией. При этом Рисунок номер 3.1.1 представлен два раза (разные рисунки, один номер), а Рисунок 3.1.2 даже три раза (одинаковый номер). Конкретнее по дублям с одинаковыми номерами:  Рисунок 3.1.1 – Схема пересечения результатов работы алгоритмов идентификации пиков  Рисунок 3.1.1 – Плотности распределений результатов идентификации РСТФ  Рисунок 3.1.2 – Плотности распределений результатов идентификации РСТФ  Рисунок 3.1.2 – Взаимосвязь различных геномных аннотаций</p>	<p>С замечаниями согласен.</p>	

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>Рисунок 3.1.2 – ROC-кривые, полученные для ChIP-seq эксперимент</p> <p>Указаны несуществующие панели рисунков, например Рисунок 3.1.1Ж.</p> <p>В то же время есть ссылки (см. Рисунок 3.2.1Б) без соответствующей панели.</p> <p>«(см. Рисунок 9)» - такого номера рисунка нет в работе.</p> <p>Подпись к рисунку 3.5.2 относится к чему-то другому – см. «Значение над медианой в “ящике с усами” указывает на количество образцов с выбранным генотипом» и далее -</p> <p>Рисунок 3.5.2 – (А)</p> <p>Рисунка номер 3.5.1 – нет, но есть следующий номер 3.5.2 и далее.</p> <p>Панели рисунков (А,Б,В,Г..) подписаны некорректно.</p> <p>Например, для рисунков 3.2.4-5, 6 и 7 указано в подписи – (АД). Следует хотя бы написать (панели А,Б,В,Г,Д), указать слово «панель» или «фрагмент»</p>	
	<p>5. Нужно сразу ставить ссылки на литературу на биологические результаты, упомянутые в тексте, Например «В последние несколько десятилетий в различных регионах мира наблюдается снижение мужского репродуктивного потенциала..» (стр.6) Где это показано, в каких регионах мира, где ссылка?</p>	<p>С замечаниями согласен.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>«Большинство известных однонуклеотидных геномных вариантов (SNV) расположено в регуляторных областях генов...» (стр.6) – Это утверждение тоже нужно подтвердить ссылкой. Можно также сказать, что большинство вариантов как раз аннотировано в белок-кодирующих частях генов. Нужна ссылка на литературу.</p> <p>«В 2022 году Suryatenggara с соавт. была опубликована статья» - нужна ссылка на эту статью, журнал (добавить ссылку в текст в месте первого упоминания)</p> <p>Нужны ссылки на литературу или интернет-ссылки (линки) на упоминаемые электронные ресурсы всюду в тексте, например проект ENCODE, ENCODE Portal, CistromeDB (стр.7).</p> <p>См также стр.11: «... базах данных: SRA, GEO и ENCODE»</p> <p>Это конкретные известные ресурсы – указать о чем идет – база данных, портал, указать конкретную ссылку, лучше на научные статьи, можно и то и другое – ссылка на статью и веб-сайт)</p> <p>Стр.12: «в соответствии с рекомендациями GATK Best Practices» - нужна ссылка на эти рекомендации, что такое GATK?</p> <p>Стр.13 – опечатки – падежи во фразе про конференции «Международная конференция... Международной конференции...». Болтгие буквы «И» в названиях конференций не нужны.</p>	

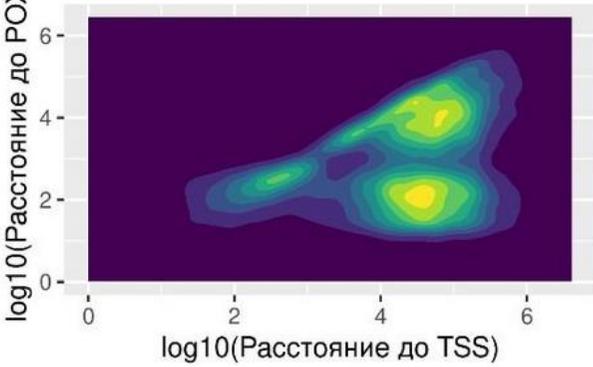
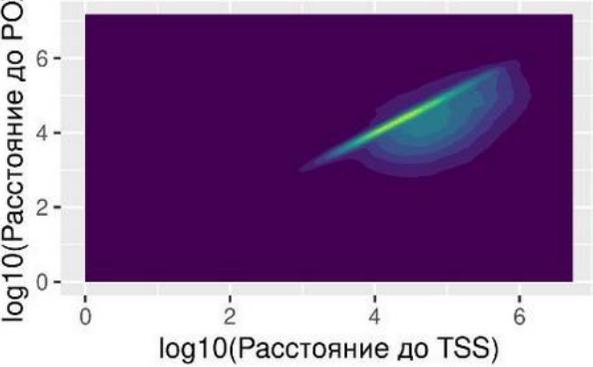
Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>Следует пронумеровать формулы в работе, в том числе указывать все параметры, стандартно выделяя параметры курсивом.</p> <p>Например «В данном контексте <math>\lambda</math> описывает ожидаемое количество прочтений в рассматриваемом окне поиска пика» - надо отметить, что такое <math>k</math> в формуле (в одной из формул это число событий, в другой число прочтений ДНК).</p> <p>Цитирование статей в тексте достаточно по фамилиям авторов и году, без инициалов, например (Gaffney D. J. et al., 2012) – инициалы автора избыточны.</p> <p>Стр. 38 – опечатки в цитировании «(Y et la., 2014) и (Zhou et Troyanskaya, 2015)» - должно быть “et al”?</p>	
	<p>6. Сокращения в Списка сокращений не всегда соответствуют терминологии, например - FN - ложно свидетельствующий об отрицательном результате (False Negative).</p> <p>Можно указать как «число ошибочных ложных предсказаний», но не как «ложно свидетельствующий». Это стандартная терминологии, непонятен новый введенный термин. То же замечание для других сокращений в списке.</p>	С замечанием согласен.
	<p>7. Есть замечания и по нумерации рисунков автореферате (например Рисунок 3.4.1) – избыточная нумерация. Формулы в автореферате представлены без детализации, не все параметры</p>	С замечаниями согласен. В данной формуле допущена ошибка: дважды повторена операция логарифмирования

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	указаны. Например, в сложной формуле $\log\log\dots$ хотя бы поставить скобки, от чего берется аргумент функции.	
от Попова Даниила Викторовича, доктора биологических наук, профессора РАН, ведущего научного сотрудника, заведующего лабораторией физиологии мышечной деятельности Федерального государственного бюджетного учреждения науки «Государственный научный центр Российской Федерации Институт медико-биологических проблем Российской академии наук», г. Москва.	1. Работа имеет ряд опечаток (особенно в разделе 3.5, посвященном идентификации и анализу однонуклеотидных геномных вариантов) и неточностей в обозначении номеров рисунков. Часть используемых в тексте аббревиатур не представлена в списке сокращений, что несколько затрудняет восприятие результатов работы.	С замечанием согласен.
	2. При описании степени разработанности темы отсутствует какая-либо информация о литературных данных по ассоциации однонуклеотидных геномных вариантов с нарушениями сперматогенеза. Более того, при обсуждении собственных данных нет описания того, насколько полученные результаты соответствуют литературе и что они добавляют в этой области.	С замечанием согласен.
	3. При описании методов, автор указал, что «процесс отбора участников (эксперимента) подробно описан в работе Osadchuk и соавторов». Отсутствие в диссертации этой важной информации затрудняет интерпретацию данных по ассоциациям с нарушением структуры сперматозоидов, представленных в работе. Помимо этого, хотелось бы видеть хотя бы краткое описание методов подготовки проб и протокола анализа для полноэкзомного секвенирования, что	Для проведения полноэкзомного секвенирования из доступных в рамках популяционного исследования мужского репродуктивного потенциала жителей РФ были выбраны 367 образцов. Исследуемая выборка состояла из 174 участников с нормоспермией и 193 участников с различными формами патоспермии. Среди них 145 человек имели олигоспермию, 26 – азооспермию, 188 – астеноспермию, и 105 – тератоспермию. У некоторых участников

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>также важно для корректной интерпретации полученных данных.</p>	<p>наблюдались комбинированные нарушения качества семенной жидкости.</p> <p>Подготовка проб и секвенирование:  Для создания экзомных библиотек был использован набор Illumina TruSeq DNA Library Prep for Enrichment с xGen® Exome Research Panel v1.0, поскольку данный набор за счёт увеличенной равномерности покрытия в целевых зонах позволяет анализировать 16 экзотов за один запуск секвенатора NextSeq 550 вместо 12. Создание геномных библиотек проводилось в основном согласно мануалу производителя. Исходные концентрации ДНК в пробах определялись спектрофотометрически, 50 мкл раствора, содержащего 2-4 мкг ДНК, подвергалось ультразвуковой фрагментации на приборе Covaris M220 с параметрами, оптимизированными для получения максимума фрагментов в диапазоне 150-200 п.н. Полученные фрагменты очищались сорбцией на магнитных частицах, растворялись в 30 мкл H<sub>2</sub>O, концентрация ДНК определялась флуориметрически на приборе Qubit набором High Sensitivity. Далее 100 нг фрагментированной ДНК в объёме 60 мкл использовалось для создания геномных библиотек: после репарации концов производилась селекция фрагментов нужной длины. Поскольку в экзомных данных нам было необходимо получение точной длины</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		<p>районов тринуклеотидных повторов в гене андрогенового рецептора, был выбран вариант создания геномных библиотек со средним размером встройки ~150 п.н. Для этого протокол селекции по длине магнитными шариками SPB был модифицирован – как первый шаг к объёму смеси добавлялось 86 мкл SPB, как второй шаг – к 180 мкл супернатанта добавлялось 125 мкл SPB. Дальнейшая процедура проводилась согласно мануалу производителя, амплификация библиотек велась 9 циклов. Качество и молярность полученных библиотек определялись с помощью биоанализатора BA2100, набор DNA 1000, точные количества библиотек определялись флуориметром Qubit. Получение собственно экзомных библиотек с панелью Illumina xGen® Exome Research Panel v1.0 выполнялось согласно мануалу производителя. Для этого взято 125 нг каждой библиотеки (расчёт 4plex), концентрирование выполнялась после добавления COTDNA добавлением 1.8 объёма AMPureXP. Гибридизация проводилась в течение 5 часов, при отмывке горячими буферами магнитный штатив также нагревался до 65 градусов цельсия. Секвенирование полученных экзомных библиотек проведено на имеющемся в нашем институте ИЦиГ СО РАН приборе NextSeq550 (400 млн. расчётных чтений, длина прямого чтения 150 п.н.) набором NextSeq 500/550 High Output Kit v2.5 (150 Cycles).</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>4. Хотелось бы видеть данные по средним размерам (и разбросам) для идентифицированных РСТФ и распределению плотности РСТФ относительно стартов транскрипции: имеется ли смещение распределения плотностей относительно стартов транскрипции?</p>	<p>В рамках диссертационной работы такого исследования не проводилось. Однако в стороннем исследовании мной изучался схожий вопрос. Наблюдалась четкая дифференцировка ТФ по степени удаленности РСТФ от TSS. В частности выделилось два кластера: в районе 100-1000 п.н. и 10000-100000 п.н. Тем не менее данные вопрос требует дальнейшего глубокого изучения.</p> <p>В рамках исследования взаимосвязи удаленности РСТФ от TSS было также показано, что некоторые ТФ обладают разной функциональной активностью. Например, РСТФ ТФ NRF2, расположенные ближе к TSS (&lt;1kb) ассоциированы с подавлением транскрипции, а более удаленные РСТФ - с активацией (&gt;1kb).</p> <p>В качестве примера приведу тепловую плотности дистанции HCNA ESR1 от TSS в клеточной линии T47D (Инвазивная протоковая карцинома (ИПК) молочной железы) и дистанция между РСТФ и ближайшим районов открытого хроматина (РОХ).</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		 <p data-bbox="1384 758 2033 901">А также для сравнения подобную тепловую карту, но вместо РСТФ исследуются мотивы связывания ESR1, предсказанные позиционной весовой матрицей (PWM)</p> 

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>5. Медианная воспроизводимость РСТФ даже в группе F4 не превышает 0,5 (Рисунок 3.1.2 Б). Чем объясняется такое относительно невысокое значение?</p>	<p>Большая часть РСТФ приходится на группу F1, т. е. районы связывания, идентифицированные только одним методом. Большой вклад в это разнообразие вносят рекомендованные разработчиками данных алгоритмов пороговые значения. Вероятно, при уменьшении пороговых значений значимости, помимо уменьшения общего числа РСТФ снизится и их вариабельность. В этом случае разница в результатах между алгоритмами идентификации пиков будет обуславливаться различиями в статистических моделях и подходах, лежащих в основе этих методов.</p>
	<p>6. На рисунке 3.4.3 представлены примеры разных динамик значений ФАФ и AUC для двух разных транскрипционных факторов (разная направленность изменений, наличие осцилляций в доле мета-кластеров в открытом хроматине для фактора JUN). Хотелось бы получить комментарии о причинах этих различий и особенностях динамики.</p>	<p>На рисунках приведены два контрастных случая: в каждом из которых почти константным остается либо отношение мета-кластеров в РОХ, либо доля мета-кластеров, содержащих мотивы связывания ТФ.</p> <p>Тенденция, наблюдаемая у NRF1 может свидетельствовать о чувствительности данного ТФ к наличию мотива-связывания в случае его высоковоспроизводимых РСТФ. Возможно эти районы связаны с общими для всех типов клеток процессами (например, энергетическим обменом). В свою очередь, мета-кластеры не содержащие мотивов связывания (или отклоняющиеся от консенсусного мотива), лежащие в районах открытого хроматина, могут тем самым снижать</p>

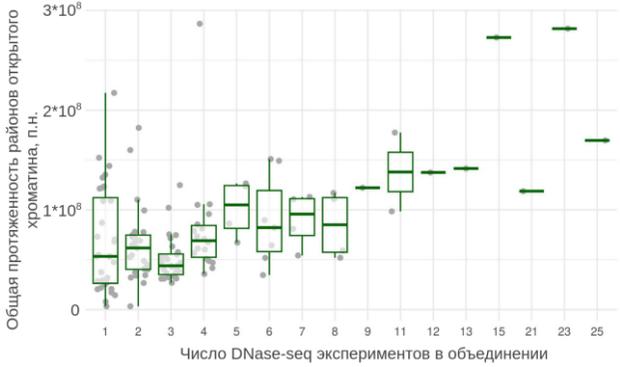
Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		<p>эффективность связывания NRF1 и обеспечивать более тонкую регуляцию экспрессии.</p> <p>Возможно, сниженная доля мета-кластеров с мотивами связывания JUN обуславливается характерной для него димеризацией с другими ТФ, а также участия в кооперативном связывании, даже в наиболее воспроизводимых РСТФ. Вероятно, волнообразный профиль кривой, отражающий долю мета-кластеров в районах открытого хроматина связан с представленностью различных клеточных типов. В связи с чем, достоверные тканеспецифичные районы связывания характерные определенной ткани дают локальные подъемы в доле мета-кластеров в РОХ. При этом данные по содержанию мотивов связывания не демонстрируют подобной тенденции. Таким образом, предложенный подход чувствителен к представленности клеточного типа или определенного условия для заданного ТФ. В связи с чем требуются дальнейшие его модификации.</p>
	<p>7. Ложноположительные результаты ChIP-seq экспериментов, связанные с наличием белок-белковых взаимодействий – это одна из ключевых проблем этого метода. Автор обсуждает эту проблему при описании высокой вариабельности в доле МСТФ и РСТФ в зависимости от транскрипционного фактора. Возможно ли</p>	<p>Действительно, наличие кооперативного связывания затрудняет картирования прямых взаимодействий ТФ, опосредованных их ДНК-связывающим доменом, как следствие, затрудняет исследования функциональной активности ТФ.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>использование каких-либо специальных подходов (при проведении ChIP-seq экспериментов или обработке их результатов) для снижения влияния белок-белок взаимодействий на результаты идентификации РСТФ. Планируется ли как-то отражать в базе данных GTRD информацию о результатах с высоким потенциальным влиянием белок-белковых взаимодействий?</p>	<p>Для выявления событий кооперативного связывания активно используется анализ ко встречаемости мотивов связывания ТФ (в данном направлении, в частности, ведет исследования группа Левицкого В.Г.), а также анализ ко-встречаемости ChIP-seq пиков. В данном контексте БД GTRD является отличным источником информации для дальнейшего анализа. Несмотря на то, что в настоящее время такие исследования нашей группой не проводятся, включение такой информации - один из самых вероятных дальнейших путей развития базы данных.</p>
	<p>8. На рисунке 3.5.3 представлена частота различных аллелей в пробах со сперматозоидами с морфологическими нарушениями. В отличие от предыдущих рисунков представлено только два генотипа. С чем связано отсутствие данных по одному из гомозиготных вариантов?</p>	<p>На рисунке представлены все три генотипа, однако в БД GTEh из-за довольно небольшого объема группы референсной гомозиготы не был построен бокс-плот/виолин-плот. Отсутствие данных связано с крайне низкой частотой гомозигот по данному аллелю. Поскольку исследуемая выборка не является репрезентативной в контексте популяции жителей РФ, по имеющимся у нас данным нельзя установить частоту аллеля в славянской популяции.</p>
<p>от Фишмана Вениамина Семеновича, кандидата биологических наук, ведущего научного сотрудника, заведующего сектором геномных механизмов онтогенеза</p>	<p>1. При разработке метода FPCM, используется распределение Пуассона. Это распределение дискретного типа случайной величины, представляющей собой число событий, произошедших за фиксированное время, при условии, что Данные события происходят с</p>	<p>На рисунке ниже представлены средние доли групп (F1-F4) в более чем 11 тысяч ChIP-seq экспериментов.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии										
1	2	3										
<p>Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» (ИЦиГ СО РАН), г. Новосибирск.</p>	<p>некоторой фиксированной средней интенсивностью и независимо друг от друга. Я считаю, что для событий обнаружения пиков разными инструментами на основе ChIP-seq-данных это предположение в корне неверно инструменты работают на одном и том же наборе данных, их результаты нельзя считать независимыми.</p>	<div data-bbox="1384 319 2016 853" data-label="Figure"> <table border="1"> <caption>Средние значения доли каждой группы</caption> <thead> <tr> <th>Группа по числу пикколтеров</th> <th>Среднее значение доли (%)</th> </tr> </thead> <tbody> <tr> <td>f1</td> <td>65</td> </tr> <tr> <td>f2</td> <td>15</td> </tr> <tr> <td>f3</td> <td>10</td> </tr> <tr> <td>f4</td> <td>8</td> </tr> </tbody> </table> </div> <p>На основе данного графика было сделано предположение (на основе визуальной схожести функции вероятности), что число истинных районов связывания в каждой из групп (F1-F4) соответствует распределению Пуассона. В качестве дополнительной проверки интересно было бы добавить еще несколько дополнительных алгоритмов идентификации пиков в ChIP-seq экспериментах, чтобы получить больше информации о наблюдаемом распределении.</p> <p>Следует обратить внимание, что разработанная характеристика базируется на оценке отклонения</p>	Группа по числу пикколтеров	Среднее значение доли (%)	f1	65	f2	15	f3	10	f4	8
Группа по числу пикколтеров	Среднее значение доли (%)											
f1	65											
f2	15											
f3	10											
f4	8											

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		<p>от распределения Пуассона. Предполагается, что в группе F1 содержится большое количество истинных событий связывания. Я согласен с замечанием, что данное распределение не самым лучшим образом описывает исследуемый объект, однако все еще позволяет детектировать отклонения в доле F1 РСТФ. Более оправданным было бы использование биномиального, или даже отрицательного биномиального распределения, что будет сделано в последующих работах.</p>
	<p>2. Связанный предыдущим вопросом: Можно ли вычислить <math>f_{1e}</math>, используя значение <math>\lambda</math>, полученное из <math>p_2</math> и <math>p_3</math> по отдельности? Повлияет ли это на результат?</p>	<p>Можно исключить из рассмотрения F2 или F3 и оценить <math>F_{1e}</math>. Однако такой анализ не проводился. Ранее был проведен достаточно поверхностный анализ качества данных, основываясь на F2, а не на F1. Были получены схожие с FPCM, основанным на F1, результаты. Однако данный вопрос требует проведения дополнительного изучения.</p>
	<p>3. С чем связан волнообразный профиль кривой для JUN на рис. 3.4.3?</p>	<p>Вероятно, волнообразный профиль кривой, отражающий долю мета-кластеров в районах открытого хроматина связан с представленностью различных клеточных типов. В связи с чем, достоверные тканеспецифичные районы связывания характерные определенной ткани дают локальные подъемы в доле мета-кластеров в РОХ. При этом данные по содержанию мотивов связывания не демонстрируют подобной</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>4. Связь пионерских факторов и РОХ. Пионерские факторы действительно способны связываться с “закрытым” хроматином в момент начала их экспрессии, что ведет к последующему ремоделингу хроматина. Но может ли это объяснить обогащение их сайтов связывания вне открытого хроматина? Кажется, такое состояние, когда ремоделлер уже связан со своим сайтом, а хроматин в этом месте ещё закрыт, будет наблюдаться в транзитном, короткоживущем состоянии клетки при переходе из одного типа в другой в процессе дифференцировки. Клеточные линии и зрелые типы клеток/тканей, представленные в Encode/GTRD, вряд ли относятся к таким переходным состояниям. Может ли тут быть другое объяснение?</p>	<p>тенденции. Таким образом, предложенный подход чувствителен к представленности клеточного типа или определенного условия для заданного ТФ. В связи с чем требуются дальнейшие его модификации.</p> <p>Действительно, связывание пионерских транскрипционных факторов с “закрытым” хроматином обычно сопровождается рекрутированием различных комплексов ремоделирования хроматина и превращения данного участка в район открытого хроматина. В работе Sherwood с соавторами 2014 года пионерные факторы делятся по их силе влияния на основании мотивов связывания и формы профилей чувствительности к ДНКазе I. На данных DNase-seq районы, ассоциированные со “слабыми” пионерскими факторами (в их число входит FOXA1) будут демонстрировать сниженный уровень сигнала DNase-seq. Таким образом, такие районы могут быть пропущены алгоритмом идентификации пиков.</p> <p>Можно предположить, что также существуют промежуточный вариант: связывание пионерского ТФ приводит к образованию “полуоткрытого” хроматина (Zaret et al., 2011). Данное состояние достаточно для рекрутирования комплекса ремоделинга хроматина и других ТФ. В статье Zaret с соавторами упоминается о роли</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		<p>FOXA1 в качестве “закладок” в хроматине для быстрого ответа клетки на определенный сигнал.</p> <p>Следует обратить внимание, что в работе для получения геномных карт открытого хроматина объединялось несколько схожих DNase-seq экспериментов (в большинстве случаев, 1-3). На графике приведены распределения общей протяженности районов открытого хроматина от числа экспериментов, участвующих в объединении.</p>  <p>Также области сниженного сигнала в DNase-seq данных, которые не детектируются алгоритмами идентификации пиков, может быть следствием клеточной гетерогенности рассматриваемых тканей.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
		<p>Zaret K. S., Carroll J. S. Pioneer transcription factors: establishing competence for gene expression //Genes &amp; development. – 2011. – Т. 25. – №. 21. – С. 2227-2241.</p> <p>Sherwood R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape //Nature biotechnology. – 2014. – Т. 32. – №. 2. – С. 171-178.</p>
	<p>5. На основании полноэкзомного анализа ассоциации было идентифицировано 135 SNP достоверно (FDR&lt;0.05) ассоциированных с морфологическими нарушениями сперматозоидов – я не нашел методическую часть, которая касалась бы поиска ассоциаций. В методах подробно описывается поиск SNV и фильтрация вариантов, но не сам анализ ассоциаций. Какова была сила эффекта для найденных ассоциаций? Часто результаты анализа представляют в виде manhattan-plot, с отметками стат. значимости - была ли выполнена такая презентация результатов?</p>	<p>Да, подобная визуализация выполнялась для всех 28 признаков. Было принято не включать данные изображения в основной текст диссертации, поскольку они носят технический характер. Однако согласен с замечанием – следовало бы включить в приложение к работе избранные построенные графики manhattan-plot и qq-plot.</p>
	<p>6. "При рассмотрении профилей выравнивания прочтений в окрестности саита связывания ТФ наблюдается бимодальность распределения выровненных прочтений (Wilbanks et Facciotti, 2010). При этом каждая мода располагается на отдельной цепи ДНК (см. Рисунок 1.2.2), а</p>	<p>В тексте литобзора следовало уточнить, что данная картина наблюдается при рассмотрении single-end данных. Данный эффект вызван процессами пробоподготовки. После очистки целевых фрагментов от ТФ происходит лигирование адаптерных и технических</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>расстояние между ними соответствует средней длине секвенируемого фрагмента ДНК." – было бы полезно чуть подробнее остановиться на этом явлении объяснить его биологический механизм.</p> <p>7. Уравнения следует нумеровать, иначе на них сложно сослаться</p>	<p>последовательностей, на 3'-концы фрагментов. Таким образом один из концов фрагмента при секвенировании будет картироваться на одну цепь ДНК, а противоположный конец - на другую. Поскольку длина фрагмента больше длины прочтения, будет наблюдаться расхождение прочтений с + цепи и с - цепи</p> <p>Для paired-end данных для определения длины секвенируемого фрагмента достаточно посчитать удаленность прочтений в паре.</p> <p>С замечанием согласен.</p>
от Антонца Дениса Викторовича, кандидата биологических наук, ведущего программиста лаборатории моделирования сложных систем ИСИ СО РАН, г. Новосибирск	1. Замечаний не поступило	–
от Пономаренко Михаила Павловича, доктора биологических наук, ведущего научного сотрудника, заведующего сектором регуляторной компьютерной геномики ИЦиГ СО РАН, г. Новосибирск	1. Замечаний не поступило	–
от Левицкого Виктора Георгиевича, кандидата биологических наук, старшего	1. Не техническим является только вопрос о термине воспроизводимость результатов картирования ChIP-seq и других данных. Не	Согласен, интересным является вопрос вклада гетерогенности (на уровне различий проведения экспериментов и клеточных типов) исследуемых

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
<p>научного сотрудника лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН, г. Новосибирск</p>	<p>понятно, какой вклад здесь вносит разнообразие конкретных биологических условий экспериментов, а какой алгоритмы обработки. Организм разделён на разные ткани, развитие происходит по стадиям, есть ещё варианты обработки самого биологического материала (например, гормонами и т.п.), я нигде в тексте не нашёл четкого разъяснения, в какой степени наблюдаемая изменчивость данных определяется самим объектом, а в какой – биоинформатической обработкой данных.</p>	<p>образов. При использовании предложенных подходов, с одной стороны, на уровне одного эксперимента происходит ранжирование РСТФ по степени их воспроизводимости и достоверности (“по мнению” каждого из алгоритмов, например, на основе приписанных РСТФ значений p-value), с другой стороны, при объединении множества экспериментов, проведенных в разных условиях в различных клеточных типах, будет наблюдаться смещение приоритизации РСТФ в сторону наиболее представленных клеточных типов. Тем не менее РСТФ с наибольшим весом будут являться достоверными районами связывания. Однако это является слабым местом предложенного подхода. Одним из вариантов решения данной проблемы может быть использование весовых коэффициентов, учитывающих представленность того или иного условия. Данный вопрос требует дальнейшего исследования</p>
	<p>2. Стр. 3 Очень резкая смена темы повествования сразу на первой странице текста, во самом первом разделе, Актуальность темы исследования: ... Поэтому актуальной является задача разработки алгоритма определения наиболее достоверных РСТФ на основе мета-анализа сходных ChIP-seq экспериментов для заданного ТФ.</p>	<p>С замечаниями согласен.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>В последние несколько десятилетий в различных регионах мира наблюдается снижение мужского репродуктивного потенциала...</p> <p>почему-то до читателя не доводится мысль о том, что основное достоинство GTRD состоит не в том, что она крупнейшая по числу экспериментов или программ обработки сырых данных, peak callers, а в том, что она ориентирована на получение биологических результатов на основе массового анализа собранных в базе данных, именно это должно быть написано при переходе от общего введения (NGS, ChIP-seq, GTRD) к вопросам массового анализа данных, в частности, от мета-анализа ChIP-seq к теме анализа однонуклеотидных геномных вариантов (SNV) и исследования мужского репродуктивного потенциала, нужно было бы осветить именно как удобство и адаптацию GTRD для получения конкретных биологических результатов.</p> <p>стр. 4 тут сразу ошибки и в логике, и в русском языке ...Также до конца нерешённым остается вопрос об интеграции имеющихся данных для ...-&gt; Также не решённым до конца я бы склонился ближе к отдельному написанию, <a href="https://gramatik.ru/ne-reshyonnyj-ili-nereshyonnyj-slitno-ili-razdelno/">https://gramatik.ru/ne-reshyonnyj-ili-nereshyonnyj-slitno-ili-razdelno/</a></p>	

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>Это следует из того, что вопрос как раз решался, но до конца не был решён, поэтому его нельзя считать нерешённым.</p> <p>Автор иногда игнорирует букву ё. нерешённым, проведён</p> <p>рис. 4 перед списком задач не написано, что это список задач.</p> <p>вводится термин однонуклеотидных геномных вариантов (SNV), а затем на стр. 5 ...В рамках диссертационного исследования были впервые идентифицированы однонуклеотидные геномные вариации... нужно быть строже в терминологии</p>	
	<p>3. стр. 6 НОСОМОСО (<a href="https://nosomoco11.autosome.ru/">https://nosomoco11.autosome.ru/</a>) чуть менее года назад вышла версия 12 <a href="https://nosomoco12.autosome.ru/">https://nosomoco12.autosome.ru/</a>) с гораздо большим числом мотивов сайтов связывания транскрипционных факторов - в 2-3 раза, чем было в версии 11, в версии не использованы данные GTRD? Насколько мне известно переход к версии 12 существенно улучшил как количество, так и качество данных, между двумя выпусками прошло 6 лет. Среди авторов версии Носомосо 12</p>	<p>В данном случае это ошибка описания использованных в работе версий баз данных. В рамках диссертационной работы использовалась БД НОСОМОСО v12.</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>соискателя найти можно (NAR 2024), а вот версии Hocomoco 11 его нет (NAR 2018):  Ivan V Kulakovskiy, Ilya E Vorontsov, Ivan S Yevshin, Ruslan N Sharipov, Alla D Fedorova, Eugene I Rumynskiy, Yulia A Medvedeva, Arturo Magana-Mora, Vladimir B Bajic, Dmitry A Papatsenko, Fedor A Kolpakov, Vsevolod J Makeev, HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale CHIP-Seq analysis, Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D252–D259, <a href="https://doi.org/10.1093/nar/gkx1106">https://doi.org/10.1093/nar/gkx1106</a>  Ilya E Vorontsov, Irina A Eliseeva, Arsenii Zinkevich, Mikhail Nikonov, Sergey Abramov, Alexandr Boytsov, Vasily Kamenets, Alexandra Kasianova, Semyon Kolmykov, Ivan S Yevshin, Alexander Favorov, Yulia A Medvedeva, Arttu Jolma, Fedor Kolpakov, Vsevolod J Makeev, Ivan V Kulakovskiy, HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors, Nucleic Acids Research, Volume 52, Issue D1, 5 January 2024, Pages D154–D163, <a href="https://doi.org/10.1093/nar/gkad1077">https://doi.org/10.1093/nar/gkad1077</a>  Возникает вопрос о том, насколько велик вклад автора в версии Hocomoco 11 и 12, лучше избегать таких вопросов, так как не все знают, что Hocomoco и GTRD плотно интегрированы, а аспирантура автора была закончена в 2020 г., и до этого он несколько лет уже работал с базой данных</p>	

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>GTRD, в частности по конвейеру обработки данных для базы данных Носомосо.</p> <p>4. стр. 7 ...Личный вклад автора База данных GTRD - результат работы большого количества аннотаторов и биоинформатиков. В ходе диссертационной работы автором лично проаннотировано 1701 DNase-seq и 1347 ChIP-seq экспериментов для человека... Необходимо указать версию GTRD, она регулярно обновляется.</p> <p>5. стр. 8 опечатка в обзоре литературе</p> <p>стр. 10 Поиск мотивов связывания ТФ (МСТФ) из БД НОСОМОСО в РСТФ не указан инструмент</p> <p>стр. 11 Автореферат заслуживает своей нумерации рисунков Рисунок 3.1.1 и т.д. - он первый в автореферате</p> <p>стр. 13 PWM - не определён термин</p> <p>стр. 14 AUC - не определён термин</p>	<p>С замечаниями согласен. В работе использовалась GTRD v21.12</p>
	6. рис. 18	“рис. 18

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	<p>...Анализ воспроизводимости РСТФ разными алгоритмами идентификации РСТФ в рамках одного ChIP-seq эксперимента показал, что в среднем полностью воспроизводится ~10% от общего числа РСТФ</p> <p>Термин воспроизводимость тут спорен, так как неконсервативность РСТФ в разных экспериментах ChIP-seq не означает, что где-то есть ошибка, я подозреваю, что данные GTRD как раз и могут помочь лучше понимать специфичность работы ТФ в разных типах клеток и на разных стадиях развития.</p> <p>Стремление получить идеальные мета-кластеры приведёт, к отмечено в автореферате, к генам домашнего хозяйства и убьёт все эффекты тканеспецифичности. Также надо автору пояснить - речь в заключении идёт о наложении результатов peak callers разных, или о наложении ChIP-seq экспериментов разных. В целом, данные агрегации данных отражают частоты экспериментов ChIP-seq, проведённых в разных тканях, стадиях.</p> <p>Некоторые ТФ узкоспецифичны по тканям (AR), другие не очень (Foxa1). Это поле для артефактных выводов из анализа данных. Этот вопрос не совсем понятен.</p> <p>7. стр. 21</p> <p>...которые расположены в наиболее воспроизводимых геномных районах связывания трёх транскрипционных факторов: AR, CTCF и</p>	<p>...Анализ воспроизводимости РСТФ разными алгоритмами идентификации РСТФ в рамках одного ChIP-seq эксперимента показал, что в среднем полностью воспроизводится ~10% от общего числа РСТФ” - полагаю, это недостаток описания рисунка в автореферате; речь идет о воспроизводимости в рамках одного ChIP-seq эксперимента. Поэтому использование термина “воспроизводимость” в данном контексте вполне оправдано.</p> <p>“В целом, данные агрегации данных отражают частоты экспериментов ChIP-seq, проведённых в разных тканях, стадиях” - тем не менее, на этапе сопоставления разных экспериментов будут исключаться ошибочно идентифицированные РСТФ. Количество артефактных выводов будет зависеть также от выбранного порогового значения для последующей фильтрации. Одним из возможных путей развития метода может быть рассмотрение полученных значений ФАФ на уровне схожих экспериментов с последующим отбором достоверных РСТФ.</p> <p>Действительно, тот факт, что в топ 100.000 РСТФ AR вошли районы связывания, ассоциированные с процессами, протекающими в органах мужской репродуктивной системы, может быть связано с преобладанием ChIP-seq экспериментов для этих</p>

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	SRBP2, участвующих в регуляции сперматогенеза, и влияют на эффективность их связывания с ДНК Это в какой степени отражает, что очень много данных ChIP-seq для AR по ткани простаты, например?	тканей. Однако это не снижает степени достоверности выявленных РСТФ и их ценности для последующей интерпретации данных.
от Лашина Сергея Александровича, кандидата биологических наук, ведущего научного сотрудника, заведующего секторами биоинформатики и информационных технологий в генетике, компьютерного анализа и моделирования биологических систем ИЦиГ СО РАН, г. Новосибирск; от Казанцева Федора Владимировича, кандидата биологических наук, старшего научного сотрудника сектора компьютерного анализа и моделирования биологических систем ИЦиГ СО РАН, г. Новосибирск	1. О том, что работа посвящена обработке ChIP-seq данных именно человека мы узнаем только в пятом абзаце. К сожалению, и в конце отсутствует обсуждение о применимости разработанного подхода для других организмов.	С замечанием согласен.
	2. Небрежное использование сокращений по тексту. Их много, они на двух языках (русском и английском), часто встречаются вперемешку. Например, «Оценка доли ложноположительных (FP) РСТФ при помощи FPCM (False Positive Control Metric)», «расположением предсказанных МСТФ в РСТФ», «анализ взаимосвязи между значениями FPCM и изменением эффективности идентификации МСТФ в полном наборе РСТФ в ответ на удаление F1 РСТФ». На странице 16 автореферата первый абзац является ярким примером смешения терминов, сокращений и английских названий, что затрудняет чтение работы.	С замечанием согласен.
	3. В тексте приводятся короткие названия транскрипционных факторов, которые теряются на фоне других сокращений. В ряде случаев только из контекста понятно, что имелось ввиду. Например, «сниженное количество F1 РСТФ в РОХ».	С замечанием согласен.

Организация, Фамилия И.О.	Вопросы/Замечания	Ответы/Комментарии
1	2	3
	4. На рисунке 3.1.1. присутствует схема разбиения РСТФ на 4-е группы F1-F4, но не расшифрован принцип деления на эти группы.	Для того, чтобы отнести РСТФ к конкретной группе выполнялась операция пересечения границ всех РСТФ, наденных четырьмя алгоритмами идентификации пиков. На основании числа пиков (кол-ва алгоритмов), попавших в это пересечение, делался вывод о принадлежности к РСТФ к определенной группе.
	5. Особняком выглядит задача 4 и соответствующей ей последний абзац в разделе «Актуальность». Хотелось бы видеть более логичную связку цели и всех задач работы.	С замечанием согласен.
от Цуканова Антона Витальевича, кандидата биологических наук, младшего научного сотрудника лаборатории эволюционной биоинформатики и теоретической генетики ИЦиГ СО РАН, г. Новосибирск	1. Из технических замечаний можно отметить, что в автореферате присутствуют опечатки.	С замечанием согласен.
	2. Так же есть дискуссионный вопрос: «Насколько корректно верифицировать РСТФ на данных открытого хроматина (ATAC-seq, DNase-seq), если эти данные обрабатываются с использованием таких же инструментов, что и ChIP-seq (например MACS2)?»	Используемый в работе для идентификации пиков в DNase-seq экспериментах алгоритм MACS2 является самым популярным алгоритмом для решения подобных задач. В дальнейшей работе планируется использовать информацию о воспроизводимости районов открытого хроматина для отбора достоверных районов открытого хроматина.

Колмыков Семён Константинович

26 октября 2024 г.