

ОТЗЫВ

члена диссертационного совета НТУ.1.5.8.01

Миронова Андрея Александровича

на диссертацию **Колмыкова Семёна Константиновича**

«Разработка методов контроля качества и построения карты геномных районов связывания транскрипционных факторов на основе сравнительного анализа ChIP-seq экспериментов»,
представленной на соискание ученой степени кандидата биологических наук по специальности 1.5.8. Математическая биология, биоинформатика

Актуальность темы. Регуляция экспрессии генов является важнейшим механизмом функционирования генома. Регуляция экспрессии происходит на разных уровнях – от структуры хроматина до стабильности РНК и белков. В настоящее время считается, что самым важной является регуляция экспрессии на уровне экспрессии РНК, которая осуществляется транскрипционными факторами, которые связываются со специфическими сайтами связывания и включают или выключают синтез РНК. Для поиска сайтов связывания в клетке используют специальные экспериментальные протоколы, в частности, ChIP-seq. В настоящее время накоплен огромный объем данных ChIP-seq. Однако, как и любой метод массового производства данных, этот метод производит достаточно большой уровень шума, как экспериментального, так и биологического. Для выделения сигнала из шума разработан ряд методов поиска сигнала (поиска пиков). Однако эти методы дают разные результаты. Более того независимые эксперименты, проведенные на тех же клетках и в тех же условиях зачастую дают противоречивые результаты. Представленная работа как раз посвящена анализу качества данных и анализу качества методов их анализа. Поэтому актуальность выбранной темы исследования не вызывает сомнения

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации. Работа, представленная в диссертации, выполнена на высоком научном уровне. Все полученные результаты научно обоснованы.

Научная новизна работы определяется тем, что в работе впервые создана уникальная коллекция из 15982 единообразно обработанных ChIP-seq экспериментов, разработаны новые программные модули для оценки качества данных, разработаны новые методы мета-анализа данных и разработаны соответствующие программные модули, На основе анализа данных полноэкзомного секвенирования впервые идентифицированы ассоциации 135 однонуклеотидных геномных вариантов с различными нарушениями морфологии сперматозоидов человека.

Теоретическая и практическая значимость работы. В работе разработаны новые подходы к анализу данных ChIP-seq экспериментов, что имеет большую теоретическую значимость. Разработанные подходы реализованы в виде программных модулей. Практическая значимость работы определяется тем, что разработанные автором подходы реализованы в виде программных модулей для биоинформатической платформы BioUML. В качестве демонстрации практической значимости диссертации в работе представлен пример применения методов к анализу реальных клинических данных – анализу однонуклеотидных геномных вариантов с различными нарушениями морфологии сперматозоидов человека и была обнаружена ассоциация 135 однонуклеотидных геномных вариантов с качеством сперматозоидов.

Степень достоверности результатов проведенных исследований. В работе достаточно полно и подробно обоснованы полученные результаты и их достоверность не вызывает сомнения.

Публикации основных результатов диссертационной работы.

По теме диссертации опубликовано 9 работ, отражающих основные положения исследования, в том числе 8 в журналах первого квартиля и 1 в журнале второго квартиля. Все журналы индексируются в международных базах Scopus и WoS и рекомендованы ВАК Минобрнауки России по искомой

специальности, причем в 3 работах С.К.Колмыков является первым автором. Опубликовано 4 статьи, индексируемые в базе Scopus. Кроме того, опубликовано 11 тезисов докладов на конференциях.

Структура диссертационной работы. Диссертационная работа изложена на 141 странице и содержит введение, где обосновываются цели и задачи исследования. В обзоре литературы приведен достаточно полный обзор существующего состояния исследований в этой области. Далее достаточно подробно изложены использованные материалы и методы исследования.

Основной раздел – результаты и обсуждения состоит из пяти подразделов, посвященных описанию задач исследования. Первый подраздел посвящен анализу воспроизводимости различных методов поиска пиков в данных ChIP-seq. Второй и третий подразделы посвящены анализу ложноположительных и ложноотрацательных результатов в программах поиска пиков. В четвертом разделе описан метод метанализа данных METARA, который позволяет объединять разнородные данные и повышать точность предсказания пиков в данных ChIP-seq экспериментов. Наконец в пятом разделе приведен пример применения разработанных методов к анализу реальных клинических данных. Отмечу, что каждый раздел завершается кратким резюме, что облегчает чтение.

Текст завершается заключением, где дано краткое описание основных результатов, а также даны рекомендации и перспективы дальнейшей разработки темы. Далее, традиционно идут выводы, список публикаций по теме работы и список литературы, насчитывающий 159 источников.

Содержание автореферата соответствует содержанию, основным положениям и результатам диссертации.

Вопросы по диссертационной работе/ Замечания

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.

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.

3. Стр. 46 «размер свободных от хроматина участков ДНК» – неудачное выражение, поскольку ДНК является непосредственным компонентом хроматина.

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.

5. Поскольку обнаружение сигнала программами поиска пиков опирается на достаточно произвольно выбранные порог, то возникает 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.

Отмеченные недостатки не снижают высокого качества исследования и не влияют на главные теоретические и практические результаты диссертации, описанные выше. Результаты оригинальны, обладают научной новизной и практически значимы.

Заключение

Диссертационная работа Колмыкова Семёна Константиновича является законченной научно-квалификационной работой, выполненной автором на

высоком научном уровне. Диссертация соответствует пп.2,5,11,12 паспорта научной специальности 1.5.8. Математическая биология, биоинформатика.

Диссертационная работа Колмыкова Семёна Константиновича «Разработка методов контроля качества и построения карты геномных районов связывания транскрипционных факторов на основе сравнительного анализа ChIP-seq экспериментов» отвечает требованиям пп.2.1–2.6 Положения о присуждении ученых степеней Автономной некоммерческой образовательной организацией высшего образования «Научно-технологический университет «Сириус» утвержденного приказом от 25 декабря 2023 г. № 350/1-ОД-У, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор, Колмыков С.К., заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8. Математическая биология, биоинформатика.

Член диссертационного совета
НТУ.1.5.8.01

профессор Факультета биоинженерии
и биоинформатика МГУ им
М.В.Ломоносова
д.б.н., профессор



Андрей Александрович
Миронов

Сведения:

ПОДПИСЬ
УДОСТОВЕРЯЮ
ЗАВ КАНЦЕЛЯРИЕЙ
Н Н СИДОРОВА

15.10.2024

Миронов Андрей Александрович, доктор биологических наук (03.00.03 – Молекулярная биология), профессор

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова», Факультет биоинженерии и биоинформатики.

Адрес организации: 119991, Российская Федерация, Москва, Ленинские горы, д. 1, Московский государственный университет имени М.В. Ломоносова, стр.73, Факультет биоинженерии и биоинформатики.

Телефон: +7 910 462 9065

e-mail: mironov@bioinf.fbb.msu.ru