

## ОТЗЫВ

члена диссертационного совета НТУ.1.5.8.01  
Фишмана Вениамина Семеновича  
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 транскрипции является фундаментальным процессом, определяющим экспрессию генов и, следовательно, функционирование клеток и организмов в целом. Транскрипционные факторы (ТФ) играют ключевую роль в этом процессе, связываясь с определенными участками ДНК и регулируя активность генов. Понимание механизмов взаимодействия ТФ с геномом имеет решающее значение для раскрытия основ развития, дифференцировки клеток и возникновения различных заболеваний.

Основным методом для экспериментального определения районов связывания транскрипционных факторов (РСТФ) является ChIP-seq (Chromatin Immunoprecipitation followed by sequencing). Однако, несмотря на широкое применение, метод ChIP-seq характеризуется высоким уровнем шума и вариабельностью результатов. Различные алгоритмы идентификации пиков, такие как MACS2, GEM, SISR, PICS и другие, могут давать существенно отличающиеся результаты при анализе одних и тех же данных. Отсутствие "золотого стандарта" для валидации правильности определения РСТФ создает серьезные препятствия для сравнения и интерпретации результатов различных исследований.

В связи с этим, создание новых методов анализа и интеграции инструментов обработки ChIP-seq-данных безусловно является важной задачей.

## **Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации.**

Выводы, положения и рекомендации полностью соответствуют полученным результатам. Форма представления материала позволяет убедиться в обоснованности научных положений, выводов и рекомендаций.

## **Научная новизна работы.**

Научная новизна работы определяется разработкой уникальных, авторских методов анализа данных, а также использованием большой коллекции экспериментов, собранных в базе GTRD. Ещё одним аспектом уникальности является использование уникальных геномных данных, полученных от мужчин с нарушением сперматогенеза.

## **Теоретическая и практическая значимость работы.**

Учитывая широкое распространение метода ChIP-seq практическая значимость работы очевидна. Важным плюсом является наличие интерфейса для доступа к аналитическим инструментам, разработанным автором. в форме веб-сервиса, что делает его доступным для широкого круга исследователей.

Что касается теоретической значимости, автор затрагивает важные вопросы как минимум в двух областях. Во-первых, при попытке определить *bona fide* ложно-положительные и ложно-отрицательные результаты ChIP-seq-эксперимента, возникает вопрос о том, можно ли вообще решить эту задачу используя лишь статистические методы обработки данных, без привлечения ортогональных подходов. Во-вторых, при анализе взаимосвязей между расположением участков связывания ТФ и других характеристик хроматина, автор обнаруживает уникальные свойства отдельных ТФ, например, обогащение сайтов связывания белков пионерских факторов в регионах, относительно редко пересекающихся хроматином, аннотированным как “открытый”. Если этот феномен не связан с техническими артефактами, он

может является значительным шагом к объяснению механизмов функционирования и роли этих ТФ.

### **Степень достоверности результатов проведенных исследований.**

Достоверность результатов не вызывает сомнений. Большая часть аналитических инструментов реализована в форме утилит на платформе BioUML и их работа может быть легко воспроизведена на этой же платформе. При обработке данных выполнен тщательный статистический анализ. Следует также отметить высокий публикационный уровень. Изложенный в профильных журналах материал прошел проверку специалистов, что также подтверждает его достоверность.

### **Публикации основных результатов диссертационной работы.**

Материалы диссертационной работы отражены в 25 научных публикациях, включая: 13 публикаций в журналах, индексируемых в международных базах данных Web of Science/Scopus, из которых 8 публикаций Q1.

### **Структура диссертационной работы.**

Диссертация построена по «классическому» типу и состоит из глав введение, обзор литературы, материалы и методы, результаты и обсуждение, выводы. Результаты и обсуждения разделены на пять подглав, соответствующих отдельным частям исследования. Диссертация изложена понятным языком, текст хорошо структурирован и иллюстрирован. Отдельно хочется отметить наличие заключений к каждой подглаве результатов — такой формат позволяет сформировать представление о наиболее важных элементах исследования по ходу прочтения и добавляет логику изложению.

Содержание автореферата соответствует содержанию, основным положениям и результатам диссертации.

### **Вопросы и замечания по диссертационной работе/ Замечания**

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 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 — т.к. инструменты работают на одном и том же наборе данных, их результаты нельзя считать независимыми.
2. Связанный предыдущим вопрос: Можно ли вычислить  $f_{le}$ , используя значение  $\lambda$ , полученное из  $p_2$  и  $p_3$  по отдельности? Повлияет ли это на результат?
3. С чем связан волнообразный профиль кривой для JUN на рис. 3.4.3?
4. Связь пионерских факторов и POX. 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 транзиторном, 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?
5. На основании полноэкзомного анализа ассоциации было идентифицировано 135 SNP достоверно ( $FDR < 0.05$ ) 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?

Небольшие замечания оформительского характера:

6. *“При рассмотрении профилей выравнивания прочтений в окрестности сайта 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 расстояние между ними соответствует средней длине секвенируемого фрагмента ДНК.”* - было бы полезно чуть подробнее остановиться на этом явлении объяснить его биологический механизм.

7. Уравнения следует нумеровать, иначе на них сложно ссылаться

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 и 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.

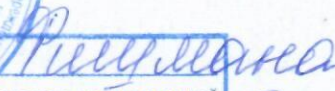
Член диссертационного совета  
НТУ.1.5.8.01  
к.б.н., внс, зав. сектором геномных  
механизмов ИЦиГ СО РАН

Фишман  
Вениамин  
Семенович



  
\_\_\_\_\_

18 октября 2024

Подпись   
удостоверяю зав. канцелярией

**ИЦиГ СО РАН**

« 18 » 10 2024

Сведения:

Федеральное государственное бюджетное научное учреждение  
«Федеральный исследовательский центр Институт цитологии и генетики  
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

Адрес:

**Адрес организации:** Российская Федерация, 630090, Новосибирск, Россия,  
пр.ак.Лаврентьева,10

Телефон: +79529008413

e-mail: minja-f@ya.ru