

ОТЗЫВ

на автореферат диссертационной работы
Пензара Дмитрия Дмитриевича
на тему: «Вычислительное предсказание эффектов мутаций в
регуляторных районах генов», представленной на соискание ученой
степени кандидата биологических наук,
по специальности 1.5.8 «Математическая биология, биоинформатика»

Диссертационная работа Пензара Дмитрия Дмитриевича посвящена очень актуальной теме предсказания генной экспрессии на основе нуклеотидной последовательности соответствующих регуляторных районов при помощи нейронных сетей.

Достоверность научных выводов и положений выносимых на защиту, подтверждается публикацией результатов исследований 6 статей в высокорейтинговых научных журналах, включая Nature, Nature Biology и Nature Communications.

Хочется отметить очень детальный и критически осмысленный обзор литературы, который представляет отдельную ценность и включает в себя 344 литературные ссылки.

К несомненным достижениям автора относится разработка лучшей в классе нейронной сети LegNet, которая заняла 1-ое место в международном конкурсе DREAM-2022 по предсказанию экспрессии гена дрожжей на основе данных MPRA эксперимента. Также выдающиеся результаты показывают разработанные на основе LegNet нейронные сети MPRALegNet и Legformer.

Тем не менее, есть ряд вопросов и замечаний к работе:

1. Для восприятия было бы легче, если бы диссертация была написана по монографическому типу, т.е. в каждой главе описан свой метод вместе со своими результатами.

2. Раздел "2.2. Примеры клинически значимых регуляторных вариантов" - не понятно, по какому принципу были отобраны примеры.

3. Раздел "2.3.1. QTL" - полезно было бы упомянуть GTEx и методы тонкого картирования казуальных вариантов SNP.

4. стр. 111:

Мы изменили формулировку задачи для машинного обучения из регрессионной в классификационную (т.н. мягкая классификация), и для этого преобразовали целевые оценки экспрессии в вероятности классов, используя приведенную ниже схему.

До этого автор в обзоре литературы показывал, что точность выше при решении задачи регрессии, чем классификации.

5. Раздел "4.3.5. Пост-конкурсная оптимизация модели LegNet" - очень интересное исследование компонентов архитектуры LegNet, но результаты представлены очень кратко, нет никаких численных оценок или графиков.

6. В разделе "4.5.5 Анализ регуляторной грамматики, выученной моделью" кажется достаточно самоочевидным, что модель сможет выучить зависимость активности региона от числа сайтов транскрипционных сайтов, так ли это? Более интересна была бы валидация предсказаний модели при помощи экспериментального нокаута.

7. Раздел "4.4. Генерация промоторных последовательностей с заданной активностью" - в обзоре автор показал, что метод генетического алгоритма не уступает алгоритмам диффузии, но в диссертации использовал метод диффузии. Почему?

8. Было бы интересно прогнать MPRA LegNet по всему геному человека и сравнить результаты с CAGE профилем.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.5.8 «Математическая биология, биоинформатика» (по биологическим наукам), а именно следующим ее направлениям «Компьютерная системная биология», «Разработка и применение новых вычислительных алгоритмов для анализа экспериментальных данных в биологии и медицине», «Разработка и применение методов машинного обучения и искусственного интеллекта для анализа и прогнозирования свойств биологических объектов на основе анализа больших биомедицинских данных», а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В.Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Пензар Дмитрий Дмитриевич заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8 «Математическая биология, биоинформатика».

доктор биологических наук,

научный руководитель направления "Вычислительная биология",

Научного центра генетики и наук о жизни,

АНОО ВО "Университет "Сириус"

Колпаков Федор Анатольевич



27.02.2025

Контактные данные:

тел.: +7(913)9431649, e-mail: kolpakov.fa@talantiuspeh.ru

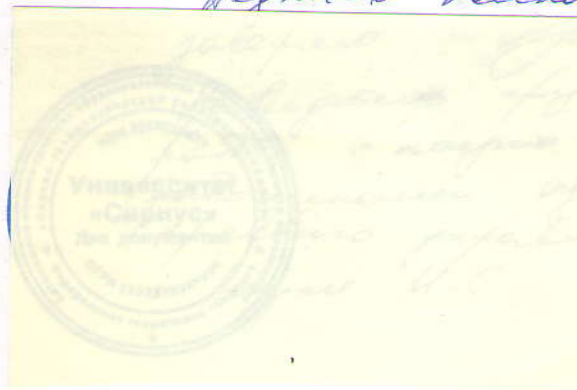
Специальность, по которой защищена диссертация:

1.5.8 «Математическая биология, биоинформатика»

Адрес места работы:

354340, Российская Федерация, Краснодарский край, федеральная
территория «Сириус», Олимпийский пр., д.1

Тел.: 8 (800) 100 41 55; e-mail: info@siriusuniversity.ru



Сергей Александрович Колпаков

Функционал по

научно-педагогической деятельности

по специальности

28.02.2025