

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»**

Работа Пензара Д.Д. направлена на изучение регуляторных последовательностей генома, в частности на поиск связи между нуклеотидным контекстом регуляторных районов и уровнем экспрессии генов. Уже давно известен генетический код, при помощи которого закодированы белки в геноме, однако до сих пор нет универсального ответа на то, как влияет нуклеотидный контекст регуляторных районов на экспрессию связанного с ним гена. Более того, основная доля мутаций приходится именно на регуляторные участки генома, из-за чего актуальным является поиск ответа на вопрос: «Каким образом мутация повлияет на уровень транскрипции?». В связи с этим актуальность работы Пензара Д.Д. «Вычислительное предсказание эффектов мутаций в регуляторных районах генов» не вызывает никаких сомнений.

Работу Пензара Д.Д. можно разделить на два блока. Первый блок посвящен проблеме обучения моделей, в частности проблеме «утечке данных», которая может приводить к завышенной оценке точности моделей. Второй блок посвящён разработке нейросетевой модели и её модификаций для предсказания уровня транскрипции генов по нуклеотидной последовательности регуляторной области, а также для генерации нуклеотидной регуляторной последовательности для заданного уровня транскрипции.

Автор разработал модель LegNet, архитектура которой основана на самой эффективной для работы с изображениями модели EfficientNetV2 на момент 2022 года, и при этом оптимизировал эту архитектуру для работы с нуклеотидными последовательностями. Считаю важным отметить, что модель

Пензара Д.Д. превзошла все остальные модели в соревновании DREME-2022 по предсказанию уровня транскрипции генов *Saccharomyces cerevisiae* по нуклеотидным последовательностям регуляторных участков. Более того модель Пензара Д.Д. лучше остальных пересказывала эффект однонуклеотидных мутаций. На следующем этапе автор модифицировал архитектуру LegNet и использовал подход «холодной диффузии», чтобы генерировать нуклеотидные последовательности для регуляторных районов *Saccharomyces cerevisiae*, которые будут иметь заданный уровень транскрипции, что является важным для генной инженерии.

Архитектура LegNet была оптимизирована для работы с данными, полученными для *Homo sapiens*. Для регуляторных районов человека, Пензар Д.Д. показал, что его модель может выделять сложные контекстные закономерности, которые оказывают влияние на уровень активности регуляторного участка. Среди таких закономерностей были, количество сайтов связывания одного фактора транскрипции, а также комбинации сайтов связывания разных ТФ. Данный результат очень важен, так как регуляция транскрипции в основном выполняется не одним белком, а комплексом, куда может входить два или более фактора транскрипции. Так же автор показал, что внесение дополнительной информации, как например эпигенетических сигналов, может улучшать возможности моделей по предсказанию влияния мутаций на уровень активности регуляторного района.

Результаты, представленные в автореферате, свидетельствуют о высоком методическом и научном уровне проведенной работы, что также подтверждают публикации в авторитетных изданиях и доклады на профильных конференциях. Разработанная автором модель позволяет не только предсказывать уровень активности регуляторных районов, но и создавать регуляторные последовательности с заданной активностью. Более того данная модель на текущий момент времени превосходит все остальные модели.

К несущественным замечанием можно отнести наличие небольшого количества опечаток и неточностей. Так же было бы интересно оценить эффективность сгенерированных с помощью LegNet регуляторных последовательностей. Особенно было бы интересно получить и проверить регуляторный участок, который бы приводил к сверхэкспрессии гена, однако это отдельная экспериментальная работа.

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».

к.б.н., м.н.с. лаборатории эволюционной биоинформатики и теоретической генетики Федерального государственного бюджетного научного учреждения

«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

Цуканов Антон Витальевич

Контактные данные:

тел.: +7 (993) 019-31-90, e-mail: tsukanov@bionet.nsc.ru

Специальность, по которой защищена диссертация:

1.5.8 – Математическая биология, биоинформатика

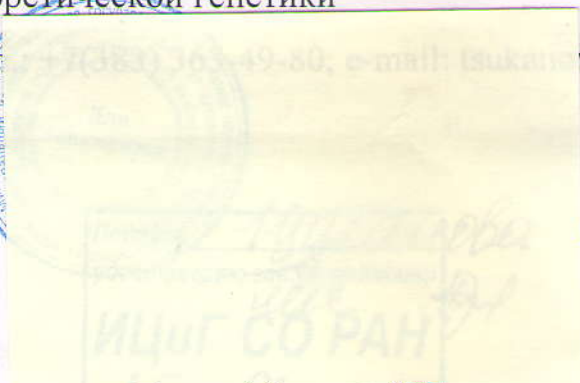
Адрес места работы:

630090, Новосибирская область г. Новосибирск, ул. Лаврентьева, д. 10

ИЦиГ СО РАН, лаборатория эволюционной биоинформатики и

теоретической генетики

Тел. +7 (993) 019-31-90, e-mail: tsukanov@bionet.nsc.ru



«ас» 1/4 2022г.