

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

к.ф.-м.н., д.б.н. Кулаковского Ивана Владимировича

на диссертационную работу

Дамарова Игоря Сергеевича

«Полногеномное выявление регуляторных SNPs человека на основе получения и анализа комплекса данных ChIP-seq и RNA-seq»,
представленную к защите на соискание ученой степени кандидата
биологических наук по специальности
«1.5.7. – Генетика (биологические науки)»

Актуальность, научная новизна, теоретическая и практическая значимость исследования. Изучение связи между индивидуальными геномными вариантами и фенотипом продолжает оставаться одной из центральных тем современной генетики. Особый интерес вызывают варианты, ассоциированные с предрасположенностью к развитию или тяжелому течению социально-значимых заболеваний, например, диабету 2го типа. В то время как варианты, непосредственно влияющие на последовательность и структуру белков, достаточно хорошо поддаются изучению и аннотации, по ним накоплен значительный объем информации. В то же время регуляторные варианты, расположенные в некодирующих областях, остаются исследованы лишь фрагментарно. Присутствие и функциональная роль таких вариантов в менее изученных популяциях, включая российскую, исследованы еще хуже. В представленной работе на клиническом материале построена панель из тысяч кандидатных регуляторных вариантов, проявляющих аллель-специфичную активность в моноклеарных клетках периферической крови российских доноров. Совокупно, это определяет **актуальность и новизну работы**. Результаты работы являются важным источником информации об однонуклеотидных заменах, локализованных в регуляторных районах, и будут безусловно полезны как при интерпретации результатов полногеномных исследований ассоциаций, так и при изучении конкретных вариантов для функциональной аннотации индивидуальных геномов в задачах медицинской генетики, что определяет **теоретическую и практическую значимость работы**.

Общая характеристика работы. Представленная диссертационная работа Дамарова И.С. построена по традиционной схеме: Введение; Обзор литературы;

Материалы и методы; Результаты; Обсуждение; Заключение и Выводы, Список литературы. Список сокращений дан на старте документа. Список цитированной литературы включает 362 источника. Работа изложена на 139 страницах печатного текста и адекватно проиллюстрирована (18 рисунков, 4 таблицы). В целом, работа хорошо изложена стилистически, подписи на основных рисунках даны на русском языке, число опечаток невелико.

Во «Введении» автор хорошо излагает актуальность исследования, его новизну и значимость, формулирует цель и задачи, приводятся сведения об апробации работы, публикациях, личном вкладе автора. В главе «Обзор литературы» удачно сгруппированы материалы по четырем связанным разделам, имеющим прямое отношение к теме работы: это классификация однонуклеотидных вариантов и их связи с патологиями человека, особая роль регуляторных вариантов, аллель-специфичный анализ омиксных данных, и, наконец, механизмы развития сахарного диабета 2го типа и особенности клеточного ответа на метформин. Обзор литературы охватывает множество публикаций, достаточно информативен и хорошо читается. В «Материалах и методах» чрезвычайно детально описан дизайн экспериментов, в подробностях приводятся использованные молекулярно-биологические методы, отдельный подраздел посвящен биоинформатическому протоколу анализа. В главе «Результаты» представлены результаты исследования от поиска регуляторных вариантов по омиксным данным ChIP-Seq и RNA-Seq до формирования итоговой панели и ее аннотации с помощью внешних ресурсов, включая привлечение данных по дифференциальной экспрессии генов. Главы «Обсуждение», «Заключение» и «Выводы» в краткой форме, но достаточно емко резюмируют успешные итоги результатов исследования соискателя.

Автореферат имеет стандартную структуру, содержит всю требуемую информацию и хорошо отражает содержание диссертации.

Обоснованность и достоверность научных положений, выводов и рекомендаций. Достоверность и обоснованность научных положений и выводов, сформулированных в работе, обусловлена адекватным подходом к экспериментальным протоколам, тщательному статистическому анализу данных (включая необходимые поправки на множественное тестирование), и проведенным сравнительным анализом с существующими литературными (публикации) и информационными (базы данных) источниками.

Результаты работы успешно апробированы на международной конференции, по результатам опубликовано 4 статьи, включая *IJMS* (журнал Q1), в одной из ключевых статей автор работы выступает в роли первого автора.

Замечания и вопросы по диссертационной работе. Представленная работа производит зрелое и позитивное впечатление. Однако, она не лишена и технических недостатков, на которые хочется обратить внимание автора.

1. Сокращения и англицизмы: Список сокращений анекдотично включает в себя ДНК и РНК, при этом в работе постоянно используются англоязычные акронимы rSNP и RBMC, которые следовало заменить русскоязычными аналогами, взяв за основу «регуляторный однонуклеотидный вариант» (ОНВ и рОНВ) и «моноклеарные клетки периферической крови» (МКПК). Другие (пусть и не все) англоязычные термины и акронимы также следовало «русифицировать».
2. Стр. 6 «более трехсот тысяч вариаций»: далее в работе упоминается в два раза большее число. Можно ожидать что на текущий момент число выросло еще более существенно в силу постоянной публикации результатов все новых полногеномных исследований ассоциаций.
3. Стр. 9 «развития различных заболеваний» было бы корректнее сформулировать спектр заболеваний, для которых релевантны полученные данные.
4. Список цитируемой литературы является избыточным для кандидатской диссертации, 362 источника это чрезмерно много.
5. Предложенная классификация вариантов «по локализации» является не вполне удачной, поскольку регуляторные варианты могут локализовываться в различных функциональных областях гена. Содержательно это отражено в тексте, но противоречит предложенной классификации (см. например стр. 16). Было бы уместнее использовать классификацию «по функции».
6. Стр. 23: здесь противопоставляется eQTL и GWAS анализ по требуемому размеру выборок. Это не вполне корректно, современный анализ eQTL по сути является GWAS с множеством количественных признаков типа «экспрессия гена».
7. Жаль, что в работе не использованы (хотя упоминается их важность на стр. 28) данные по аллель-специфичной доступности хроматина из базы данных UDACHA (<https://udacha.autosome.org>).
8. Материалы и методы содержат дословный перевод инструкций к коммерческим наборам реактивов. Достаточно было бы указать «анализ проводился в соответствии с инструкциями производителя».
9. Стр. 53: Trimmomatic реализован на Java (кросс-платформенный язык программирования и виртуальная машина), не вполне корректно

говорить о том, что это инструмент командной строки операционной системы Linux.

10. На стр. 64 не приведено информации о доле уникально картируемых прочтений.

К работе есть и несколько содержательных замечаний и вопросов, которые требуют прямых комментариев соискателя.

1. В работе использованы образцы МКПК небольшого числа доноров. Насколько полна, по мнению авторов, созданная панель? Насколько она могла бы вырасти при увеличении числа доноров? Увеличении числа повторностей экспериментов? Расширении спектра омиксных методов?
2. Стр. 19: «негативная регуляторная область». Что имеется в виду? Сайленсер?
3. Стр. 53: для выравнивания прочтений использовался Bowtie 2. Эта программа хорошо себя зарекомендовала, но существенно устарела. Почему не был выбран более современный и точный метод картирования прочтений, например, BWA? Для поиска пиков наоборот (стр. 54) использовался наиболее современный (и менее отлаженный, чем предыдущие версии) метод MACS3.
4. Стр. 55: при поиске гетерозиготных вариантов фильтруются позиции с частотой минорного аллеля менее 20%. Кажется, что это избыточно жесткий порог, поскольку наиболее интересные «экстремальные» аллель-специфичные варианты, по определению, будут иметь малую частоту минорного аллеля. Такой порог можно было бы использовать для поиска вариантов по данным полногеномного ДНК-секвенирования, где не ожидается выраженного функционального аллельного имбаланса.
5. Стр. 56: для идентификации аллель-специфичных вариантов используется простая биномиальная модель, не учитывающая сверхдисперсию, часто наблюдаемую в аллельных частотах для вариантов, детектируемых в омиксных данных. Не является ли такой подход избыточно оптимистичным и (как следствие) увеличивает число малоинтересных вариантов в итоговой панели?
6. Стр. 69: оценка обогащения проведена относительно «всех гетерозиготных нерегуляторных вариантов». Было бы более корректно провести такую оценку для вариантов, не являющихся аллель-специфичными, но локализованных в тех же геномных областях (например, промоторах). В противном случае можно подозревать, что

эффект обогащения вызван не аллель-специфичностью, а геномной локализацией вариантов.

Тем не менее, указанные замечания не снижают высокого уровня работы, не влияют на научную достоверность сделанных выводов и не смазывают общего позитивного впечатления от знакомства с содержанием работы и автореферата.

Таким образом, Диссертационная работа Дамарова Игоря Сергеевича «Полногеномное выявление регуляторных SNPs человека на основе получения и анализа комплекса данных ChIP-seq и RNA-seq», представленная на соискание ученой степени кандидата биологических наук по специальности «1.5.7 – Генетика (биологические науки)» – **является завершенным научно-квалификационным исследованием**, а полученные результаты имеют важное значение в области медицинской генетики человека. Совокупно, принимая во внимания актуальность выбранной темы, научную новизну, достоверность основных положений и представленных результатов и сделанных выводов, представленная **диссертационная работа отвечает требованиям п.п. 9-14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства РФ №842 от 24 сентября 2013 г. (с последующими редакциями), предъявляемых к диссертациям на соискание ученой степени кандидата наук, а ее автор, Дамаров И.С., безусловно заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.7 – Генетика (биологические науки).**

Официальный оппонент:

ведущий научный сотрудник группы регуляции биосинтеза белка
Федерального государственного бюджетного учреждения науки
Института белка Российской академии наук
к.ф.-м.н., д.б.н.

И.В. Кулаковский
27 августа 2025

Контактная информация:

142290 Московская область, г. Пущино
ул. Институтская 4, Институт белка РАН
+7 4967 318441, ivan.kulakovskiy@gmail.com

ПОДПИСЬ

УДОСТОВЕРЯЮ

З.А.В. КАНЦЕЛЯРИИ

ИБ РАН

АКСЕНОВА Т.Н.

27.08.2025

