

## ОТЗЫВ

официального оппонента на диссертацию Цепилова Якова Александровича  
«Разработка и применение новых моделей в полногеномном анализе ассоциаций»,  
представленную на соискание ученой степени кандидата биологических наук по  
специальности 03.02.07 – генетика

**Актуальность исследования.** Проблема генетического картирования многофакторных заболеваний (МФЗ) и сложнопреднаследуемых признаков является одной из важнейших и актуальнейших задач современной генетики человека. Несмотря на впечатляющий прогресс в области разработки новых методов генотипирования (биологические микрочипы на сотни тысяч и миллионы полиморфных маркеров, полногеномное и полноэкзомное секвенирование) и в области биоинформационных и генетико-статистических подходов к анализу больших массивов данных (big data), современная наука не может похвастаться сколь-либо всеобъемлющим знанием генетических основ распространенных болезней и их эндофенотипов (связанных с патогенезом заболевания количественных признаков). Это хорошо иллюстрирует феномен так называемой «упущенной (или «потерянной», как называет ее автор диссертации), наследуемости», который заключается в том, что лишь небольшая часть наследуемой компоненты того или иного МФЗ или другого сложного признака может быть атрибутирована к известным генетическим локусам или маркерам болезни. Обычно доля генетической изменчивости, объясняемая выявленными в ходе ассоциативных исследований маркерами или локусами, не превышает 10%.

Более того, практическая ценность генетических маркеров, ассоциированных с нормальным или патологическим сложным фенотипом (болезнью) для диагностики или предсказания риска развития болезни пока не может конкурировать с традиционными клиническими, параклиническими или анамнестическими данными. Так, автор приводит пример зависимости роста детей на основе роста родителей, который оказывается значительно более точным предиктором, чем предсказание на основе многолокусных генотипов по известным и достоверно ассоциированным с этим чрезвычайно сложным признаком генетическим маркерам. Такая же ситуация характерна и для многофакторных заболеваний. Например, уровень липидов крови и соотношение концентрации их фракций является более надежным критерием риска развития ишемической болезни сердца, чем совокупность ассоциированных с заболеванием генетических маркеров.

Проблема «упущенной наследуемости» и пути ее решения широко обсуждаются в научной литературе. И среди множества возможных траекторий движения к ее решению

особое и чрезвычайно важное место занимает разработка новых методов анализа ассоциаций. Действительно, большая часть имеющихся в литературе данных по полногеномному анализу ассоциаций (ПГАА) получена на базе использования аддитивных моделей наследования, которые отнюдь не всегда отражают биологическую реальность. В связи с этим диссертационное исследование Цепилова Я.А. «Разработка и применение новых моделей в полногеномном анализе ассоциаций», является чрезвычайно актуальным и обладает научной и практической новизной.

**Степень обоснованности и достоверности полученных научных положений и выводов, сформулированных в диссертации, их достоверность и новизна.** Обоснованность положений диссертационной работы Я.А. Цепилова обусловлена адекватной целью и задачами исследования, современными методами, большими массивами данных, использованных для разработки и проверки представленных в исследовании новых методов полногеномного анализа ассоциаций. Положения, выносимые на защиту, и выводы исследования обоснованы и вытекают из материалов диссертации.

**Научная новизна исследования** заключается в том, что в работе впервые предложены методы геномного контроля и полногеномного анализа ассоциаций с использованием неаддитивных моделей наследования, выявлены новые маркеры, ассоциированные с концентрацией некоторых метаболитов в сыворотке крови человека, демонстрирующие неаддитивные механизмы в составе генетической компоненты сложных фенотипов.

**Значение выводов и положений работы для науки и практики** заключается в том, что полученные результаты расширяют представления о полигенном контроле сложнопредопределенных признаков; предлагают новые или существенно модифицированные подходы к анализу данных полногеномного поиска ассоциаций в рамках более широкого спектра моделей аллельных взаимодействий; разработано программное обеспечение, реализующее методы геномного контроля для неаддитивных моделей наследования.

**Оценка содержания диссертации.** Диссертационная работа оформлена в традиционном стиле с соблюдением требований Высшей аттестационной комиссии Российской Федерации. Диссертация изложена на 101 странице машинописного текста и содержит разделы «Оглавление», «Список сокращений», «Обзор литературы», «Материалы и методы», «Результаты», «Обсуждение», «Заключение», «Выводы», «Список литературы». Дополнительные материалы исследования представлены в 4 приложениях. Список литературы включает 110 источников.

Во введение представлена актуальность исследования, раскрыта степень проработанности темы, сформулированы цель и задачи работы, представлены положения о научной новизне и научно-практической ценности исследования. Раскрыт личный вклад автора, сформулированы положения, выносимые на защиту, охарактеризованы публикации по теме диссертации. Формулировка цели работы корректна и соответствует содержанию исследования. Задачи логично вытекают из цели работы. Следует отметить, что информации об апробации работы на научных форумах (а таковая, как следует из автореферата и списка публикаций, проведена в рамках 4 международных конференций по биоинформатике, математической генетике и вычислительной инженерии) не приведена.

Глава «Обзор литературы» содержит краткий анализ современных представлений о генетической эпидемиологии, методах генетического картирования фенотипических признаков человека с акцентом на полногеномный анализ ассоциаций, характеристику моделей неаддитивных эффектов генов и подходов к геномному контролю в ПГАА. В обзоре литературы автору удалось найти хороший баланс между краткостью, четкостью и ясностью изложения, с одной стороны, и обоснованностью и полнотой информации, основанной на анализе современного состояния исследований в области анализа ассоциаций, с другой.

Мне кажется, что обзор литературы в несколько расширенном виде мог бы быть оформлен в отдельную обзорно-методическую публикацию, ориентированную, прежде всего, на молодых исследователей - генетиков и клиницистов в недостаточной степени знакомых с методическими основами анализа ассоциаций и ПГАА, в частности. Это было бы особенно актуальным с учетом недостатка соответствующей литературы на русском языке.

Обзор литературы автор начинает с самых корней генетической эпидемиологии, которые, как и истоки почти любой современной научной дисциплины, можно найти в античные времена. Автор ссылается на воззрения Гиппократов, Аристотеля, Левенгука, Пастера и Коха. Если автор пользовался первоисточниками, то их следовало бы привести в списке литературы. Если информация о взглядах предшественников современной генетической эпидемиологии была почерпнута из историко-обзорных работ, то стоило сослаться на эти работы.

Глава «Материалы и методы» содержит информацию о выборках, использованных в исследовании и методах, использованных для моделирования и симуляции данных, методов, примененных для теста кодоминантной модели, поиска неаддитивных эффектов генов и полногеномного анализа ассоциаций. Исследование проведено на симуляционных

данных и реальных данных трех крупных международных исследований – ERF, KORA и TwinsUK. Поиск неаддитивных эффектов генов в формировании концентрации 151 метаболита сыворотки крови осуществлялся на данных консорциума KORA. Общий объем реальных данных (более 5000 индивидов, прогенотипированных по 300 тыс. – 1.7 млн. однонуклеотидных маркеров (SNP)) – представляет собой огромный массив генотипов, который предоставляет широкое поле для тестирования и апробации разработанных в исследовании методов и поиска маркеров с неаддитивными эффектами.

Для анализа симулированных и реальных данных в полногеномном поиске ассоциаций использовались стандартные современные подходы, адекватные объему данных и задачам исследования. Современность выбранных методов, их соответствие поставленным задачам, большой объем выборок, прогенотипированных маркеров и метаболических фенотипов позволяет не сомневаться в результатах исследования и выводах работы.

Глава «Результаты» представлена на 21 странице и содержит два раздела, посвященных методам геномного контроля и поиску неаддитивных эффектов генов на метаболоме человека. В первой части главы описаны полученные в работе методы геномного контроля при полногеномном анализе ассоциаций. Суть полученных результатов заключается в получении аналитических выражений фактора инфляции тестовой статистики для разных моделей наследования. Значимость этого результата обуславливается тем, что предложенные методы позволяют проводить статистические тесты при ПГАА для любых вариантов аллельных взаимодействий, не ограничиваясь аддитивными, что значительно расширяет информативность подходов, применяемых в ПГАА. Полученные в этом разделе результаты являются новыми, оригинальными и обладают перспективой их ассимиляции и дальнейшего использования научным сообществом.

Во втором разделе главы «Результаты» представлены итоги апробации разработанных методов в исследовании ассоциаций с метаболическими показателями на данных крупных международных популяционных исследований. Было обнаружено 23 маркера, ассоциированных на полногеномном уровне с показателями метаболома человека. 4 из них демонстрируют наличие эффектов неаддитивных взаимодействий. Далее 17 ассоциаций (включая 3 маркера с неаддитивными эффектами) были реплицированы на независимой выборке. Этот раздел работы кроме того что является демонстрацией работоспособности и применимости разработанных в диссертации методов анализа ассоциаций, что являлось его главной задачей, имеет и ряд других важных особенностей. Во-первых, часть ассоциированных с метаболическими

фенотипами локусов выявлена в работе впервые (ассоциация в ПГАА на данных KORA не была показана в предыдущих работах по этому массиву данных). Во-вторых, в некоторой степени результаты работы можно считать иллюстрацией возможного соотношения аддитивных и неаддитивных эффектов в наследственной компоненте некоторых количественных признаков человека, являющихся эндофенотипами распространенных многофакторных заболеваний.

В кратком разделе «Обсуждение», который изложен на 8 страницах автор оценивает результаты диссертационного исследования в контексте современного состояния аналитического аппарата ПГАА и обсуждает применимость аддитивных и неаддитивных моделей в анализе данных, генерируемых современными «омиксными» технологиями.

В «Заключении» автор предельно лаконичен и формирует 5 основных пунктов, отражающих наиболее важные результаты исследования. Выводы соответствуют цели и содержанию работы, отражают основные результаты исследования, сформулированы корректно. В качестве пожелания можно отметить, что хотелось бы видеть лучшее соответствие структуры выводов структуре задач.

Основные результаты исследования полностью отражены в научной печати (6 печатных работ, включая 2 статьи в журналах, входящих в Web of Sciences и 4 тезисов международных конференций). Содержание автореферата соответствует основным разделам и положениями диссертационной работы. Текст диссертации достаточно иллюстрирован рисунками, а данные - таблицами.

В ходе знакомства с работой возникло несколько вопросов по исследованию, которые носят характер дискуссии и не являются указанием на недостатки работы.

1. Чем объясняется тот факт, что при проведенном в работе повторном анализе массива данных, уже подвергавшегося анализу ранее (KORA), были выявлены новые ассоциации генетических маркеров с метаболическими показателями, не зафиксированные ранее?

2. Учитывая результаты поиска неаддитивных эффектов на метаболоме человека и мнение автора о невысокой степени неаддитивности генетического контроля количественных признаков, выявляемых «омиксными» данными, какие фенотипы, анализируемые методами полногеномного анализа ассоциаций, могут представлять собой резерв для поиска «упущенной наследуемости» путем анализа неаддитивных эффектов?

3. Поскольку структурный анализ геном-феномных отношений в формировании признаков человека (прежде всего, болезней и их эндофенотипов) переходит на технологии массового параллельного секвенирования, и учитывая большую вероятность

выявления неаддитивных взаимодействий при анализе тесного неравновесия по сцеплению на данных с высоким геномным покрытием (MPS/NGS) возникает вопрос - насколько применимы разработанные модели для полногеномных или полноэкзомных данных, а не для данных по отдельным точкам генома?

**Заключение.** Диссертационная работа Цепилова Якова Александровича «Разработка и применение новых моделей в полногеномном анализе ассоциаций» является законченным самостоятельным научным квалификационным исследованием, содержит новые научно-практические данные в области генетико-математических моделей полногеномного анализа ассоциаций и полностью соответствует требованиям п. 9 Положения «О порядке присуждения ученых степеней» ВАК Министерства образования и науки РФ (Постановление Правительства РФ от 24 сентября 2013 г. №842), предъявляемых к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.02.07 - генетика.

#### **Официальный оппонент**

Временно исполняющий обязанности директора,  
Заведующий лабораторией эволюционной генетики  
ФГБНУ «Научно-исследовательский институт  
медицинской генетики»,  
доктор биологических наук, профессор

Степанов Вадим Анатольевич  
06.06.2016



(ФГБНУ Научно-исследовательский институт медицинской генетики), 634050, Томск,  
Набережная р. Ушайки, 10. <http://www.medgenetics.ru>; тел. (3822)512228; e-mail  
[vadim.stepanov@medgenetics.ru](mailto:vadim.stepanov@medgenetics.ru)

Ученый секретарь ФГБНУ «Научно-исследовательский институт медицинской генетики»,  
кандидат биологических наук

Хитринская Ирина Юрьевна