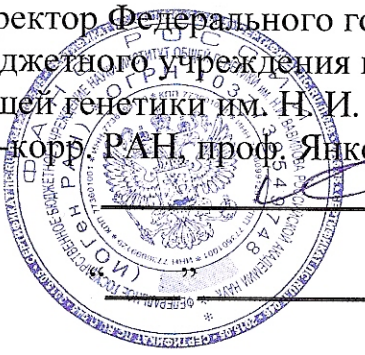
**“УТВЕРЖДАЮ”**

директор Федерального государственного бюджетного учреждения науки Институт общей генетики им. Н. И. Вавилова РАН, чл.-корр. РАН, проф. Янковский Н.К.



_____

“___” _____2016 г.

# ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки «Институт общей генетики им. Н.И. Вавилова» РАН о научно-практической ценности диссертации Цепилова Якова Александровича на тему «Разработка и применение новых моделей в полногеномном анализе ассоциаций» на соискание учёной степени кандидата биологических наук по специальности 03.02.07 – Генетика

**Актуальность темы диссертационного исследования.**

Представленная на соискание учёной степени кандидата биологических наук диссертация посвящена исследованию актуальной проблемы – методам анализа неаддитивных эффектов генов, в том числе методам геномного контроля (ГК) при обработке результатов GWAS в случае использования неаддитивных моделей взаимодействия аллелей (доминантная, рецессивная, кодоминантная и сверхдоминантная модели наследования). Под ГК имеется в виду использование заведомо нейтральных полиморфных сайтов для коррекции результатов GWAS на присутствие генетической структурированности исследуемых выборок (например, наличие близких родственников или этническая гетерогенность). Подобные методы широко используются в случае сравнения выборок по частотам аллелей (аддитивная модель), однако совершенно не разработаны в общем случае. Интерес к этим методам в условия GWAS обусловлен повышенной мощностью тестов, основанных на доминантных/рецессивных схемах наследования по сравнению с аддитивной моделью.

**Научная новизна и практическая значимость.**

В диссертации Я.А. Цепилова впервые разработаны методы ГК для моделей неаддитивных аллельных взаимодействий (кодоминантной, доминантной, рецессивной, сверхдоминантной) применительно к анализу результатов GWAS. Получены аналитические выражения для фактора инфляции тестовой статистики для неаддитивных моделей наследования в условиях генетической гетерогенности выборки. Показано, что в отличии от аддитивного случая в общем случае фактор инфляции дисперсии (VIF) существенно зависит от частоты аллеля и выбранной модели наследования.

На основе полученных аналитических выражений разработано соответствующее программное обеспечение, реализующее методы ГК для неаддитивных моделей. Предложен набор статистических тестов, позволяющих установить значимость отклонений от аддитивной модели наследования. Проведена оценка ошибок первого рода и мощности разработанных тестов с использованием моделированных и реальных данных.

В рамках апробации впервые осуществлен неаддитивный анализ полногеномных ассоциаций концентраций метаболитов сыворотки крови человека. В результате были идентифицированы четыре локуса, обладающих значимыми неаддитивными эффектами.

Предложенные в диссертации подходы по поиску неаддитивных эффектов могут быть использованы при полногеномном анализе широкого спектра признаков.

**Достоверность полученных результатов**

Работа выполнена на современном методическом уровне. Для анализа симулированных и реальных данных использовались тесты, реализованные в пакете MixABEL, который является частью известного набора программ GenABEL. Обзор литературы, постановка цели и задач исследования, его проведение и обсуждение результатов свидетельствуют о хорошей теоретической и методической подготовленности диссертанта. Выводы,

сформулированные в диссертации, обеспечиваются качественным анализом научной литературы по изучаемой проблематике и применением адекватных статистических методов обработки данных.

Разработанные методы ГК были успешно апробированы на данных популяционного исследования KORA (1785 человек, 1717498 SNP, уровень мочевой кислоты и концентрации 151 метаболита в сыворотке крови) и близнецового исследования TwinsUK (843 человека).

Основные положения диссертации доложены на Международных конференциях в Германии, Нидерландах и Новосибирске. По теме диссертации опубликовано 6 печатных работ, из них 2 статьи в престижных зарубежных журналах с высоким импакт-фактором, реферируемых в ISI Web of Science и рекомендованных ВАК Минобрнауки России. Автореферат и опубликованные работы полностью соответствуют материалам диссертации.

Таким образом, полученные Я.А. Цепиловым научные результаты и выводы являются обоснованными и достоверными.

**Структура диссертации**

Диссертация включается в себя введение, четыре главы, обсуждение, заключение, выводы и библиографический список литературы, состоящий из 110 источников. Основной текст диссертации изложен на 107 страницах, сопровождается 8 рисунками и 14 таблицами.

Во введении **(первая глава)** раскрыта актуальность темы исследования, определены цели и задачи, сформулированы основные положения, обладающие научной новизной и выносимые на защиту.

**Вторая глава** диссертации посвящена обзору литературы. Обзор затрагивает все ключевые работы по тематике исследований, хорошо читается и представляет самостоятельный интерес для широкого круга генетиков, эпидемиологов и биоинформатиков. Вместе с тем, по мнению рецензента, едва ли стоило столь подробно, начиная с Гиппократа, излагать основы эпидемиологии, генетического картирования и сцепления. Более уместно было бы сосредоточиться на собственно математических методах обработки GWAS. Например, бурно обсуждаемый метод GCTA (см.

критическое выступление Маркуса Фелдмана) удостоился лишь беглого упоминания (статья Дж. Янга и др.), хотя речь идет о принципиально новых попытках решения проблемы missing heritability.

В **третьей главе** представлены описания использованных баз экспериментальные данных и перечень методов, применявшиеся при поиске неаддитивных эффектов генов

Основные результаты работы представлены в **четвертой главе**. Здесь приводится вывод аналитических выражений для VIF в случае произвольных схем наследования, даются оценки параметров VIF и рассматриваются различные аппроксимации. Далее приводятся результаты апробации оценок на симулированных и реальных данных.

В **пятой главе** проводится обсуждение полученных результатов и рассматривается применимость аддитивных моделей для полногеномных ассоциативных исследований «омиксных» признаков.

**Замечания**

Принципиальных замечаний к работе нет. Работа написана предельно ясно и хорошо иллюстрирована. Имеются немногочисленные редакционные замечания: местами читаем *m1* вместо $m_1$ и т.п. Однако хотелось бы сделать ряд общих необязательных замечаний дискуссионного характера.

1. Название работы представляется излишне многообещающим. В работе решена конкретная и важная задача – построить и апробировать алгоритмы ГК применительно к доминантным, рецессивным и пр. моделям наследования в условиях GWAS. Однако представленные исследования едва ли соответствуют понятию «новая модель». Новая модель – это то, чем увлекается команда Питера Виссчера, которая утверждает, что их вычисления снимают проблему missing heritability [Yang J., Lee S.H., Goddard M.E., Visscher P.M. GCTA: a tool for genome-wide complex trait analysis // Am. J. Hum. Genet. 2011. V. 88(1)]. В представленной работе речь идет о вполне традиционных подходах. Доминантные/рецессивные модели обладают большей мощностью и почти всегда используются при анализе ассоциативных исследований с небольшим числом полиморфных сайтов.

GWASы действительно почти всегда обрабатывались через сравнения частот аллелей (что эквивалентно аддитивной модели наследования), в частности, из-за отсутствия адекватных методов ГК в общем случае. Теперь исследования автора ликвидируют этот досадный пробел.

2. На стр. 94 (и не только!) читаем: «Отсутствие различных неаддитивных эффектов в используемых моделях может быть одной из причин, по которым нам пока не удается объяснить наследуемость в широком смысле». Рецензент сомневается, что окончательное решение проблемы missing heritability может быть хоть как-то связано с разработкой новых математических методов обработки GWAS. Увы, мы занимаемся генетикой, а не астрофизикой. Существенные прорывы, скорее всего, будут возможны лишь с открытием новых видов генетической/эпигенетической изменчивости, а также регуляторных механизмов.

3. Термин «неаддитивные модели» используется в работе для обозначения доминантных, рецессивных и прочих моделей наследования, связанных со сравнением частот генотипов (но не аллелей) одного сайта. Формально здесь все правильно. Однако более привычно полагать, что «неаддитивная модель» - это то, что описывает неаллельные взаимодействия разных генов, т.е. эпистаз.

4. Вывод 3 (эффективность поиска неаддитивных локусов зависит от плотности генетических маркеров) как-то слабо подкреплен соответствующими рассуждениями в тексте диссертации.

Приведенные замечания не влияют на общее исключительно благоприятное впечатление от диссертационной работы Я.А. Цепилова.

**Заключение**

Диссертационная работа Цепилова Я.А. на тему «Разработка и применение новых моделей в полногеномном анализе ассоциаций» является законченной научно-квалификационной работой, выполненной под руководством доктора биологических наук Аульченко Ю.С, содержащей новое решение актуальной научной задачи – разработка методов геномного

контроля для неаддитивных моделей при полногеномном анализе ассоциаций, имеющей существенное значение для специальности 03.02.07 - Генетика. По актуальности, научной новизне, теоретической и практической значимости, достоверности полученных результатов и обоснованности выводов диссертационная работа Цепилова Якова Александровича соответствует требованиям п. 9 «Положения о порядке присуждения учёных степеней», утверждённого постановлением Правительства РФ № 842 от 24.09.2013 г., предъявляемым к кандидатским диссертациям, ее автор – Цепилов Яков Александрович – заслуживает присуждения учёной степени кандидата биологических наук по специальности 03.02.07 - Генетика.

Зав. лаб. экологической генетики ФГБУН ИОГен РАН
Доктор биологических наук Рубанович Александр Владимирович

Отзыв утвержден на межлабораторном семинаре «Мутагенез и генетическая безопасность» ИОГен РАН 19.05.16 (протокол № 42).

Рук. семинара зам. директора ФГБУН ИОГен РАН
Доктор биологических наук, профессор Абилев Серикбай Каримович

Подпись А.В. Рубановича заверяю:
ученый секретарь ФГБУН ИОГен РАН
доктор биол. наук Огаркова Ольга Александровна

«23» мая 2016 г.

Федеральное государственное бюджетное учреждение науки
Институт общей генетики им. Н.И. Вавилова Российской академии наук
119991, Москва, ГСП-1, ул. Губкина, 3
Тел.: 499 135-62-13, E-mail: iogen@vigg.ru

Избранные работы сотрудников Института за последние 5 лет по теме, соответствующей данной диссертации

1. Protasova MS, Grigorenko AP, Tyazhelova TV, Andreeva TV, Reshetov DA, Gusev FE, Laptenko AE, Kuznetsova IL, Goltsov AY, Klyushnikov SA, Illarioshkin SN, Rogaev EI. Whole-genome sequencing identifies a novel ABCB7 gene mutation for X-linked congenital cerebellar ataxia in a large family of Mongolian ancestry. // Eur J Hum Genet. 2015 Aug 5.

2. Poliakov, E., Cooper, D. N., Stepchenkova, E. I., & Rogozin, I. B. (2015). Genetics in Genomic Era. Genetics research international. dx.doi.org/10.1155/2015/364960

3. Veytsman B, Wang L, Cui T, Bruskin S, Baranova A. Distance-based classifiers as potential diagnostic and prediction tools for human diseases. BMC Genomics. 2014 Dec 19;15 Suppl 12:S10.

4. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P, Pagani L, Tyler-Smith C, Xue Y, Cucca F, Schurr TG, Gaieski JB, Melendez C, Vilar MG, Owings AC, Gómez R, Fujita R, Santos FR, Comas D, Balanovsky O, Balanovska E, Zalloua P, Soodyall H, Pitchappan R, Ganeshprasad A, Hammer M, Matisoo-Smith L, Wells RS; Genographic Consortium. Geographic population structure analysis of worldwide human populations infers their biogeographical origins // Nat Commun. 2014 Apr 29;5:3513. doi: 10.1038/ncomms4513. - IF 10.742

5. Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. Exploring massive, genome scale datasets with the GenometriCorr package // PLoS Comput Biol. 2012 May; 8(5):e1002529.

6. Rubanovich A.V., Khromov-Borisov N.N. Theoretical analysis of the predictability indices of the binary genetic tests // Russ. J. Genet.: Appl. Res. 2014. V. 4. No. 2. P. 146–158. DOI: 10.1134/S2079059714020087.

Список работ сотрудников Института теме, диссертации, заверяю:

ученый секретарь ФГБУН ИОГен РАН

доктор биол. наук Огаркова Ольга Александровна