

## Отзыв

официального оппонента на диссертационную работу

**Романова Дмитрия Евгеньевича**

**«Анализ связи между структурно-функциональной организацией генома в окрестностях регулирующих рост генов и морфо-физиологическими характеристиками млекопитающих»,**

представленную на соискание ученой степени кандидата биологических наук по специальности 03.02.07 – «генетика».

**Актуальность темы исследования.** Получение полногеномных последовательностей для большого числа видов животных, в т.ч. млекопитающих, открыло возможности для разнообразных корреляционных анализов, направленных на описание закономерностей структуры и эволюции генома между видами. Такие анализы быстро выявили сложную сеть корреляций. Оказалось, что координаты, взаимное положение, нуклеотидный состав, консервативность и прочие характеристики различных геномных элементов находятся в сложной взаимосвязи друг с другом. Ясно, что эти взаимосвязи определяются многими силами, действовавшими в эволюционной истории соответствующих видов: естественным отбором, в т.ч. действующим скоординировано на различные элементы; неравномерностью мутагенеза; генетическим дрейфом; рекомбинацией и др. Характер многих из этих сил остается неизвестным. Для того, чтобы точно определить функциональную значимость наблюдаемых корреляций, необходимы эксперименты, в которых наблюдаемые паттерны (например, взаимное положение двух функциональных элементов) нарушаются, и исследуются вызываемые такими нарушениями изменения фенотипа; однако на масштабе полных геномов сегодня такие исследования невозможны. Поэтому корреляционные биоинформатические исследования остаются актуальными.

Предлагаемая к защите работа относится к этой категории. В ней анализируется возможная значимость взаимного расположения и удаленности функциональных и консервативных элементов генома видов млекопитающих. При этом рассматривается набор из 17 генов, участвующих, по литературным данным, в регуляции роста. Выявление новых функциональных взаимосвязей между элементами генома является, безусловно, актуальной научной задачей, которая имеет в том числе и прикладное значение.

**Структура диссертации.** Диссертационная работа Д.Е.Романова с точки зрения оформления и подачи материала соответствует требованиям, предъявляемым к

кандидатским диссертациям. Представленная работа изложена на 123 страницах и включает в себя список сокращений и условных обозначений, введение (где, в том числе, излагается актуальность проблемы, цели и задачи исследования, его новизна и практическая значимость), обзор литературы, материалы и методы, результаты, обсуждение, заключение, выводы, список литературных источников из 145 наименований, список иллюстративного материала и список таблиц. Работа включает 11 рисунков и 36 таблиц. Все разделы написаны хорошим понятным языком, не перегружены жаргоном, их приятно читать.

**Научная новизна и значимость результатов.** В представленной работе выявлена корреляция между некоторыми характеристиками видов млекопитающих: фенотипическими (массой, размером тела, продолжительностью жизни, периодом полового созревания) и геномными (расстояние между консервативными элементами в окрестностях определенных генов, расстоянием от гена до ближайшей теломеры). Полученные результаты обсуждены с точки зрения их возможного биологического значения. Эти результаты являются новыми и потенциально интересными. Кроме того, разработан впечатляющий арсенал биоинформатических инструментов; часть из них играют техническую роль, но программа для построения множественных дот-плотов не имеет известных мне аналогов и выглядит полезной. В целом работа представляется мне научно новой и потенциально значимой.

**Степень обоснованности и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации.** Результаты разносторонне и систематически обсуждены, делаемые из них выводы подробно обоснованы. Сильной стороной работы являются четко сформулированные на основании полученных результатов модели (с. 90 и 93), из которых следуют конкретные предсказания. Часть этих предсказаний можно проверить по имеющимся данным, например, по экспрессии соответствующих генов в различном возрасте.

#### **Замечания.**

1. Работа во многом основана на корреляционных анализах с участием большого числа видов млекопитающих (напр. таб. 3.2, 3.3, рис. 3.2, 3.3 и ниже). При этом, по-видимому, никак не учтен тот факт, что отдельные виды млекопитающих не являются независимыми экспериментальными точками. Можно привести следующую аналогию. Проанализировав большое число видов млекопитающих и

птиц, можно выявить сильную статистически значимую ассоциацию между кормлением детей молоком и наличием зубов. Но эта ассоциация будет связана не с функциональной связью между этими двумя признаками, а с тем, что они оба присутствуют у млекопитающих и отсутствуют у птиц. Эта статистическая проблема известна как «филогенетическая псевдорепликация». Меня беспокоит, что она могла внести вклад в наблюдаемые в данной работе сильные корреляции. Действительно, многие из анализируемых пар видов являются эволюционно достаточно близкими. Например, два грызуна (мышь и крыса) на рис. 3.1 имеют очень близкое расстояние между консервативными элементами MYCN(-6893) и MYCN(74), также как и два кита (кит-полосатик и косатка); естественно, массы тела в этих парах также близки. Однако можно предположить, что близость расстояний между элементами в парах связана с эволюционной близостью соответствующих видов, а не с необходимостью в одинаковой регуляции. Для обращения с наборами данных, характеризующимися филогенетической структурой, существуют специальные методы ([https://en.wikipedia.org/wiki/Phylogenetic\\_comparative\\_methods](https://en.wikipedia.org/wiki/Phylogenetic_comparative_methods)); использование подобных методов сделало бы полученные результаты более убедительными.

Остальные замечания к работе – более мелкие.

2. Одним из результатов работы стала разработка программного конвейера на основе bash-скриптов для автоматизации частых биоинформатических операций. Хотелось бы уточнить, существуют ли у этого конвейера аналоги и если да – чем предпочтителен именно авторский вариант.
3. Как были выбраны 17 генов для анализа?
4. Можно предположить, что основным фактором, который должен определять расстояния между консервативными элементами, должен быть размер генома: в более компактных геномах функционально важные участки будут расположены ближе друг другу. Наблюдалась ли подобная зависимость?
5. Некоторые выводы – слишком смелые, например – о форме зависимости для полученных корреляций (с. 77 и 89). Чтобы показать, что зависимость именно экспоненциальная (или хотя бы – что она ближе к экспоненциальной, чем к линейной), нужны специальные статистические тесты, которые не приведены. В другом месте (с. 81) статистически незначимые подтверждающие результаты (с  $p=0.06$  и  $p=0.17$ ) по принятому авторами порогу значимости (0.05) обсуждаются так, как если бы они были значимыми.

