

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
НАУЧНОЕ УЧРЕЖДЕНИЕ

«ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ
ЦИТОЛОГИИ и ГЕНЕТИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК»
(ИЦиГ СО РАН)

Пр-т. Академика Лаврентьева, д. 10, Новосибирск, 630090
Телефон: (383) 363-49-80
Факс (383) 333-12-78
E-mail: icg-adm@bionet.nsc.ru
<https://www.icgbio.ru>
ИНН 5408100138/КППИ 540801001
ОКПО 03533895 ОГРН 1025403657410

от 10.08.2024 № 15345-29-38/740
на № _____ от _____



ОТЗЫВ

ведущей организации о диссертационной работе **Давитавяна Сурена Самвеловича на тему «Характеристика молекулярного разнообразия рака молочной железы и глиомы на основе транскриптомных, геномных и эпигенетических данных»**, представленной на соискание ученой степени кандидата биологических наук по специальности 03.00.02 Биофизика, биоинформатика в специализированном совете ВАК РА.

Актуальность темы диссертационной работы

Диссертационная работа Давитавяна С.С. посвящена разработке интегрированного биоинформатического подхода для анализа мультиомиксных данных и обобщения знаний в области канцерогенеза для изучения разнообразия на молекулярном уровне рака груди и глиом низкой степени злокачественности. Актуальность работы обоснована сразу несколькими факторами.

Во-первых, онкологические заболевания являются одной из наиболее частых причин смертности в мире. Для борьбы с ними активно разрабатываются эффективные технологии, основанные на детальном знании молекулярных механизмов рака. Это позволило в последнее время создать новые подходы к лечению раковых заболеваний, увеличить общую выживаемость пациентов в результате терапии, минимизировать осложнения, оценивать исход лечения.

Во-вторых, наши знания показывают, что рак является сложным многофакторным заболеванием. Существует гетерогенность злокачественных опухолей на молекулярном уровне, и, следовательно, разнообразие различных типов рака. Опухолевые клетки демонстрируют изменчивость на уровне генома, эпигенома и транскриптома. Эти особенности раковых клеток характерны в частности для рака груди и глиом низкой степени злокачественности. Разнообразие опухолевых клеток затрудняет лечение, так как потенциальные способы воздействия на конкретные типы опухолей могут быть не учтены. Это может привести к выбору неоптимальной стратегии для лечения пациентов.

В-третьих, успехи современных экспериментальных молекулярно-биологических методов позволяют получать огромное количество данных на разных молекулярных уровнях: геномном, эпигеномном, транскриптомном, протеомном и т.п. Для этих данных обеспечена аннотация и доступ для широкого круга исследователей, что позволяет осуществлять их массовую обработку, выявлять функциональные связи между генами, особенностями опухолей и их развития, выявлять важные молекулярные маркеры, в том числе и прогностические. Анализ и обработка этих данных требует развития новых биоинформатических методов, которые заключаются в интеграции и представлении результатов в виде, удобном для биологической интерпретации.

Необходимо отметить, что несмотря на то, что в мире разработано большое количество программ для решения задач подобного сорта, не всегда их можно использовать в случае опухолей конкретного типа. Для этого требуется реорганизация и подстройка потоков обработки данных для специфических целей, адаптация алгоритмов.

В соответствии с вышесказанным, цель диссертационной работы Давитавяна С.С., а также задачи, поставленные для ее достижения, являются актуальными.

Научная новизна и практическая значимость работы

Работа Давитавяна С.С. построена по классической для биоинформатических работ схеме: (1) предложен оригинальный компьютерный метод для решения комплексной биологической задачи; (2) с помощью разработанного метода в результате широкомасштабного анализа данных получены новые биологические результаты.

В диссертации разработан оригинальный подход для интеграции и анализа мультиомиксных данных на основе самоорганизующихся карт (Self-Organizing maps, SOM). Это популярный метод машинного обучения «без учителя», преимущество которого заключается в том, что он не требует процедуры разметки исходных данных экспертом (в случае с мультиомиксными данными это весьма затруднительно). Он заключается в отображении свойств биологических объектов, описанных множеством различных характеристик на двумерную сеть нейронов. Интенсивность нейронов позволяет описывать

внутреннюю структуру данных в виде т.н. «портретов». Группы соседних нейронов соответствуют группам генов, сходных по молекулярным характеристикам. Наиболее активно этот метод используется в задачах классификации больших данных. Однако, существующие методы могли использовать лишь однородные входные данные для классификации. Давитавян С.С. модифицировал алгоритм SOM, реализовав возможность подачи данных на вход алгоритма в виде множественных слоев (ml-SOM). Эта модификация алгоритма впервые позволила использовать мультиомиксные данные для описания и классификации раковых опухолей. Эти данные могут включать динамические экспрессионные (транскриптомные и протеомные), а также регуляторные данные (метиломные, вариации числа копий генов, одиночные вариации нуклеотидов).

Автор также предложил оригинальный подход, который позволяет отображать новые данные в пространство карт SOM без дополнительного обучения алгоритма. Этот новый метод (sup-SOM) использует метод регрессии на основе машины опорных векторов (SVMR) и существенно экономит вычислительные ресурсы при анализе огромного массива мультиомиксных данных. Этот подход позволяет получать карты SOM без обучения сети с высокой точностью (0.9-0.99).

Для интерпретации полученных классификаций опухолей на основе молекулярных данных автор предложил модель, основанную на карте фенотипа.

Предложенные подходы, включая методы предварительной обработки данных и интеграции метаинформации об образцах опухолей из базы данных TCGA, были реализованы в виде уникального вычислительного конвейера. Таким образом, в области разработки методов анализа мультиомиксных данных, связанных с описанием молекулярных характеристик различных типов рака, диссертационная работа, несомненно, обладает высокой степенью новизны.

Разработанные методы были впервые использованы для описания мультиомиксного ландшафта опухолей рака груди и глиом низкой степени злокачественности. На их основе были выявлены новые молекулярные особенности различных типов этих опухолей.

Для набора образцов опухолей рака молочной железы, классифицированного ранее по методу PAM50 на 7 типов, удалось показать значимые различия в картах SOM этих типов и выявить 18 групп генов (генные модули A-R), обладающих разными функциональными характеристиками. Наиболее представленными оказались группы генов, связанных с такими функциями как клеточный цикл, метастазирование (A), функции микроРНК и репарации ДНК (C), сплайсинг (R) и др. По стандартному протоколу проведен детальный функциональный анализ основных модулей генов. Проведена оценка влияния характеристик по каждому из типов мультиомиксных данных на выживаемость пациентов для разных типов рака. Выявлены

молекулярные особенности опухолей разного типа, ассоциированные с выживаемостью пациентов. В результате впервые фактически был создан «атлас» генов рака молочной железы, предоставляющий максимально полную информацию об их функционировании в пространстве мультиомиксных данных.

Аналогичная работа была проведена и для глиом низкой степени злокачественности. Был проведен анализ 4 известных субтипов этих опухолей. Анализировались профили экспрессии и метилирования генов, а также вариации числа их копий. Выделено 11 генных модулей (А-К) и проведена их функциональная аннотация. Показано, что описание молекулярных особенностей глиом на основе разработанного подхода дает лучшее разделение субтипов и прогностические результаты, чем классификация WHO на основе генетических данных.

Полученные результаты имеют высокую практическую значимость, поскольку они позволяют описывать молекулярные механизмы злокачественных опухолей с большей детализацией, более точно выявлять маркерные гены и оценивать выживаемость пациентов.

Обоснованность и достоверность научных положений, выводов и заключений

Достоверность полученных результатов обусловлена несколькими факторами. Прежде всего, автор тщательно подошел к выбору данных в своих исследованиях, использовал наиболее авторитетный ресурс по молекулярно-биологическим данным в области онкологии – базу данных TCGA. Данные на первом этапе анализа были предварительно обработаны стандартными общепризнанными методами (нормализация, устранение смещения и т.п.). При разработке методов автор опирался на программное обеспечение в области анализа данных (opusSOM), который ранее успешно зарекомендовало себя при решении самых разных биологических задач. При проверке статистических гипотез автор использовал широкий спектр статистических методов для проверки гипотез, реализованных в пакете R (регрессионные модели, корреляционный анализ, тест Даннетта, Z-статистика и др.).

Основные биологические результаты по выявлению молекулярных особенностей опухолей различных типов в основном хорошо согласуются с существующими представлениями о функциях генов, ассоциированных с развитием злокачественных опухолей. Все это свидетельствует об обоснованности сделанных на основе анализа выводов и заключений.

Оценка содержания диссертационной работы и ее завершенности

Материалы и результаты работы изложены последовательно и аргументировано. Диссертация построена по традиционному плану и включает следующие разделы: введение, литературный обзор, материалы и методы, результаты и обсуждение, заключение, выводы, список литературы, приложение. Приведен список сокращений. Работа изложена на английском языке на 100 страницах и содержит 29 рисунков и 2 таблицы. Список литературы

содержит 266 библиографических ссылок. Диссертационная работа, как и автореферат, оформлены в соответствии с установленными требованиями. Автореферат в необходимой и достаточной степени отражает результаты диссертационной работы.

Литературный обзор описывает молекулярное разнообразие двух типов злокачественных опухолей: рака груди и глиом низкой степени злокачественности. Дает описание современных молекулярно-биологических ресурсов по омиксным данным в области изучения рака. Приводится описание основных типов методов машинного обучения, более детально описан метод самоорганизующихся карт и их применения к анализу биологических данных.

В главе 2 «Материалы и методы» описаны источники экспериментальных данных для анализа, процедуры их предварительной обработки, описана оригинальная модификация метода SOM для использования мультиомиксных данных. Детально описан протокол и методы статистического анализа больших данных, метод описания фенотипа и проекции новых данных на существующие SOM.

В главе 3 «Результаты и обсуждение» детально описан вычислительный конвейер для решения задачи массового анализа мультиомиксных данных для раковых опухолей разного типа. Описаны результаты анализа рака груди и глиом низкой степени злокачественности. Детально рассмотрены характеристики для выявленных модулей генов. Обсуждаются их функции в сравнении с существующими представлениями о молекулярных особенностях рака обоих типов.

В «Заключении» обобщаются полученные результаты работы, преимущества и недостатки новых методов, возможности его использования для более широкого круга анализа больших данных в области молекулярной биологии.

Текст работы полностью соответствует поставленным целям исследования. Таким образом, поставленные цели и задачи выполнены в полном объеме, что позволяет говорить о завершенности поставленной работы. Заключение и выводы являются обоснованными и аргументированными.

Замечания по содержанию диссертации

К работе есть ряд незначительных замечаний.

1. Автор указывает, что недостатком существующих алгоритмов SOM является их невозможность добавлять новые данные без последующего переобучения сети. Это требует временных затрат, а предлагаемый метод, supSOM, более эффективен. Хотелось бы в этой связи получить оценки, сколько времени занимает обучение SOM и использование метода supSOM.

2. В работе не хватает указания, где можно получить программное обеспечение, которое было разработано в процессе выполнения диссертации, либо указания, что оно является частью или

дополнением к уже существующим пакетам. Полагаю, что доступ к ПО был бы полезен для широкого круга исследователей.

3. Некоторые рисунки имеют довольно низкое качество (например, рис. 10, 25, 26, 27). Особенно сложно разобрать мелкий шрифт. На рисунке 12 неудачно подобраны цвета шрифтов для классов опухолей LumB и normal-like, их очень трудно разобрать.

Эти недостатки, однако, не снижают высокий уровень представленной работы и основных научных положений и выводов диссертации.

Подтверждение опубликования полученных результатов диссертации в научной печати

По материалам диссертационной работы опубликовано пять работ. Три из них в рецензируемых журналах, одна в Российском журнале («Вестник РАУ»), две других в зарубежных (Cancers, BioMedInformatics). Опубликованные по теме диссертации работы в достаточной степени отражают ее основные положения. Также основные положения и результаты диссертационной работы были представлены на международных и российских научных конференциях и опубликованы в сборниках трудов этих конференций.

Заключение

Диссертационная работа Давитавяна С.С. является законченной и выполнена соискателем на высоком научном уровне. Представленные в работе исследования достоверны, представленные алгоритмы хорошо описаны, биологические результаты, полученные с помощью этих методов, также являются достоверными. Выводы и заключения обоснованы.

По актуальности, научной новизне, теоретической и практической значимости представленных результатов, диссертационная работа Давитавяна С.С. «Характеристика молекулярного разнообразия рака молочной железы и глиомы на основе транскриптомных, геномных и эпигенетических данных» соответствует требованиям, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.00.02 Биофизика, биоинформатика.

Отзыв был заслушан и утвержден на межлабораторном семинаре по молекулярной генетике, клеточной биологии и биоинформатике Института цитологии и генетики СО РАН. «ФИЦ Институт цитологии и генетики СО РАН» 3 июня 2024 года.

зав. лабораторией эволюционной
биоинформатики и теоретической генетики
«ФИЦ Институт цитологии и генетики СО РАН»,
д.б.н., доцент
(e-mail ada@bionet.nsc.ru)

Ученый секретарь ИЦиГ СО РАН, к.б.н.



Афонников Д.А. Афонников Д.А.

Г.В.

Орлова Г.В.