

ОФИЦИАЛЬНЫЙ ОТЗЫВ

на диссертационную работу Аракеляна Арсена Арташесовича «Разработка биоинформатических подходов для изучения молекулярных механизмов развития хронических заболеваний человека», представленную на соискание ученой степени доктора биологических наук по специальности 03.00.03. «Молекулярная и клеточная биология».

Компьютерная наука и информационные технологии являются мощными источниками получения информации, без которой, в настоящее время, не мыслимо развитие науки, в том числе биологической. Не случайно, по мнению многих знаменитых ученых-естественников наиболее выдающиеся достижения в области физико-химической биологии, молекулярно-клеточной биологии, биомедицины в недалеком будущем будут получены в исследованиях с применением математических и биоинформатических подходов. Важность развития биоинформатики трудно переоценить хотя бы потому, что она позволяет исследовать сложнейшие проблемы молекулярной биологии и генетики без применения длительных и трудоемких экспериментальных методов, дорогостоящего оборудования, и что особенно важно, посредством оригинальных и, во многих случаях, более достоверных подходов, таких как компьютерное программирование и математическое моделирование. Использование специфических алгоритмов, биоинформатических методов дает возможность анализа больших объемов данных, что особенно важно в работах преимущественно с геномными и белковыми последовательностями. Биоинформатические, модельные теоретические исследования важны также в аспекте оптимизации экспериментов, выявления наиболее эффективного пути экспериментального изучения того или иного процесса, что позволяет экспериментатору сэкономить время и средства для разрешения стоящей перед ним сложной научной задачи.

Вышеизложенное свидетельствует о значимости подобных исследований, об актуальности работ по изучению молекулярных механизмов разнообразных, сложнейших клеточных процессов, происходящих в норме и в патологии с применением биоинформатических подходов. Обсуждаемая диссертационная работа является типичным примером таких исследований.

Основная цель работы – изучение молекулярных механизмов развития хронических неинфекционных заболеваний человека с использованием

биоинформатических алгоритмов и программных пакетов. Для достижения этой цели диссертанту необходимо было во-первых разработать эти алгоритмы и пакеты программ (первая часть работы), а уже затем исследовать глобальную экспрессию генов и активность биологических путей при хронических неинфекционных заболеваниях (второй этап исследования). Перед диссертантом были поставлены две трудоемкие задачи. Первая - разработать биоинформатические подходы для анализа экспрессии генов и идентификации потенциальных биомаркеров и вторая, не менее трудоемкая задача, – разработать биоинформатические методы для моделирования активности биологических путей.

Для выполнения этих задач диссертантом проведен целый ряд оригинальных исследований. Среди них: исследования по оценке влияния топологии биологических путей на результаты функционального анализа общей экспрессии генов /1/, по анализу топологической резистентности биологических путей к мутациям /2/, по изучению общих и специфических особенностей активации биопутей при онкологических и хронических воспалительных заболеваниях легких /3/, по сравнительному анализу профилей активации биопутей при аутоиммунных и аутовоспалительных заболеваниях /4/, по исследованию влияния единичных мутаций экспрессии генов и активации биопутей при моногенных аутовоспалительных заболеваниях /5/, по оценке дифференциальной экспрессии генов при посттравматическом стрессовом расстройстве /6/. Этот простой перечень проведенных исследований свидетельствует об огромном объеме работы, охватывающей солидный спектр неинфекционных хронических заболеваний человека.

Обсуждаемая диссертация является сугубо теоретическим исследованием, по всей вероятности, первой в Армении диссертационной работой на соискание ученой степени доктора биологических наук, со структурой, соответствующей теоретическим работам. Она состоит из введения, четырех глав («Обзор литературы», «Разработка логико-комбинаторных подходов анализа экспрессии генов», «Разработка биоинформатических методов для моделирования активности биологических путей», «Исследование молекулярных механизмов патогенеза комплексных заболеваний человека»), заключения, выводов и списка цитируемой литературы.

Во введении автор останавливается на актуальности проблемы, аргументирует цель и основные задачи работы, излагает научную новизну, теоретическое значение и научно-практическую ценность работы, приводит сведения о публикациях по теме работы и об апробациях работы.

Главы диссертационной работы содержат большое количество понятий и терминов, присущих, с одной стороны области информационных технологий и биоинформатики, а с другой - молекулярно-клеточной биологии и генетики. Эта специфика работы, с одной стороны, вынуждает автора порой давать более подробные объяснения и описание терминов и явлений, что разумеется несколько увеличивает объем работы, с другой стороны делает работу более доступной для специалистов смежных наук. Учитывая это, при обсуждении глав диссертации, постараюсь по возможности кратко останавливаться лишь на, с моей точки зрения, более значимых результатах.

Первая глава диссертации - Обзор литературы, охватывающий 37 страниц машинописи, состоит из четырех разделов. В первом разделе подробно приводятся имеющиеся в литературе данные о молекулярных механизмах развития неинфекционных заболеваний человека, причем как моногенных, так и полигенных. Классифицируются моногенные заболевания: аутосомно-доминантные и рецессивные, Х-сцепленные доминантные и рецессивные, Y-сцепленные и митохондриальные. Даются основные характеристики полигенных заболеваний. Особо информативна и важна приводимая диссертантом таблица о наиболее популярных базах данных и хранилищ по генетическим ассоциациям. Интересен и второй раздел Обзора, где биология характеризуется как наука о данных. Примечательно, что объем биологической информации, относящийся к различным аспектам жизнедеятельности живых систем, растет с невероятной скоростью и к 2024 году достигнет до 1 зеттабайта / 10^{21} байта/, что сделает биологию лидером по объему всех имеющихся данных! Это весомый аргумент в пользу необходимости применения информатических подходов в биологических исследованиях. Третий и четвертый разделы Обзора посвящены описанию литературных данных по использованию анализа транскриптома при изучении молекулярных механизмов развития заболеваний и по анализу экспрессии генов, имеющих особо важное значение для дальнейшей интерпретации результатов, полученных диссертантом. В целом, Обзор литературы оставляет весьма благоприятное впечатление, он достаточно информативен и читается с большим интересом. Чувствуется, что диссертант досконально знаком с соответствующей литературой и обладает необходимыми знаниями по обсуждаемым проблемам.

Вторая глава диссертации, охватывающая 36 страниц, посвящена разработке методических подходов анализа экспрессии генов. Глава состоит из четырех разделов, посвященных основным типам задач обработки данных в анализе транскриптома /1/,

функциональному анализу экспрессии генов /2/, алгоритму растущих опорных множеств /3/ и валидации этого алгоритма /4/.

В первом разделе автор описывает основные типы задач при анализе транскриптома. При этом, в зависимости от поставленной задачи о нахождении функциональных зависимостей между генотипом и фенотипом, характеризует три основные типы сравнительного анализа дифференциальной экспрессии генов. Из них диссертант, логично обосновывая, отдает предпочтение задаче с наиболее распространенным дизайном анализа транскриптома, задаче по «определению эффективных групп генов для сравнения классов». Диссертант формулирует основную гипотезу, лежащую в основе этого типа анализа транскриптома, относительно того, насколько и какие флуктуации в уровнях экспрессии генов связаны с фенотипическими различиями исследуемых классов, а не вызваны случайными факторами и для ее решения справедливо обращается к методам математической статистики. По мнению диссертанта при наличии достаточно большой выборки можно на вероятностном уровне сделать выводы о свойствах исследуемых объектов, однако допущение нормальности распределения, независимости исследуемых переменных, зачастую малый размер выборки делают применение этих методов неадекватными и уменьшают эффективность исследования. Диссертант обсуждает также возможность применения других, альтернативных методов классической статистики, подходов для анализа экспрессии генов – «логико-комбинаторное машинное обучение». Досконально обсуждая ряд методов, их плюсы и минусы в отношении их применимости для анализа экспрессии генов, автор констатирует, что для анализа транскриптома необходимо учитывать его специфику: в нем множество объектов имеет минимальный или нормальный объем, однако число признаков слишком велико. По справедливому мнению диссертанта, именно в этом заключается причина того, что помимо методов математической статистики, машинного обучения необходимо разработать новые подходы и найти алгоритмические решения эффективного анализа данных экспрессии генов.

Второй раздел главы посвящен функциональному анализу экспрессии генов, которая определяется количеством синтезированных копий молекул РНК с матрицы гена в клетке в определенных условиях и в определенный момент времени. Трудно не согласиться с основным выводом диссертанта о том, что «для решения проблемы корректной идентификации дифференциально экспрессируемых генов механистические алгоритмы, основанные на математических и статистических

особенностях анализируемых данных несостоятельны». Диссертант справедливо отмечает, что «они должны быть дополнены алгоритмами, основанными на использовании накопленных знаний о биологической природе исследуемого состояния и о связи генов и их экспрессии с ним». Только в этом случае возможен истинный функциональный анализ экспрессии генов. В конце раздела автор приводит ряд литературных источников и публичных баз данных по изучаемой проблематике.

Третий раздел главы посвящен предложенному диссертантом алгоритму растущих опорных множеств, т.е. алгоритму определенного набора признаков, который предоставляет классификатор необходимой точности, тогда как под- и надмножества этого множества менее и не более точны, соответственно. В четвертом разделе диссертант описывает валидацию разработанного им алгоритма на примере использования набора данных по экспрессии генов при остром лимфобластном и остром миелобластном лейкозах. Примечательно, что предложенный автором алгоритм обладает определенными преимуществами, что было подтверждено функциональным анализом, который показал ассоциацию идентифицированных генов с развитием лейкемий. Примечательно также то, что список генов, полученных с помощью алгоритма автора, позволил получить информацию о биологических процессах, лежащих в основе этих заболеваний.

В третьей главе диссертационной работы, на 42 страницах описываются разработанные автором биоинформатические методы для моделирования активности биологических путей, т.е. серий взаимодействий между молекулами в клетке, приводящих к синтезу определенного продукта или к изменению функционального статуса клетки. Приводится информация о визуализации и редактировании биологических путей баз данных KEGG /Kyoto Encyclopedia of Genes and Genomes/pathways, представляющих собой набор карт регуляторных, сигнальных и метаболических путей. В главе описывается также разработка нескольких программных пакетов, в том числе: программного пакета KEGGParser для среды MATLAB и SciLab, программного пакета CyKEGGParser для среды Cytoskope, разработка алгоритма оценки сигнальных потоков PSW /Pathway Signal Flow/ и разработка приложения Pathway Signal Flow Calculator для Sytoscape. Разработка этих программных пакетов является важным результатом, так как впервые была внедрена возможность полуавтоматической коррекции взаимодействий между компонентами биологических путей, учитывающую тканевую специфичность. По-мнению автора, это существенно повышает точность результатов биоинформатического анализа, с чем нельзя не согласиться. Приведенные

выше разработанный диссертантом алгоритм и программные пакеты нацелены на переход от по-генного метода анализа дифференциальной экспрессии генов к анализу на уровне сигнальных путей. Очевидно, что это значительно улучшает точность определения биологических процессов, являясь существенным достижением методологической части диссертационной работы. Интересные результаты были получены диссертантом при исследовании сигнальных и метаболических путей и их устойчивости к мутациям. Оказалось, что хотя эти пути в общем устойчивы к мутациям благодаря множественным дублирующим ответвлениям, разветвленным топологиям, однако в них могут быть также «узкие» места, мутации в которых могут резко влиять на общую активность пути.

Четвертая глава диссертационной работы самая объемная /64 страницы/ и посвящена исследованию молекулярных механизмов патогенеза широкого спектра моногенных и полигенных заболеваний человека с помощью разработанных биоинформатических алгоритмов и программных пакетов, использованных для анализа доступных баз данных по глобальной экспрессии генов. Диссертантом были использованы наборы данных из хранилища Gene Expression Omnibus, содержащие профили экспрессии генов в различных тканях при псориазе, рассеянном склерозе, ишемическом инсульте. Были выбраны 24 биологических пути из базы данных KEGG, которые были вовлечены в развитие указанных заболеваний. Был проведен GSEA /Gene Set Enrichment Analysis/ с использованием классического алгоритма. Оценка активности биологических путей была осуществлена с использованием алгоритма PSF.

Разработанные диссертантом алгоритмы и биоинформатические пакеты были использованы им в исследованиях молекулярных механизмов развития целого ряда заболеваний человека. Среди них – острый лейкоз, хронические и онкологические болезни легких, аутоиммунные и аутовоспалительные нарушения, посттравматическое стрессовое расстройство.

Обобщая рецензирование этой интересной, оригинальной и объемистой диссертационной работы считаю необходимым в кратце акцентировать ее главные достижения как методологические, так и исследовательские. В методологическом аспекте: 1. Разработанные алгоритм и метод опорных множеств. Метод позволяет выявить группы генов, вовлеченные в патомеханизмы заболеваний и идентифицировать ключевые гены, нарушения экспрессии которых играют существенную роль в патогенезе заболеваний. 2. Разработанные программные пакеты KEGGParser (для вычислительной среды MATLAB) и CyKEGGParser (для платформы


визуализации и расчетов Cytoscape), которые позволяют создавать библиотеки биологических путей на основе информации, хранящейся в базе данных KEGG Pathway. 3. Разработанный алгоритм моделирования распространения сигнальных потоков в биологических путях PSF /«Pathway Signal Flow»/, позволяющий переход от по-генного к системному анализу, возможность использования программы как самостоятельно, так и в комбинации с другими методами, оценки изменений активности как на уровне пути в целом, так и в комбинации с другими методами и др. Главные исследовательские достижения заключаются в следующем: 1. Проведенные исследования позволили автору характеризовать основные причины развития рака легких и хронических заболеваний легких. Если в первом случае это активация биологических путей, связанных с пролиферацией клеток, то хронические заболевания легких сопровождаются изменениями, связанными с иммунным ответом и фиброзным ремоделированием легочной ткани. 2. При изучении патогенеза аутоиммунных и аутовоспалительных заболеваний диссертант заключает, что в основе патогенеза лежит смещение дисфункций иммунного и воспалительного ответов. Примечательно, что патогенез ряда аутоиммунных заболеваний характеризуется вовлечением биологических путей, связанных с аутовоспалением. 3. Примечательным является также вывод диссертанта о том, что при моногенных аутовоспалительных синдромах генетическая гетерогенность и различия в экспрессии генов не отражаются на профиле активации биологических путей, связанных с иммунным и воспалительным ответами. По мнению диссертанта, это может объяснить сходную клиническую картину при разных мутациях, затрагивающих один и тот же ген.

Таким образом, результаты диссертационной работы значительно расширяют наши знания о патогенезе ряда хронических, онкологических, аутоиммунных и воспалительных заболеваний. Вместе с тем, они убедительно выявляют преимущества использования биоинформатических подходов при анализе больших объемов экспериментальных данных, позволяющих всесторонне исследовать молекулярные механизмы функционирования биологических систем как в норме, так и в патологии.

По теме диссертации имеются 34 публикации, среди них 22 статьи в отечественных и в зарубежных авторитетных журналах, 2 обзорные главы в зарубежных изданиях и 8 тезисов конференций.

По работе нет существенных замечаний. Вместе с тем, мне кажется, что разработанные диссертантом программные пакеты корректнее называть «пакетами прикладных программ» («application softwares»), т.е. комплексы программ,

предназначенные для решения определенного класса задач, основанные на определенной вычислительной среде. Вопрос касается терминологии и, потому не принципиален. Автореферат диссертации в основном отражает содержание диссертационной работы, хотя в нем отсутствует раздел диссертации «Заключение», видимо из-за ограниченности допустимого объема автореферата. Имеются также некоторые упущения (скажем, в автореферате, в описании структуры работы нет названия главы 3, хотя название главы 2 дважды упоминается) и т.д. Естественно, что отмеченное никак не умаляет значимость диссертационной работы. Диссертационная работа «Разработка биоинформатических подходов для изучения молекулярных механизмов развития хронических заболеваний человека» актуальна, насыщена научной новизной, имеет важное прикладное значение, открывает новые перспективы исследований, а ее автор – Арсен Арташесович Аракелян, несомненно заслуживает искомой степени доктора биологических наук по специальности «Молекулярная и клеточная биология». Вместе с тем, уверен, что диссертационная работа могла быть успешно защищена также по специальности «Биоинформатика».

Официальный оппонент: 

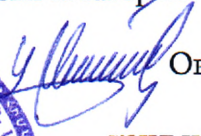
Геворкян Э.С.

док.биол.наук, чл.-корр. НАН РА, профессор

Подпись Геворкяна Эмиля Сосевича заверяю:

Ученый секретарь ЕГУ





Овсепян Л.С.

канд.ист.наук

29.11.2019.