

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ



Федеральное государственное бюджетное учреждение науки
Институт общей генетики им. Н.И. Вавилова
Российской академии наук
(ИОГен РАН)

ул. Губкина, д. 3, г. Москва, ГСП-1, 119991
Тел.: (499) 135-62-13, (499) 135-20-41
Факс: (499) 132-89-62

E-mail: iogen@vigg.ru
http: www.vigg.ru

№ 92

На №

УТВЕРЖДАЮ

Директор ФГБУН Института Общей Генетики
им. Н.И. Вавилова Российской Академии Наук,
доктор биологических наук,
А.М. Кудрявцев

«20» сентябрь

2018 г.



ОТЗЫВ

ведущей организации федерального государственного учреждения науки
института общей генетики им. Н. И. Вавилова российской академии наук
на диссертационную работу Сайк Ольги Владимировны на тему «Выявление
ассоциативных связей между заболеваниями и генетической регуляцией метаболических
процессов на основе автоматического анализа текстов и баз данных» представленную к
защите в диссертационном совете Д. 003.011.01 на базе ФГБУН «Федеральный
исследовательский центр институт цитологии и генетики сибирского отделения российской
академии наук» по специальности 03.01.09 – математическая биология, биоинформатика.

Актуальность исследования

Исследование метаболических нарушений в человеческом организме является
чрезвычайно важным для дальнейших исследований в области молекулярной медицины.
Поиски соответствий различных биомедицинских, генетических и молекулярно-
биологических термов, таких как «генотип-фенотип» или же «ген-транскрипционный

фактор» является довольно нетривиальной задачей. Кроме того, даже на появление таких алгоритмов, как word2vec, агрегация данных из разных баз данных представляет значительную сложность. Это связано как с неоднородностью состава этих баз, так и с огромными их объемами. Стоит отметить, что и отсутствие единой стандартизации для всех этих баз может вносить путаницу.

Научная новизна работы

Диссертантом впервые был произведен автоматический текстовый анализ огромного числа различных по природе описываемых явлений баз данных. На основе всей этой информации была создана БД Promedia. Подход к описанию агрегированных данных так же нов — был использован метод фреймов, взятый из теории реляционных баз данных и теории искусственного интеллекта, ранее не используемый при анализе биологических баз данных. С его помощью, между различными терминами устанавливается связь типа «объект-отношение-субъект». Используя этот метод, а также методы машинной обработки текстов, диссертант смогла соединить воедино большое количество баз данных, выстроить между ними ассоциативные связи и свести все в единую базу данных.

Стоит отметить также широкий спектр клинически значимых примеров, приведенных в работе. С помощью разработанного метода поиска ассоциаций, были обнаружены наиболее приоритетные при генотипировании гены при коморбидных состояниях астмы/гипертонии, выявлено 13 летучих метаболитов различных заболеваний, предложены новые гены-кандидаты, возможно связанные с различными социально-значимыми заболеваниями.

Теоретическая и практическая значимость исследования

Результаты, полученные диссертантом могут найти широкое применение в медицинской генетике и вычислительной молекулярной биологии, в таких практически важных задачах, как постановка диагнозов и поиск биомаркеров различных заболеваний.

Стоит также отметить, что результаты диссертационной работы уже используются в курсе лекций и практических занятий на кафедре информационной биологии факультета естественных наук Новосибирского государственного университета, что свидетельствует о большом практическом потенциале.

Структура диссертационной работы

Диссертационная работа состоит из оглавления, списка сокращений, введения, обзора литературы, главы с описанием материалов и методов, главы с описанием разработанной базы данных, трех глав, в которых описывается применение разработанной базы данных, заключения, выводов и списка литературы.

Текст диссертации изложен на 185 страницах печатного текста, включает 47 рисунков и 14 таблиц. Список литературы включает 435 источников.

Введение содержит обоснование актуальности темы, степень проработанности задач, определение целей исследования, способов решения задач. Так же описана научная новизна темы, теоретическая и практическая значимость, внедрение полученных результатов, основные положения, выносимые на защиту, апробация материалов, список статей и конференций диссертанта.

В обзоре литературы дается очень подробное описание различных методов диагностики, приводится определение «метода фреймов», дается очень полное и объемное описание источников данных, используемых для построения модели. Стоит отметить чрезвычайную проработанность и подробность этого раздела.

Во второй главе описываются методы, используемые для построения базы данных Promedia.

В третьей главе описывается база данных Promedia, структура её таблиц, а так же производится анализ связей термов, связанных с метаболитами. После их подсчета автор производит поиск ассоциаций между самыми распространенными термами метаболитов и термами заболеваний, подкрепляя каждую найденную ассоциацию ссылками на экспериментальные работы.

В четвертой главе, с помощью разработанной базы данных, производится поиск транзитивности ассоциации заболеваний с другими термами из БД Promedia, отбор термов, для которых гипотеза транзитивности выполняется. После производится их статистический анализ.

В пятой главе приводятся конкретные примеры поиска ассоциаций для набора социально-значимых заболеваний, таких как первичная открыто-угольная глаукома, лимфедема, презклампися, астма, гепатит С.

В шестой главе приводится пример, как разработанная база данных может быть использована при планировании молекулярно-биологического эксперимента, а также для поиска возможных биомаркеров.

В заключении подводятся итоги выполненного исследования, кратко описаны наиболее важные для практических приложений результаты.

По результатам проведенных исследований было сформулировано пять выводов, которые прямо вытекают из результатов выполненного исследования и соответствуют поставленным задачам.

Достоверность результатов исследования

При подготовке к защите диссертационной работы диссертантом было опубликовано большое количество статей в рецензируемых, что, вкупе с хорошо выполненной диссертационной работой, говорит об огромном объеме проведенных исследований и достоверности представленных результатов. При разработке методов было использовано 14 различных баз данных, каждая из которых дает достаточно материала для статистической обработки.

Все разработанные методы подкреплены достаточным числом тестов, результаты которых можно увидеть в соответствующих таблицах в тексте диссертации. Ассоциированность значимых генов, иллюстрируется ссылками на рецензируемые источники. Кроме того, был создан web-сайт, на котором любой желающий может удостовериться в корректности и достоверности полученных диссертантом результатов.

Автореферат диссертационной работы Сайк О.В. и её опубликованные работы полностью отражают содержание диссертации.

В работе, однако, имеются некоторые недостатки:

1. На рисунке 3.1 присутствует подробная схема базы данных Promedia, однако, она неполна, некоторые поля скрыты. Лучшим решением было бы вынести её в отдельное приложение, расположенное после основного текста диссертации.

2. Раздел «Материалы и методы» недостаточно подробный. В данном разделе хотелось бы увидеть более подробное описание модели, с помощью которой извлекалась текстовая информация из источников данных.

Перечисленные замечания, однако, не снижают научной ценности представленной диссертационной работы.

Рекомендации по использованию работы

Результаты работы Сайк Ольги Владимировны могут быть использованы в научно-исследовательских институтах медицинского профиля (Первый МГМУ им. Сеченова), занимающихся исследованием связи наборов фенотипических признаков и геномных полиморфизмов. Кроме того, работа будет интересна организациям, которые занимаются исследованиями в области вычислительной молекулярной биологии (Институт Общей Генетики им. Н.И. Вавилова РАН, Институт Молекулярной Биологии им. В.А. Энгельгардта РАН).

Заключение

Диссертация Сайк Ольги Владимировны «Выявление ассоциативных связей между заболеваниями и генетической регуляцией метаболических процессов на основе автоматического анализа текстов и баз данных» выполнена на высоком научном уровне, носит фундаментальный характер и имеет широкий спектр применения как в теоретических, так и в практических исследованиях, связанных с генетическими заболеваниями человека. Актуальность темы, новизна результатов, теоретическая и практическая значимость результатов работы полностью соответствуют критериям пп. 9-14 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г., №842, предъявляемого к диссертациям на соискание степени кандидата наук. Считаю, что автор диссертации — Сайк Ольга Владимировна заслуживает присуждения ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Отзыв ведущей организации рассмотрен и одобрен на заседании научного семинара лаборатории системной биологии и вычислительной генетики ФГБУН Института общей генетики им. Н.И. Вавилова Российской академии наук, 24 сентября 2018 г., протокол № 7

Данные об организации:

Федеральное государственное бюджетное учреждение науки Институт общей генетики им. Н.И. Вавилова Российской академии наук

Тел.: (499) 135-62-13, Факс: (499) 132-89-62

Почтовый адрес (официальный): 119991, ГСП-1 Москва, ул. Губкина, д. 3

Email: iogen@vigg.ru

Сайт: vigg.ru

Отзыв составил:

Научный сотрудник лаборатории системной биологии и вычислительной генетики Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук, кандидат физико-математических наук,
Л.А. Урошлев



Подпись к.ф.-м.н. Л.А. Урошлева удостоверяю:

Ученый секретарь ФГБУН ИОген РАН,

Доктор биологических наук

О.А. Огаркова



Сведения о составителе отзыва

Урошлев Леонид Андреевич, кандидат физико-математических наук, научный сотрудник лаборатории системной биологии и вычислительной генетики Федерального государственного бюджетного учреждения науки Института общей генетики им. Н.И. Вавилова Российской академии наук.

Адрес: 119991, ГСП-1 Москва, ул. Губкина, д. 3

Телефон: (499)132-89-64

E-mail: leonid.uroshlev@vigg.ru