

ОТЗЫВ

на автореферат диссертации Сайк Ольги Владимировны на тему «Выявление ассоциативных связей между заболеваниями и генетической регуляцией метаболических процессов на основе автоматического анализа текстов и баз данных», представленной на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика

Диссертационная работа посвящена системному анализу генетической регуляции метаболических процессов, ассоциированных с заболеваниями. В работе проведен поиск закономерностей связи ряда социально-значимых заболеваний с такими молекулярно-генетическими объектами как транскрипционные факторы, микроРНК, ферменты и метаболиты. Предложен подход, основанный на анализе ассоциативных генных сетей, который позволяет предсказывать новые маркеры заболеваний и выявлять гены-кандидаты для дальнейшего генетического тестирования. Для построения ассоциативных генных сетей, описывающих механизмы развития патологий, были использованы методы автоматического извлечения знаний из научных текстов системы PubMed. Следует отметить, что такой масштабный поиск связей между генетической регуляцией метаболических процессов и заболеваниями является актуальной задачей современной биомедицины.

Автором создана база данных фреймовых моделей Promedia, в которой интегрирована информация по связи генетической регуляции метаболических процессов с более чем 16 тысячами заболеваний. В работе приведены примеры использования разработанной базы данных для поиска новых маркеров таких заболеваний как болезнь Паркинсона, гепатит С, ревматоидный артрит и другие. Исследованы молекулярно-генетические особенности коморбидности астмы и гипертонии и предложены гены *IL10*, *TLR4* и *CAT* в качестве кандидатов для генотипирования.

Следует указать следующие недостатки в тексте автореферата. В разделе «2.2. Анализ транзитивных ассоциаций...» при описании алгоритма построения псевдослучайных фреймовых моделей не освещён ряд вопросов, препятствующих полному пониманию процесса. А именно – судя по описанию, количество заболеваний, общих для псевдослучайных наборов в точности подчиняется гипергеометрическому распределению, а значит нет смысла строить случайные наборы болезней. Во-вторых, некоторые болезни могут и встречаются гораздо чаще других болезней в базах данных благодаря большей степени изученности и популярности. При статистических расчётах следует делать на это поправку. Учесть данный баес можно, например, применяя не случайный выбор болезней из списка, где все болезни встречаются ровно один раз, а формировать псевдослучайные списки болезней путём случайного перемешивания болезней между изначальными моделями. В-третьих, ничего не упоминается на стр. 6 о том, проводилась ли коррекция на множественное тестирование. В качестве стилистических замечаний можно указать на частое повторение одних и тех же предложений в тексте автореферата. Также вызывает дискомфорт отсутствие пояснения термина «фреймовая» в отношении моделей. Данные замечания не являются существенными с точки зрения оценки качества и значимости работы, не затрагивают основные выводы и положения, выносимые на защиту.

Диссертационная работа Сайк О.В. является законченным научным исследованием. Автореферат диссертации дает адекватное представление о проведенной работе. Результаты проведенных исследований опубликованы в рецензируемых журналах, а также были представлены на различных научных конференциях. Диссертационная работа Сайк О.В. соответствует требованиям ВАК РФ, предъявляемым к кандидатским диссертациям по

специальности 03.01.09 – математическая биология, биоинформатика, а ее автор заслуживает присуждения степени кандидата биологических наук.

Старший научный сотрудник,
ФГБУН Институт систем информатики
им. А.П. Ершова СО РАН (ИСИ СО РАН)
кандидат физ.-мат. наук

Штокало Дмитрий Николаевич

01.10.2018

