

КОМПЬЮТЕРНЫЙ АНАЛИЗ И ОБРАБОТКА ДАННЫХ ЭКСПРЕССИИ ГЕНОВ

Спицина А.М., Орлов Ю.Л.¹

Новосибирский государственный университет, Россия, 630090, г. Новосибирск,
ул. Пирогова, д. 2, +7(952)9200699, anastasia.spitsina@gmail.com

¹Институт Цитологии и Генетики СО РАН, Россия, 630090, г. Новосибирск,
пр. ак. Лаврентьева, д. 10, orlov@bionet.nsc.ru

Данные об уровнях экспрессии генов, полученные с помощью ДНК-микрочипов, находят новое применение в области биоинформатики. Эти данные представляют большое значение, как со стороны медицины, так и статистики – их обработка и анализ помогут ответить на вопросы, связанные с проблемами поведения, образования и развития различных заболеваний, в том числе онкологических и нейродегенеративных. С развитием технологии накапливается объем данных, не поддающийся ручной обработке. Необходимо создание программных комплексов для анализа и визуализации данных. Анализ включает несколько стадий – сопоставление численных и текстовых данных, фильтрация, статистический анализ и визуализация. Использовалась серия микрочипов Affymetrix U133, данные по геному человека и модельным геномам мыши и крысы. Разработан программный комплекс, который включает в себя различные опции для работы с большим массивом микрочиповых данных:

- сопоставление данных экспрессии генов из нескольких баз;
- фильтрация данных: удаление проб-дублей; удаление изоформ гена;
- расчет матрицы корреляций по группе генов (заданная или случайная выборка);
- построение графика тканеспецифичности;
- визуализация генных сетей на основе корреляции генов;
- поиск и анализ расположения генов относительно петель сайтов связывания транскрипционных факторов.

Программа написана на языке C++ в среде программирования Microsoft Visual Studio 2010 с использованием скрипта на языке JavaScript. Взаимодействие с пользователем происходит посредством текстового меню. Построены выборки генов с повышенной экспрессией в тканях мозга, построены профили тканеспецифичности этих выборок. С помощью данной программы были проанализированы уровни экспрессии и корреляции экспрессии генов в составе генных сетей.

Литература.

1. Спицина А.М., Орлов Ю.Л., Подколотная Н.Н., Свичкарев А.В., Дергилев А.И., Чен М., Кучин Н.В., Черных И.Г., Глинский Б.М. Суперкомпьютерный анализ геномных и транскриптомных данных, полученных с помощью технологий высокопроизводительного секвенирования ДНК // Программные системы: теория и приложения. 2015, **Том 6**, номер 1(23), с. 157–174.