

GEOParser – WEB-ресурс для анализа экспериментов NCBI GEO

Научный руководитель – Мустафаев Орхан Нариман

Huseynova Ulkar Elman

Студент (магистр)

Национальная академия наук Азербайджана, Биология, Баку, Азербайджан

E-mail: ulker.huseynova97@gmail.com

GEOParser - WEB-ресурс для анализа экспериментов NCBI GEO

Гусейнова У.Э.

Магистр, 2 курс

Институт генетических ресурсов,

Национальная академия наук Азербайджана,

Баку, Азербайджан

E-mail: ulker.huseynova97@gmail.com

Gene Expression Omnibus (GEO) - это обширный ресурс, который архивирует и свободно распространяет миллионы экспериментов по изучению экспрессии генов, метилирования и других аспектов функциональной геномики, поставленных на модельных организмах. GEO содержит необработанные, обработанные и описательные метаданные. Информация о каждом эксперименте и использованных образцах (метаданные) представлена в виде нестандартного текстового описания, а результаты экспериментов - в виде отдельных текстовых файлов для каждого образца. Стоит отметить, что информация о проведенном эксперименте, включая использованные платформы (протоколы) и описания образцов, также может быть получена в различных форматах, таких как MINIML или SOFT. Однако для анализа содержащейся в них информации требуется ее предварительная обработка автоматизированными программными средствами. Для решения этой задачи был разработан ряд программных пакетов, таких как GEO2ENRICH (https://amp.pharm.mssm.edu/g2e/), ImaGEO (http://bioinfo.genyo.es/imageo/)[1], BART (http://igc1.salk.edu:3838/bart/)[2], однако они имеют ряд недостатков: GEOQUERY и BART требуют навыков в работе в программной среде R, а ImaGEO разработан для проведения сравнительного анализа нескольких экспериментов, в результате чего нет возможности просмотреть данные одного эксперимента.

В этой связи нами был разработан WEB-ресурс, позволяющий пользователю проводить базовый анализ результатов экспериментов, размещенных на сайте GEO. В отличие от поиска информации непосредственно в GEO, где пользователю необходимо скачивать и обрабатывать файлы большого объема, разработанная нами система позволяет обходить сложности со скачиванием и сопоставлением данных содержащихся в файлах. При введении искомой серии эксперимента система выводит исследователю список использованных в нем платформ и образцов с возможностью выбора необходимой информации. На следующем этапе пользователю предоставляется список полей, относящихся к образцам - например, идентификатор локуса или уровень экспрессии. Полученные результаты могут быть проанализированы непосредственно в окне браузера: для этого предусмотрены сортировка, фильтрация и графическое представление результатов, выявление дифференциально экспрессируемых локусов (в случае экспрессии генов), а также экспортированы с целью их дальнейшего анализа в других программных пакетах.

Разработанная система будет доступна по адресу <https://geo.bioset.org/>.

Источники и литература

- 1) D. Toro-Domínguez et al., “ImaGEO: Integrative gene expression meta-analysis from GEO database // Bioinformatics. 2019 vol. 35, no. 5, pp. 880–882.
- 2) M. L. Amaral, G. A. Erikson, and M. N. Shokhirev, “BART: Bioinformatics array research tool // BMC Bioinformatics. 2018, vol. 19, no. 1, pp. 1–6.