

Определение генетических рисков развития шизофрении у армянской и еврейской популяции с использованием методов машинного обучения

Научный руководитель – Аракелян Арсен Арташесович

Давитавян С.С.¹, Никогосян М.А.²

1 - Российско-Армянский (Славянский) университет, Институт математики и высоких технологий, Кафедра биоинженерии, биоинформатики и молекулярной биологии, Ереван, Армения, *E-mail: surendavitavyan98@gmail.com*; 2 - Российско-Армянский (Славянский) университет, Институт математики и высоких технологий, Кафедра медицинской биохимии и биотехнологии, Ереван, Армения, *E-mail: marianikoghosyan@gmail.com*

Новые высокопроизводительные технологии геномики открыли новую эру в анализе «больших биологических данных». Анализ больших данных способствует развитию персонализированной медицины, следовательно, есть необходимость в новых подходах к анализу накопленных данных. Алгоритмы машинного обучения и нейронных сетей доказали свою способность эффективно анализировать большие данные. В настоящее время наиболее частыми причинами смерти в мире являются комплексные заболевания, такие как метаболические и сердечные заболевания, нейродегенерация и др. .

Комплексные заболевания являются следствием сочетания генетических факторов риска и окружающей среды [1]. Наиболее распространенным подходом к идентификации генетических маркеров, например однонуклеотидных полиморфизмов (ОНП), которые ассоциированы с комплексными заболеваниями, являются полногеномные исследования ассоциации (GWAS). Однако исследования GWAS имеют серьезное ограничение. ОНП, ассоциированные с заболеваниями в одной популяции, могут иметь разные эффекты для других [2]. Генетический риск, связанный с одним и тем же заболеванием, значительно различается среди людей и групп населения. Следовательно, существует необходимость в разработке новых подходов к интеллектуальному анализу данных, которые будут способствовать эффективной диагностике заболеваний, прогнозированию и мониторингу их течения.

В данном исследовании определили популяционные генетические маркеры, связанные с шизофренией, и сравнили их распределение у армян и евреев-Ашкенази, а также получили портреты распределения генетического риска заболевания на уровне популяции. Данные ОНП пациентов из еврейского населения ашкенази и здоровых людей были получены из базы данных Gene Expression Omnibus [3], данные об армянских образцах были получены в результате полноэкзомного секвенирования. Мы использовали алгоритм самоорганизующейся карты (SOM) [4] для кластеризации и уменьшения размерности. Такой подход позволил выделить наборы ОНП, ассоциированных с шизофренией, для обеих популяций, а с другой стороны, дал нам информацию о распределении популяционно-специфических ОНП для армянской и еврейской выборки.

Источники и литература

- 1) Craig J (2008) Complex diseases: Research and applications// Nature Education 1.1: 184.
- 2) Norrgard K (2008) Genetic variation and disease: GWAS// Nature Education 1.1: 87.
- 3) <https://www.ncbi.nlm.nih.gov/geo/> (Gene Expression Omnibus)
- 4) Löffler-Wirth H, Kalcher M, Binder H. (2015) oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor// Bioinformatics 31.19: 3225- 3227.