

Прогнозирование риска развития шизофрении и биполярного расстройства в разных популяциях с использованием подходов 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». Эта информация способствует прогнозированию и мониторингу заболеваний, что является важной частью персонализированной медицины. 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]. Следовательно, существует потребность в новом подходе к оценке генетического риска развития болезни, который будет скорректирован в зависимости от принадлежности исследуемых людей к определенной популяции.

В данном исследовании мы определили популяционно специфические генетические варианты, ассоциированные с комплексными заболеваниями и создали портреты генетического риска заболеваний на уровне населения, с использованием метода машинного обучения самоорганизующихся карт (Self-Organizing Maps (SOM)) [3]. Для этого мы использовали данные генотипов для 51 популяций, полученных в рамках проекта Human Genome Diversity Project (HGDP) [4]. На основе полученных портретов, мы оценили риск развития шизофрении и биполярного расстройства в популяциях Среднего Востока, которые были адаптированы в зависимости от наличия популяционно специфических маркеров. Для этого мы использовали метод опорных векторов (support vector machines) [5].

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. 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.
4. Cavalli-Sforza LL (2005) The human genome diversity project: past, present and future// Nature Reviews Genetics 6.4: 333-340.

5. Кесман В (2005) Basics of machine learning by support vector machines// Real World Applications of Computational Intelligence, Springer, Berlin, Heidelberg, pp 49-103.