**Вариант №4: исследование и оптимизация метода универсальной гармонизации генно-экспрессионных профилей.**

В биоинформатике актуальной является проблема гармонизации и сравнения генно-экспрессионных профилей, поученных на разных установках, использующих методы высокопроизводительного секвенирования мРНК и микрочиповой гибридизации мРНК.

Недавно нами был предложен метод Shambhala (Borisov et al., 2019), опирающийся на более ранний метод XPN (Shabalin et al., 2008).

Метод Shambhala зависит от двух массивов профилей – промежуточного (P) и дефинитивного (Q).

В качестве критерия качества гармонизации будем использовать две метрики – 1) watermelon multisection metric, описанную в работе (Zolotovskaia et al., 2020): WM-метрика кластерной дендрограммы для деления по типу биологических образцов должна быть больше, чем WM-метрика кластерной дендрограммы для деления по типу экспериментальной платформы; 2) метод анализа однородности на основе теста Стьюдента (Aliev and Borisov, 2018): однородной внутри одинаковых типов образцов должна быть больше, чем внутри одинаковой платформы после гармонизации.

Исследование провести для семи типов промежуточных массивов профилей и двух типов дефинитивных массивов, попарно для каждого типа экспериментальных установок (задача 1), а также для всего массива профилей вместе (задача 2).

**Литература**

Aliev, R. O., and Borisov, N. M. (2018). The Method for Analysis of Expression Data Homogeneity Based On the Student Test. *Math. Biol. Bioinforma.* 13, 50–67. doi:10.17537/2018.13.50.

Borisov, N., Shabalina, I., Tkachev, V., Sorokin, M., Garazha, A., Pulin, A., et al. (2019). Shambhala: a platform-agnostic data harmonizer for gene expression data. *BMC Bioinformatics* 20, 66. doi:10.1186/s12859-019-2641-8.

Shabalin, A. A., Tjelmeland, H., Fan, C., Perou, C. M., and Nobel, A. B. (2008). Merging two gene-expression studies via cross-platform normalization. *Bioinformatics* 24, 1154–1160. doi:10.1093/bioinformatics/btn083.