

## Тестовое задание по биоинформатике №4

### Подготовка

С сайта [www.cbioportal.org](http://www.cbioportal.org) скачать данные по когорте больных меланомой — нажать кнопку download. Когорта находится по пути Skin → Melanoma → Cutaneous Melanoma → Skin Cutaneous Melanoma (TCGA, Firehouse Legacy). В архиве содержится информация об экспериментах и аннотации, например:

- RNAseq dataRNA\_Seq\_v2\_expression\_median.txt),
- клинические аннотации
  - data\_bcr\_clinical\_data\_sample.txt – по образцам,
  - data\_bcr\_clinical\_data\_patient.txt – по пациентам.

Внимание! У пациента ID длиной в 12 символов, у образца – 16 символов.

### Задание №1

Воспроизвести классификацию на 3 экспрессионных типа: MITF-low, Keratin, Immune из статьи [Akbari et al. 2015](#) (doi:10.1016/j.cell.2015.05.044), описывающей этот датасет.

На момент написания статьи было доступно ~330 образцов. К какому классу они относятся можно найти в аннотации из статьи — первый supplementary файл, таблица S1D (лист в excel файле, колонка RNASEQ-CLUSTER\_CONSENSIER). Можно использовать эту классификацию, а остальных пациентов доклассифицировать или проделать процедуру из статьи заново.

### Задание №2

Воспроизвести классификацию, описанную в статье [Jönsson et al. 2010](#), на 4 класса: Pigmentation, Proliferative, Normal-like, High-immune response. В supplementary data к этой статье на последней табличке есть список генов, по которым это можно сделать.

Важно учесть, что платформы экспериментов разные — данные второго эксперимента получены при помощи микрочипов, поэтому нельзя переносить абсолютные значения центроидов на NGS данные, то есть на меланомную когорту из пункта «Подготовка».

## Задание №3

Сравнить полученные классификации. Какая из классификаций лучше разделяет пациентов по группам риска? Как вы думаете почему? Какая из классификаций лучше отражает биологию опухолей? Почему? Какую из классификаций вы бы взяли для работы и почему?

Описать полученный результат. Получилось ли у вас воспроизвести результаты авторов? Почему?

## Комментарий

Работа будет оцениваться по трём пунктам — техническая реализация, биологическое обоснование, ясность донесения результата.

Не обязательно делать каждый пункт идеально, но обязательно пояснить каждый шаг. Блок кода можно пояснить небольшим комментарием, а промежуточный результат — картинкой.

Работу нужно выполнить в Jupyter Notebook на языке Python или R, сохранить и прислать в форматах .html и .ipynb.