

## Задание 6

Выполнить комплексный анализ биологического датасета, включая выравнивание, оценку качества, дифференциальную экспрессию генов и функциональное аннотирование. Все выбранные инструменты и методы должны быть обоснованы.

### Цель анализа

Определить признаки, разделяющие критических и некритических пациентов из датасета GSE172114.

### Основные задачи

1. Выполнить выравнивание сырых fastq файлов к референсному геному на основе GENCODE v29. Выбор метода выравнивания не принципиален, допускается использование псевдовыравнивания.
2. Провести базовый анализ метрик качества данных (QC). На этом этапе необходимо оценить качество данных, чтобы удостовериться в их пригодности для последующего анализа.  
*Если по технической или иной причине выполнение 1 и 2 задачи невозможно, то мы можем предоставить обработанные данные по запросу. В этом случае необходимо подготовить один файл или полностью расписать пайплан.*
3. Провести анализ дифференциальной экспрессии генов и построить вулкан-плоты. Вулкан-плоты помогут визуализировать результаты этого анализа, показывая значимость и степень изменения экспрессии генов.
4. Провести анализ обогащения с использованием баз mSigDb. Этот шаг включает в себя идентификацию и интерпретацию биологических путей и процессов, которые наиболее существенно представлены в наборе дифференциально экспрессируемых генов.

### Дополнительные задачи

5. Выполнить типирование HLA II типа. Типирование HLA II типа необходимо для изучения генетического разнообразия в области генов, ответственных за иммунные реакции.
6. Провести деконволюцию для оценки клеточного состава образцов. Это позволяет получить более детальную картину клеточного состава и понять, какие типы клеток могут быть связаны с наблюдаемыми биологическими изменениями.
7. Выполнить анализ взвешенных коррелятивных сетей (WGCNA) для выявления модулей коэкспрессии генов. WGCNA позволяет идентифицировать группы генов с похожими паттернами экспрессии, которые могут быть функционально связаны или участвуют в общих биологических процессах.

### Критерии оценки

- Грамотный выбор и использование методов. Баланс между трудозатратами и результатом.
- Гибкость пайплайна. Удобное добавление дополнительных тулов или замена датасета.
- Наличие значимых результатов анализа не являться обязательным условием.
- Вывод должен содержать четкие положения по каждой выполненной задаче.

### Требования к оформлению

- Оформите весь пайплайн в Jupyter Notebook.
- Отдавайте предпочтение библиотекам на Python.
- Включите весь код запуска используемых инструментов и методов.
- Обоснуйте выбор каждого инструмента и метода на каждом этапе анализа.