

“Анализ данных single cell секвенирования образцов костного мозга для изучения его состава и микроокружения и поиска прогностически значимых маркеров для пациентов с лейкозом”

Анализ единичных клеток - мощный инструмент, позволяющий "смотреть" на свойства клеток по отдельности, не усредняя результат (как например в обычном, так называемом bulk РНК секвенировании). Такой подход очень важен в эмбриологии и в изучении дифференцировки стволовых клеток. Для задач онкологии метод single cell секвенирования также представляет большой интерес. Выживаемость пациентов во многом зависит от опухолевого микроокружения, а исследовать клеточные подтипы очень удобно с помощью анализа единичных клеток, что вам и предстоит сделать. В предложенной задаче у вас будет возможность опробовать данный инструмент для изучения микроокружения костного мозга и поиска новых маркеров, которые имеют прогностическую значимость для пациентов с лейкозом.

В данной задаче первым шагом вам нужно будет проанализировать single cell датасет с образцами костного мозга. Если вы никогда не работали (или мало работали) с single cell данными, то полезным может быть для начала посмотреть запись [семинара](#) от BostonGene (вам будут актуальны первые ~40 минут, где рассказывается про решение похожей задачи). Для ознакомления с single cell в общем можно посмотреть статью **“The triumphs and limitations of computational methods for scRNA-seq”**, которая также лежит на гугл диске. Также вам в помощь дан ноутбук ([Bio\\_tech\\_sc.ipynb](#)), который пересекается с семинаром, но содержит некоторые дополнения. В анализе вы можете использовать любые необходимые на ваш взгляд приемы из ноутбука/семинара, или же можете поискать что-то еще. Целью анализа является воспроизведение разметки предложенного датасета по клеточным типам (типичные клеточные типы костного мозга указаны на рисунке 1, но если возникнет желание, то можете попробовать типировать датасет глубже)

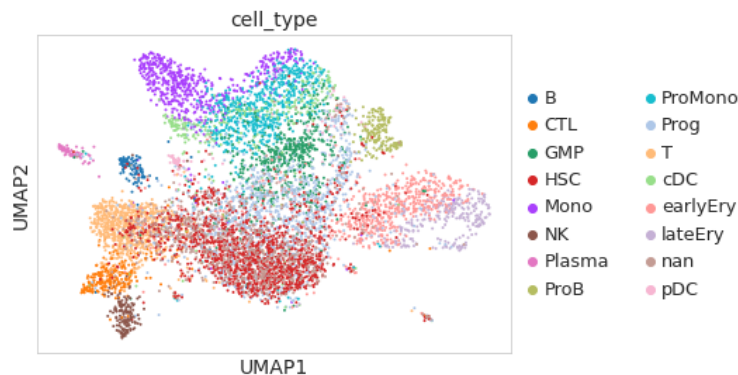


Рисунок 1. Разметка датасета по клеточным типам

и поиск генов, наилучшим образом (наиболее специфично) описывающих каждый из клеточных типов. Датасет single cell, состоящий из 6 образцов, лежит в папке **sc\_data**.

Дальше можно переходить к следующей части задания, в которой нужно будет работать с RNA-seq датасетом из образцов костного мозга пациентов с лейкозом (**expressions\_leukemia.tsv**). После выполнения первой части задания для каждого клеточного типа у вас есть список генов, наиболее специфично его описывающий. По этим генам для образцов пациентов с лейкозом можно косвенно судить о количестве данного типа клеток (хотя бы качественно, не обязательно искать точные проценты клеток) в образце. (Ваша задача в том числе придумать, как именно это лучше всего сделать).

Финальной задачей будет проанализировать, как именно связано количество тех или иных типов клеток в образцах пациентов с лейкозией с их выживаемостью, которая дана в аннотации к образцам пациентов (**annotation\_leukemia.tsv**) в графе OS (overall survival) в днях. (Рекомендуется обратить внимание на кривые Каплана-Мейера)

P. S. Дополнительно вам предоставлены данные RNA-seq нескольких здоровых пациентов (**expressions\_healthy.tsv**)

Куратор - Элеонора Белых  
Биоинформатик BostonGene,  
студент магистратуры ФБМФ  
МФТИ



Контакты:  
Telegram: @miptbioexorcist  
Vk: @white1201