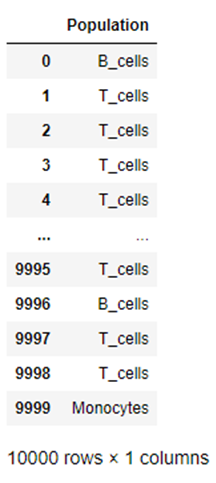
**Тестовое задание в группу анализа данных цитометрии**

В этом задании вам предстоит обработать небольшой фрагмент результата эксперимента многоцветной проточной цитометрии.

В вашем распоряжении данные по измерениям интенсивности флюоресценции ряда маркеров на ~ 34.000 клетках. Необходимо определить, какие из клеток в этой смеси являются моноцитами, T- и B- клетками.



**Входные данные:**

На вход дана таблица в формате tsv. В ней содержится ~ 34.000 строк. Каждая строка соответствует одной клетке, а каждая колонка определенному маркеру. В ячейках содержатся значения интенсивности флуоресценции маркера на данной клетке. По тому, какую интенсивность флуоресценции имеют маркеры в файле на данной клетке, можно определить, к какому типу относится клетка.

**Результат:**

*В качестве ответа на задание пришлите jupyter-ноутбук с решением и таблицу в формате tsv следующего вида:*

Таблица должна содержать одну колонку “Population”, в которой для каждого индекса (клетки) из входной таблицы прописан клеточный тип, к которому эта клетка относится. Если в файле оказались клетки, которые не являются моноцитами, T- и B- клетками, назвать их можно “Other\_cells”.

Важно: индексация в этой таблице должна полностью совпадать с индексацией в исходной таблице с интенсивностями.

Пример итоговой таблицы приведен слева.

Решение **необходимо сопроводить** подробным текстовым документом с описанием хода ваших действий.

**Ход решения**

Для успешного решения задания необходимо выполнить следующие шаги:

1) Построить по входным данным tSNE для отображения клеток в двумерном пространстве. Для каждой из 10000 клеток вам необходимо получить двумерные координаты.

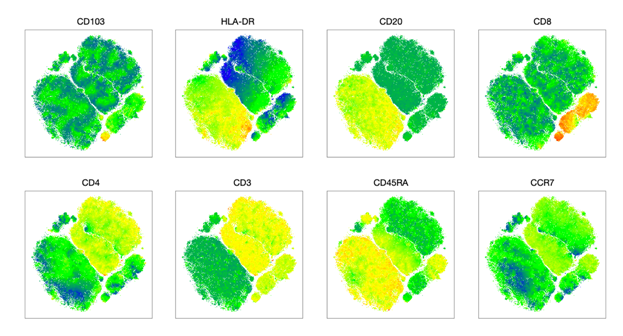
2) Сделать кластеризацию клеток. Выбор алгоритма кластеризации остается за вами. В данном случае для решения подойдут и самые простые алгоритмы, однако применение пакетов кластеризации для данных цитометрии будет высоко оценено. Необходимо для каждой клетки получить номер кластера, к которому она относится. Число кластеров нужно подбирать исходя из данных.

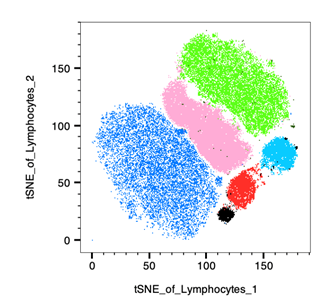
3) Для присвоения клетке определенного типа нужно построить два графика по координатам tSNE:

I. Постройте scatter-plot по двумерным координатам tSNE и покрасьте точки по признаку принадлежности к кластеру. Вы должны увидеть, как отдельные кластера точек, удаленные в пространстве друг от друга, окрасились разными цветами.

II. Постройте несколько scatter-plotов по числу маркеров (колонок) в исходной таблице. Каждый из этих графиков необходимо покрасить в соответствие с интенсивностью маркера в таблице. Вы увидите, что некоторые кластера яркие по одному маркеру и тусклые по другому маркеру.

Полученные графики с окраской по маркерам могут выглядеть как на картинке ниже (картинки из обучающего текста по ссылке -<https://marissafahlberg.com/a-basic-overview-of-using-t-sne-to-analyze-flow-cytometry-data/>, поэтому маркеры отличаются и цветовая шкала может быть подобрана лучше).



4) Далее нужно разобраться, клетки какого типа находятся в каждом из кластеров. Предполагается, что клетки одного типа похожи друг на друга, и, следовательно, попадут в один кластер. Поэтому на данном этапе задача сводится к присвоению названия клеточного типа кластеру.

tSNE с окраской по кластерам будет похож на картинку слева. Для того, чтобы присвоить название кластеру, нужно посмотреть какого цвета этот кластер по интересующим маркерам.

К примеру голубой, красный и черный кластера, судя по картинке из примера выше, яркие по CD8 и по CD3, поэтому могут считаться CD8+ T-клетками. А зеленый и розовый кластера яркие по CD4 и CD3, поэтому могут называться CD4+ T-клетками.

5) Разобраться с тем, по каким маркерам можно отличить T-клетки, B-клетки и моноциты вам необходимо самостоятельно, используя Pubmed, Google Scholar, Википедию и другие источники.

6) После этого необходимо сопоставить кластерам названия и сгенерировать итоговую таблицу в формате, описанном в разделе “Результаты” выше. В ноутбуке с решением также посчитайте сколько клеток какого типа удалось найти в файле.

7) Нанесите на tSNE полученные результаты (покрасьте по принципу принадлежности к клеточному типу).

При решении задания можно пользоваться любыми материалами, а также любыми методами и библиотеками языка ***python.***