Тестовое задание BostonGene

Во вложении test\_image\_data.zip в папке tiled\_data приведены изображения, полученные при съемке миндалин человека с помощью методики мультиплексной иммунофлюоресценции. Особенностью этого типа данных является довольно плотное расположение клеток.

Изображения сняты с перекрытием в 50% между каждым отснятым квадратом (назовём их тайлами), сторона тайла - 1024 пикселя (итого сетка 4 на 4 тайлов). Размерность снятого фрагмента ткани - квадрат со стороной 1548 пикселей. Папка tiled\_data имеет подпапки, названные по имени маркеров, для детекции которых происходило окрашивание: DAPI - ядерный маркер, Ki67 - маркер делящихся популяций клеток, CD3e, CD4, CD8 - маркеры T-клеточных популяций, экспрессирующиеся на мембранах клеток, первый экспрессируется на всех Т-клетках в данном примере, два последних - маркеры субпопуляций T-клеток.

В каждой из подпапок расположены тайлы в формате TIFF, индексация ведется с 1, тайлы были сняты последовательно строка за строкой, слева направо.

Задания:

1. Необходимо провести сшивку (stitching) имеющихся тайлов для каждого из маркеров, получив таким образом полные изображения;
2. Необходимо провести сегментацию отдельных клеток на данном фрагменте. Выбор технологии сегментации клеток свободный, можно использовать как трешхолдирование с вотершедом, так и иные методы на усмотрение. Protip: имеется набор мембранных маркеров, они могут помочь в процессе сегментации.
3. На сшитом стеке видно, что не вся площадь отснятого слайда покрыта клетками, необходимо провести обрезку интересующего ROI автоматически.
4. Полученные клетки можно охарактеризовать по среднему значению экспрессии в пределах ее сегмента. Нужно рассчитать средние значения экспрессии в сегментах и провести кластеризацию клеток по их видам. Затем вывести описательные статистики для набора данных: процентный состав кластеров, средняя величина интенсивности экспрессии маркеров в кластере, частоты контактов разных типов клеток в кластере. Набор статистик можно расширить по желанию, это будет являться плюсом при выполнении задачи.

Прим.:  Решение ожидается в виде архива репозитория git. При выполнении плюсом будет являться оформление в виде пайлпайна (прим. используя Snakemake или Nextflow, либо любой другой workflow engine на усмотрение), выполнение которого позволяет получить все результаты работы воспроизводимым образом. Для работы можно использовать ImageJ и скрипты на нем. Ход работы и расчет различных статистик для набора данных можно описывать в Jupyter Notebook или RMarkdown, при желании, можно оформить и в виде docx документа, в зависимости от удобства. При выполнении работы не обязательно делать каждый пункт “идеально”, главное последовательно описать каждый этап и почему были приняты те или иные решения при работе. Для невыполненных задач желательно объяснить причины и предположения, как их можно было бы решить при наличии большего времени.