

# Пакет Scater из Bioconductor

Попов Иван

11.29.2022

# Введение

**scater** - пакет из Bioconductor для визуализации транскриптимных данных уровня единичных клеток

Пакет основан на S4 классе `SingleCellExperiment`, представляющий собой геномные данные набора клеток в виде матрицы, у которой элемент  $a_{ij}$  хранит значение  $i$ -ой фичи  $j$ -ой клетки

# Графики для контроля качества

**Контроль качества** (quality control) в целях удаления поврежденных клеток и плохо отсековированных библиотек - типичная часть пайплана анализа клеточных данных.

**scater** предоставляет функции, позволяющие легко построить графики для такого контроля качества. Чтобы посчитать различные метрики контоля качества (чтобы потом построить графики по ним), можно воспользоваться пакетом **scuttle**, автоматически загружаемым при загрузке **scater**

# Графики для контроля качества

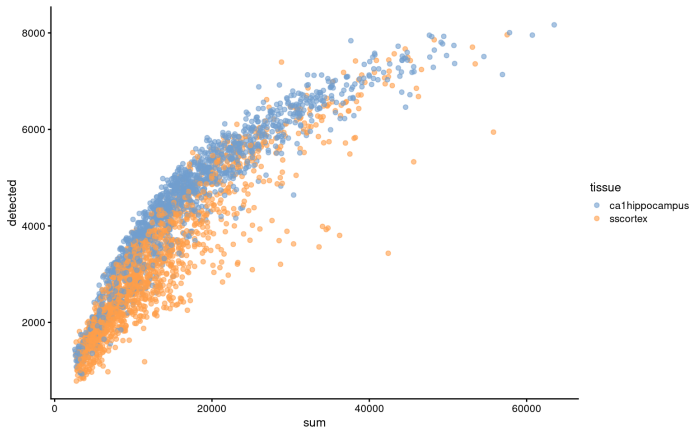
С помощью `plotColData` можно построить график по метаинформации для каждой клетки, которая вычисляется через `addPerCellQC`.

```
1 library(scater)
2 example <- addPerCellQC(example,
3   subsets=list(Mito=grep("mt-", rownames(example))))
4 plotColData(example, x = "sum",
5   y="detected", colour_by="tissue")
```

**Listing:** Вычисление quality control метаданных и построение по ним графика

## Графики для контроля качества

Мы ожидаем, что с ростом кол-ва клеток будет расти и кол-во распознанных в них генов - в данном случае данные соотносятся с нашими предположениями



**Figure:** Кол-во распознанных генов в зависимости от кол-ва клеток какого-то типа

# Визуализация данных об экспрессии

Пакет также содержит функцию `plotExpression`, позволяющую строить графики для значений экспрессии каких-то генов или признаков в зависимости от каких-то значений метаданных или от других генов (для построения профиля экспрессии среди клеток)

# Визуализация данных об экспрессии

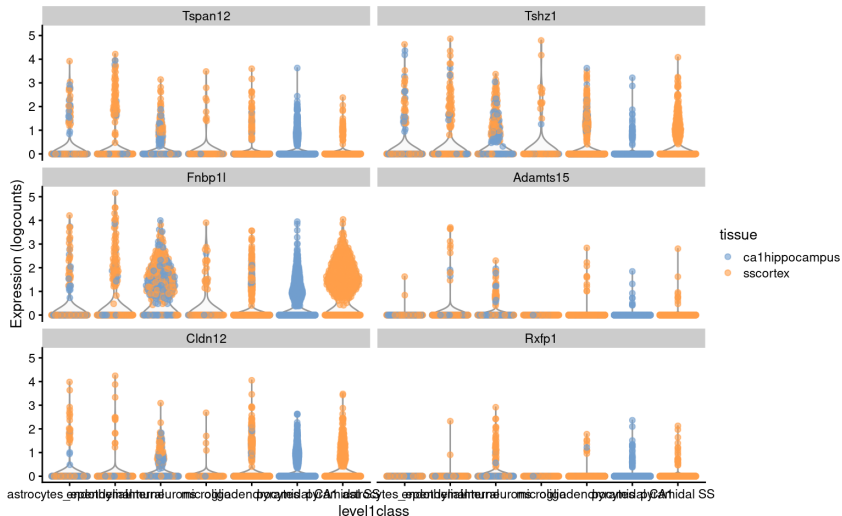
Параметр `x` указывает независимые переменные(covariates), по которым строится график - т.е. значения на оси X.

При выборе непрерывной переменной будет строиться scatter-график, а при категориальной - т.н. **violin plot**(приближают дискретную функцию плотности непрерывной)

```
1 plotExpression(example, rownames(example)[1:6],  
2               x = "level1class", colour_by="tissue")
```

**Listing:** Экспрессия 6 первых генов от категориальной переменной level1class

# Визуализация данных об экспрессии



**Figure:** Экспрессия 6 первых генов от категориальной переменной level1class



# Понижение размерности и визуализация

scatter также позволяет выполнять понижение размерности алгоритмами PCA и t-SNE и визуализировать результаты.

```
1 example <- runPCA(example ,  
2     name="PCA2" ,  
3     subset_row=rownames(example)[1:1000] ,  
4     ncomponents=25)
```

**Listing:** Понижение размерности PCA-2 и построение графика

# Понижение размерности и визуализация

На графике каждая точка - клетка, а раскраска идет по значению level1class(метаданные, не входят в фичи, которые понижает PCA)

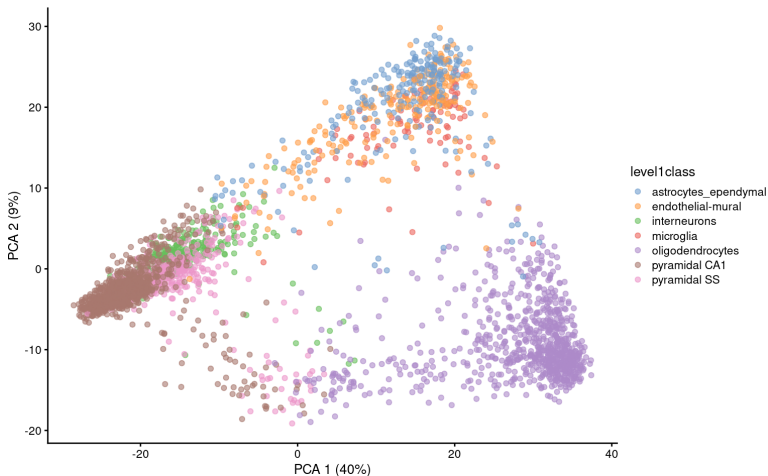


Figure: PCA график