1 Single Cell Quality Checking

This module allow to remove bad quality cells from dataset.

- **Internal name :** scqualitychecking01
- **Avalaible :** local mode
- **Input Ports :**
  - initial count matrix (tsv)
  - initial cells metadata (tsv)
  - genes metadata (tsv)
- **Output Ports :**
  - filtered count matrix (tsv)
  - filtered cells metadata (tsv)
- **Optional parameters :**

Parameter	Type	Description	Default Value
plot_option	string	Option for plots generation: All all plot Dots only scatter plots Saturation only saturation plots	All
detection_threshold	integer	Minimal number of reads to consider a feature as detected	10
expression_threshold	integer	Minimal number of detected features	4000
expression_option	string	Type of feature to detect (Endogenous, Nuclear or All)	Endogenous
reads_threshold	integer	Minimal number of mapped reads to keep a cell	200 000
reads_option	string	Type of features to consider for reads counting (Endogenous, Nuclear or All)	Endogenous
prop_mt	float	Maximum proportion of reads mapping to mitochondrial features	0.1
prop_sp	float	Maximum proportion of reads mapping to exogenous features	0.5

- **Configuration example:**

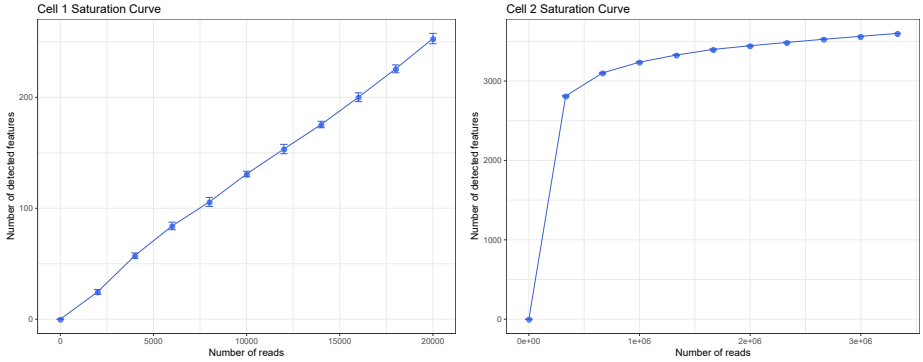
```
<step id="QC" skip="false">
        <module>scqualitychecking01</module>
                <parameters>
                        <parameter>
                                <name>plot_option</name>
```

```
                <value>Dots</value>
            </parameter>
            <parameter>
                <name>reads_threshold</name>
                <value>500000</value>
            </parameter>
            <parameter>
                <name>expression_threshold</name>
                <value>1500</value>
            </parameter>
        </parameters>
</step>
```

2 Interpreting output files

2.1 Saturation Plots

For each cell, the module produces a saturation plot. Briefly, considering count for a cell, the module resamples increasing number of reads and count the number of unique feature detected.
After this sampling two main profiles can be observed :

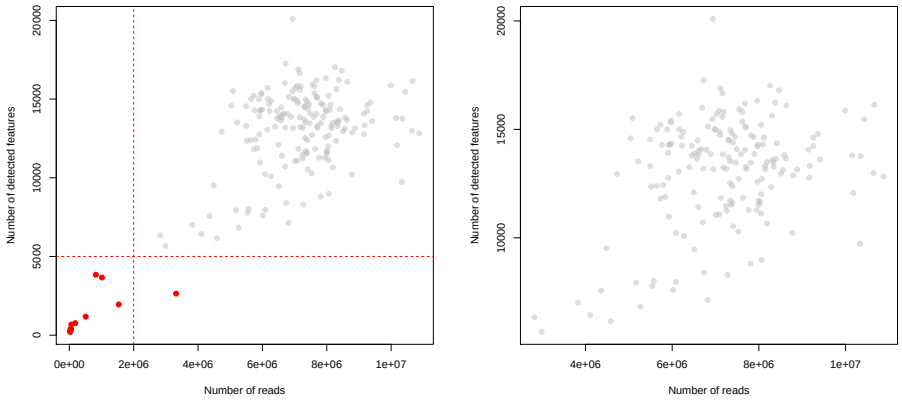


Left : unsaturated cell, Right : saturated cell

The idea is to exclude unsaturated cells and keep saturated ones, for most of the information from saturated cells has been captured, whereas unsaturated cells miss a piece of information.

2.2 Scatter Plot

After cleaning data, the module produces two scatter plot, showing all cells in term of number of feature (y-axis) and number of reads (x-axis).



The first one, show all cells, and filtering threshold. Cells in red are those being eliminated. The second one shows cells remaining after filtering. At the end of the filtering, cells should behave like a mixture of gaussian, i.e. you can wrap them in a given number of ellipses.