

Immediately after clicking finishing data import, you will be shown the summary of per spot/cell gene count metrics. This summary shows the number of counts and genes per spot (or cell, depending on the spatial transcriptomics technology), for each sample. Here a gene is counted if at least one read was detected for that gene in at least one spot/cell. This summary can be downloaded to a comma-delimited file if necessary (red arrow).

This tutorial will start by clicking the **Filter data** tab.

QC & Data Transformation

- **Filter data:** Remove spots or genes from the data set. The filter can be executed on all samples or a subset.
- **Normalize data:** Transform gene expression counts, with log-normal or SCTransform as options. The normalization procedure does not require filtering, but removal of low-count spots/cells or low count genes is recommended.
- **Pseudobulk analysis:** Explore sample-level gene expression. Pseudobulk analysis combines gene counts across all spots/cells and produce pseudo-RNAseq samples. This analysis allows exploration of sample-level transcriptional patterns. Does not require execution of the Filter Data or Normalize Data sections.
- **Quilt plot:** Generates spatial (quilt) plots showing spot/cell-level metadata. Users can assess QC metrics such as gene counts per spot or total counts per spot and detect potential technical artifacts.

Original Summary **Filter data** Normalize data Pseudo-bulk analysis Quilt plot

1

Per Spot/Cell Metrics

Sample Name	Min Counts	Average Counts	Max Counts	Min Genes	Average Genes	Max Genes
ductal_carcinoma	253	13,320	43,555	225	4,776	8,414
breast_cancer_section1	524	18,606	72,315	393	5,063	9,594
breast_cancer_section2	96	18,425	80,748	81	5,003	10,020

Download summary

In the **Filter data** tab, please click the “Select samples to apply this filter” text. Blue boxes with the uploaded samples will appear. You can click individual samples that you do not wish to apply the filter (red arrows). In this tutorial, however, no samples will be excluded from the filter as most users will probably want to apply filters to all samples in a project.

Please click in the next section **Filter spots/cells**.

Original Summary **Filter data** Normalize data Pseudo-bulk analysis Quilt plot

Select samples to apply this filter 2

Click to add/remove a sample

All samples selected

DUCTAL_CARCINOMA

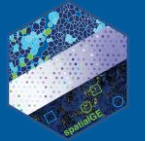
BREAST_CANCER_SECTION1

BREAST_CANCER_SECTION2

Filter Spots/Cells 3

Filter genes

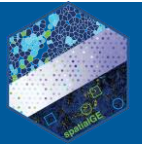
APPLY FILTER



The **Filter spots/cells** section removes ROIs, spots, or cells based on the number of counts or genes per ROI/spot/cell. The first filter in this section, **Keep spots/cells with this number of counts** defines the minimum and maximum gene counts an ROI/spot/cell should have to be kept in the data set after filtering. The user is encouraged to think about reasonable values for this filter. Since each project is different, and factors such as quality of the sample, sequencing depth, level of necrosis, affect the number of counts per ROI/spot/cell, establishing default values is difficult. From the **Original summary** tab, it can be seen that the spot each Visium sample has spots with gene counts as low as 96. For the purpose of this example, spots with less than 2000 counts per spot will be removed, which should be enough to remove low-quality spots. For that, the user can either type in the number in the left box or move the left slider button to the right until 2000 shows in the left box. The maximum number of counts per spots will be left untouched. Nevertheless, for technologies that use image segmentation to define cells, the user might want to limit the maximum number of counts per cell in order to remove poorly segmented cells (“doublets”).

Next, the **Keep spots/cells with this number of expressed genes** filter removes ROIs, spots, or cells based on the number of expressed genes. Here, a gene is considered expressed if at least one count was detected in at least one spot. In general, spots with low number of expressed genes are also considered low-quality. From the **Original summary** tab, it can be seen that the sample “breast_cancer_section2” has spots with as low as 81 genes. To remove those spots, please write in the left box the number 200. Similarly, the maximum number of genes will be left untouched, but the same previous considerations apply for image segmentation-based spatial transcriptomics.

The spots can also be filtered based on the percentage of mitochondrial or ribosomal counts. **Under Keep spots/cells by percentage of counts**, the user can specify the minimum and maximum percentage of counts from mitochondrial or ribosomal genes. Here, spots will be filtered based only on mitochondrial gene counts by checking the “**Mitochondrial genes (^MT-)**” box. Next, move the slider button on the right to the 20% mark or



type 20 in the right box. Keep the left slider button in zero. With this filter, spots that contain 0 to 20% of mitochondrial reads are kept. A 20% cutoff for mitochondrial counts is commonly used in many single-cell studies. Here, this filter aims to remove spots from areas with cellular stress or death which may have higher expression of mitochondrial genes.

Note: The “^MT-” in the **Mitochondrial genes** check box indicates that spatialGE searches for genes with names starting by “MT” followed by a hyphen in the gene names. This filter works in Visium as mitochondrial genes are named that way.