After creating a project, you will be taken to the **Import data** module. This is the starting point of your data analysis. In spatialGE, many data formats are accepted in order to enable analysis from many spatial transcriptomics technologies. If you would like to read more about the accepted formats, please go to "How to get started". In this tutorial, a Visium data set containing three samples will be used as example. The data sets are the "Human Breast Cancer (Block A Section 1)", "Human Breast Cancer (Block A Section 2)", and "Human Breast Cancer: Ductal Carcinoma In Situ". Please, note that you may be required to create an account to download these data sets from here.

Samples in spatialGE are added one at a time. Hence, the following steps should be repeated until all samples are imported to the project.

In this example project, which has been named "TestProject", start by assigning a name to the sample. If you choose not to assign a name to the first sample, the name "Sample01" will be automatically assigned. In this example, the first sample that will be imported is the ductal carcinoma sample. The name "ductal_carcinoma" will be typed in the **Sample name** box.

Next, the gene expression data should be added. Since this example data set comprises Visium samples, an .h5 file can be dragged from a directory in the computer to the center of **Gene expression** box. Alternatively, you can click the box, which will open a window for you to find and select the .h5 file. The same process should be repeated with the **Coordinates** box, which takes the comma- or tab-delimited file containing the Visium spot coordinates.
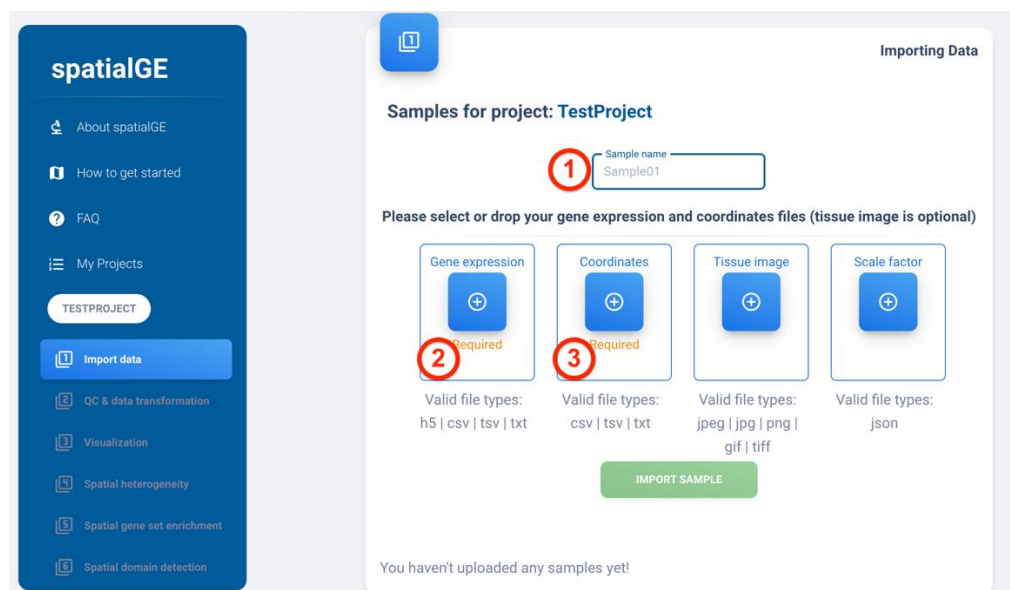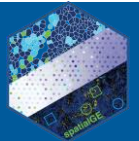

Optionally, you can import a tissue image accompanying the expression and coordinate data. In this example, H&E images are available, which you can drag and drop to the **Tissue image** box. In the case of Visium, a JSON file can also be uploaded using the **Scale factor** box. The JSON file is a default output of the Space Ranger algorithm used to process raw Visium data.

1
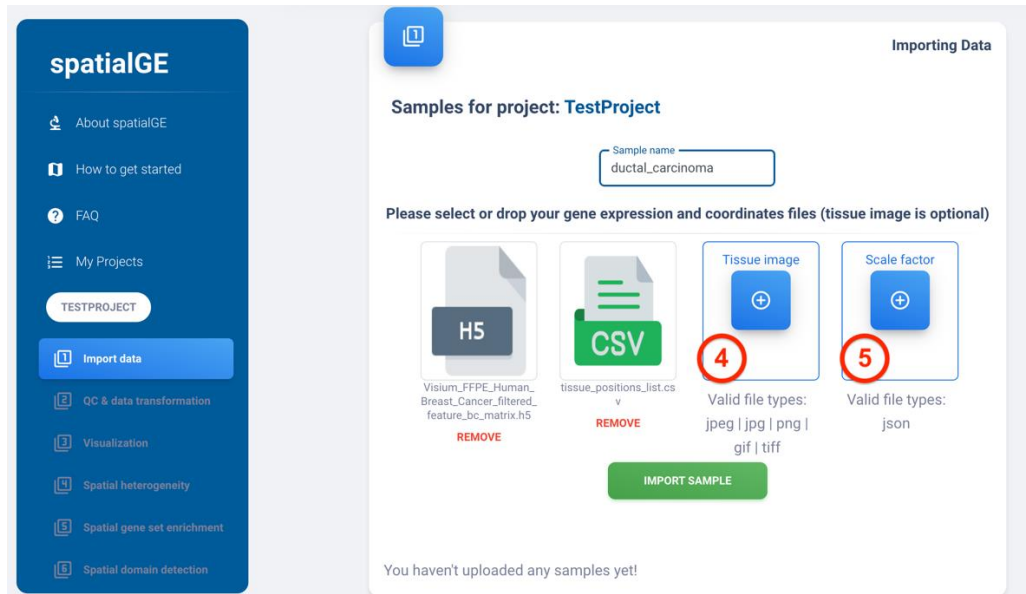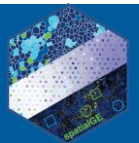
To finalize the upload of this sample, click the green button **IMPORT SAMPLE**.
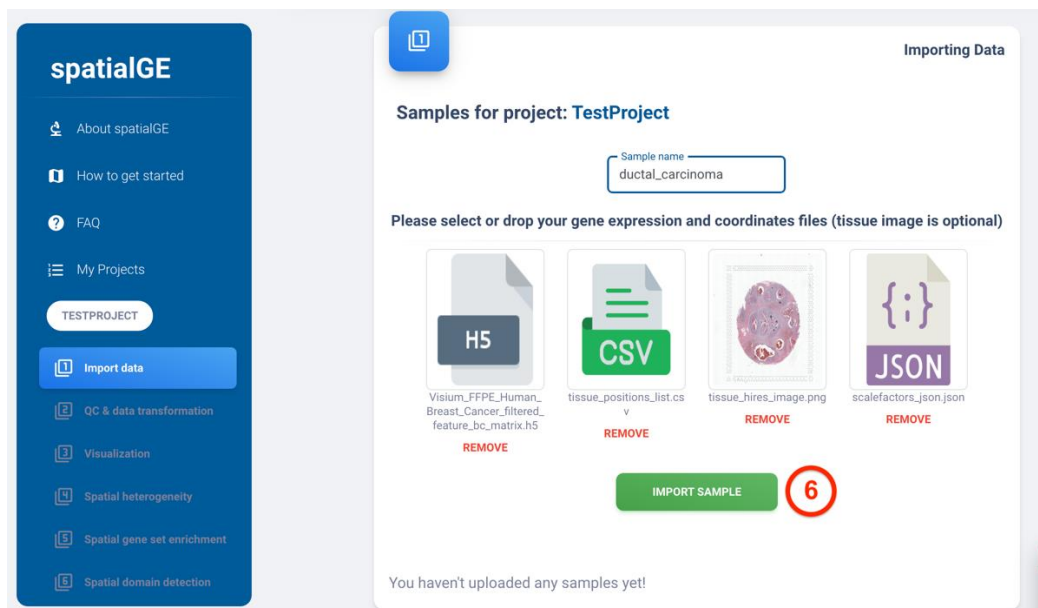
At this point, you will be directed to the project's main page. Two more samples were dowloaded from the Visium public data repository. Thus, click the **ADD NEW SAMPLE** button, and repeat the previous steps (1-6 in the figures) for the other two samples.

Now that spatial data for all samples have been uploaded, you have the option to input sample-level metadata. This metadata can be any variable(s) that are associated to the sample or the source of the sample. Examples include tissue type, donor, treatment, mutational status, etc. Two options are available to input sample metadata. Option 1 involves upload of a file containing a table with the metadata. The first column of this file contains the same sample names used to upload the data files. These sample names are case-sensitive. Subsequent columns contain the sample metadata. For more details about this file, please refer to "How to get started".

Adding metadata is optional, and users can proceed directly to click **IMPORT DATA**. In this example, Option 2: **ADD METADATA MANUALLY** will be used. Please click Option 2.
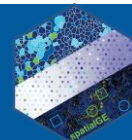
Next, please click the **ADD NEW METADATA COLUMN** button. An empty column appears in the "Sample/Metadata" table. Now, you can add a name for the first metadata variable and the value of that variable for each sample. The process is repeated for as many times as metadata variables are available.

Now that metadata input has been completed, please click the green button **METADATA COMPLETE**. Finally, click the **IMPORT DATA**. A small pop-up will show asking if you would like to receive an email when the data import is finished. For a small project like this with only three samples, this might not be necessary.

When the pop-up disappears, you will be taken to the next module **QC and data transformation**. Please click here to see the documentation for this module.
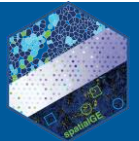
spatialGE

- About spatialGE
- How to get started
- FAQ
- My Projects

TESTPROJECT

1. **Import data**
2. **QC & data transformation**
3. Visualization
4. Spatial heterogeneity
5. Spatial gene set enrichment
6. Spatial domain detection
7. Differential expression
8. Spatial gradients

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 SCTransfom 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 |

Per Spot/Cell Metrics

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

Download summary