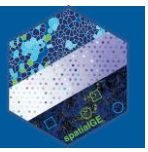


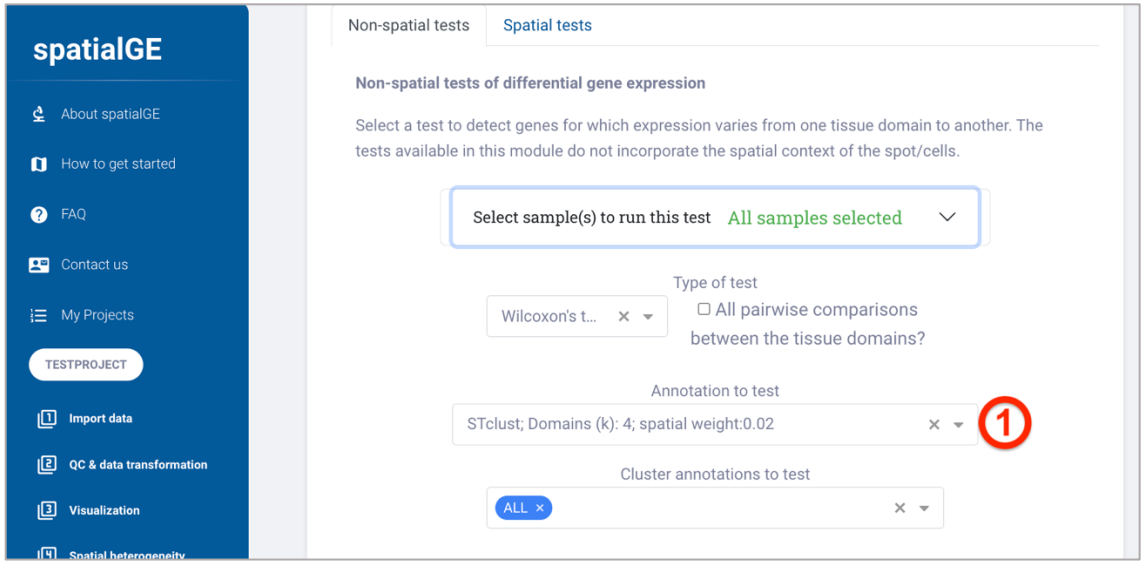
After tissue domain detection has been completed (**Spatial domain detection** module), the **Differential expression** module becomes available. The **Differential expression** module allows users to identify genes with differences in expression among the detected domains (i.e., Gene_A is expressed significantly higher in Domain_1 compared to Domain_2). Users can perform differential expression analysis using one of the tests available: Wilcoxon's Rank Test, T-test, or mixed models.

The **Differential expression** module can be used in two modes:

- The first mode which will be called in this tutorial "marker detection", tests if a **gene is differentially expressed in one of the domains compared to the rest of the tissue**. Using this mode, users can detect genes that "transcriptomically define" a given domain. Most users will probably be interested in using this mode.
- The second mode, or "pairwise testing", allows users to identify genes that are different between two given domains. This testing mode will allow you to detect fine-grain differences, as **each test will consider only the regions of interest (ROIs), spots, or cells assigned to the two domains being tested**.

Note: The current **Differential expression** module the spatialGE web application does not account for spatial autocorrelation (i.e., two close ROIs/spots/cells are likely to be more transcriptomically similar compared to distant ROIs/spots/cells). Work is being conducted to add support for spatially-informed differential expression.

To begin with the analysis, it may be worth looking for a moment to the results in the **Spatial domain detection** module. In that module, identify the clustering solution that represents in the best way the biological findings you want to extract. Then go back to the **Differential expression** module. For the purpose of this tutorial, the K=4 solution with spatial weight=0.02 (*STclust; Domains (k): 4; spatial weight:0.02*) will be selected in the **Annotation to test** drop-down of the **Differential expression** module.

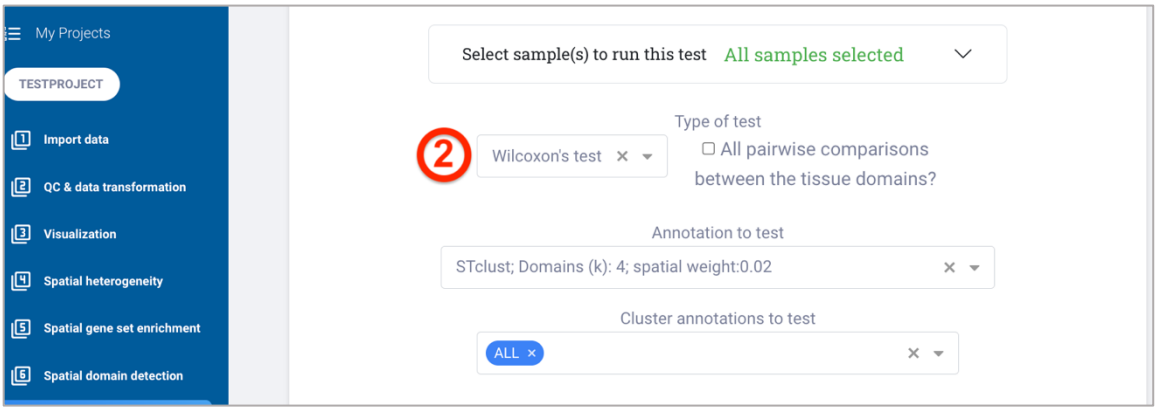


Notice that each sample can be tested separately using the **Select sample(s) to run this test** section. This functionality is useful for when you consider each sample is best represented by different clustering solutions,

however, you will need to save the results each time in the form of a spreadsheet (explained later in this document). For the purpose of this tutorial, all samples will be included.

The **Type of test** section has two components. In the first one, the user can select the one of Wilcoxon's test, T-test, or mixed models. The *Wilcoxon's test* will be used in this tutorial. This test is one of the most used differential expression tests in single-cell analysis (e.g., Seurat's FindAllMarkers). The T-test is another option, especially when domains are composed of tens to thousands of spots/cells. This scenario is typical of Visium or single-cell spatial data sets. The mixed model option is applicable to similar situations as the T-test, but its implementation is more similar to the classical analysis package limma for bulk-RNAseq.

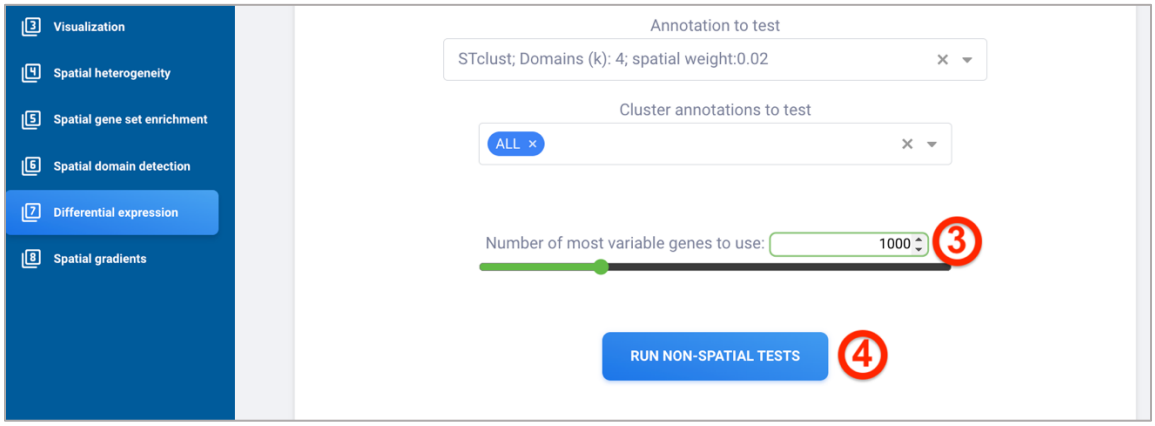
The **All pairwise comparisons between the tissue domains?** checkbox will be left unchecked in this example, which instructs spatialGE to run the "marker detection" mode. If the checkbox is checked, the "pairwise testing" mode is activated. See the beginning of this documentation for more details on the differences among these two modes.



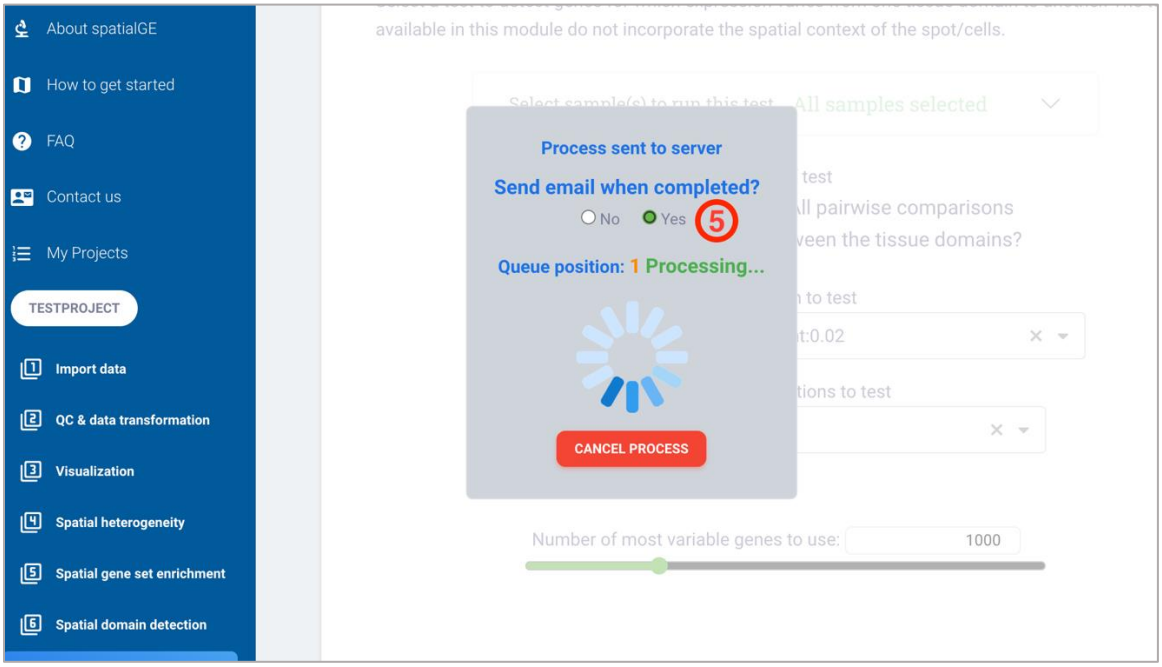
In the **Cluster annotations to test** dropdown, it is possible to select a subset of domains to be involved in the analysis. By default, *ALL* the domains are included in the analysis. If the user wants to exclude a specific domain, For example, domain 3, then domains 1, 2, and 4 should be selected from the drop-down. In this example, all domains will be included. Please, keep the default (*ALL*).

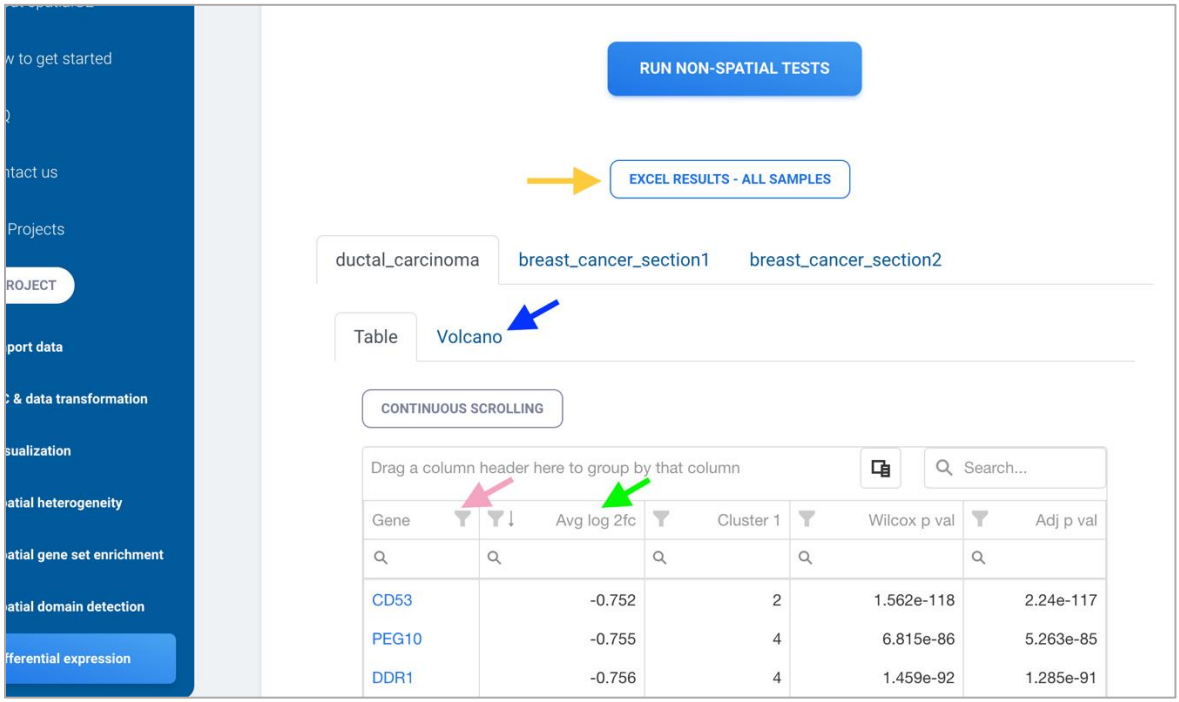
By default, spatialGE will perform the differential expression test on the top 100 genes with the highest variation. If the user wishes to include more genes, the **Number of most variable genes to use** slider should be used. In this example, the top 1000 genes will be included, and to do so, please write the number 1000 in the textbox next to the slider.

To execute the analysis, please click the **RUN NON-SPATIAL TESTS** button.



The differential expression analysis is time-consuming, and it is recommended to use the email notification functionality and return to the project once the analysis is completed. Please, click the **Yes** radio button under the **Send email when completed?** option of the pop-up window.



Once the analysis is finished, the results are available in the two forms: The first is a table with the results of each gene x domain test in rows. By clicking a gene name, a web browser window is opened displaying the corresponding GeneCards record. Filtering and sorting of this table is available by clicking on the filter icon (pink arrow) or column names (green arrow) respectively. Notice that this table can be downloaded as an Excel file by clicking the **EXCEL RESULTS - ALL SAMPLES** button (yellow arrow). The second option to visualize results is a volcano plot, which can be activated by clicking the **Volcano** sub-tab (blue arrow) within each sample tab. A volcano plot is generated for each domain (or each two-domain comparison if pairwise testing mode was requested). These plots can be downloaded in different formats.

Module: Differential expression