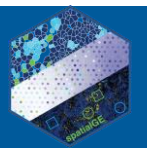
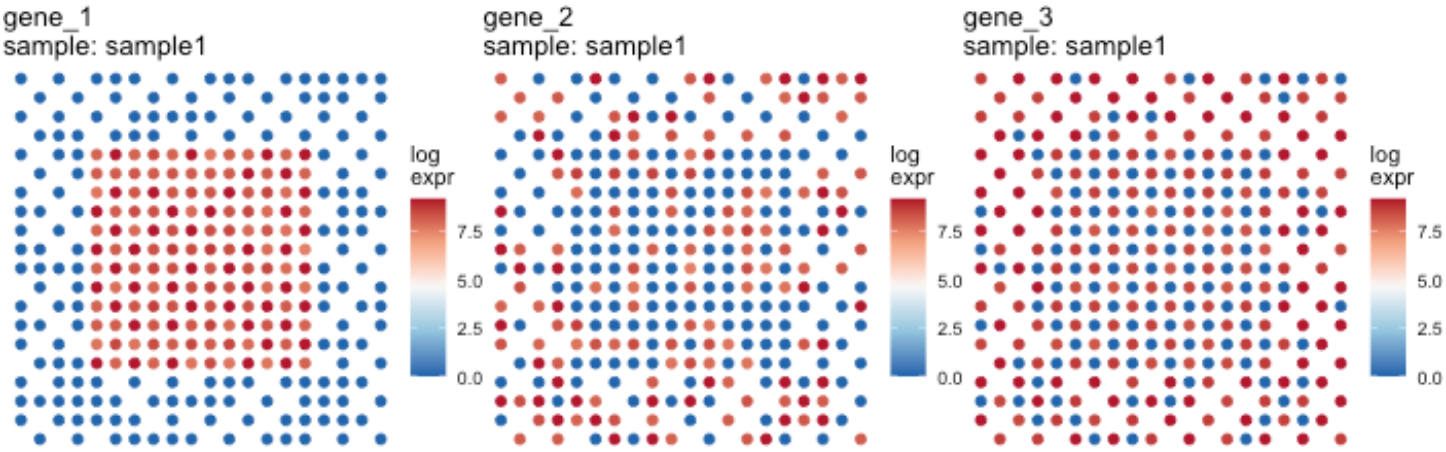


The **Spatial heterogeneity** module is automatically enabled after the normalization procedure (QC & data transformation module) is completed. **Spatial heterogeneity** is one of the modules in spatialGE that applies spatial statistics to spatial transcriptomics data. Specifically, this module enables the user to quantify the level of spatial autocorrelation in a series of genes at each sample. In other words, spatial autocorrelation is a measure of the tendency of gene expression to be arranged in "hotspots". Once spatial statistics have been calculated, and if sample-level metadata is available. the **Spatial heterogeneity** module also allows the user to look for associations between the level of spatial autocorrelation and sample-level data (e.g., clinical outcomes, therapy, tissue type, etc.).

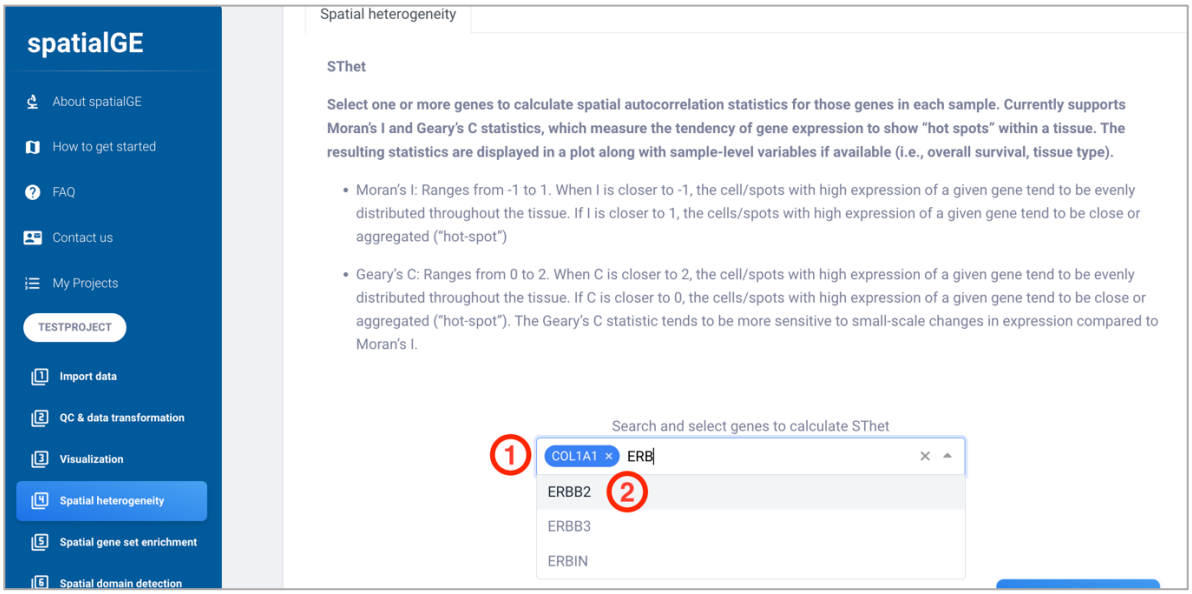
To illustrate the concept of spatial autocorrelation, below is a simulated spatial transcriptomics sample with three genes. The empty areas result from the automatic removal of zero-count spots that occurs during data import in spatialGE. The first gene ("gene_1") is an example of positive spatial autocorrelation in gene expression. An expression hotspot can be observed in the center of the sample. The second gene ("gene_2") contains the same gene counts as in "gene_1" but have been randomly assigned to spots. In this case, no spatial pattern is observed and hence no spatial autocorrelation. Finally, for the third gene ("gene_3"), the expression has been arranged in a way that spots with no expression of "gene_3" are equidistantly distributed (i.e., high expression spots alternating with low expression spots). In the case of "gene_3", expression is uniformly distributed and a pattern of negative spatial autocorrelation.



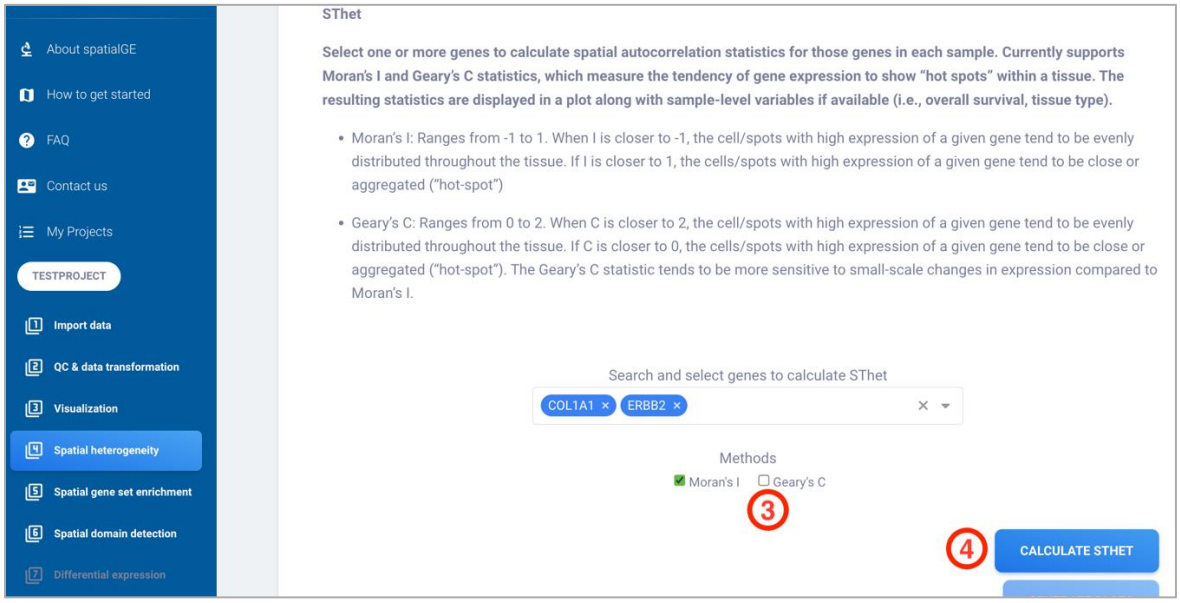
The spatial autocorrelation measurements included in spatialGE are Moran's I and Geary's C. The Moran's I statistics ranges from -1 (negative autocorrelation) to 1 (positive autocorrelation), with zero indicating no spatial autocorrelation. The Geary's C statistic ranges from 0 (positive autocorrelation) to 2 (negative autocorrelation), with 1 indicating no spatial autocorrelation. For the simulated tissues above, the Moran's I values were 0.19, 0.01 and -0.01 for "gene_1", "gene_2", and "gene_3" respectively. The Geary's C estimates were 0.83, 0.98, and 1.03. Notice that the Geary's C estimates for "gene_2" and "gene_3" are almost the same and indicate no spatial autocorrelation, which probably results from Geary's C being more sensitive to differences at smaller spatial scales.

When estimating spatial statistics on real samples, the user might find that genes rarely produce negative spatial autocorrelation patterns. In the context of biological tissues, it is difficult to imagine that cells (or spots) with high expression of a gene alternate equidistantly with cells/spots with low expression of the same gene, as in the artificial pattern in "gene_3".

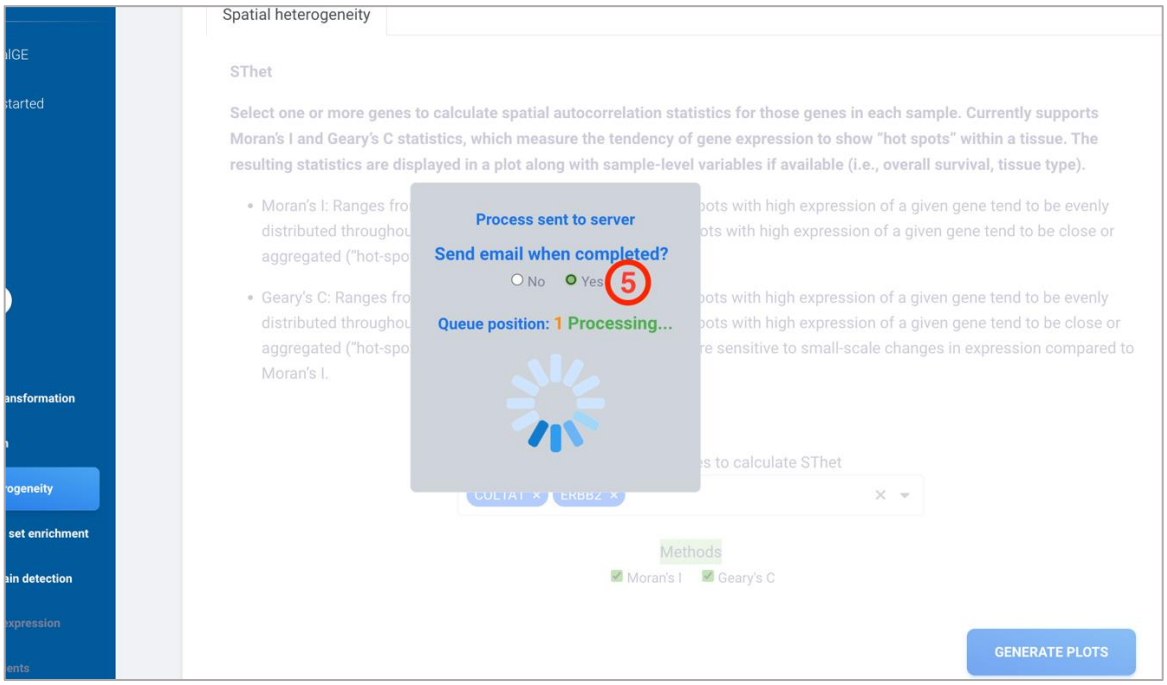
Now that the concept of spatial autocorrelation has been briefly explained, in the **Spatial heterogeneity** module please locate the **Search and select genes to calculate SThet** textbox. Next, type the names of genes to calculate spatial statistics. The textbox features auto-complete, and by typing a letter genes with names beginning with that letter show up. Once the name of the gene appears in the drop-down, please click on it to add it to the analysis. In this example, spatial statistics for genes *COL1A1* and *ERBB2* will be calculated. Please, remember that the textbox is case-sensitive.



Next, check the **Geary's C** checkbox to calculate this statistic in addition to **Moran's I**. And click on the **CALCULATE STHET** button. This button triggers execution of the spatialGE's function *SThet*, which uses the R package *spdep*.

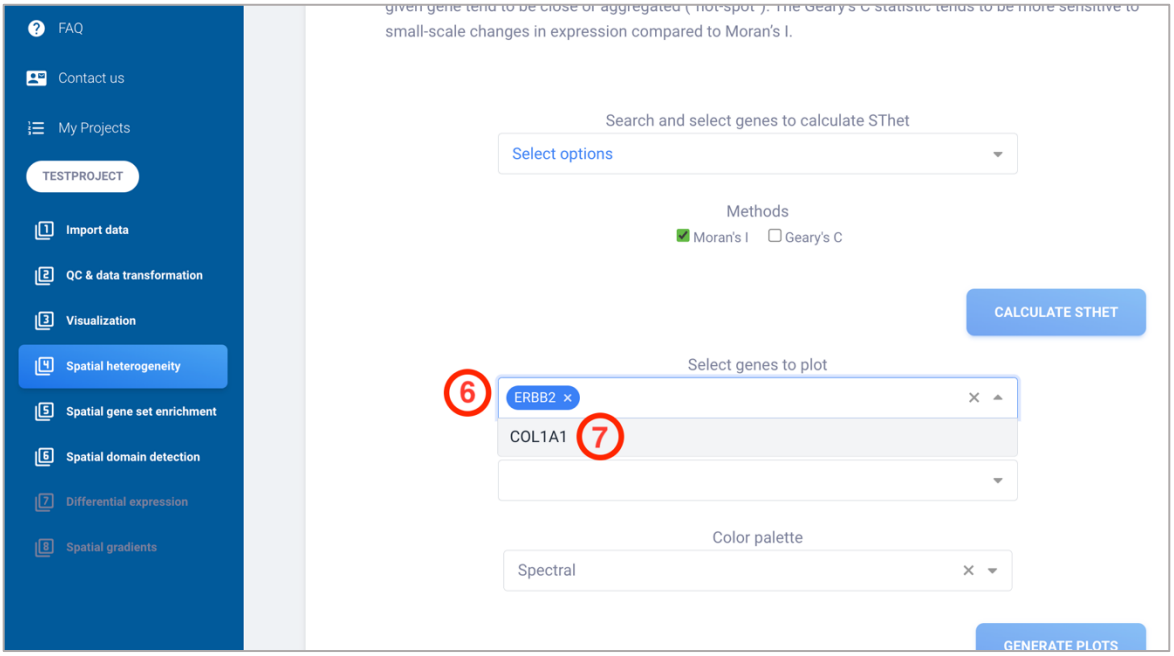


Depending on the number of spots/cells in your samples, calculating spatial autocorrelation statistics can be time-consuming. In this case, the user can check the radio button for e-mail notifications, which becomes available immediately after clicking the **CALCULATE STHET** button. Please, click **Yes** under the **Send email when completed** option. You will receive an email message letting you know when the process has been completed. At this point, the user can close the web browser without losing progress.

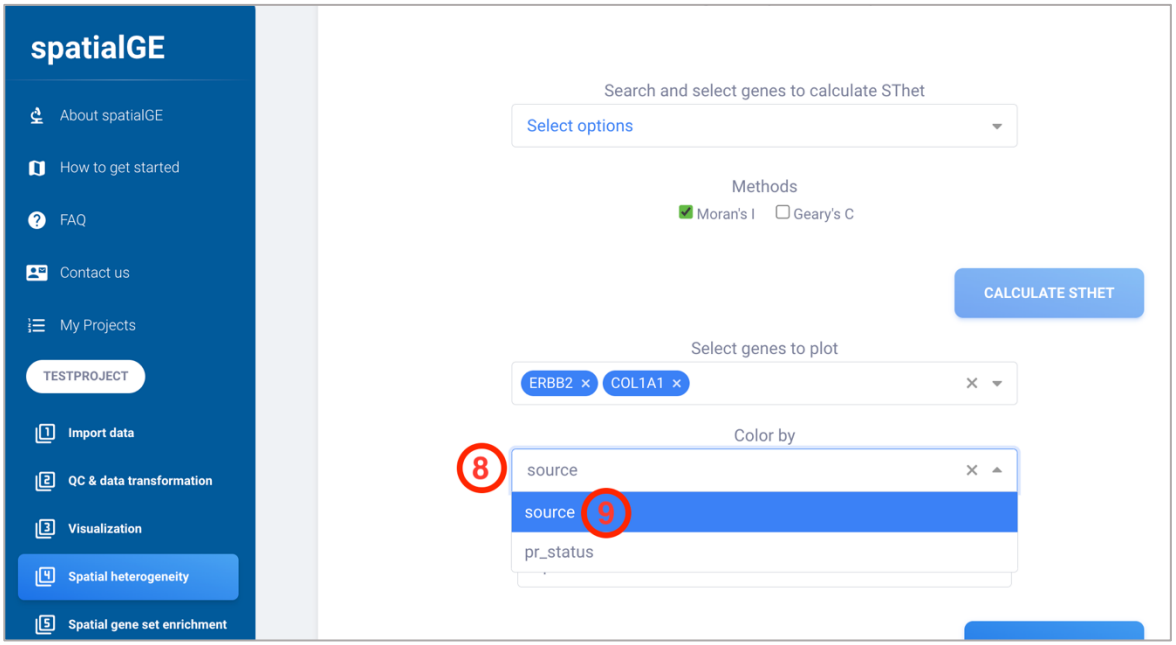


Once the calculation of statistics is completed, new functionality is enabled in the **Spatial heterogeneity** module (regardless of election to receive a notification via email). Below the **CALCULATE STHET** button, a new drop-down appears. In the **Select genes to plot** drop-down, please click the small, inverted triangle to see the list of

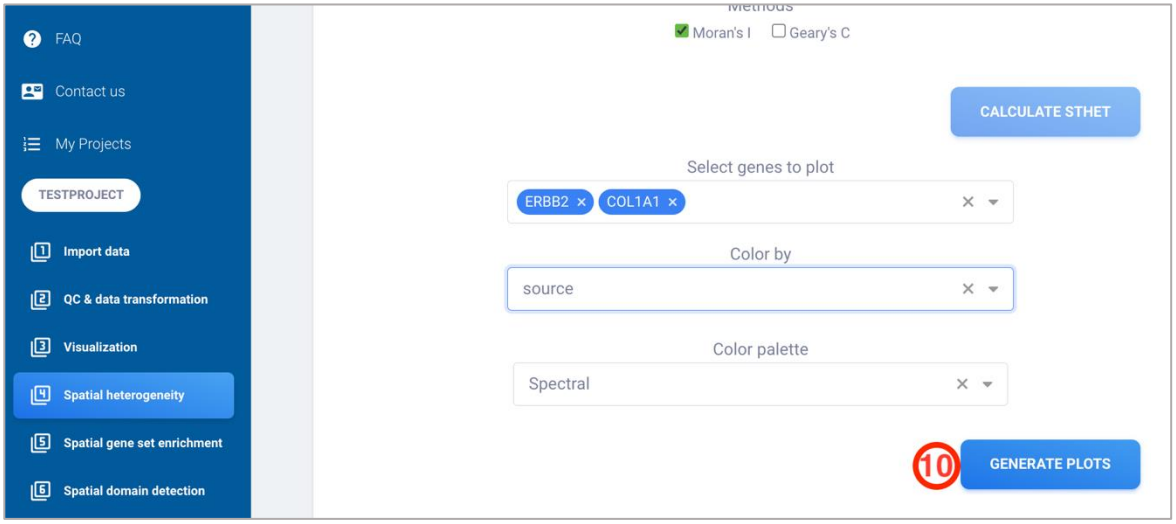
genes for which spatial statistics have been calculated. Please, add the two genes (*COL1A1* and *ERBB2* in this example) by clicking on their names.



If sample-level metadata is available, as in this example, please click the **Color by** drop-down. In this drop-down, the variables included as metadata in during data import are displayed. Given the small size of this data set, associations between the metadata variables and spatial statistics are unlikely to be observed. Nevertheless, for illustration purposes, please select "source" (one of the variables entered during data import in this tutorial).



The default **Color palette** will be used ("Spectral"). Now, please click the **GENERATE PLOTS** button.



The plot generated shows the estimates of Moran's I and Geary's C for each sample (each point represents a sample). This plot enables the user to discover patterns in which spatial patterns could be related to clinical variables. The user can download this plot by clicking the desired format (**PDF**/**PNG**/**SVG**, green arrow). If the numeric estimates are needed, the user can download a file containing a table with the estimates (pink arrow).