Module: Spatial domain detection – SpaGCN

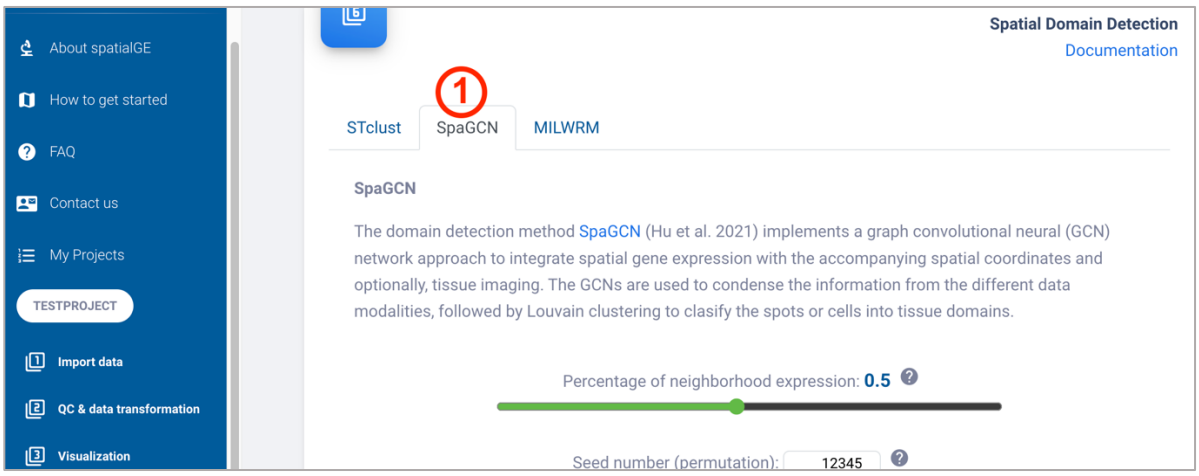
Researchers studying single-cell gene expression commonly use some flavor of clustering algorithm to group cells based on their transcriptomic similarity. When studying spatial transcriptomics (ST) data, this type of analysis acquires an additional purpose: Identify tissue niches or domains. Multiple options are available for clustering of ST data, with some treating regions of interest (ROIs), spots, or cells as spatially independent, and others explicitly using the spatial information. The method SpaGCN falls in this second category, and is provided as an alternative to STclust (the method developed as part of the spatialGE R package).

The method SpaGCN (Hu et al. 2021) is implemented in the **Spatial domain detection** module. SpaGCN features the use of graph convolutional networks (GCNs) to conduct unsupervised clustering of the spots or cells in a ST sample. In brief, SpaGCN constructs a graph from the spatial coordinates to describe the spatial relationships between the spots/cells. A GCN is constructed to summarize gene expression of neighboring spots/cells. Domains are identified from the expression graph using Louvain clustering followed by iterative clustering. If the data set contains Visium samples, the user optionally can refine the domains assignments, resulting in domains that are more spatially continuous (see Hu et al. 2021 for details).

Some aspects to consider about the spatialGE implementation of SpaGCN:

- SpaGCN supports the use of tissue images to better inform spatial relationships. This feature is not implemented yet in the **Spatial domain detection** module of spatialGE but will be available in the future.
- Spatially variable genes can be detected for each domain in SpaGCN. This feature will be available in the **Spatial domain detection** module of SpaGCN.
- A **seed number** argument is available in this implementation. The user is encouraged to run SpaGCN using different seed values and saving the results in order to check for consistency in tissue domain predictions.

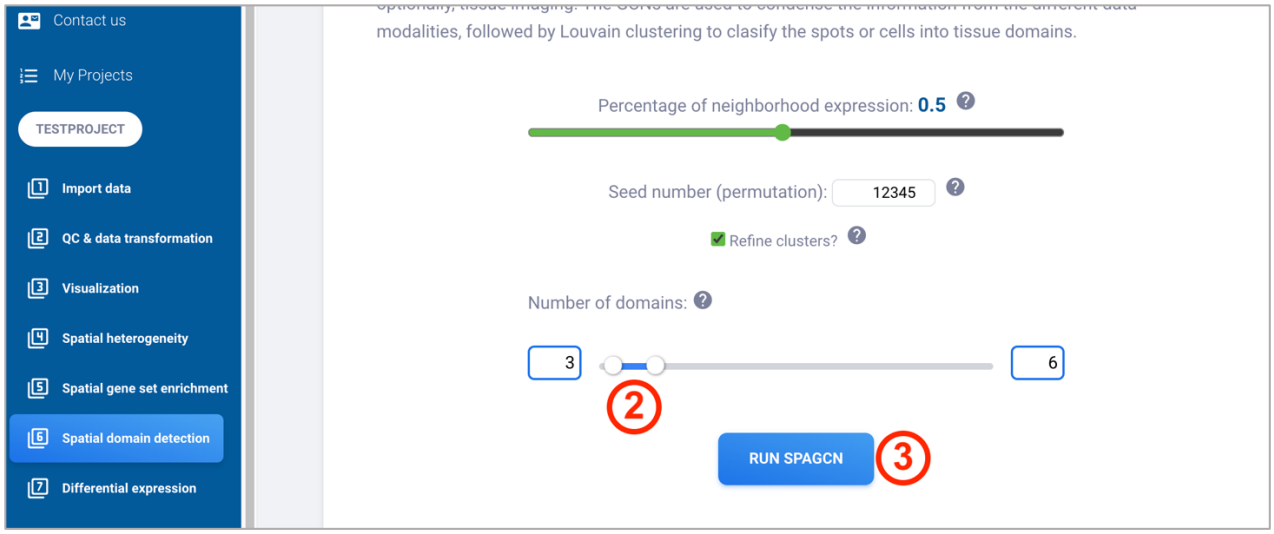
To begin domain detection, click the **SpaGCN** tab next to the STclust tab.



One of the parameters to define when running SpaGCN is the **Percentage of neighborhood expression** contributing to the expression profile of a given spot/cell. The authors of SpaGCN state that an appropriate setting of this parameter for Visium experiments is 0.5 (default). If data from single-cell spatial transcriptomics is used,

this parameter should be set to a higher value, so that the summarized expression is more affected by neighboring cells. Since the data set used here was generated with Visium (see Import data), the **Percentage of neighborhood expression** parameter will be left in its default (0.5).

Next, use the **Number of domains** slider to specify a range of K values (i.e., number of expected domains). SpaGCN will be run for each of the selected K values. For this tutorial, move the slider so that the range of K values goes from 3 to 6. Finally click on the **RUN SPAGCN** button to perform tissue domain detection.



Once the SpaGCN has been completed, a series of tabs (K=3 to K=6) are presented in the interface below the **RUN SPAGCN** button. Each tab contains a representation of the spatial locations of domains within the tissue samples (and the tissue images if available). Within each tab, additional sub-tabs contain the plots for each sample. These tabs contain the results from SpaGCN when for the K value in the first tab level, as well as the *refined* clustering if the **Refine clusters** checkbox was activated. Plots can be downloaded by clicking the appropriate file format button (**PDF/PNG/SVG**, green arrow). The tissue images (if available) can be downloaded along the tissue domain plots by checking the **Quilt plot with H&E image** checkbox (red arrow).

Module: Spatial domain detection – SpaGCN