Python

The solver is implemented in Python v3.8 and the following main python packages have been used:

- matplotlib=3.4.3
- numpy=1.21.2
- opencv=4.0.1
- pandas=1.3.3
- skimage=0.18.1

For further details concerning the used Python packages see the conda environment.yml file.

Structure of the project

The structure of the directories and files should be as following and is also highlighted via the 2 example cases (SAMPLE_FOLDER) in the data folder:

```
PROJECT_FOLDER
 |--gtc
 |--data
 |  |--pathways.csv
 |  |--SAMPLE_FOLDER_A/
 |  |  |--Gtc_Parameters_1A.txt
 |  |  |--Gtc_Parameters_2A.txt
 |  |  |--Data/
 |  |--SAMPLE_FOLDER_B/
 |  |  |--Gtc_Parameters_1B.txt
 |  |  |--Gtc_Parameters_2B.txt
 |  |  |--Data/
 |  |--etc.
```

Gtc_Parameters.txt

Is a dictionary which contains the:

- necessary information for the analysis of the in situ sequencing results,
- parameters for the tissue compartment building and
- formatting properties of the plots.

One sample directory must contain at least one and can contain up to several Gtc_Parameters.txt files with different gene sets or settings for tissue compartment building. The various files can be named Gtc_Parameters_xxx.txt, where xxx is an arbitrary string.

Parameters and Entries:

- **disease_history_relaps**: ['no', 'yes']
  Did the patient suffer a relapse?
- **files**:
  Input files for each sample.

    - **dapi_image_tif**: ['*.png', '*.tif']
      Absolute path to the DAPI-stained image. Has the same pixel size as the FITC-stained image.
    - **fitc_image_tif**: ['*.png', '*.tif']
      Absolute path to the FITC-stained image. Has the same pixel size as the DAPI-stained image.
    - **cell_pos_csv**: ['*.csv']
      Absolute path to the as csv-file saved data frame containing the cell coordinates in relation to the image sizes.
    - **gene_pos_csv**: ['*.csv']
      Absolute path to the as csv-file saved data frame containing the gene coordinates in relation to the image sizes.
    - **drawn_tissue_mask_png**: ['*.png']
      Absolute path to image containing the tissue compartments as classified by the pathologist specialist.
      RGB-colour coding:
        - (*144, 194, 38*) non-neoplasitc (green)
        - (*196, 47, 26*) neoplastic (red)
        - (*30, 218, 230*) remove transcripts within (light blue)
- **img_res_in_um_per_px**: [*float*]
  The microscope image resolution in micrometer per pixel.
- **img_scale_factor**: [*integer*]
  Scale factor to shrink images in order to accelerate the analysis. The coordinates are adapted analogously.
- **move_in_secure_seal**: [*integer*]
  Length in pixel to move in from the image border for the secure area. All transcripts outside the secure area are neglected.
- **tissue_masks**:
  The tissue compartments.
    - **composite**:
      Tissue compartment resulting from the union of the neoplastic and non-neoplastic tissue compartments.
        - **method**: ['union', 'threshold']
          How is this tissue compartment build?
    - **neoplastic**:
      Tissue compartment.
        - **method**: ['union', 'threshold']
          How is this tissue compartment build?
        - **genes**: [*string list*]
          Genes whose transcripts are used for construction of the neoplastic tissue compartment, e.g. [*'EREG', 'MET', 'BIK'*].
    - **dot_radius**:
      Size of the kernel representing each gene transcript and cell.
        - **union**: [*integer*]
          Kernel size used for the method union.
        - **threshold**: [*integer*]
          Kernel size used for the method threshold.

    - **threshold_value**: [*float*]
      Cut-off value for binary tissue compartment building. The value lies between 0 and 1. If the value is 0, then the method is the same as union.
    - **kernel_shape**: ['uniform', 'gaussian']
      Shape of the kernel for tissue compartment building. Has either a uniform or a Gaussian shape.
    - **morph_operations**:
      Morphological operations used in compartment building.
        - **blurring**: [True, False]
          Distorts the details.
        - **closing**: [True, False]
          A dilation followed by an erosion.
        - **opening**: [True, False]
          An erosion followed by a dilation.
- **tissue_region_masks**:
  Split the tissue compartment into a boundary region of specified size and a center region and perform gene analysis for each region.
    - **tissues**: ['neoplastic', 'non-neoplastic']
      Which tissue compartments should be split into regions?
    - **boundary_size**: [*integer*]
      The thickness of the boundary region.
- **custom_gene_sets**: [*string: string list*]
  Define sets of genes for comparative analyses, e.g.
  -- '*Enteroendocrine cells*': [*'GAST', 'CHGA', 'GIP', 'MLN'*]
  -- '*Cancer associated fibroblasts*': [*'DCN', 'FSTL1', 'MMP2'*]
- **barplot_props**:
  Formatting properties for the bar plots.
    - **num_genes_per_subplot**: [*integer*]
      Number of different genes per each row of the bar plot.
    - **size_subplot**:
      Resolution of a row of the bar plot.
        - **x**: [*float*]
          Horizontal resolution.
        - **y**: [*float*]
          Vertical resolution.
    - **sort_genes_by**: ['value', 'letter']
      Order in which the genes in the bar plots are sorted – either they are arranged by decreasing counts of transcripts or alphabetically by the gene name.
- **plot_props**:
  Formatting properties for further plots.
    - **output_image_dim**: [*integer*]
      Saving resolution of the obtained tissue compartments in (*image_dim, image_dim*).
    - **dot_radius**: [*integer*]
      Representation size of a transcript in some plots.
    - **num_hist2d_bins**: [*integer*]
      Heat map plot are performed with (*hist2d_bins, hist2d_bins*) compartments.