

rasp-mibi: Recurrence And Survival Prediction via Multiplexed Ion Beam Imaging

This repository contains the code for the article, "Multiplexed Analysis of the Tumor-Immune Microenvironment Reveals Predictors of Outcome in Triple-Negative Breast Cancer"

Software Requirements

This code was developed in the following settings.

OS

- MacOS 10.14.6

Processor

- Intel Core i5

Dependencies

- python (3.7.3)
- numpy (1.16.4)
- pandas (0.24.2)
- opencv-python (3.4.2.16)
- Pillow (6.0.0)
- scikit-image (0.15.0)
- scipy (1.4.1)
- matplotlib (3.1.0)
- lifelines (0.24.0)
- seaborn (0.10.1)
- statsmodels (0.11.1)
- pysurvival (0.1.2)
- scikit-image (0.15.0)
- shap (0.37.0)

Installation

To install the required packages, you can download the required packages individually using: `pip3 install package_name` or, alternatively, use the requirements.txt file: `pip3 install -r requirements.txt`

Download the repository as a whole to run the demos. Install time should be less than 10 minutes.

Demo

1: Data Collection

Images can be downloaded from: <https://www.angelolab.com/mibi-data>.

Only step 2 requires these images directly—all other parts of analysis can be run without it, as intermediate data is provided.

2: Preliminary Features

Note: this step requires the original image dataset in order to be run. If the reviewer chooses to download the images, please edit the paths in the code.\ `python3 calculate_cell_prevalence.py` \ Purpose: calculate the proportion of cells of each cell type in each patient's image.\ Output: a CSV file in the `intermediate_data/` folder indicating the prevalence of each cell type in each patient's image.\ `python3 calculate_protein_expression.py` \ Purpose: Calculate protein expression in each cell of each patient's image and assign positivity to each cell based on a threshold.\ Output: CSVs of expression levels in the `intermediate_data/protein_expression/` folder and positivity assignments in `intermediate_data/created_protein_positivity/`

3: Immune Composition

`python3 immune_composition.py` \ Purpose: determine whether immune composition is associated with recurrence or survival.\ Ouput: two results CSVs in `results/`.

4: Protein Expression

`python3 protein_expression.py` \ Purpose: determine whether the expression of functional proteins is associated with recurrence or survival.\ Output: two results CSVs in `results/`.

5: Protein Co-expression

`python3 calculate_coexpression.py` \ Purpose: Calculate instances of co-expression between proteins.\ Output: `intermediate_data/created_coexpression_matrices/`. The reader can compare this output to `coexpression_matrices/` to ensure reproducibility.\ Estimated time: 10 minutes.

`python3 protein_coexpression.py` \ Purpose: determine whether protein co-expression patterns are predictors of recurrence and survival.\ Output: two KM curves with log-rank test p-value in `results/`.

6: Cell-to-cell Interactions

`python3 voronoi_interactions.py` \ Purpose: calculate cell-to-cell interactions using Voronoi diagrams.\ Output: interaction matrices in `intermediate_data/created_interaction_matrices`. The reader can compare this output to `interaction_matrices/` to ensure reproducibility. \ Estimated time: 40 minutes.

`python3 functional_protein_interactions.py` \ Purpose: determine whether interactions involving functional proteins are predictors of recurrence and survival.\ Output: two KM curves with log-rank test p-value in `results/`.

`python3 functional_protein_interactions.py` \ Purpose: determine whether interactions involving immunoregulatory proteins are predictors of recurrence and survival.\ Output: two KM curves with log-rank test p-value in `results/`.

7: Multivariate Analysis

`python3 multivariate_cox_regression.py` \ Purpose: perform multivariate Cox regression.\ Demo Note: the reader should change the type of cluster to be examined based on the options given in the comments.\ Output: a model summary printed to output.

`python3 random_survival_forest.py` \ Purpose: build a random survival forest to evaluate importance and measure model accuracy.\ Output: an importance plot in `results/` and a concordance index printed to output.\ Estimated time: 2 minutes.

Citation

Multiplexed Analysis of the Tumor-Immune Microenvironment Reveals Predictors of Outcome in Triple-Negative Breast Cancer. \ Aalok Patwa, Rikiya Yamashita, Jin Long, Michael Angelo, Leeat Keren, Daniel Rubin