Project 1

Quantification of tumor-infiltrating lymphocytes from imaging and sequencing data

Data overview

quanTIseq (https://icbi.i-med.ac.at/quantiseq) is a deconvolution method that estimates the fraction of ten different immune cell types from RNA sequencing (RNA-seq) data [1]. It has been used to analyze more than 8,000 patients across 19 cancer types of The Cancer Genome Atlas (TCGA). The results of this analysis for four cancer types are available in the files:

- TCGA_SKCM_quanTIseq.txt (melanoma)
- TCGA_SKCM_quanTIseq.txt (breast cancer)
- TCGA_SKCM_quanTIseq.txt (lung adenocarcinoma)
- TCGA_SKCM_quanTIseq.txt (lung squamous cell carcinoma)

A parallel study [2] has analyzed the H&E images from TCGA patients to estimate the percentage of tumor-infiltrating lymphocytes (TILs). The results of this analysis are available in the Supplementary Material of the original publication, as an Excel file available from the following link: https://ars.els-cdn.com/content/image/1-s2.0-S2211124718304479-mmc2.xlsx. The file reports various features, including the patients' identifiers ("ParticipantBarcode" column), the cancer types ("Study"), and the percentage of TILs estimated from images ("til_percentage").

Analysis to be performed

For each cancer type, select the patients' identifiers (e.g. "TCGA-34-2596") in common between the two datasets.

For each patient, compute the fraction of lymphocytes ("fTIL") estimated by quanTIseq for these patients by summing up the cell fractions of the following cell types:

- B.cells
- NK.cells
- T.cells.CD4
- T.cells.CD8
- Tregs

For each of the selected cancer types, compare fTIL (derived from RNA-seq data) with the til_percentage (derived from images) using scatterplots and Pearson's correlation.

Discuss the results considering that RNA-seq data and image data are usually derived from different tumor portions. Discuss your results also in the light of the results obtained in the original publication using a previous deconvolution method (see Figure 4C).

Report

The analysis above should be described in a short report that will be evaluated and considered for the final grade. The report should:

- Contain the **code** implemented to run the analysis above, together with the **results** (as tables, plots, or just numbers reported within the text) and their **description/discussion**.
- Be at **maximum 6 pages** long.
- Be saved in a Word doc named **"Surname_Name_Project1.doc"**.
- Be sent by e-mail to Dr. Finotello not later than **June 03, 2019**, using **"RProject1_2019"** as e-mail object.

References

[1] Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, Sopper S, Ijsselsteijn M, Brouwer TP, Johnson D, Xu Y, Wang Y, Sanders ME, Estrada MV, Ericsson-Gonzalez P, Charoentong P, Balko J, de Miranda N, Trajanoski Z. ***Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data***. Genome Medicine (in press).

[2] Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Samaras D, Shroyer KR, Zhao T, Batiste R, Van Arnam J; Cancer Genome Atlas Research Network, Shmulevich I, Rao AUK, Lazar AJ, Sharma A, Thorsson V. ***Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images.*** Cell Reports, 23(1), 181–193.e7.