DIGITAL EPIDEMIOLOGY AND PRECISION MEDICINE Project

Differential Analyses of Gene Expression to Investigate Sexual Dimorphism in Liver Hepatocellular Carcinoma[1]

A. Student, B. Student, C. Student

Abstract. Brief and clear paragraph: it should be a miniature of the manuscript including a brief description of: scientific issue and aim, methods, main result and conclusion.

INTRODUCTION

This section includes: the definition of the scientific issue, the summary of main results in the scientific literature (state of the art) and of principles needed for the comprehension of the new hypothesis; the identification of the hypothesis, the disclosure of the relevance of the results and the overview of contents of following sections. In the introduction the aims of the study must be clearly stated.

It does not include conclusions and recommendations

MATERIALS AND METHODS

Describe in this section the experimental procedures and resources, data analysis procedures and statistical methods. Give enough detail to replicate the experiment but do not overwhelm the reader with too many details.

To do:

1. Data
    1.1. Download gene expression data from https://portal.gdc.cancer.gov/ [2] (data category: *Transcriptome Profiling*; data type: *Gene Expression Quantification*; workflow type: *STAR - Counts*) selecting the project ID: TCGA-LIHC in GDC data portal.
    1.2. Select only samples related to the patient IDs in *Annex1* maintaining both tissues (cancer, normal).
    1.3. Select only genes in gene-list2.txt.

[1] a new title can be proposed (it is not required). A good title should be concrete, clear, and brief.
[2] see: https://classroom.google.com/w/NjI1ODE4OTcyMjYw/tc/NjMyMjg4MjE5NTQz

1.4. Separate the dataset into two subsets based on gender: female and male.

2. Differentially Expressed Genes (DEGs)

2.1. Identify DEGs (cancer vs normal) for each sex and specify the thresholds setting; p-value threshold should be less than or equal to 0.05 (it is suggested to apply a correction for multiple comparisons) and Fold Change (FC) threshold: $|FC| \geq 1.2$.

2.2. Evaluate the overlap between the sets of female-DEGs and male-DEGs.

3. Differential Co-expressed Network

3.1. *Cancer vs Normal (for each sex)*

- *Computation*. Compute the differential co-expression network (cancer vs normal) following the procedure described in the slides (DEPM_6_Diff Co-Expression Analysis.pdf). Binary adjacency matrix with $a_{ij}=0$ if $|Z|<3$ [3].
- *Analysis.* Binary and unsigned network: compute the degree index, check if the network is a scale free network and find the hubs (5% of the nodes with highest degree values)
- Evaluate the overlap between the sets of female-hubs and male-hubs.

3.2. *Female vs Male (only for cancer tissue)*

- *Computation*. Compute the differential co-expression network (female vs male) following the procedure described in the slides (DEPM_6_Diff Co-Expression Analysis.pdf). Binary adjacency matrix with $a_{ij}=0$ if $|Z|<3$ [3].
- *Analysis.* Binary and unsigned network: compute the degree index, check if the network is a scale free network and find the hubs (5% of the nodes with highest degree values)

4. Patient Similarity Network (PSN)

4.1. Compute the Patient Similarity Network using cancer gene expression profile

4.2. Perform the community detection (e.g. apply Louvain algorithm to the PSN)

5. OPTIONAL TASKS

5.1. Compute task 3.2 for normal tissue.

5.2. Analysis of differential network obtained with task 3.2. Binary and signed network: extract the two subnetworks (network with only positive links and network with only negative links) and distinguish between hubs with highest negative degree values and hubs with highest positive degree values.

[3] this is a suggestion, but if with this threshold the obtained network is composed of many disconnected components or if it is too dense, a different value can be applied.

RESULTS AND DISCUSSION

Describe the obtained results with the help of figures and tables.

Expected figures:

- Volcano plot
- Degree distribution of the networks
- Subnetwork plot of the *most relevant* genes (e.g the node with the highest degree value in differential co-expression network)

Provide a discussion of the results (e.g. opinions, interpretation, relationship of your findings with other published findings) and implications that can be drawn from your findings.

(if possible) for one or two genes identified in task 3 (e.g. *hubs of the differential network*) provide some reference papers where that(those) gene(s) is(are) studied in the context of the same phenomenon.

Produce a short report (up to 6 pages) following the guidelines. No template is provided. Use the freedom in choosing your format wisely. Body text font sizes smaller than 11pt should be avoided.

Please specify the used programming language, the available functions and, if necessary, those implemented. Do acknowledge any source you used, such as software code, third party figures, cited text, contribution by non-authors, etc.

Evaluation criteria

The mark will be based on: clarity of writing, accuracy of the methods' description, completeness and appropriate presentation of results (including quality of figures and tables), appropriate discussion of the outcome of comparisons, overall structure and format of the report.

Submission procedure

Each student will submit (by email) two files:

i. a PDF file containing the report;

ii. a compressed archive (zip, rar, 7z) containing the software code used to perform the analyses. The latter archive may contain intermediate results (unedited figures, raw output files, etc), if appropriate. Do not include downloaded data in the archive.

The files must be named according to the following scheme:

DEPM-1_proj-<nn >.<ext>,

where <nn> is the StudentID and <ext> is either ‘pdf’ or {‘zip’|‘rar’|‘7z’}.

Delivery date: 2 days before oral exam

Annex1

TCGA-DD-A1EG
TCGA-DD-A3A1
TCGA-FV-A3I0
TCGA-DD-A3A4
TCGA-EP-A26S
TCGA-EP-A12J
TCGA-FV-A23B
TCGA-BC-A10T
TCGA-DD-A116
TCGA-DD-A113
TCGA-DD-A1EH
TCGA-BC-A110
TCGA-DD-A3A6
TCGA-DD-A1EB
TCGA-DD-A11A
TCGA-BC-A10Z
TCGA-DD-A119
TCGA-DD-A3A5
TCGA-DD-A39V
TCGA-EP-A3RK
TCGA-DD-A11B
TCGA-BD-A3EP
TCGA-BC-A10Q
TCGA-FV-A3I1
TCGA-DD-A1EI
TCGA-DD-A39X
TCGA-BD-A2L6
TCGA-BC-A10W
TCGA-DD-A39Z
TCGA-DD-A39W
TCGA-G3-A3CH
TCGA-DD-A11C
TCGA-BC-A10X
TCGA-DD-A3A2
TCGA-DD-A1EC
TCGA-DD-A1EJ
TCGA-BC-A10Y
TCGA-DD-A3A3
TCGA-DD-A3A8
TCGA-DD-A118
TCGA-BC-A216
TCGA-DD-A1EL
TCGA-DD-A11D
TCGA-BC-A10U
TCGA-DD-A114
TCGA-FV-A2QR
TCGA-DD-A1EE
TCGA-ES-A2HT
TCGA-BC-A10R