Data Science 10593052 Bioinformatics & Network Medicine 2020-2021

Date: DD Month YYYY

Network Medicine Project report

|  |
| --- |
| Network Medicine@Data Science A.Y. 2020-2021  Manuscript Title  Laurenti Laura, Mandica Paolo, Testa Lucia  Group no. 1 Abstract Max 150 words. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. |

**Part 3 – Reporting**

----

**3.1) Summarize the following information in a short report which includes:**

• very short intro (10 lines max) about the pathophysiological condition (i.e. the seed

genes context) and, if any, issues with gene IDs

• a table with seed genes information (point 1b; omit “protein description”)

• a summary table of interaction data (point 1.2c)

• the charts of the enrichment analysis from Enrichr (point 1.4)

• a table with global measures of the disease interactome LCC  
• a figure of the LCC (do not forget figure captions)

• a table with the first 20 highest ranking genes for betweenness (include in the table

also all other calculated centrality measures as from 1.2b) for the LCC

• summary table of the putative disease modules found (*for each module: no. of seed*

*genes in each module, total no. of genes in each module, ratio no. seed genes/total*

*genes in the module, p-value of the enrichment using the hypergeometric test*)

• the list of the first 30 genes identified by the DIAMOnD tool and charts from Enrichr.

• notes and comments on the method followed, discrepancies, lack of data, any other

point worth to be mentioned.

Notes: all tables and figures must have a caption (i.e. they must be self-consistent); a report  
template is provided.

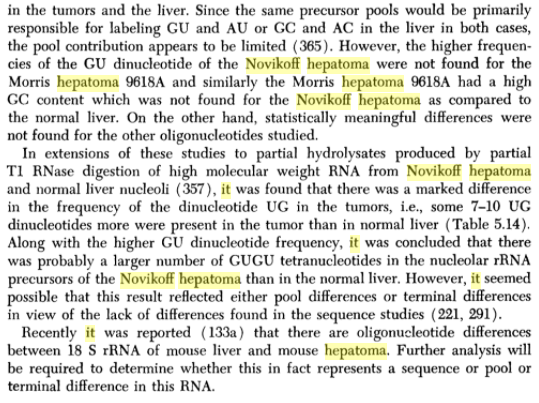
Basic introduction about the disease/process

Max 200 words. Add here few basic information about the disease/process under scrutiny.

DA RIASSUMERE

Novikoff hepatoma mitochondria were studied with regard to their functional and structural organization. The activities of cytochrome oxidase and malate dehydrogenase in mitochondria purified on sucrose gradients were similar in organelles derived from normal liver and hepatoma tissue. However, the activities of reduced nicotinamide adenine dinucleotide oxidase and succinate oxidase in hepatoma mitochondria were reduced to only 10 and 35%, respectively, of the activities in liver mitochondria. Also, monoamine oxidase and rotenone-insensitive reduced nicotinamide adenine dinucleotide-cytochrome c reductase (enzymes localized in the outer mitochondrial membrane) have significantly reduced activities in hepatoma mitochondria. The structural changes in hepatoma mitochondria might be correlated with differences in the banding patterns of liver and hepatoma mitochondria in sucrose gradients. While liver mitochondria banded sharply at a density of 1.187 g/ml, as evidenced by marker enzyme activity and protein assay, hepatoma mitochondria were heterogeneous, banding over a density range of 1.144 to 1.161 g/ml. The addition of 1% bovine serum albumin to the isolation medium increased the density of hepatoma mitochondria in sucrose gradients to 1.171 g/ml and resulted in sharp, homogeneous bands. This increase in density is due in part to protein binding, as evidenced by bovine serum albumin-14C-binding experiments. In addition, hepatoma mitochondria were shown to be heterogeneous by their separation into three density classes on discontinuous bovine serum albumin gradients.

Highly malignant tumor cells have been shown to be well coupled or to possess large numbers of gap junctions. These include Novikoff hepatoma cells. Moreover, other studies have analyzed gap junctions during tumor progression in skin tumors and in hepatocarcinoma and have found that, while gap junctions or coupling may disappear during the transformation of normal cells by cancer-causing agents (such as tumor promoters), their loss seems to follow, rather than lead the changes which transform these normal cells to tumor cells.



Seed genes

Explain briefly the methods you followed to get the information about the seed genes and add the related table (see table format below).

To collect the set of seed genes we started by filtering the “Curated gene-disease associations” dataset from DisGeNet in order to find all the genes associated with hepatoma, making sure they were all human genes.

Subsequently, we used the REST API of HGNC to fetch the status (approved or not) of each seed gene. All the 110 genes collected from DisGeNet resulted approved on HGNC, so we parsed the Uniprot dataset and collected the information requested in 1.1.b. We found that only 80 of the 110 seed genes resulted officially reviewed on Uniprot.

All the data collected has been stored in a *.tsv* file.

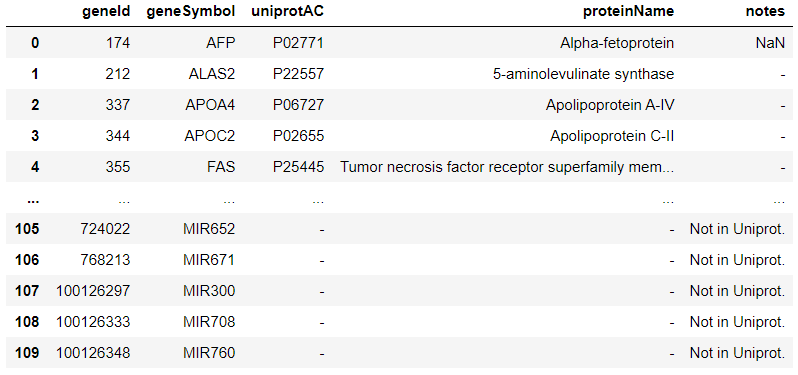


Table 1. Summary table of seed genes without protein description.

Summary on interaction data

Explain briefly the methods you followed to get the information about the interaction data and add the related table. Refer clearly to different files (i.e. when necessary.

Sample numbered list, if necessary.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

1. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

* Sample bullet list, if necessary.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Table 1.**Sample table. This should be the table format, add/remove columns and rows according to the data to be shown.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |S| | Predicted cost | Timing | Predicted speed | Speed |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

# Interactomes data

Explain briefly the methods you followed to build the intersection interactome and add the related tables/charts.

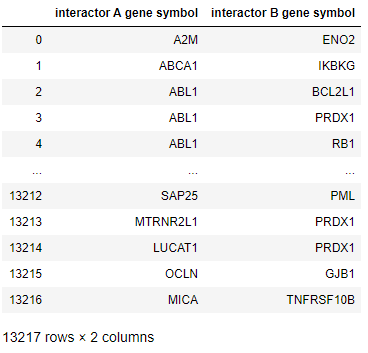
# Enrichment analysis

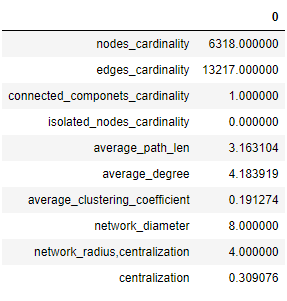
Explain briefly the methods you followed to carry out the enrichment analysis and add the related tables/charts.

**Table with the first 20 highest ranking genes for betweenness (include in the table**

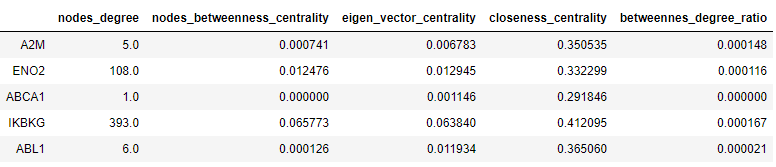
**also all other calculated centrality measures as from 1.2b) for the LCC**

We create the graph using a dataframe with the following informations.



The nodes cardinality is 6318 and the edges cardinality is 13217 (all the rows in the dataframe). No node is isolates, in fact the isolated nodes cardinality is 0 and the connected componets cardinality is 1.   
Since the whole graph is connected, the LCC is the graph itself.   
The other measures are in the table below.  


And here an example:



**Summary table of the putative disease modules found (*for each module: no. of seed***

***genes in each module, total no. of genes in each module, ratio no. seed genes/total***

***genes in the module, p-value of the enrichment using the hypergeometric test*)**

In this disease, using MCL we found 62 clusters (using the default hyperparameters: expansion=2, inflation=2, loop\_value=1, iterations=100, pruning\_threshold=0.001, pruning\_frequency=1, convergence\_check\_frequency=1), and 53 of them are longer than 10.

In the following table the final results of the method:

AGGIUNGERE TABELLA

# Notes and comments

References (if any, this is the format to be used)

Alexandrescu,A. (2001) Modern C++ Design: Generic Programming and Design Patterens Applied. Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.