Differential Analyses of Gene Expression

Sexual Dimorphism in Liver Hepatocellular Carcinoma

Digital Epidemiology and Precision Medicine

Di Nino Leonardo, Di Poce Giuseppe

Academic Year 2023/2024

SAPIENZA
UNIVERSITÀ DI ROMA

This project aims at conducting a study on sexual dimorphism in liver hepatocellular carcinoma leveraging some tools from bioinformatics. In particular, the purpose was to identify Differentially Expressed Genes between cancer condition and control ones in both male and female patients, trying to pick out patterns and structures arising between the two groups. Further tools from network science have been deployed to provide deeper investigations, paying special attention to co-expression network and patient similarity networks: the analysis of the degree distribution and the community structures on these networks unveils properties only encoded in pairwise relations between entities.

Table of Contents

1 Materials and methods

▶ Materials and methods

▶ Results and discussions

Data

1 Materials and methods

- TCGA-LIHC repository from on The Cancer Genome Atlas
- Gene expression for transcriptome profiling
- Four experimental conditions: male, female, cancer tissue, normal tissue
- DESeq2 normalization pipeline using *median of ratios* method applied within each condition

Differentially Expressed Genes (DEGs)

- *Fold change*, which quantifies differences in expression for a gene G across two different experimental conditions:

$$FC(G)_M = \log_2 \frac{\sum_{i=1}^{|M|} e(G)_{C,i}}{\sum_{i=1}^{|M|} e(G)_{N,i}}, FC(G)_F = \log_2 \frac{\sum_{j=1}^{|F|} e(G)_{C,j}}{\sum_{j=1}^{|F|} e(G)_{N,j}}$$

- p-value for t-test between averages across patients between the two conditions for each gene
- We considered a gene to be differentially expressed between conditions C and conditions N if $|FC(G)| \geq 1.2$ and $p(G) \leq 0.05$.

Differentially co-expressed network

We retrieved differential co-expression network related to all different cross-conditions:

- Cancer compared with normal expression for male population;
- Cancer compared with normal expression for female population;
- Male expression compared with female expression under cancer condition;
- Male expression compared with female expression under normal condition;

Differentially co-expressed network

1 Materials and methods

The procedure to retrieve these network has been the following:

- Compute the correlation matrix between data under hypotesis A and data under hypotesis B;
- Apply Fisher z-transform:

$$z_{i,j} = \frac{1}{2} \log \frac{1 + \rho_{i,j}}{1 - \rho_{i,j}} \tag{1}$$

- Compute Z-scores:

$$Z_{i,j} = \frac{z^A_{i,j} - z^B_{i,j}}{\sqrt{\frac{1}{n^A - 3}} + \sqrt{\frac{1}{n^B - 3}}}$$

- Threshold the Z-scores to retrieve an adjacency matrix defining the graph of interest $A_{i,j} = 1(|Z_{i,j}| \geq \overline{Z})$.

Patients Similarity Network

1 Materials and methods

Under cancer condition, we build a data collecting all patients irrespectively to their gender and collected the correlation matrix among patients.
Then we thresholded it $A_{i,j} = 1(|C_{i,j}| \geq \overline{C})$ to retrieve an adjacency matrix. We applied Louvain clustering algorithm to it in order to identify communities in the network.

Table of Contents

2 Results and discussions

▶ Materials and methods

▶ Results and discussions

Differentially Expressed Genes (DEGs)

2 Results and discussions

From our analysis on differential expression:

- We retrieved 67 DEGs among the male population (31 are up-regulated, 36 are down regulated);
- We retrieved 61 DEGs among the female population (34 are up-regulated, 27 are down regulated).

ENSG00000004799.8	ENSG00000067082.15	ENSG00000073756.12	ENSG00000091879.14	ENSG00000095002.15	ENSG00000097046.13
ENSG00000101412.13	ENSG00000104635.15	ENSG00000106366.9	ENSG00000106462.11	ENSG00000106993.12	ENSG00000107566.14
ENSG00000107864.15	ENSG00000109805.10	ENSG00000113594.10	ENSG00000114346.14	ENSG00000116128.11	ENSG00000117399.14
ENSG00000117632.23	ENSG00000125257.16	ENSG00000130164.14	ENSG00000131747.15	ENSG00000134243.12	ENSG00000134294.14
ENSG00000136158.12	ENSG00000138160.7	ENSG00000138180.16	ENSG00000139318.8	ENSG00000140044.13	ENSG00000147889.18
ENSG00000150907.10	ENSG00000154065.17	ENSG00000155090.15	ENSG00000164045.12	ENSG00000164761.9	ENSG00000165757.9
ENSG00000171848.16	ENSG00000176597.12	ENSG00000177606.8	ENSG00000178999.13	ENSG00000185652.12	ENSG00000187498.16

Table: List of DEGs overlapping between men and women data

Differentially Expressed Genes (DEGs)

2 Results and discussions

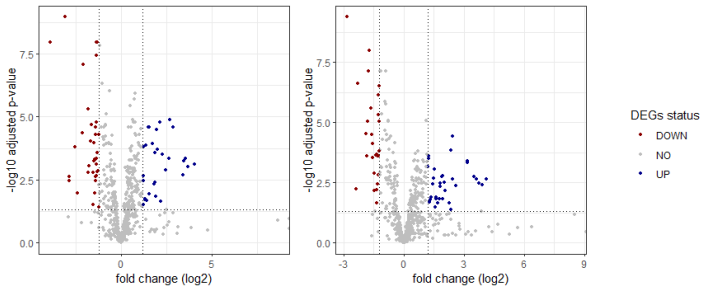


Figure: Volcano plot for DEGs with respect to fold change and p-value for t-statistics between cancer and normal tissue for male patients (on the left) and female patients (on the right)

Differentially co-expressed network

2 Results and discussions

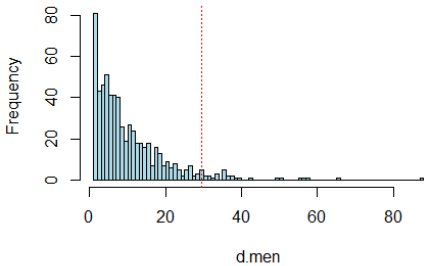


Figure: Male network degree distribution;

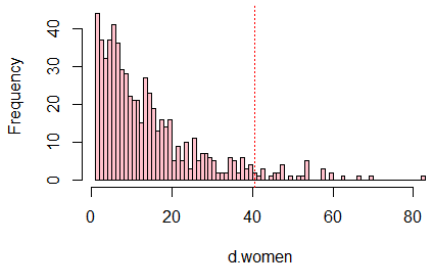


Figure: Female network degree distribution;

Differentially co-expressed network

2 Results and discussions

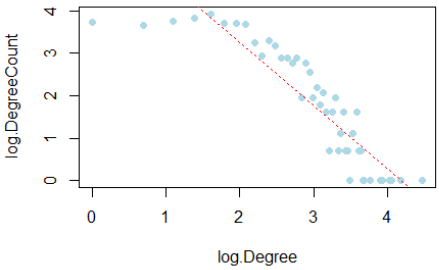


Figure: Male degree distribution in log-log scale

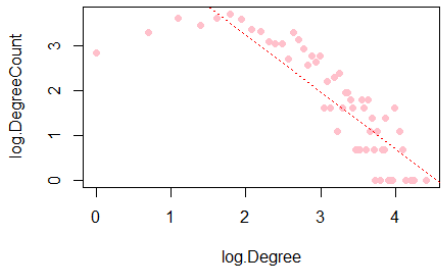


Figure: Female degree distribution in log-log scale

Differentially co-expressed network

2 Results and discussions

ENSG00000059758.8	ENSG00000069702.11	ENSG00000100644.17	ENSG00000101187.16	ENSG00000101413.12	ENSG00000105835.12
ENSG00000106993.12	ENSG00000108799.13	ENSG00000109320.13	ENSG00000113916.18	ENSG00000114861.23	ENSG00000115758.13
ENSG00000116984.14	ENSG00000117500.13	ENSG00000124216.4	ENSG00000126261.13	ENSG00000134294.14	ENSG00000136158.12
ENSG00000136448.13	ENSG00000137193.14	ENSG00000138685.17	ENSG00000138814.17	ENSG00000147649.10	ENSG00000152661.9
ENSG00000153147.6	ENSG00000155096.14	ENSG00000164961.16	ENSG00000165312.6	ENSG00000169398.19	ENSG00000173812.11
ENSG00000177606.8	ENSG00000184203.8	ENSG00000189376.12			

Table: List of hubs in male data

ENSG00000068878.15	ENSG00000076826.10	ENSG00000091136.15	ENSG00000095002.15	ENSG00000103222.20	ENSG00000107560.12
ENSG00000113594.10	ENSG00000113916.18	ENSG00000115392.12	ENSG00000116127.19	ENSG00000116128.11	ENSG00000120008.16
ENSG00000124496.12	ENSG00000136379.12	ENSG00000137193.14	ENSG00000138078.16	ENSG00000139793.18	ENSG00000140443.15
ENSG00000141582.15	ENSG00000144580.14	ENSG00000147251.15	ENSG00000156515.24	ENSG00000162711.18	ENSG00000164331.10
ENSG00000164961.16	ENSG00000169554.22	ENSG00000171492.14	ENSG00000171791.14	ENSG00000173801.17	ENSG00000180447.7
ENSG00000182481.10	ENSG00000185652.12	ENSG00000196628.20	ENSG00000198218.11		

Table: List of hubs in female data

Differentially co-expressed network

2 Results and discussions

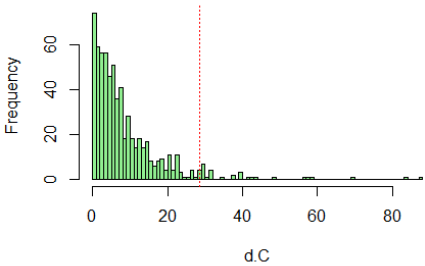


Figure: Cancer degree distribution;

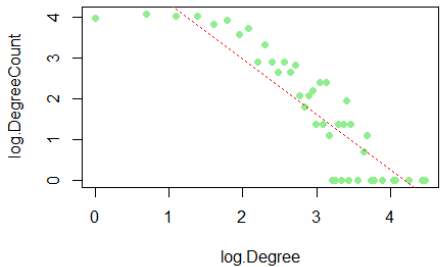


Figure: Cancer degree distribution in log-log scale;

Differentially co-expressed network

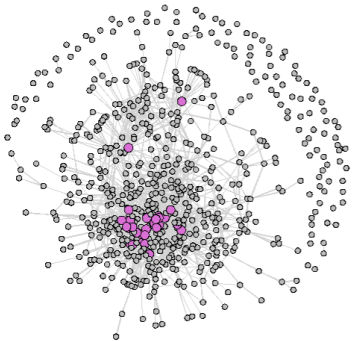
2 Results and discussions

ENSG00000070718.12	ENSG00000073921.18	ENSG00000076826.10	ENSG00000078177.14	ENSG00000085871.9	ENSG00000091136.15
ENSG00000095002.15	ENSG00000099203.7	ENSG00000099783.12	ENSG00000105976.16	ENSG00000111361.13	ENSG00000117500.13
ENSG00000132341.12	ENSG00000134250.20	ENSG00000136068.15	ENSG00000149136.9	ENSG00000151233.11	ENSG00000157404.16
ENSG00000158270.12	ENSG00000163249.13	ENSG00000164062.13	ENSG00000164331.10	ENSG00000164466.13	ENSG00000166598.15
ENSG00000169710.9	ENSG00000170365.10	ENSG00000170558.10	ENSG00000176171.11	ENSG00000196628.20	ENSG00000213551.7
ENSG00000229807.12	ENSG00000239264.9				

Table: List of female-male hubs in DEnetwork under cancer condition

Differentially co-expressed network

2 Results and discussions



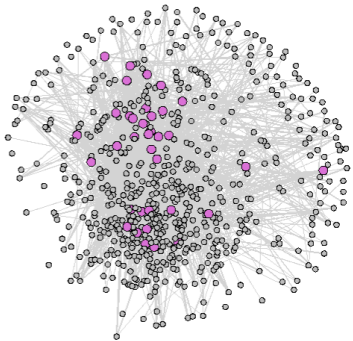


Figure: Positive (left) and negative (right) signed network;

Differentially co-expressed network

2 Results and discussions

ENSG00000070718.12	ENSG00000136068.15	ENSG00000151233.11	ENSG00000169710.9
ENSG00000170558.10	ENSG00000213551.7	ENSG00000229807.12	

Table: Hubs overlapping between positive and negative subnetworks

Patients Similarity Network

2 Results and discussions

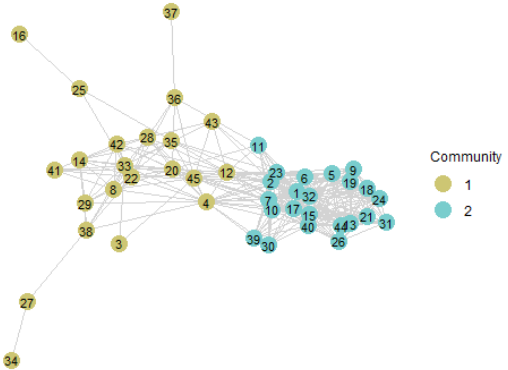


Figure: Patient similarity network highlighting the two communities retrieved by Louvain algorithm

Patients Similarity Network

2 Results and discussions

	Alive	Dead	Stage I	Stage $>$ II	Female	Male
Community 1	40.91 %	59.09 %	26.32 %	73.68 %	59.09 %	40.91 %
Community 2	21.74 %	73.91 %	55.56 %	44.44 %	26.09 %	73.91 %

Table: A table with some insights on the two communities (inconsinstencies might be due to misreporting)