DIFFERENTIAL GENE EXPRESSION ANALYSIS

THYROID CANCER

Cruoglio Antonella Mascolo Davide Napoli Mario

Digital Epidemiology and Precision Medicine - A.Y. 2022/2023
MSc in Data Science - Sapienza University of Rome

TABLE OF CONTENTS

1

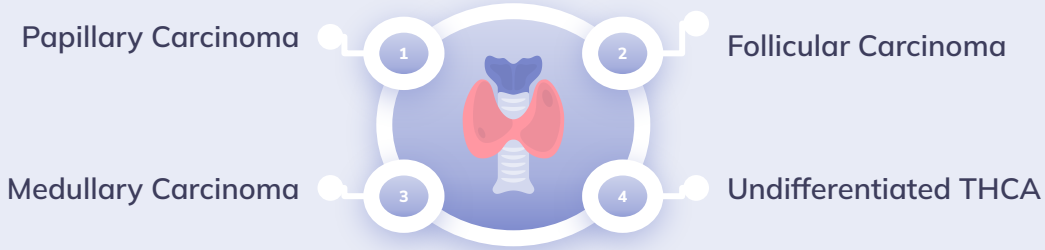
INTRODUCTION

The goal of the project and the involved analysis

THCA

Thyroid carcinoma (THCA) is a common endocrine malignant cancer, with an incidence rate that is increasing over the years.

THCA is mainly divided into four categories:



CAUSES AND TREATMENTS

The development of this tumor may depend by different factors among which a **genetic change** of the THCA.
For this reason, it has become very important in recent years to study the genetic expression to find some characteristics that could be used to improve the **therapies**.

The goal of the study is to analyze the expression of 18323 genes over 59 patients using RNA sequencing data in order to identify hub genes involved in this disease.

WORKFLOW OF THE ANALYSIS

DEGs

Identify subset of genes respecting a specific level of significance (p-value) and a specific threshold (FC).

1

2

Co-Expression Network

Build networks using a given similarity measure between pair of genes and using a threshold (Hard/Soft).

Differential Co-Expression

Build network considering the different expression in both groups.

3

4

Patient Similarity Network

Compute PSN and perform Community Detection

2

DATA

Gene Expression Data

DATA COLLECTION

Using The Cancer Genome Atlas (TCGA), we extracted two datasets: one about **tumor tissue** and the other about **normal tissue**.

Each row of the datasets represents the expression of a gene registered for different patients that are represented by the columns

PRE-PROCESSING

- Consider only the common patients and common genes between the two conditions.
- Remove the genes that were not expressed with a significant level.
- Check that there were no missing values.

3

DEGs

Differentially Expressed Genes

GOAL

Identify Differentially Expressed Genes:
those genes with a statistically significant change in expression between Tumor and Normal condition.

TOOL

R, “DESeq2” package.

ANALYSIS

Input: genes for which there are enough reads (at least 10 reads)

Criteria:

- $|LFC| \geq 1.2$.
- P-value ≤ 0.05 with FDR (Benjamini-Hochberg) adjustment.

RESULTS

- We obtained 669 genes, of which **520 up-regulated** genes and **149 down-regulated** genes.

4

CO-EXPRESSION NETWORK

Significant co-expression relationship between genes

GOAL

Build a network in which the nodes are genes activities and edges represent significant associations between them.

ANALYSIS

Input: DEGs (669 genes)

- Compute Correlation matrices for both condition
  - Pearson
  - Spearman
- Build Adjacency matrices
  - Hard-thresholding ($|\rho| \geq 0.7$)
  - Soft-thresholding (power adjacency function).

SOFT-THRESHOLDING

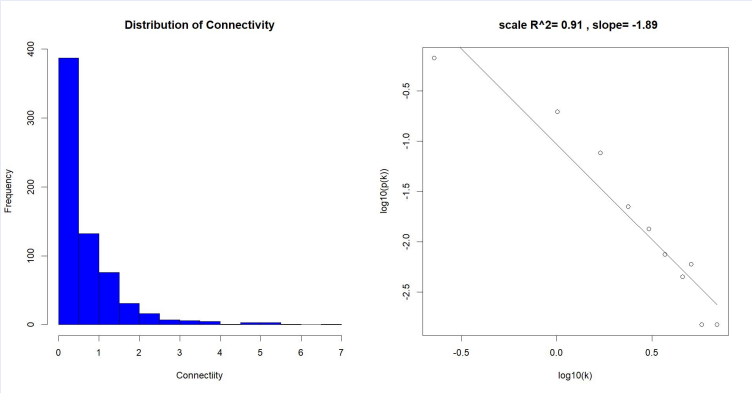
$$A_{ij} = power(s_{ij}, \beta) = |s_{ij}|^{\beta}$$

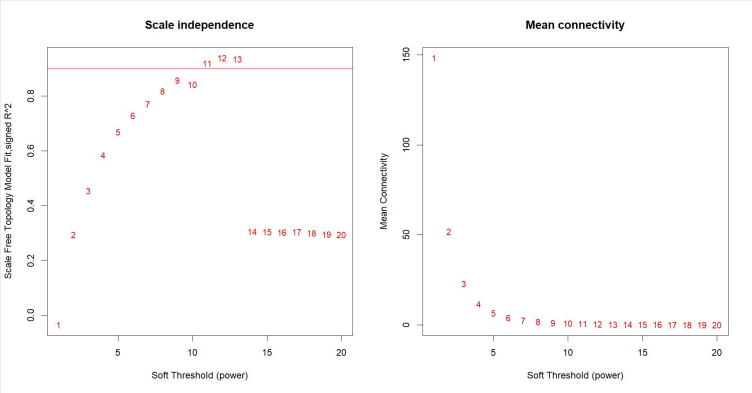
TOOL R package ”WGCNA”

RESULTS

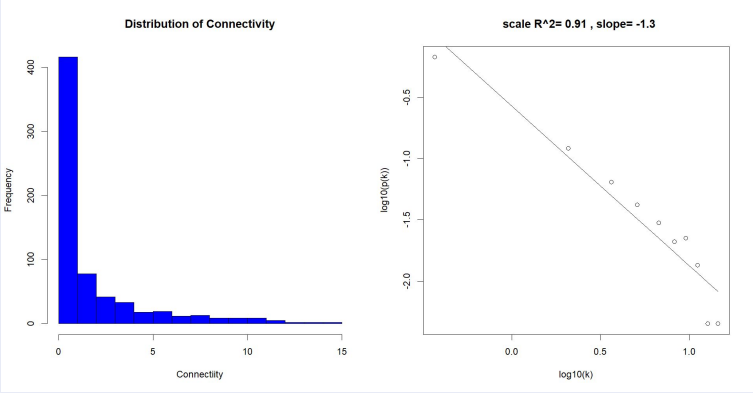
- Cancer: **β = 11** with a scale free **R^2 = 0.92**.
- Normal: **β = 8** with a scale free **R^2 = 0.91**.

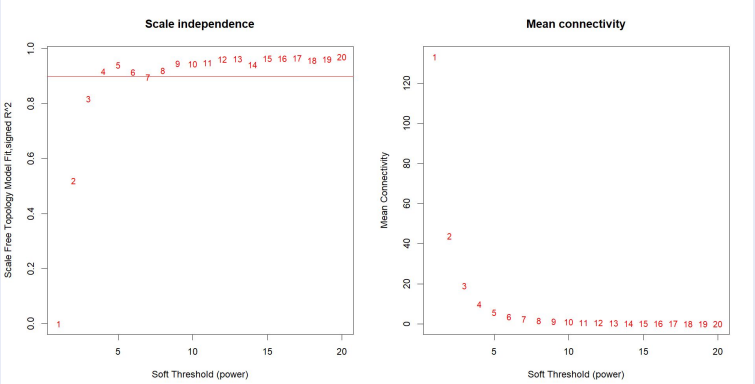
Cancer





Normal





DIFFERENT CENTRALITY INDEX

Compute different centrality indices and check the overlap between the 5% of the nodes with highest CI values and the degree-based hubs.

CI	Common Genes
Betweenness	CLIP3, NECTIN4, PDLIM4, SLC4A4, LGALS3, RIN1
Closeness	ITGA3, ENTPD2
Eigenvector	MIR31HG, CLIP3, MDK, NPAS1, VAC14-AS1, PROC, CCDC33, TMEM58L, DOK7, CASC15, ENTPD2, SPOCD1, LINC02981, PIANP, RIN1, C1QTNF12

Table 1: Overlap for cancer condition

CI	Common Genes
Betweenness	KCNN4, SLC1A5, STING1, RUNX1, HES6, CDKN2A, PDZK1IP1, DUSP4, STRA6, SYTL1
Closeness	NFE2L3, ALOX5, BUD, RASGRF1
Eigenvector	NFE2L3, KCNN4, TMC6, SLC1A5, STING1, ALOX5, CTSH, BID, WNT10A, RASGRF1, RUNX1, SPOCK2, DUSP4, STRA6, SYTL1, TMEM163

Table 2: Overlap for normal condition

DIFFERENT SIMILARITY MEASURE

Spearman Correlation

Condition	Common Genes
Cancer	KRT19, TMPRSS4, ITGA3, EVA1A, NECTIN4, PERP, RUNX1, SLC4A4
Normal	NFE2L3, KCNN4, TMC6, STING1, ALOX5, RUNX1, DUSP4

Table 3: Intersection between Pearson and Spearman Hubs

The **hubs** characterizing only the cancer network are:

TMPRSS4, FN1, CD55, ELF3, EVA1A, GRB7, AHNAK2, KRT19, CRYBG2, MUC1, TMPRSS6, ITGA3, MRO, PERP, B3GNT3, MPPED2, SERPINA1, NECTIN4, SLC4A4, MET, ERBB3.

4

DIFFERENTIAL CO-EXPRESSION NETWORK

Significant co-expression change

GOAL

Test if the change in co-expression is significant by encoding the changes in connections among nodes between the two conditions.

ANALYSIS

- **Fisher Z-transformation** on the correlation coefficients in each condition (z1 and z2)

$$z_{1or2} = \frac{1}{2} ln \frac{1+\rho_{1or2}}{1-\rho_{1or2}}$$

- Compute the **overall Z-Score**

$$Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}, \quad where\ n_i = sample\ size\ for\ the\ condition\ i.$$

- Build adjacency matrix with a **threshold** of 3.

5

PATIENT SIMILARITY NETWORK

Community Detection

GOAL

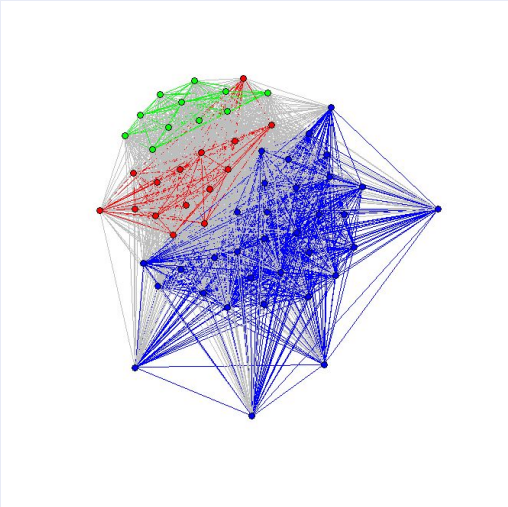
Compute Patient Similarity Network and perform Community Detection.

ANALYSIS

- Euclidean Distance transformed into a similarity measure
- **Louvain algorithm** to make community detection

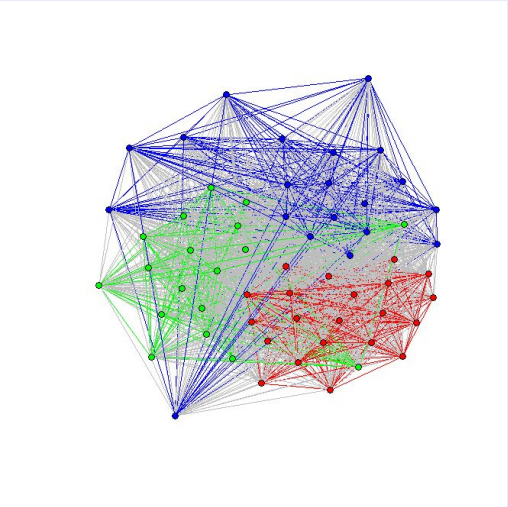
RESULTS

Cancer



Modularity = 0.12

Normal

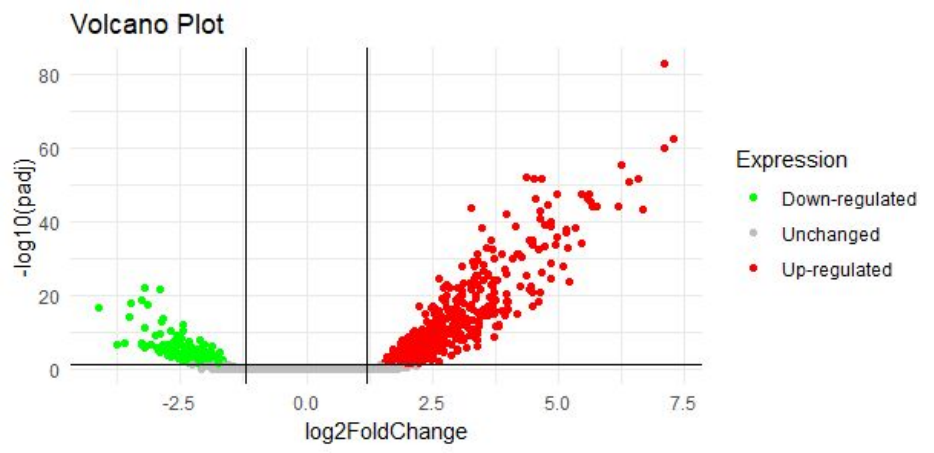


Modularity = 0.082

6

RESULTS AND DISCUSSION

VOLCANO PLOT

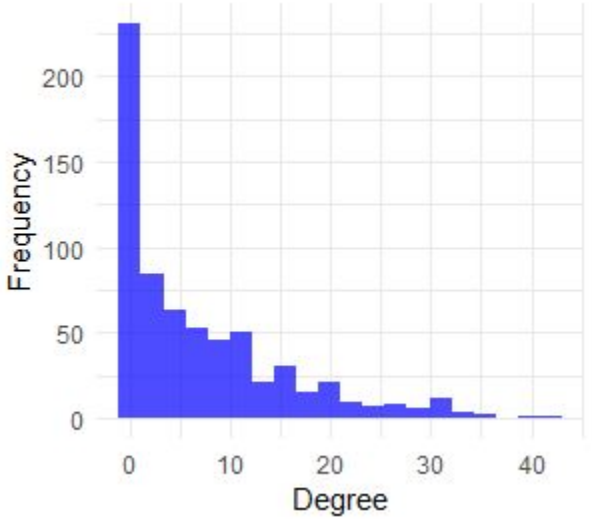


669 DEGs:

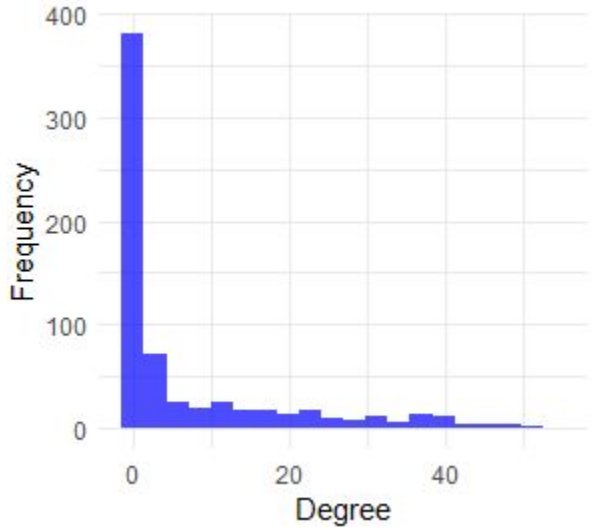
- **520 up-regulated**
- **149 down-regulated**

DEGREE DISTRIBUTION FOR CO-EXPRESSION NETWORKS

Cancer



Normal



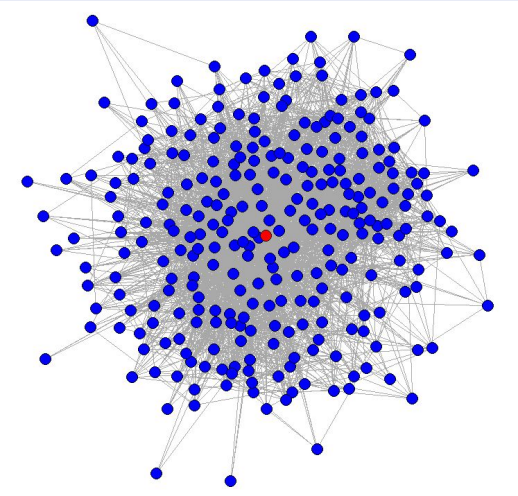
HUBS OF THE CO-EXPRESSION NETWORK

Condition	Hubs Genes
Cancer	MIR31HG, CLIP3, MDK, NPAS1, KRT19, VAC1-AS1, PROC, TMPRSS4, CCDC33, TMEM59L, ITGA3, DOK7, CASC15, ENTPD2, EVA1A, SPOCD1, NECTIN4, LINC02981, PERP, PDLIM4, PIANP, RUNX1, SLC4A4, LGALS3, RIN1, C1QTNF12

Table 5: Hubs characterized Cancer condition

We have 26 hubs for the cancer network and 23 hubs for the normal network and they have only one hub in common, that is: **RUNX1**.

DIFFERENTIAL CO-EXPRESSION NETWORK



The most connected hub is: **PRSS22**, that is a gene that encodes a member of the trypsin family of serine proteases.

These are the **hubs** in the differential co-expressed network:

PRSS22, ITGA3, PSD, ERBB3, RSPO4, MET, GALNT7, PERP, EVA1A, ECE1, TGFBI, APOE, NPAS1, ARNTL, C5AR2, ENTPD1, PDE5A, GRB7, XPR1, MXRA8, BNIPL, LMOD1, SPINT1, SCG5, PIP5KL1, ZFPM2, KCNS3, ETV4, H2AW, DMD, NPTXR, ELFN1, MEX3A, KCNJ2-AS1.

RELEVANT GENES AND REFERENCE PAPERS

- **MIR31HG**: is the most connected hub for the co-expression cancer network and it is over-expressed in the human thyroid cancer. [4]

- **MDK**: also this gene is over-expressed in the human thyroid cancer and is the third most connected hub for the co-expression cancer network. [5]

- **SERPINA1**, **FN1**: these two genes characterize the co-expression cancer network obtained with the Spearman correlation. [6][7]

- **ITGA3**: this is a very interesting gene and it is central to the development of the disease. It characterizes the co-expression cancer network using both correlations (Pearson and Spearman) and also using the Closeness centrality. Furthermore, it is also present in the differential co-expressed hubs. [8]

RELEVANT GENES AND REFERENCE PAPERS

- **LGALS3**: this gene is one of the most connected hubs in the co-expression cancer network and with higher betweenness. [7]

- **MET**: this gene is present in the co-expression cancer hubs using Spearman correlation and in the differential co-expression network hubs. [7]

- **SLC4A4**: this gene is down-regulated and it is present in the co-expression cancer network using both correlations and is also one of the genes with highest betweenness centrality measure. [9]

REFERENCES

[1] Kim K, Jeon S, Kim TM, Jung CK. Immune Gene Signature Delineates a Subclass of Papillary Thyroid Cancer with Unfavorable Clinical Outcomes. Cancers (Basel), 2018 Dec 5.

[2] Liu R, Cao Z, Pan M, Wu M, Li X, Yuan H, Liu Z. A novel prognostic model for papillary thyroid cancer based on epithelial-mesenchymal transition-related genes. Cancer Med, 2022 May 24.

[3] Wang Y, Huang H, Hu F, Li J, Zhang L, Pang H. CITED1 contributes to the progression of papillary thyroid carcinoma via the Wnt/β-catenin signaling pathway. Onco Targets Ther, 2019 Aug 21.

[4] Peng S, Chen L, Yuan Z, Duan S. Suppression of MIR31HG affects the functional properties of thyroid cancer cells depending on the miR-761/MAPK1 axis. BMC Endocr Disord, 2022 Apr 20.

[5] Jee YH, Celi FS, Sampson M, Sacks DB, Remaley AT, Kebebew E, Baron J. Midkine concentrations in fine-needle aspiration of benign and malignant thyroid nodules. Clin Endocrinol (Oxf), 2015 Dec.

[6] Vierlinger K, Mansfeld MH, Koperek O, N¨ohammer C, Kaserer K, Leisch F. Identification of SERPINA1 as single marker for papillary thyroid carcinoma through microarray meta analysis and quantification of its discriminatory power in independent validation. BMC Med Genomics, 2011 Apr 6.

[7] Huang, Ying, et al. “Gene Expression in Papillary Thyroid Carcinoma Reveals Highly Consistent Profiles.” Proceedings of the National Academy of Sciences of the United States of America, vol. 98, no. 26, 2001.

[8] Zhang G, Li B, Lin Y. Evaluation of ITGA3 as a Biomarker of Progression and Recurrence in Papillary Thyroid Carcinoma. Front Oncol, 2022 Jan 31.

[9] Huang Y, Ling J, Chang A, Ye H, Zhao H, Zhuo X. Identification of an immune-related key gene, PPARGC1A, in the development of anaplastic thyroid carcinoma: in-silico study and in-vitro evaluation. Minerva Endocrinol (Torino), 2022 Jun.

THANK YOU FOR THE ATTENTION