

**M2 BIM - RESYS Project:   
Master Regulators for Metastatic behavior in Osteosarcoma**

By Alexis TRANG and Yann ZHONG

Contents

[**Introduction** 1](#_Toc87890334)

[**Background** 2](#_Toc87890335)

[**The dataset and conclusions from the paper** 2](#_Toc87890336)

[**Master regulators signaling** 3](#_Toc87890337)

[**Integrative network analysis of regulators of genetic risk of breast cancers** 3](#_Toc87890338)

[**RedeR: a Bioconductor R package for representation of modular structures** 3](#_Toc87890339)

[**Methods** 3](#_Toc87890340)

[**Installations and preprocessing** 4](#_Toc87890341)

[**Using RTN to infer regulatory networks** 4](#_Toc87890342)

[**Using RedeR to visualize the preliminary output** 4](#_Toc87890343)

[**Discussion** 5](#_Toc87890344)

[**Bibliography** 5](#_Toc87890345)

# **Introduction**

**General description, copied from project prompt**:

An osteosarcoma (OS) or osteogenic sarcoma (OGS) (or simply bone cancer) is a cancerous tumor in a bone. Specifically, it is an aggressive malignant neoplasm that arises from primitive transformed cells of mesenchymal origin (and thus a sarcoma) and that exhibits osteoblastic differentiation and produces malignant osteoid. Osteosarcoma is the most common histological form of primary bone cancer. It is most prevalent in teenagers and young adults. Overall survival of patients with metastatic disease is approximately twenty percent. Mechanisms behind the development of metastases in osteosarcoma are unknown. To identify gene signatures that play a 4 role in metastasis, a study performed genome-wide gene expression profiling on pre-chemotherapy biopsies of osteosarcoma patients who developed metastases within 5yrs and patients who did not develop metastases within 5yrs.

In genetics, a master regulator is a gene at the top of a gene regulation hierarchy, particularly in regulatory pathways related to cell fate and differentiation. When analyzing the signature of a specific behavior in a disease, you can obtain the transcription factors that are master regulators for that phenomenon, that is, responsible for the behavior you see. In this case, we’re talking about metastatic behavior. RTN is an R package specialized at inferring gene regulatory networks, based on ARACNe.

**Step by step, as according to project description:**

1. Download [gene expression data](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse21257).
2. Install [RTN](https://www.bioconductor.org/packages/release/bioc/html/RTN.html) package, and infer regulatory network based on downloaded gene expression data.
3. Install [snow](https://cran.r-project.org/web/packages/snow/index.html) package for parallel processing in RTN.
4. Install [RedeR](https://bioconductor.org/packages/release/bioc/html/RedeR.html) package to better visualize inferred network.
5. [Perform differential gene expression analysis](https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/rna-sequencing/performing-a-rna-seq-experiment/data-analysis/differential-gene-expression-analysis/) between metastatic and non-metastatic biopsies, obtaining a signature for metastatic behavior of this cancer in the patients.
6. [Run a Master Regulator Analysis](https://rdrr.io/bioc/RTN/man/tna.mra.html) to infer putative Master Regulators by using inferred network and obtained signatures.
7. Get biological insight into the putative Master Regulators at the [Human Protein Atlas](https://www.proteinatlas.org/) and at [Gene Cards](https://www.genecards.org/).

# **Background**

The project revolves around using gene expression data downloaded from GEO (the Gene Expression Omnibus), a large online database for genomics data, notably data from microarray or high throughput sequencing experiments for use in determining gene expression activity.

In this project, the dataset we are interested in is the GSE21257 dataset, which corresponds to the following title: **“Genome-wide gene expression profiling on pre-chemotherapy biopsies of osteosarcoma patients who developed metastases within 5yrs (n=34) and patients who did not develop metastases within 5yrs (n=19)”**[1]. This dataset became public on the 11th of February 2011.

## **The dataset and conclusions from the paper**

The data we are looking at is the gene expression data from a total of 53 patients who were diagnosed with osteosarcoma (bone cancer) and who were due for chemotherapy. Of those patients, 34 developed metastases (the spread of the cancer throughout the rest of the body, which essentially indicates the development of the cancer) within 5 years, and 19 did not. The paper associated to this study is **Buddingh et al. (2011)**, and sequencing was done through Microarray technology (and not next generation sequencing, which is become increasingly the norm).

To analyze said data, Buddingh et al. performed DE (Differential Expression) analysis on the 53 patients – DE consists of looking at a control group (in this case, the patients who did develop metastasis) and a “differential” group (in this case the patients who did no develop metastasis). This analysis brings us to the fundamental concept of gene regulatory networks, notably upregulated and downregulated genes.   
 An **upregulated** cell will have its gene products such as RNA/proteins more strongly expressed and produced in response to an external stimulus, whereas a **downregulated** gene will have its gene products less expressed. A hallmark of cancer is also DNA damage, which is caused by the inability of the body to repair natural DNA mutations. Such inability is, in the case of cancers such as bladder cancer, stomach cancer, or thyroid cancer, caused by downregulation of the MGMT gene (a DNA repair gene). [2][3][4]

What was found was that from the two cohorts, there were a total of 139 significantly differentially expressed genes, with **125 being upregulated and 14 downregulated**. Example of such genes that the authors further confirmed experimentally through the gold standard of RT-qPCR are **CD14** and **HLA-DRA**, which were found to be barely present patients who did exhibit metastases. Of all of those, **roughly half (20+25 = 45%)** **were associated (strongly or indirectly) to upregulation of macrophage or immunological functions**.  
 These macrophages are mainly M2 macrophages but can also become M1 – going from tumor’s “friend” to “foe” (M1 is better). TAMs are as such a mix of M1 antitumor and M2 protumor macrophages, being a heterogeneous cell population. M2 “alternatively activated TAMs (Tumor Associated Macrophages)” are associated with worse patient prognostics in numerous cancer types. However, here, it seems that they have shown that the larger presence of infiltrating TAMs correlated with better survival odds for patients with osteosarcoma.

The paper’s authors conclude that that treatment with MTP (or L-MTP-PE): a macrophage activating agent, could potentially yield beneficial results relative to cancer treatment. The next step for us was then to try and visualize a regulatory network out of this dataset using RTN.

### **Master regulators signaling**

RTN is a Bioconductor package which was developed by **Fletcher et al. (2013)** [5], in an effort to visualize networks associated to genes and regulators that were likely to play a role in breast cancer development, more notably the FGFR2 (fibroblast growth factor receptor 2) locus and their respective Master regulators (MRs).

Primarily, what interests us in this paper is the rough background but mainly the methods, as RTN is what we will be using to infer a regulatory network for our own osteosarcoma dataset.

Some more stuff to add to background.

Regulon.

Probe.

### **Integrative network analysis of regulators of genetic risk of breast cancers**

**Castro et al. (2015)**. [6]

Risk TFs.

Master regulators.

Cluster identifications.

Figures 3 and 5.

### **RedeR: a Bioconductor R package for representation of modular structures**

**Castro et al. (2012)**. [7]

Why RedeR was designed, what it’s made to overcome.

Focus on examples that can benefit us.

System reaching equilibrium via the relax command.

Different modes of visualization (circle, hierarchical, etc.).

Clustering.

Illumina beadchip using nuIDs.

# **Methods and discussion**

The entirety of our work shall be presented in a R markdown file, which will also be attached a knitted html file.

### **Installations and preprocessing**

We start the code by installing BiocManager if required, and using this Bioconda package in order to install all the rest of the necessary packages in RTN, snow, RedeR, limma, Biobase and GEOquery. Following this, the packages are simply activated by usage of the *library* function.

Then, we need to preprocess the input data before it can be used as input to the RTN package functions. This code, which was provided to us, starts by getting GEO data remotely with the *getGEO* function from the GEOquery package.

Illumina beadchip using nuIDs.

Set names and stuff.

### **Using RTN to infer regulatory networks**

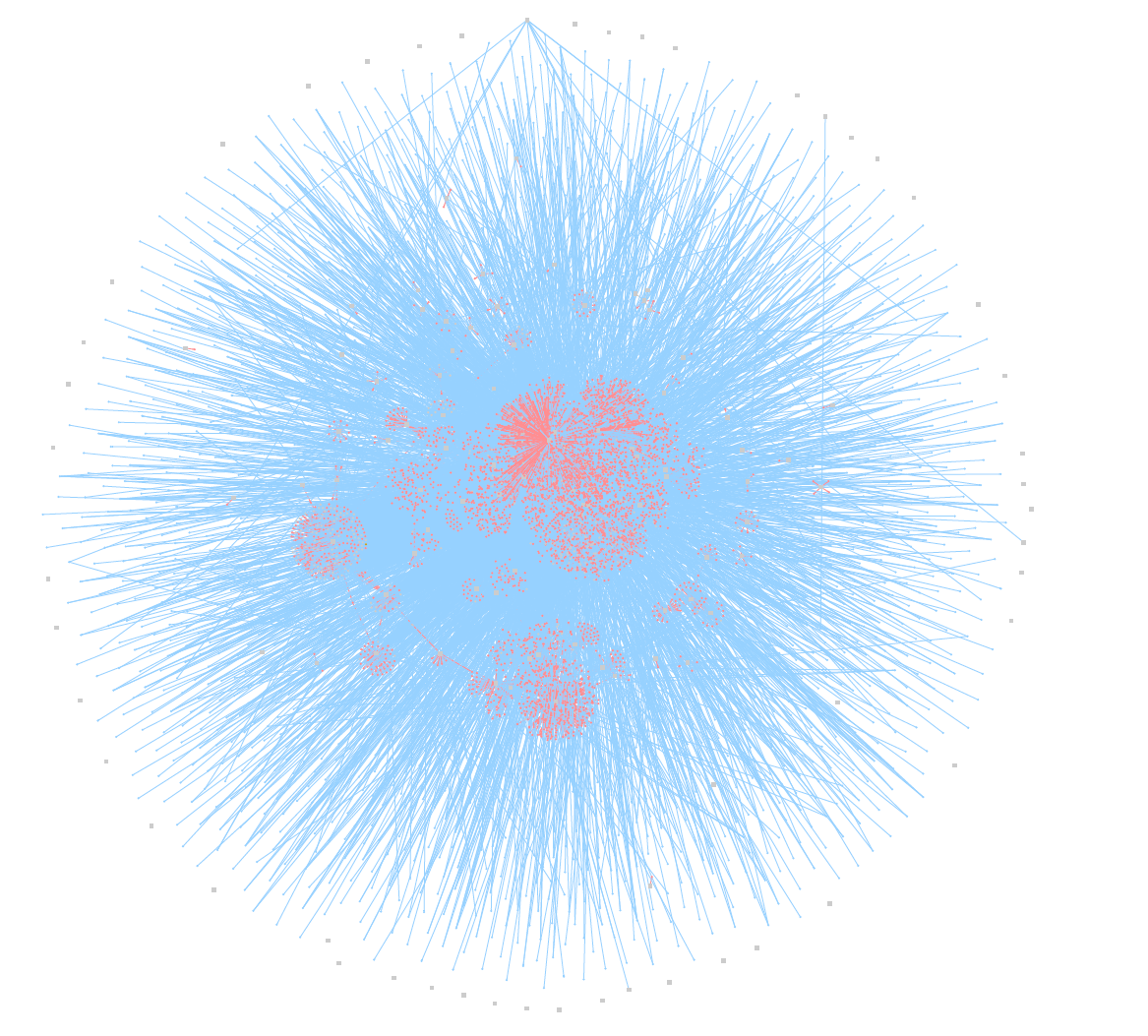
Once the input data has been successfully preprocessed, using the RTN vignette as a reference, we start inferring the regulatory networks. We start by running a permutation analysis (talk more about this) with 1000 permutations total, as the documentation recommended running this analysis with a number of permutations equal to greater than 1000.  
 Following this, we use two distinct methods to remove both unstable and weak interactions: bootstrap and ARACNE respectively (talk more about this).

To get a general overview of the inferred network, we run the tni package *summary* function on our network, indicating a network comprised of 140 regulons. (Analyze regulon summary output).

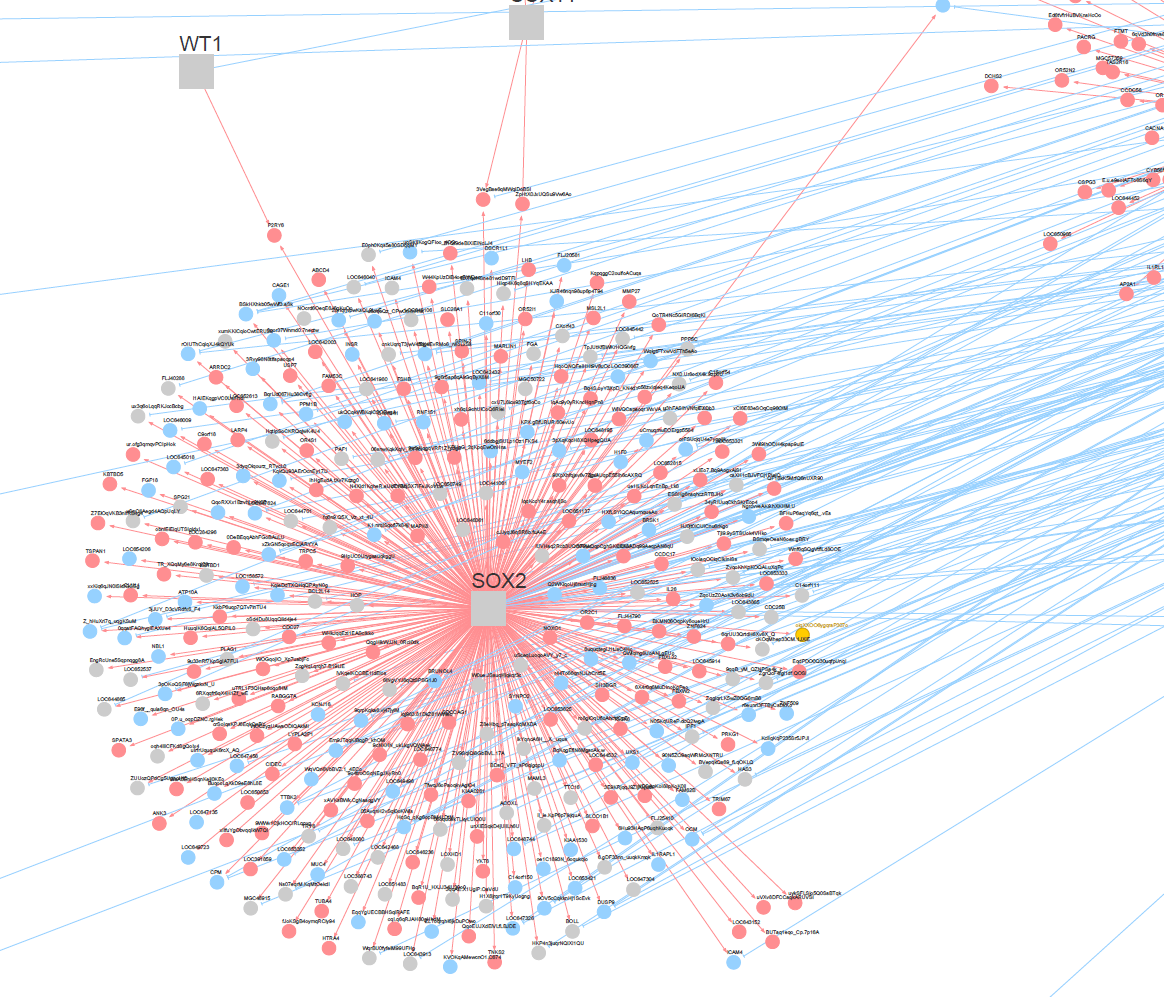
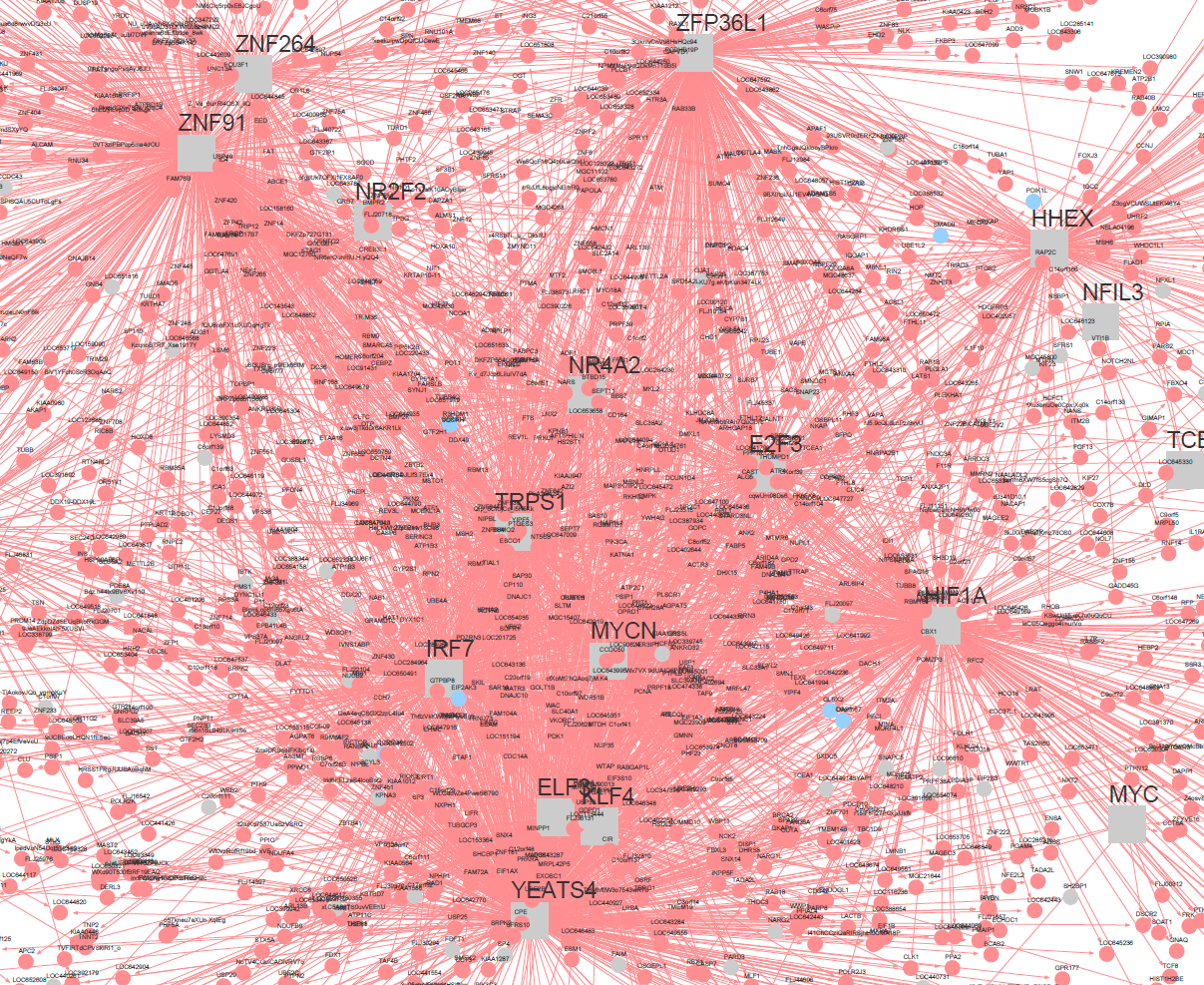
### **Using RedeR to visualize the preliminary output**

RedeR is a package which was designed to link R to Java in order to have an interactive network for us to visualize. In order to do so, an Apache server is built with the *RedPort* function and assigned to a rdp object. Then, we call the object to launch the RedeR interface and use *addGraph* as well as *addLegend* in order to plot out our inferred network object. The *relax* function [8] is used to allow the complex network edges to be considered as springs that can exert repulsive or attractive forces based on a set of parameters, such as target length, stiffness, repel factor, cooling factor…

In our case, we plot the full inferred network with the regulons all displayed on one graph, giving a very (very) global view of our network:

  
*Full view of inferred network after relaxation*

With a more zoomed in view on various parts:



*Left: zoom of central cluster (YEATS4, MYCN, etc.); Right: zoom of leftmost cluster SOX2*

In fact, looking at the *regulon* object in R-Studio interface gives us a good indicator of how many elements the top 10 genes actually have a link to, out of the 140:

|  |  |  |  |
| --- | --- | --- | --- |
| **Symbol** | **Full name** | **Type** | **Nb of links** |
| YEATS4 | YEATS Domain Containing 4 | Protein Coding Gene | 1580 |
| ZNF91 | Zinc Finger Protein 91 | Protein Coding Gene | 845 |
| HIF1A | Hypoxia Inducible Factor 1 Subunit Alpha | Protein Coding Gene | 709 |
| IRF7 | Interferon regulatory factor 7 | Protein Coding Gene - RF | 638 |
| SOX2 | SRY-Box Transcription Factor 2 | Protein Coding Gene - TF | 480 |
| PAX7 | Paired Box 7 | Protein Coding Gene - TF | 480 |
| TRPS1 | Transcriptional Repressor GATA Binding 1 | Protein Coding Gene - TF | 464 |
| MYCN | MYCN Proto-Oncogene, BHLH Transcription Factor | Protein Coding Gene | 419 |
| IRF8 | Interferon Regulatory Factor 8 | Protein Coding Gene - RF | 358 |
| ZNF264 | Zinc Finger Protein 264 | Protein Coding Gene | 342 |

\*: RF = Regulatory Factor; TF = Transcriptional Factor

The above information was procured from GeneCards[9], which is a comprehensive database that details the functions of genes as well as links and information relevant to said gene. Of the top 10 hits with the most links (be they positive – upregulated, in red, or negative – downregulated, in blue), 5 are simple genes that code for proteins (many of them commonly found in tumors or have oncogenic links), while 3 others are transcription factors and 2 are regulatory factors.

Find out how those top 10 hits are linked to the results from the paper.

### **Using limma and rta to infer the Master Regulators**

To add to when Alexis has managed?

# **Conclusion**

In conclusion, we got nice results (I hope).

# **Bibliography**

[1] E. P. Buddingh *et al.*, ‘Tumor-Infiltrating Macrophages Are Associated with Metastasis Suppression in High-Grade Osteosarcoma: A Rationale for Treatment with Macrophage Activating Agents’, *Clin Cancer Res*, vol. 17, no. 8, pp. 2110–2119, Apr. 2011, doi: 10.1158/1078-0432.CCR-10-2047.

[2] ‘Gene regulatory network’, *Wikipedia*. Oct. 19, 2021. Accessed: Nov. 13, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Gene\_regulatory\_network&oldid=1050666872

[3] ‘Downregulation and upregulation’, *Wikipedia*. Jun. 01, 2021. Accessed: Nov. 13, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Downregulation\_and\_upregulation&oldid=1026374055

[4] ‘Regulation of gene expression’, *Wikipedia*. Oct. 26, 2021. Accessed: Nov. 13, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Regulation\_of\_gene\_expression&oldid=1051856995

[5] M. N. C. Fletcher *et al.*, ‘Master regulators of FGFR2 signalling and breast cancer risk’, *Nat Commun*, vol. 4, no. 1, p. 2464, Dec. 2013, doi: 10.1038/ncomms3464.

[6] M. A. A. Castro *et al.*, ‘Regulators of genetic risk of breast cancer identified by integrative network analysis’, *Nat Genet*, vol. 48, no. 1, pp. 12–21, Jan. 2016, doi: 10.1038/ng.3458.

[7] M. A. Castro, X. Wang, M. N. Fletcher, K. B. Meyer, and F. Markowetz, ‘RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations.’, *Genome Biol*, vol. 13, no. 4, p. R29, 2012, doi: 10.1186/gb-2012-13-4-r29.

[8] ‘relax: relax in RedeR: Interactive visualization and manipulation of nested networks’. https://rdrr.io/bioc/RedeR/man/relax.html (accessed Nov. 15, 2021).

[9] ‘GeneCards - Human Genes | Gene Database | Gene Search’. https://www.genecards.org/ (accessed Nov. 15, 2021).