Gene Expression Analysis of Rheumatoid Arthritis using GEOQuery Dataset

Bioinformatics Course

Master's Degree in Computer Science

Francesco Fortunato (1848527)

Academic Year 2022/2023

SAPIENZA
UNIVERSITÀ DI ROMA

Table of Contents

Rheumatoid Arthritis

The disease

Rheumatoid Arthritis (**RA**) is an autoimmune disease, that typically causes damage to the synovial membrane, cartilage, and bone and includes a wide range of symptoms, such as painful and swollen joints

The dataset

Description

The expression profiles of **GSE93272** was downloaded from the **Gene Expression Omnibus** database with the **R language** thanks to the *GEOquery* library. The dataset includes gene expression profiling by array for whole blood samples from RA patients with or without drug treatments (methotrexate, infliximab, and tocilizumab), as well as healthy controls. It consists of **275 samples**, including **232 RA samples** and **43 healthy controls**.

The dataset

Excerpt

title	geo_accession	status	submission_	last_update	type	ch	source_name_organism_ch1	characteristics_ch1	characteristics_ch1.1
Whole blood from healthy control(HC003_1)	GSM2449608	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC003	disease state: healthy control
Whole blood from healthy control(HC004_1)	GSM2449609	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC004	disease state: healthy control
Whole blood from healthy control(HC005_1)	GSM2449610	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC005	disease state: healthy control
Whole blood from healthy control(HC006_1)	GSM2449611	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC006	disease state: healthy control
Whole blood from healthy control(HC007_1)	GSM2449612	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC007	disease state: healthy control
Whole blood from healthy control(HC008_1)	GSM2449613	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC008	disease state: healthy control
Whole blood from healthy control(HC009_1)	GSM2449614	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC009	disease state: healthy control
Whole blood from healthy control(HC010_1)	GSM2449615	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC010	disease state: healthy control
Whole blood from healthy control(HC011_1)	GSM2449616	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC011	disease state: healthy control
Whole blood from healthy control(HC012_1)	GSM2449617	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC012	disease state: healthy control
Whole blood from healthy control(HC013_1)	GSM2449618	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC013	disease state: healthy control
Whole blood from healthy control(HC015_1)	GSM2449619	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC015	disease state: healthy control
Whole blood from healthy control(HC016_1)	GSM2449620	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC016	disease state: healthy control
Whole blood from healthy control(HC017_1)	GSM2449621	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC017	disease state: healthy control
Whole blood from healthy control(HC018_1)	GSM2449622	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC018	disease state: healthy control
Whole blood from healthy control(HC019_1)	GSM2449623	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC019	disease state: healthy control
Whole blood from healthy control(HC020_1)	GSM2449624	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC020	disease state: healthy control
Whole blood from healthy control(HC021_1)	GSM2449625	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC021	disease state: healthy control
Whole blood from healthy control(HC022_1)	GSM2449626	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC022	disease state: healthy control
Whole blood from healthy control(HC023_1)	GSM2449627	Public on Jul 16 2018	Jan 06 2017	Jul 16 2018	RNA	1	Whole blood f	Homo sapiens	individual id: HC023	disease state: healthy control

Figure: metadata.txt

Table of Contents

2 The analysis

The analysis

Overview

The analysis conducted includes several steps:

- **Pre-processing**
- **Filtering**
- **Statistical significance**
- **Principal Component Analysis**
- **Functional Enrichment Analysis**
- **miRNA**
- **Literature-based research**

More in detail, Pre-processing, Filtering, Statistical significance, PCA and Functional Enrichment Analysis were conducted with the **R language**. The source code can be reachable at this GitHub repository. The miRNA Analysis, instead, were conducted thanks to the **miRTarBase**. Let's go deeper in the analysis.

Table of Contents

Pre-processing

3 Pre-processing, Filtering and Statistical Significance

- Total number of genes: **23520**
- Mean filtering:
  - the search of the overall mean equal to 0 didn't return any gene
- Logarithmic transformation:
  - $\log_2(data + 1)$
- IQR:
  - the IQR value is applied for each gene.

In order to not discard too many genes in this phase, the **10th** percentile of the IQR was chosen, passing from **23520** to **21168** genes.

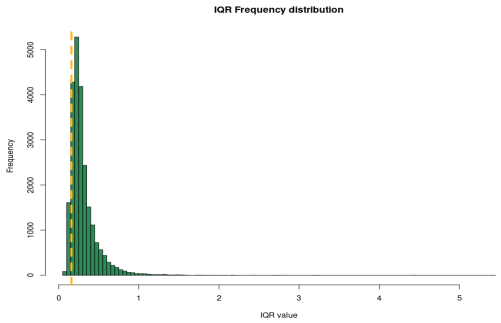


Figure: IQR frequency distribution

Fold Change Filtering

In the data filtering step, the **Log Fold Change** is calculated to assess the significance of gene expression changes. Based on the values observed in the graph, a LogFC threshold of $\log_2(1.2)$ was chosen as a reasonable cutoff. After applying this threshold, the number of genes remaining in the analysis is **1939**.

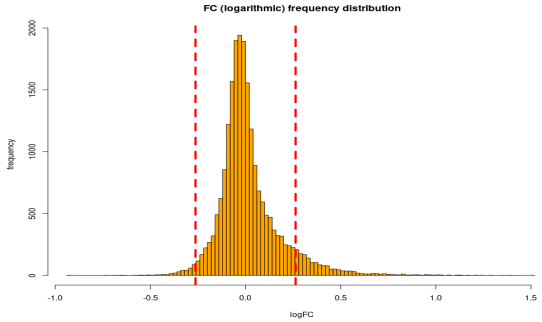


Figure: LogFC frequency distribution

Statistical Significance

For the analysis, the assumption that the data follows a **normal distribution** was considered due to a sufficiently large number of samples ($n > 30$), **from the central limit theorem**. To compare the gene expression between the case and control groups, a **Student's t-test** were performed.

The t-test allows to test two hypotheses:

- **Null Hypothesis**: The gene expression levels in the case and control groups are drawn from the same underlying distribution, and their means are equal.
- **Alternative Hypothesis**: The gene expression levels in the case and control groups are from different distributions, and their means are different.

To determine the statistical significance, the **p-values** was computed for each gene and it was applied the **Benjamini-Hochberg (BH) correction**, also known as the **False Discovery Rate (FDR)** correction. A significance level of $p < 0.05$ was chosen, resulting in the rejection of the null hypothesis for **9 genes**.

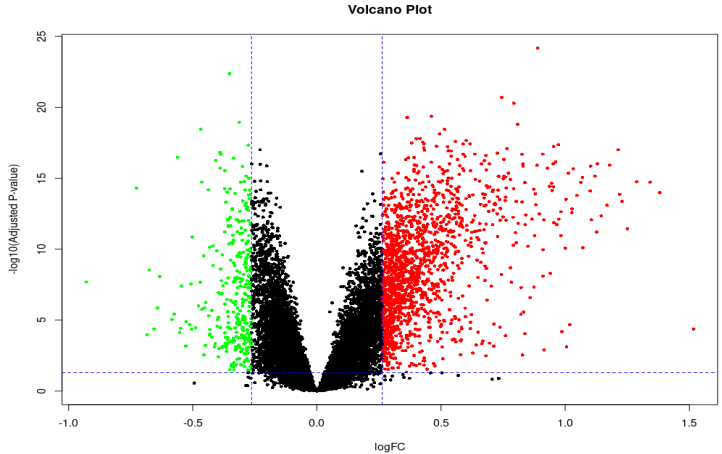
Visualizing the Data

Volcano plots

The following two slides represents the **Volcano plots** of the differentially expressed genes between RA and healthy control before and after this first step of the analysis. **Black points** represent the adjusted p-value $> 0.05 \vee |logFC| < \log_2(1.2) = 0.263$. **Green points** represent adjusted p-value < 0.05 and **down-regulated** genes. **Red points** represent adjusted p-value < 0.05 and the **up-regulated** genes.

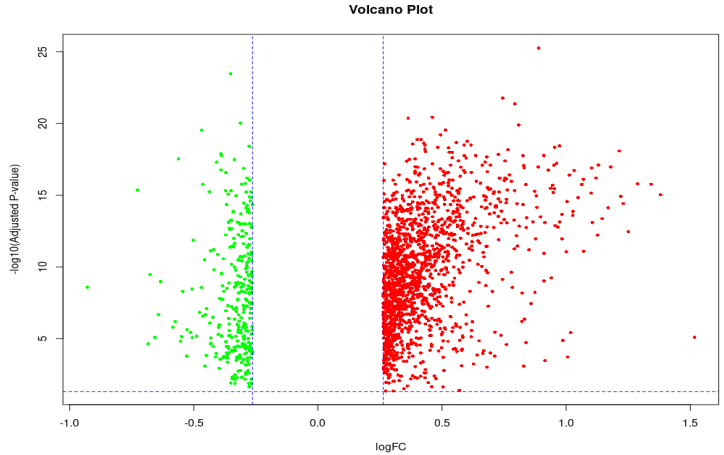
Visualizing the Data

Volcano plot before the first phases



Visualizing the Data

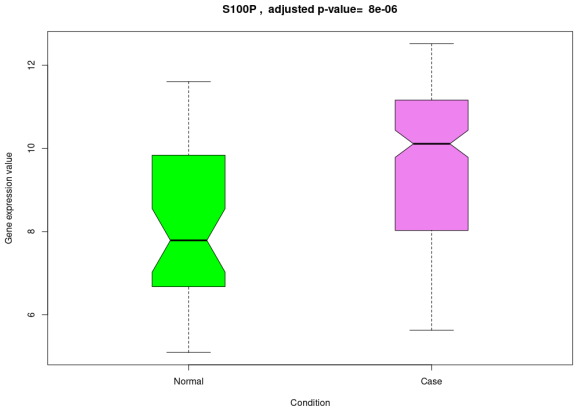
Volcano plot after the first phases



Visualizing the Data

Most upregulated gene

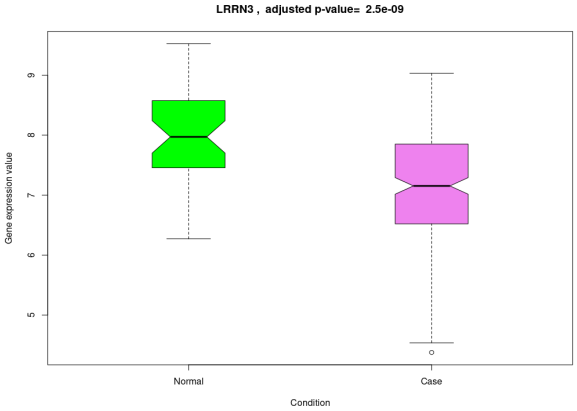
This box plot represents the expression levels of the most up-regulated gene **S100P** between the RA and healthy control groups. It visualizes the distribution of gene expression highlighting the differences between the two groups.



Visualizing the Data

Most downregulated gene

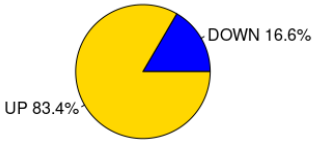
This box plot represents the expression levels of the most down-regulated gene **LRRN3** between the RA and healthy control groups.



Visualizing the Data

Pie chart of up and downregulated

The pie chart illustrates the distribution of differentially expressed genes based on their up/down regulation.



Visualizing the Data

Heatmap

The heatmap displays the gene expression levels for a subset of genes across different samples. It provides a visual representation of gene expression patterns, where the color intensity reflects the expression level of each gene in a particular sample.

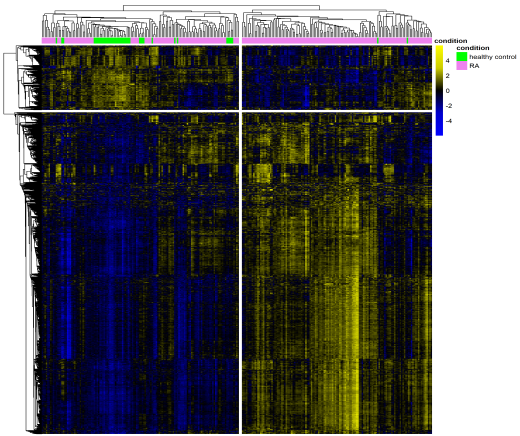


Table of Contents

4 PCA Analysis

PCA Analysis

Score Plot

The score plot shows the distribution of samples in the PCA space. Each point represents a sample, and their positions are determined by the principal components PC1 and PC2.

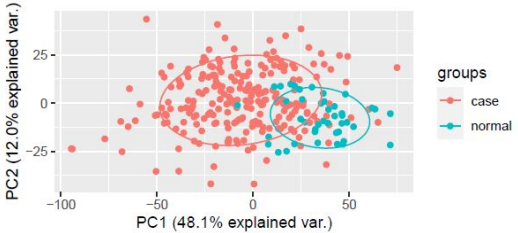


Figure: Score Plot

PCA Analysis

Pareto Chart and Scree Plot

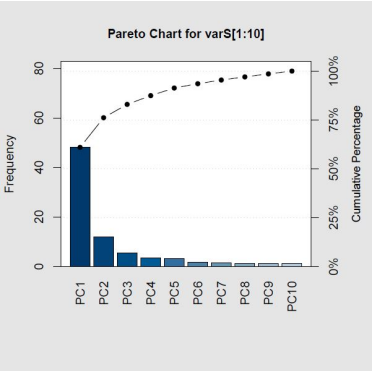


Figure: Pareto Chart

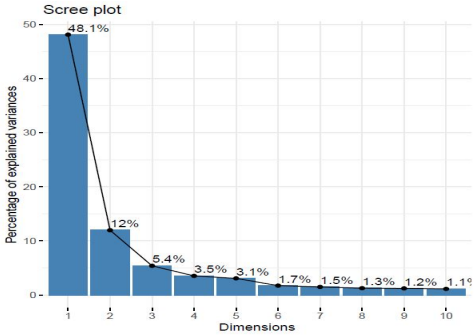


Figure: Scree Plot

PCA Analysis

Contributions of Components

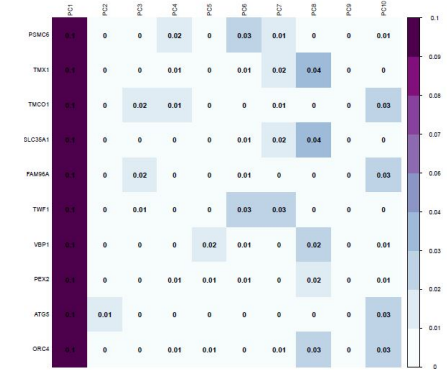


Figure: Contributions of Variables

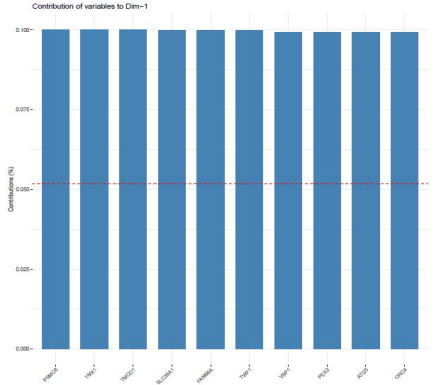


Figure: Scree Plot

PCA Analysis

Contributions of Components

This is the Scree plot representing the composition of the first three components. Is it possible to note that the first component is significantly more influential than the others and that the genes **SRP9** and **RCN2** play a central role according to PCA.

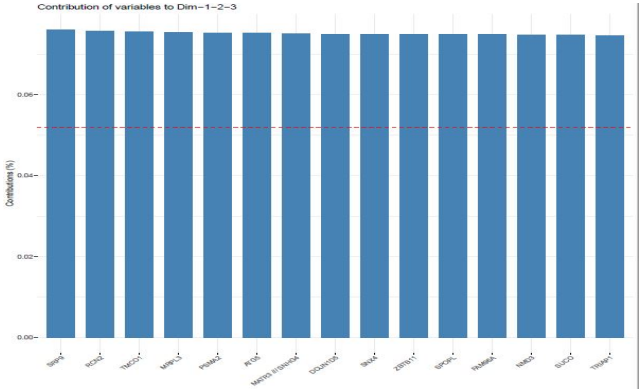


Table of Contents

5 Functional Enrichment Analysis

Functional Enrichment Analysis

Overview

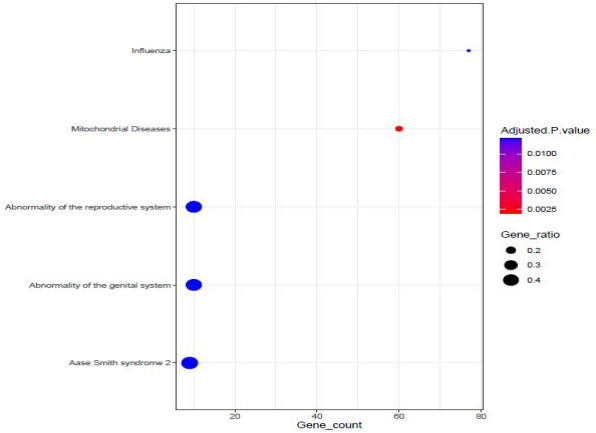
A Fucntional Enrichment Analysis was also conducted from the obtained list of differentially expressed genes using the *enrichR* R package to identify functional categories or pathways that are significantly enriched among the Differentially Expressed Genes (DEGs) associated with RA. It were performed two separate enrichment analyses: for upregulated genes and downregulated genes. The databases used are

- **DisGeNET** database:
  - **DisGeNET**: Disease-gene associations.
- **Gene Ontology** (**GO**) databases:
  - **GO Biological Process 2021**: Gene Ontology Biological Process terms.
- **KEGG** (**Kyoto Encyclopedia of Genes and Genomes**) database:
  - **KEGG 2021 Human**: KEGG pathways for human genes.
- **TRANSFAC and JASPAR PWMs** database:
  - Transcription factor binding motifs and regulatory elements.

Functional Enrichment Analysis

DisGeNET - Upregulated

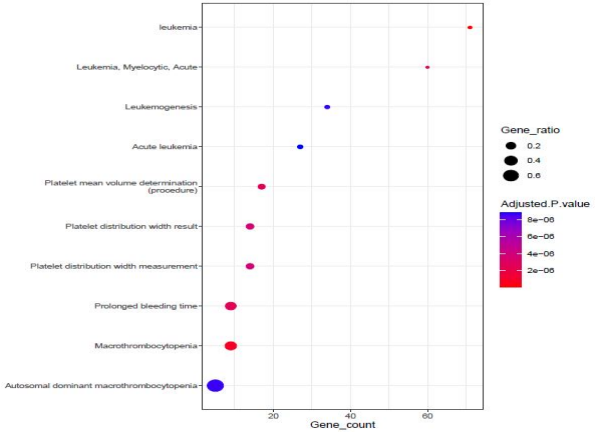
- Notice the high gene ratio of **abnormality of the reproductive and genital systems**.



Functional Enrichment Analysis

DisGeNET - Downregulated

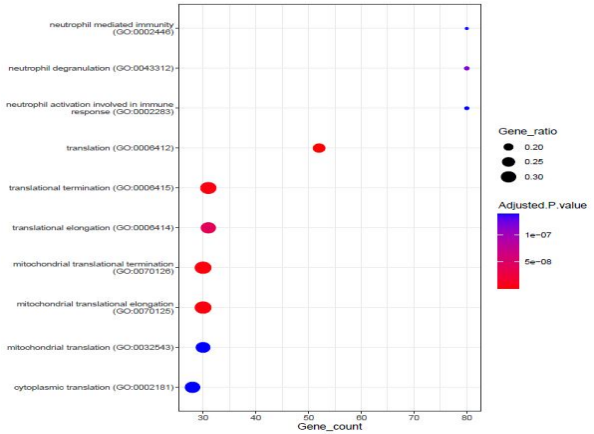
- Notice many voices associated with **leukemia**.



Functional Enrichment Analysis

GO Biological Process - Upregulated

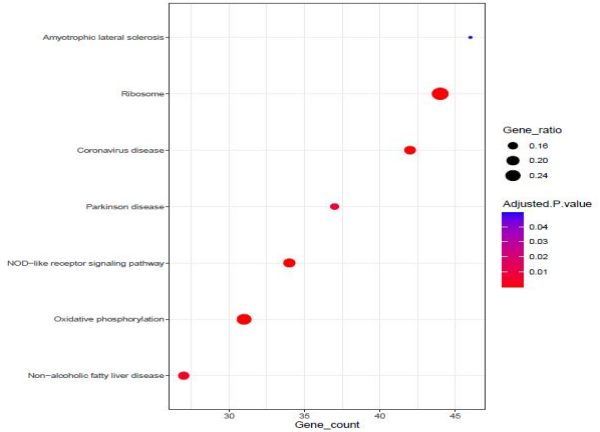
- The first three results are all related to **neutrophils**
- The other results are all related to **mitochondrial translational processes** and **general translational processes**



Functional Enrichment Analysis

KEGG Human - Upregulated

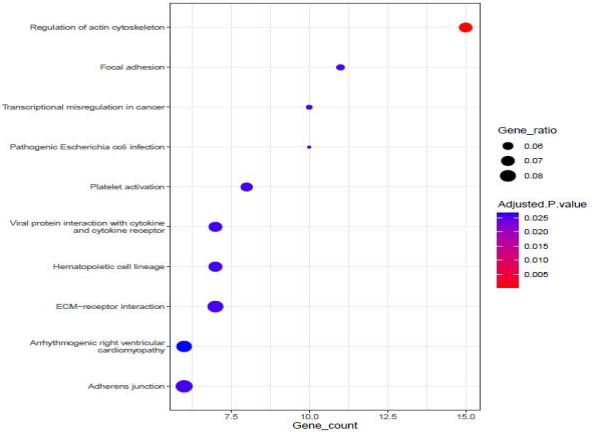
- We can notice the high gene ratio and low adjusted p-value of **Ribosome**-related pathways.
- We can notice several disease: **Coronavirus**, **amyotrophic**, etc.



Functional Enrichment Analysis

KEGG Human - Downregulated

- We can notice a important relevance of **regulation of actin cytoskeleton**.



Functional Enrichment Analysis

TRANSFAC and JASPAR PWMs - Upregulated

- Here we can notice a important relevance of **TCF4, MYB, GATA6**

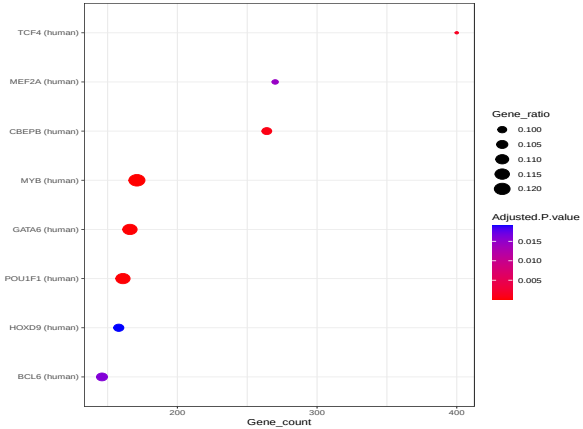


Table of Contents

6 miRNA Analysis

miRNA Analysis

Upregulated

- The research on the **miRTarBase** of the up-regulated genes returned 504 results

					Validation methods									
					Strong evidence			Less strong evidence						
ID	Species (miRNA)	Species (Target)	miRNA	Target	Reporter assay	Western blot	qPCR	Microarray	NGS	pSILAC	Other	CLIP-Seq	Sum	# of papers
MIRT000081	Homo sapiens	Homo sapiens	hsa-miR-328-3p	ABCG2	✔	✔	✔				✔		4	4
MIRT000197	Homo sapiens	Homo sapiens	hsa-miR-200c-3p	BMI1	✔	✔	✔	✔			✔		5	4
MIRT000266	Homo sapiens	Homo sapiens	hsa-miR-16-5p	BMI1	✔	✔					✔		3	3
MIRT000280	Homo sapiens	Homo sapiens	hsa-miR-15a-5p	BMI1	✔	✔	✔	✔			✔		5	3
MIRT000536	Homo sapiens	Homo sapiens	hsa-miR-16-5p	ACVR2A	✔	✔			✔		✔	✔	5	3
MIRT000709	Homo sapiens	Homo sapiens	hsa-miR-196a-5p	ANXA1	✔	✔	✔				✔		4	2
MIRT000998	Homo sapiens	Homo sapiens	hsa-miR-519c-3p	ABCG2	✔	✔	✔				✔		4	3
MIRT001571	Homo sapiens	Homo sapiens	hsa-miR-155-5p	ARFIP1	✔					✔	✔		3	2
MIRT001801	Homo sapiens	Homo sapiens	hsa-miR-16-5p	CCNT2	✔				✔		✔		3	2

miRNA Analysis

Downregulated

- The research on the **miRTarBase** of the downregulated genes returned 197 results

Page of 7 ‹ Prev 1 2 ... 7 Next ›

Download search result

Filter for miRNA and target Search Example

					Validation methods									
					Strong evidence			Less strong evidence						
ID	Species (miRNA)	Species (Target)	miRNA	Target	Reporter assay	Western blot	qPCR	Microarray	NGS	pSILAC	Other	CLIP-Seq	Sum	# of papers
MIRT000020	Homo sapiens	Homo sapiens	hsa-miR-148a-3p	DNMT1	✔	✔	✔	✔			✔		5	6
MIRT000038	Homo sapiens	Homo sapiens	hsa-miR-532-5p	RUNX3	✔	✔	✔				✔		4	3
MIRT000121	Homo sapiens	Homo sapiens	hsa-miR-24-3p	MYC	✔	✔	✔	✔			✔		5	1
MIRT000189	Homo sapiens	Homo sapiens	hsa-miR-204-5p	MEIS1	✔	✔		✔			✔		4	2
MIRT000196	Homo sapiens	Homo sapiens	hsa-miR-200c-3p	TUBB3	✔	✔	✔	✔			✔		5	4
MIRT000287	Homo sapiens	Homo sapiens	hsa-miR-155-5p	MEIS1	✔	✔	✔				✔		4	2
MIRT000475	Homo sapiens	Homo sapiens	hsa-let-7g-5p	MYC	✔	✔	✔				✔		4	3
MIRT000482	Homo sapiens	Homo sapiens	hsa-miR-17-5p	JAK1	✔	✔	✔	✔	✔		✔		6	3

miRNA Analysis

Most Upregulated

According to the **miRTarBase**, the most upregulated gene, **S100P**, is targeted by miRNA **hsa-miR-495-3p**, which is associated with the following diseases:

					Validation methods										
					Strong evidence			Less strong evidence							
ID	Species (miRNA)	Species (Target)	miRNA	Target	Reporter assay	Western blot	qPCR	Microarray	NGS	pSILAC	Other	CLIP-Seq	Sum	# of papers
MIRT736285	Homo sapiens	Homo sapiens	hsa-miR-495-3p	S100P	✔	✔	✔						3	0

Figure: Result of the S100P research

Figure: Disease associated with hsa-miR-495-3p

miRNA Analysis

Most Downregulated

According to the **miRTarBase**, the most downregulated gene, **LRRN3**, is targeted by miRNA **hsa-miR-181a-5p**, which is associated with the following diseases:

					Validation methods											
					Strong evidence			Less strong evidence								
ID	Species (miRNA)	Species (Target)	miRNA	Target	Reporter assay	Western blot	qPCR	Microarray	NGS	pSILAC	Other	CLIP-Seq	Sum	# of papers
MIRT025063	Homo sapiens	Homo sapiens	hsa-miR-181a-5p	LRRN3				✔					1	1

Figure: Result of the LRRN3 research

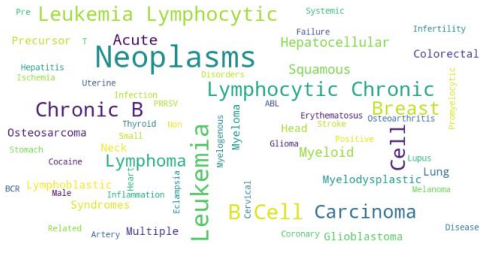


Figure: Disease associated with hsa-miR-181a-5p

Table of Contents

7 Literature-based research

Literature-based research

Overview

The Literature-based research is based on the results of the analysis obtained. The genes and diseases studied were:

- Genes:
  - **S100P** (The most upregulated)
  - **LRRN3** (The most downregulated)
  - **SRP9** and **RCN2** (most influential from the PCA analysis)
- Cell:
  - **Neutrophils**
- Transcription Factors:
  - **MYB**
  - **GATA6**

Literature-based research

S100P

The **S100 calcium binding protein P**, is a **protein coding** gene. Location is 4p16.1. The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. [NCBI].

- Wu YY, Li XF, Wu S, Niu XN, Yin SQ, Huang C, Li J. **Role of the S100 protein family in rheumatoid arthritis** Arthritis Res Ther. 2022 Jan 31;24(1):35. doi: 10.1186/s13075-022-02727-8. PMID: 35101111; PMCID: PMC8802512.

Literature-based research

S100P

The paper highlights the **significant role of S100 proteins in rheumatoid arthritis (RA) pathogenesis**. **These calcium-binding proteins**, belonging to the S100 protein family, **are overexpressed in RA and contribute to synovial hyperplasia, inflammatory cell infiltration, and joint destruction**.
The upregulated levels of S100 proteins in the serum and synovial fluid are closely associated with the **severity** of RA. **They serve as useful biomarkers for disease assessment**. Additionally, **S100 proteins have been implicated in the formation of pannus, a pathological tissue in RA**. Targeting S100 proteins may offer therapeutic opportunities for managing chronic inflammation in RA. **Importantly, since the effects of targeted treatments primarily occur at local inflammatory sites, the risk of systemic adverse reactions is minimized**.

Literature-based research

LRRN3

LRRN3 (**Leucine Rich Repeat Neuronal 3**) is a Protein Coding gene. Location is 7q31.1. Predicted to act upstream of or within positive regulation of synapse assembly. Predicted to be integral component of membrane. Predicted to be active in extracellular matrix and extracellular space. [provided by Alliance of Genome Resources, Apr 2022] [NCBI].

- Póliska, S., Besenyei, T., Végh, E. et al. **Gene expression analysis of vascular pathophysiology related to anti-TNF treatment in rheumatoid arthritis**. Arthritis Res Ther 21, 94 (2019). https://doi.org/10.1186/s13075-019-1862-6

Literature-based research

LRRN3

The analysis of the differentially expressed genes (DEGs) in rheumatoid arthritis (RA) revealed that **Leucine-rich repeat neuronal 3 (LRRN3) gene exhibited significant downregulation in patients with RA**. This finding is **in line with the paper**, which also **highlights the downregulation of LRRN3 in association with abnormal intima-media thickness (IMT) and pulse wave velocity (PWV) in RA patients**.

The paper further supports the notion that various genes, including LRRN3, play **essential roles in the pathogenesis of RA**. Notably, **the downregulation of LRRN3 in RA patients with abnormal IMT and PWV suggests its potential involvement in the vascular and cardiovascular manifestations of the disease**.

Further research on these genes' specific roles and mechanisms may hold the key to developing targeted therapeutic strategies for RA patients at risk of cardiovascular complications.

Literature-based research

SRP9

SRP9 (**Signal Recognition Particle 9**) is a Protein Coding gene. Location is 1q42.12. Predicted to enable RNA binding activity and signal recognition particle binding activity. Predicted to be involved in SRP-dependent cotranslational protein targeting to membrane. Predicted to be located in cytosol. Predicted to be part of signal recognition particle, endoplasmic reticulum targeting. [provided by Alliance of Genome Resources, Apr 2022] [NCBI].

- No work linked to RA was found.

Literature-based research

RCN2

RCN2 (**Reticulocalbin 2**) is a Protein Coding gene. Location is 15q24.3. The protein encoded by this gene is a calcium-binding protein located in the lumen of the ER. The protein contains six conserved regions with similarity to a high affinity Ca(+2)-binding motif, the EF-hand. This gene maps to the same region as type 4 Bardet-Biedl syndrome, suggesting a possible causative role for this gene in the disorder. Alternatively spliced transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Nov 2012] [NCBI].

- No work linked to RA was found.

Literature-based research

Neutrophils

Neutrophils are **immune cells** with unusual biological features that furnish potent antimicrobial properties. These cells phagocytose and subsequently kill prokaryotic and eukaryotic organisms very efficiently. [...] They can harness their noxious machinery in several contexts, like cancer. Inappropriate or dysregulated neutrophil activation damages the host and contributes to autoimmune and inflammatory disease. [Burn GL, Foti A, Marsman G, Patel DF, Zychlinsky A. **The Neutrophil**. Immunity. 2021 Jul 13;54(7):1377-1391. doi: 10.1016/j.immuni.2021.06.006. PMID: 34260886.].

- Zhang L, Yuan Y, Xu Q, Jiang Z, Chu CQ. **Contribution of neutrophils in the pathogenesis of rheumatoid arthritis**. J Biomed Res. 2019 Oct 24;34(2):86-93. doi: 10.7555/JBR.33.20190075. PMID: 32305962; PMCID: PMC7183296.

Literature-based research

Neutrophils

Neutrophils, as major innate immune cells, **contribute to RA pathogenesis through the formation of neutrophil extracellular traps (NETs) and release of citrullinated antigens**. They exhibit prolonged activation, enhanced cytotoxic potential, and regulatory functions similar to macrophages and dendritic cells. **Depletion of neutrophils reduces arthritis severity, and various therapeutic agents affecting neutrophils show efficacy in RA treatment**. **Dysregulated neutrophils in RA contribute to inflammation, tissue destruction, and dysregulated adaptive immune responses**. Their involvement in citrullination and release of citrullinated proteins further amplifies disease processes. Understanding the factors driving neutrophil dysregulation in RA can be crucial for developing targeted therapies.

Literature-based research

MYB

MYB (**MYB Proto-Oncogene, Transcription Factor**) is a Protein Coding gene. Location is 6q23.3. This gene encodes a protein with three HTH DNA-binding domains that functions as a transcription regulator. This protein plays an essential role in the regulation of hematopoiesis. This gene may be aberrently expressed or rearranged or undergo translocation in leukemias and lymphomas, and is considered to be an oncogene. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2016] [NCBI].

- Comertpay B, Gov E. **Identification of key biomolecules in rheumatoid arthritis through the reconstruction of comprehensive disease-specific biological networks**. Autoimmunity. 2020 May;53(3):156-166. doi: 10.1080/08916934.2020.1722107. Epub 2020 Feb 3. PMID: 32013628.

Literature-based research

MYB

- **MYB**, a transcription factor, shows **high gene ratio** and very **low adjusted p-value** in rheumatoid arthritis (RA).
- Integration of gene expression among with literature analysis reveals the **involvement of MYB** as a **key biomolecule in the molecular network of RA**.
- MYB is identified as a **transcription factor** (TF) **in the disease-specific biological networks of RA**.
- It is suggested that **MYB**, along with other hub proteins, receptors, and miRNAs, plays a role in the **pathogenesis** of RA and may serve as potential drug targets or diagnostic markers.

Literature-based research

GATA6

GATA6 (**GATA Binding Protein 6**) is a Protein Coding gene. Location is 18q11.2. This gene is a member of a small family of zinc finger transcription factors that play an important role in the regulation of cellular differentiation and organogenesis during vertebrate development. This gene is expressed during early embryogenesis and localizes to endo- and mesodermally derived cells during later embryogenesis and thereby plays an important role in gut, lung, and heart development. Mutations in this gene are associated with several congenital defects. [provided by RefSeq, Mar 2012] [NCBI].

- Zhao J, Chen B, Peng X, Wang C, Wang K, Han F, Xu J. **Quercetin suppresses migration and invasion by targeting miR-146a/GATA6 axis in fibroblast-like synoviocytes of rheumatoid arthritis**. Immunopharmacol Immunotoxicol. 2020 Jun;42(3):221-227. doi: 10.1080/08923973.2020.1742732. Epub 2020 Mar 26. PMID: 32216502.

Literature-based research

GATA6

- **GATA6**, a transcription factor, exhibits **high gene ratio** and very **low adjusted p-value** in rheumatoid arthritis (RA).
- The **upregulation of GATA6** in RA patients suggests its **potential role in joint dysfunction and pathogenesis of RA**.
- Further investigation is needed to determine the specific mechanisms by which **GATA6 influences the migration and invasion of fibroblast-like synoviocytes** (FLSs) in RA.

Table of Contents

Conclusion

Final Considerations

The analysis conducted in this study on the rheumatoid arthritis (RA) **is coherent** with existing literature. Several key factors, including **S100P, LRRN3, MYB, GATA6**, and **neutrophils**, have been identified as **influential** in RA pathogenesis.

S100P has been consistently associated with RA and is **closely linked** to **inflammatory processes and joint damage**. Its **overexpression in RA patients is indicative of its potential role in the pathogenesis of the disease**.

Neutrophils, as key components of the immune response, play a critical role in RA. Their **excessive activation and infiltration into the synovial tissue contribute to chronic inflammation and joint destruction in RA patients**.

While other genes (LRRN3, MYB, GATA6) also seem to exhibit significant relevance to RA, **further research is required to fully understand their specific contributions and interactions within the complex network of RA pathogenesis**.

Gene Expression Analysis of Rheumatoid Arthritis using GEOQuery Dataset

Thank you for listening!