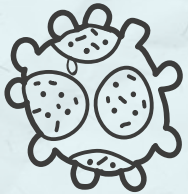
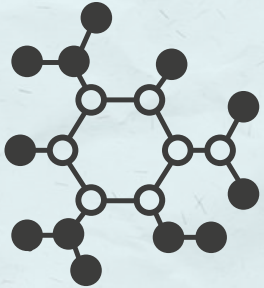
Determining Potential Biomarkers to Identify Drug Resistant Tuberculosis

Group 14





A Revisit of Our Goal

There has been a lack of efficient diagnostic strategies used to identify Drug-Resistant TB from the usual TB Which often misguides treatment of TB-patients.

The goal of our project is to identify potential biomarkers that could be used to identify whether a TB patient has drug resistance to one or more of the TB drugs.

Biomarkers are measurable indicators such as molecules, physical characteristics and genetic mutations which can distinguish between normal and Multi-Drug Resistant TB.

Dataset

GSE157657

Title Blood transcriptomics for diagnosis, risk, and treatment monitoring in tuberculosis reveal the evolution and resolution of TB disease: Does one signature capture all?

Organism Homo sapiens

Experiment type Expression profiling by high throughput sequencing

Groups Created:

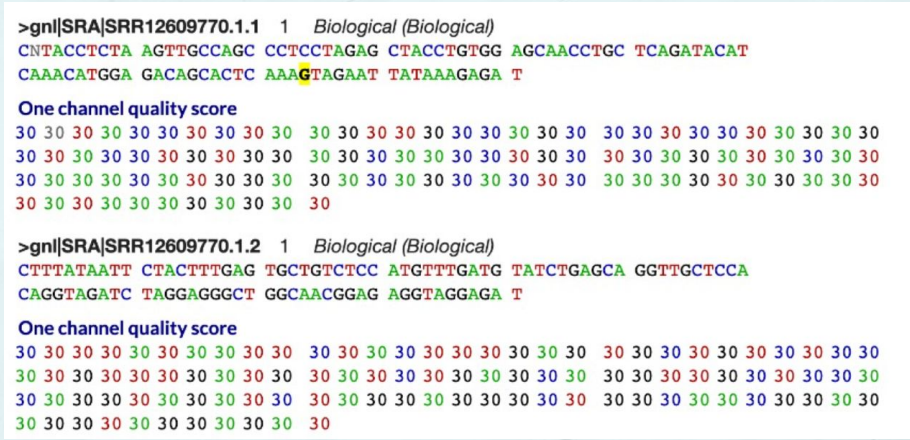
1) Drug-Resistant TB
2) Culture Positive TB Patients

The given data set is an RNA sequencing dataset which is a powerful technique to measure expression levels of genes in a biological sample.

Quality Control

Quality Control is a necessary step of RNA sequencing to assess quality and reliability of sequencing data. To do so we download the FASTQ file from NCBI and input it into FASTQC software. Based on the quality assessment, low-quality bases or adapter sequences may need to be trimmed or removed.

The data given is already quality controlled and trimmed to remove poor quality reads. The image on the right shows how only good quality sequences have been retained in the data.



Alignment

Sequence alignment is the process of comparing two or more biological sequences (such as DNA, RNA, or protein sequences) to identify regions of similarity, homology, or differences.

RNA sequence alignment is a crucial step in differential expression analysis (DEA) because it enables the comparison of gene expression levels between different samples or conditions.

The given data was aligned with the reference human genome given.

Reference Transcriptome:

https://www.ncbi.nlm.nih.gov/nuccore/NC_000023.11?report=fasta&from=73792205&to=73829231

Read Counting

Read counting is the step where the number of sequencing reads that are aligned to specific genomic locations are determined.

This step of preprocessing had already been performed for the given dataset as NCBI uses raw count matrices as input to DeSeq2.

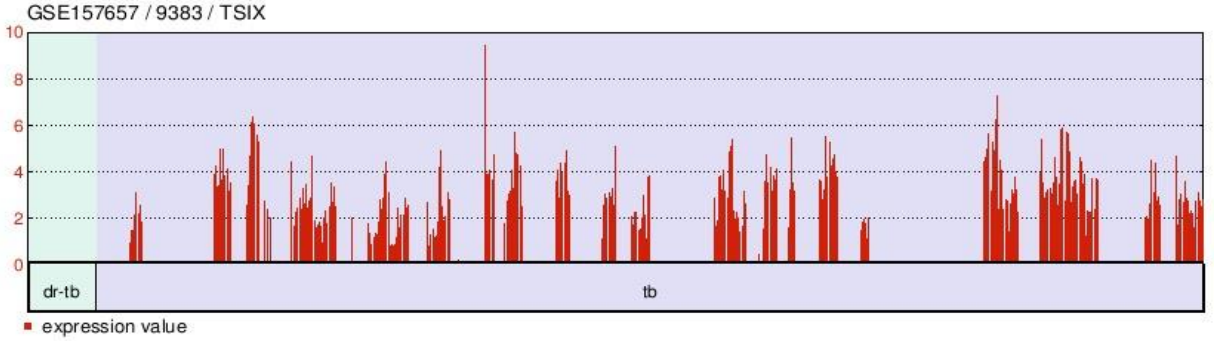
How the Differential Expression Analysis is executed

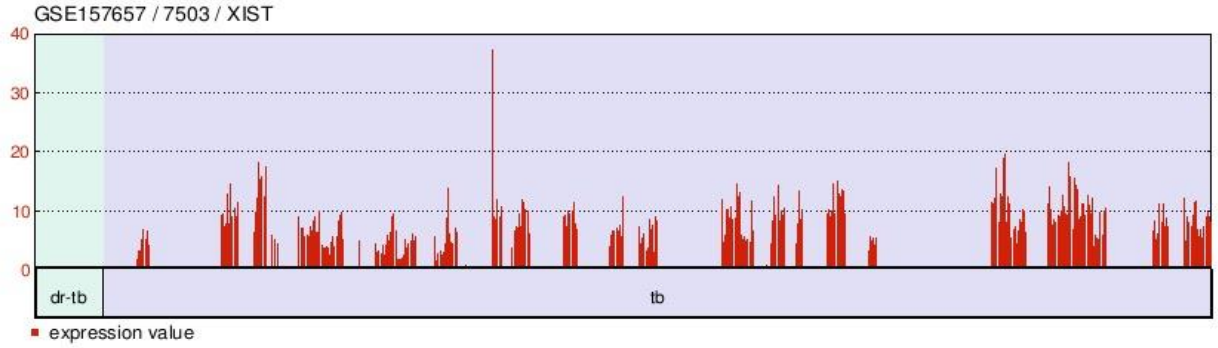
- The primary task is analysis of count data from RNA-seq is the detection of differentially expressed genes.
- The count data are presented as a table which reports, for each sample, the number of sequence fragments that have been assigned to each gene.
- An important analysis question is the quantification and statistical inference of systematic changes between conditions.
- The package DESeq2 provides methods to test for differential expression.

Observation-DEA

Software used:
GEO2R

Genetic Biomarkers

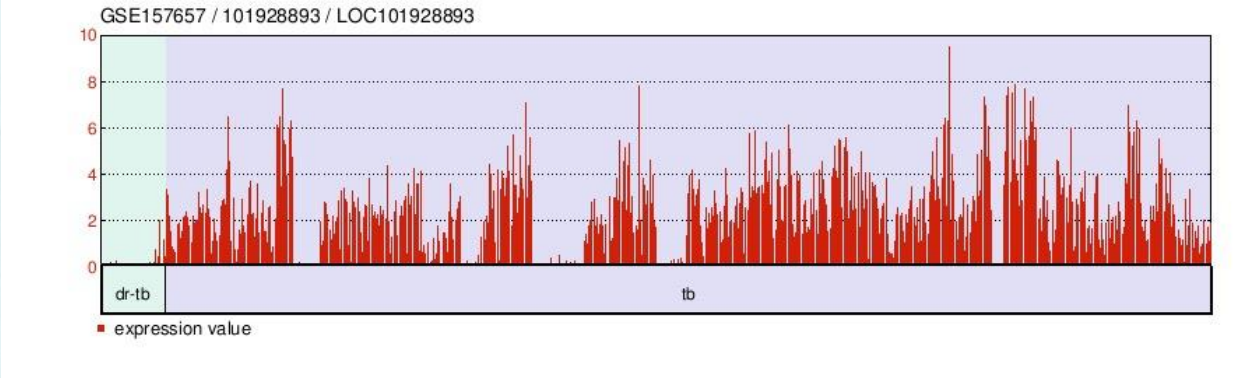




We can say that these genes have turned off in people suffering from drug-resistant tb.

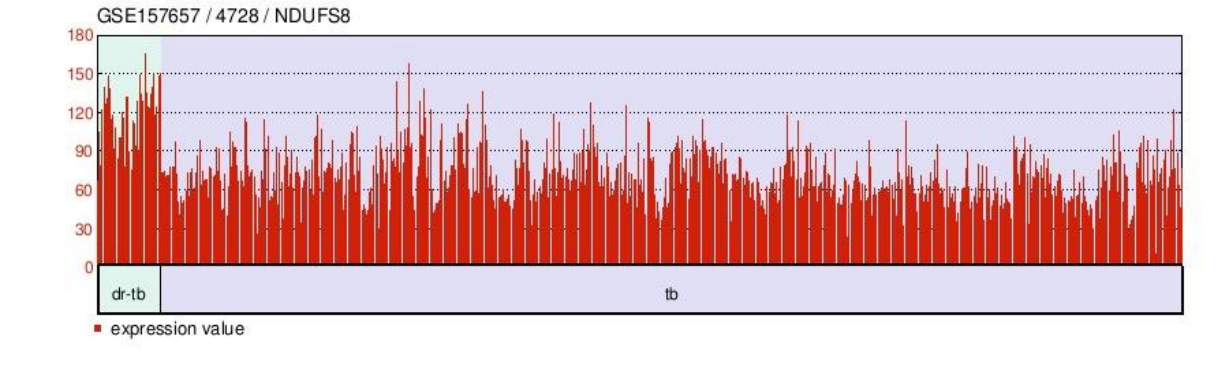
Observation-DEA

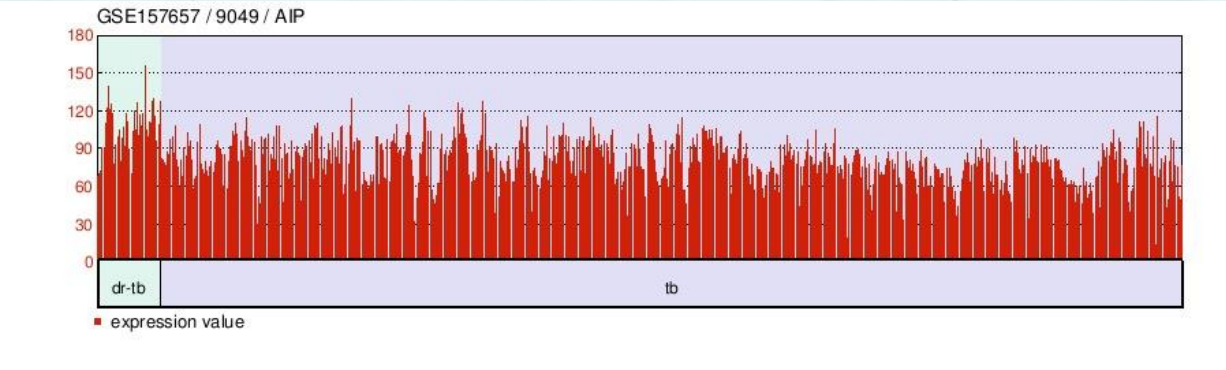
Expression Biomarker (Downregulated Gene)



Observation-DEA

Expression Biomarkers (Upregulated Gene)

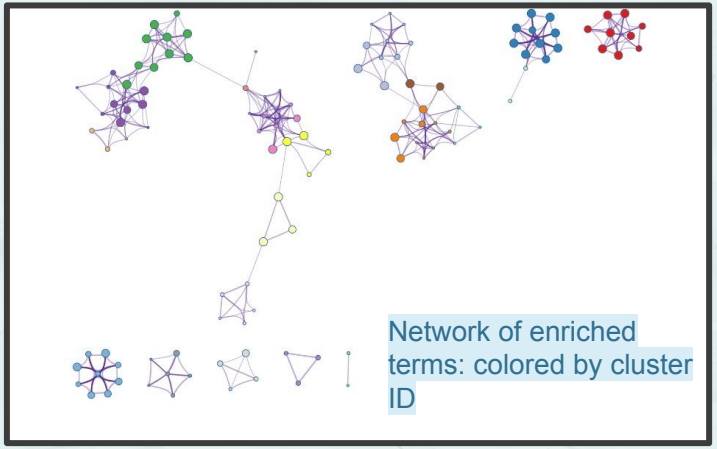




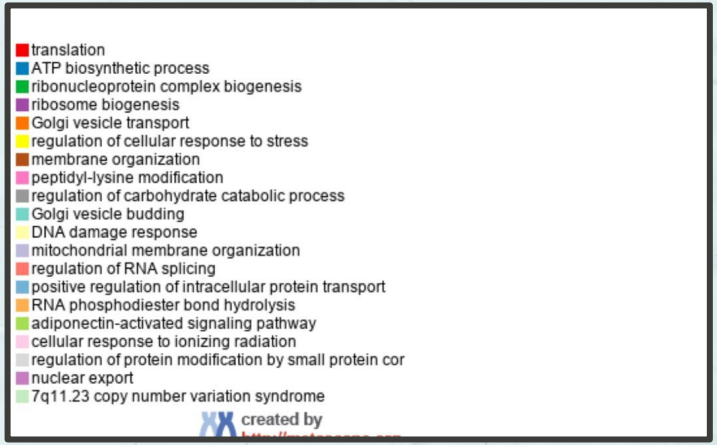
Observation-Pathway Analysis

Software used:
Metascape

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0006412	GO Biological Processes	translation	39	8.07	-19.11	-14.90
GO:0006754	GO Biological Processes	ATP biosynthetic process	20	4.14	-17.26	-13.53
GO:0022613	GO Biological Processes	ribonucleoprotein complex biogenesis	32	6.63	-10.68	-8.14
GO:0042254	GO Biological Processes	ribosome biogenesis	21	4.35	-7.43	-5.04
GO:0048193	GO Biological Processes	Golgi vesicle transport	18	3.73	-6.12	-3.78
GO:0080135	GO Biological Processes	regulation of cellular response to stress	30	6.21	-5.91	-3.57
GO:0061024	GO Biological Processes	membrane organization	31	6.42	-5.65	-3.32
GO:0018205	GO Biological Processes	peptidyl-lysine modification	16	3.31	-5.33	-3.00
GO:0043470	GO Biological Processes	regulation of carbohydrate catabolic process	8	1.66	-5.25	-2.93
GO:0048194	GO Biological Processes	Golgi vesicle budding	4	0.83	-4.90	-2.60
GO:0006974	GO Biological Processes	DNA damage response	29	6.00	-4.85	-2.55



Network of enriched terms: colored by cluster ID



Conclusion

Our analysis yields two genetic biomarkers - TSIX and XIST genes which were found to be unexpressed in Drug Resistant Tuberculosis.

TSIX and XIST

Even though there is no direct evidence linking the TSIX or XIST gene to tuberculosis (TB).
Recent studies have shown that non-coding RNAs, including long non-coding RNAs (lncRNAs) like TSIX, can play important roles in the regulation of immune responses and the pathogenesis of infectious diseases, including TB.

Conclusion

LOC101928893 (Downregulated)

While the specific role of the LOC101928893 gene in drug-resistant TB is not well understood, it is possible that dysregulation of lncRNAs like LOC101928893 may contribute to altered immune responses and the development of drug resistance in TB. Further research is needed to investigate this possibility.

Conclusion

Several expression biomarkers were also identified which included the

NDUFS8 (Upregulated)
A study published in the Journal of Proteome Research in 2012 analyzed the proteome (the entire set of proteins) of macrophages infected with M. tuberculosis and identified changes in the expression of various mitochondrial proteins, including NDUFS8. The study suggested that the upregulation of NDUFS8 in infected macrophages may be related to the increased metabolic demands required to support host defense mechanisms against M. tuberculosis.

This indicates a possible impairment or dysfunction of the host's macrophages in their ability to combat tuberculosis in patients with DR-TB. They could be suppressed due to the drugs given to the patient.

However, a conflicting study published in the journal PLOS ONE in 2012 found that the expression of genes involved in mitochondrial function, including NDUFS8, was downregulated in drug-resistant strains of M. tuberculosis compared to drug-susceptible strains. The study suggested that this downregulation may be related to the adaptation of M. tuberculosis to the stressful conditions induced by antibiotic treatment.

Conclusion

AIP (Upregulated)
The AIP gene, also known as FKBP5 (FK506 binding protein 5), is located on chromosome 6 in humans. The protein encoded by this gene is a member of the immunophilin protein family and functions as a co-chaperone in the regulation of steroid hormone receptor function.

There is some evidence supporting our finding which suggest that the AIP gene may be related to drug-resistant tuberculosis (TB). A study published in the Journal of Proteomics in 2014 found that AIP was one of the proteins differentially expressed in multi-drug resistant (MDR) TB compared to drug-susceptible TB. AIP was upregulated in the MDR-TB group and the authors suggested that it may be involved in the stress response and drug resistance mechanisms of the bacteria.

References

https://www.ncbi.nlm.nih.gov/gene/7503

https://metascape.org/gp/index.html#/reportfinal/ti14i2c1z

http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#the-deseqdataset

https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html

Contributions

Arnav Agarwal(2021235) : Identification of Dataset, Data Preprocessing

Anubhav Patel(2021017) : Pathway and Process enrichment Analysis and Literature Review

Mayank Gupta(2021065): Pathway and Process enrichment Analysis

Raj Pratap Singh(2021084): Differential Expression Analysis

Arpan Kumar(2021020): Differential Expression Analysis

Manya Tyagi(2021064): Data Preprocessing

Janesh Kapoor(2021466): Background Research

Thankyou!