To study SUM159 Cell Lines Untreated and Treated with drug Mebendazole and its effect on Transcriptome Level (Rna-Seq Analysis).

[1]*Pramodkumar P Gupta, [1]Mala Parab, [1]Janhavi Tripathi, [1]Mayur Sonkusare, [1]Hridhya Nair, [1]Mrunal Gokhale, [2]Amit Kumar Shrivastava, [1]Prerona Boruah, [1]Debjani Dasgupta

[1]School of Biotechnology and Bioinformatics, D Y Patil Deemed to be University, Plot 50, Sector 15, CBD Belapur, Navi Mumbai 400614, Maharashtra, India
[2]Department of Pharmacology, Universal College of Medical Sciences, and Teaching Hospital Bhairahawa, Rupandehi 32900 Nepal
Corresponding author: Dr Pramodkumar P Gupta, pramodkumar785@gmail.com / pramod.gupta@dypatil.edu

Abstract:
Breast cancer is a leading cause of cancer-related deaths worldwide. While significant progress has been made in its diagnosis and treatment, it remains a major public health concern. This study aimed to investigate the transcriptomic effects of Mebendazole, an antiparasitic drug, on SUM159 cell lines, a model for triple-negative breast cancer (TNBC). RNA-Seq analysis was conducted to identify differentially expressed genes (DEGs) between untreated and Mebendazole-treated cells. Our analysis revealed significant transcriptional alterations in Mebendazole-treated SUM159 cells, with data collected from the NCBI GEO database. The GALAXY server NGS data analysis frame work was used, followed by FASTQC, FASTP, HISAT2, SAMTOOL_Sort, SAMTOOLs_Dataset, the Uploading GTF for Humans standard file from UCSC, Hiseq-Count and analysis in R-reported DEGs. The list of 820 DEGs were analysed for enrichment using EnrichR, David tools, to understand the involvement of DEGs in biological processes. DEGs were enriched in pathways related to cell cycle regulation, apoptosis, survival signaling, and metabolism, suggesting that Mebendazole exerts its effects through multiple mechanisms. Notably, the identified DEGs were associated with various diseases, including breast cancer and neurodegenerative disorders. The downregulated genes from the analysis reported to be involved in liver cirrhosis, atherosclerosis, neurological disorder, cellular adhesion, the extracellular matrix, the transcription factor, and the assembly of collagens in humans, whereas the upregulated genes were more involved in pathways of breast cancer, ovarian cancer, the cell cycle and the cell division process. These findings highlight the potential for Mebendazole to be repurposed as a therapeutic agent beyond its traditional use in parasitic infections. Further research is needed to validate these in vitro findings in vivo and explore the clinical implications of Mebendazole for cellular mechanisms, the signal cascade, TNBC and neurodegenerative diseases.

Background:
- Breast cancer is a world-wide dreadful disease.
- Numerous experimental and known drugs are tested to achieve the targeted therapy and fails due to the poorly identified mechanism (it can be achieved by genomic, transcriptomic and proteomic based approaches) .
- A high-throughput technology that enables rapid sequencing of hundreds of thousands of genes or whole genomes.
- Popular methods include Illumina Sequencing, Ion Torrent Sequencing, and Long-Read Sequencing.

What is TNBC?
- TRIPLE NEGATIVE BREAST CANCER.
- Accounts for 10-15% of all breast cancers.
- Cancer cells lack estrogen or progesterone receptors.
- Cells produce HER2 protein "negative" on all three tests.
- More common in women under 40, with BRCA1 mutation.

What is SUM159 cell line?
- Its is Triple negative breast cancer cell line.
- Highly useful for various studies.
- Grow well in approriate medium with suppliments.
- Use to investigate effective novel signalling pathways.

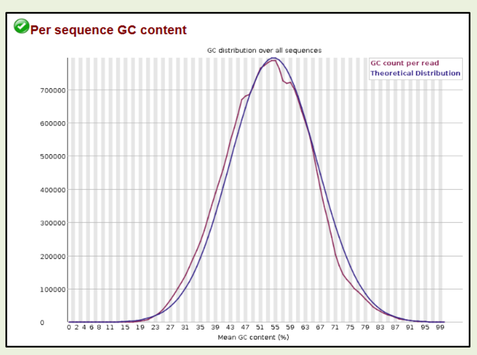
Problem Formulation
To understand the effect of Mebendazole drug and its effect on SUM 159 cell lines and identification of DEGs and involvement in metabolic pathway

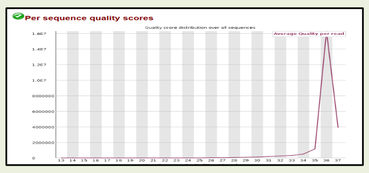
Materials and Method

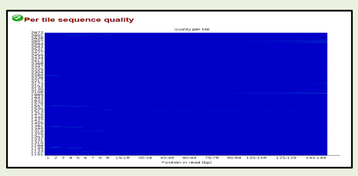
Retrieving data from NCBI and ENA Databases → Galaxy PROJECT → **Uploading data in GALAXY webserver** → **Running FASTQC** performs quality control → **FASTP** performs quality control (for pre-filtering data and post-filtering data) → **HISAT2** alignment program for mapping next-generation sequencing reads → **SAMTOOL SORT** widely used programs for processing and analyzing high throughput sequencing data. → **Samtools Idxstats** command prints stats for the BAM index file. → **GTF Files is uploaded in galaxy for Human hg38 with was downloaded from UCSC Genome Browse.** → **Htseq-Count** Used for RNA-Seq data analysis compare BAM file and a GTF or GFF file with gene models. It counts for each gene how many aligned reads overlap its exons. → **The Htseq-count data was then downloaded and opened in excel file.** → **Downloaded data of Htseq-count taken to R PROGRAMMING for obtaining DEG'S** → **Packages installed for obtaining DEG's in R**
1. BiocManager
2. Deseq
3. Pasilla

Results

FASTQC REPORT

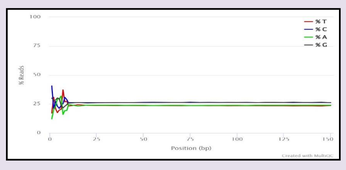


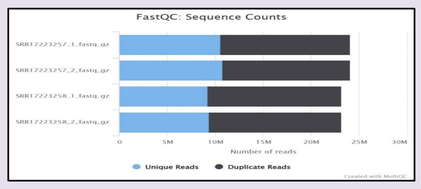


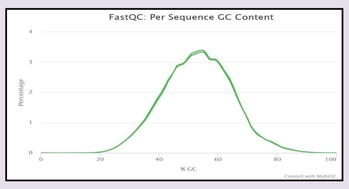


MULTIQC REPORT

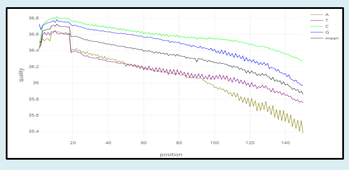
232	% Dups	% GC	M Seqs
SRR17223257_1_fastq_gz	56.1%	52%	24.1
SRR17223257_2_fastq_gz	55.3%	52%	24.1
SRR17223258_1_fastq_gz	60.1%	52%	23.2
SRR17223258_2_fastq_gz	59.5%	52%	23.2

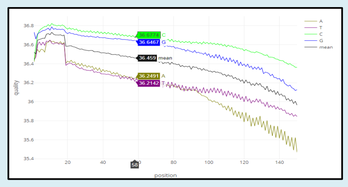






FASTP REPORT





HT SEQ Count REPORT

Geneid	Samtools sort on data 29
A1BG	186
A1BG-AS1	168
A1CF	0
A2M	0
A2M-AS1	0
A2ML1	1
A2MP1	0
A3GALT2	4
A4GALT	575
A4GNT	2
AA06	0
AAAS	786
AACS	515

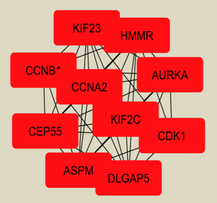
Up-Regulated Genes

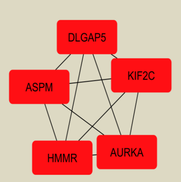
id	Fold Change	log2Fold Change	pval	Padj
ABCA10	8.945055	3.16109	0.01479	0.64907
ABCA13	7.333333	2.874469	0.001266	0.14337
ACOX2	11.21154	3.486912	0.017845	0.717637
AJAP1	4.496703	2.168868	0.019168	0.747947
APOBEC3B	11.09066	3.471273	4.82E-05	0.014849
APOBEC3B-AS1	4.865385	2.282554	0.024768	0.859233
ARC	11.63462	3.540352	0.015536	0.666964
AREG	8.615385	3.106915	0.006989	0.42505
ASPM	5.747144	2.522845	0.000459	0.072867
ATAT1	24.53846	4.616973	0.023736	0.845786

Down-Regulated Genes

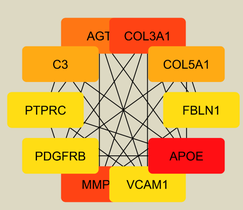
id	Fold Change	log2Fold Change	pval	padj
ABCA8	0.105112	-3.25	0.001081	0.128299
ABCA9	0.226804	-2.14048	0.042005	1
ACAN	0.124842	-3.00182	0.018423	0.730453
ACKR1	0.049107	-4.34792	1.02E-05	0.003943
ACTA2-AS1	0.204244	-2.29163	0.011333	0.567177
ACTBL2	0.097633	-3.35649	0.047919	1
ADAMTS2	0.240358	-2.05674	0.001926	0.190699
ADGRL3	0.094017	-3.41093	0.0002	0.038593
ADORA1	0.05641	-4.1479	0.017988	0.72043
AGT	0.075763	-3.72236	1.89E-07	0.000149

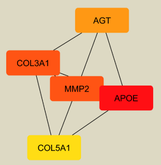
Up-Regulated Hub Genes



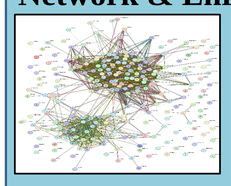


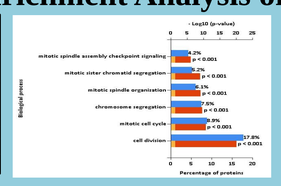
Down-Regulated Hub Genes

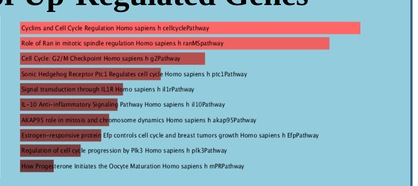




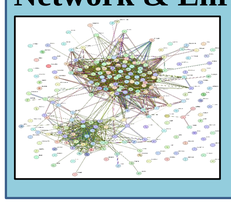
Network & Enrichment Analysis of Up-Regulated Genes

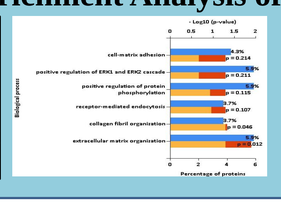


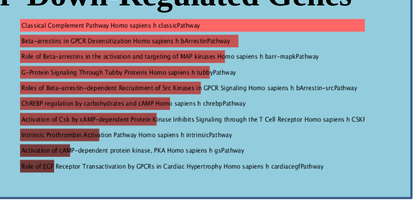




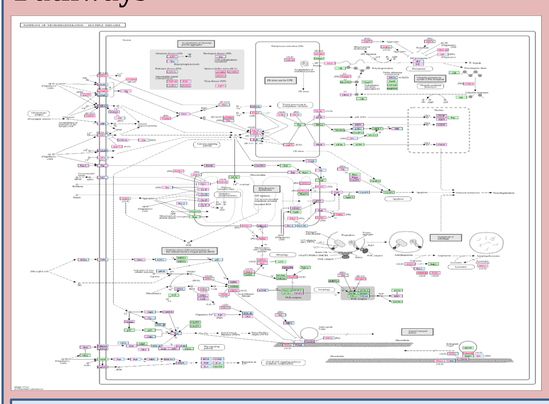
Network & Enrichment Analysis of Down-Regulated Genes

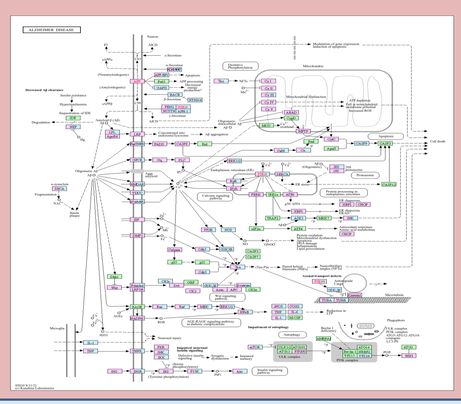


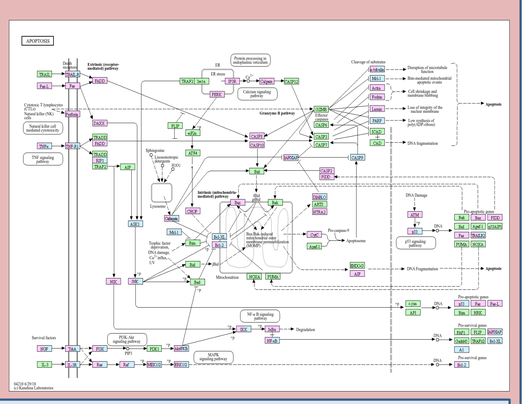




Pathways







Other Key Pathways such as: **P**arkinson's **P**athway, **P**rion **D**isease **P**athway, **T**ight **J**unction **P**athway, **H**untington **D**isease, **P**hagocytosis **P**athway, **G**ap **J**unction, **M**otor **P**roteins **P**athway, **C**ell **C**ycle, **C**yclin & **C**ell **C**ycle **R**egulation, **F**ocal **A**dhesion **PI3K Akt mTOR Signaling P**athway, **R**egulatory **C**ircuits of the **S**TAT3 **S**ignaling **P**athway, **C**omplement **S**ystems in **N**euronal **D**evelopment, **G-P**rotein **S**ignaling, **C**omplement and **C**oagulation **C**ascade, **I**nsulin **R**esistance, **R**elaxing **S**ignaling **P**athway and Many more…..

Conclusion:

This study employed Next-Generation Sequencing (NGS), a powerful technology, to investigate the effects of Mebendazole (MBZ) on the gene expression profile of SUM159 cells. NGS allowed for the rapid and comprehensive analysis of thousands of genes simultaneously. Total Differentially Expressed Genes (DEGs): 18,429 genes were identified as differentially expressed upon MBZ treatment compared to the control group. Upregulated Genes: 170 genes displayed a log2 fold change greater than or equal to 2. These genes are potentially activated by MBZ and might play a role in its mechanism of action or the cellular response to the drug. Downregulated Genes: 42 genes exhibited a log2 fold change less than or equal to -2, suggesting their suppression by MBZ. These genes could be involved in normal cellular functions that MBZ disrupts.
Refined Analysis (p-value <= 0.005):
Applying a more stringent p-value threshold (0.005) resulted in a smaller but potentially more reliable set of DEGs. Total DEGs: 1,130 genes were identified, indicating a focus on highly significant changes in gene expression. Upregulated and Downregulated Genes: The analysis identified 80 upregulated and 30 downregulated genes with a log2 fold change greater than or equal to 2 (up) or less than or equal to -2 (down), respectively. These genes represent strong candidates for further investigation into their functional roles in the context of MBZ treatment. Treatment-Specific Genes: Nine genes remained treatment-specific, highlighting their potential importance in MBZ's mechanism of action.

• **References:** Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), Article 10. https://doi.org/10.1186/gb-2010-11-10-r106. ; Bano, A., Stevens, J. H., Modi, P. S., Gustafsson, J.-Å., & Strom, A. M. (2023). Estrogen Receptor β4 Regulates Chemotherapy Resistance and Induces Cancer Stem Cells in Triple Negative Breast Cancer. *International Journal of Molecular Sciences*, *24*(6), Article 6. https://doi.org/10.3390/ijms24065867. ; Chai, J.-Y., Jung, B.-K., & Hong, S.-J. (2021). Albendazole and Mebendazole as Anti-Parasitic and Anti- Cancer Agents: An Update. *The Korean Journal of Parasitology*, *59*(3), Article 3. https://doi.org/10.3347/kjp.2021.59.3.189