Effects of p53 mutations on 3D changes in Chromatin conformation in cancers and its effect on Genome functioning

A project submitted to the

Bioinformatics Centre

Savitribai Phule Pune University

For the degree of

M. Sc. in Bioinformatics

By

Mohd. Naveed Hasan

Under the guidance of

Dr. Abhijeet Kulkarni

(Guide)

Bioinformatics Centre, SPPU

MAY – 2024

CERTIFICATE

This is to certify that the project entitled **"Effects of p53 mutations on 3D changes in chromatin conformation during cancer and its effect on genome functioning",** submitted by **Mr. Mohd. Naveed Hasan** in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics, has been carried out satisfactorily by him at the Bioinformatics Centre, Savitribai Phule Pune University.

Date: 18 /05 /23

Dr. Sangeeta Sawant

Place: Pune

Director
Bioinformatics Centre
Savitribai Phule Pune University,
Pune 411007

CERTIFICATE

This is to certify that the project entitled **“Effects of p53 mutations on 3D changes in Chromatin conformation during cancer and its effect on Genome functioning”**, submitted by M**r. Mohd Naveed Hasan** in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics, has been carried out satisfactorily by him at the Bioinformatics Centre, Savitribai Phule Pune University, under my guidance and supervision.

Dr. Abhijeet Kulkarni
(Guide)
Bioinformatics Centre, SPPU

Date: 18 /05 /23
Place: Pune

DECLARATION & UNDERTAKING

I hereby declare that the project entitled, "**Effects of p53 mutations on 3D changes in Chromatin conformation during cancer and its effect on Genome functioning**", submitted in partial fulfillment of the requirements of the degree of Master of Science in Bioinformatics, has been carried out by me at Bioinformatics Centre, Savitribai Phule Pune University under the guidance of **Dr. Abhijeet Kulkarni** (Guide). I further declare that the project work or any part there of has not been previously submitted for any degree or diploma of any University.

I also declare that to the best of my ability, I have ensured that the submission made herein, including the main text, supplementary data, deposited data, database entries, software code, figures, does not contain any plagiarized material, content or ideas, and that all necessary attributions have been appropriately made and all copyright permissions obtained, cited and acknowledged.

I also declare that any further extension, continuation, publication, patenting or any other use of this project (either in full or in part), if any, shall be undertaken with prior written consent from the Director, Bioinformatics Centre, Savitribai Phule Pune University and the Project Supervisor/s.

I further state that I shall explicitly mention, "Bioinformatics Centre, Savitribai Phule Pune University" as "place of work" and acknowledge "the M.Sc. Bioinformatics training programme at Savitribai Phule Pune University for infrastructure and facilities" in the publication (print and online)/patent based on this work. I shall also acknowledge DBT for providing studentship.

Date: 18 /05 /2024 Mohd. Naveed Hasan

Place: Pune BIM-2022-14

ACKNOWLEDGEMENTS

I am deeply grateful to Dr. Abhijeet Kulkarni, my research guide, for bestowing upon me his trust and granting me the invaluable opportunity to work under his guidance. His unwavering mentorship and consistent support have played a pivotal role in shaping this thesis. I am indebted to him for sharing his extensive knowledge and expertise throughout this journey and also for motivating me when the motivation was lacking.

I would also like to express my sincere appreciation to Dr. Sangeeta Sawant, Director of the Bioinformatics Centre at Savitribai Phule Pune University, Pune. Her unwavering assistance and provision of all necessary requirements have been instrumental in the successful completion of this project.

I express my deepest gratitude to Ms. Akshita Upretri for insightful discussions, patience, and invaluable feedback. Her willingness to give her time and an always ready to explain attitude has been very much appreciated. Apart from her domain knowledge, her positive and analytical thinking made the project happen so productively and efficiently. I would also like to extend my gratitude to my senior and a constant support in this project, Ms. Amisha Gupta for always being so understanding and supportive of me all throughout the project, her confidence in me has been a real key for completion of this project in time.

I want to thank the Department of Biotechnology, Government of India for providing me the studentship and resources to work on this project. I express my heartfelt gratefulness to all the non-teaching and technical staff for their continuous support.

I would also like to thank my friends for constant help and support, without which I would have been sulking every day, at last I would like to extend by biggest thanks to my family who encouraged me in everything, every time.

Mohd. Naveed Hasan

BIM-2022-14

TABLE OF CONTENTS

ABSTRACT

P53 is a tumor suppressor protein known to play ‘Genome Guardian’ role by checking genome integrity and deciding cell fate. Mutations in its DNA-Binding domain (DBD) are linked with a variety of patho-physiologies related to cancers. This is in principle because of loss of its function carried out through its DNA Binding. However, many studies indicate that mutant p53 can still bind to DNA in a non-canonical way via Gain of Function (GOF), acting like a MAR Binding protein. MARs are DNA element that keep the chromatin in a specific 3D shape. Their binding to mutant p53 leads to different chromatin folding, accessibility and therefore genome function. Overall changes in genome-wide gene expression program caused due to such changes in genome topology mediated by p53 mutations lead to cancer phenotypes. The present study aims investigate and determine effects of p53 DBD mutations on genome-wide topological changes and establishing its link with genome functionality. To investigate this, chromatin capture data, i.e., HI-C data of wild-type p53 and mutant specific p53 cell lines, are curated from GEO database and analyzed to learn that mutations in p53 indeed causes significant topological changes in genome conformation causing genes to shift from active chromatin domains to inactive chromatin domains. It also relates the genome functioning as determined by gene expression analysis performed using RNA-Seq in same conditions. These findings are presented and discussed in the present report.

Keywords – p53, Mut p53, Chromatin conformation, MARs, Cancer

❖ List of Abbreviations

Abbreviation	Description
Mut	Mutant
WT	Wild-Type
HIC	High-throughput Chromosome Conformation Capture
FANC	Feature rich framework for the analysis and visualization of chromosome conformation capture data
TAD	Topologically Associating Domains
S/MARs	Scaffold/Matrix Associated Regions
GO	Gene ontology
SAM	Sequence Alignment Map
BAM	Binary Alignment Map
BED	Browser Extensible Data
ENA	European Nucleotide Archive
DGE	Differentially Expressed Genes
PANTHER	Protein Analysis Through Evolutionary Relationships

❖ List of Tables

❖ List of Figures

INTRODUCTION

P53 is widely dubbed as the “Guardian of Genome”, the protein p53, encoded by the TP53 gene, is situated on the short arm of chromosome 17. Cellular perturbations, including but not limited to DNA damage, leads to genomic anomalies such as mutations, deletions, and translocations culminating in genomic instability and the potential inception of oncogenesis. To counteract these threats, an efficacious stress response is needed for the preservation of genomic integrity and to avert the emergence of such instabilities. The transcription factor p53 is pivotal in this defensive mechanism. Typically maintained at very low levels, p53 resides in a state of functional dormancy, governed by proteasomal degradation, predominantly orchestrated by the E3 ubiquitin protein ligase MDM2.

Upon encountering DNA damage, p53 is subjected to post-translational modifications, including phosphorylation and acetylation, which catalyze the nuclear accumulation of p53.This cascade of modifications releases p53 from the regulation of MDM2, thereby activating it. The activated p53 then transactivates a curated repertoire of target genes, which govern cell cycle arrest or apoptosis induction, depending upon the magnitude and nature of the DNA damage. Cell cycle arrest facilitates DNA repair, after which cellular proliferation may proceed unimpeded. In instances of extensive DNA damage, p53 initiates apoptosis or programmed cell death, eliminating cells with extensive genetic impairment and therefore stopping these defects to be translated to the next generation. Thus, p53 is integral to the maintenance of genomic integrity by eliciting appropriate cellular responses to DNA damage.

The DNA-binding domain of p53 engages with the tandem repeat of the p53-responsive element, characterized by the sequence **RRRCWWGYYY**—where R denotes G/A, W signifies A/T, and Y represents C/T.

These motifs are found within the promoter regions of p53 target genes, interspaced by no more than 13 base pairs. In response to diverse cellular stresses, such as DNA damage and energetic stresses, p53 induces cell cycle arrest and/or apoptosis via the transactivation of its target genes, thus elucidating its role in the cellular functioning that preserves life’s genetic continuum.

Mutations in the tumor suppressor gene p53 are prevalent in human cancers, with over 50% of cases exhibiting such alterations. The majority of these mutations are missense mutations,

leading to the buildup of mutant p53 within tumors. These mutated forms often acquire oncogenic functions, intensifying the malignant traits of cancer cells. Typically, these mutations affect the DNA-binding domain of the protein, disrupting its usual DNA-binding capabilities compared to the wild-type protein. Mutant p53 can roughly be divided into two categories: DNA contact (class I) mutants, where mutations affect amino acids directly involved in binding to the p53-responsive element in DNA, and conformational (class II) mutants, where mutations change the structure of p53, preventing it from binding to DNA in the same way as the wild-type protein. (Parrales A & Iwakuma T, 2015)

Moreover, mutant p53 shows oncogenic gain-of-function (GOF) activities, such as enhanced rate of tumor progression, metastatic potential, and drug resistance. One of these GOF abilities of the mutant protein is its ability to bind to the Nuclear Matrix/Scaffold Associated Region and inducing DNA conformation changes within the nucleus.

Chromosome are arranged in their respective localized areas within the nucleus and this induced conformation change can affect the chromatin accessibility of various transcription factors, regulating gene expression via a conformational mechanism. It is now evident that mutant p53 not only suppresses those genes that act as cell cycle checkpoints but also activate the expression of many growth-promoting and oncogenic genes which may lead to progression of aggressive cancers.

The array of TP53 mutations in human tumors displays remarkable diversity and tissue specificity. Mutations are notably concentrated in the DNA binding domain (DBD), with a predominance of missense mutations (~80%), including six "hotspot" codons (R175, R213, G245, R248, R273, and R282), which collectively contribute to approximately 25% of all TP53 mutations. Conversely, mutations occurring outside the DBD are more inclined to be nonsense or truncating mutations (~67%) rather than missense mutations (Hainaut P & Pfeifer GP, 2016). Furthermore, besides acquiring a TP53 mutation in one allele, most tumors undergo loss of the second allele through deletion or copy neutral loss of heterozygosity (Shirole et al., 2017). The degree to which this allelic variation arises from the underlying mutational mechanisms or biological selection for functionally relevant mutations, or a combination of both, remains partially understood.

In this study we have used Chromatin conformation as well as gene expression data from both cell types with wild-type and mutant p53 to find a consensus relation as to how mutant

p53 is affecting the 3D-conformation of the genome and how that is related to the changes in the gene expression leading to possibility of cancer progression through means of conformational gene expression regulation by the differential DNA-binding of mut-p53.

We have also tried to identify and characterize the genes that are switching compartments from transcriptionally active regions of the genome and going to the transcriptionally inactive region of genome in the mutant p53 phenotype as compared to their initial localization in the wild-type p53 phenotype.

Chromatin conformation

Chromatin conformation refers to the three-dimensional arrangement of chromatin within the nucleus of eukaryotic cells, which plays a fundamental role in regulating gene expression, genome stability, and other nuclear processes. The organization of chromatin into higher-order structures is crucial for the spatial and temporal control of gene activity, as well as for the maintenance of genome integrity. Understanding the principles underlying chromatin conformation has thus become a central focus of modern molecular biology research. One of the key features of chromatin conformation is the formation of topologically associating domains (TADs). TADs are relatively stable regions of the genome characterized by increased interactions among loci within the same domain and decreased interactions with loci outside the domain. These domains serve as structural units that facilitate the regulation of gene expression by bringing enhancers and promoters into close spatial proximity. TADs are thought to be delimited by boundary elements such as insulator proteins, which prevent the spread of chromatin modifications and help maintain the integrity of individual domains. TADs, or topologically associating domains, represent self-interacting genomic regions, indicating that DNA sequences within a TAD exhibit more frequent physical interactions with each other compared to sequences outside the TAD. The boundaries on both ends of these domains remain conserved across various mammalian cell types and even across species. These boundaries are notably enriched with CCCTC-binding factor (CTCF) and cohesin (Long et al., 2022).

Furthermore, the spatial organization of the nucleus is characterized by the segregation of chromatin into distinct compartments, known as A and B compartments. A compartment is enriched in active chromatin marks and gene-rich regions, whereas B compartment is associated with repressive chromatin marks and gene-poor regions. The

compartmentalization of chromatin into A and B compartments reflects the functional organization of the genome, with transcriptionally active regions preferentially localized in the A compartment and transcriptionally silent regions concentrated in the B compartment.

In this study, our primary source of data was chromatin conformation data generated through Hi-C experiments of wt and mut p53 cell lines, which were individually analyzed and inferred to find the compartmentalization of the genome specifically focusing on the A and B compartmentalization of the genome and elucidating the genes in these compartments to infer the overall functionality with respect to active and inactive genes. This organization, we hypothesized was an effector function of the mutation in p53 DNA-Binding abilities.

Differential Chromatin regulation of p53 and mutant p53

Unlike the wt p53, the GOF mutant p53 does not bind to the consensus DNA sequence motif of the wild type, it is hypothesized that the various missense mutations in the wild type protein affects the DNA-Binding motif of the wt p53. Mutant p53 has a strong affinity for MAR-DNA elements (MARs), hinting at the possibility that these differential binding in mut p53 may be linked to its enhanced oncogenic properties. Mutant p53 recognizes Matrix Attachment Regions (MARs) that have a high AT content and contain variations of the "DNA-unwinding motif" AATATATTT. This motif contributes to the dynamic nature of chromatin structure, facilitating the local unwinding of DNA strands. Mutant p53 has a specific interaction with oligonucleotides derived from MARs that carry these unwinding motifs. This interaction leads to the separation of DNA strands, especially when the motif is located within a structurally flexible sequence environment.

MARs organize cellular chromatin into topologically independent loops. This organization provides a structural foundation for the independent spatial and temporal regulation of gene expression and the initiation of DNA synthesis. This type of regulation is believed to form a higher-order regulatory mechanism for controlling development and differentiation.

The differential chromatin regulation of both the proteins can lead to the activation and inactivation of numerous oncogenic and tumor-suppressor genes, respectively. The identification of a mutant p53-specific activity that could potentially be involved in the modulation of chromatin structure and function might provide insights to better understand the activities of mutant p53 related to its "gain of function" properties.

In this study, we adopted HI-C and RNA-Seq publicly available data for wt p53 and mut p53 to investigate the correlation between the changes in chromatin conformation and the gene expression regulation as an effect of this change. We assumed that the changes in the compartmentalization of the genes from the conformation capture technique would also be evident in the RNA-seq gene expression analysis of the wt and mutated p53 cell lines.

MATERIALS & METHODS

Hi-C data analysis

This section describes various steps performed to analyze and interpret the Chromatin conformation generated through Hi-C experiments. An extensive data mining approach was taken to find relevant data from the cancer cell lines with wt and mut p53.

The entire workflow for HiC data analysis can be simplified in the following steps:

1. Data collection.
2. Mapping FASTQ files to the reference genome.
3. Creating a non-redundant dataset from the mapped files.
4. Generating and filtering Read pairs.
5. Generating, binning and filtering HiC objects
5. Analyzing HiC matrices

Data Collection

Sequence reads obtained from the GEO and SRA database were retrieved by utilizing their corresponding accession IDs to download the data from ENA in FASTQ format from wt p53 and mut p53 (Table 1) specific cell lines.

Table 1: Dataset for HI-C Analysis for Wild-type p53

Cell Line	**GSE**	**GSM**	**SRR**	**Read-Type**
SW780	GSE148079	GSM4453813	SRR11478969	Paired
HUVEC (CC-2517)	GSE63525	GSM1551631	SRR1658713	Paired

HUVEC (CC-2517)	GSE63525	GSM1551631	SRR1658714	Paired
HCT116	GSE158007	GSM4784017	SRR12646280	Paired

Table 2: Dataset for HI-C analysis for Mutant p53

Cell Line	**GSE**	**GSM**	**SRR**	**Read-Type**
TNBC	GSE167150	GSM5098064	SRR13755463	Paired
TNBC	GSE167150	GSM5098065	SRR13755464	Paired
TNBC	GSE167150	GSM5098068	SRR13755467	Paired
FHC	GSE133928	GSM3930267	SRR9675751	Paired

Read Mapping

The filtered reads were mapped to the reference genome (hg38 Built) downloaded from UCSC Browser using Bowtie2 v2.4.5. Bowtie 2 is an ultra-fast and memory-efficient tool for aligning sequencing reads to long reference sequences and it has three modes: gapped, local and paired-end alignment. Bowtie2 gives alignment output in .sam format (Langmead et al.,2012). Alignment files from Wild-type and Mutant p53 samples were merged into one BAM file to create a non-redundant dataset of the reads from all the cell lines.

Generating FANC pairs

The fanc pairs command handles the creation and modification of Pairs objects, which represent the mate pairs in a Hi-C library mapped to restriction fragments. The input given were 2 BAM files (of a single paired-end reads) sorted by read name and output was a FANC pairs file. The fanc hic command is used to generate fragment-level and binned Hi-C matrices from these pairs file.

Analysing HiC matrices

1. Computing AB compartments matrix

In a Hi-C matrix, regions are typically categorized as either active (A) or inactive (B) compartments. These designations stem from a correlation matrix, where each entry i, j represents the Pearson correlation between row i and column j of the Hi-C matrix (derived from the .hic file generated earlier). The eigenvector of this correlation matrix determines the compartment type and strength for each bin in the matrix. Generally, regions with positive eigenvector values are labeled as belonging to the 'A' compartment, while those with negative values are labeled as 'B' compartments. The "fanc compartments" command is utilized to generate a correlation matrix (AB compartment) object from a FAN-C matrix file. Eigen vectors are computed to assign genome bins into A or B compartments.

2. Generating AB compartment domains

The AB compartment file generated from the hi-c pairs were taken as an input to generate a BED file for the respective A and B domains. The domains are written to the BED file in the general BED format with 4 distinct columns, The starting base, The end base and the feature. Feature is generally either A or B compartment and a Quality Score in the fourth column. The domains files generated from the fanc compartment file was used to annotate the individual AB compartments along with their genomic locations. Consecutive matrix bins with the same eigenvector sign are considered part of a “domain.”

The domains BED file merges all consecutive bins in the same domain, which is why A and B are always alternating. It contains the domain type in the “name” field and the average eigenvector entry values of all bins in the domain in the “score” field. The Negative score corresponds to B compartment whereas Positive score corresponds to A compartment.

3. Analysis of the domain files

The domain files were used as an input for structural annotation of the genome compartments. The genomic locus of each compartment was extracted individually from the domains file and was given as an input to various annotation software and packages.

We used **PeakAnalyzer** (Salmon-D et al,2010) and **HOMER** (Hypergeometric Optimization of Motif EnRichment) (Heinz S et al, 2010). PeakAnalyzer is a JAVA based software which has its own GUI and HOMER is a perl-based package for motif enrichment and discovery. Both tools were used to find a consensus gene list which corresponds to the genomic locations of our individual A and B compartments.

To do this, all the genes from the Human genome which overlapped with our regions identified from the domains file were extracted and made a list for further analysis. We did this for all 4 categories of data, viz genes from wt p53 A and B compartment and genes from mutant p53 containing A and B compartment. Distinct set of genes overlapping with the feature were identified and taken for further analysis.

RNA-Sequencing data analysis

This section describes the various steps to perform RNA-Seq data analysis and interpretation. RNA-seq data was used to quantify the gene expression data from the Cancer cell lines of mut p53. The entire workflow can be simplified as:

1. Data collection
2. Data Quality Control
3. Mapping
4. Gene Expression Quantification
5. Normalization
6. Differential Gene Expression Analysis using DeSeq2
7. Downstream Analysis

Data Collection

Sequence reads were retrieved from the SRA (https://www.ncbi.nlm.nih.gov/sra) and ENA Browser (https://www.ebi.ac.uk/ena/browser/) by utilizing their corresponding accession IDs in FASTQ format for p53 and mutant p53 cancer cell lines.

Table3: Dataset for RNA-Seq data for wt p53

Cell Line	**GSE**	**GSM**	**SRR**	**Replicates**
HCT 116 +|+	GSE124993	GSM3560679 GSM3560680 GSM3560681	SRR8435995 SRR8435996 SRR8435997	3
IMR 90	GSE202664	GSM6128029 GSM6128030 GSM6128031	SRR19159300 SRR19159299 SRR19159298	3
U20S	GSE232592	GSM7356394 GSM7356395	SRR24572365 SRR24572364	2

Table4: Dataset for RNA-Seq data for mut p53

Cell Line	**GSE**	**GSM**	**SRR**	**Replicates**
HCT 116 +|-	GSE124993	GSM3560676 GSM3560677 GSM3560678	SRR8435992 SRR8435993 SRR8435994	3
SW480	GSE231706	GSM7296438 GSM6128039 GSM6128040	SRR24442521 SRR24442520 SRR24442519	3
MDA-MB-231	GSE220931	GSM6829540 GSM6829541	SRR22729523 SRR22729522	3

		GSM6829542	SRR22729521	
MDA-MB-231 LML	GSE217287	GSM6710491 GSM6710492 GSM6710493	SRR22192981 SRR22192979 SRR22192978	3

Data Quality Control

Following the retrieval of sequence reads, the reads underwent quality control checks. The quality assessment of all sequenced reads was performed using FASTQC version 0.12.0, which generated a FASTQC report. Subsequently, based on the quality report, poor quality reads having contaminating adapter sequences were subjected to trimming using cutAdapt version 1.12. The FASTQC program provides a set of green/orange/red flags that correspond to good/warning/failure messages.

Read Mapping

The filtered reads were mapped to the human reference genome. The genome assembly taken as a reference for this study is Hg38 everywhere. To perform the RNA-Seq alignment HISAT2 (Kim, D., et al, 2019) was used as the aligner as it is a very fast and efficient tool to perform alignment of the transcripts. HISAT2, similar to BWA and Bowtie, employs the Burrows-Wheeler Transform (BWT) for genome compression, reducing memory storage needs significantly. The alignment process of HISAT2 occurs in two stages: initial mapping locates exons for unspliced reads, followed by independent alignment of split unmapped reads to identify exon junctions.

Gene Expression Quantification

After the Read Mapping and alignment process, the next step in RNA-Seq analysis is the Read counts. Read counts allows us to measure and compare gene expression patterns at an unprecedented resolution. The alignment file contains millions of reads, Gene Expression Quantification involves computationally mapping the “reads” onto a reference genome to reveal a “transcriptional map,” where the number of reads aligned to each gene gives a measure of its level of expression.

To do the read count, we used the HTSeq command line tool. HTSeq is a high throughput sequence analysis tool used to quantify gene expression from RNA-Seq data.

Normalizing the Reads and Differential Gene Expression Analysis using DESeq2

It is important to do normalization of the gene count data obtained from HTSeq count to account for the biases generated from varying sequencing depth between the samples and to account for changes in gene length. The longer a gene is, the more reads are going to be mapped against it but count of read mapping against that gene would not be an accurate representation of expression of that gene, therefore we need to perform normalization of the raw read counts across all samples.

Knight Ruiz Normalization is incorporated with the DESeq2 package, we did not normalize each sample count individually but leveraged the ability of DESeq2 to that by itself.

DESeq2 (Love M et al,2014) is a R package designed for differential gene expression analysis based on the negative binomial distribution. It estimates the gene-wise dispersions (variance) and uses empirical Bayes technique to shrink these estimates. It fits a negative binomial model to the counts, which allows for greater variability than expected under a Poisson process. Deseq2 also provides several functions for visualizing the data and the results.

For DESeq2 analysis, we had total 8 samples from wt p53 and 12 samples from mut p53 (including replicates), the condition set was such that, all the wild-type samples were compared against all the mutant samples as it can handle complex experimental designs, such as multiple groups with multiple replicates.

RESULTS

P53 mutation causes changes in chromatin topology and function

Through the utilization of Hi-C data, we were able to gain a comprehensive understanding of the three-dimensional chromatin structure of the genome, organized into distinct compartments. Upon visual inspection and subsequent analysis of the generated plots, an observable difference in the chromatin structural organization was observed between the wild-type and p53 mutant samples. This observed difference was represented as a heatmap, with each bin corresponding to 1 Mb of genomic size. This process was carried out for all 22 autosomes and the sex chromosomes (X and Y), and a contact map for AB enrichment was plotted for visual analysis of the differential chromatin organization. About ~30% change in the Genome-Wide Active-Inactive topology between the wild-type and mutant condition was observed.

Presented here is a heatmap for Chromosome 17 in both the wild-type and mutant conditions. The rationale behind the selection of Chromosome 17 for display is twofold: the impracticality of presenting plots for all chromosomes, and the relevance of Chromosome 17 due to the location of the TP53 gene on this chromosome.

Wild-type p53:

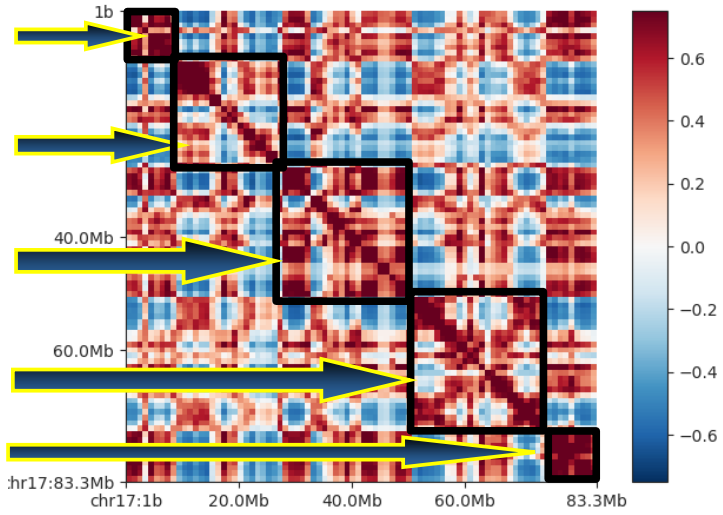

Figure1: Contact Heatmap of wt p53 domains depicting the TADs present

The blue regions show the inactivated regions of the chromosome, i.e. B compartment whereas red regions show activated regions, i.e. A compartments in the genome. The gradient vertical bar shows the enrichment scores. From this heatmap, we can observe Topologically associating domains (TAD) at the position between 1mb- ~10mb, 10mb-30mb, 35mb – 50mb, ~50Mb - 75Mb, ~75mb – 83.3mb.

Mutant p53:

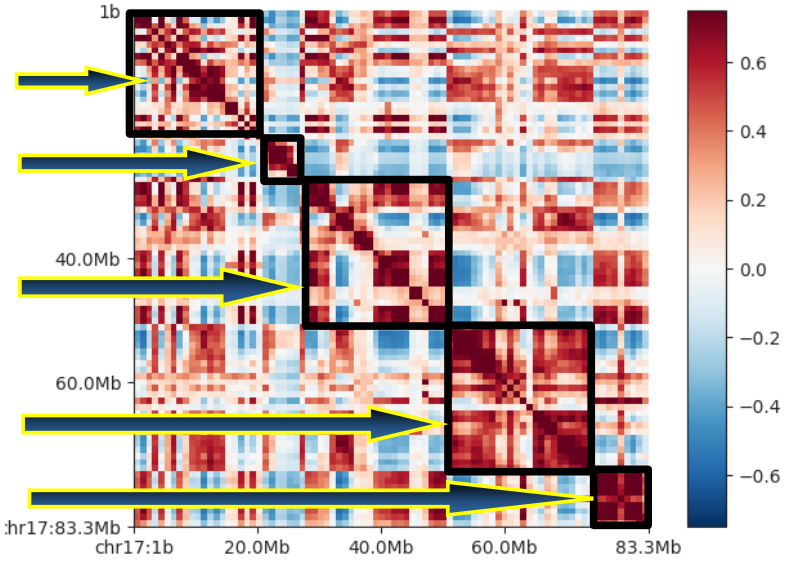

Figure2: Contact Heatmap of mut p53 domains depicting the TADs present

The blue regions show the inactivated regions of the chromosome, i.e. B compartment whereas red regions show activated regions, i.e. A compartment in the genome.

From this contact map of mut p53 we can see the TAD between the 55Mb- ~75Mb is highly enriched for interactions than the wild type plot. The initial TAD between 1Mb-10Mb is also effectively lost. There is an extra small TAD being observed between the positions ~25mb-~35mb. The TAD between 35mb – 55Mb is also seen considerably reduced for interactions than the wt TAD at the same position.

TADs here are thought to act as functional units in the genome. They co-localize genes and their regulatory elements, forming the unit of genome switching between active and inactive compartments. Most studies indicate that TADs regulate gene expression by limiting the enhancer-promoter interaction to each TAD (Long et al,2022).

Pertaining to these observed changes we identified set of genes falling on these regions across genome which show differential chromatin structural organisation between the two instances, viz, containing wt p53 expression or mutant p53 expression condition

In our investigation, we identified a set of 'Compartment switching genes.' These genes exhibit dynamic behaviour, transitioning between distinct compartments. Specifically:
1) Genes shift from active topologies in wild-type phenotype to inactive topologies in the mutant phenotype
2) Conversely, other genes move from the inactive topologies in wild-type phenotype to active topologies in mutant phenotype.

To achieve this, we employed a straightforward comparison technique. Our initial input consisted of gene sets from each compartment across multiple samples. Total number of genes extracted from each category were as given below.

Table 5: Gene set extracted from domain files

Wild-type p53 A compartment genes	**10692 genes**
Wild-type p53 B compartment genes	**9021 genes**
Mutant p53 A compartment genes	**10224 genes**
Mutant p53 B compartment genes	**9367 genes**

We then compared the commonalities between wt A compartment genes and mut B compartment genes, and vice versa. This analysis yielded a list of genes shared between the two compartments. For instance, genes present in both the wt A and mut B compartments were identified as 'Compartment switching genes.'

Similarly, we performed a comparative analysis for genes between the wt active compartment and the mut inactive compartment. As a result, we identified a total of **5758 genes** transitioning from the wild-type A compartment to the mutant B compartment, while **5343 genes** shifted from the wild-type B compartment to the mutant A compartment.

Variations in Genes expression in mutant p53 as compared with the wild-type samples

Utilizing RNA-seq data, we discerned differentially expressed genes in both wild-type and mutant p53 conditions. These alterations reflect differential gene expression across the entire transcriptome, rather than being solely attributed to conformational changes in chromatin. To visualize these differential gene expression patterns, we generated plots for enhanced clarity.

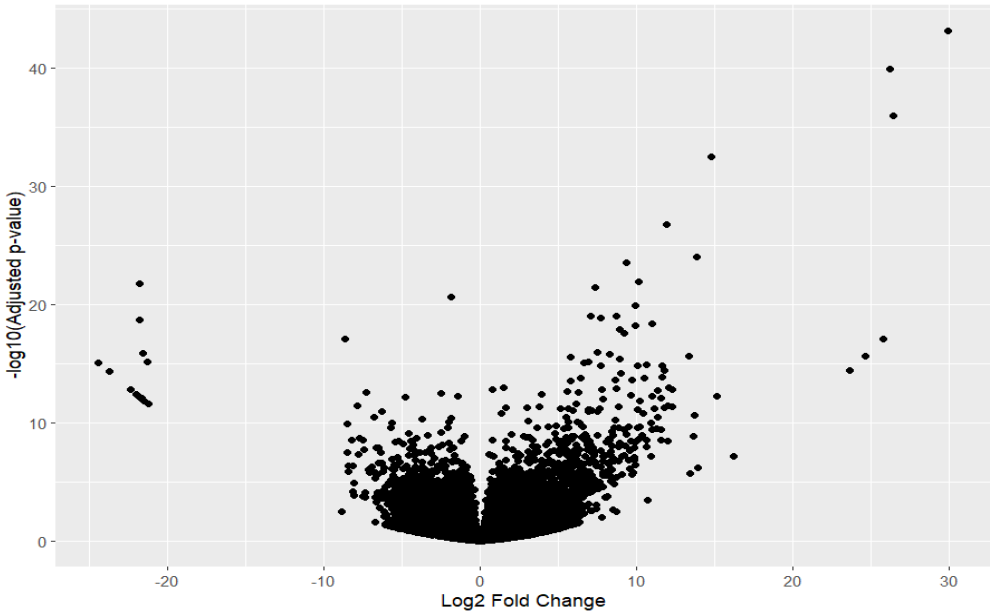

Figure3: plot of Log2fold changes vs adjusted p value

Each individual dots represent a gene deviating from the log2fold change value of 0, i.e no change in gene expression.

From this plot we can see that some genes have high negative log2fold change values (<-20) which are about 15 in number whereas there are also 6 genes which have a very high positive values of log2fold change indicating >20-fold upregulation of these genes in the mut p53 samples.

Principal Component Analysis

Furthermore, we also did the principal component analysis to plot the variances within our samples. We expected high variance from the sample because we intentionally took

different cell lines to account for cell-line specific changes and to minimize that. We wanted to make sure that the changes in the expression that we are seeing here are not because of the cell line specific variations within the cancer cells.

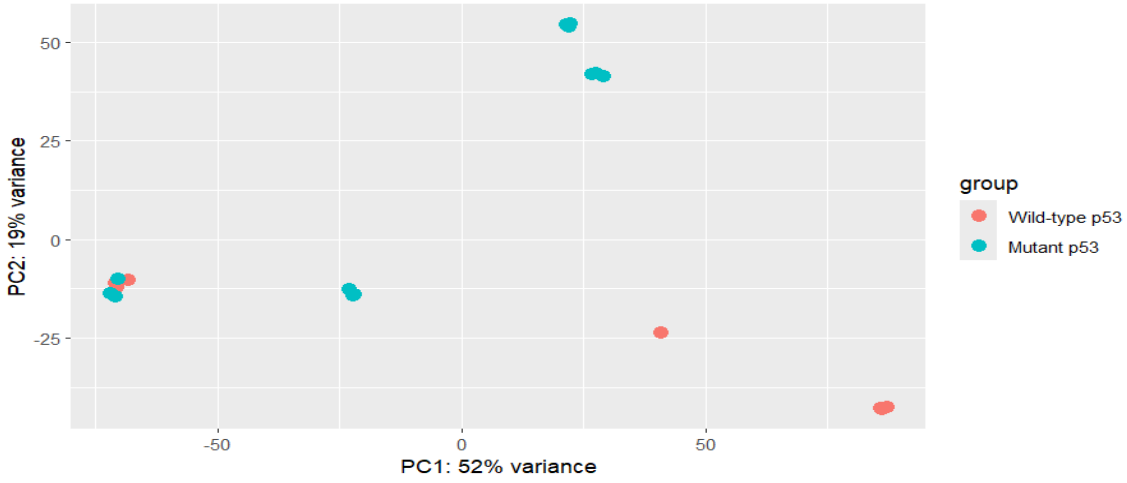

Figure4: Plot of variances in PC between wt and mut rna-seq samples

From the RNA-seq data analysis we identified a total of 2769 **genes** that were getting downregulated in mut p53 with a (cutoff -1), as compared to wt p53.

Total upregulated genes (cutoff +1) in mut p53 with respect to the wt sample were **2647 genes.**

Table 6: The top 10 upregulated genes from DGE analysis of RNA-seq data

Gene symbols	**Name**
IGFBP5	insulin like growth factor binding protein 5(IGFBP5)
CXCL12	C-X-C motif chemokine ligand 12(CXCL12)
FAM162B	family with sequence similarity 162 member B(FAM162B)
NDN	necdin, MAGE family member (NDN)
MXRA5	matrix remodeling associated 5(MXRA5)
CHRM2	cholinergic receptor muscarinic 2(CHRM2)

HGF	hepatocyte growth factor (HGF)
ZNF677	zinc finger protein 677(ZNF677)
TRPA1	transient receptor potential cation channel subfamily A member 1(TRPA1)
COL1A2	collagen type I alpha 2 chain (COL1A2)

Table 7: The top 10 downregulated genes from DGE analysis of RNA-seq data

Gene Symbols	**Name**
CD8B2	CD8B family member 2(CD8B2)
NANOG	Nanog homeobox(NANOG)
DEFA5	defensin alpha 5(DEFA5)
GSX1	GS homeobox 1(GSX1)
CCNA1	cyclin A1(CCNA1)
LINC01169	long intergenic non-protein coding RNA 1169(LINC01169)
LAPTM5	lysosomal protein transmembrane 5(LAPTM5)
LINC00868	long intergenic non-protein coding RNA 868(LINC00868)
NRROS	negative regulator of reactive oxygen species(NRROS)
MT-TV	mitochondrially encoded tRNA valine (MT-TV)

Connecting the link between conformation change and gene expression regulation

With our hic data analysis we got 2 different sets of "Compartment Switching Genes". These gene list contained 5758 and 5343 genes respectively.

Functional annotation and pathway enrichment was done for these gene sets using **Enrichr , PANTHER** and **ShinyGO**

Table8: Important functionality of compartment switching genes in wild-type and mutant p53 phenotypes

P53	**Mut P53**
Calcium-Dependent Cell-Cell Adhesion Via Plasma Membrane Cell Adhesion Molecules (GO:0016339)	Antigen processing and presentation of endogenous peptide antigen (GO:0002483)
Cellular Response to Zinc Ion (GO: 0071294)	MHC class II protein complex assembly (GO:0002399)
CXCR Chemokine Receptor Binding (GO:0045236)	Peptide antigen assembly with MHC protein complex (GO:0002501)
Sulfuric Ester Hydrolase Activity (GO:0008484)	Negative Regulation of hormone secretion (GO:0046888)
Arylsulfatase Activity (GO:0004065)	RNA pol II Cis-Regulatory and Trancription Regulatory Region sequence-specific DNA Binding (GO:0000977)

Alcohol Dehydrogenase Activity – Zinc Dependent (GO:0004024)	Deoxycytidine Deaminase Activity (GO:0047844)
Adrenergic Receptor and Histone Deacetylase Binding (GO:0031690)	NK-cell Lectin like Receptor Binding (GO:0046703)
Cysteine Type Endopeptidase Inhibitor Activity (GO:0004869)	Inhibitory MHC class I Receptor Activity (GO:0032396)
Acid-Thiol Ligase Activity (GO:0016878)	Cis Regulatory region sequence-specific DNA Binding (GO:0000987)

Pertaining to these observed Gene Ontologies there are multiple studies validating that these processes are involved in Oncogenesis and Cancer progression.

Calcium-Dependent Cell-Cell Adhesion: This process is crucial for maintaining the integrity of tissues and can be disrupted in cancer. Changes in cell adhesion molecules like cadherins can lead to increased tumour invasiveness and metastasis (Peng et al, 2020).

Cellular Response to Zinc Ion: Zinc ions play a role in various cellular functions, including DNA synthesis and repair, and cell proliferation. Abnormal zinc metabolism has been associated with cancer, where zinc transporters and levels can influence tumour growth. (Sugimoto et al,2024)

CXCR Chemokine Receptor Binding: CXCR4 receptor binding plays a key role in tumour growth, survival, angiogenesis, metastasis, and therapeutic resistance. (Bianchi ME & Mezzapelle R, 2020)

Sulfuric Ester Hydrolase Activity: Sulfatase enzymes like SULF2 are implicated in cancer progression and are correlated with tumour-associated macrophages in gastric cancer. (Wang T et al, 2021)

Arylsulfatase Activity: Arylsulfatase B has been suggested as a biomarker in prostate cancer and is involved in the degradation of chondroitin sulfate, which impacts the tumour microenvironment (TME). (Feferman et al, 2013)

Alcohol Dehydrogenase Activity – Zinc Dependent: Alcohol dehydrogenases are associated with the prognosis of various cancers, including hepatocellular carcinoma. (Liu et al,2020)

Adrenergic Receptor Histone Deacetylase Binding: Beta-adrenergic receptor signalling is involved in various tumour environments and is associated with cancer progression and metastasis (Mravec et al,2020). Aberrant Histone deacetylases (HDACs) activity is linked to cancer development. (Bruna B & Maribel P,2012)

Cysteine Type Endopeptidase Inhibitor Activity: Type II cystatin genes, which include cysteine protease inhibitors, have been associated with the prognosis of gastric cancer and other malignancies. (Daher B et al,2020). While specific studies on acid thiol ligase activity in cancer were not found, thiol-containing amino acids like cysteine are crucial for redox balance and are implicated in ferroptosis, a form of cell death. (Daher B et al,2020)

We used ShinyGO to annotate the role of compartment switching genes **from inactive topologies to active topologies** in mutant p53 phenotype in **pathways related to cancer.**

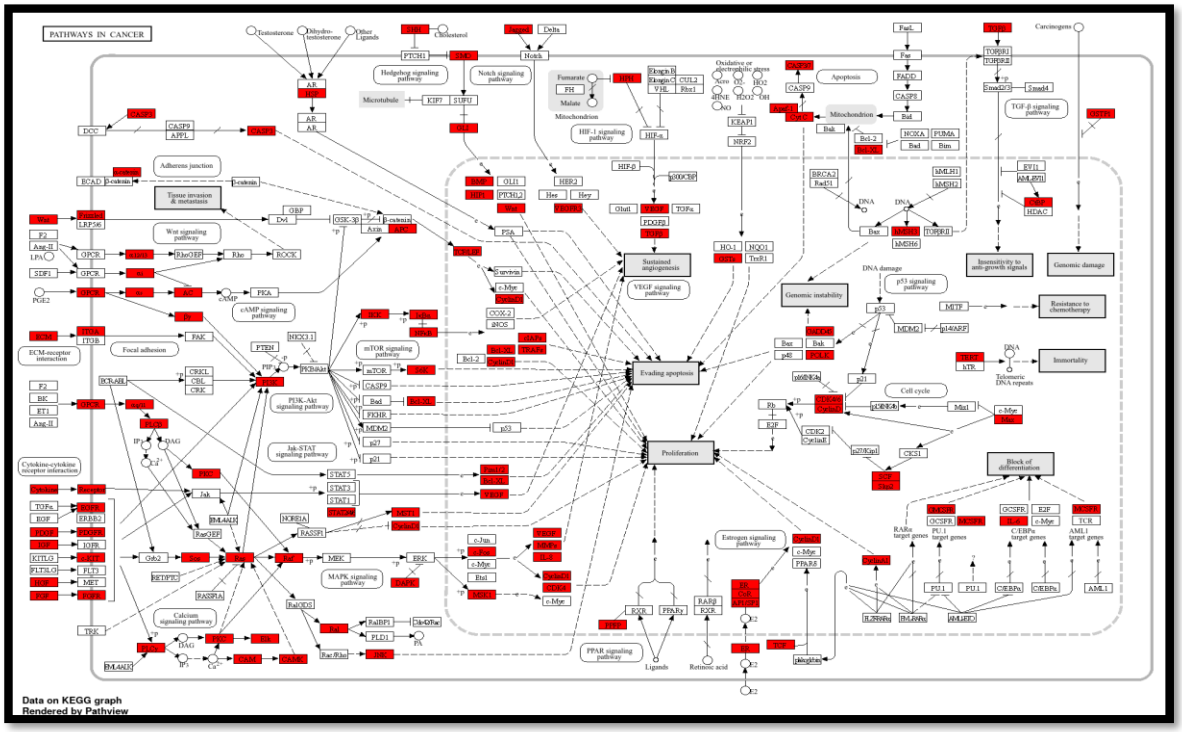

Figure5: Pathway map from KEGG, highlighted in red are the compartment switching genes from inactive to active topology in mutp53 phenotype.

We then used ShinyGO to annotate the role of compartment switching genes **from active topologies to inactive topologies** in mutant p53 phenotype in **pathways regulating cell cycle**

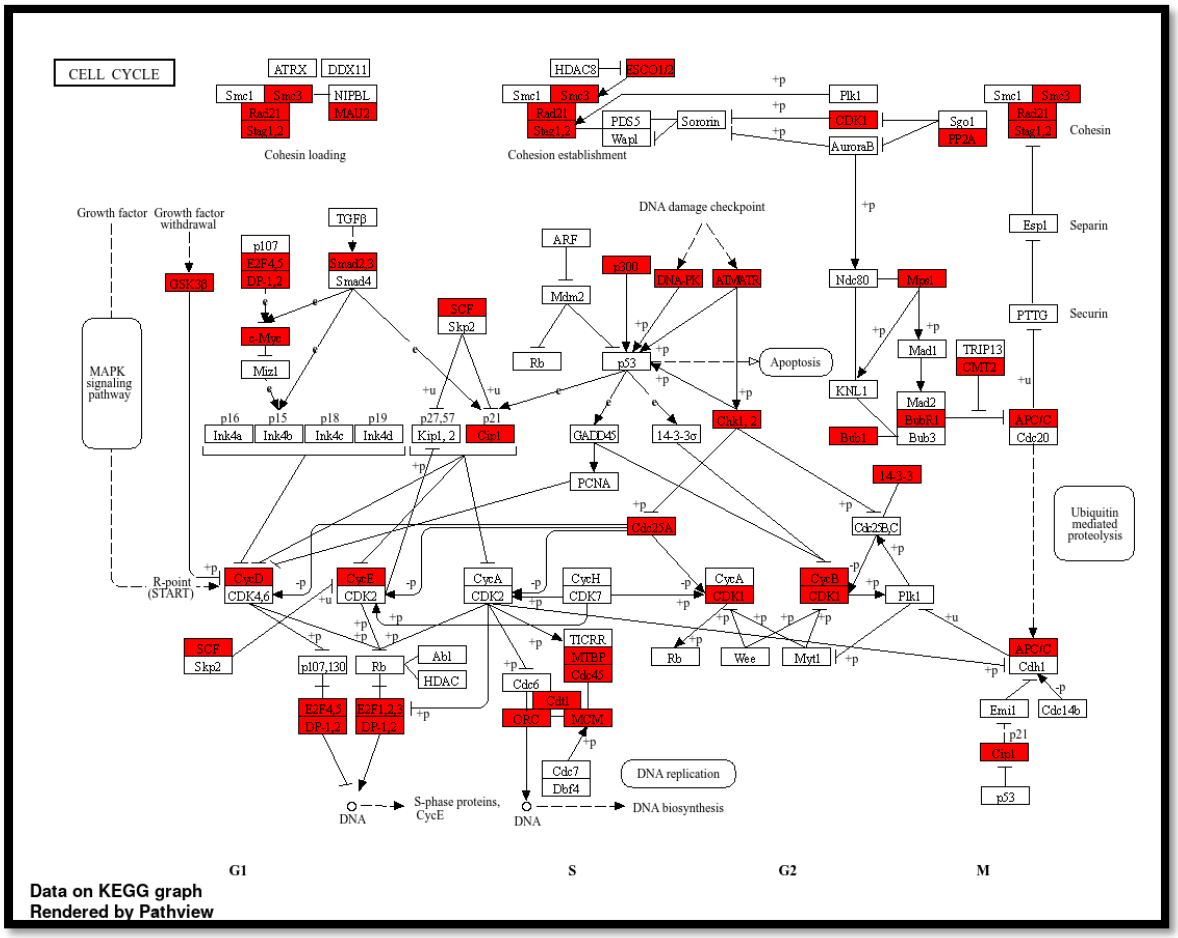

Figure6: Pathway map from KEGG, highlighted in red are the compartment switching genes from active to inactive topology in mutp53 phenotype.

Correlating the chromatin topology and genome functioning

In this segment of the research, we attempted to correlate gene expression data in mutant p53 with chromatin organization. This involved a comparative analysis to ascertain the overlap between compartment switching genes and differentially expressed genes. The premise of our comparison was predicated on the hypothesis that genes transitioning from **inactive regions in wt p53 phenotype to active regions in mut p53 phenotype** would be **upregulated**, while those transitioning from **active regions in wt p53 phenotype to inactivated regions in mut p53 phenotype** would be **downregulated**.

For this comparative analysis, we extracted lists of upregulated and downregulated genes from our DGE results, which were obtained via RNA-seq data analysis using DESeq2. We then assessed for commonalities between our initial gene list—comprising genes transitioning from wild-type B to mutant A—and the list of upregulated genes. A parallel comparison was conducted for downregulated genes and genes transitioning from wild-type A to mutant B, with the common genes being duly recorded.

We identified a total of **791 genes** that were common between the wild-type B to mutant A transition and the upregulated gene sets. In a similar fashion, **1898 genes** were found to be common between the wild-type A to mutant B transition and the downregulated gene sets. Subsequently, these gene sets were functionally annotated using Gene ontology and pathway enrichment analysis to decipher the relation between chromatin organisation and Gene expression.

1) Genes going from wt B to mut A and upregulated

Biological processes

Top biological processes identified sorted by p-value ranking were:

- Cell-Cell adhesion via plasma membrane adhesion molecules (GO:0098742)
- Calcium-Dependent Cell-Cell adhesion via plasma membrane cell adhesion molecules (GO:0016339)

- Synapse assembly (GO:0007416)
- Anterograde Trans-synaptic signalling (GO:0098916)

GO molecular Function

- Arylsulfatase Activity (GO:0004065)
- Sulfuric Ester Hydrolase Activity (GO:0008484)
- Cytokine Receptor Activity (GO:0004896)
- Voltage-Gated Potassium Channel Activity (GO:0005251)

Upon conducting a Gene Ontology analysis of the 791-gene set, we uncovered evidence for biological processes and molecular functions that are relevant to oncogenesis and cancer progression. The genes mapped against the Jensen DISEASES database also showed high confidence of these genes being involved in various cancer types.

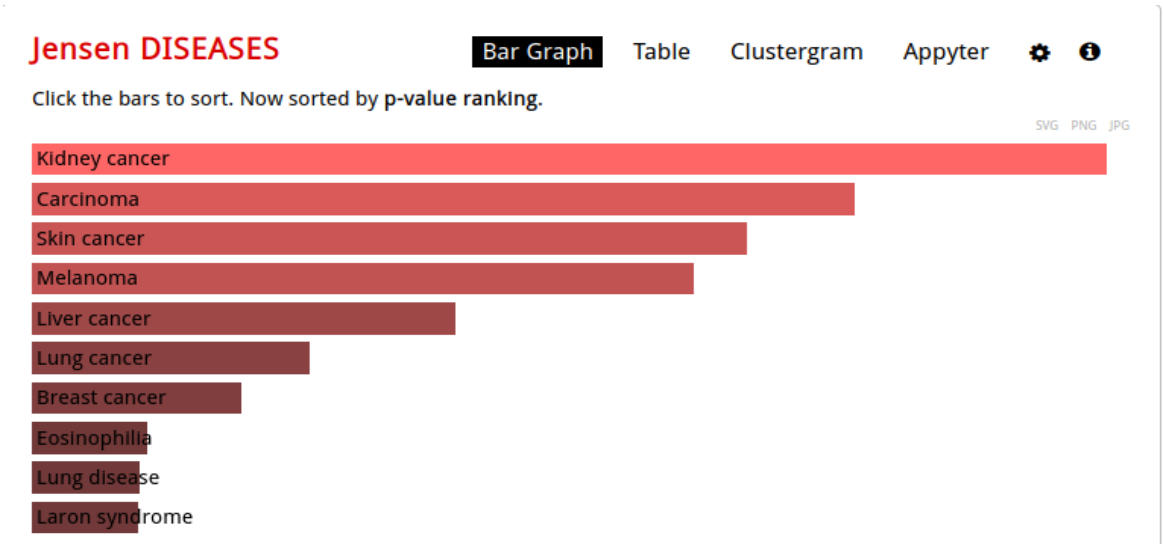

Figure9: Top hits obtained from Jensen DISEASE database

Pathway enrichment analysis using PANTHER

The significant pathways enriched for our gene set were for:

- Cadherin Signalling Pathway in Homo sapiens (P00012) and
- Wnt Signalling in Homo sapiens (P00057).

It has been observed that disruption of cadherin signalling pathway can lead to tumorigenesis, tumour progression, angiogenesis and altered tumour immune response. The Wnt/ Beta- catenin signalling pathway is also involved in cell proliferation, differentiation, and stem cell renewal. Dysregulation of this pathway contributes to the development and progression of various cancers. In the canonical Wnt pathway, the stabilization and accumulation of β-catenin in the nucleus activate transcription of target genes that promote cellular proliferation and survival. Abnormal activation of the Wnt pathway is a recognized driver of cancer, particularly colorectal cancer. (34)

Both pathways are interconnected, with cadherin-catenin complexes influencing Wnt signalling. For example, the loss of E-cadherin can lead to the release of β-catenin, which then participates in Wnt signalling, promoting oncogenic transcription.

We then analysed our gene set against MsigDB Hallmark gene-sets. MsigDB is a comprehensive knowledge base for biochemical pathways, signalling cascades and other biological concepts. Hallmark gene sets are consistently co expressed across multiple biological conditions. The most significant result of such an enrichment was – **Epithelial-mesenchymal transition (EMT).** EMT is a fundamental biological process where epithelial cells acquire mesenchymal traits, leading to enhanced migratory capacity, invasiveness, and increased resistance to apoptosis. In the context of cancer, EMT is a critical event that contributes to cancer progression and metastasis. It allows carcinoma cells to suppress their epithelial characteristics and adopt mesenchymal features, which facilitates their mobility and the ability to migrate from the primary tumour site.

2) Genes going from wt A to mut B and down-regulated

This gene set consists of **1898** unique genes that were found common between the compartment switch from activated regions in wt p53 to inactivated regions in mut p53 respectively.

GeneOntology Analysis:

Biological processes

The top-most significant hits sorted by p-value, were.

- Regulation of DNA-templated DNA Replication (GO:0090329)
- Amino Acid Catabolic Process (GO:0009063)
- Regulation of Mitotic Nuclear Division (GO:0007088)
- Regulation Of DNA-templated DNA Replication Initiation (GO:0030174)
- Positive Regulation Of Nuclear Division (GO:0051785)

Molecular functions

The top-most significant hits sorted by p-value, were:

- Single-Stranded DNA Helicase Activity (GO:0017116)
- Chemokine Receptor Activity (GO:0004950)
- UDP-galactose:beta-N-acetylgalactosamineBeta-1,3-Galactosyltransferase Activity (GO:0008499)
- Magnesium Ion Binding (GO:0000287)
- G Protein-Coupled Chemoattractant Receptor Activity (GO:0001637)
- Nucleoside Triphosphate Diphosphatase Activity (GO:0047429)
- Beta-1,3-Galactosyltransferase Activity (GO:0048531)

Pathway enrichment analysis using PANTHER:

Pathway enrichment analysis was done, results of which are shown below:

- Apoptosis signalling pathway (P00006)
- 2-arachidonoylglycerol biosynthesis (P05726)
- Axon guidance mediated by semaphorins (P00007)
- Thyrotropin-releasing hormone receptor signalling pathway (P04394)
- B cell activation (P00010)

- T cell activation (P0053)

Pathway enrichment analysis and GO top hit corresponds to Apoptosis signalling pathway which is an indicative of roles of these genes in cell death mediation, thereby giving more confidence to our hypothesis that mutation in p53 affects the 3D chromatin organisation which in turn affects various gene functioning and indeed helps in cancer progression

The pathway enrichment analysis for these genes also strengthened our hypothesis that the genes switching to inactivated regions and getting downregulated are involved in crucial checkpoint regulators for cancer proliferation. These results suggest that these genes which are involved in normal cell cycle progression and cell response to DNA damage are getting inactivated in the mut p53 indicating that mut p53 is changing the expression of these genes through process of chromatin remodelling.

We did the MsigDB Hallmark gene sets, and the results are shown below:

- Allograft Rejection
- INF alpha response
- INF Gamma Response
- KRAS signaling downregulation
- G2-M checkpoints

The outcomes of this analysis indicate that the identified gene set plays a pivotal role in diverse cellular immune responses, thereby regulating the cell cycle and preserving the coherence of intercellular interactions. The observed downregulation of KRAS signalling implies the activation of a tumour-suppressive mechanism that ensures cellular regulation, thereby inhibiting cellular transformation into tumour formations.

Discussions

The phenomenon of ‘Compartment switching genes can have significant implications in cancer. The spatial organization of the genome into A and B compartments is crucial for regulating gene expression. Disruption of TAD boundaries have been found to be associated with a wide range of diseases including cancer. The mechanisms underlying TAD formation are complex and not yet fully elucidated, though several protein complexes and DNA elements are associated with TAD boundaries. Compartment switching may disrupt this organization, leading to aberrant gene expression patterns. Changes in chromatin compartmentalization are associated with cell differentiation and cancer progression. 30% of genomic regions change compartments during these processes, thus influencing gene regulation, DNA damage repair, replication, and the physical state of the cell.

Abnormal phase separation and the resulting misfolding of chromatin loops can potentiate oncogene activation. It was shown that the normal compartment structures of chromatin are compromised in tumors, indicating that compartment switching may be a hallmark of tumor progression.

In summary, compartment switching genes may contribute to the development and progression of cancer by altering the three-dimensional organization of the genome, affecting gene expression in mutant p53 phenotypes.

CONCLUSIONS

1. Mutations in p53's DBD transforms it into a MAR-Binding protein which can induce chromatin remodeling upon binding with MAR elements in the DNA.

2. Mutant p53 was shown to bring about changes in the Genome Organization as observed from the chromatin conformation capture data analysis.

3. These changes are suggestive of oncogenic driver mechanisms and signaling cascades getting activated where anti-tumor activity and normal cell growth is affected.

FUTURE WORK

1. Selection of genes from the identified gene set which has MAR sequence in their promoter region or within 5kb distance.

2. Selection of genes which have mutant p53 binding sequence in their promoter or within 5kb distance.

3. RT qPCR to check the transcript level of those genes which we get from the RNA seq results.

References

1. Hu, J., Cao, J., Topatana, W. et al. Targeting mutant p53 for cancer therapy: direct and indirect strategies. J Hematol Oncol 14, 157 (2021). https://doi.org/10.1186/s13045-021-01169-0

2. Parrales A and Iwakuma T (2015) Targeting Oncogenic Mutant p53 for Cancer Therapy. Front. Oncol. 5:288. doi: 10.3389/fonc.2015.00288

3. Hainaut P, Pfeifer GP. Somatic TP53 Mutations in the Era of Genome Sequencing. Cold Spring Harb Perspect Med. 2016 Nov 1;6(11):a026179. doi: 10.1101/cshperspect.a026179. PMID: 27503997; PMCID: PMC5088513.

4. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010 Jan;2(1):a001008. doi: 10.1101/cshperspect.a001008. PMID: 20182602; PMCID: PMC2827900.

5. Shirole NH, Pal D, Kastenhuber ER, Senturk S, Boroda J, Pisterzi P, Miller M, Munoz G, Anderluh M, Ladanyi M, Lowe SW, Sordella R. TP53 exon-6 truncating mutations produce separation of function isoforms with pro-tumorigenic functions. Elife. 2016 Oct 19;5:e17929. doi: 10.7554/eLife.17929. Erratum in: Elife. 2017 Feb 01;6: PMID: 27759562; PMCID: PMC5092050.

6. Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J, Spitzer B, Bosbach B, Kastenhuber ER, Baslan T, Ackermann S, Cheng L, Wang Q, Niu T, Schultz N, Levine RL, Mills AA, Lowe SW. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. Nature. 2016 Mar 24;531(7595):471-475. doi: 10.1038/nature17157. Epub 2016 Mar 16. PMID: 26982726; PMCID: PMC4836395.

7. Kruse, K., Hug, C.B. & Vaquerizas, J.M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol 21, 303 (2020). https://doi.org/10.1186/s13059-020-02215-9

8. Dhaka B, Sabarinathan R. Differential chromatin accessibility landscape of gain-of-function mutant p53 tumours. BMC Cancer. 2021 Jun 5;21(1):669. doi: 10.1186/s12885-021-08362-x. PMID: 34090364; PMCID: PMC8180165.

9. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc. 2015 Apr 13;2015(11):951-69. doi: 10.1101/pdb.top084970. PMID: 25870306; PMCID: PMC4863231.

10. Salmon-Divon, M., Dvinge, H., Tammoja, K. et al. PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics 11, 415 (2010). https://doi.org/10.1186/1471-2105-11-415

11. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PMID: 20513432; PMCID: PMC2898526.

12. Long, H.S., Greenaway, S., Powell, G. et al. Making sense of the linear genome, gene function and TADs. Epigenetics & Chromatin 15, 4 (2022). https://doi.org/10.1186/s13072-022-00436-9

13. Peng, Y., Wu, D., Li, F. et al. Identification of key biomarkers associated with cell adhesion in multiple myeloma by integrated bioinformatics analysis. Cancer Cell Int 20, 262 (2020). https://doi.org/10.1186/s12935-020-01355-z

14. Ensieh Farahani, Hirak K. Patra, Jaganmohan R. Jangamreddy, Iran Rashedi, Marta Kawalec, Rama K. Rao Pariti, Petros Batakis, Emilia Wiechec, Cell adhesion molecules and their relation to (cancer) cell stemness, Carcinogenesis, Volume 35, Issue 4, April 2014, Pages 747-759, https://doi.org/10.1093/carcin/bgu045

15. De Santis, M., Mantovani, A. & Selmi, C. The other side of the innate immune system: humoral arms favoring cancer. Cell Mol Immunol 17, 1024-1025 (2020). https://doi.org/10.1038/s41423-020-0512-x

16. Sugimoto, R., Lee, L., Tanaka, Y. et al. Zinc Deficiency as a General Feature of Cancer: a Review of the Literature. Biol Trace Elem Res 202, 1937-1947 (2024). https://doi.org/10.1007/s12011-023-03818-6

17. Guertin, D.A., Wellen, K.E. Acetyl-CoA metabolism in cancer. Nat Rev Cancer 23, 156 172 (2023). https://doi.org/10.1038/s41568-022-00543-5

18. Hao Y, Yi Q, XiaoWu X, WeiBo C, GuangChen Z and XueMin C (2022), Acetyl-CoA: An interplay between metabolism and epigenetics in cancer. Front. Mol. Med. 2:1044585. doi: 10.3389/fmmed.2022.1044585

19. Bianchi ME and Mezzapelle R (2020) The Chemokine Receptor CXCR4 in Cell Proliferation and Tissue Regeneration. Front. Immunol. 11:2109. doi: 10.3389/fimmu.2020.02109

20. Tong-Bo Wang, Ze-Feng Li, Xiao-Jie Zhang et al. SULF2 Is a Prognostic Biomarker and Correlated with Tumor Associated Macrophages in Gastric Cancer, 19 November 2021 https://doi.org/10.21203/rs.3.rs-1083793/v1

21. Feferman, L., Bhattacharyya, S., Deaton, R. et al. Arylsulfatase B (N-acetylgalactosamine-4-sulfatase): potential role as a biomarker in prostate cancer. Prostate Cancer Prostatic Dis 16, 277 284 (2013). https://doi.org/10.1038/pcan.2013.18

22. Guo, J., Song, Z., Muming, A. et al. Cysteine protease inhibitor S promotes lymph node metastasis of esophageal cancer cells via VEGF-MAPK/ERK-MMP9/2 pathway. Naunyn-Schmiedeberg's Arch Pharmacol (2024). https://doi.org/10.1007/s00210-024-03014

23. Daher B, Vuceti ˇ c M and Pouysségur J ´(2020) Cysteine Depletion, a Key Action to Challenge Cancer Cells to Ferroptotic Cell Death. Front. Oncol. 10:723. doi: 10.3389/fonc.2020.00723

24. Daher B, Vucetic M and Pouysségur J ´(2020) Cysteine Depletion, a Key Action to Challenge Cancer Cells to Ferroptotic Cell Death. Front. Oncol. 10:723. doi: 10.3389/fonc.2020.00723

25. Liu, X., Li, T., Kong, D. et al. Prognostic implications of alcohol dehydrogenases in hepatocellular carcinoma. BMC Cancer 20, 1204 (2020). https://doi.org/10.1186/s12885-020-07689-1

26. Mravec, B.; Horvathova, L.; Hunakova, L. Neurobiology of Cancer: The Role of -Adrenergic Receptor Signaling in Various Tumor Environments. Int. J. Mol. Sci. 2020, 21, 7958. https://doi.org/10.3390/ijms21217958

27. Barneda-Zahonero Bruna, Parra Maribel, (2012), Histone deacetylases and cancer, Molecular Oncology, 6, doi: https://doi.org/10.1016/j.molonc.2012.07.003.

28. Hai R, He L, Shu G and Yin G (2021) Characterization of Histone Deacetylase Mechanisms in Cancer Development. Front. Oncol. 11:700947. doi: 10.3389/fonc.2021.700947

29. Zhang, Y., Wang, X. Targeting the Wnt/ -catenin signaling pathway in cancer. J Hematol Oncol 13, 165 (2020). https://doi.org/10.1186/s13045-020-00990-3

30. Ramos FS, Wons L, Cavalli IJ, Ribeiro EM (2017) Pressurized Intra Peritoneal Aerosol Chemotherapy (PIPAC) in patients suffering from peritoneal carcinomatosis of biliary carcinoma. Integr Cancer Sci Therap. 4: DOI: 10.15761/ICST.1000243.

31. Dhouha Grissa, Alexander Junge, Tudor I Oprea, Lars Juhl Jensen, Diseases 2.0: a weekly updated database of disease–gene associations from text mining and data integration, Database, Volume 2022, 2022, baac019, https://doi.org/10.1093/database/baac019

Project Report:

32. Differential regulation of chromatin by p53 and mutant p53: Possible role of chromatin conformation in carcinogenesis, Amisha Gupta,2023

URLs:

33. GEO: https://www.ncbi.nlm.nih.gov/geo/

34. ENA: https://www.ebi.ac.uk/ena/browser/

35. Enrichr: https://maayanlab.cloud/Enrichr/

36. PANTHER: https://pantherdb.org/

37. ShinyGO; http://bioinformatics.sdstate.edu/go/

38. Ensembl: https://www.ensembl.org/info/data/ftp/index.html?redirect=no

Tools:

- FANC: Kruse, K., Hug, C.B. & Vaquerizas, J.M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. Genome Biol 21, 303 (2020). https://doi.org/10.1186/s13059-020-02215-9
- fastQC (v0.12.0): https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Cutadapt (v1.12): DOI:10.14806/ej.17.1.200
- Bowtie2 (v2.5.2): Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.
- Hisat2: Kim, D., Paggi, J.M., Park, C. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37, 907–915 (2019)
- HTSeq: G Putri, S Anders, PT Pyl, JE Pimanda, F Zanini, Analysing high-throughput sequencing data in Python with HTSeq 2.0 (2022).https://doi.org/10.1093/bioinformatics/btac166
- PeakAnalyzer: Salmon-Divon M, Dvinge H, Tammoja K, Bertone P (2010) PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics.
- HOMER: Heinz S, Benner C, Spann N, Bertolino E et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 2010 May 28;38(4):576-589. PMID: 20513432
- DESeq2: doi:10.18129/B9.bioc.DESeq2

Appendix

FANC – Protocols used for HI-C data

Read Mapping –

For generating HiC matrices, the paired end reads are treated as single end reads and are mapped iteratively. All the default parameters were used.

Command: bowtie2 –x <Index_name> -U <Input_filename.fastq> –S <output_filename.sam>

The resulting SAM file which store reads aligned to a reference genome was converted to its indexed , block-compressed binary format, BAM that provides fast, direct access to any part of the file, using SAMtools (Li, Handsaker et al. 2009).

Command: samtools view –bS <input_filename.sam> -o <Output_filename.bam>

The resulting BAM files were sorted by coordinates using “sort” option with the command.

Command: samtools sort –n <input_filename.bam> > <output_sort.bam>

Command: samtools merge -b -n <input_bam_files.bam> -o <Merged.bam>

The resulting merged BAM file were again sorted using the samtools sort command.

All the default parameters were used.

The aforementioned procedures were executed on all downloaded sequence reads

pertaining to both wt p53 and mut p53

Generating FANC pairs-

The input given were 2 BAM files (of a single paired-end reads) sorted by read name (-n option was used in the previous step) and output was a FANC pairs file.

Command: fanc pairs R1.bam R2.bam output.pairs –r HindIII –g GRCh38.p13.genome.fa

-g: Path to genome file (FASTA, folder with FASTA, HDF5 file), which will be used in conjunction with the type of restriction enzyme to calculate fragments directly. Here, FASTA file of Hg38 assembly was downloaded from GENCODE.

-r: Name of the restriction enzyme used in the experiment, e.g. HindIII

All the default parameters were used.

Generating HiC objects

Command: fanc hic <output.pairs> output.hic –b 1mb

-b : Bin size in base pairs. Human-readable formats, such as 10k, or 1mb. If omitted, the command will end after the merging step. Rest all default parameters were used.

Computing AB compartments matrix

Command: fanc compartments output.hic output.ab –g GRCh38.p13.genome.fa

Input: hic object file

Output: AB compartment matrix file

-g : Genome file. Used to "orient" the eigenvector values (change sign) using the average GC content of domains. Possible input files are FASTA, folder with FASTA, comma separated list of FASTA) used to change sign of eigenvector based on GC content

Generating AB compartment domains

The " fanc compartments" command can also write the AB domains in the bed file when it is coupled with the -d command.

Command: fanc compartments -d output.ab domains.bed –g GRCh38.p13.genome.fa

Input: AB compartment matrix file

Output: Domains file (BED)

HOMER - annotatePeaks.pl domains.bed

Protocols in RNA-Sequencing Data Analysis

Data Quality Control

Command: : fastqc <Input_filename> Cutadapt –a <Adapter_sequence> -o <output_file-name> <input_filename>

Read Mapping

The reference genome used was Hg38. The first step was to build an index of the reference genome. To do this, the following command was used.

Command : hisat2-build <reference_genome.fa> <index_name>

Input: Reference genome assembly Hg38

Output: Indexed genome files

After building the index for the reference genome, we performed the alignment using the hisat2 command.

Command for single-end FASTA reads DNA alignment :

hisat2 -f -x <index_name> -U <reads in fastq or.fa format> -S output.sam --no-spliced-alignment

Command for paired-end FASTQ reads alignment :

hisat2 -x <index_name> -1 <read 1 in fastq or.fa format> -2 <reads2 in fastq or.fa format> -S output.sam

All the default parameters were used

Input: Paired end/Single end FASTQ read file(s)

Output: SAM file

Command: samtools view –bS <input_filename.sam> -o <Output_filename.bam>

Input: SAM file generted by HISAT2

Output: BAM file

Command: samtools sort –n <input_filename.bam> > <output_sort.bam>

-n parameter sorts the SAM/BAM file based on the read names.

Input: Unsorted BAM file

Output : Sorted BAM file

Command: samtools index sorted-input.bam

Input: sorted BAM file

Output: .BAI file

Gene Expression Quantification

Command : htseq-count -f bam -s no -a 10 input.bam genes.gtf > output.txt

-f bam specifies that the input file is in BAM format

-s no specifies that the sequencing library is not strand-specific

-a 10 filters out all reads with a mapping quality less than 10

Input: Sorted BAM file and a feature file (GTF/GFF)

Output: Csv file containing Gene identifiers and number of reads mapped alongside it.

The feature file for this analysis was downloaded from ensemble