Differential regulation of chromatin by p53 and mutant p53: Possible role of chromatin conformation in carcinogenesis

A project submitted to the

Bioinformatics Centre

Savitribai Phule Pune University

For the degree of

M. Sc. in Bioinformatics

By

Amisha Gupta

Under the guidance of

Dr. Abhijeet Kulkarni

(Guide)

Bioinformatics Centre, SPPU

May 2023

CERTIFICATE

This is to certify that the project entitled **"Differential regulation of chromatin by p53 and mutant p53: Possible role of chromatin conformation in carcinogenesis",** submitted by **Ms. Gupta Amisha** in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics, has been carried out satisfactorily by her at the Bioinformatics Centre, Savitribai Phule Pune University.

Date: 22/05/23
Place: Pune

Dr.Sangeeta Sawant
Director
Bioinformatics Centre
Savitribai Phule Pune University,
Pune 411007

CERTIFICATE

This is to certify that the project entitled **"Differential regulation of chromatin by p53 and mutant p53: Possible role of chromatin conformation in carcinogenesis"**, submitted by **Ms. Gupta Amisha** in partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics, has been carried out satisfactorily by her/him at the Bioinformatics Centre, Savitribai Phule Pune University, under my guidance and supervision.

Dr. Abhijeet Kulkarni
(Guide)
Bioinformatics Centre, SPPU

Date: 22/05/23
Place: Pune

DECLARATION & UNDERTAKING

I hereby declare that the project entitled, **"Differential regulation of chromatin by p53 and mutant p53: Possible role of chromatin conformation in carcinogenesis"**, 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 thereof 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.

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 want to thank the DBT for providing studentship.

Date – 22/05/23 Amisha Gupta

Place : Pune BIM-2021-09

ACKNOWLEDGEMENT

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.

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 Upreti for insightful discussions, patience, and invaluable feedback. Her willingness to give his time has been very much appreciated. Her positive and analytical thinking made it all happen so productively. I want to thank the DBT for providing their constant support. I express my heartfelt greatness to all the non-teaching and technical staff for their continuous support.

I would like to thank my friends for constant help and support and my family for their encouragement.

Amisha Gupta

BIM-2021-09

TABLE OF CONTENTS

List of Abbreviations

Abbreviation	Description
Mut	Mutant
WT	Wild Type
GEO	Gene Expression Omnibus
ENA	European Nucleotide Archive
GO	Gene Ontology
SAM	Sequence Alignment Map
BAM	Binary Alignment Map
ChIP	Chromatin Immuniprecipitation
BED	Browser Extensible Data
FANC	feature-rich framework for the analysis and visualisation of chromosome conformation capture data
ChIP-Seq	Chromatin Immunoprecipitation coupled with high throughput sequencing
MACS	Model based analysis of ChIP-Seq

List of Tables

List of figures

ABSTRACT

p53, a nuclear transcription factor, plays a vital role in regulating the cell cycle during DNA damage and other stress conditions. As a tumor suppressor, p53 acts as a key regulator of essential cellular processes, including cell cycle arrest, DNA repair, apoptosis, and senescence. Its significance in maintaining cellular homeostasis and preventing DNA damage accumulation has made it a central focus of cancer research.

Frequent mutations in p53 not only abolish its tumor suppressor capabilities but also confer various gain-of-function (GOF) activities. These activities profoundly impact molecules and pathways that are now recognized as crucial for tumor development and progression. Mutant p53 exhibits a high affinity for a variety of MAR-DNA elements, leading to conformational changes in chromatin. This provides a structural basis for the independent spatial and temporal regulation of gene expression and initiation of DNA synthesis.

Deciphering how mutations in wild-type p53 induce conformational changes in chromatin and understanding their contribution to carcinogenesis can significantly enhance therapeutic intervention strategies. To investigate this, we performed analyses of ChIP-seq and HiC data using p53 and mutant p53-specific control cell lines. Through these analyses, we successfully identified unique genes and pathways regulated by p53 and mutant p53, shedding light on their potential roles in tumor suppression and development/progression, respectively.

Furthermore, we examined the 3D conformation of the genome in these cell lines, unraveling how differential chromatin regulation by these proteins can activate or inactivate oncogenic and tumor-suppressor genes, thereby providing valuable insights for therapeutic interventions.

Keywords – p53, Mut p53, chromatin conformation, cancer

1. INTRODUCTION

p53: guardian of the genome

p53 is encoded by the TP53 gene located on the small arm of chromosome 17.

Cellular stresses, including DNA damage, lead to genomic aberrations such as mutations, deletions, and translocations, resulting in genomic instability and the potential development of cancers. To safeguard against this, cells require an effective stress response to maintain genomic integrity and prevent malignant transformation. The transcription factor p53 plays a crucial role in this process. Normally expressed at low levels, p53 is kept in a functionally latent state through proteasomal degradation mediated primarily by the E3 ubiquitin protein ligase MDM2. However, upon DNA damage, post-translational modifications such as phosphorylation and acetylation trigger p53 to accumulate in the nucleus. This modification process releases p53 from MDM2, activating it. Activated p53 transactivates a specific set of target genes involved in cell cycle arrest and/or apoptosis induction, depending on the extent and types of DNA damage. Cell cycle arrest allows for DNA repair, after which cells can resume normal cell cycle progression. In cases of severe DNA damage, p53 initiates apoptosis, eliminating cells with extensive DNA damage and preventing the transmission of such damage to future generations of cells. Thus, p53 plays a crucial role in maintaining genomic integrity by promoting appropriate cellular responses to DNA damage.

The DNA-binding domain of p53 interacts with the tandem repeat of the p53-responsive element (consisting of the sequence RRRCWWGYYY, where R represents G/A, W represents A/T, and Y represents C/T). These elements are found within the promoter regions of p53 target genes and are separated by a maximum of 13 base pairs. In response to a variety of cellular stresses, such as DNA damage and energetic stress, p53 induces cell cycle arrest and/or apoptosis through the transactivation of its target genes.

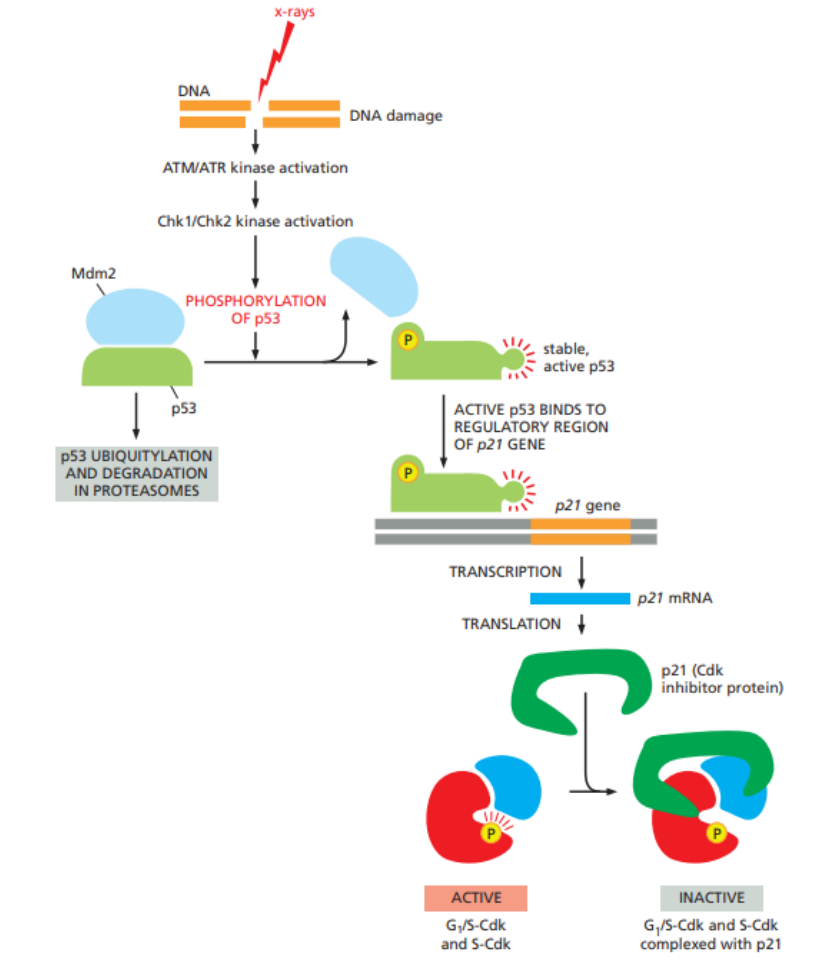


Figure 1: Pathway showing activation of p53 in response to DNA damage and its role as cell cycle checkpoint regulator (Adopted from Molecular Biology of THE CELL- 6th edition by Bruce Alberts).

Mutant p53

Mutational inactivation is a prevalent mechanism leading to the dysfunction of p53, with loss of function mutations of p53 found in over half of human cancers. Specifically, 95% of these mutations are localized within exons 5-8, which encode the DNA-binding domain of p53. Analysis of mutation profiles revealed that certain amino acid residues, including **Arg-175, Gly-245, Arg-248, Arg-249, Arg-273, and Arg-282**, are frequently mutated in human cancers. These mutations within the DNA-binding domain disrupt the proper

conformation of p53, rendering the mutant p53 unable to activate transcription through the wild-type p53-binding consensus element. Furthermore, mutant p53 exhibits a dominant-negative effect on wild-type p53 by forming hetero-tetramers and possesses oncogenic potential. Consensus binding motif mutant p53 has not yet been identified.

It is now seen that mutp53 does not only suppress that genes that act as cell cycle checkpoints and stop the cell cycle in case of damaged DNA but also activate the expression of many growth-promoting and oncogenic genes which can lead to the progression of the aggressive cancers.

In this study we tried to identify the genes that are uniquely regulated by p53 and mutp53 in order to understand the unique pathways that can lead to cancer development and progression. We also tried to identify the consensus binding motif of mut p53 via computational tools.

Chromatin Conformation

Chromatin, the complex structure formed by DNA and proteins, plays a crucial role in organizing the genome within the three-dimensional (3D) nuclear space. It efficiently packages the DNA strands of each cell's genome within the limited volume of the nucleus while facilitating proper gene expression and replication. The organization of chromatin within the nucleus is of significant importance for both gene-specific biological functions and broader nuclear processes.

Chromatin can be visually distinguished into two distinct regions: a darker, electron-dense region known as heterochromatin, and a lighter region called euchromatin. The positioning of genes within the 3D chromatin architecture is closely associated with their activity levels (Shachar and Misteli, 2017). The space occupied by an individual chromosome within the nucleus is referred to as a "chromosomal territory" (CT) Cremer and Cremer, 2001; Dixon et al., 2016). Additionally, the genome exhibits hierarchical organization at multiple scales: at the megabase level, there are A/B compartments, and at the sub-megabase level, there are topologically associated domains (TADs) (Dixon et al., 2016).

A compartments constitute the Active region of the genome where as B compartments constitute the inactive region of the genome.

Differential chromatin regulation of p53 and mutant-p53

Expression of mutant (mut) p53 is not equivalent to a p53 “null” situation , as genetic and biological evidence indicates that mut p53 exerts oncogenic functions of its own. Mutant p53 has a strong affinity for various MAR-DNA elements (MARs), unlike the wild-type p53, suggesting that the binding of mutant p53 to MARs may be linked to its dominant-oncogenic activities. MARs recognized by mutant p53 share a high AT content and contain variations of the AATATATTT "DNA-unwinding motif." This motif enhances the dynamic nature of chromatin structure and facilitates local unwinding of DNA strands. Mutant p53 specifically interacts with MAR-derived oligonucleotides that carry these unwinding motifs, leading to DNA strand separation, particularly when this motif is situated within a structurally flexible sequence environment. MARs organize the cellular chromatin into topologically independent loops, thereby providing a structural basis for the independent spatial and temporal regulation of gene expression and initiation of DNA synthesis. Such a regulation is thought to form a higher-order regulatory mechanism for controlling development and differentiation.

Differential chromatin regulation of both the proteins can result in activation and inactivation of many oncogenic and tumor-suppressor genes respectively, identification of a mut p53-specific activity that possibly could be involved in the modulation of chromatin structure and function might provide clues to better understand the activities of mut p53 related to its “gain of function” properties.

We therefore adopted HiC and ChIP-Seq techniques to determine binding pattern of this protein and to check it has effects on 3D organization of genome within nucleus. Towards this, existing data mininf was employed as a strategy to investigate possible roles of Mut p53 in carcinogenesis via alteration in 3D chromatin organization.

ChIP-Seq:

Chromatin immunoprecipitation (ChIP) is an experimental method to study protein-DNA interactions. It can be used to determine the locations of binding sites on the genome of transcription factors and epigenetic factors such as modified histones.

By integrating chromatin immunoprecipitation (ChIP) assays with sequencing techniques, ChIP sequencing (ChIP-Seq) represents a potent approach for identifying DNA binding sites of transcription factors and other proteins across the entire genome. In ChIP-Seq, DNA-bound proteins are initially immunoprecipitated using specific antibodies, followed by the co-precipitation, purification, and subsequent sequencing of the bound DNA molecules.

The utilization of next-generation sequencing (NGS) technology in ChIP has revolutionized our understanding of gene regulatory mechanisms implicated in various diseases and biological processes, including developmental processes and cancer advancement. ChIP-Seq facilitates comprehensive investigation of protein-nucleic acid interactions at a genome-wide level, providing valuable insights into the intricate dynamics and functional implications of these interactions.

Chromatin Conformation capture techniques

These techniques are used to study the physical structure of chromosomes in the cell nucleus. These methods identify areas of DNA that are brought together in close proximity as a result of looping.

Hi-C

The Hi-C method identifies physical proximity between pairs of chromosomal loci in an unbiased manner on a genomic scale, without relying on binding to any specific protein or immunoprecipitation (Lieberman-Aider et al. 2009) . The experimental method relies on the use of formaldehyde to form DNA-protein-DNA bridges between DNA loci that are near each other in the nucleus. The DNA is digested with a restriction enzyme; the cleaved ends are labeled with biotin, and then ligated under conditions that favor ligation events

between the cross-linked DNA fragments that were originally in close proximity in the nucleus, marked with biotin at the junction.

The DNA is sheared, biotin-containing fragments are selected, and a sequencing library is constructed.

The fragments must be sequenced as paired-ends to identify the two different chromosomal locations that were cross-link-mediated ligation.

The Hi-C interaction data indicate a pattern of chromosome looping that correlated quite closely with linear genomic distance. Chromatin region preferentially interact with nearby regions, and inter-chromosomal interactions are much more rare.

(Adopted from Next-Generation DNA Sequencing Informatics- II Edition by Stuart M. Brown)

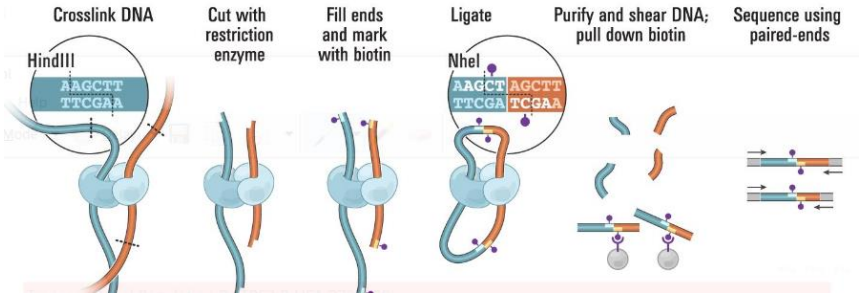


Figure 2: Overview of Hi-C protocol (Lieberman-Aiden et al. 2009)

2. MATERIALS & METHODS

2.1 ChIP-Seq data analysis

This section describes various steps to perform ChIP-Seq data analysis and interpretation. The entire workflow can be simplified as:

1. Data collection
2. Data Quality Control
3. Mapping
4. Peak Calling
5. Non-redundant merged peak coordinates
6. Peak Annotation
7. Downstream Analysis

2.1.1 Data Collection

Sequence reads obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) were retrieved by utilizing their corresponding accession IDs to download the data from ENA (https://www.ebi.ac.uk/ena/browser/home) in FastQ format for p53(Table 1) and mut p53(Table 2) specific cell lines. Control cell lines were used for this study. Corresponding Input DNA control pr mock IgG control reads were also downloaded for peak calling.

Dataset : Table 1- Dataset for p53

Cell Line	**GSE**	**GSM**	**SRR**	**GSM(Input)**	**SRR(Input)**
HCT116	GSE125106	GSM356354	SRR8443719	GSM3563536	SRR8443721
IMR90	GSE42728	GSM1048850	SRR628591	GSM1048853	SRR628594
U2OS	GSE46641	GSM1133488	SRR847016	GSM1133489	SRR847017
Erythroid cells	GSE157208	GSM4758996	SRR12556664	GSM4758998	SRR12556666
Human neonatal foreskin keratinocytes	GSE56674	GSM136689	SRR1231841	GSM1366685	SRR1231837
Lymphoblastoid cells	GSE46991	GSM1142697	SRR1231847	GSM1142699	SRR851810
A382	GSE158549	GSM4802678	SRR12708831	GSM4802677	SRR12708830

Table 2 – Dataset for mut p53

Cell Line	**GSE**	**GSM**	**SRR**	**GSM(Input)**	**SRR(Input)**
HCT116	GSE125106	GSM3563533	SRR8443718	GSM3563535	SRR8443720
SW480	GSE102796	GSM2746539	SRR594462 & SRR594462	GSM2746548	SRR5944081
MDA MB 231	GSE66543	GSM1624706	SRR1829691	GSM1624707	SRR18299692
MDAH087 fibroblast cell line	GSE36749	GSM900168	SRR448039	GSM900167	SRR448038
MDA-MB-231 LM2	GSE95305	GSM2501568	SRR5282136	GSM2501570	SRR5282138

2.1.2 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 0.11.3, 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. During ChIP-seq analysis many of these warnings are ignored such as over-represented sequences because these warnings assume a sample of random and diverse DNA fragments.

Command: fastqc <Input_filename>

Cutadapt –a <Adapter_sequence> -o <output_filename> <input_filename>

2.1.3 Read Mapping

The filtered reads were mapped to the reference genome (hg38 Built) downloaded from UCSC Browser using Bowtie2 v2.2.6. **Bowtie 2** is an ultrafast 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) . The indexed file for the reference genome GRCh38 was obtained from https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip.

For unpaired reads, following command was used:

Command: bowtie –x <Index_name> -U <Input_filename> -S <Output_filename.sam>

For paired reads, following command was used:

Command: bowtie2 –x <Index_name> -1 Read1.fastq -2 Read2.fastq –S <output_filename.sam>

All the default parameters were used.

The resultant 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 version 1.7 which uses htslib 1.7 (Li, Handsaker et al. 2009).

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

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

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

All the default parameters were used.

The aforementioned procedures were executed on all downloaded sequence reads pertaining to both p53 and mut p53, as well as their corresponding input samples separately.

2.1.4 Peak Calling

This step intended to find the genomic regions that indicate binding sites of p53 and mut p53. To identify regions exhibiting notable enrichment of ChIP signals in comparison to the background (DNA input or mock IgG samples), peak calling was performed on the sequencing reads aligned to the genome using Model-based Analysis for ChIP-Seq (MACS2), version 2.1.1 (Zhang, Liu et al. 2008) tool.

Command: macs2 callpeak –t <Treated_sample.bam> -c <input_sample.bam> -f BAM –n File_name_output

All the default parameters were used.

Following this procedure, a BED file containing the chromosome name, peak start position, and peak end position is obtained.

2.1.5 Creating non-redundant dataset

This step intended to create a non-redundant BED file for both p53 and mut p53, i.e., combining overlapping or "book-ended" features in an interval file into a single feature which spans all of the combined features using "merge" option of BedTools, version 2.25.0 (Aaron et al.).

For this first all the BED files were sorted by chromosome and then by Start position in ascending order using "sort" option of BedTools.

Command: bedtools sort –i <Input_filename.bed> > sorted.bed

After sorting all the BED files, non-redundant data set for p53 and mut p53 were created using "bedtools merge" which merge overlapping repetitive elements into a single entry.

Command: bedtools merge –i <names of all BED files> > merged.bed

Non-redundant BED files for p53 and mut p53 were retrieved after this step.

All the default parameters were used.

2.1.6 Peak Annotation

After obtaining the BED coordinates which indicate putative binding sites of p53 and mut p53 , the subsequent step was to annotate the enriched regions in relation to genomic features throughout the entire genome in the sequencing reads. This annotation process aimed to associate biological significance and contextual relevance to the identified features.

An R package called ChIPseeker (Yu, Wang et al. 2015) was used to annotate the peaks that took input in the form of 3 columns BED file.

2.1.7 Functional Analysis

Functional enrichment analysis of the gene set that are uniquely regulated by p53 and mut p53 was done using Enrichr (E Y Chen et al.) to identify key pathways and ontologies associated with the gene set.

The nucleotide sequences corresponding to the peak coordinates were downloaded from the UCSC genome browser (Kent, Sugnet et al. 2002) using an in-house perl script and was used as an input to MEME-ChIP (Ma, Noble et al. 2014) to identify consensus binding sequence or motif sequence in peak region.

2.2 HiC data analysis

This section describes various steps to perform HiC data analysis and interpretation. FANC (v.0.9.26b) (Kruse, K., Hug, C.B. & Vaquerizas, J.M.): a feature-rich framework for the analysis and visualisation of chromosome conformation capture data tool was used for this purpose.

The entire workflow can be simplified as:

1. Data collection
2. Mapping FASTQ files to the reference genome
3. Generating and filtering Read pairs
4. Generating, binning and filtering HiC objects
5. Analysing HiC matrices

2.2.1 Data Collection

Sequence reads obtained from the GEO database were retrieved by utilizing their corresponding accession IDs to download the data from ENA in FastQ format for p53 and mut p53 (Table 3) specific cell lines. Control untreated cell lines were used for this study.

Table 3: Dataset for HiC Analysis

Cell Line	**p53 status**	**GSE**	**GSM**	**SRR**	**Read-type**
HCT116	WT	GSE158007	GSM4784017	SRR12646280	Paired
FHC	Mut	GSE133928	GSM3930267	SRR9675751	Paired

2.2.2 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 ultrafast 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) . The indexed file for the reference genome GRCh38 was obtained from https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip.

For generating HiC matrices, the paired end reads are treated as single end reads and are mapped iteratively.

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

All the default parameters were used.

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>

All the default parameters were used.

The aforementioned procedures were executed on all downloaded sequence reads pertaining to both p53 and mut p53

2.2.3. 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 was 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.

2.2.4. Generating HiC objects

The fanc hic command is used to generate fragment-level and binned Hi-C matrices.

Fanc pairs file was used as the input

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.

2.2.5 Analysing HiC matrices

1. Computing AB compartments matrix

Regions in a Hi-C matrix can generally be assigned to either the active or the inactive compartment, also called ‘A’ and ‘B’ compartments, respectively. Compartments are derived from a correlation matrix, in which each entry i, j corresponds to the Pearson correlation between row i and column j of the (Hi-C) matrix (.hic file generated in the previous step). The eigenvector of the correlation matrix is used to derive compartment type and strength for each matrix bin. Generally, regions with positive values are assigned the ‘A’, regions with negative values the ‘B’ compartment.

The “fanc compartments” command can produce a correlation matrix (AB compartment) object from a FAN-C matrix file.

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

Input: hic object file

Output: AB compartment matrix file

Eigen vectors are calculated which is used to assign the genome bins into A or B compartments

-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.

2. Plotting AB compartments map for Visualization

AB compartments maps are plotted for all the 23 chromosomes for both the p53 and mutp53 compartment matrix generated to visualize any changes in the compartmentalization of the genome due to mutation in the p53 protein.

Command: fancplot -o plot.png chr<n> \-p square output.ab \-vmin -0.75 -vmax 0.75 -c RdBu_r

3. RESULTS

3.1 ChIP Seq Data Analysis

3.1.1 Quality checking of sequencing reads

All the sequences passed the quality criteria, with few warnings as during ChIP-seq analysis many of these warnings are ignored such as over-represented sequences because these warnings assume a sample of random and diverse DNA fragments.

3.1.2 Read Mapping

The mapping generated output files in human human non-readable BAM format with alignment percentage as shown below:

Table 4: Alignment percentage of samples when reads mapped to reference human genome(Hg38)

Cell Line	**SRR**	**Alignment percentage**	**SRR (Input)**	**Alignment percentage**
HCT116	SRR8443719	94.08%	SRR8443721	95.91%
IMR90	SRR628591	87.56%	SRR628594	83.48%
U2OS	SRR847016	92.02%	SRR847017	96.45%
Erythroid cells	SRR12556664	98.07%	SRR12556666	89.71%
Human neonatal foreskin keratinocytes	SRR1231841	91.69%	SRR1231837	91.09%
Lymphoblastoid cell	SRR1231847	97.54%	SRR851810	92.49$
A382	SRR12708831	72.04%	SRR12708830	78.76%

Table 5: Alignment percentage of samples when reads mapped to reference human genome(Hg38)

Cell Line	**SRR**	**Alignment percentage**	**SRR(Input)**	**Alignment percentage**
HCT116	SRR8443718	93.46%	SRR8443720	92.34%
SW480	SRR594462 & SRR594462	97.64%	SRR5944081	98.87%
MDA MB 231	SRR1829691	68.82%	SRR18299692	72.35%
MDAH087 fibroblast cell line	SRR448039	89.64%	SRR448038	83.46%
MDA-MB-231 LM2	SRR5282136	92.07%	SRR5282138	94.68%

3.1.3 Peak Calling

This analysis resulted in the discovery of genomic regions exhibiting an abundance of aligned reads, indicating potential binding sites for transcription factors. The output files produced encompassed various formats, including R, xlsx, BIGWIG, and other peak calling files, each providing distinct information such as statistical significance in terms of FDR, enriched values, and tags associated with the aligned reads.

To consolidate the data, the bed files from different p53 cell lines and mutant p53 were merged, resulting in the creation of two non-redundant files containing peak coordinates.

3.1.4 Peak Annotation

Annotating peaks led to the identification of upstream and downstream elements of enriched peak regions. We retrieved the nearest genes around the peaks and used the genes for further downstream analysis.

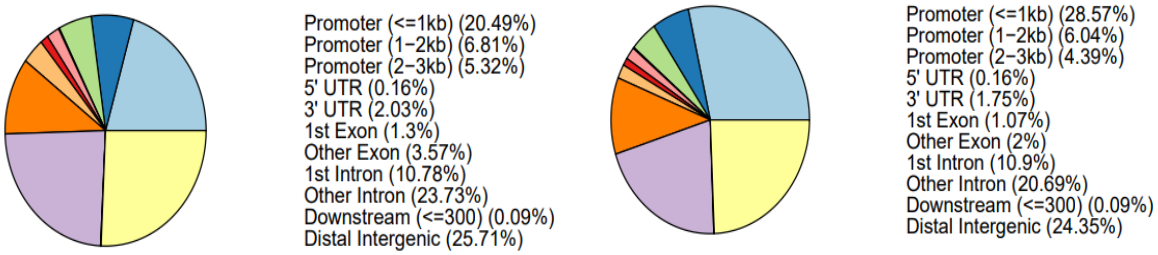
Promoter (<=1kb)	1	12010	13670	1661	1	1E+08	ENST00000450305.2	10	ENSG00000223972	DDX11L1
Promoter (<=1kb)	1	17369	17436	68	2	1.02E+08	ENST00000619216.1	541	ENSG00000278267	MIR6859-1
Promoter (<=1kb)	1	17369	17436	68	2	1.02E+08	ENST00000619216.1	0	ENSG00000278267	MIR6859-1
Promoter (<=1kb)	1	35245	36073	829	2	645520	ENST00000461467.1	955	ENSG00000237613	FAM138A
Promoter (<=1kb)	1	89295	120932	31638	2	1.01E+08	ENST00000466430.5	-413	NA	LOC100996442
Promoter (<=1kb)	1	120725	133723	12999	2	1.01E+08	ENST00000610542.1	748	NA	LOC100996442
Promoter (<=1kb)	1	139790	140339	550	2	729737	ENST00000493797.1	731	NA	LOC729737
Promoter (<=1kb)	1	165491	169210	3720	2	1.01E+08	ENST00000662089.1	0	NA	LOC100996442
Promoter (<=1kb)	1	182696	184174	1479	1	1.03E+08	ENST00000624431.2	-733	ENSG00000223972	DDX11L17
Promoter (<=1kb)	1	182696	184174	1479	1	1.03E+08	ENST00000624431.2	448	ENSG00000223972	DDX11L17
Promoter (<=1kb)	1	187891	187958	68	2	1.02E+08	ENST00000612080.1	870	ENSG00000273874	MIR6859-2
Promoter (<=1kb)	1	185217	195411	10195	2	1.03E+08	ENST00000623083.4	116	NA	WASH9P
Promoter (<=1kb)	1	365395	368450	3056	2	1.12E+08	ENST00000453935.1	-951	NA	LOC112268260
Promoter (<=1kb)	1	494475	495368	894	2	1E+08	ENST00000642074.1	0	NA	LOC100132287
Promoter (<=1kb)	1	586071	612813	26743	2	1.05E+08	ENST00000634833.2	896	NA	LOC105378947
Promoter (<=1kb)	1	627377	631150	3774	2	1.02E+08	ENST00000452176.2	861	ENSG00000230021	LOC101928626
Promoter (<=1kb)	1	627377	631150	3774	2	1.02E+08	ENST00000452176.2	681	ENSG00000230021	LOC101928626
Promoter (<=1kb)	1	627377	631150	3774	2	1.02E+08	ENST00000452176.2	183	ENSG00000230021	LOC101928626
Promoter (<=1kb)	1	627377	631150	3774	2	1.02E+08	ENST00000452176.2	0	ENSG00000230021	LOC101928626
Promoter (<=1kb)	1	594198	631204	37007	2	1.05E+08	ENST00000641296.1	-13	NA	LOC105378947
Promoter (<=1kb)	1	594198	631204	37007	2	1.05E+08	ENST00000641296.1	-770	NA	LOC105378947
Promoter (<=1kb)	1	594191	633129	38939	2	1.05E+08	ENST00000440196.3	836	NA	LOC105378947

Figure 3. Snapshot of annotated file containing all the information about TFBS and target genes.

3.1.5 Functional Analysis

1. Genomic Feature distribution of peaks

The analysis revealed the comprehensive distribution pattern of transcription factor binding sites (TFBS) across various genomic locations relative to the transcription start site of the corresponding genes. Notably, significant difference was observed in the peaks within the promoter region of both p53 and mutp53, indicating that mutation in the p53 promoter increases the affinity for many promoters. Therefore, the mutation in p53 not only leads to the loss of its tumor suppressor function but also confers a gain of function, thereby imparting additional activities that may play a crucial role in cancer progression.



p53 Mut p53

Promoter = 32.62% Promoter = 39%

Figure 4: The overall distribution of genomic elements in p53 and mutp53 samples.

2. Identifying unique genes regulated by p53 and mut p53

Once the distribution of different genomic elements was known, peaks lying in the proximity to the promoter region within the range of 3kb upstream and 1 kb downstream were taken for further analysis to identify target genes uniquely regulated by p53 and mut p53. We held an assumption that if the distance between the peak and TSS lies within a defined range, then only genes were to be considered as target genes.

Out of the total 11,531 genes identified, 8926 genes were common to both the proteins, whereas 3677 genes were uniquely regulated by p53 and 7854 genes were uniquely regulated by mut p53. This again validated our hypothesis that the mutation in p53 not only leads to the loss of its tumor suppressor function but also confers a gain of function, thereby imparting additional activities that may play a crucial role in cancer progression.

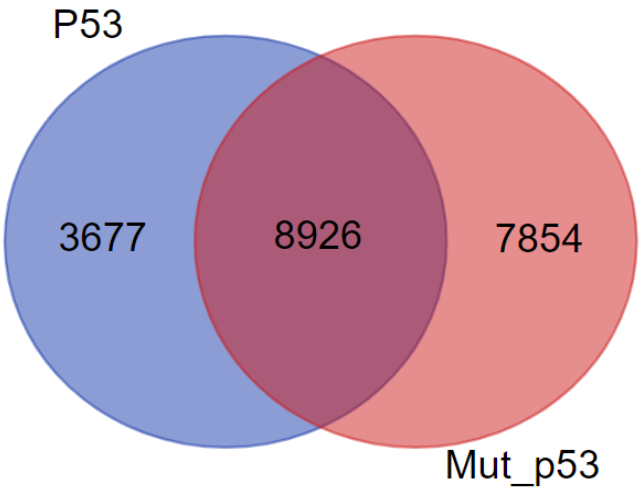


Figure 5: Venn diagram showing proportion of target genes common and unique to both p53 and mut p53

Common	P53	MutP53
ELMO2	MMP2	RPS11
CREB3L1	LOC653712	LOC101930053
PNMA1	SALL1	TMEM216
C10orf90	MIR6500	TRAF3IP2-AS1
ERCC5	LOC105371055	ZHX3
RXFP3	GP1BA	DECR1
APBB2	XAB2	RPS18
H2AZ2	TBX10	BRIX1
NAPRT	LOC112268039	CHD8
PDCL3	LINC00683	LOC101928894
AEN	MIR3156-3	MMP7
CADM4	LINC01685	MIR618
LOC105377763	PRKAG2-AS1	RTN1
SLC10A7	GRM2	ADAM21P1
PXYLP1	GARIN4	KBTBD8
SUMO1	LOC105374945	ZEB1
UQCR11	LOC105375115	FAT4
DDB1	LOC112268276	BANK1
MYO9B	SLC22A2	SFTA2
NIFK-AS1	BTNL10P	SPIN2B
RNVU1-14	SLC25A46	EVC
ZBTB12	MT1JP	CHD1
ZAR1L	LOC401357	SYN3
DTNBP1	RNF223	TMEM274P

Figure 6: Snapshot showing list of target genes common and unique to both p53 and mut p53

3. Gene Ontology

To uncover significant biological pathways linked to two distinct sets of genes regulated exclusively by p53 and mut p53, we conducted functional enrichment analysis of the gene sets using Enrichr. Our objective was to determine whether there are any difference in the pathways regulated by mut p53 that may provide added advantages for cancer cell survival and cancer progression.

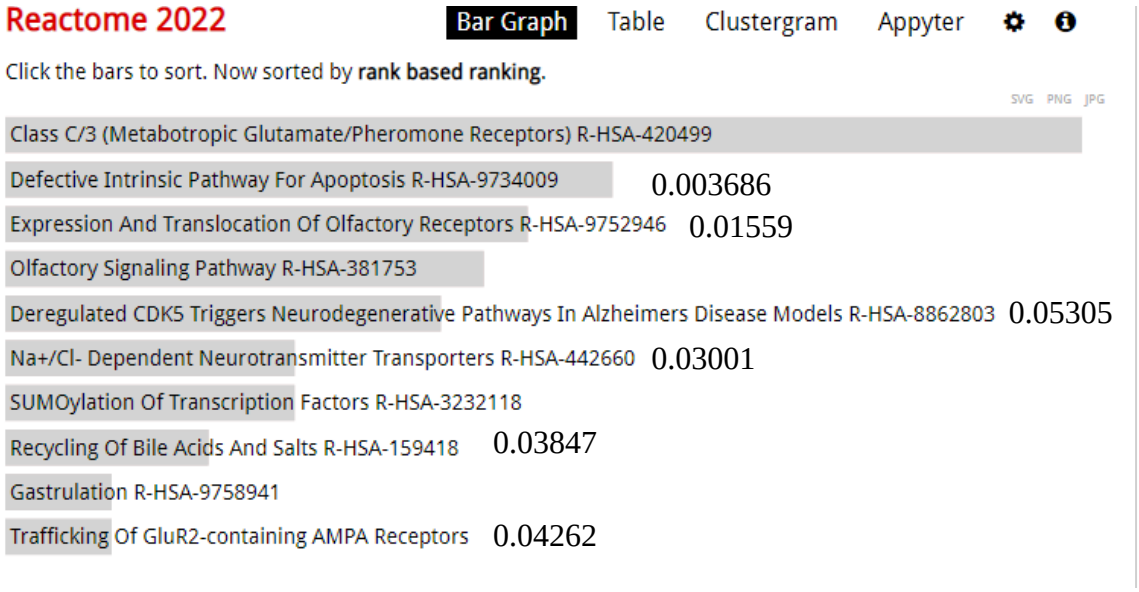


Figure 7: Top 10 pathways regulated by mutp53

Biological Pathways :

Table 6: Important opposing pathways regulated by p53 and mutp53

P53	**MUT P53**
Sodium coupled sulphate di and tri carboxylate transporters	Sodium coupled phosphate co transporters
Neuroactive liganad receptor interaction	Neuroregulin recpetor degradation protein 1 control ErBb3 receptor cycling
Mitotic G2-G2M Phases	DNA Replication initiation
Host-pathogen interaction of human coronaviruses - MAPK signaling	ER stress response in coronavirus infection
Nitrogen and Phenylalanine metabolism	Propanoate and Riboflavin Metabolism
Cortisol synthesis and secretion	CoA biosynthesis
Protein digestion and absorption	Vitamin digestion and absorption
Fibroblast Growth Factor Receptor (FGFR) signaling	epidermal growth factor receptor (EGFR) signalling
Genes **up**-regulated by **KRAS** activation	Genes down-regulated by **KRAS** activation
Myogenesis	Angiogenesis
Heme Metabolism	Fatty Acid Metabolism
GPCR/EPO Signalling pathway	RAS Signalling Pathway
TGF Beta Signalling Pathway	IL 4/ Insulin Signalling Pathway
pyrimidine deoxyribonucleotides de novo biosynthesis	purine nucleotides de novo biosynthesis
Gluconeogenesis	Glycogenolysis
IL27-mediated signaling	IL12 signaling by STAT
Aspartate and Asparagine Biosynthesis	Serine and Glycine Biosynthesis
Hypoxia Response	Oxidative Stress Response
Dopamine receptor mediated signaling	EGF receptor signaling
MAP2K5/MAP2K1	MAP3K12/MAP3K7

CAMK2B	CAMK1
JAK2 activemutant 178	JAK1 activemutant 61
CDK19/7 knockdown	CDK8 knockdown
IGF1R druginhibition	EGFR druginhibition
SNARE complex (VAMP3, VAMP4, STX16)	RC complex (Replication competent complex)

Biological Processes :

Table 7: Gene ontologies/Biological Processes associated with p53 and mutp53

p53	mut p53	
Dopamine uptake and transport	Dopamine secretion	GO Biological Process 2021
monoamine transmembrane transporter activity	inorganic phosphate transmembrane transporter activity	GO Molecular Function 2021
Double Stranded RNA Binding	Ribosomal Large Subunit Binding	GO Molecular Function 2021
integral component of postsynaptic density membrane	integral component of mitochondrial inner membrane	Go Celuular Component 2021
intrinsic component of postsynaptic density membrane	intrinsic component of mitochondrial inner membrane	Go Celuular Component 2021
Late endosome lumen	Early Endosome Lumen	Go Celuular Component 2021
Cation channel complex	Chloride Channel Complex	Go Celuular Component 2021
Abnormal limb development	Abnormal muscle development	MGI Mammalian Phenotype 4
Integral Component Of Postsynaptic Membrane	Extrinsic Component Of Postsynaptic Membrane	SynGO 2022

Diseases and Drugs

Table 8: Diseases and drugs associated with p53 and mutp53 in different cell types

p53	mut p53	Cell Types
Infantile dystonia-parkinsonism	Adult onset dystonia parkinsonism	Orphanet Augmented 2021
ENS Neurons in intestine	Vascular endothelial cells in Heart	Descartes Cell Types and Tissue 2021
Metanephric cells in kidney	Mesangial cells in Kindney	Descartes Cell Types and Tissue 2021

4. Motif Analysis

De-novo enrichment motif analysis further identified the consensus DNA binding sequences of p53 and mut p53 (Fig. 9)

A. p53

Consensus binding motif: **RRRCWWGYY**; where R=A/G, W= A/T, Y = C/T

nsites= 9559
E= 1.7e-418

nsites= 8595
E = 5.33e-400

nsites = 10841
E = 1.5e-370

Figure 8: MEME predicted motifs for p53 with 3 highest e-values

B. Mut p53

Consensus binding motif : **NRTGASTCAYN**; where N =A/T/G/C, R=A/G , Y=C/T,

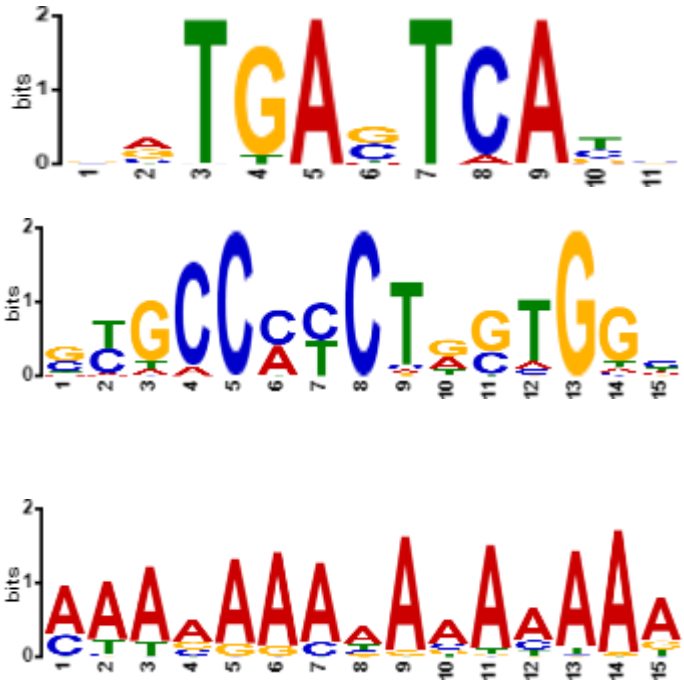


Figure 9: MEME predicted motifs for mut p53 with 3 highest e-values

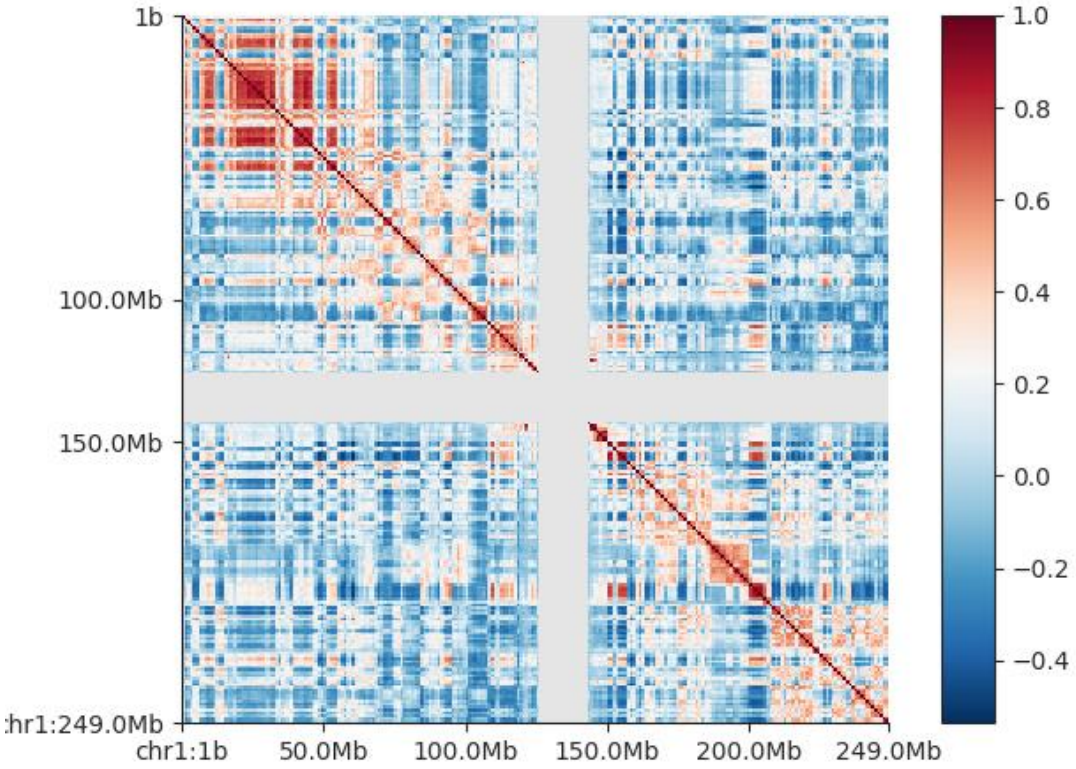
3.2 HiC Data Analysis

3.2.1 Hi-C matrix generation: from raw sequencing output to chromatin contacts

After obtaining the raw reads in FastQ format for cell lines specific to p53 “HCT116” and and mut p53 “FHC”

AB chromatin compartments were created by following the steps described in the previous section (2.2). A and B segregation was further used to quantify the level of compartmentalization in the whole genome. A/B compartment maps of p53 and mut p53 specific cell lines for each chromosome were plotted for visualization.

WT p53: HCT116



Mut p53 : FHC

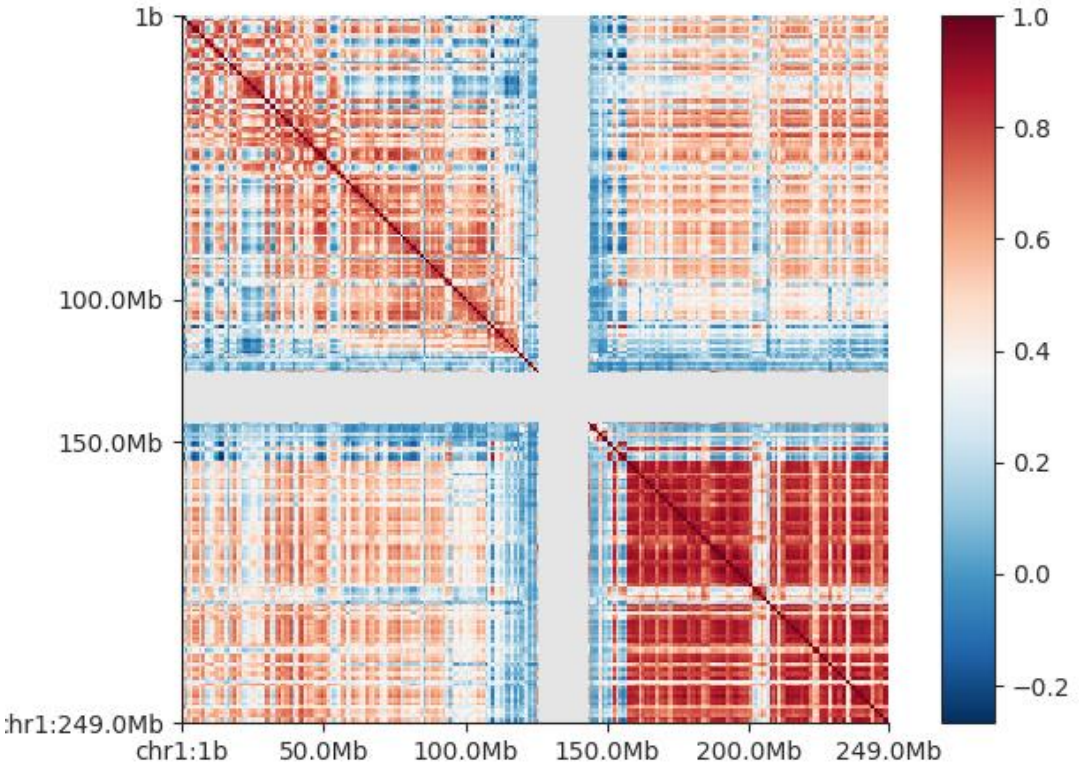


Figure 10: Contact maps of chromosome 1 showing A/B compartmentalization for WT p53 specific cell lines(HCT116)and mut p53 specific cell line (FHC). Positive Score(Red) show A (Active) compartment and Negative scores (Blue) show B (inactive compartment)

Note – Chromatin maps for other chromosomes are added in the appendix.

3.2.2 Functional Analysis

After generating the A/B compartment maps and binning the whole genome into A and B compartments, Subsequently, we identified genes that are regulated by p53, mut p53 or both located within the Active and Inactive regions of the genome, i.e. into A or B compartment. We then proceeded to analyze whether the activation or inactivation of these

genomic regions, resulting from chromatin looping, led to the activation or inactivation of crucial pathways that potentially contribute to cancer progression.

We determined the compartmentalization of 24 extensively researched genes associated with cancer studies, as documented in existing literature.

Table 9: compartmentalization of genes in p53 and mutp53 specific cell lines.

Gene	**P53**	**Mut p53**	**Biological Role**
ATM	A	B	regulators of cell cycle checkpoint signaling pathways that are required for cell response to DNA damage and for genome stability.
ATR	A	B	
MDM2	B	A	The encoded protein can promote tumor formation by targeting tumor suppressor proteins, such as p53, for proteasomal degradation.
BANP (SMAR1)	A	B	This gene encodes a protein that binds to matrix attachment regions. The protein forms a complex with p53 and negatively regulates p53 transcription, and functions as a tumor suppressor and cell cycle regulator
BBC3	A	B	Plays role in apoptotic pathway. This gene encodes a protein that binds to matrix attachment regions. The protein forms a complex with p53 and negatively regulates p53 transcription, and functions as a tumor suppressor and cell cycle regulator
WRN	B	A	The encoded nuclear protein is important in the maintenance of genome stability and plays a role in DNA repair, replication, transcription and telomere maintenance. This gene is not regulated by p53 but the mutant isoform can bind to it and regulate its transcriptional activity.
AEN	A	B	Involved in intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator and response to ionizing radiation.
GADD45A	B	A	The protein encoded by this gene responds to environmental stresses by mediating activation of the p38/JNK pathway via MTK1/MEKK4 kinase. The DNA damage-induced transcription of this gene is mediated by both p53-dependent and -independent mechanisms.
PLKB3	B	A	This gene has also been implicated in stress responses and double-strand break repair. In human

			cell lines, this protein is reported to associate with centrosomes in a microtubule-dependent manner, and during mitosis, the protein becomes localized to the mitotic apparatus.
IGF-1	B	A	The protein encoded by this gene is similar to insulin in function and structure and is a member of a family of proteins involved in mediating growth and development.
PTEN	A	B	It negatively regulates intracellular levels of phosphatidylinositol-3,4,5-trisphosphate in cells and functions as a tumor suppressor by negatively regulating AKT/PKB signaling pathway.
WT1	A	B	It has an essential role in the normal development of the urogenital system, and it is mutated in a small subset of patients with Wilms tumor.
E2F8	A	B	This gene encodes a member of a family of transcription factors which regulate the expression of genes required for progression through the cell cycle. The encoded protein regulates progression from G1 to S phase by ensuring the nucleus divides at the proper time
MYBL2	B	A	MYBL2 is a transcription factor that plays a crucial role in cell cycle control and DNA replication, and its dysregulation has been implicated in various cancers, including breast, lung, and prostate cancer.
TICRR	A	B	TICRR is an essential factor for DNA replication initiation and maintenance of genomic stability, making it critical for proper cell division. Dysregulated TICRR expression has been linked to cancer development and progression, highlighting its potential as a target for therapeutic interventions.
RBP7	B	A	RBP7, a retinol-binding protein, is involved in the transport and metabolism of retinoids.
VEGFB	A	B	The VEGF family members regulate the formation of blood vessels and are involved in endothelial cell physiology.
INSL3	A	B	INSL3, a member of the insulin-like peptide family, has been implicated in testicular descent during embryonic development.
ALX4	A	B	This gene encodes a paired-like homeodomain transcription factor expressed in the mesenchyme of developing bones, limbs, hair, teeth, and mammary tissue.

NOTCH3	A	B	NOTCH3 is a receptor involved in cell signaling and development.
HDAC1	B	A	HDAC1 is a histone deacetylase that regulates gene expression by modifying chromatin structure
RPPH1	B	A	RPPH1 is a long non-coding RNA that has been observed to have altered expression in several cancer types, although its precise function in cancer is not fully understood. Its dysregulation suggests a potential involvement in cancer biology.
NKRF	A	B	This gene encodes a transcriptional repressor that interacts with specific negative regulatory elements to mediate transcriptional repression of certain nuclear factor kappa B responsive genes.
IPO8	B	A	The encoded protein binds to the nuclear pore complex and, along with RanGTP and RANBP1, inhibits the GAP stimulation of the Ran GTPase.

Gene Ontology:

Biological processes:

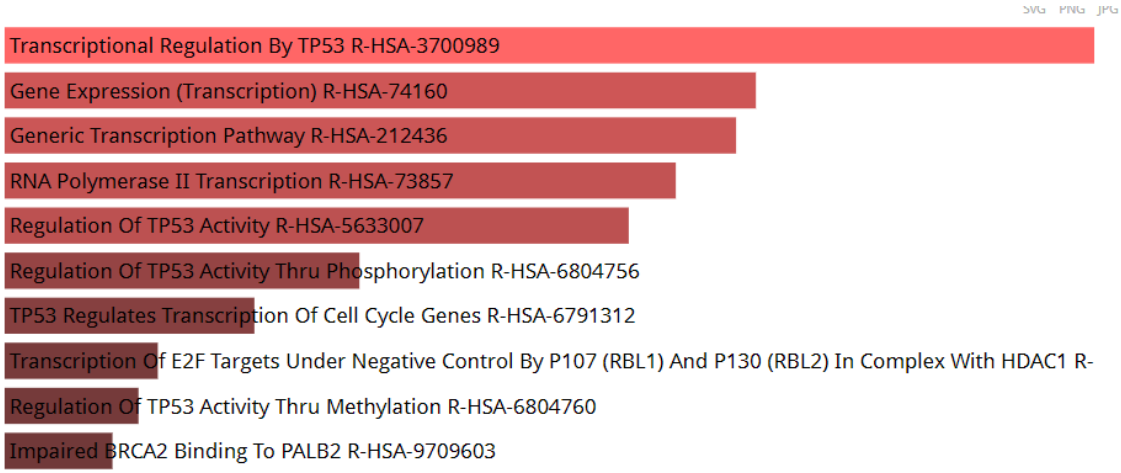


Figure 10: Top 10 pathways associated with these genes

4. Discussion

p53 is a key tumor suppressor, and loss of p53 function is frequently a prerequisite for cancer development. The p53 gene is the most frequently mutated gene in human cancers; p53 mutations occur in >50% of all human cancers and in almost every type of human cancers. In addition to the loss of its original tumor-suppressive role, mutated p53 proteins often acquire new oncogenic activities, independent of the wild-type p53, thus promoting cancer progression—a phenomenon known as gain-of-function (GOF). The accumulation of mutp53 proteins at high levels within cancer cells is critical for their GOF. Considering the high mutation rate of the p53 gene and the oncogenic activities of mutp53 in cancer, there has been significant interest in developing therapies targeting mutp53.

In our current study, we collected and analyzed ChIP-seq data from p53 and mutant p53-specific cell lines obtained from the GEO database. Our analysis clearly demonstrates that p53 mutations not only suppress its tumor suppressor activities but also confer gain-of-function activities. The consensus binding motif of mutant p53 changes, enabling it to regulate a broader range of target genes that provide cancer cells with survival advantages. We performed functional enrichment analysis of the identified target genes, revealing intriguing results. Both p53 and mutp53 were found to be involved in regulating opposing pathways. For instance, p53 controls gene up-regulation triggered by KRAS activation, gluconeogenesis, whereas mutp53 regulates gene down-regulation upon KRAS inactivation, glycogenolysis etc. Additionally, p53 governs numerous pathways critical for cell cycle regulation and acts as a checkpoint in response to DNA damage, such as regulating the transition from mitotic G2 to M phase and GPCR signaling pathways. Conversely, mutp53 regulates pathways that confer survival advantages to cancer cells, including DNA replication initiation, insulin signaling pathways, and Co-A biosynthesis. These findings validate the notion that p53 mutations bestow additional functions upon the mutant p53 protein.

Recently, a novel hypothesis has emerged, suggesting that mutant p53's ability to bind to the matrix attachment regions (MAR) differentially regulates chromatin structure, leading

to the activation or inactivation of oncogenic or tumor suppressor genes, respectively. However, this area remains largely unexplored. So, next part of our study dealt with understanding how p53 and mutant p53 regulate chromatin structure. We analyzed the HiC data from p53 and mutant p53 control cell lines. Chromatin contact maps were generated for each chromosome. For chr 1 as shown in the figure 10, the chromatin contact maps of p53 and mut p53 shows difference in the A/B compartmentalization as evident from the blue and red region of the maps. Next step was to see if these changes in the compartmentalization patterns has any effect on the expression or suppression of genes that play any part in carcinogenesis. We identified 24 most studied genes in cancer studies, retrieved its location on chromosomes from NCBI and mapped them on the contact maps. Much to our surprise, important genes such as ATM, ATR which are responsible for cell cycle regulation in case of DNA damage were found in the inactive compartment in the mutp53 cell line. Genes such as MDM2 which degrades p53 was found in the active compartment in the mutp53 cell line. Similarly other genes were studied and it was seen that genes that are responsible for maintaining integrity of the genome are inactivated in mutant p53 cell lines, whereas genes responsible for carcinogenesis are activated or found in the transcriptionally active compartments.

5. Conclusion

- In summary, mutation in the p53 protein, confer additional gain-of-function activities to the mutant p53
- We have successfully identified distinct target genes and pathways controlled by both p53 and mutp53, which have the potential to contribute to the control or development of cancer.
- Based on the involvement of genes in specific processes as provided by gene ontology results, we have been able to conclude that p53 regulates processes that are important for cell cycle regulation where as mutant p53 regulates pathways that are important for survival of cancer cells and suppressing pathways responsible for tumor suppression.
- Motif analysis using MEME
    - Consensus motif for p53 analyzed using MEME-Chip: RRRCWWGYY; where R=A/G, W= A/T, Y = C/T
    - Consensus motif for mut p53 analyzed using MEME-Chip: NRTGASTCAYN; where N =A/T/G/C, R=A/G , Y=C/T
- Contact maps generated using HiC data analysis showed differences in A/B compartmentalization of p53 and mut p53 specific cell lines leading to differential activation or inactivation of genes corresponding to these regions. This validated our hypothesis that differential regulation of chromatin structure by p53 and mut p53 have possible roles in carcinogenesis.

6. Future Work

- Analysis of more p53 and mutant p53 specific cell lines to get a consensus compartmentalization pattern for which data is available on GEO.
- Validation through wet lab experiments.

7. References

- Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. Genes Cancer. 2011 Apr;2(4):466-74. doi: 10.1177/1947601911408889. PMID: 21779514; PMCID: PMC3135636.

- Mantovani F, Collavin L, Del Sal G. Mutant p53 as a guardian of the cancer cell. Cell Death Differ. 2019 Jan;26(2):199-212. doi: 10.1038/s41418-018-0246-9. Epub 2018 Dec 11. PMID: 30538286; PMCID: PMC6329812.

- GEO: Edgar R, Domrachev M, Lash AE.
  Gene Expression Omnibus: NCBI gene expression and hybridization array data repository
  Nucleic Acids Res. 2002 Jan 1;30(1):207-10

- https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software and the SRA Toolkit Development Team.

- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

- Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.

- Twelve years of SAMtools and BCFtools
  Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng LiGigaScience, Volume 10, Issue 2, February 2021, giab008, https://doi.org/10.1093/gigascience/giab008

- Zhang, Y., Liu, T., Meyer, C.A. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008). https://doi.org/10.1186/gb-2008-9-9-r137

- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2. doi: 10.1093/bioinformatics/btq033. Epub 2010 Jan 28. PMID: 20110278; PMCID: PMC2832824.

- Human genome assembly: The Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921.
- Heberle, H., Meirelles, G.V., da Silva, F.R. et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 16, 169 (2015). https://doi.org/10.1186/s12859-015-0611-3
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 Apr 15;14:128. doi: 10.1186/1471-2105-14-128. PMID: 23586463; PMCID: PMC3637064.
- Philip Machanick, Timothy L. Bailey, MEME-ChIP: motif analysis of large DNA datasets, Bioinformatics, Volume 27, Issue 12, June 2011, Pages 1696–1697, https://doi.org/10.1093/bioinformatics/btr189
- Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, Chuan Deng, Thawfeek Varusai, Eliot Ragueneau, Yusra Haider, Bruce May, Veronica Shamovsky, Joel Weiser, Timothy Brunson, Nasim Sanati, Liam Beckman, Xiang Shao, Antonio Fabregat, Konstantinos Sidiropoulos, Julieth Murillo, Guilherme Viteri, Justin Cook, Solomon Shorser, Gary Bader, Emek Demir, Chris Sander, Robin Haw, Guanming Wu, Lincoln Stein, Henning Hermjakob, Peter D'Eustachio, The reactome pathway knowledgebase 2022, Nucleic Acids Research, 2021;, gkab1028, https://doi.org/10.1093/nar/gkab1028

URLs

- **GEO: https://www.ncbi.nlm.nih.gov/geo/**
- **ENA: https://www.ebi.ac.uk/ena/browser/home**
- **MEME ChIP: https://meme-suite.org/**
- **Enrichr: https://maayanlab.cloud/Enrichr/**

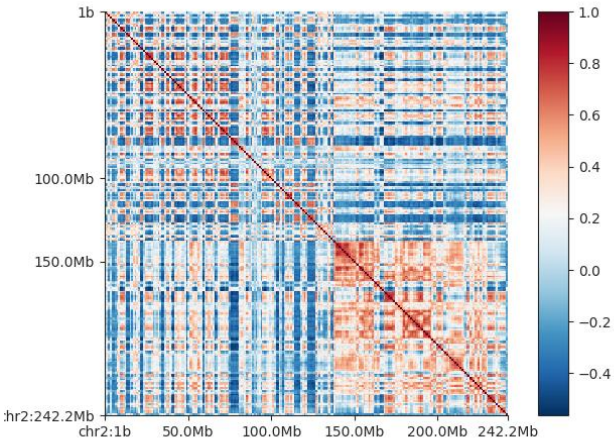
Tools

- fastQC (v0.11.3): https://www.bioinformatics.babraham.ac.uk/
- Cutadapt (v1.12): https://doi.org/10.14806/ej.17.1.200
- Bowtie2 (v2.2.6): Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923. PMID: 22388286; PMCID: PMC3322381.
- Samtools (v1.7): Heng Li and others, The Sequence Alignment/Map format and SAMtools, Bioinformatics, Volume 25, Issue 16, August 2009, Pages 2078–2079, https://doi.org/10.1093/bioinformatics/btp352
- MACS2 (v2.1.1): Zhang, Y., Liu, T., Meyer, C.A. et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008). https://doi.org/10.1186/gb-2008-9-9-r137
- Bedtools (v2.25.0): Aaron R. Quinlan , Ira M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, Volume 26, Issue 6, March 2010, Pages 841–842, https://doi.org/10.1093/bioinformatics/btq033
- FANC(v0.9.26): 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

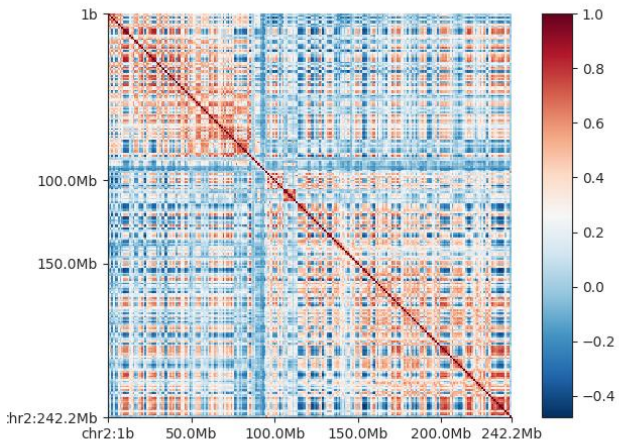
8. Appendix

Contact Maps of different chromosomes showing A/B compartmentalization in p53 and Mut p53 specific cell lines

Chr2 :

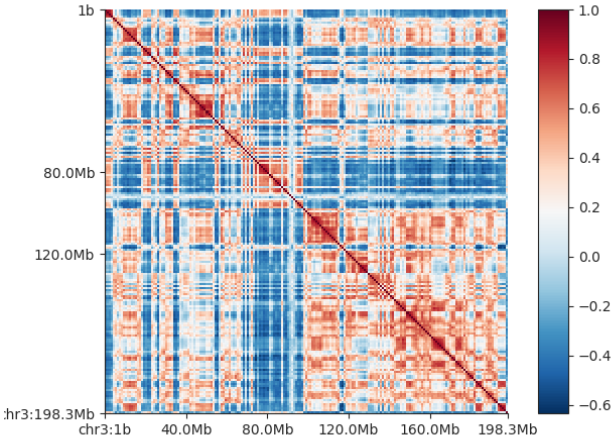


WT p53 : HCT116

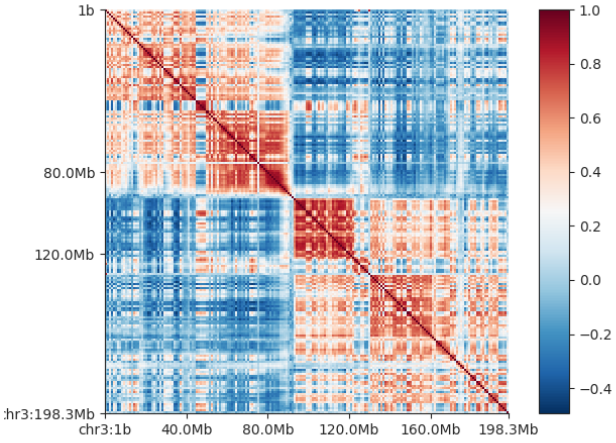


Mut p53 : FHC

Chr3

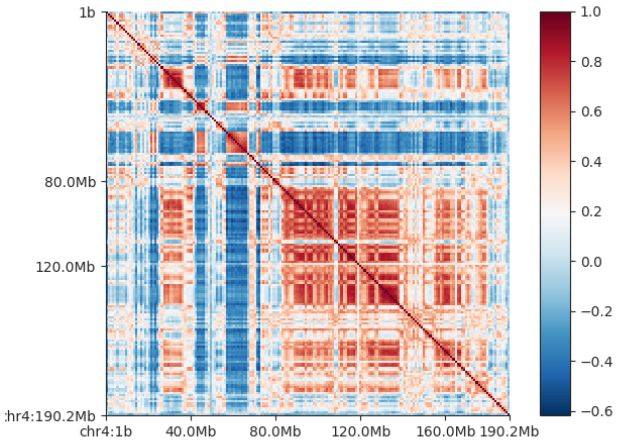


WT p53 : HCT116

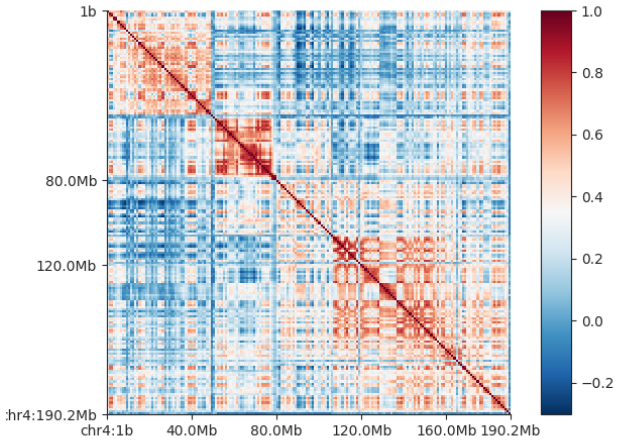


WT p53 : FHC

Chr4

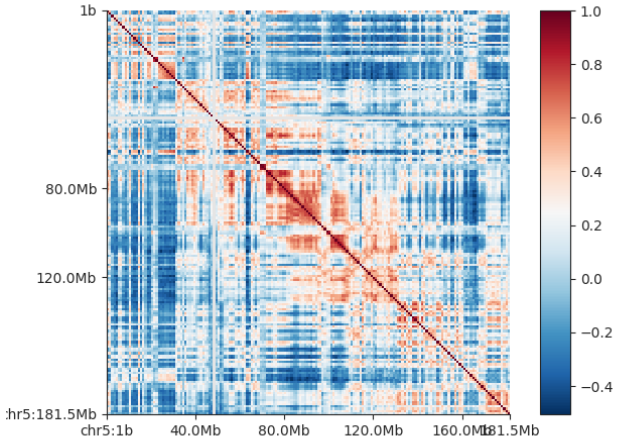


WT p53 : HCT116

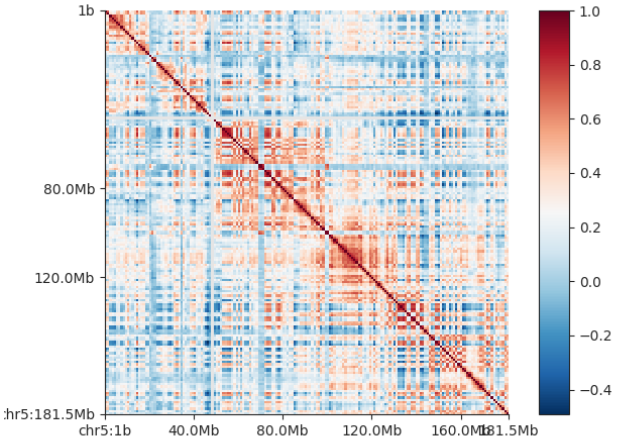


WT p53 : FHC

Chr5

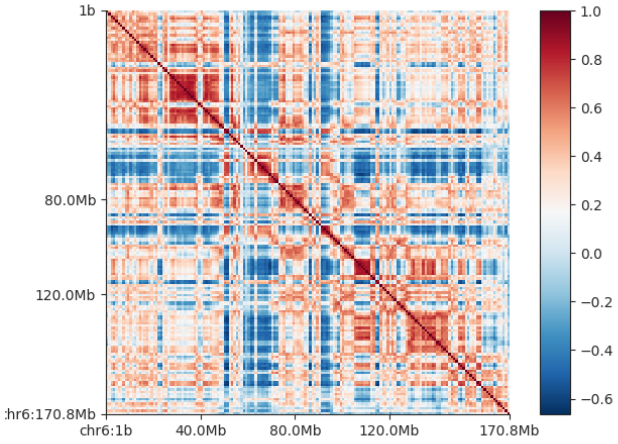


WT p53 : HCT116

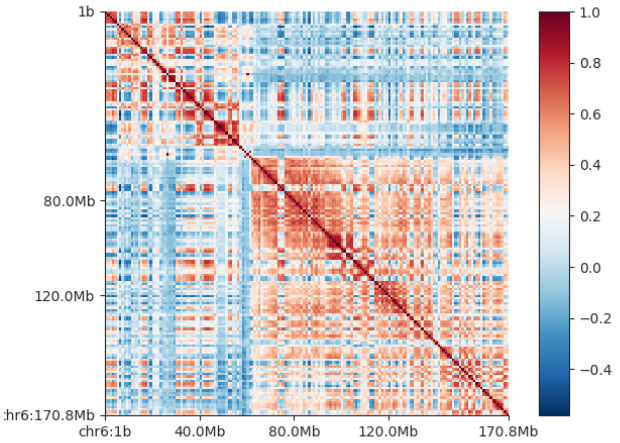


MUT p53 : FHC

Chr 6

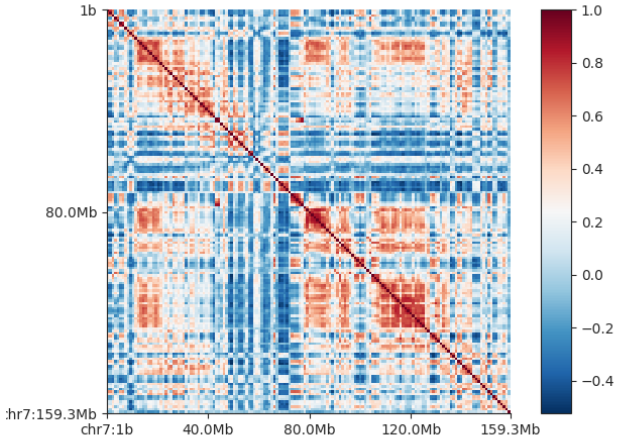


WT p53 : HCT116

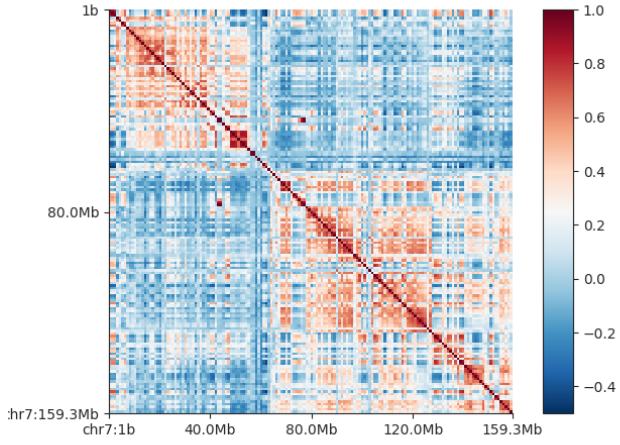


WT p53 : FHC

Chr7:

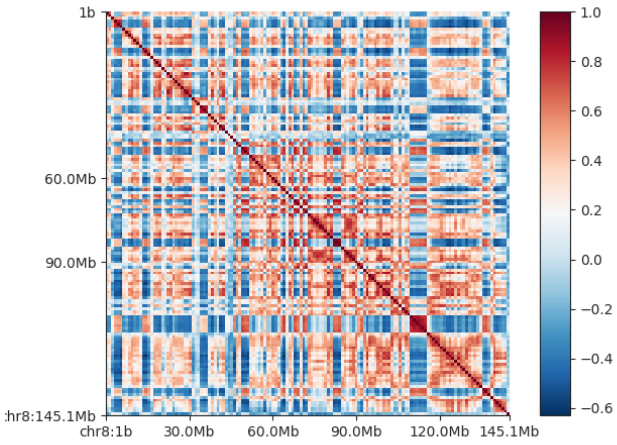


WT p53 : HCT116

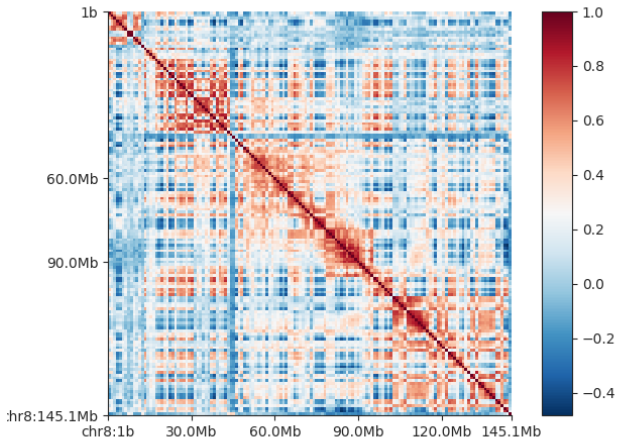


MUT p53 : FHC

Chr8:

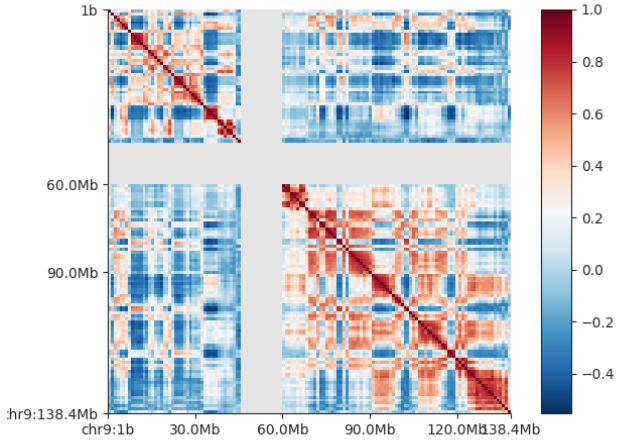


WT p53 : HCT116

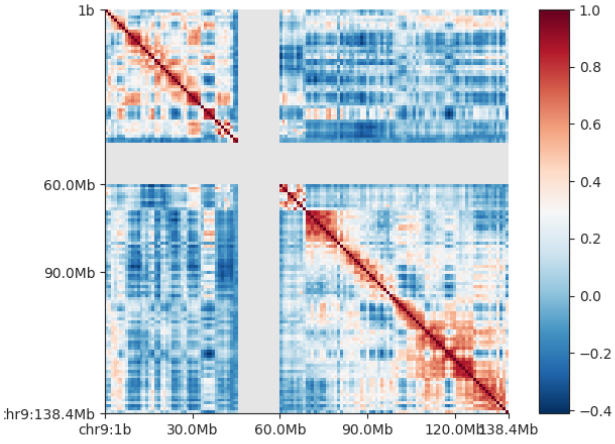


MUT p53 : FHC

Chr9

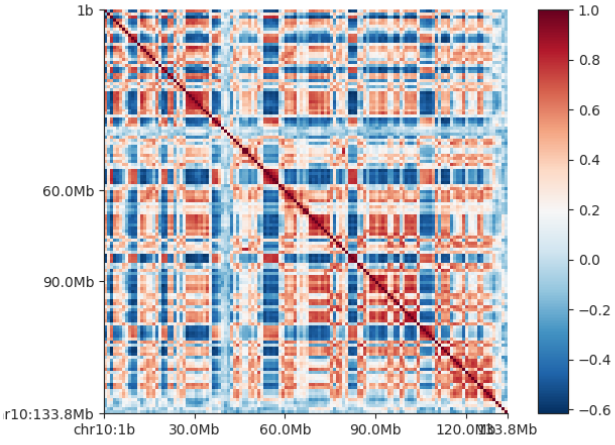


WT p53 : HCT116

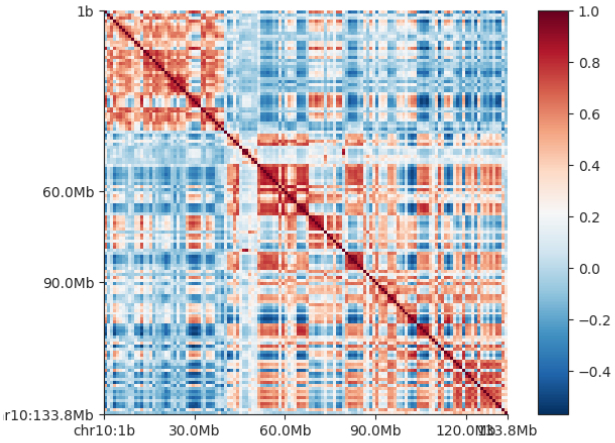


MUT p53 : FHC

Chr10:

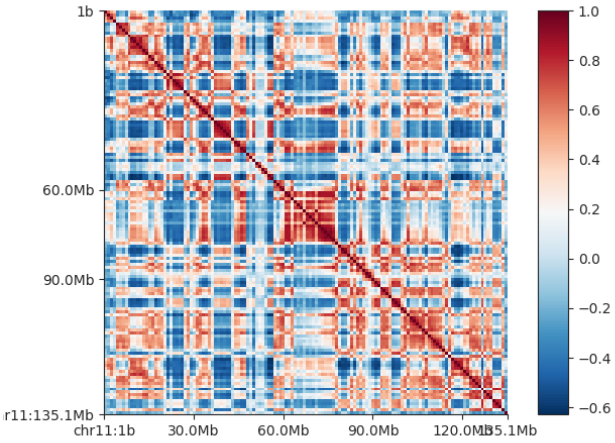


WT p53 : HCT116

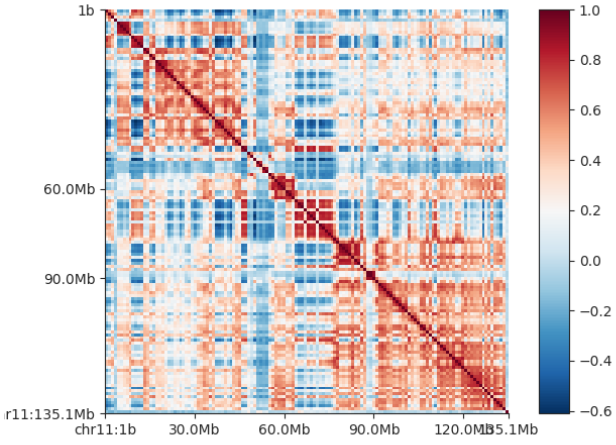


MUT p53 : FHC

Chr11

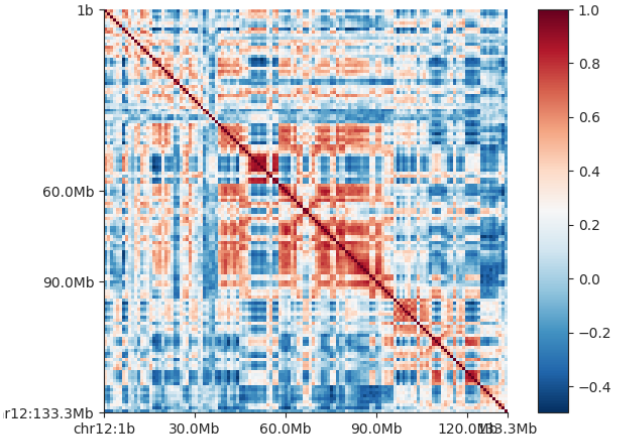


WT p53 : HCT116

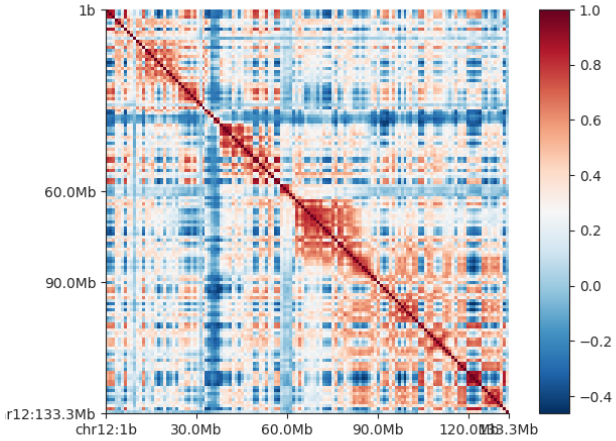


MUT p53 : FHC

Chr12

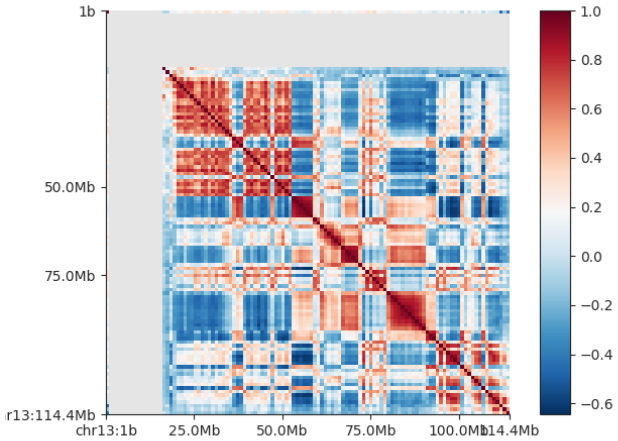


WT p53 : HCT116

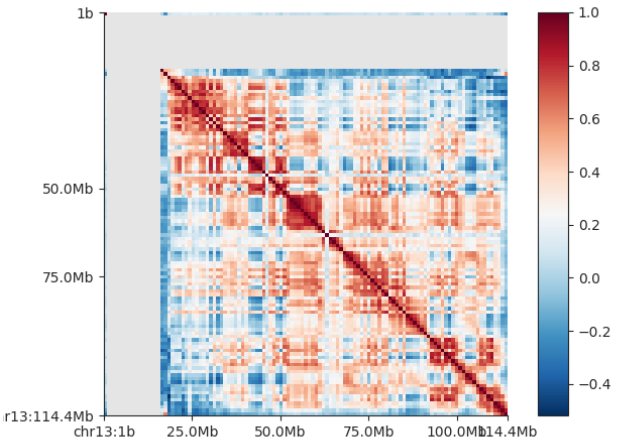


MUT p53 : FHC

Chr13

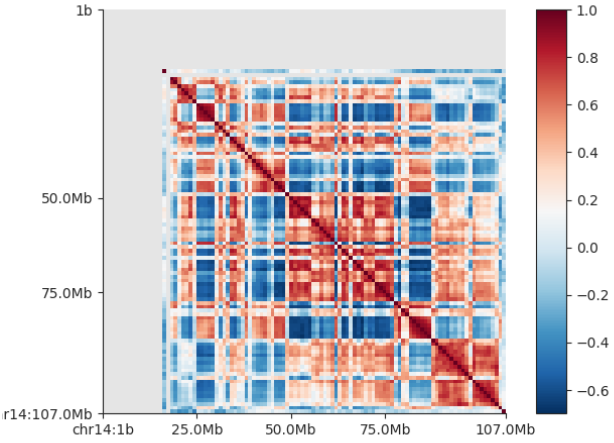


WT p53 : HCT116

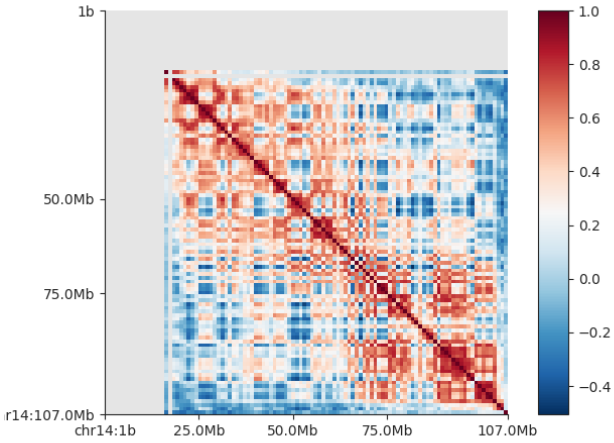


MUT p53 : FHC

Chr14

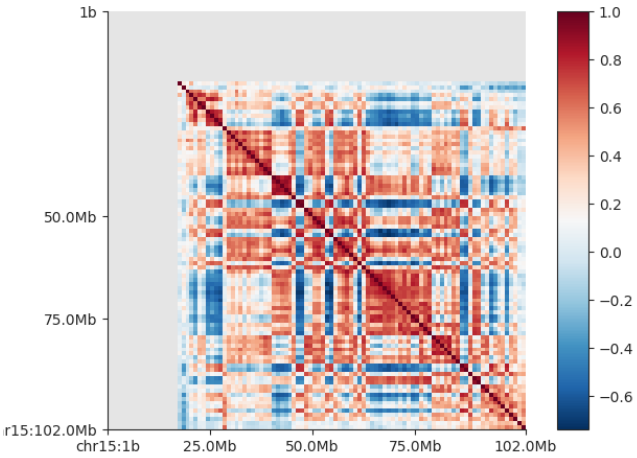


WT p53 : HCT116

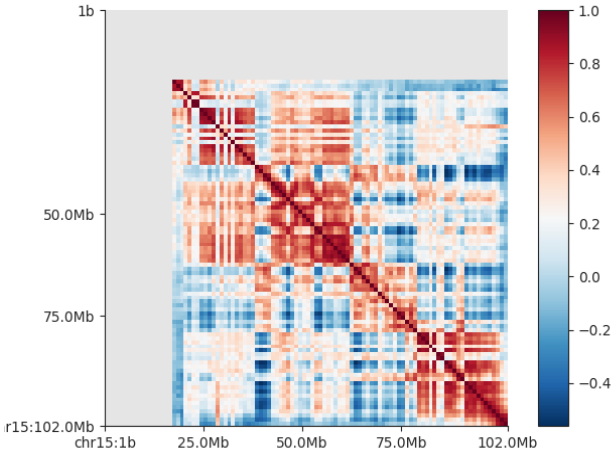


MUT p53 : FHC

Chr15

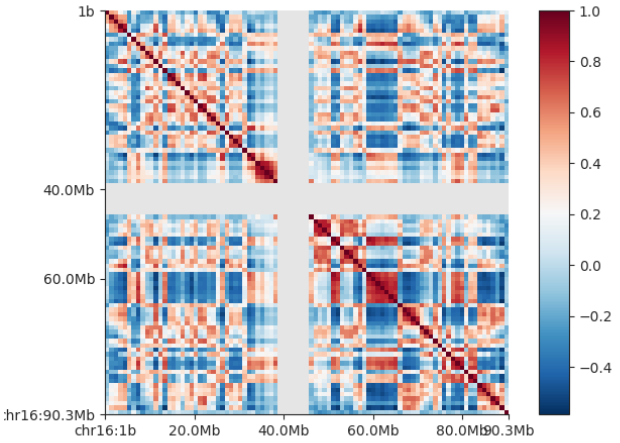


WT p53 : HCT116

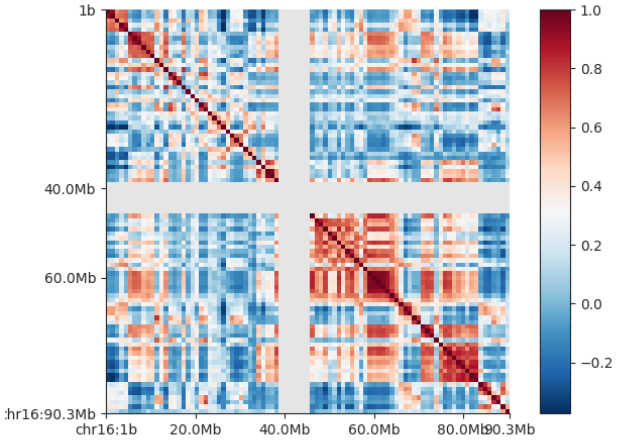


MUT p53 : FHC

Chr16

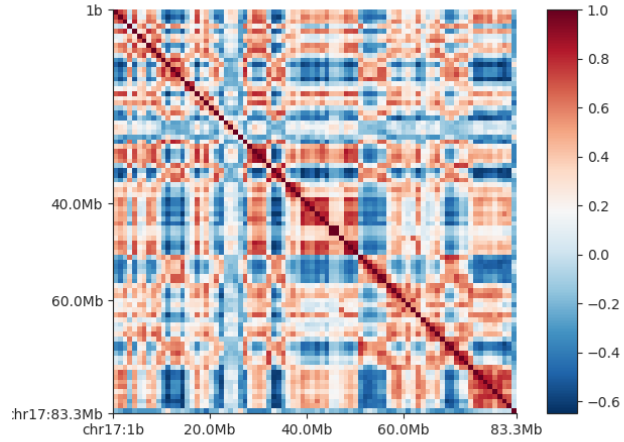


WT p53 : HCT116

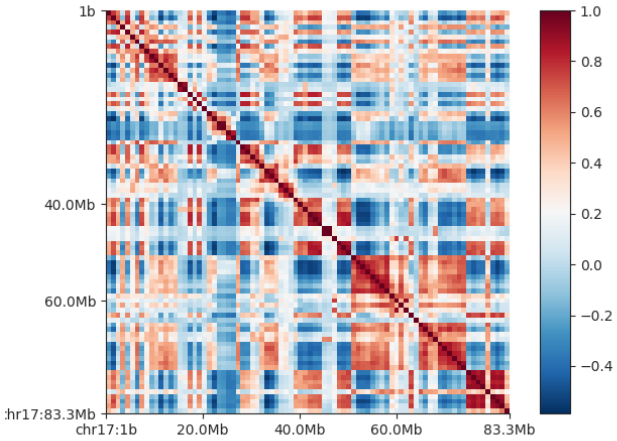


MUT p53 : FHC

Chr17

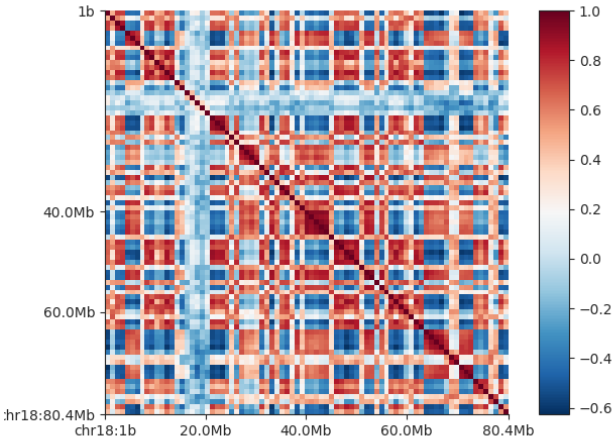


WT p53 : HCT116

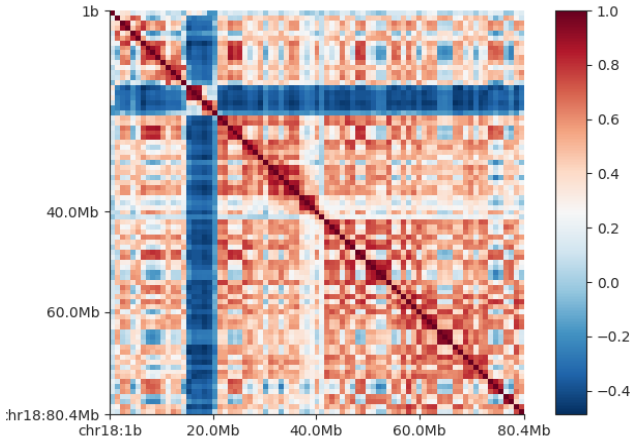


MUT p53 : FHC

Chr18

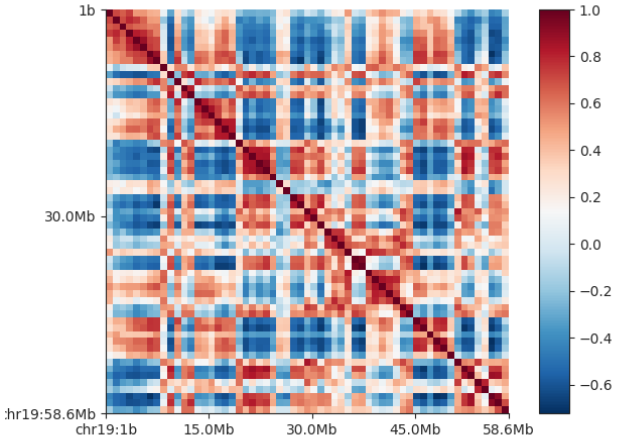


WT p53 : HCT116

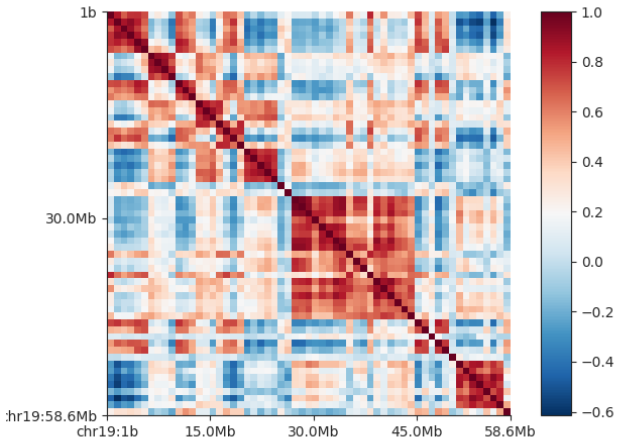


MUT p53 : FHC

Chr19

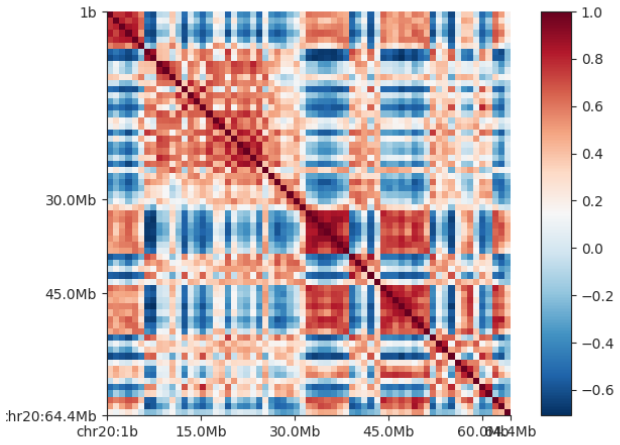


WT p53 : HCT116

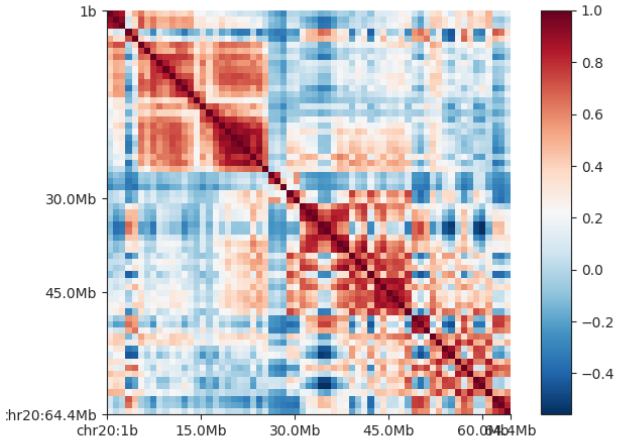


MUT p53 : FHC

Chr 20

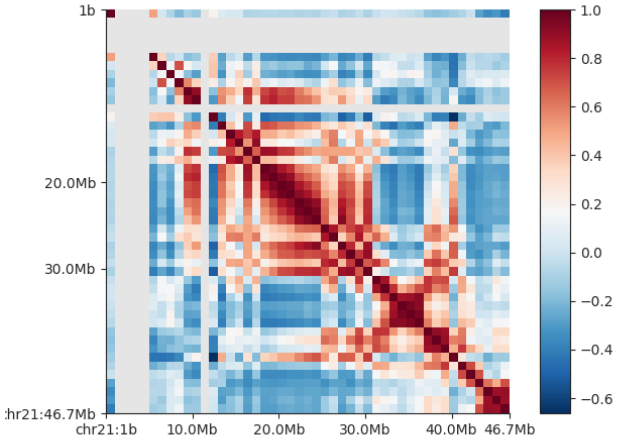


WT p53 : HCT116

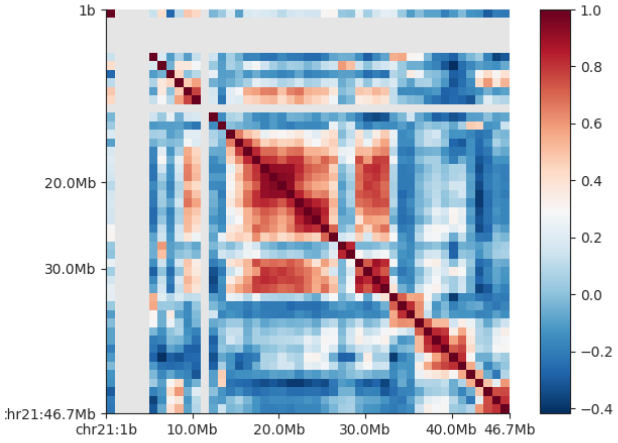


MUT p53 : FHC

Chr21

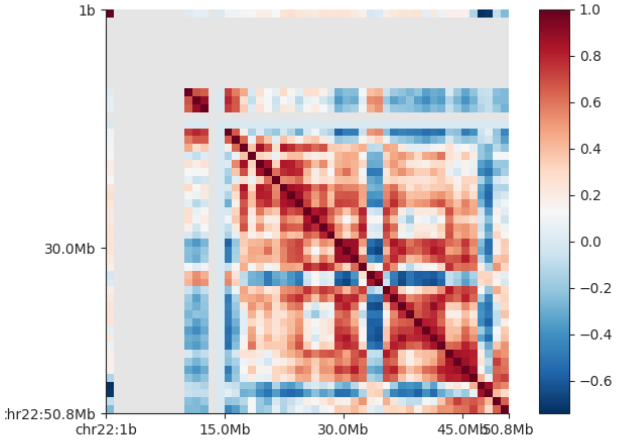


WT p53 : HCT116

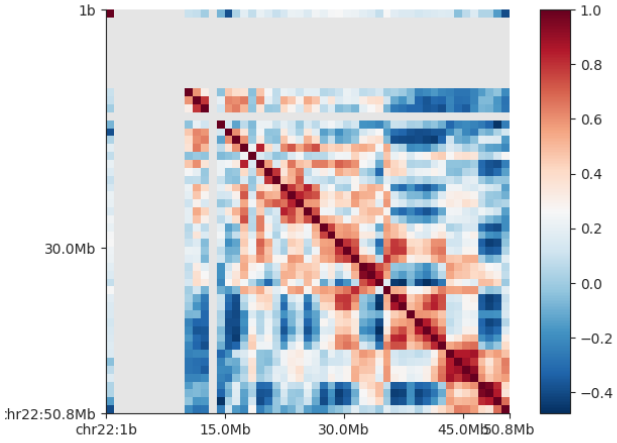


MUT p53 : FHC

Chr22:

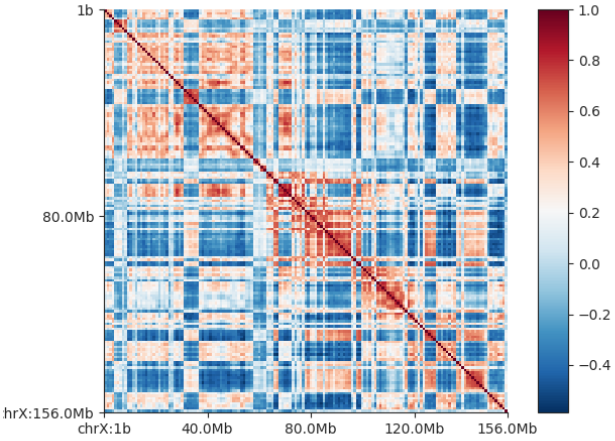


WT p53 : HCT116

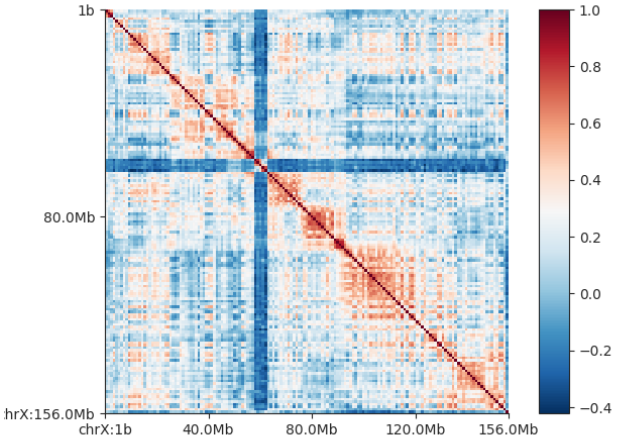


MUT p53 : FHC

ChrX

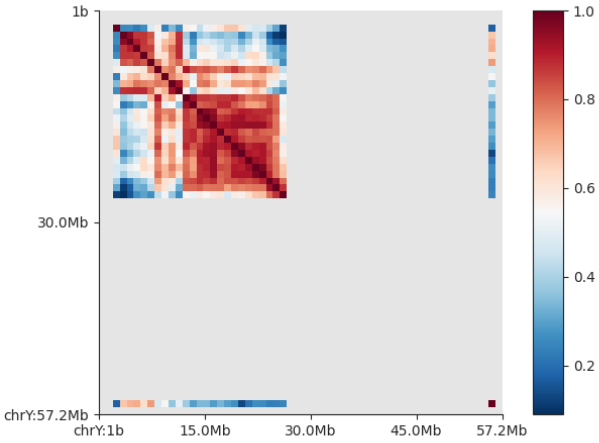


WT p53 : HCT116

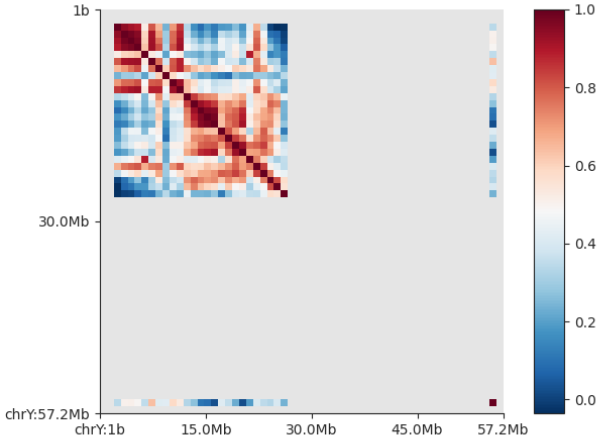


MUT p53 : FHC

ChrY



WT p53 : HCT116



MUT p53 : FHC