UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Doctoral Program in Bioinformatics

School of Mathematics and Statistics

Exploring the impact of p53 activation on spatio-temporal genome topology

submitted by

Mónica Cabrera Pasadas

for the academic degree of

Doctor of Philosophy (PhD) in Bioinformatics

developed under the supervision and authorize to deposit by

Prof. Alfonso Valencia
Director

Dr. Biola M. Javierre
Co-director

Dr. Alexandre Perera
University tutor

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Institut de Recerca CONTRA LA LEUCÈMIA Josep Carreras

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Barcelona, October 2023

I was just guessing at numbers and figures
Pulling the puzzles apart
Questions of science, science and progress
Do not speak as loud as my heart
But tell me you love me, come back and haunt me
Oh and I rush to the start

Nobody said it was easy
Oh, it's such a shame for us to part
Nobody said it was easy
No one ever said it would be so hard
I'm going back to the start

The Scientist
Coldplay

Acknowledgments

I used to consider this section somewhat clichéd and believed there was no necessity to delve into it extensively. However, as my journey reaches its end, I have come to the realization that despite this work is carrying my name, numerous individuals have played indispensable roles in this thesis—individuals I wish to express my gratitude and remember forever. To do so, I have chosen to adopt a 3D organizational metaphor to symbolize the multifaceted nature of their contributions. Enjoy your character!

On the largest scale, I would like to express my appreciation to the A/B compartments of my PhD journey, Alfonso Valencia and Biola Javierre. These two compartments, situated in distinct cities and characterized by distinct working methodologies, have allowed me the privilege of witnessing two unique perspectives on science, leadership, and life itself. This duality has enriched my understanding in ways I will forever remember. I am truly fortunate to have had the opportunity to use and learn coding on the supercomputer Mare Nostrum, but also witness the passion and dedication exhibited day after day in every experiment in the wet lab.

One of the criteria I was looking before starting a PhD journey was the opportunity to have a young female scientist as a role model, and I am happy I have been under Biola's supervision. There were occasions in the laboratory when witnessing your arrival with boundless energy left me drained, but also served as a source of profound motivation. It was during these moments that I learned a valuable life lesson: the indispensable role of determination in the pursuit of one's aspirations. This wisdom, instilled by your example, is something I will forever cherish. Thank you Biola.

Despite the challenging circumstances imposed by COVID-19 not allowing me to spend more time at the A compartment, I am deeply grateful for the opportunities and help that Alfonso provided. Conferences and computational group meetings have been instrumental in allowing me to feel more in my ‘salsa’ as a bioinformatician. My admiration for you has been since I first learned about your significant contributions to the field of bioinformatics during my time in Madrid and the intersection of our paths in Barcelona has had a profound impact on my journey. I am genuinely delighted to acknowledge that this PhD opportunity has played a pivotal role in shaping the person I have become today. Thanks Alfonso.

Zooming into the B compartment, I am immensely grateful to have encountered key regulators who enhanced both my life and this thesis. Without a doubt, I crossed paths with a pioneer transcription factor like p53, Laureano Tomás-Daza. Laure, you can join whatever compartment and augment everyone's life, orchestrating a cascade of positive energy. You have been a constant since day one of my PhD journey; the car rides to Can Ruti, our initial solitary meals, the COVID-19 challenges, meetings preparations, roasting sessions, las fiestas del pueblo, pawrenthood, and countless other moments. All of them hold a special place in my heart. Thanks Laure.

Then, super-enhancers who brightened my daily existence came slowly steady to the B compartment: Llorenç, Andrea, Paula, Blanca U, Blanca V, Lucia, Ainoa, Raul, Xavi, Alvaro. You have been able to create a pleasant environment making the everyday life in the Can Ruti mountain a comfortable space to work making the lab a warm place even at 18 degrees ;)

On the other side, the elements of the A compartment are constantly in dynamism. Nevertheless, I interacted with people that has been a privilege to witness their growth during these years and some of them went through the same time-course as me making me feel less alone at the BSC building—Jon, Vicky, Iker, Davide, Bea, Alba, and many more, thanks. Also, thanks to Gonzalo, who came just in time to make me have one of my most happy experiences during my PhD, be the ESCS chair with many RSGs help: Ina, Adrian, Mirko, Carla, Irene and many more showing me the lesson: Fake it until you make it.

Between the two compartments, I found a TAD border, Fransua, that because of COVID-19 he has been appearing and disappearing in my screen countless times. It hasn’t been easy, but the project we talked about in 2018 is coming to an end thanks to your help.

However, much like in cancer, it is imperative to recognize that a PhD success is not solely about the internal elements of the lab but also the microenvironment, both the one we are born in and the one we consciously choose with our actions.

En cuanto al microambiente en el que he nacido, me faltan las palabras para expresar lo que significan para mí mis padres Araceli y Javier. Ellos dos son literalmente la base de quien soy, quienes, a pesar de no comprender por completo mi carrera o lo que hago, siempre me han brindado un apoyo incondicional. Habéis sido de vital importancia para que, a pesar de los desafíos que se han presentado a nivel más básico, lo malo no haya logrado avanzar. Mi idea era siempre investigar para curar el cáncer y gracias a vuestro apoyo lo he podido intentar.

Luego está mi hermanita Araceli, quien siempre será la pequeñita, incluso después de haberme dado una sobrinita. Tus regalos de temática friki o científica siempre me han ayudado a sentirme mejor. Estoy inmensamente orgullosa de que seas mi hermana y de la mujer en la que te has convertido. Siempre compartiremos una parte significativa de nuestro ADN y, en última instancia, somos más parecidas de lo que a veces habríamos querido. Gala, la princesita de la casa, eres tan pequeña y, a la vez, tan preciosa y despierta que solo puedo desearte lo mejor en la vida. Estaré siempre a tu lado para presenciar tus logros y ayudarte en las asignaturas de ciencia.

Aunque recientemente se les haya reconocido legalmente como miembros de la familia, para mí ya lo era mucho antes Blanca, mi perrita, quien ha estado conmigo desde el primer día como científica empezando la carrera de Biología Humana. Fuiste testigo de cómo fui cumpliendo mis sueños, aunque lamentablemente no pudiste llegar a ver el final de esta etapa. Supongo que no podías quedarte para siempre, pero siempre recordaré haberme confiado 3 meses a tu lado y poder disfrutarte 24/7.

Finalmente, quiero expresar mi gratitud a todos los demás miembros de mi familia: abuelos, tíos y primos que después de tantos años ya no preguntan qué hago, sino que se alegran si salgo en la tele de rebote o hago una conferencia.

Respecto el microambiente que he elegido, Maanas Mediratta, mera bachito, cuando te conocí recuerdo que no paraba de hablar de mi siguiente paso haciendo un doctorado y estuviste conmigo cuando por fin lo empecé. Desde el inicio de esa tesis has estado dándome tu apoyo, me has animado a crecer, a no rendirme y no dejarlo, a manejar situaciones elegantemente, te has alegrado por cada mini logro y motivado en ayudarme en mis proyectos aun sin entender que es un gen. Gracias a ti he podido desconectar de la tesis en lugares increíbles que una estudiante precaria de doctorado no podría imaginarse, aunque la pandemia nos hizo bajar esa vida loca ayudándonos a convertirnos en adultos de verdad. Al final, poca gente va a poder

decir que ha sobrevivido a una relación a distancia, una pandemia, a un cachorrito hiperactivo, la compra de una vivienda, la mudanza y un doctorado. Espero seguir creciendo contigo en nuestro nuevo capítulo como padres primerizos de Baby M.
Mani, mani manito pesadito bonito, pensábamos que tener un cachorrito iba a ser tener un peluchito al que acariciar mientras teletrabajábamos, pero me obligaste a salir a la calle a todas horas y socializar en el pipican. En muchos momentos me superó, pero ahora reconozco que en realidad llegaste justo en el mejor momento para ayudarme a no encerrarme ni física ni mentalmente.
Milan, llegaste justo al comienzo de la redacción de esta tesis y estarás al caer cuando se acabe. Supongo que desde el principio de tu corta vida te he enseñado a lidiar con las hormonas del estrés. Me llenas de alegría todos los días desde que me levanto sabiendo que estás conmigo y ahora aún más notando tus pataditas. Me muero de ganas por vivir todo lo que nos espera juntos mi pequeño gran regalo de Navidad :)

Tampoco sé qué habría hecho sin tener una red de amigas que me han permitido ver la vida desde perspectivas diversas. Neus, Rocío, Izabella, Marina, Anna Llopart, Salma… sois únicas y estoy muy contenta de que forméis parte de mi día a día. Me habéis motivado a buscar otras vías para seguir creciendo y sobre todo escoger algo que me haga feliz. Sois las personas a las que puedo enviar infinitos audios de updates y sentir como que no pasa el tiempo.
David Montero x2, tú también has estado literalmente desde el primer día de esta tesis, aquel día en el que fuimos a Mordor con Laure en tu mini coche gris. Hay personas que son la pareja de X, pero tú te has hecho un huequito como amigo, aunque sin mantilla bordada, lo siento. Gracias por tu sinceridad en todo momento.
Miguel Madrid, apenas he podido compartir este camino del doctorado en persona contigo porque te marchaste a Toulouse cuando lo empecé, sin embargo, si pudimos compartir el camino previo a ello: un verano en el CNIO patinando por el retiro y nuestro inicio a la vez en el BSC. Quizás eres el culpable de que acabase haciendo el PhD ya que gracias a ti pude conocer a Alfonso y su grupo, tanto hablarme de proteínas me gustó.

Me alegra que todos los nombrados hayáis podido formar parte de este camino. Espero poder seguir contando con todos vosotros a mi lado por muchos más años.

Abstract

The tumor suppressor protein p53 plays a pivotal role in cellular responses to stress. Under non genotoxic or oncogenic stressed conditions, p53 cellular levels are kept low by the MDM2 protein that promotes ubiquitination and proteasome-mediated degradation. However, upon encountering cellular stress, the p53-MDM2 interaction is disrupted leading to p53 accumulation and activation. Once active, p53 promotes critical cell fate decisions such as cell cycle arrest, apoptosis, or senescence by transactivating an array of target genes, leading to effective tumor regression. Consequently, dysregulation of the p53 pathway is strongly associated with cancer susceptibility. Approximately, half of all human cancers carry mutations or deletions of the p53 gene, while the other half have disruptions to the p53 signaling pathway. Nevertheless, despite decades of extensive research, the precise molecular mechanisms governing p53-mediated gene regulation and its tumor-suppressive efficacy remain incompletely understood.

Gene regulation is associated with the three-dimensional (3D) packing of the chromatin within the nucleus. Chromatin loops that involve cis-regulatory elements, such as gene promoters and distal regulatory elements (e.g., enhancers or silencers), are key to control gene transcription. However, the mechanisms by which p53 engages the genome in the context of 3D chromatin to activate transcription are unexplored.

This doctoral thesis investigates p53's ability to reconfigure the genome and regulate gene transcription by inducing alterations in the temporal and spatial genome structure. It is the hypothesis of this thesis that p53 exerts its influence by modulating physical connections between cis-regulatory elements via the manipulation of chromatin loops.

To assess this hypothesis, a comprehensive investigation into the potential effects of p53 activation in HCT116 cell lines achieved through pharmacologically inhibiting MDM2 with Nutlin-3a drug has been performed. This approach synergistically

integrates diverse genomic datasets, encompassing Hi-C, Promoter Capture Hi-C, RNA-seq, and ChIP-seq methodologies.

The results of this thesis demonstrate that p53 induces rapid changes in the different hierarchical levels of the genome, impacting the p53 transcriptional program.

By deciphering the nexus between 3D genome organization and p53-driven transcriptional regulation, this thesis aspires to help in the unravel avenues for therapeutic intervention in the realm of cancer treatment.

Key words: 3D chromatin organization, p53, NGS, Genomics, Bioinformatics

Resumen

La proteína supresora de tumores p53 desempeña un papel fundamental en las respuestas celulares al estrés. Bajo condiciones de estrés no genotóxico u oncogénico, los niveles celulares de p53 se mantienen bajos gracias a la proteína MDM2, que promueve la ubiquitinación y la degradación mediada por el proteasoma. Sin embargo, al haber estrés celular, la interacción p53-MDM2 se interrumpe, lo que lleva a la acumulación y activación de p53. Una vez activa, p53 promueve decisiones críticas en el destino celular, como la detención del ciclo celular, la apoptosis o la senescencia, al activar una serie de genes que conducen a una efectiva regresión tumoral. En consecuencia, la desregulación de la vía de p53 está fuertemente asociada con la susceptibilidad al cáncer. Aproximadamente, la mitad de todos los cánceres humanos son debidos a mutaciones o deleciones del gen p53, mientras que la otra mitad tiene alteraciones en la vía de señalización de p53. No obstante, a pesar de décadas de investigación extensa, los mecanismos moleculares precisos que regulan la expresión génica mediada por p53 y su eficacia en la supresión tumoral aún no se comprenden completamente.

La regulación génica está asociada con el empaquetamiento tridimensional (3D) de la cromatina dentro del núcleo. Los bucles o lazos de cromatina que involucran elementos cis-reguladores, como promotores génicos y elementos reguladores distales (por ejemplo, potenciadores o silenciadores), son clave para controlar la transcripción génica. Sin embargo, los mecanismos por los cuales p53 interactúa con el genoma en el contexto de la cromatina 3D para activar la transcripción aún no se han explorado.

Esta tesis doctoral investiga la capacidad de p53 para reconfigurar el genoma y regular la transcripción génica al inducir alteraciones temporales y espaciales en la estructura del genoma. Es la hipótesis de esta tesis que p53 ejerce su influencia al modular las conexiones físicas entre elementos cis-reguladores a través de la manipulación de los bucles de cromatina.

Para evaluar esta hipótesis, se ha llevado a cabo una investigación exhaustiva sobre los posibles efectos de la activación de p53 en líneas celulares HCT116 mediante la inhibición farmacológica de MDM2 con el fármaco Nutlin-3a. Este enfoque integra de manera sinérgica diversos conjuntos de datos genómicos, que incluyen metodologías como Hi-C, Promoter Capture Hi-C, RNA-seq y ChIP-seq.

Los resultados de esta tesis demuestran que p53 induce cambios rápidos en los diferentes niveles jerárquicos del genoma, afectando el programa de transcripción de p53.

Al descifrar la conexión entre la organización tridimensional del genoma y la regulación de la transcripción impulsada por p53, esta tesis aspira a contribuir en la identificación de posibles intervenciones terapéuticas en el ámbito del tratamiento del cáncer.

Palabras clave: Organización de la cromatina 3D, p53, NGS, genómica, bioinformática.

Thesis overview

In 2019, the year I embarked on this thesis, marked both the 40th anniversary of p53's discovery and the 30th anniversary of its recognition as a tumor suppressor. Despite the elapse of time, there are still unanswered questions. This thesis aims to investigate how p53 functions through the lenses of 3D chromatin structure, exploring for the first time how it modifies the topology of the genome to control transcriptional programs and serve as a 'guardian of the genome'.

To provide the readers a comprehensive understanding of the findings presented in this thesis, the introduction covers various aspects of the genome as its structure, function, regulation, organization, and variation together with fundamental insights about the p53 protein. Subsequent sections detail the methods used for data analysis, providing insights into the experimental techniques and computational approaches employed, present key findings and their significance, discuss the implications of the results obtained, and conclude by summarizing the main research outcomes. By structuring the thesis in this manner, I hope the reader will be guided through a comprehensive journey to the specific investigation of p53 I carried out during these years of my PhD.

The results I am about to present are entirely my own work and are part of the publication below:

p53 rapidly restructures 3D chromatin organization to trigger a transcriptional response. François Serra, Andrea Nieto-Aliseda, Llorenç Rovirosa, **Mónica Cabrera-Pasadas**, Lucía Fanlo, Aleksey Lazarenkov, Blanca Urmeneta, Andrei L. Okorokov, Peter Fraser, Jose Luis Sardina, Alfonso Valencia and Biola M. Javierre. Manuscript accepted under major revisions at Nature Communications.

- Journal Impact Factor: Nature Communications, 17.694 (2023)
- Candidate Contributions: Performed preliminary bioinformatic analysis of the PCHi-C experimental data generated at the Babraham Institute by Biola M Javierre before the establishment of her laboratory as a PI, to understand the potential of the project for a PhD thesis. Contributed to the project design for

further experiments after the confirmation of its potential. Performed data processing, quality control and downstream analysis to analyze the chromatin structure data (Hi-C and new PCHi-C experiments), the omics data (ChIP-seq, RNA-seq) as well as its integration. The Hi-C part was supervised by François Serra, postdoc expert in the field to process and analyze Hi-C data. Interpreted the results with my supervisors guidance. Created the figures of the Hi-C section, PCHi-C, omics QCs and integration although some of them were improved on inkscape by another person. Written figure legends and methods and uploaded the data into GEO.

In addition, I contributed to the following publication and projects that are not detailed in this thesis during the COVID-19 lockdown and subsequent months due to the limitations to send or receive experiments about my project on p53:

- **The HDAC7–TET2 epigenetic axis is essential during early B lymphocyte development.** Alba Azagra, Ainara Meler, Oriol de Barrios, Laureano Tomás-Daza, Olga Collazo, Beatriz Monterde, Mireia Obiols, Llorenç Rovirosa, Maria Vila-Casadesús, **Mónica Cabrera-Pasadas**, Mar Gusi-Vives, Thomas Graf, Ignacio Varela, José Luis Sardina, Biola M Javierre, Maribel Parra. *Nucleic Acids Research*, Volume 50, Issue 15, 26 August 2022, Pages 8471–8490, Published: https://doi.org/10.1093/nar/gkac619
    - Journal Impact Factor: Nucleic Acids Research, 14.9 (2023)
    - Candidate Contributions: Performed the ATAC-seq bioinformatic analysis.

- Construction of 3D models depicting a 50Mb region on chromosome 17 within Monocytes, Naive B cells, and Naive CD4 cells with guidance of TADdyn developer Julen Mendieta (Marti-Renom Lab). These models were developed as part of an application for an ERC grant within the Javierre laboratory.

- Collaborative project with Nicolas Simonet, a former postdoc from Alex Vaquero's lab and now a postdoc with Jeannie Lee at Harvard University where I analyzed his Hi-C data on chromosome X and integrate it with histone marks to understand the role of Sirtuin 7 in chromosome X inactivation with guidance François Serra.

List of figures

List of tables

Table of contents

Introduction

1.The genome

A genome is the complete set of genetic material stored in every cell of a living organism encompassing all the hereditary information that guides an organism's development, functioning, growth, and reproduction.

1.1.The genome structure

Deoxyribonucleic acid, or **DNA**, is the genetic material of almost every organism. DNA comprises three essential components at its most fundamental level: a phosphate group, a sugar molecule called deoxyribose, and a nitrogen base. These three constituents —the sugar, phosphate, and any of the four nitrogen bases— are known as **nucleotides**. In the DNA, there are four distinct nucleotides, each characterized by a different nitrogenous base: Adenine (often abbreviated as "A" in scientific literature), Cytosine ("C"), Guanine ("G"), and Thymine ("T") **(Fig 1).**

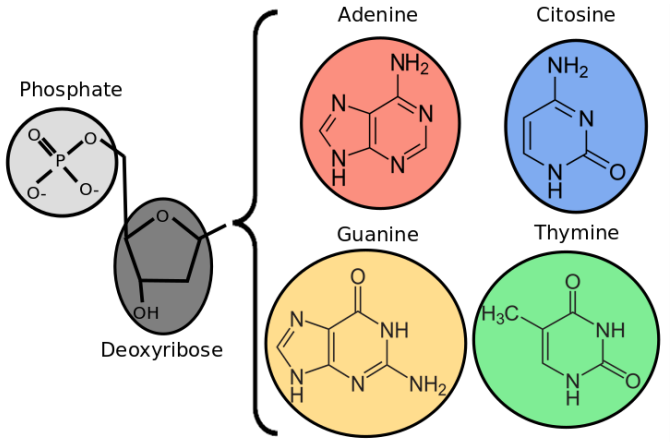


Fig 1. The structure of a nucleotide. The arrangement of a nucleotide encompasses the covalent bonding of a phosphate group (in light gray) to a deoxyribose unit (in dark gray), linked to a nitrogenous base (A, C, G, or T, in red, blue, orange, or green respectively).

Nucleotides can link together sequentially to form a structure known as a **polynucleotide**, with the arrangement of nucleotides ultimately determining the unique characteristics of each organism (**Fig 2A**). Nevertheless, DNA assumes its most stable form when two polynucleotides arrange; the nitrogen bases of one polynucleotide are attached to the nitrogen bases of the other polynucleotide, forming a **double-stranded DNA**. This base-to-base bonding is not random; the four bases always pair up like A with T and C with G, creating a **base pair unit (Fig 2 B)**. Therefore, the information on one DNA strand is redundant to that on the other,

ensuring the recovery of damaged strands using its complementary template. In the case of humans, the genome is over $3 \cdot 10^9$ base pair (bp) long.

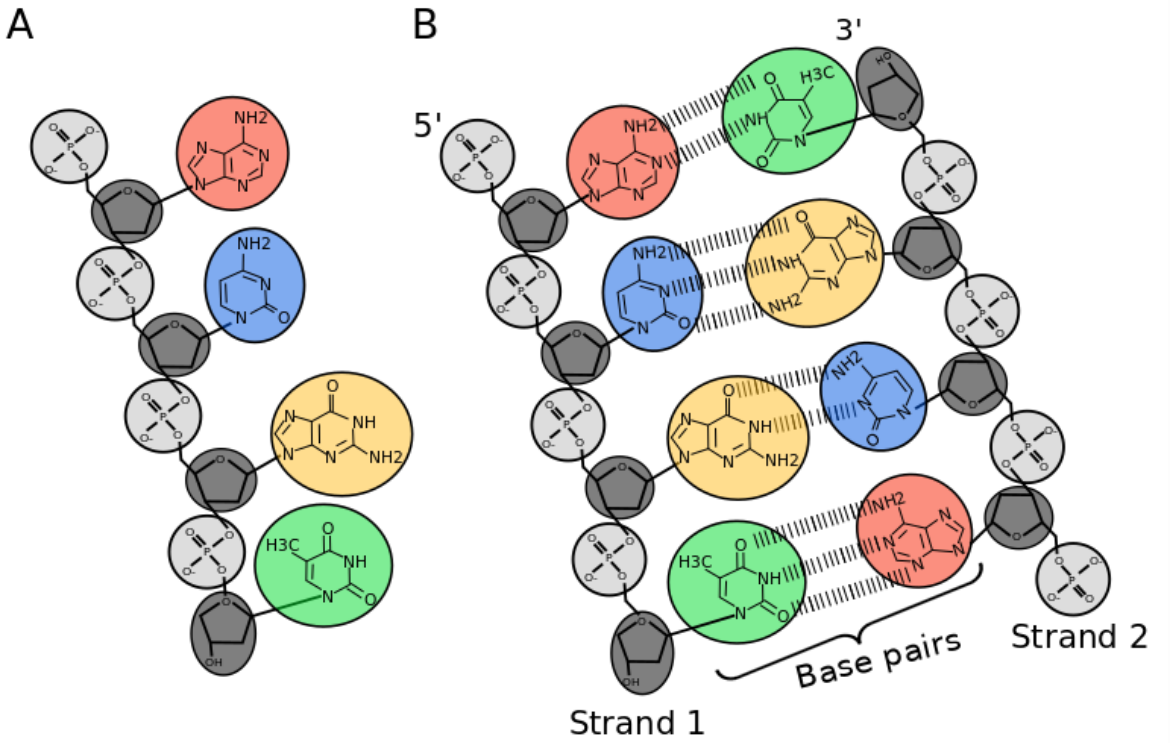


Fig 2. The molecular structure of DNA. A) Nucleotides binding each other through the phosphate molecule generating a polynucleotide **B)** Simplified diagram of the DNA model to focus on the base pairing. Two strands of DNA oppositely oriented (direction marked by the 5' and 3') hold together through electrostatic forces (hydrogen bonds represented by the lines between the nitrogen bases).

For cells with a nucleus (eukaryotes), most DNA is in the nucleus (where it is called nuclear DNA), and a small amount of DNA can be located in the mitochondria (mitochondrial DNA or mtDNA). In cells without a nucleus (prokaryotes) DNA is found freely inside the cell. In both cases, the electrostatic repulsion between adjacent phosphates of the DNA structure makes DNA a stiff, negatively charged polymer that cells cannot accommodate as a straight piece(1). To overcome this issue, in eukaryotes, specialized proteins called **histones** play a crucial role in packaging DNA within the nucleus. Histones belong to a family of small, positively charged proteins designated as H1, H2A, H2B, H3, and H4. They form a **histone octamer**, consisting of two copies each of H2A, H2B, H3, and H4, which wraps approximately 147 base pairs (bp) of DNA (**Fig 3**). The H1 histone will bind to this core to further stabilize it.

The histone octamer, the H1, and the DNA wrapped together are referred to as **nucleosome**.

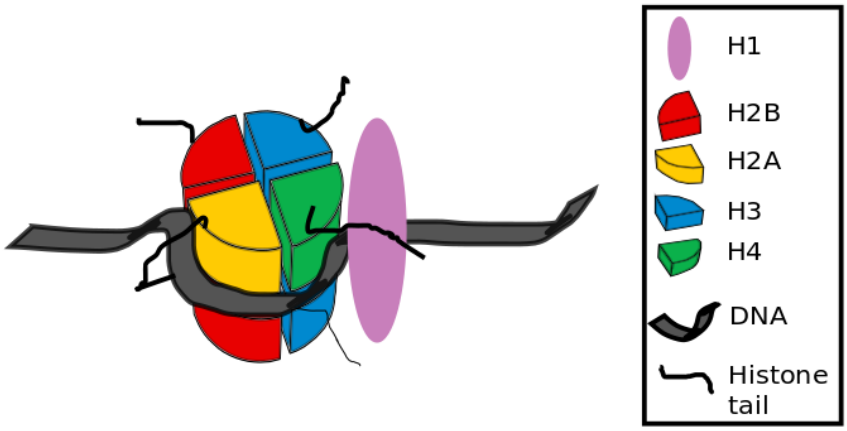


Fig 3. Nucleosome structure. Two copies of H2A, H2B, H3 and H4 histones form the nucleosome core particle and the H1 binds to the nucleosome core.

Nucleosomes, which manifest a continuous mobility oscillating and adapting its positions, associate with each other to form a higher order structure called the **chromatin fiber**. Depending on the compaction of the nucleosomes, the chromatin fiber results in the differentiation of two states: euchromatin and heterochromatin (**Fig 4**). In the state of euchromatin, the chromatin fiber adopts a less compacted and therefore more open and relaxed configuration that allows the DNA to be accessible. Conversely, within the state of heterochromatin, the chromatin fiber assumes a compacted and guarded stance making the DNA less accessible.

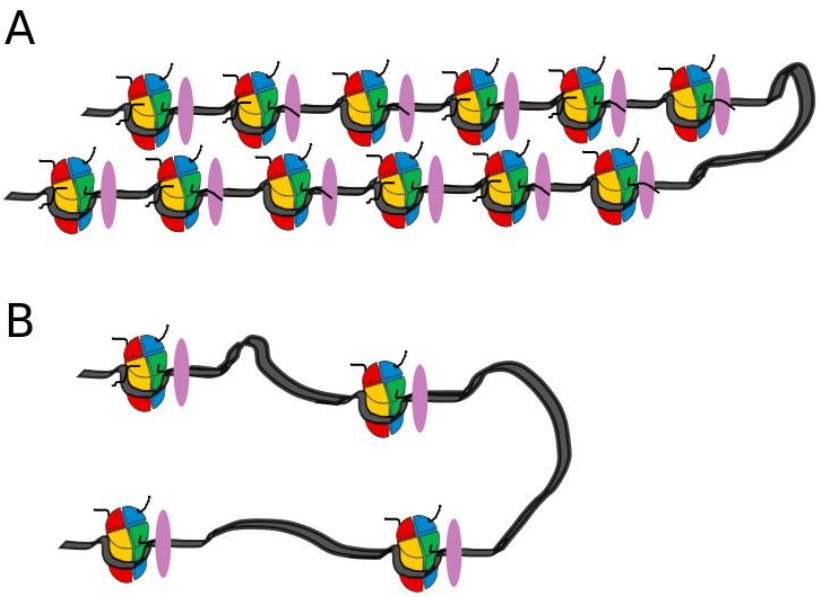


Fig 4. Chromatin fiber. Two states of chromatin fiber regarding its accessibility. **A)** Heterochromatin, highly compacted chromatin fiber **B)** Euchromatin, loosely compacted chromatin.

1.2. The genome function

The unique structure of DNA is linked to its function as it enables the transfer of genetic information and its transformation into functional products. This fundamental principle on how the genetic information flows within a biological system is often referred to as the central dogma of molecular biology (**Fig 5**).

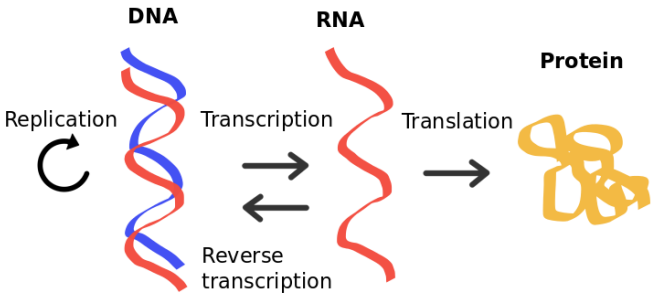


Fig 5. Central dogma of molecular biology. The genetic information flows from existing DNA to make new DNA (replication) or RNA (transcription), and from RNA to make a functional product, a protein (translation).

1.2.1. Replication

Each double-stranded DNA molecule unzips by an enzyme called helicase which breaks the hydrogen bonds and separates them into two single strands. One of the strands is oriented in the 3' to 5' directions while the other is oriented in the 5' to 3' direction. Therefore, the two strands are replicated differently as result of their different orientations. In the 3' 5' strand, a short piece of RNA called **primer** becomes the starting point of DNA synthesis. Then the **DNA polymerase** adds new complementary nucleotide bases to the strand of DNA in the 5' to 3' direction. While in the 5' 3' strand, numerous RNA primers are made by the primase enzyme and bind at various points along the strand. Finally, an enzyme called **DNA ligase** seals up the sequence of DNA into two continuous double strands. The result of DNA replication is two identical copies of the DNA molecule.

Cell replication is tightly regulated and occurs through a series of sequential events called the **cell cycle**. In eukaryotes, the cell cycle consists of four phases: G1 (Gap phase 1), S (Synthesis phase), G2 (Gap phase 2), and M (Mitosis phase). During the M phase, the chromatin condenses into tightly packed structures named **chromosomes** to facilitate the separation of the genetic material.

1.2.2. Functional products

Only approximately 1% of the human genome is **coding DNA**, that is, encodes proteins(2). In contrast, the remaining 99% of the human genome is **non-coding DNA**, which researchers historically considered "junk DNA" due to the absence of apparent function. However, extensive research over the past few decades has unveiled the essential roles of non-coding DNA in gene regulation, genome organization, and various cellular processes. Non-coding DNA includes a diverse array of sequences, such as: promoters and enhancers or non-coding RNAs (ncRNAs) further explained below.

The process of going from DNA to a functional product is known as **gene expression**, and it involves two significant steps:

- **Transcription**: The process involves copying a specific DNA segment (known as a **gene**) to produce an RNA molecule in the nucleus of eukaryotic cells and the cytoplasm of prokaryotic cells, with three major stages: initiation, elongation, and termination.

As important note, **reverse transcription** (the transfer of information from RNA to make new DNA) can occur but it only happens in retroviruses such as HIV.

- **Translation**: This process takes place in ribosomes and comprises four distinct stages: initiation, elongation, termination, and recycling. During translation, the messenger RNA (mRNA) sequence is decoded in sets of three nucleotides, known as codons, each of which corresponds to a specific amino acid. Transfer RNA (tRNA), a type of non-coding RNA, plays a crucial role by transporting the appropriate amino acids and facilitating their binding to the codons through complementary base pairing. This orchestrated process ensures the accurate assembly of amino acids in the correct sequence, ultimately giving rise to a functional protein with a specific biological function.

1.3. The genome regulation

Through a series of tightly controlled steps, cells within the same organism, harboring identical DNA, can exhibit diverse responses to internal and external cues adapting their gene expression patterns to changing conditions and developmental stages.

1.3.1. Transcriptional regulation

Transcriptional regulation plays a crucial role in controlling which genes are transcribed in different cell types and in response to various environmental conditions. This process entails regulatory proteins, such as **transcription factors (TFs)**, binding to specific non-coding DNA sequences known as **cis-regulatory elements (CREs)**. These regulatory elements can enhance or repress gene transcription by interacting with the transcriptional machinery. The focus of this thesis is on two types of CREs: promoter and enhancers. However, there are other types of CREs, such as super-enhancers, silencers, and super-silencers among others.

1.3.1.1. Promoters

Promoters are non-coding DNA regions situated at the initiation site of a gene, known as the **transcription start site (TSS).** These regions serve as recruiting platforms for TFs and the **RNA polymerase II (RNAPol II)**, the enzymatic machinery responsible for gene transcription. The assembly of these elements forms the **pre-initiation complex (PIC)**, the basic transcription machinery required to start the transcription of a gene. The engagement of RNAPol II and the functionality of the PIC can undergo alteration through the influence of TFs.

Promoters encompass two distinct categories: constitutive promoters, which maintain activity across all cells and tissues, and tissue-specific promoters, selectively active within specific cellular contexts, tissues, or in response to a particular stimulus.

1.3.1.2. Enhancers

Enhancers are non-coding DNA regions that play a crucial role in controlling the gene expression of their target genes. These regions can be located either upstream or downstream of the gene they regulate, and they can function over long distances, at

various distances from their target genes(3)(4). The relationship between enhancers and promoters is not linear. An enhancer can regulate multiple gene promoters, and multiple enhancers can regulate a promoter (**Fig 6)**. Per cell type, about four enhancers contact an active gene(5). This non-linear relationship suggests that enhancers function as modular units of gene expression, exhibiting seemingly additive and redundant effects on their target genes. This redundancy ensures the robustness and reliability of gene expression.

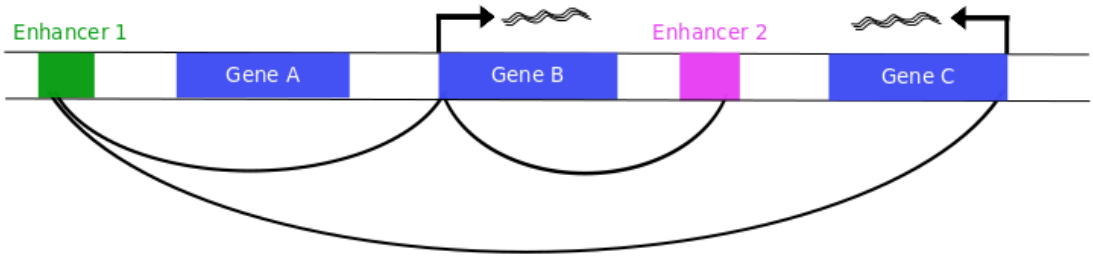


Fig 6. Enhancer regulation diagram. Interactions among enhancers and gene promoters (illustrated as arcs). A single gene can be controlled by multiple enhancers (Gene B with Enhancer 1 and 2), and multiple genes can be regulated by the same enhancer (Gene C and B with Enhancer 1). Enhancer 1 jumps over a gene that does not regulate (Gene A).

1.3.2. Post-transcriptional regulation

Post-transcriptional regulation controls how available the mRNA is to ribosomes and encompasses a series of essential processes that contribute to the maturation, modification, and ultimate destiny of RNA molecules. These processes introduce an additional layer of complexity to gene expression, enabling the generation of multiple mRNA isoforms, precise regulation of protein synthesis in terms of timing and location, and the integration of diverse cellular signals.

Firstly, the newly transcribed RNA molecule, known as the primary transcript or pre-mRNA, undergoes a process called capping. A modified nucleotide is added into the beginning of the RNA molecule to provide stability and help protect the RNA from degradation. Additionally, the cap is essential for the recognition and binding of the mRNA to the ribosome during translation. Following capping, the pre-mRNA molecule undergoes splicing. Introns, non-coding regions withing the gene sequence, are removed from the pre-mRNA, and the remaining exons, which contain the protein-coding information, are joined together. This process is mediated by a complex

molecular machinery called the spliceosome. Alternative splicing can occur, leading to the production of different mRNA isoforms from a single gene, thereby increasing the diversity of proteins that can be generated. After splicing, the mRNA molecule undergoes polyadenylation. A long stretch of adenine nucleotides, known as the poly(A) tail, is added to the 3' end of the mRNA. This poly(A) tail plays a crucial role in mRNA stability and efficient translation. Once the mRNA molecule has undergone capping, splicing, and polyadenylation, it is considered mature and can be exported from the nucleus to the cytoplasm. In the cytoplasm, the mature mRNA serves as a template for protein synthesis during the process of translation. The ribosomes recognize the mRNA and decode its sequence to produce the corresponding protein.

It is important to note the difference between the **nascent transcript** and the **mature mRNA**. The nascent transcript refers to the RNA molecule being synthesized during transcription, which includes both the exons and introns. It is the primary transcript before any processing occurs. In contrast, the mature mRNA is the processed and modified version of the nascent transcript, where introns have been removed, and the exons have been joined together. The mature mRNA is the final product that carries the genetic information from the gene to the ribosomes for protein synthesis.

1.3.3. Translational regulation

Several stresses can lead to changes in gene expression patterns resulting in selective recruitment of ribosomes to mRNAs whose protein products are required for responding to stress. Translational regulation is achieved through a variety of mechanisms that control the rate and efficiency of protein synthesis during translation. These mechanisms act at different stages of the translation process, primarily during initiation but also during elongation and termination.
miRNAs and siRNAs, ncRNA, can bind to target mRNAs and prevent their translation by inhibiting ribosome binding or promoting mRNA degradation.

1.3.4. Post-translational regulation

After the protein has been synthesized, through a diverse array of biochemical reactions, post-translational modifications (PTMs) can dynamically and reversibly modify the structure and properties of proteins adjusting their activities to fulfill cellular

requirements. Each PTM introduces specific chemical changes to the protein, resulting in distinct functional consequences. Several examples of such modifications include **Phosphorylation**, **Ubiquitination**, **Acetylation** and **Methylation**.

1.3.4.1. Histone PTMs

Histone tails contain many amino acids that can be post-translationally modified(6). These modifications are a type of the so called "**epigenetic**" modifications, a diverse range of molecular alterations that influence gene expression without altering the DNA sequence(7). The unique combinations and patterns of histone PTMs, often referred to as **chromatin signatures**, hold the key to deciphering the transcriptional states of enhancers and promoters(8). By examining their distribution and co-occurrence within the genome, we can gain insights into gene regulation.

This thesis centers its focus on two crucial histone modifications:

- **Histone H3 Lysine 4 Monomethylation:** abbreviated as **H3K4me1,** involves the addition of a single methyl group to a lysine residue at position 4 on the histone H3 protein. H3K4me1 is instrumental in marking enhancers in a "**primed state**", meaning they possess the potential to become active upon receiving specific signals or stimuli.
- **Histone H3 Lysine 27 Acetylation**: abbreviated as **H3K27ac**, involves the addition of an acetyl group to a lysine amino acid at position 27 on the histone H3 protein. Acetylation typically leads to a more relaxed chromatin structure, rendering the DNA in that region more accessible for transcription. H3K27ac is particularly significant as it is strongly associated with active gene expression. Its presence at specific locations along the genome serves as a marker for the "**active state**" of enhancers and promoters.

Therefore, in this thesis these two PTMs serve as valuable indicators to detect primed enhancers with the presence of H3K4me1, active enhancers exhibiting a co-occurrence of H3K4me1 and H3K27ac, underscoring their indispensable role in gene expression(9), and active promoters distinguished by the presence of the H3K27ac mark.

1.4.The genome organization

The organization of the genome is a complex and dynamic process that determines how genetic information is stored, accessed, and regulated within the nucleus of a cell. Contrary to the notion of a linear and randomly packed DNA sequence, it has become increasingly clear that the genome is intricately organized within the nucleus, with distinct levels of chromatin organization and higher-order structures (**Fig 7**).

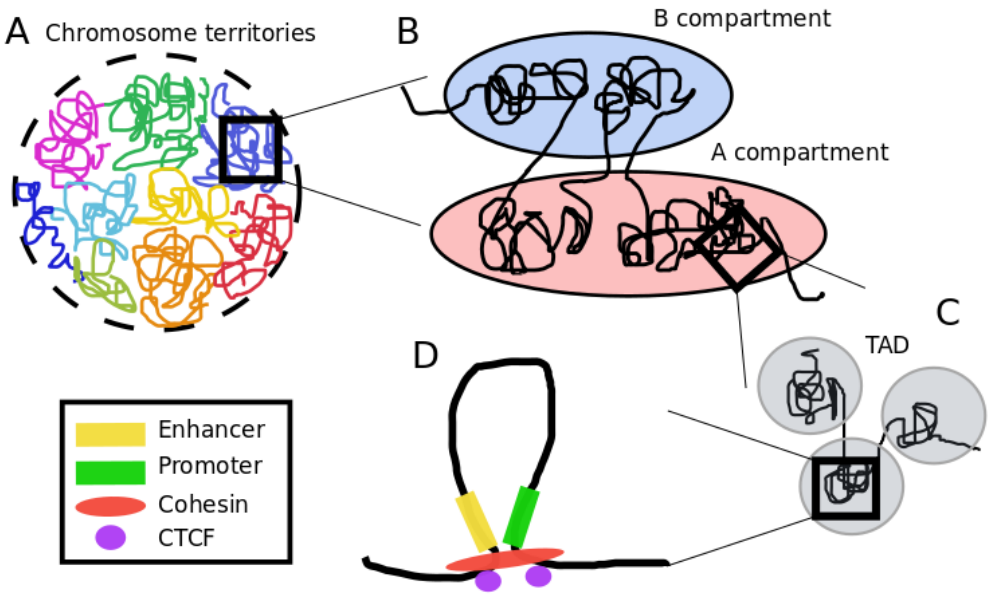


Fig 7. Hierarchical organization of the 3D genome in the nucleus. The 3D genome units of chromatin in the nucleus are represented in descending order: A) Chromosomal territories (CTs): Each chromosome occupies its own distinct and non-overlapping region within the nucleus. This arrangement promotes interactions within the same chromosome (cis interactions) more than interactions between different chromosomes (trans-interactions). **B)** Chromatin compartments within a single chromosome: These compartments are divided into two categories, A and B compartments, which exhibit distinct spatial patterns within the genome's structure. **C)** Compartment subdivision into topologically associated domains (TADs): TADs play a crucial role in folding the genome and guiding the regulation of long-range interactions. **D)** Chromatin loops: Typically established through interactions between distant elements, such as promoters and enhancers facilitated by specific protein complexes, including CTCF (CCCTC-binding factor) and cohesin.

1.4.1. Chromosome territory

At the largest scale, known as the chromosome level, within an interphase nucleus, each chromosome occupies a non-random position called a chromosome territory(10)(11). This spatial organization of chromosomes has been observed in many organisms and cell types, highlighting its fundamental role in how the genome functions, and is regulated. This distribution suggests that the spatial distribution of the chromosomes can influence their accessibility to the transcriptional machinery and regulatory factors, ultimately impacting gene expression levels.
It is important to note that the positioning of chromosomes within chromosome territories is not a static phenomenon but can dynamically change in response to cellular processes, developmental stages, and environmental cues(12).

1.4.2. A/B compartments

At a smaller scale, in the range of megabases, one notable characteristic of chromosome organization is the spatial separation of transcriptionally active and inactive chromatin, labeled arbitrarily to as A and B compartments, respectively(13). Gene-rich regions tend to be positioned toward the interior of the nucleus, forming the A compartment. In contrast, the nuclear periphery contains gene-poor regions, constituting the B compartment. As chromosome territories, the concept of compartments relies on the observation that genomic loci within a compartment contact each other more frequently than between compartments.

Studies performed in human(13–15), mouse(14,16), Drosophila(17,18), and several other species, have shown that compartmental switches occur during development and differentiation processes. This dynamic nature of A/B compartments reflects the adaptive capabilities of the genome in response to different stimuli.

1.4.3. Topological Associating Domains

At the kilobase scale, chromatin is organized into a substructure within the A/B compartments called topologically associated domains (TADs) that is characterized by a higher frequency of interaction within the region compared to interactions withing regions outside of the TADs(19,20). Recent studies with high sequencing depth and

improved resolution allowed the appreciation of domain sizes ranging from 40kb to 3Mb(21).

TADs are bordered by low interaction regions called **TAD borders**(22). TAD borders are essential for preserving the integrity of genomic domains, facilitate cell-type-specific enhancer-promoter interactions, and prevent undesired interactions between regulatory elements in neighboring regions. Perturbations in these borders can lead to aberrant rewiring of enhancer-promoter interactions, contributing to developmental disorders and cancer progression(14,23). The loss of TAD borders alone may not always be adequate to dysregulate gene expression in nearby genes. Instead, it should be accompanied by specific rewiring of enhancer-promoter contacts to elicit functional consequences(24,25).
Unlike compartments, TAD borders are relatively conserved (50%-90%) between species and tissues(14). However, individual cells of the same type do not consistently define the same TAD borders, even though they can be statistically identified in cell populations(26). Therefore, within a population, the TAD border positions have cell-to-cell heterogeneity, but each cell will likely possess globular structures with start and end coordinates that partially overlap with the TADs.

1.4.4. Chromatin Loops

Chromatin loops are the highest level of chromatin organization and play a crucial role in bringing distant regulatory elements, such as enhancers and promoters, into proximity, thereby facilitating their long-range interactions and enabling the regulation of gene activity(15,27). It is important to note that chromatin loops exhibit high dynamism and can undergo remodeling in response to developmental signals, environmental stimuli, and alterations in cellular states. This dynamic nature allows for precise and context-dependent regulation of gene expression (28).

Disruption or dysregulation of chromatin loops has been implicated in various diseases, including cancer, where aberrant interactions between regulatory elements can lead to abnormal gene expression patterns and contribute to disease progression(24).

1.4.5. Architectural proteins

Some of the key players in maintaining genome architecture are:

- The **CCCTC binding factor (CTCF)** is a zinc finger protein that binds to specific DNA sequences in the genome, known as CTCF-binding sites. CTCF is involved in diverse genomic functions, including insulation, chromatin looping, and enhancer-blocking. By binding to CTCF-binding sites, CTCF helps establish and demarcate boundaries between genomic regions, preventing the spreading of chromatin modifications and facilitating the formation of topologically associating domains (TADs).

- The **cohesin complex** is a ring-shaped protein complex formed by SMC1A/B, SMC3, STAG1/2 and RAD21 able to bind to the DNA and reel the chromatin fiber through it. Cohesin depletion vastly reduces chromatin looping(29).

1.5.Decoding the genome

Decoding the genome refers to the process of understanding the organization, arrangement, and functional elements of the DNA within an organism's genome.
Early approaches to capture the information stored in the genome relied on **DNA sequencing**, a technology designed to determine the precise order of the four nucleotide bases within a DNA strand. The pioneering method known as **Sanger sequencing** was developed by Frederick Sanger in 1977.

Due to the limitations of DNA sequencing methods, it is not feasible to sequence the entire genome as a single unit. To overcome this challenge, the genome is fragmented into smaller pieces, known as **reads**. These reads are sequences of DNA that can be individually analyzed and sequenced. By breaking the genome into these smaller fragments, it becomes more manageable to sequence and process the DNA. Once the reads are obtained through sequencing, the next crucial step is to reconstruct the original sequence of the genome by aligning and overlapping the reads. This process is known as genome assembly and requires sophisticated computational algorithms and bioinformatic techniques.

Over time, in the early 2000s, technological advancements and the advent of **Next-Generation Sequencing (NGS)** methods have transformed the field of DNA sequencing(30). NGS technologies have greatly enhanced the speed, throughput, and cost-effectiveness of sequencing by enabling the simultaneous sequencing of millions to billions of DNA fragments in a single experiment. These methods use massively parallel sequencing platforms that generate vast amounts of sequencing data, revolutionizing the ability to capture genomic information at an unprecedented scale.

The field of **bioinformatics** has played a crucial role in leveraging the enormous amount of genomic data generated by NGS technologies. Bioinformatics tools and pipelines are indispensable for processing, analyzing, and interpreting the massive datasets produced by NGS.

1.5.1. Techniques to study 3D genome structure

The pioneer method that allowed the detection of physical contacts between specific loci providing insight into their spatial proximity was the Chromosome conformation capture, or **3C**, technique(31). It involves four steps: 1) formaldehyde fixation to crosslink proteins to DNA and other proteins allowing to capture the regions where these molecules come into physical contact at a certain timepoint, 2) digestion of the chromatin with a restriction enzyme, such as *Hind*III(31), *Bgl*II(32) or *Dpn*II (33) among others; 3) ligation of the cross-linked fragments, 4) detection by PCR and analysis of ligation junctions to determine contact frequencies of the regions investigated. Building upon 3C, additional techniques such as 4C, 5C and High-throughput Chromosome Conformation Capture (**Hi-C**) were developed to broaden the scope of chromatin interaction analysis.

In Hi-C, a crucial step involves labeling all genomic fragments with biotinylated nucleotides before ligation, a process that serves to mark the ligation junctions. These marked junctions can be efficiently isolated and purified by streptavidin-coated magnetic beads, enriching the library for ligation products that can be detected by NGS (**Fig 9**). It allowed the identification for the first time of A/B compartments and TADs and is even widely used up until today. Between 2012 and 2015, several modifications to the Hi-C protocol took place. One of the improved Hi-C is the ***in situ*** **Hi-C**(15), performing ligation within the nucleus effectively minimized inter-chromosomal interactions and promoted the formation of intra-crosslinked complexes. As a result, this method became the established standard for the ligation step in all subsequent 3C techniques.

Hi-C analysis involves specialized bioinformatic tools to construct a list of DNA-DNA contacts, collectively referred to as a Hi-C Map, derived from Hi-C experiments. This process involves segmenting the linear genome into fixed-size loci, allowing the Hi-C map to be represented as a contact matrix. Principal Component Analysis (PCA) is applied to the Hi-C matrix to identify patterns or structure in the data. PCA finds linear combinations of the rows and columns of the matrix, called eigenvectors, that capture the most significant variation in the data dividing the genome into two compartments based on the sign of its entries. The first eigenvector (PC1) typically captures the most

prominent large-scale structural features. Boundary changes between the two compartments occur where the entries of the first eigenvector change sign. The sign of this eigenvector is arbitrary. By plotting PC1 as a heatmap, the segregation of chromatin into A (active) and B (inactive) compartments within the nucleus can be visualized. The resolution of the Hi-C map reflects the finest discernible scale at which local features can be reliably identified (**Fig 8**).

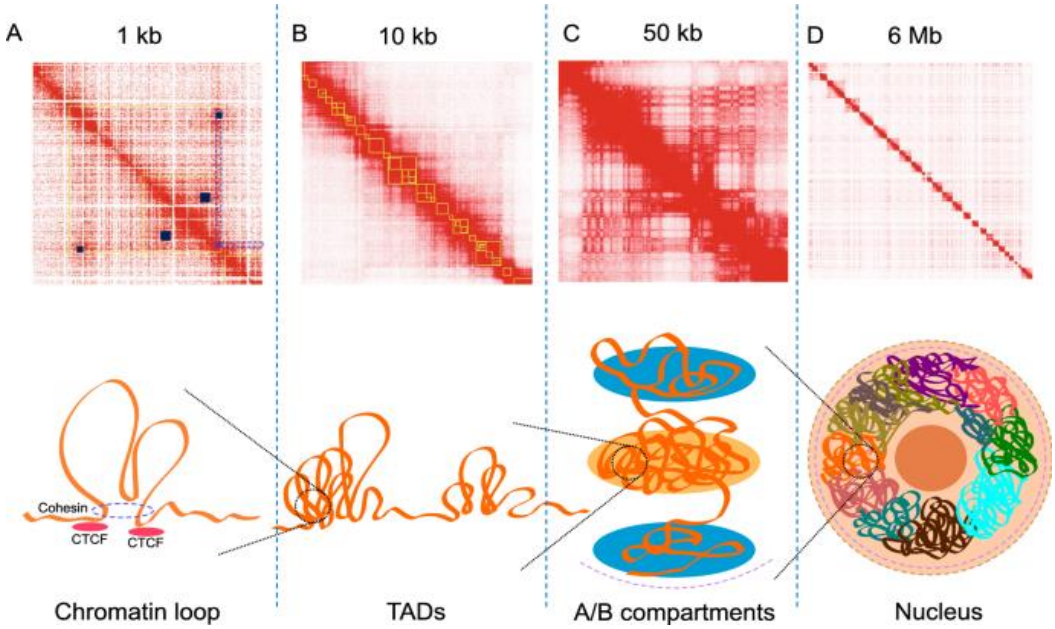


Fig 8. Visual representation of the structural organization of the genome and the corresponding Hi-C interaction maps. Upper Panel: Interaction heatmaps A, B, C, and D are displayed at various resolutions (measured in kilobases or megabases per pixel) to align with the 3D structure illustrations presented in the lower panel. In heatmaps A and B, TADs are denoted by yellow boxes, and chromatin loops are indicated by small blue boxes within heatmap A. The prominently interacting region highlighted by the purple box in heatmap A exhibits a characteristic "V" shape pattern, depicted with purple dotted lines. Lower Panel: Depictions of 3D genome structures. **Source**:(34).

However, achieving high-resolution contact maps for studying specific interactions, such as those between enhancers and promoters, requires deep sequencing with high cost associated. To address this problem, a variation of the Hi-C technique was created to focus on the interactions involving gene promoters called Promoter Capture Hi-C(35). Promoter Capture Hi-C (**PCHi-C**) combines the power of Hi-C with targeted enrichment of promoter regions, allowing for higher-resolution analysis interactions of each promoter in a cell type with other DNA elements of the genome, such as

enhancers or other promoters (**Fig 9**). PCHi-C can enrich more than 28,000 human and 27,00 mouse gene promoters using an RNA probe capture system allowing to define the promoter interactome.

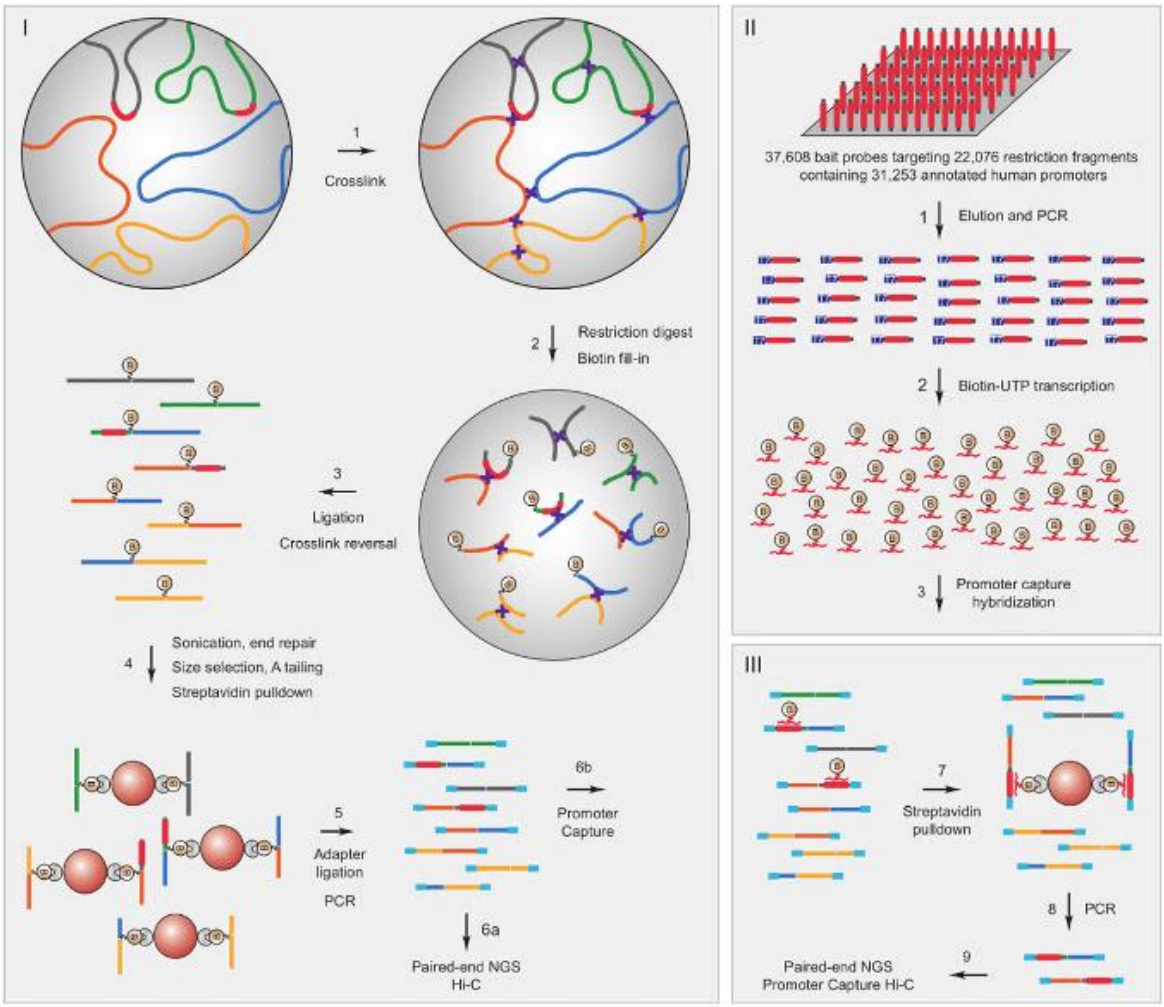


Fig 9: Hi-C and PCHi-C experimental strategy: (I) Generation of the Hi-C libraries as described in Lieberman-Aiden et al., 2009(13), subjected to Promoter Capture (step 6b). Chromosomal regions are depicted in blue, green, grey, orange, and yellow, and promoters are depicted in red. Biotin moieties are symbolized by an encircled 'B', and formaldehyde crosslinks are represented by purple crosses. (II) To generate biotinylated RNA capture baits, the corresponding DNA fragments are synthesized on microarrays, eluted and PCR-amplified with primers containing a T7 promoter sequence (blue) for subsequent in vitro transcription in the presence of biotin-UTP, essentially as previously described(36). The RNA baits are designed to target the sequences at the ends of HindIII restriction fragments containing promoter elements. RNA molecules are represented by red fragments connected to a biotin moiety (encircled 'B'). (III) For Promoter Capture, the Hi-C library ('pond') is hybridized to the RNA

capture library (‘bait’) in solution, followed by streptavidin pull-down of Hi-C library ligation products containing promoters targeted by the biotin-RNA baits. The resulting Promoter Capture library is analyzed by massively parallel paired-end sequencing. **Source**:(35).

The analysis of PCHi-C data follows similar principles to standard Hi-C data analysis with further steps (they will be mentioned in the methods section). Since PCHi-C specifically targets promoter regions, the analysis can be tailored to prioritize the identification and interpretation of promoter-enhancer interactions. The results can include two types of interactions:

- **Promoter-promoter interactions:** interactions that involve physical contacts between promoters of different genes, which can provide insights into potential regulatory relationships, co-regulation, or functional coordination between genes.
- **Promoter-other end interactions:** interactions between gene promoters and other genomic regions, such as enhancers, insulators, or other regulatory elements. These interactions reveal the spatial proximity and potential functional interactions between gene promoters and distal regulatory elements, playing a crucial role in gene regulation and transcriptional control.

1.5.2. Techniques to study DNA-protein interactions

Even though there are several and different techniques to describe, analyze and compare DNA–protein interactions, considering genome-wide studies, chromatin immunoprecipitation followed by sequencing (**ChIP-seq**)(37) has been the elected method for many researchers. Two of the most used applications of ChIP-seq are the identification of TF and histone PTMs. ChIP-seq involves covalently crosslinking proteins to DNA by fixation with formaldehyde, shearing the DNA in a random manner by sonication, binding an antibody to the protein of interest, pulling-down the region where the antibody is bound, purify the DNA, and then analyzing the enriched DNA to determine binding sites (**Fig 10a**). The output is a genome-wide map of protein-DNA interactions, revealing the binding sites and distribution patterns of the studied protein throughout the genome.

Like all techniques, ChIP-seq has some limitations, prompting the development of alternative methods in recent years. One such method is cleavage under targets and release using nuclease, abbreviated as **CUT&RUN**(38). CUT&RUN eliminates two ChIP-seq time consuming steps which are the need for cell fixation and sonication (**Fig 10b**). More recently, an even more advanced iteration of this protocol, known as **CUT&Tag** (Cleavage Under Targets and Tagmentation) has emerged(39). CUT&Tag eliminates the need for a separate library preparation step as by using a hyperactive transposase tethered to protein A it incorporates NGS adapters directly while cleaving the DNA (**Fig 10c**). However, CUT&Tag has shown reduced efficiency when profiling TF occupancy and, therefore, its use has been centered on histone PTM profiling.

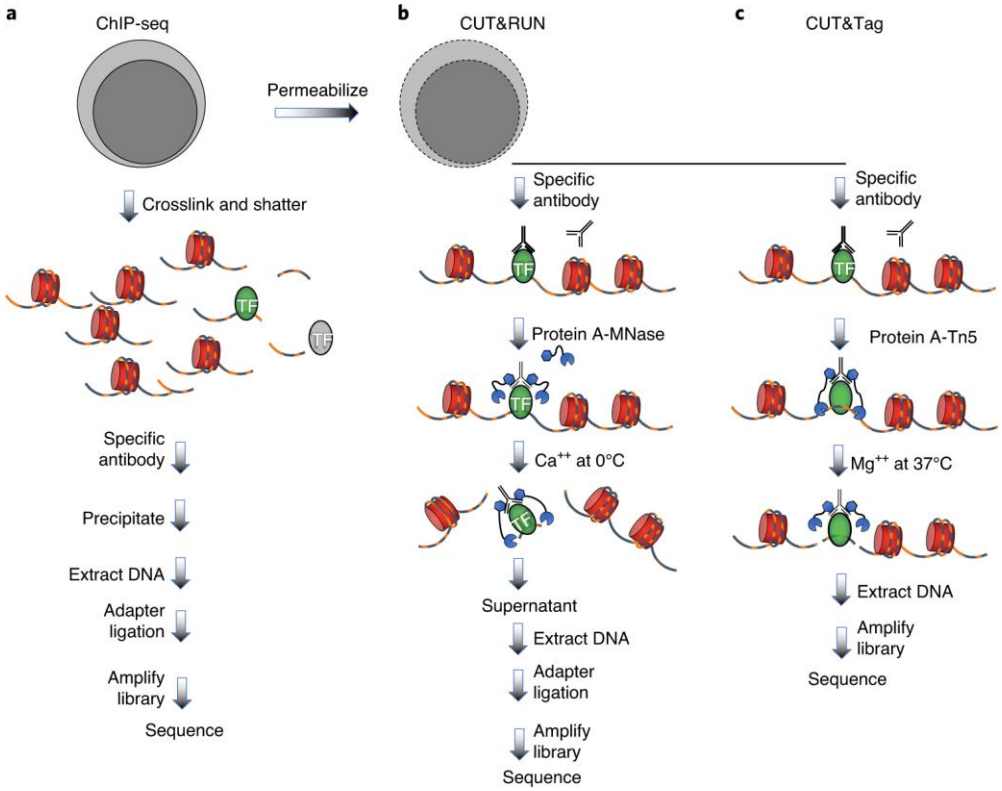


Fig 10. DNA-protein strategies. a) ChIP-seq, **b)**, CUT&RUN, **c)**, CUT&Tag. Cells are indicated in gray, chromatin as red nucleosomes, and a specific chromatin protein in green. See text for details of each procedure. **Source**:(40)

1.5.3. Techniques to study transcription

Techniques for studying transcription play a pivotal role in unraveling the intricacies of gene expression. **RNA-seq**(41) has emerged as the preferred technology for the measurement of transcript sequences and their abundance. The basic steps of an RNA-seq experiment involve RNA extraction, RNA fragmentation, cDNA generation,

library amplification, and sequencing (**Fig 11**). It is critical that the RNA sample being sequenced is enriched for the RNA type of interest. Without enrichment, most sequencing reads will correspond to ribosomal RNA (rRNA), which makes up over 90% of the cell's RNA. The two methods used to achieve enrichment in eukaryotic cells are rRNA removal and/or poly(A) selection of mRNA using oligo (dT) primers, which requires high-quality, high-abundance mRNA. The results are either aligned to a reference genome or transcripts, or they are assembled de novo without a genomic sequence to produce a genome-scale transcription map that consists of both the transcriptional structure and the level of expression for each gene.

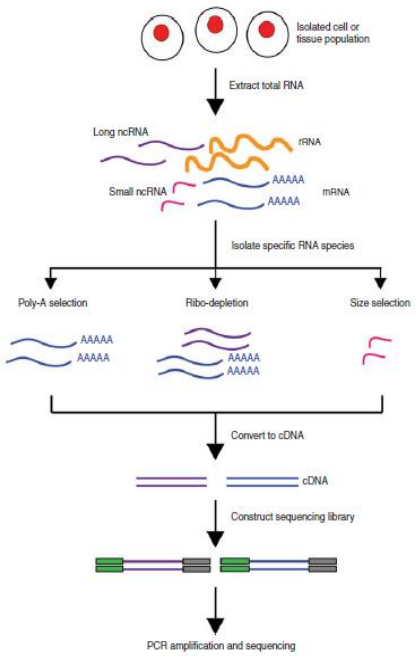


Fig 11. Overview of RNA-Seq. First, RNA is extracted from the biological material of choice (e.g., cells, tissues). Second, subsets of RNA molecules are isolated using a specific protocol, such as the poly-A selection protocol to enrich for polyadenylated transcripts or a ribo-depletion protocol to remove ribosomal RNAs. Next, the RNA is converted to complementary DNA (cDNA) by reverse transcription and sequencing adaptors are ligated to the ends of the cDNA fragments. Following amplification by PCR, the RNA-Seq library is ready for sequencing. **Source**:(42)

In RNA-seq, the same absolute amount of RNA is sequenced for every sample. This means that the relative abundance of a given mRNA is evaluated regardless of if the gene is being actively transcribed. Recently a method termed Global Run-On followed by sequencing (**GRO-Seq**)(43) has enabled the quantification of nascent transcripts. Capturing the actively transcribed mRNAs provides a direct measure of RNA polymerase II activity.

1.6.Variations in the genome

Variations in the genome, known as **mutations**, are alterations in the DNA sequence that can disrupt normal cellular functions and contribute to the initiation and progression of various diseases, including cancer(44).
In addition to genetic mutations, disruptions in the three-dimensional (3D) chromatin organization can lead to aberrant gene expression patterns, dysregulation of signaling pathways, and altered cellular functions, all of which contribute to tumorigenesis(24).

Among the many genes pivotal in orchestrating cancer suppression, the TP53 gene, responsible for encoding the p53 protein, stands as a sentinel. This guardian of genomic integrity regulates a multitude of cellular processes through its ability to induce the expression of key genes. Given its central role, it is not surprising that the p53 gene is either mutated or deleted in nearly half of all human cancers(45), with the p53 signaling pathway perturbed in the remaining cases. Remarkably, p53 mutations are particularly prevalent in some of the most aggressive cancer types. Despite several decades of research focused on targeting the p53 pathway, the design of effective therapies has remained an enduring challenge. Nevertheless, many ongoing clinical trials continue to explore this pathway as a potential therapeutic target across a range of human malignancies.

Therefore, it becomes imperative to delve deeper into the study of the p53 protein and its profound significance in the context of cancer to help in the research of new and effective therapies against cancer due to p53 mutations.

2.p53

2.1.Discovery of p53

The discovery of p53 can be traced back to the late 1970s when multiple groups discovered a cellular protein that was frequently overexpressed in cancer cells(46,47). The protein, which had a molecular weight of approximately 53kDa, was named p53 (protein p53). Further studies in the late 1980s revealed that when the p53 was introduced into cells that had lost their ability to undergo programmed cell death (apoptosis), the cells were able to undergo apoptosis suggesting that the p53 gene was able to detect damaged DNA and trigger apoptosis as a protective mechanism against cancer, being characterized as a tumor suppressor(48). In 1990, the status of p53 as a tumor suppressor was firmly established through the discovery that individuals afflicted with Li–Fraumeni syndrome, a condition predisposing them to various forms of cancer, had inherited mutations in the TP53 gene(49).

2.2.Structure of p53

In humans, the TP53 gene is located on the short arm of chromosome 17 (17p13.1) (50,51) and contains 11 exons and 10 introns(52). In addition to the full-length protein (p53α), the human TP53 gene encodes at least 15 protein isoforms, ranging in size from 3.5 to 43.7 kDa(53). The wild type full-length p53 active protein is composed of a tetramer, four identical chains of 393 amino acids with 3 functional domains(54):

- The **N-terminal region** contains the transactivation domain (TAD), and the Src homology 3-like (SH3) domain. The TAD domain is the region of p53 responsible for initiating transcription of target genes by interacting with various transcriptional factors and components of the basal transcriptional machinery. It is divided into two subdomains with complementary functions(55), the TAD subdomain 1 (TADI), provides the generic transactivation functions, while the second one (TADII) may play a role in the selection of specific transactivation targets. The SH3 domain, is a proline-rich domain required for interaction of p53 with SIN3, which protects p53 from degradation(56).

- The **DNA-binding domain (DBD)** is the central (core) domain (p53 core domain, p53C), and enables sequence-specific DNA binding allowing p53 to recognize and bind to specific response elements within target gene promoters. This region is pivotal for the transcriptional activity of p53. Numerous studies have demonstrated that the core DNA-binding domain of p53 is highly conserved across different species, indicating its essential role in maintaining genomic integrity. This conservation suggests that the binding specificity of p53 is preserved in various cell types(57,58).

- The **C-terminal regulatory domain** is the region of p53 responsible for regulating the activity of the protein. It contains multiple functional regions involved in protein-protein interactions and several post-translational modification sites, including phosphorylation and acetylation sites, which can alter the stability, subcellular localization, and transcriptional activity of p53. It contains the tetramerization domain, responsible for the formation of a tetrameric complex. The tetramerization domain contains a coiled-coil structure that mediates the formation of homotetramers of p53.

Over 80% of mutations located in the p53 gene in cancerous cells commonly occur in its DNA-binding domain as these mutations produce a protein with a reduced capacity to bind to a specific DNA sequence that regulates the p53 transcriptional pathway(59) followed by the tetramerization domain (3.4%), proline-rich domain (2.3%), the transactivation domain (1%) and the regulatory C-terminal end (0.3%)(60).

2.3.Regulation of p53

In healthy, non-stressed cells, the tumor suppressor protein p53 is maintained at low levels by its primary negative regulators **MDM2** (Mouse Double Minute 2), also known as **HDM2** (Human Double Minute 2) in humans, along with **MDM4**. These regulators interact with p53's transactivation domain (TAD) to inhibit its ability to carry out transcriptional functions. Notably, p53 itself triggers the transcription of MDM2, creating a self-regulating feedback loop that limits its own activity. However, when cells experience stressors such as DNA damage, oncogene-driven hyper-proliferation, oxidative stress, hypoxia, exposure to ultraviolet (UV) radiation, viral infections, reactive oxygen species (ROS), or nutrient deprivation, this prompts the activation of

kinases ATM and ATR that phosphorylate p53 primarily at its amino terminus, disrupting the binding of MDM2 and MDM4 and consequently leading to an increase in p53 levels(61,62).

2.4.Activation of p53

Upon phosphorylation, p53 becomes stabilized and activated, and accumulates in the cell nucleus, where it binds to specific DNA sequences called **p53 response elements (REs)** located in the promoters or enhancers of target genes (**Fig 12)**. A large proportion of direct p53-induced target genes are regulated through p53 binding to proximal p53REs, but recent research demonstrates that the expression of a few p53-regulated genes may be directly influenced by distal p53RE binding events (63–67). This binding initiates the transcription of a broad array of target genes that play pivotal roles in various cellular processes(68).

The activation of p53 is tightly regulated to prevent excessive or prolonged activation, as such uncontrolled activation could potentially disrupt normal cellular processes. Predicting the specific transcriptional response of p53 in various conditions is challenging. Nonetheless, a simplified model has been proposed, dividing p53-inducible genes into two distinct categories: the first category includes genes that are swiftly induced in response to low levels of stress, typically associated with damage that can be repaired; while the second category comprises genes that are activated in response to higher levels of p53 or more severe stress conditions, typically linked to damage that is beyond repair(69).

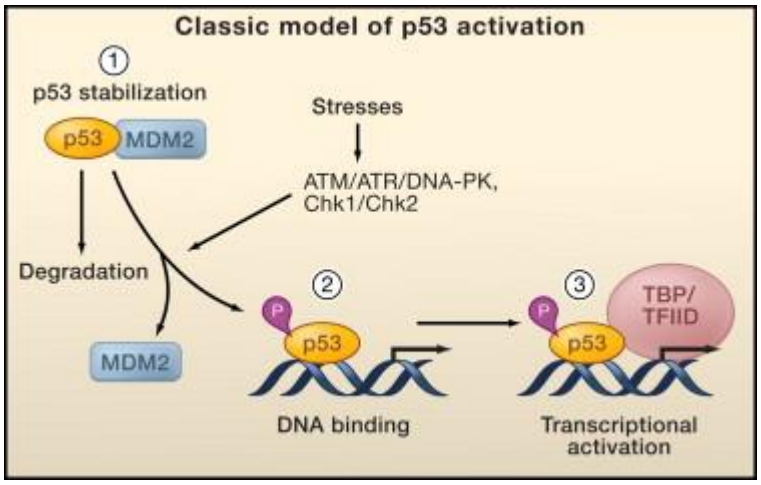


Fig 12. Classical Model of p53 Activation. The classical model for p53 activation generally consists of three sequential activating steps: (1) stress-induced stabilization mediated by phosphorylation (P), (2) DNA binding, and (3) recruitment of the general transcriptional machinery. **Source:**(70)

2.5.Functions of p53

In response to different types or levels of cellular stress, p53 can initiate temporary cell-cycle arrest to facilitate DNA damage repair, trigger cellular senescence or apoptosis, which respectively halt the spread of damage and eliminate compromised cells (71). Beyond its transcriptional functions, p53 also exerts non-transcriptional roles by interacting with other proteins in the cytoplasm or mitochondria. This interaction allows p53 to modulate processes like apoptosis, autophagy, and metabolic pathways. These non-transcriptional activities further underscore the versatile role of p53 in coordinating a wide range of responses to diverse cellular stresses.

A series of meta-analyses has suggested that p53's binding to the genome remains invariant(72) proposing that p53 operates independently to stimulate the expression of a core tumor suppressor network across all types of cells. This conclusion is drawn from observations such as the limited presence of other transcription factor motifs at enhancer sites where p53 is bound and the ability to target a TF binding site (TFBS) within a nucleosome, characteristic for a special class of TFs known as **pioneer factors**(72–74).

While there is no denying the indispensable role of p53-DNA binding in p53's functionality, it has become evident that p53 binding alone does not provide a comprehensive explanation for the observed variations in gene expression. Instead, there is a proposition that the divergent activity of regulatory regions may hold a pivotal position in governing p53-dependent gene expression(75,76). It is worth noting that enhancers and promoters exhibit pronounced cell type-dependent characteristics, underscoring their significance in this context (77,78). Recent research has indicated that in many instances, p53 is already bound to its target promoters without necessarily triggering transcription. Nevertheless, it remains plausible that acetylation may play a significant role in enhancing DNA binding(79). The implications of p53 acetylation underscore the significance of the timing of acetylation events in various regions of p53 for precise p53 regulation and the determination of cellular outcomes. It is highly probable that distinct mechanisms are responsible for activating different categories of p53 targets as part of the stress response process(70).

2.6. Pharmacological inhibitors of p53-MDM2 interaction

Given the critical significance of the interaction between Mdm2 and p53, blocking or inhibiting this interaction is considered a promising avenue with therapeutic potential. One approach for p53 activation involves the use of small molecules that inhibit the MDM2–p53 interaction called Nutlins. Nutlins specifically bind to the p53-binding pocket on the surface of the MDM2 molecule, effectively disrupting the interaction between MDM2 and p53. This disruption leads to the stabilization and activation of wild type p53(80,81). Nutlin-3, the most potent compound among the three existing nutlins, is a racemic mixture of nutlin-3a (active enantiomer) and nutlin-3b (inactive enantiomer). The binding affinity for **nutlin-3a** to MDM2 is 150-fold higher than nutlin-3b(80) activating p53 in cell culture at a concentration of 5–10 μM, inducing cell cycle arrest or apoptosis in cancer cells expressing wild-type p53(82), and inhibiting tumor growth when administered orally at a dose of 200 mg/kg(83).

2.7. The p53 family

The p53 family consists of three homologous proteins—p53, p63, and p73. Experiments with p63 and p73 knockout model mice indicated that p63 is important to epithelial development and p73 plays a profound role during embryonic neuronal development (84).The p63 gene, TP63, is located at chromosome 3q27-26 and contains 265822 nucleotides spanning 14 exons(85)(86). The gene gives rise to a transcription factor with a domain structure like p53: an N-terminal transactivation domain, a central DNA binding domain and a tetramerisation domain. Particularly the DNA binding domain is highly homologous to p53 with 58% identical amino acids. The full-length p63 protein is considerably longer than p53 (641 amino acids) and contains a sterile α motif (SAM) at its C-terminus, which is not present in p53. The p73 gene is located on chromosome 1p36.3 and includes 80728 nucleotides spanning 14 exons. Like p63, the p73 gene also gives rise to a transcription factor highly homologous to p53 with 62% identical amino acids in the DNA binding domain, 18% in the N-terminal transactivation domain and 22% in the tetramerization domain(87).

Aim of the thesis

Although being the most extensively studied protein in history, the molecular mechanisms by which p53 triggers the transcription of hundreds of genes upon cellular stress are not completely understood.

It is the hypothesis of this thesis that p53 dynamically rewires the genome organization to facilitate the formation of chromatin loops to regulate gene transcription and influence crucial cellular processes such as DNA repair or apoptosis.

The specific objectives to test my thesis hypothesis are:

1) Investigate the various levels of chromatin organization after p53 activation including A/B compartments, TADs, and DNA loops to understand how p53 activation influences chromatin organization.

2) Evaluate the impact of p53 activation in the epigenetic landscape to elucidate how regulatory elements modulate gene expression.

3) Explore the impact of p53 activation on gene expression.

Methods

***Disclosure*:** It's important to note that, in my capacity as a bioinformatician, my involvement primarily revolved around the computational aspects rather than direct participation in the experimental procedures.

In this part of the thesis, I will present the workflow executed to conduct the different analyses, accompanied by the commands so the readers can replicate the findings. Nevertheless, the data and scripts responsible for generating the figures presented in the results section are accessible on my GitHub repository: https://github.com/MonicaCabreraP/PhDthesis

To comprehensively explore the implications of p53 activation on 3D genome organization, we exposed the HCT116 cell line—an established model characterized by its wild-type p53 response- to 10μM Nutlin-3a to disrupt the p53-MDM2 interaction and induce p53 activation in response to cellular stress.

We conducted an in situ Hi-C time course experiment, allowing us to construct genome-wide chromosome conformation maps at five distinct time points following p53 activation: 1, 4-, 7-, 10-, and 24-hours post-treatment. We also treated cells with the drug vehicle (i.e., DMSO) and used these cells as a control of basal conditions without p53 activation (also referred to in the thesis as 0 hours of Nutlin-3a) Furthermore, we generated high-quality Promoter Capture Hi-C (PCHi-C) libraries using samples collected at 1 and 10 hours post-Nutlin-3a treatment, as well as control cells treated with the drug vehicle (0 hours), ChIP-seq libraries of H3K4me1 and H3k27ac histone marks, and poly(A)-enriched RNA-seq libraries. For all experiments, we performed 2 biological replicates (**Fig 13**).

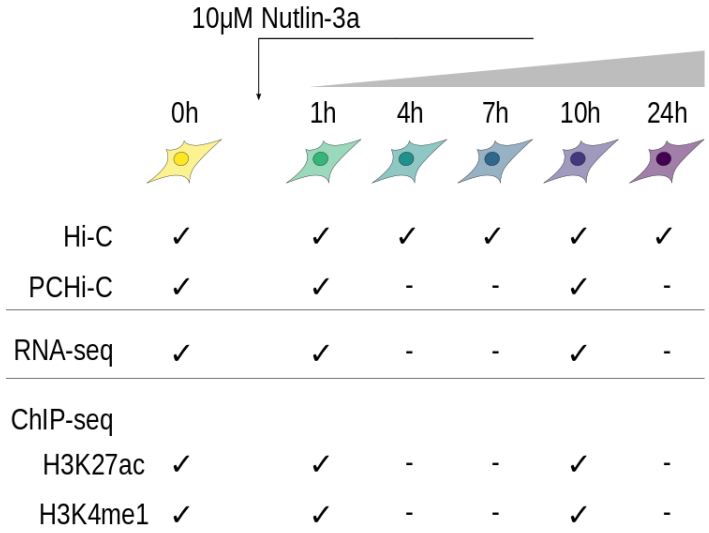


Fig 13. Experimental multi-omics framework of the work. 0h means control cells treated with drug vehicle characterized by no p53 activation.

Experimental analysis

Cell culture

HCT116 were cultured in RPMI1640 medium with Glutamax (Fisher Scientific #61870044) supplemented with 10% FBS (Life Technologies #10270106) and penicillin-streptomycin (Dutscher #L0022-100) in a humid incubator at 37ºC and 5% CO2. To activate p53 response, cells were supplemented with 10µM of Nutlin-3a (Quimigen #HY-10029) dissolved in DMSO and incubated for the required time. Appropriate negative controls for each condition were established by incubating the cells with the drug vehicle (DMSO).

In situ Hi-C and PCHi-C libraries generation

Cells were fixed in DMEM medium supplemented with 10% FBS and 2% methanol-free formaldehyde for 10 minutes rotating at room temperature. After quenching the formaldehyde with 0.125M glycine and washing the cells with 1X PBS, nuclei were extracted by lysing the cells in 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% IGEPAL CA630 (SIGMA-ALDRICH #18896-50ML), 1× complete EDTA-free protease inhibitor cocktail. Chromatin was digested overnight with HindIII enzyme and the cohesive restriction fragment ends were filled-in with dCTT, dTTP, dGTP and biotin-14-dATP (Invitrogen #19524-016) nucleotides. After blunt-end ligating the restriction fragment ends, the chromatin was decrosslinked by incubating overnight with proteinase K and purified by phenol:chloroform:isoamyl alcohol 25:24:1 (SIGMA-ALDRICH #P3803-100ML) extraction.

Non-informative biotin at restriction fragment ends was removed by incubating the samples with dATP and T4 DNA polymerase (New England Biolabs #M0203S). After purifying the DNA again with a phenol:chloroform:isoamyl alcohol extraction, 10µg of the chromatin was sheared using a Covaris M220 focused ultrasonicator in 130µl cuvettes using the following parameters: 20% duty factor, 50 peak incident power, 200 cycles per burst, 65 seconds. The ends were end-repaired and dATP-tailed, followed by a biotin-pulldown using Dynabeads MyOne streptavidin C1 paramagnetic beads (Thermo Fisher #65001) to enrich for those DNA fragments which contain information of a chromatin loop. PE Illumina adapters were ligated to the sample and the library

was amplified for 8 cycles. Finally, the library was purified using a SPRI bead double-sided selection (0.4-1 volumes). Size and concentration of the finished Hi-C libraries was assessed by DNA ScreenTape Analysis (Agilent #5067-5582) on an Agilent 2200 Tapestation.

For PCHi-C libraries, 500-1000ng of Hi-C library were captured using the SureSelectXT Target Enrichment System for the Illumina Platform (Agilent Technologies) as instructed by the manufacturer and the promoter capture system previously published(88). Captured library was amplified a total of 4 cycles, and the size and concentration of the finished PCHi-C libraries was assessed by high-sensitivity DNA ScreenTape Analysis (Agilent #5067-5584) on an Agilent 2200 Tapestation.

ChIP-seq library preparation

Cells were crosslinked in 1X PBS supplemented with 1% methanol-free formaldehyde (Thermo Fisher #28908) for 10 minutes rotating at room temperature. After quenching the formaldehyde with 0.125M glycine for 5 minutes rotating at room temperature. After washing the cells with 1X PBS, pelleted cells were lysed in 1% SDS (AppliChem #A0676,0250), 10mM EDTA (Invitrogen #AM9260G), 50mM Tris-Cl pH 8'1 (Invitrogen #AM9855G) at 4ºC for 20 minutes. Samples were sonicated using a Covaris M220 focused ultrasonicator at a concentration of $20 \cdot 10^6$ cells/ml using the following parameters: 10% duty factor, 75 peak incident power, 200 cycles per burst, 15 minutes. After centrifuging the samples at 14000rpm to remove cell debris and recovering the chromatin-containing supernatant, 33µl of sonicated chromatin were prepared for immunoprecipitation by adding 267µl of buffer containing 1% Triton (AppliChem #A4975,0100), 1'2mM EDTA, 16'7mM Tris pH 8, 167mM NaCl (Invitrogen #AM9760G), 1X cOmplete protease inhibitor cocktail (Merck cat. #11873580001) and the appropriate amount of antibody (1ug of α-H3K27ac and 0.5ug of α-H3K4me1; Diagenode #C15410196 and #C15410194 respectively). Samples were incubated rotating at 4ºC overnight. Chromatin was immunoprecipitated by adding 10µl of both protein A and protein G-conjugated paramagnetic beads (Invitrogen #1001D and #1003D, respectively) and incubated at 4ºC for 1h. After washing the beads, chromatin was decrosslinked overnight at 65ºC using proteinase K (ThermoFisher Scientific

#EO0491) and purified using 1.1 volumes of SPRI beads according to the manufacturer's instructions (CleanNA #CNGS-0050). For ChIP inputs, an equivalent amount of sonicated chromatin was directly decrosslinked and purified as before.

ChIP-seq libraries were performed using the KAPA HyperPrep Kit (Roche #07962363001) according to the manufacturer's instructions using Truseq Illumina adapters and PCR primers described in Supplementary Data 9. Samples were sequenced to reach a minimum number of either 20M or 45M valid paired-read for H3K27ac and H3K4me1 histone marks respectively.

RNA-seq library preparation

RNA was extracted from 200,000 frozen cell pellets using the RNAeasy Mini Kit (Qiagen #74104) following the manufacturer's instructions. Total RNA's integrity and concentration was assessed using RNA ScreenTape Analysis (Agilent #5067-5576) on an Agilent 2200 Tapestation. Samples were sequenced to reach a minimum number of 30M unique valid paired reads.

Computational analysis

Hi-C analysis

The analysis of Hi-C data was performed using the TADbit pipeline(89) with a specific version available at https://github.com/fransua/TADbit/tree/p53_javierre. The steps are the following:

1. **Quality control:** To compute the quality of the experiment the PHRED score and the number of unidentified nucleotides (Ns) in the read sequence were addressed as follows:

```
from pytadbit.utils.fastq_utils import quality_plot

r_enz = 'HindIII'

cell = 'DMSO'
repl = 'BR1'

quality_plot('.fastq'.format(cell,      repl),      r_enz=r_enz,
nreads=1000000)
```

2. **Mapping**: The GEM3 mapper(90) was used to map the reads from the FASTQ files to the GRCh37 reference genome (hg19) obtained from UCSC (http://genome.ucsc.edu).

```
from pytadbit.mapping.full_mapper import full_mapping

cell = 'DMSO'
rep = 'BR1'

! mkdir -p results/iterativ/$cell\_$rep
! mkdir -p results/iterativ/$cell\_$rep/01_mapping

# for the first side of the reads
full_mapping(mapper_index_path='genome/Human_GRCh37_contigs.ge
m',

out_map_dir='results/iterativ/{0}_{1}/01_mapping/mapped_{0}_{1
}_r1/'.format(cell, rep),
fastq_path='FASTQs/%s_%s_1.fastq.dsrc' % (cell,rep),
frag_map=False,  clean=True, nthreads=8,
```

```
windows=((1,25),(1,35),(1,45),(1,55),(1,65),(1,75)),

temp_dir='results/iterativ/{0}_{1}/01_mapping/mapped_{0}_{1}_
```

#Repeat for the second side of the read-end (2.fastq.dsrc)

3. **Read filtering:** Mapped reads were filtered using nine different filters in TADbit, including self-circles, dangling ends, errors, extra dangling-ends, over-represented, too short, too long, duplicated, and random breaks as follows:

```
from pytadbit.mapping.filter import filter_reads

masked = filter_reads(
'results/fragment/{0}_{1}/03_filtering/reads12_{0}_{1}.tsv'.fo
rmat(cell, rep),
    max_molecule_length=750, over_represented=0.005,
max_frag_size=100000,
    min_frag_size=50, re_proximity=5, min_dist_to_re=1000)

from pytadbit.mapping.filter import apply_filter

apply_filter('results/fragment/{0}_{1}/03_filtering/reads12_{0
}_{1}.tsv'.format(cell, rep),
'results/fragment/{0}_{1}/03_filtering/valid_reads12_{0}_{1}.t
sv'.format(cell, rep), masked,filters=[1, 2, 3, 4, 6, 7, 9,
10])

# Save to bam file
from pytadbit.parsers.hic_bam_parser import bed2D_to_BAMhic

bed2D_to_BAMhic('results/fragment/{0}_{1}/03_filtering/valid_r
eads12_{0}_{1}.tsv'.format(cell, rep),valid=True, ncpus=8,

outbam='results/fragment/{0}_{1}/03_filtering/valid_reads12_{0
}_{1}'.format(cell, rep),frmt='mid', masked=None)

hic_data.filter_columns(draw_hist=True, min_count=10,
by_mean=True, silent=True)
```

4. **Hi-C data normalization**: The Hi-C data were normalized using ICE normalization(91)

```
hic_data.normalize_hic(iterations=100, max_dev=0.00001)

hic_data.save_biases('results/fragment/{0}_{1}/03_filtering/va
lid_reads12_{0}_{1}_ICE_{2}kb.biases'.format(cell, repl, reso
// 1000))
```

5. **<u>Interaction matrix generation</u>:** Interaction matrices were generated at two resolutions, 100 kb for A/B compartments and 50 kb for TADs.

- <u>A/B compartment analysis</u>: A/B compartments were determined independently for each chromosome and time-point at a resolution of 100 kb. The ICE-corrected interaction matrices were distance corrected and transformed into Pearson correlation matrices(13). The A compartment was assigned to genomic bins with positive first principal component (PC1), while the B compartment was assigned to genomic bins with negative PC1. The resulting transformed eigenvectors were referred to as compartment scores.

```
hic_data = load_hic_data_from_bam(base_path.format(cell),
resolution=reso,biases=bias_path.format(cell, reso // 1000),
ncpus=8)

corr =
hic_data.find_compartments(show_compartment_labels=True,
show=True, crms=[chrname], vmin='auto', vmax='auto',
rich_in_A=rich_in_A,savedata='results/fragment/{0}_both/05_seg
menting/compartments_{1}_{2}.tsv'.format(cell, chrname, reso),
savedir='results/fragment/{0}_both/05_segmenting/eigenvectors_
{1}_{2}'.format(cell, chrname, reso))
```

- <u>TAD identification</u>: TADs were identified at a resolution of 50 kb. TAD borders were assigned scores ranging from 1 to 10 based on their robustness in the TAD border detection. The strength of a TAD border was calculated by the number of times it appeared in the optimal pathway. TAD borders were also identified using the insulation score(92).

```
from pytadbit import Chromosome
from pytadbit.parsers.hic_parser import load_hic_data_from_bam
base_path =
'results/fragment/{0}_both/03_filtering/valid_reads12_{0}.bam'
bias_path =
'results/fragment/{0}_both/04_normalizing/biases_{0}_both_{1}k
b.biases'
```

```
reso = 100000
cel = 'DMSO'

hic_data =
load_hic_data_from_bam(base_path.format(cel),resolution=reso,r
egion='chr3',biases=bias_path.format(cel1, reso //
1000),ncpus=8)

#Insulation score

hic_data1.normalize_hic()
hic_data1.normalize_expected()

wsize = (1, 4)

insc, delta = insulation_score(hic_dataa, [wsize],
resolution=100000, normalize=True, delta=2)
```

Note: the execution of these steps should be repeated for each of the replicates

PCHi-C processing

1. **Mapping and Filtering:** The raw sequencing reads were processed using HiCUP (0.8.2)(93)**.** There are several pre-steps to perform before running the pipeline:

    a. Genome Digest: To filter out common experimental artifacts, HiCUP requires the positions at which the restriction enzyme used in the protocol cut the genome. In our case we used the restriction enzyme HindIII which target sequence is A^AGCTT. To look for the target sequence in the genome and generate all the restriction fragments for this specific restriction enzyme the hicup_digester function was used as follows:

    ```
    hicup_digester --genome Human_GRCh37 --re1 A^AGCTT,HindIII
    *.fa
    ```

    The digest output file will be labeled as the genome ‘Human_GRCh37’.

b. Bowtie2 Index: The bowtie2-build function was used to perform the genome indexing as follows:

```
bowtie2-build 1.fa,2.fa,...,MT.fa Human_GRCh37
```

c. Run the pipeline: To run HiCUP a config file is needed. A template of it can be obtained using the function `hicup --example` and fill it with the data. Once is filled, HiCUP was run with the following command:

```
hicup --config config_file.txt
```

After running HiCUP we will have a filtered bam file with the valid unique di-tags (paired reads) and a report in html format of all the process.

2. **Capture efficiency:** The reads must be splitted by the ones captured by the capture system and the non-captured reads. To perform this step, we use a miscellaneous script inside HiCUP. To use this script, we need the bam file from HiCUP and a bed file with the coordinates of the captured restriction fragments. For this purpose, a bed file with the coordinates of the captured fragments was generated.

```
perl HICUP/0.8.2/Misc/hicup_capture --baits baits.bed hicup.bam
```

After running this step, a bam file with only those read with at least one end captured is obtained and the percentage of the capture efficiency can be counted to know the efficiency and therefore, the quality of the data.

3. **Downsampling:** In this thesis, the datasets were downsampled to the same sequencing depth as in reference to the sample with less reads using samtools view.

4. **Interaction calling**: Interaction confidence scores were computed using the R package CHiCAGO (1.24.0)(94). Interactions with a CHiCAGO score ≥ 5 were considered high-confidence interactions. To run the pipeline, there are several

files to prepare. To create all these files we can use several scripts (Chicago Tools) available in their bitbucket repository: https://bitbucket.org/chicagoTeam/chicago/src/master/chicagoTools/ , together with well documented information about each file and script usage. Furthemore, the bam file must be edited to a specific format used by Chicago as input, called chinput.

After all the files are prepared, the pipeline was run using the runChicago.R script as follow:

```
Rscript runChicago.R --design-dir DESIGN-DIR <input-files> <output-prefix>
```

5. **Weight Recalibration of Interaction Calling**: Default parameters for interaction calling are calibrated on high-confidence calls from seven human Macrophage data sets (i.e., interactions that pass the p-value threshold in all seven samples). As our data set is from a cancer cell line, we recalibrated these parameters using our control data. To do this, we obtained the weight recalibration numbers running the fitDistCurve function from ChicagoTools as follows:

```
fitDistCurve.R --inputs 1stFile.Rda,2ndFile.Rda…
```

This procedure generates a file called cellType_summaryInput.Rda. This file was then used to re-run the CHiCAGO pipeline adding to the --settings-file the path where the cellType_summaryInput.Rda is.

6. **Peak matrix**: Finally, CHiCAGO results from multiple experiments can be summarized in the form of a 'peak matrix' using makePeakMatrix.R from the chicagoTools.

ChIP-seq analysis

1. **Raw read Quality Control**: Quality control (QC) measures such as assessing read distribution, fragment size estimation, and duplicate read removal to ensure the reliability and accuracy of the subsequent analysis steps was done using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/):

```
fastqc -t 2 $FASTQ1 $FASTQ2 -o $OUTDIR/quality
```

2. **Trimming adapters:** Removal of low-quality reads, adapter sequences and other artifacts was done using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with these parameters:

```
trim_galore $FASTQ1 $FASTQ2 --output_dir $FASTQ_DIR --paired --
basename $NAME -a $ADAPTERS -A $ADAPTERS --fastqc_args --outdir
$OUTDIR/quality --cores $THREADS
```

3. **Alignment to a reference genome:** the alignment algorithm Bowtie(95) was used as followed:

```
bowtie2 -x $INDEX -1 $FASTQ1 -2 $FASTQ2 --very-sensitive -k 2 -
t -p $THREADS -S $OUTDIR/$NAME.sam
```

We obtained a sam file with all the reads aligned. Then we transform the sam to bam and sort it, and we generate an index for the bam file, all these using both samtools and sambamba:

```
samtools view -@ $THREADS -bS $OUTDIR/$NAME.sam >
$OUTDIR/$NAME.bam

sambamba sort -t $THREADS -o $OUTDIR/$NAME.sort.bam
$OUTDIR/$NAME.bam

sambamba index -t $THREADS -p $OUTDIR/$NAME.bam
```

4. **Filtering**: We followed the criteria established by the ENCODE Consortium in their Data Standards: https://docs.google.com/document/d/1lG_Rd7fnYgRpSIqrIfuVlAz2dW1VaSQThzk836Db99c/edit#. There are several steps using samtools and sambamba :

a. Filtering unmapped, mate_is_unmapped, secondary and duplicates, and for paired end read the properly paired:

```
FILT_ARG="'proper_pair and not unmapped and not
mate_is_unmapped and not secondary_alignment and not
duplicate'"

sambamba view -h -t $THREADS -f bam -F $FILT_ARG $BAM >
${PATH_PREFIX}.clean.bam

samtools stats ${PATH_PREFIX}.clean.bam >
${PATH_PREFIX}.2b_filt1.stats

samtools flagstat ${PATH_PREFIX}.clean.bam >
${PATH_PREFIX}.3b_filt1.flagstat
```

b. Filtering Blacklist together with chr MT

```
bedtools intersect -v -abam ${PATH_PREFIX}.clean.bam -b
$BLACKLIST > ${PATH_PREFIX}_filtered.bam

samtools stats ${PATH_PREFIX}_filtered.bam >
${PATH_PREFIX}.2c_blacklist.stats

samtools flagstat ${PATH_PREFIX}_filtered.bam >
${PATH_PREFIX}.3c_blacklist.flagstat
```

c. Fix mate and repeat filtering:

```
sambamba sort -t $THREADS -n -o ${PATH_PREFIX}_namesort.bam
${PATH_PREFIX}_filtered.bam
```

```
samtools fixmate -m ${PATH_PREFIX}_namesort.bam ${PATH_PREFIX}_fixmate.bam

sambamba view -h -t $THREADS -f bam -F $FILT_ARG ${PATH_PREFIX}_fixmate.bam > ${PATH_PREFIX}_fixmate.clean.bam

samtools stats ${PATH_PREFIX}_fixmate.clean.bam > ${PATH_PREFIX}.2d_filt2.stats

samtools flagstat ${PATH_PREFIX}_fixmate.clean.bam > ${PATH_PREFIX}.3d_filt2.flagstat
```

d. Mark duplicates:

```
sambamba sort -t $THREADS -o ${PATH_PREFIX}_fixmate.sort.bam ${PATH_PREFIX}_fixmate.clean.bam

samtools markdup -@ $THREADS ${PATH_PREFIX}_fixmate.sort.bam ${PATH_PREFIX}.markdup.bam
```

e. Report library complexity

```
sambamba sort -n -t $THREADS -o ${PATH_PREFIX}.markdup.sort.bam ${PATH_PREFIX}.markdup.bam

REPORT="bedtools bamtobed -bedpe -i ${PATH_PREFIX}.markdup.sort.bam | awk 'BEGIN{OFS=\"\t\"}{print \$1,\$2,\$4,\$6,\$9,\$10}' | sort | uniq -c | awk 'BEGIN{mt=0;m0=0;m1=0;m2=0} (\$1==1){m1=m1+1} (\$1==2){m2=m2+1} {m0=m0+1} {mt=mt+\$1} END{printf \"%d\t%d\t%d\t%d\t%f\t%f\t%f\n\",mt,m0,m1,m2,m0/mt,m1/m0,m1/m2 }' > ${PATH_PREFIX}.encode.qc"
$REPORT
```

f. Remove duplicates

```
samtools markdup -r -@ $THREADS ${PATH_PREFIX}_fixmate.sort.bam
${PATH_PREFIX}.markdup.bam 2> ${PATH_PREFIX}.4e_markdup.txt

sambamba view -h -t $THREADS -f bam -F $FILT_ARG
${PATH_PREFIX}.markdup.bam | sambamba sort -t 10 /dev/stdin -o
${PATH_PREFIX}.filt.nodup.bam

samtools stats ${PATH_PREFIX}.filt.nodup.bam >
${PATH_PREFIX}.2e_markdup.stats

samtools flagstat ${PATH_PREFIX}.filt.nodup.bam >
${PATH_PREFIX}.3e_markdup.flagstat
```

5. **Visualization of the coverage:** we generated a bigwig file with the profile of our ChIP-seqs with DeepTools (ref) just for visualization purposes.

```
bamCoverage --binSize 10 -b ${PATH_PREFIX}.filt.nodup.bam -o
${PATH_PREFIX}.bw
```

6. **Peak calling:** To identify the genomic regions (peaks) significantly enriched with the protein of interest we used MACS(96). Different variations used in this project are shown below:

```
# Calling peaks in default (narrow) mode for histone mark
H3K27ac.
macs2 callpeak -t $BAM -f $FORMAT -c $BAMINPUT -g $ORG -B -n
${PREFIX} --outdir $OUTDIR/peaks

# Calling peaks in broad mode for histone mark H3K4me1.
macs2 callpeak --broad -t $BAM -f $FORMAT -c $BAMINPUT -g $ORG
-B -n ${PREFIX} --outdir $OUTDIR/peaks

# Calling merge/consensus peaks for one sample with many
replicates
macs2 callpeak -f $FORMAT -t $BAM1 $BAM2 $BAM3  -c $BAMINPUT1
$BAMINPUT2 $BAMINPUT3  -g $ORG -B -n ${PREFIX} --outdir
$OUTDIR/peaks
```

7. **Generation of the peak matrix:** We generated a matrix of all peaks of all our samples to quantify and compare peaks between the samples following these steps:

a. Choose the samples: For this, you need to have a complete vector of paths to your sample's bam files (obtained after the filtering step) and corresponding peak files (obtained after the peak calling step)

```
library(csaw)
library(GenomicRanges)
library(tidyverse)

path_bams <- arg[1] # list of paths to your bam files of interest
path_peaks <- arg[2] # list of paths to your corresponding peak files of interest
```

b. Read in peaks: We read all peak files and generated one object with the complete list of all significant peaks found in all samples of interest as follows:

```
peaks <- GRanges()
for (i in peak_files)
{
  a <- data.table::fread(i)[,1:4]
 aGR <- makeGRangesFromDataFrame(a, seqnames.field = "V1",
start.field = "V2", end.field = "V3", keep.extra.columns = T)
  peaks <- c(peaks, aGR)
}
colnames(elementMetadata(peaks)) <- "name"
peaks$ID <- 1:length(peaks)
```

c. Disjoin peaks: If two peaks overlap, they are split into three regions, with region 1 corresponding to the region unique to peak1, regions 2 being the region where both peaks overlap, and region 2 corresponding to the region unique to peak 3.

```
peak_disj <- disjoin(peaks,with.revmap=T)
```

d. Count reads in regions: Here we quantified the number of reads falling in each region per bam file (per sample)

```
param <- readParam()

peak_count <- regionCounts(bam.files = bam_files, regions = peak_disj, ext = NA, param = param, BPPARAM = BatchtoolsParam(workers = length(bam_files)*2))
```

e. Format and save the peak matrix: We formatted the resulting quantification and save it as the peak matrix where each row reflects a region and each column represents a bam file (sample).

```
peak_df <- as.data.frame(assay(peak_count))
colnames(peak_df) <- colData(peak_count)$bam.files
peak_df$region <- paste(seqnames(rowRanges(peak_count)),ranges(rowRanges(peak_count)), sep = ":")

data.table::fwrite(peak_df[,c(ncol(peak_df),1:(ncol(peak_df)-1))], file=paste0(out_dir,"/peak_matrix_counts.tsv"), row.names = F, col.names = T, quote = F, sep = "\t")
```

8. Generating Background Matrix

Repeat steps 7a-c first.

d. Count background reads: to quantify background noise in 10kb bins.

```
param <- readParam(discard=peak_disj) # discard disjoined peaks from background counting

bg_count <- windowCounts(bam.files = bam_files, width = 10000, spacing = 10000, ext = NA, filter = 0, param = param, BPPARAM = BatchtoolsParam(workers = length(bam_files)*2)) # count background noise
```

e. Format and save background matrix: We formatted the resulting background quantification and save it as a background matrix, where each row reflects a 10kb bin and each column represents a bam file (sample).

```
bg_df <- as.data.frame(assay(bg_count))
colnames(bg_df) <- colData(bg_count)$bam.files
bg_df$region <-
paste(seqnames(rowRanges(bg_count)),ranges(rowRanges(bg_count)
), sep = ":")

data.table::fwrite(bg_df[,c(ncol(bg_df),1:(ncol(bg_df)-1))],
file=paste0(out_dir,"/background_counts.tsv"), row.names = F,
col.names = T, quote = F, sep = "\t")
```

9. **Differential Analysis:** The output is a table with regions found with a significant differential deposition of a histone mark (gain or loss). Steps:

a. Estimate size factors of background matrix: Size factors (normalization factors) of background matrix must be computed in order to perform the background correction to our peak matrix for proper normalization before differential analysis.

```
library(DESeq2)

# read background matrix
background_matrix <-
read.table("path/to/background_matrix.txt",header=T)
colnames(background_matrix) <- c("regions","Sample1_1",
"Sample1_2", "Sample2_1", "Sample2_2")
rownames(background_matrix) <- background_matrix$regions
background_matrix$regions <- NULL
# prepare metadata
coldata <- data.frame(id=c("Sample1_1", "Sample1_2",
"Sample2_1", "Sample2_2") , tissue=c("Sample1", "Sample1",
"Sample2", "Sample2"), rep=c(1, 2, 1, 2))
```

```
rownames(coldata) <- coldata$id
dds <- DESeqDataSetFromMatrix(countData = background_matrix,
colData = coldata, design = ~ tissue)

# estimate size factors
dds <- estimateSizeFactors(dds)
sf <- sizeFactors(dds)
```

b. Background correction: Correct peak matrix for background signal using the background matrix's size factors.

```
count_matrix <- read.table("path/to/peak_matrix.txt",header=T)
colnames(count_matrix) <- c("regions","Sample1_1", "Sample1_2",
"Sample2_1", "Sample2_2")
rownames(count_matrix) <- count_matrix$regions
count_matrix$regions <- NULL

dds <- DESeqDataSetFromMatrix(countData = count_matrix,
                              colData = coldata,
                              design = ~ tissue)

# Apply pre-compute size factors with background regions -
background correction
sizeFactors(dds) <- sf
```

c. Differential Analysis:

```
dds <- DESeq(dds)
res <- results(dds,contrast=c("tissue","Sample1","Sample2"))
```

RNA-seq analysis

1. **Quality and trimming**: This two steps were performed as in the ChIP-seq analysis using FASTQC and Trim Galore!.

2. **Alignment:** We aligned the reads to the reference genome using STAR(97) as follows:

```
STAR --runThreadN $THREADS --genomeDir $INDICES --readFilesIn
$FASTQ1 $FASTQ2 --outFileNamePrefix $OUTNAME --outFilterType
BySJout --readFilesCommand zcat --outFilterMultimapNmax 20 --
alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --
outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04
--alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax
1000000 --outStd SAM | samtools view -bS - > $BAMOUT
```

3. **Counting**: We quantified the reads falling in each gene using its GTF file:

a. Sorting the BAM file:

```
sambamba sort -t 15 -o RNA_1.sort.bam _RNA_1.bam
```

b. Count reads: The featureCounts tool (98) was employed to count the number of reads that map to each gene or transcript, providing quantitative expression data.

```
featureCounts -a $GTF -p -B -C -T 15 -o $OUTFILE.csv $BAM1 $BAM2
```

4. **Differential Analysis:**

a. Create count matrix

```
files <- list.files(path = "/path/to/featureCounts/out", pattern
= "*.csv$", recursive = T, full.names = T)
```

```
countdata <- matrix(data.table::fread(files[1])$Geneid)
colnames(countdata) <- "Geneid"
for (i in files)
{
  a <- data.table::fread(i)
  a <- a[ ,-(2:6)]
  colnames(a)[-1]        <-        gsub("\\.[sb]am$",        "",
basename(colnames(a)[-1]))
  a <- as.matrix(a)
  countdata <- merge(countdata,a, by="Geneid")
}

countdata[,-1] <- apply(countdata[,-1], 2, as.numeric)

data.table::fwrite(countdata,              file              =
"count_table.txt",col.names = T, row.names = F, quote = F, sep
= "\t")
```

5. Differential Expression Analysis:

```
count_matrix <- read.table("path/to/count_matrix.txt",header=T)
colnames(count_matrix) <- c("geneID","Sample1_1", "Sample1_2",
"Sample2_1", "Sample2_2")
rownames(count_matrix) <- count_matrix$geneID
count_matrix$geneID <- NULL

dds <- DESeqDataSetFromMatrix(countData = count_matrix, colData
= coldata,design = ~ tissue)
dds <- estimateSizeFactors(dds)
dds <- DESeq(dds)
res <- results(dds,contrast=c("tissue","Sample1","Sample2"))
```

Definition of functional p53 binding sites

As p53's chromatin binding pattern remains remarkably consistent across diverse cell types and stimuli (76), it is essential to recognize that not every binding translates into subsequent transactivation. This underscores the significance of discerning which binding events bear functional importance. Within this context, in this thesis we classified p53 bindings via their overlap with enriched regions of H3K27ac and defined a functional subset of p53 binding sites. A p53 binding site is deemed functional when it overlaps with the presence of H3K27ac mark in the time points corresponding to p53 activation after 1 or 10 hours of Nutlin-3a treatment. This approach ensures that the selected binding sites are more likely to possess regulatory roles and enables a targeted comparison of the cis-regulatory elements across the three time points.

The p53 ChIP-seq used is a publicly available ChIP-seq data for p53 in the HCT116 cell line from a study with the accession number GSE86164(73). To extract time-point-specific functional p53 binding sites, we refined this dataset by intersecting it with our H3K27ac consensus peaks using the R package GenomicRanges (version 1.50.2) (99).

Definition of promoters and enhancers

Gene promoters were defined as regions spanning 1,000 base pairs upstream and 200 base pairs downstream of their transcriptional start site. The Ensembl gene annotation GRCh37 release 87 was employed to determine the transcriptional start sites. To identify enhancers, we initially focused on the intersection between consensus H3K4me1 and H3K27ac peaks at three time points: Nut 0h, Nut 1h, and Nut 10h. Subsequently, we delineated enhancer activity by annotating the presence of H3K4me1 and H3K27ac peaks for each time point. Specifically, the presence of only H3K4me1 indicated a primed enhancer state, while the presence of both H3K4me1 and H3K27ac denoted an active enhancer state.

Statistical methods

HiCRep algorithm(100) was implemented in TADbit to compute the stratum-adjusted correlation coefficient (SCC) between the two replicates.

PCA of TAD borders and compartments was performed using the scikit-learn API v1.1.3 (101). Statistical tests and complex numerical treatments were performed with SciPy v1.10.1 (102) and NumPy v.1.24.2 (103). PCA for RNA-seq, ChIP-seq, and PCHi-C samples were generated using prcomp function from stats (4.2.1) R package **(R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.).**

Correlation analyses were done using ggpubr (0.6.0) **(package version 0.6.0, <https://CRAN.R-project.org/package=ggpubr>.)**. Data manipulation and complex numerical treatments were performed using tidyverse (1.3.2) (**DOI: 10.21105/joss.01686**).

Means were compared using a two-sided Wilcoxon's rank sum test. Data manipulation and integration with PCHi-C data sets were performed using HiCaptuRe (https://zenodo.org/badge/latestdoi/488882077).

Results

Quality control assessment: High quality libraries

As described in the methods sections, all computational analysis had a quality control step at the beginning to assess the library quality.

Prior to conducting genome segmentation analysis, an assessment of reproducibility between biological replicates demonstrated a high level of consistency, indicating a good reproducibility of our Hi-C experiment (**Fig.14**). This consistency justifies the merging of biological replicates for downstream analysis.

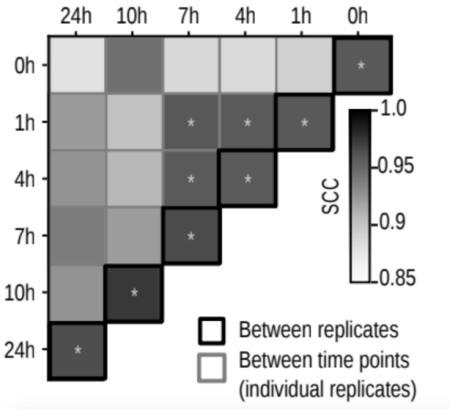


Fig 14. Stratum Adjusted Correlation Coefficient (SCC). Similarity between Hi-C interaction matrices calculated using the SCC value; 1 means perfect correlation, values above 0.95 (represented by a white star) are expected for biological replicates.

The **Hi-C** libraries successfully passed all quality control metrics, confirming the acquisition of high-quality data and ensuring reliable results (**Table 1**).

	Nut 0h	Nut 1h	Nut 4h	Nut 7h	Nut 10h	Nut 24h
Self-circle	4.828.613	1.302.097	1.225.115	1.821.086	3.944.101	1.128.166
Dangling-end	13.468.542	2.916.999	3.635.722	4.087.437	15.073.344	3.859.660
Error	3.515.613	4.095.813	4.076.862	5.370.533	3.280.623	3.689.919
Extra dangling-end	99.689.992	71.792.219	70.807.804	97.252.310	101.343.586	69.542.093
Too close from RES	134.058.783	84.120.583	82.039.771	115.923.800	155.404.384	82.166.782
Too short	5.889.809	3.639.020	3.577.031	4.998.276	6.235.726	3.837.703
Too large	129.057	157.091	151.235	209.138	156.168	156.252
Duplicated	37.071.570	38.846.937	39.287.149	62.942.034	44.210.476	36.995.913
Random breaks	8.661.093	2.644.983	2.953.859	3.531.457	9.532.653	3.382.397
Valid-pairs	275.752.742	178.866.432	171.686.675	237.889.042	290.240.484	163.550.288
Valid-pairs cis-close	60.389.715	21.775.777	20.863.052	29.985.442	53.929.544	19.361.406
Valid-pairs cis-far	196.925.331	137.555.415	133.099.889	182.453.829	211.463.277	125.767.543
Valid-pairs_trans	18.437.696	19.535.240	17.723.734	25.449.771	24.847.663	18.421.339

Table 1. TADbit quality control parameters. Columns indicate the samples, rows indicate the filtered reads quantified from “Self-circles” to “Random breaks”, and the number of reads that remained after filtering (“Valid pairs”).

The **PCHi-C** libraries successfully passed all quality control metrics, confirming the acquisition of high-quality data and ensuring the reliability and accuracy of subsequent analysis. Summary of the key number from all samples is listed in **Table 2**.

	Biological replicates	Processed reads	Capture Unique di-tags
Nut0h	BR1	116.906.405	70.289.872
	BR2	115.216.237	96.428.391
Nut1h	BR1	167.735.584	39.507.208
	BR2	149.076.169	36.829.919
Nut10h	BR1	138.784.892	104.193.388
	BR2	123.785.344	90.251.467

Table 2. Key characteristics of PCHi-C samples. Processed reads are raw reads before filtering. Unique di-tags are read pairs after mapping and filtering.

Prior to conducting interaction calling, biological replicates were merged as due to the nature of the data particularly at longer distance ranges, may lead to significant sensitivity issues when using thresholded interactions calls in comparative analysis (94). **Fig 15** shows the number of interactions among all biological replicates above a CHiCAGO score ≥ 5 reinforcing the merging of samples.

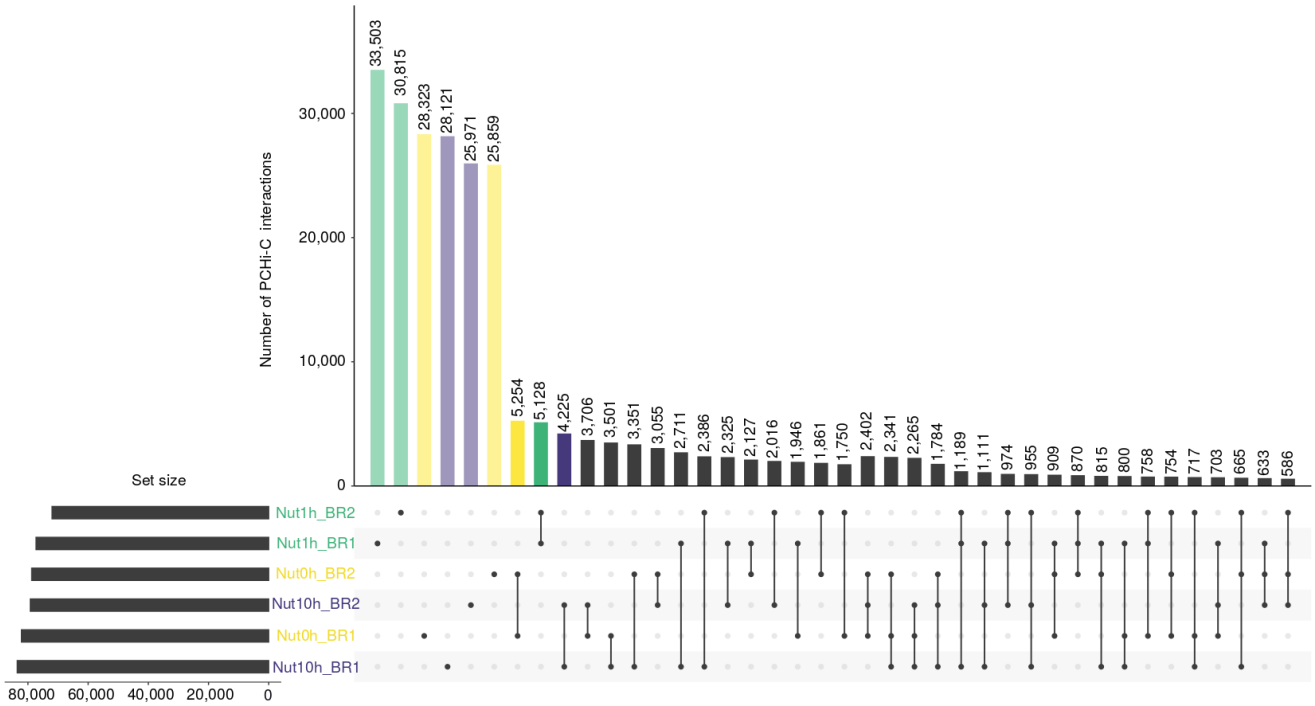


Fig 15. PCHi-C replicability. Upset plot showing the number of shared interactions between different time-point and replicates.

Both **ChIP-seq** libraries H3K4me1 (**Table 3**) and H3K27ac (**Table 4**) demonstrated high-quality libraries:

H3K4me1	Biological replicates	Total Reads	Dup %	FragL	N peaks
Nut0h	BR1	46.799.878	0,00	233.6	91.730
	BR2	70.732.722	0,00	234.6	86.207
Nut1h	BR1	126.818.792	0,00	264.2	92.943
	BR2	82.963.955	0,00	273.1	86.750
Nut10h	BR1	51.691.545	0,00	231.3	84.492
	BR2	60.578.238	0,00	240.4	89.231

Table 3. Summary of ChIP-seq filtering and quality metrics for H3K4me histone mark. Dup% represents the percentage of duplicated reads in each sample, FragL indicates the average fragment length observed, and N peaks refers to the number of final peaks identified after filtering and peak calling.

H3K27ac	Biological replicates	Total Reads	Dup %	FragL	N peaks
Nut0h	BR1	65.536.634	0	225.8	49.011
	BR2	46.329.256	0	234.4	48.253
Nut1h	BR1	43.399.072	0	252.9	45.759
	BR2	49.659.129	0	279.6	43.236
Nut10h	BR1	63.432.073	0	236.9	49.127
	BR2	63.917.155	0	233.6	46.638

Table 4. **Summary of ChIP-seq filtering and quality metrics for H3K27ac histone mark.**

The **RNA-seq** libraries successfully passed all quality control confirmed high-quality of the libraries, demonstrating reliable and reproducible results (**Table 5**).

	Biological replicates	Raw reads	Uniquely mapped	Counts
Nut0h	BR1	40.853.593	39.115.581	35.473.485
	BR2	40.722.423	39.014.296	35.141.106
Nut1h	BR1	38.869.311	36.619.441	31.731.267
	BR2	31.269.612	29.619.401	26.068.076
Nut10h	BR1	40.877.956	38.713.920	34.977.386
	BR2	41.005.490	38.863.127	34.972.083

Table 5. Summary of RNA-seq filtering and quality metrics.

Results I: p53 activation alters 3D chromatin structure

1.1. A/B compartment distribution undergoes reorganization during p53 activation

By examining the first eigenvector values derived from PCA, the genome was effectively classified into distinct A and B compartments, with positive eigenvector values serving as indicators of active compartment (A compartment) and negative eigenvector as inactive compartment (B compartment). The two-dimensional clustering plot based on PC1 and PC2 scores revealed distinct clustering patterns and recapitulates the p53 activation trajectory, identifying three primary groups: Group X, Group Y, and Group Z. Group X consisted of Hi-C datasets with early p53 activation (1h, 4h, and 7h). Group Y included datasets prior to p53 activation (0h) and after 10 hours. Group Z comprised datasets at 24 hours post-activation of p53, exhibiting distinct genomic characteristics and spatial separation from the other clusters (**Fig.16A**). Hierarchical clustering analysis confirmed the observed grouping. The dendrogram demonstrated clear separation between clusters X and clusters Y and Z at a higher dissimilarity level, indicating their distinct compartment genomic characteristics (**Fig.16B**).

Overall, the genome was split into ~60% A compartment and ~40% B compartment at each time point (**Fig.16C**). Most compartment assignments were invariant during differentiation, however, p53 activation induced compartmental switches in 12.46% of the genome (**Fig.16D**). Clustering analysis identified seven primary clusters representing switching compartments in response to p53 activation (**Fig.16E**). Notably, 23.21% of the genome underwent decompaction, exhibiting a transitory switch from B to A compartment at early stages, and then switching back to B (C1 and C2 **Fig.16E**). Conversely, 39,5% of the genome experienced compaction, shifting unidirectionally from A to B compartment one time at late stages (C3, C4, and C5 **Fig.16E**). The other 2 clusters switched compartments more than once during p53 activation.

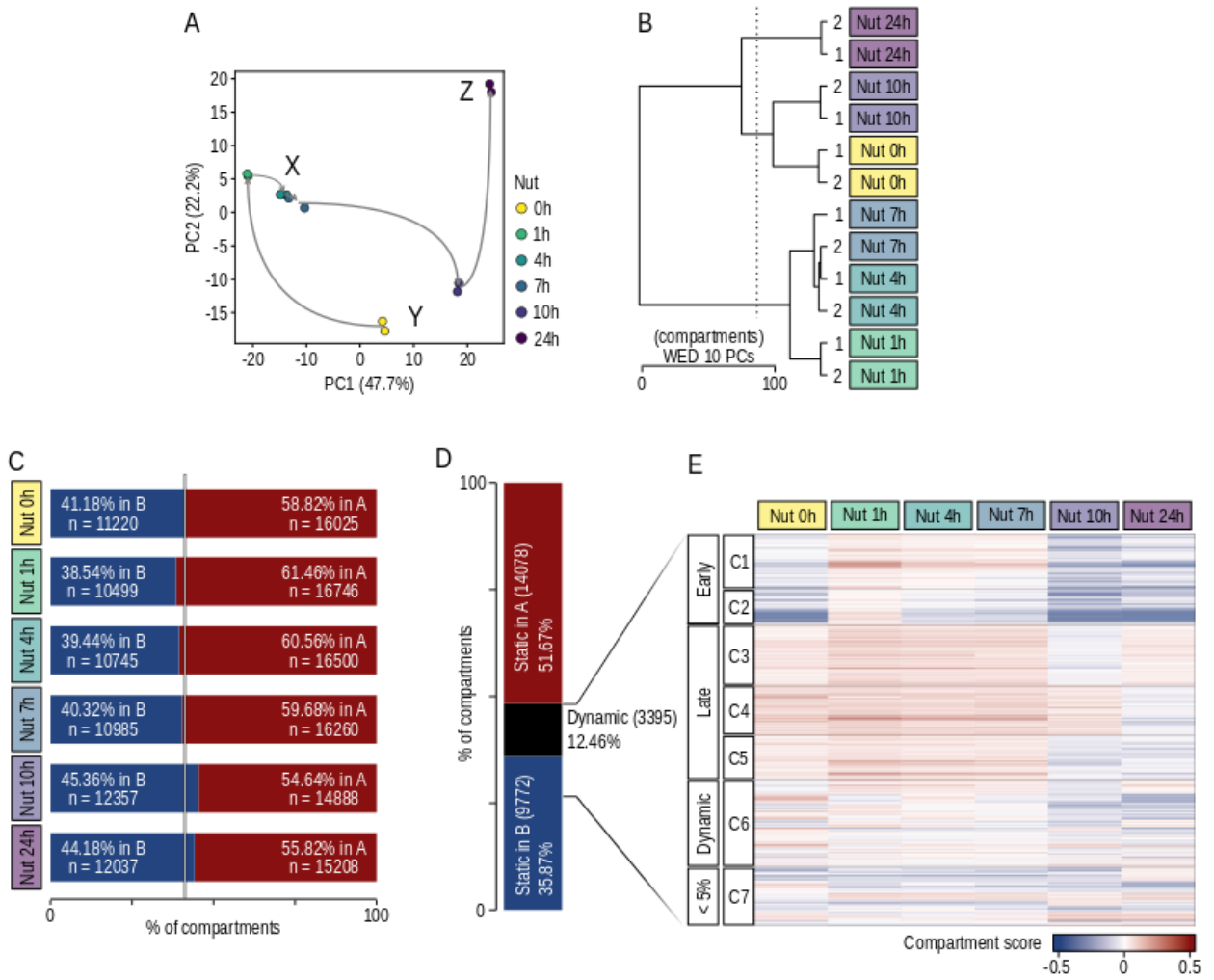


Fig 16. A/B compartment dynamism along p53 activation. A) Principal Component Analysis of the compartment scores for each Hi-C individual biological replicate across p53 activation. Principal components (PC) 1 and 2 were taken into consideration. Numbers in parentheses represent the percentage of variance explained by each PC. **B)** Hierarchical clustering analysis (Ward criterion) of compartment values throughout p53 activation reflecting the degree of dissimilarity between biological replicates of different samples. The distances between replicates were measured by applying weighted Euclidean distance (WED) of the 10 first principal components. **C)** Percentage of 100kb genomic regions defined as A or B compartments for each time point. **D)** Percentage of compartment categories according to their dynamics along p53 activation. Dynamic compartments are defined as those that change compartment between any two time points. **E)** Compartment scores of the dynamic compartments in response to p53 activation, with each row representing a 100kb bin and each column representing a time point. The heatmap color scale represents the compartment score assigned to each 100kb bin with positive values indicating A compartment and negative values indicating B compartment.

Additionally, statistically significant positive associations were observed for the A compartment in samples with early p53 activation at 1h, 4h, and 7h (represented in shades of red **Fig.17A column A**), while statistically significant negative associations were observed for the A compartment in samples with late p53 activation at 10h and 24h (represented in shades of blue **Fig.17A column compartment A**) and vice versa for the B compartment (**Fig.17A column compartment B**). **Fig. 17B** shows the compartment scores for a region of chromosome 12 with a discernible switch from B to A compartment at early p53 activation stages.

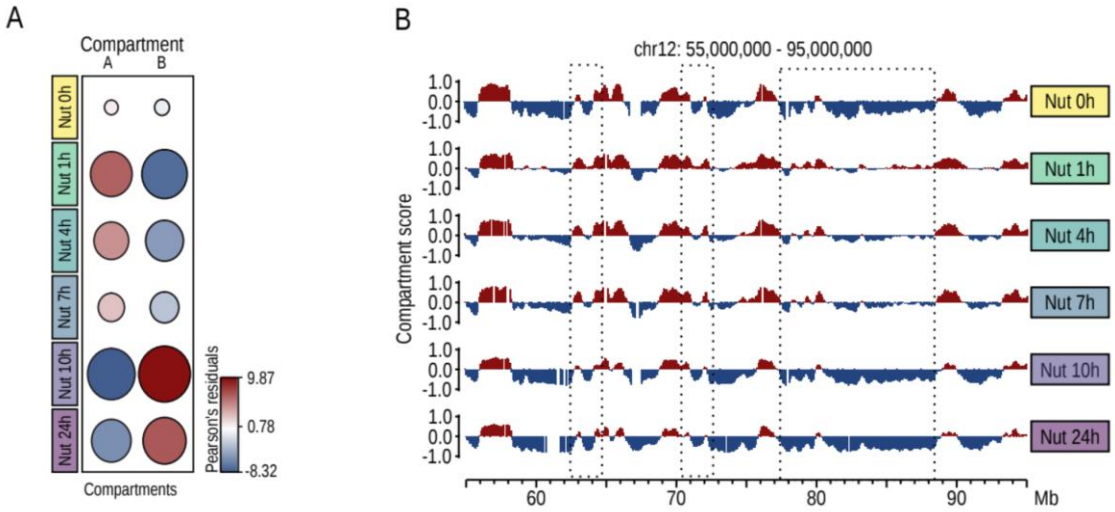


Fig 17. **A/B compartment changes due to p53 activation. A)** Correlation plot resultant from Pearson's residuals Chi-square test of independence for the contingency table time-course *vs.* compartment category (X-squared = 410.52, df = 5, p-value < 2.2e-16). Positive residuals are red, suggesting a positive association between the corresponding row and column and negative residuals are blue, suggesting a negative association. For a given cell, the size of the circle is proportional to the amount of the cell contribution. **B)** Variation of eigenvector values along a 50 Mb genomic region of chromosome 12. The genomic bins are colored according to their compartment type (A compartment red and B compartment blue). Discontinued dots highlight regions transitioning from B to A due to early p53 activation (1h).

1.2. TAD borders are dynamically remodelled during p53 activation

To analyze the TAD borders we used two methods: insulation score and TADbit score(89,92). PCA and hierarchical clustering of TADbit scores revealed remarkable reproducibility between biological replicates and demonstrated progressive changes in TAD borders that correlated with p53 activation (**Fig 18 A and B**) like the groups observed previously with the compartments.

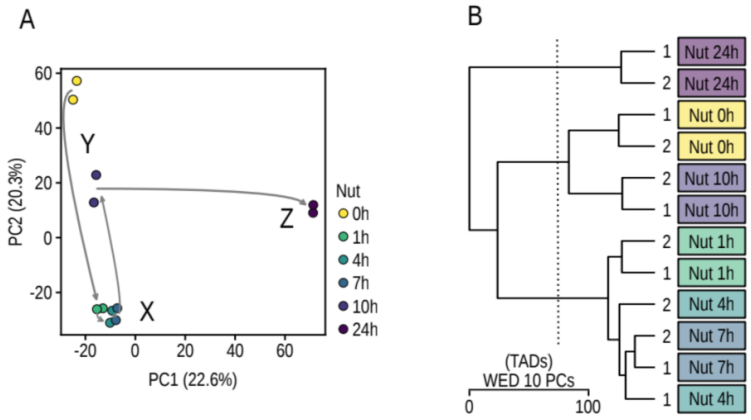


Fig 18. TAD dynamics along p53 activation. A) Principal Component Analysis of TADbit TAD border scores for all Hi-C biological replicates across p53 activation. Principal components 1 and 2 were taken into consideration. Numbers in parentheses represent the percentage of variance explained by each principal component. **B)** Hierarchical clustering analysis (Ward criterion) of TADbit TAD border scores across p53 activation reflecting the degree of dissimilarity between biological replicates of different samples. The distances between replicates were measured by applying weighted Euclidean distance (WED) of the 10 first principal components.

Using a stringent cutoff of the TADbit score >4, we identified between 3963 and 2820 TAD borders (**Fig 19A**). Among the identified TAD borders, a subset of 2091 remained stable across all stages examined, indicating a conserved spatial organization within the genome. However, a significant number of TAD borders exhibited dynamic changes in response to p53 activation. Upon 1 hour of p53 activation, a striking loss of TAD borders was observed, coinciding with an enlargement of the TAD size. The mean size of TADs ranged from 494 to 672kb (**Fig 19B**) suggesting a disruption in the organization of TADs. This disruption may reflect a re-configuration of chromatin

interactions, allowing for more extensive interaction within expanded regions. Interestingly, the loss of TAD borders continued until 10 hours of p53 activation, characterized by a pronounced acquisition of the same TAD borders before p53 activation (**Fig 19 C cluster C4**). However, at 24 hours of treatment, a global erasure of TAD borders was observed. **Figure 19 D** illustrates a chromosome 1 region displaying a TAD border loss caused by early p53 activation, followed by a subsequent regain of the TAD borders prior to p53 activation.

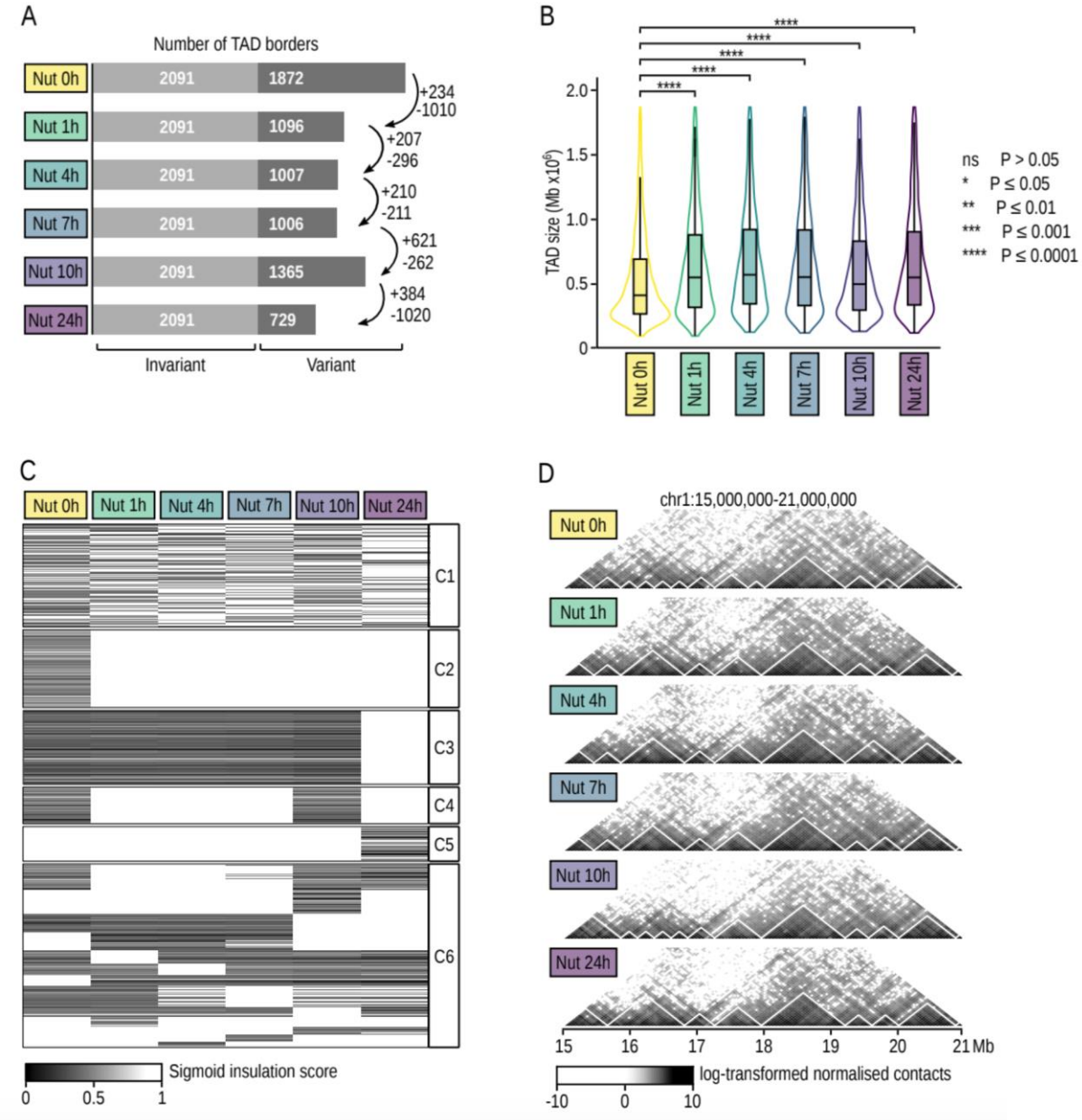


Fig 19. TAD changes along p53 activation. A) Bar plot displaying the number of variant and invariant TAD borders along p53 activation. Invariant TAD borders are those detected at all time points. Positive/negative numbers near arrows represent the number of TAD borders gained/lost between the corresponding time points. **B)** Distribution of TAD sizes along p53

activation. A two-sided Wilcoxon test was used to test whether distributions of TAD sizes differed throughout p53 activation compared to control. Nut 0h refers to control cells without p53 activation. **C)** Clustering of TAD borders based on their TAD insulation scores and the patterns of temporal changes observed throughout p53 activation. Only TAD borders characterized by a TADbit score > 4 were included. TAD borders were manually clustered considering their dynamism throughout p53 activation. **D)** Hi-C contact maps of a 6Mb genomic region of chromosome 1 where TAD is marked by white lines.

1.3. Chromatin loops are rewired during p53 activation

Since Hi-C results revealed two clear waves of dynamism at 1 hour and 10 hours after p53 activation with Nutlin-3a, to further explore these dynamics at the chromatin loop scale, we generated two biological replicates of high quality and reproducible PCHi-C libraries of the 3 time points: 0h, 1h and 10h.

Promoter interactomes were highly reproducible within each biological replicate (**Fig 20A**). In line with prior Hi-C results, correlation analysis and hierarchical clustering analysis reveals a clear partitioning of the samples into two distinct clusters where the 1-hour activation of p53 exhibits substantial dissimilarities in the promoter interactome compared to the 0-hour and 10-hour time points. (**Fig 20B-C**).

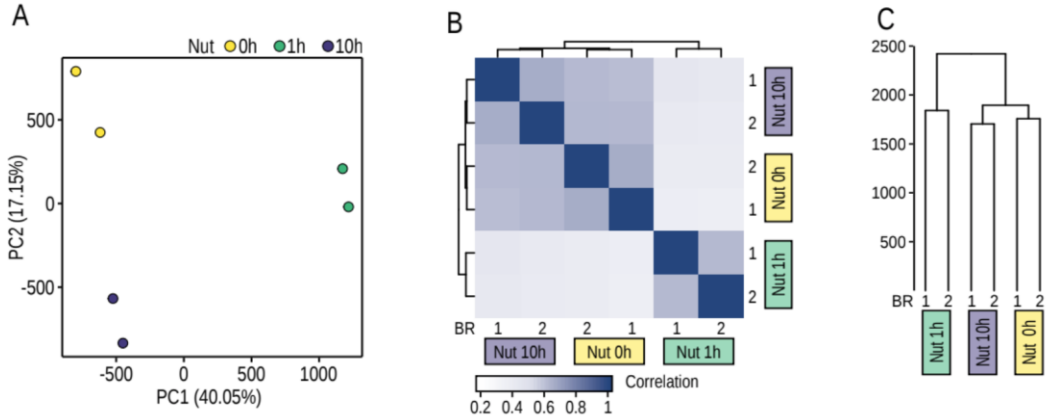


Fig 20. PCHi-C dynamism along p53 activation. A) Principal Component Analysis of promoter interactomes throughout p53 activation obtained by PCHi-C at the biological replicate level. Numbers in parentheses represent the percentage of variance explained by each principal component. **B)** Pearson Correlation between promoter interactomes obtained by PCHi-C of biological replicates at Nut 0h, Nut 1h and Nut 10h time points. Samples were clustered hierarchically inputting Euclidean distances to the “complete” agglomeration method. **C)** Hierarchical clustering of biological replicates (1 and 2) of promoter interactomes

obtained by PCHi-C at Nut 0h, Nut 1h and Nut 10h time points. Euclidean distances were measured between samples, and the “complete” agglomeration method was used.

We detected an average of 78,832 significant interactions per sample, and an average of 83% of promoter-to-non-promoter interactions (P-OE), underscoring the prevalence of this specific type of interaction in the dataset (**Fig 21 A-B**). Detailed number of interactions per sample can be found in **Table 6**.

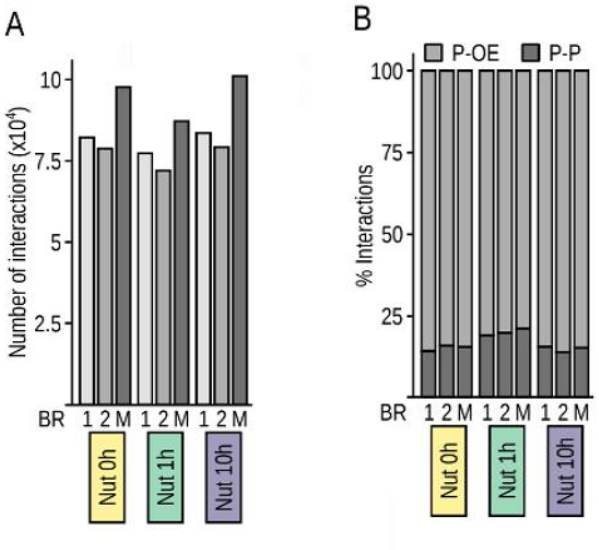


Fig 21. PCHi-C interactions statistics. A) Total number of significant interactions (CHICAGO score ≥ 5) of biological replicates (1 and 2) and merged samples (M) **B)** Percentage of promoter-promoter (P-P) and promoter-other end (P-OE) interactions of biological replicates (1 and 2) and merged samples (M).

	Biological replicates	Significant interactions	P-P interactions	P-OE interactions
Nut0h	BR1	82.165	11.486	70.679
	BR2	78.751	12.327	66.424
	merged	97.691	14.884	82.807
Nut1h	BR1	77.310	14.607	62.703
	BR2	71.994	14.083	57.911
	merged	87.160	18.285	68.875
Nut10h	BR1	83.562	12.822	70.740
	BR2	79.207	10.897	68.310
	merged	101.076	15.266	85.810

Table 6. Key characteristics of PCHi-C samples. Significant interactions are significantly interacting restriction fragments. Promoter-promoter are interactions between baited fragments. Promoter-other are interactions between a baited and a non-baited fragment.

Most of the interactions are specific of each time point, although there is a 16,8% of the interactions maintained along the three time points (**Fig 22A**). The remaining interactions were either newly formed or lost as a result of p53 activation at 1h or 10h and can be clustered as: gained at 1 hour (C1), gained at 10 hours (C2), common to

both time points (C7), lost during early (C3-4) and late (C6) p53 activation, and lost interactions re-established at the 10-hour mark of Nutlin-3a treatment (C4) (**Fig 22B**). Early activation of p53 resulted in a loss of interactions exceeding 200Kb (**Fig 22C**). Conversely, late activation of p53 led to an increase in long interactions, specifically those exceeding 300Kb (**Fig 22C**).

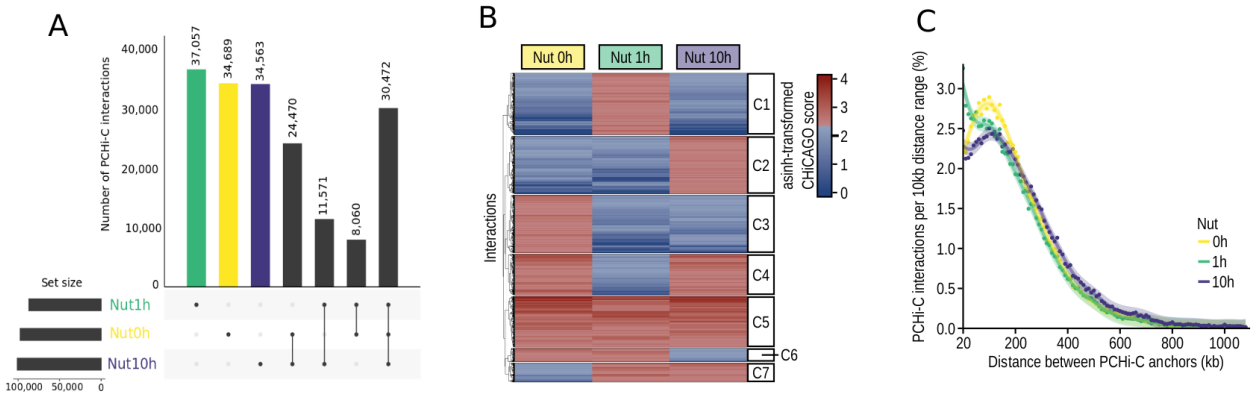


Fig 22. Promoter interactions shift during p53 activation. A) Upset plot illustrating the count of specific interactions per sample, as well as the shared interactions among two or all samples. **B)** Heatmap of asinh-transformed CHiCAGO scores. Only interactions found significant in at least one time point were considered. Clusters were defined based on presence or absence of interaction significance in each time point, and hierarchical clustering was performed withing each predefined cluster to order the interactions order the clusters from C1-C7. Euclidean distances were measured between interactions and clusters, and the "complete" agglomeration method was used. **C)** Observed number of interactions with CHICAGO score ≥ 5 ranked according to the genomic distance between their anchor points. Distances were grouped into 10kb bins and converted to percentages to facilitate the comparison between samples.

Based on the chi-squared test, we determined that the observed association between distances and p53 activation time points were statistically significant (**Fig 23A**). Statistically significant positive associations for short-distance interactions at 1 hour of activation, as indicated by shades of red in **Fig 23A column Nut 1h** and statistically significant negative associations were observed for short-distance interactions in samples at 0 hours and 10 hours, represented by shades of blue in **Fig 23A column Nut 0h and Nut 10h**. The opposite pattern was observed for long-distance interactions. **Fig 23B** provides an example of PCHi-C interactions, showcasing the dynamic changes in the promoter interactome resulting from p53 activation.

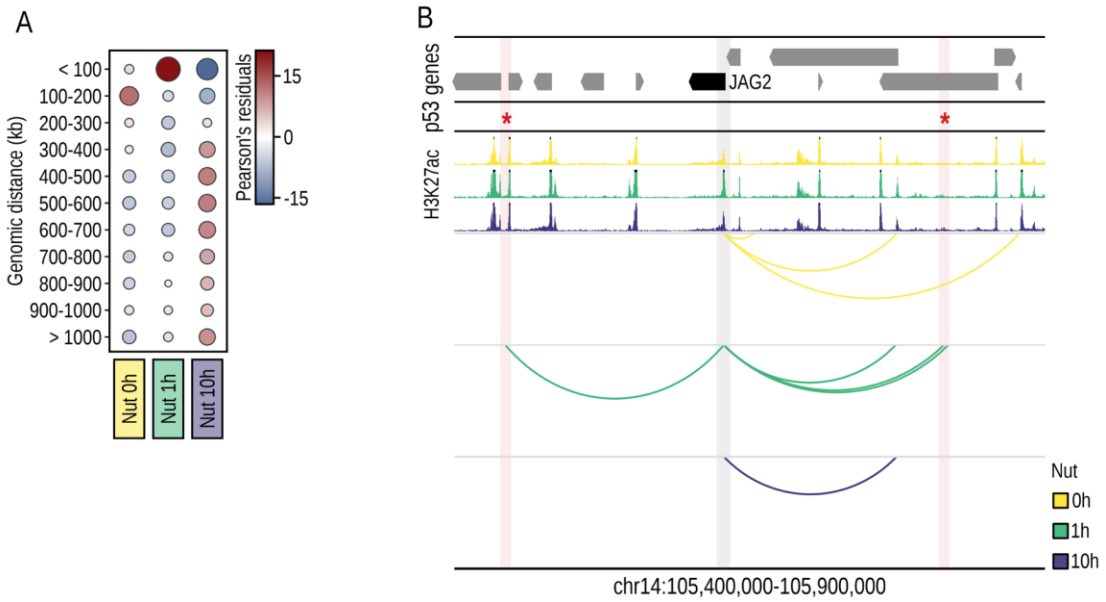


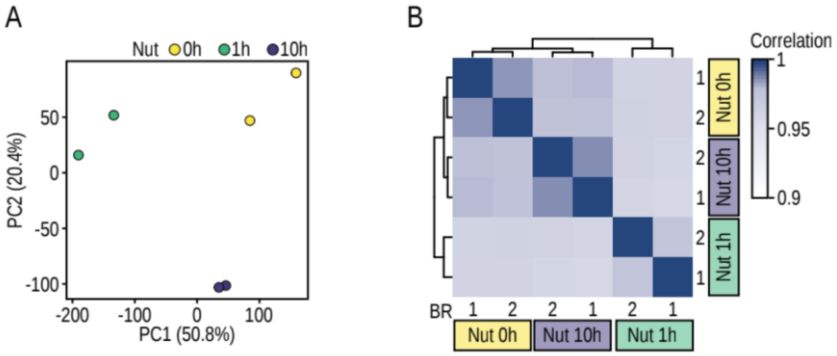
Fig 23. PCHi-C changes due to p53 activation. A) Correlation plot for resultant Pearson's residuals from a Chi-square test of independence for the contingency table time-course *vs.* genomic distance bracket (X-squared = 2096.3, df = 20, p-value < 2.2e-16). Bubble size indicates the number of interactions associated. Blue to red corresponds to an increasing association with genomic distance bracket. **B)** Interaction landscape of *JAG2's* gene promoter (grey shade) showing dynamic interactions (arcs) with two p53-bound enhancers (red asterisk and pink shade) located 186kb and 185kb away. H3K27ac profiles along p53 activation are represented at the top. Peak tips colored in black represent a peak going over the scale limit. Arrows symbolize gene placement and orientation along the genomic window.

Results II: The epigenetic landscape is impacted by p53 activation

To further investigate the consequences of p53 activation, we performed comprehensive epigenetics analyses at the three time points previously identified as crucial in capturing the dynamic changes in chromatin structure: 0, 1, and 10 hours of Nutlin-3a treatment. We used the 0-, 1- and 10-hours' time points to generate high quality chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) libraries of two histone modifications (H3K4me1 and H3K27ac), enabling the identification of primed enhancers (H3K4me1), active enhancers (H3K4me1 and H3K27ac) and active promoters (H3K27ac). Besides, we also used already processed publicly available ChIP-seq data of p53's binding profile in the same cell type subjected to similar activation.

2.1. Enhanced deposition of H3K4me1 and H3K27ac marks

The PCA revealed distinct sample clustering based on their H3K4me1 and H3K27ac deposition profiles. Principal components 1 and 2 explained a significant proportion of the variance in the data, highlighting their importance in capturing the major sources of variation. Notably, clustering patterns align with our previous findings on chromatin structure. Nut0h and Nut10h samples exhibited similar deposition patterns, suggesting a degree of consistency in H3K4me1 (**Fig 24 A-B**) and H3K27ac (**Fig 24 C-D**) levels during these time points. Conversely, the deposition pattern observed in Nut1h differed significantly from the other conditions, indicating a distinct response to p53 activation at this specific time point.



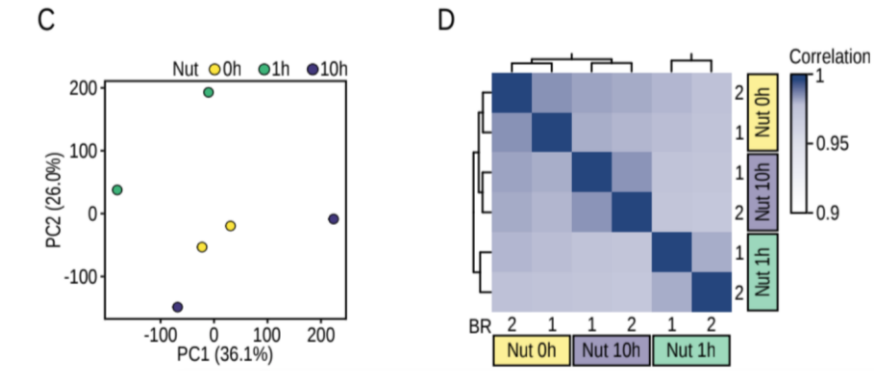
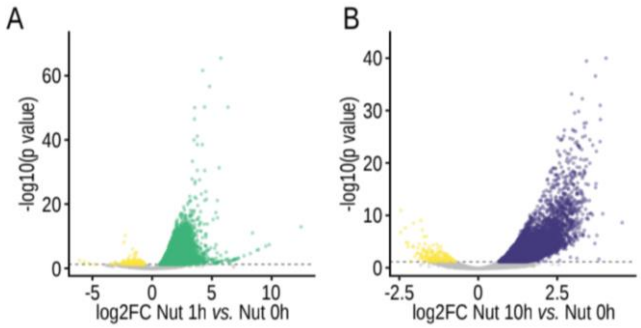


Fig 24. H3K4me1 and H3K27ac dynamics along p53 activation. A) Principal Component Analysis of H3K4me1 biological replicates along p53 activation. Numbers in parentheses represent the percentage of variance explained by each principal component. **B)** Pearson Correlation between biological replicates (1 and 2) of H3K4me1 samples along p53 activation. Samples were clustered hierarchically inputting Euclidean distances to the "complete" agglomeration method. **C) and D)** Same as A and B but for H3K27ac.

The activation of p53 with Nutlin-3a resulted in an increase on H3K4me1 levels at both 1 hour and 10 hours after treatment compared to the control group (**Fig 25 A-B**). At 10 hours after Nutlin-3a treatment, the deposition became more pronounced highlighting the cumulative and time-dependent influence of p53 activation on H3K4me1 deposition. Similarly, the activation of p53 with Nutlin-3a led to an increase in H3K27ac deposition compared to the control group. As time progressed, at 10 hours after Nutlin-3a treatment, the effects of p53 activation on H3K27ac became even more pronounced, with a further significant increase in H3K27ac levels compared to the control group (**Fig 25 C-D**). These findings show the cumulative impact of p53 activation on H3K4me1 and H3K27ac deposition and emphasize its sustained effect over time.



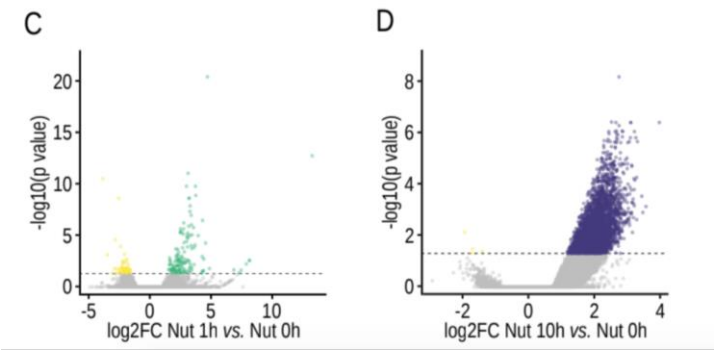


Fig 25. H3K4me1 and H3K27ac changes due to p53 activation with Nutlin-3a. A) Differential expression analysis between 1 hour (Nut 1h) and 0-hour (Nut 0h) time points. Differentially expressed peaks are those with log2 fold change (log2FC) ≠ 0 and adjusted P value ≤ 0.05 (horizontal dotted line). Increase in deposition of H3K4me1 peaks are in green, loss of deposition H3K4me1 peaks are in yellow. **B)** Same as A but for Nut 10h instead of 1h. Increase in deposition of H3K4me1 peaks are in purple. **C) and D)** Same as A and B but for H3K27ac.

2.2. Augmented p53 occupancy and preference binding for established enhancers upon p53 activation

We observed a significant increase in p53 binding upon activation compared to the basal state. Before activation, p53 exhibited a relatively low occupancy across the genome (657 peaks). However, following p53 activation, we observed a widespread increase of p53 binding (5667 peaks) (**Fig 26A**). Most of the p53 binding sites prior Nutlin-3a treatments were also conserved after p53 activation (655 out of 657 peaks).

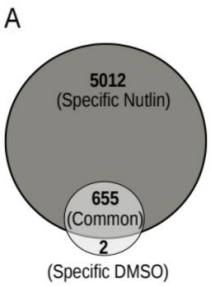


Fig 26. p53 bindings before and after p53 activation with Nutlin-3a. Overlap between p53 binding before p53 activation (DMSO), in this thesis DMSO is referred as Nut0h, and p53 bindings after p53 activation using Nutlin-3a during 12h.

We characterized a set of 2105 functional p53 binding sites as those p53 binding sites harboring activating H3K27ac histone mark under p53 activation (at 1 or 10 hours) (**Fig 27A**) that were consistently distributed across chromosomes (**Fig 27B**).

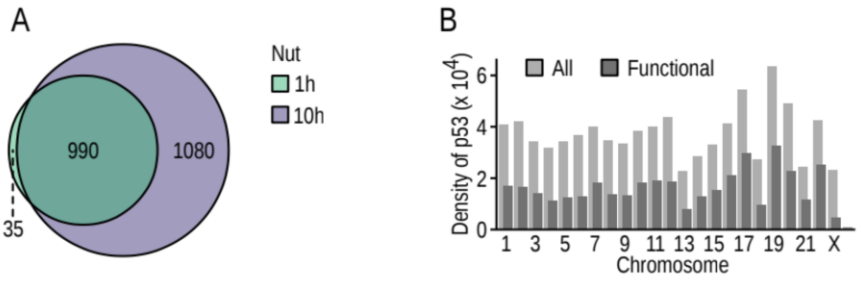


Fig 27. Functional p53 bindings. A) Overlap between functional p53 binding sites defined at Nut 1h and Nut 10h time points. **B)** Number of all (light grey) and functional (dark grey) p53 binding sites per chromosome.

Only a small proportion, 137 out of 2105 (6.5%), of these functional p53 bindings were found at promoters (*i.e.,* -1000 +200bp of any TSS). The remaining functional p53 binding events were distal from TSS. Interestingly 901 out of 2105 (42.8%) sites were located at active enhancers (*i.e.,* co-presence of H3K4me1 and H3K27ac at 1 or 10 hours of Nutlin-3a treatment), suggesting that p53's regulatory influence extends beyond the traditional promoter region (**Fig 28A**). Interestingly most of these active enhancers were already established at basal conditions (**Fig 28B**) and increased their H3K27ac levels as consequence of p53 activation (**Fig 28C-D**).

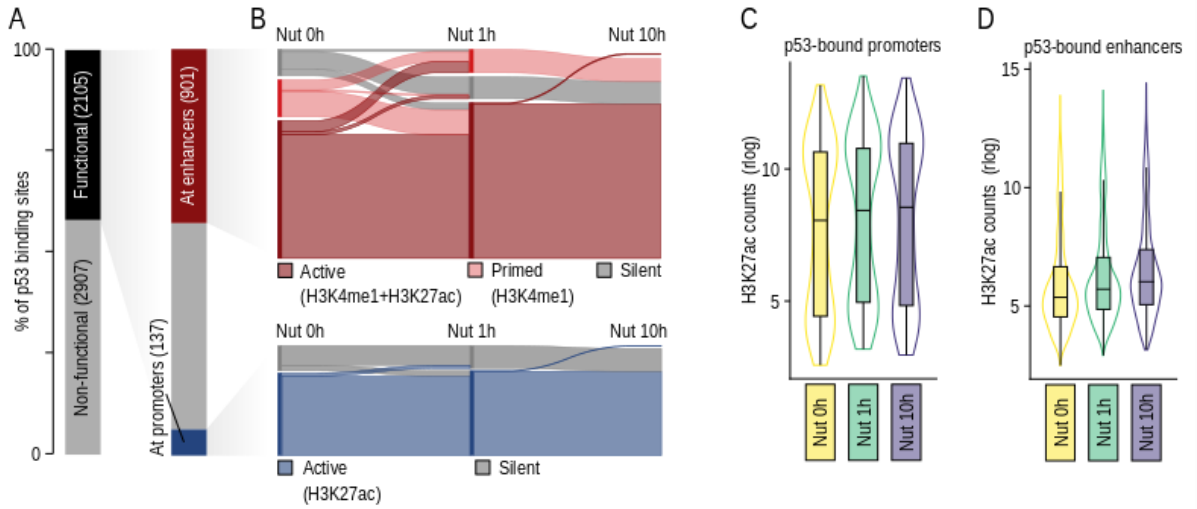


Fig 28. Distribution of the functional p53 binding sites. A) Bar plots showing the proportion of functional p53 binding sites (black) found at either promoters (blue) or enhancers (red). **B)** Sankey plot with varying states of activity over defined time points (Nut 0h, Nut 1h and Nut 10h). **C)** Distribution of normalized and rlong transformed H3K27ac counts at active promoters over the three time points (Nut 0h, Nut1h, Nut10h) **D)** Distribution of normalized and rlong transformed H3K27ac counts at active enhancers over the three time points.

Results III: Gene transcription is altered by p53 activation

Cells collected at the 3 time points: 0,1 and 10h were used to generate RNA-seq libraries to enable the genome-wide profiling of gene transcription. The transcriptome profiles of Nut0h and Nut10h samples as observed in both the PCA and correlation analysis exhibited a striking similarity, suggesting a common response to p53 activation. Conversely, the transcriptome pattern observed in Nut1h samples demonstrated significant divergence from the other conditions, indicating a distinct transcriptional response specifically at this time point (**Fig 29 A-B**). The volcano plots showed changes in gene expression profiles induced by p53 activation at both 1 and 10h, reflecting cellular responses to p53 activation (**Fig 29 C-D**).

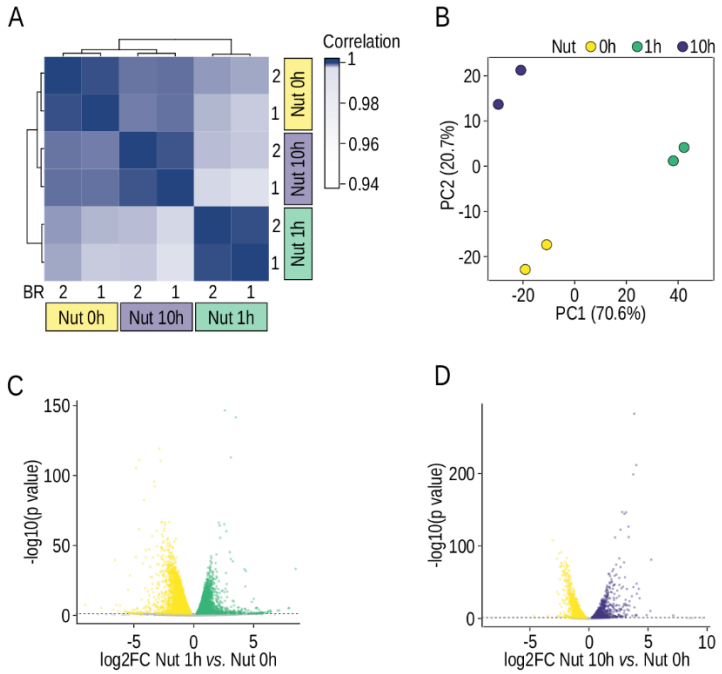


Fig 29. **Transcriptional changes due to p53 activation with Nutlin-3a. A)** Pearson Correlation between biological replicates of RNA-seq samples along p53 activation. Samples were clustered hierarchically inputting Euclidean distances to the "complete" agglomeration method. **B)** PCA of RNA-seq biological replicates along p53 activation. Numbers in parentheses represent the percentage of variance explained by each principal component. **C)** Differential expression analysis between 1 hour (Nut 1h) and 0 hour (Nut 0h). Differentially expressed genes are those with log2 fold change (log2FC) ≠ 0 and adjusted P value ≤ 0.05 (horizontal dotted line). Upregulated genes are in green or purple, downregulated genes are in yellow. **D)** Same as C but for 10-hour (Nut 10h) and 0-hour (Nut 0h) time points.

Results IV: Genome organization changes are triggered by p53 binding directly and indirectly

We conducted an integration analysis to determine if the binding patterns of functional p53 sites exerted any influence, directly or indirectly, on the genome organization.

Firstly, we observed a consistent preference of functional p53 bindings for A compartments (**Fig 30A**). This suggests that p53 may play a role in modulating the organization of chromatin compartments, favoring interactions within the A compartment. Additionally, we observed a notable enrichment of p53 binding near the TAD borders (**Fig 30B**). This finding implies that p53 binding may have a role in shaping the boundaries of these structural domains, potentially influencing the interactions and regulatory relationships between genomic regions. Furthermore, we discovered a significant enrichment of p53 binding in compartments that underwent a switch from the B compartment to the A compartment during p53 activation (**Fig 30C**). Specifically, after 1 hour and 10 hours of p53 activation, 5.7% (61 out of 1,078; Fisher's odds ratio [O.R.] 3.5, p-value 0.007) and 13.1% (22 out of 168; Fisher's O.R. 3.9, p-value 4e-6) of the transitioning compartments from B to A displayed functional p53 binding (depicted by the pink area in **Fig 30D** for 1h and **Fig 30E** for 10h). Interestingly, when we examined the impact of functional p53 binding on compartments transitioning from A to B during p53 activation, we observed a smaller proportion of these compartments harboring functional p53 binding sites (1.7% after 1 hour, 7 out of 411; 3.7% after 10 hours, 48 out of 1281) (depicted by the light blue area in **Fig 30D** for 1h and **Fig 30E** for 10h). However, these compartments displayed a more moderate decrease in compartment score compared to compartments without functional p53 binding (depicted by the light blue area in **Fig 30F** for 1h and **Fig 30G** for 10h). These findings suggest that the influence of p53 on compartments transitioning from A to B during p53 activation is relatively less pronounced compared to the B to A switch. Nevertheless, it is important to note that even in these transitioning compartments, p53 still plays a role in reducing their level of inactivity.

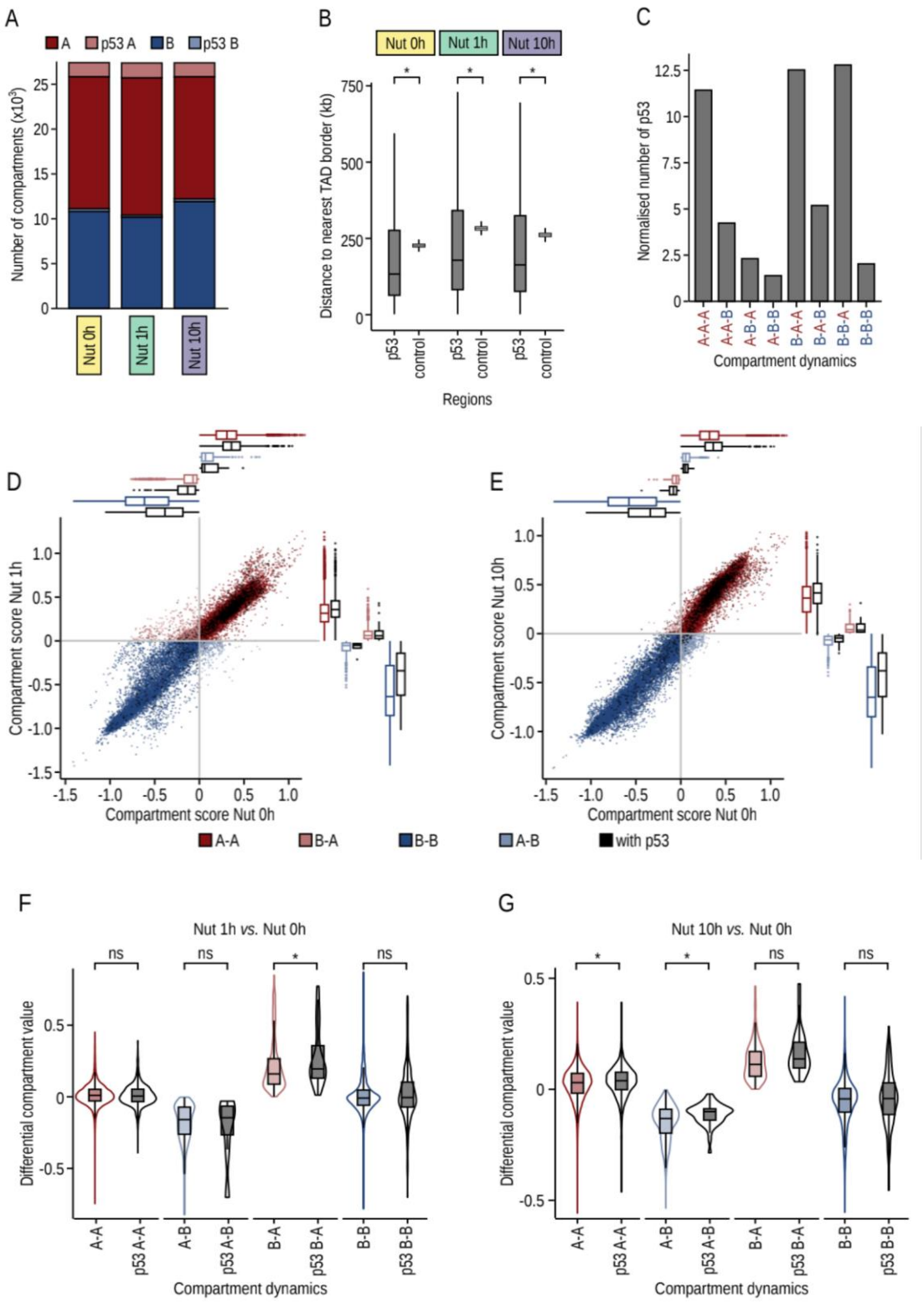


Fig 30. p53 binding to chromatin leads to direct changes in 3D genome topology. A) Proportion of functional p53 bindings in A/B compartments along p53 activation. p53 A/ p53 B are 100 kb bins in A/B compartment with at least one p53 functional binding. **B)** Distribution of the distances between functional p53 binding sites or regions without p53, and the nearest TAD border. Regions without p53 binding are defined as a random set of regions of similar

genomic length to a p53 binding site and a similar distribution across the genome. **C)** Proportion of p53 bindings in the different categories of A/B compartments dynamics. Data is normalized by the total number of 100kb regions considered in each sample. **D)** Scatter plot showing changes in compartment score between compartments defined at the control time point (Nut 0h) and after 1 hour of p53 activation (Nut 1h). Each dot represents a 100kb region. Dots are colored black if the region contains at least one p53 binding site. Box plots represent the distribution of the compartment scores for each defined compartment dynamic, with or without the presence of p53. **E)** As in panel D, but between compartments defined at the control time point (Nut 0h) and after 10 hours of p53 activation (Nut 10h). **F)** Distribution of differential compartment scores between compartments defined after 1 hour of p53 activation and control compartments (Nut 1h – Nut 0h). 100kb regions were classified according to their compartment dynamism and according to the presence or absence of functional p53 binding sites. **G)** As in panel F, but differential compartments scores are between compartments defined after 10 hours of p53 activation and control compartments (Nut 10h – Nut 0h).

Irrespective of compartment dynamics, compartments harboring functional p53 binding sites exhibited a notable increase in activating H3K27ac mark both at 1h and 10h of p53 activation (**Fig 31A**) indicating enhanced chromatin accessibility and permissiveness for transcriptional activation. Furthermore, after 10h of p53 activation regions harboring functional p53 binding sites exhibited a notable increase in transcriptional activity (**Fig 31B**) compared to regions lacking functional p53 binding.

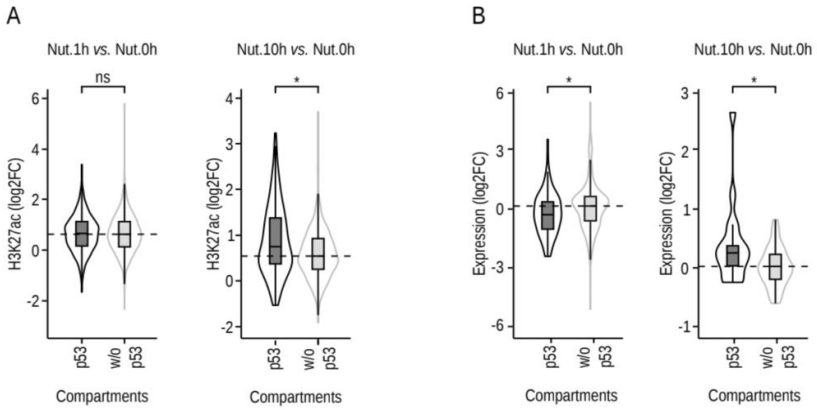


Fig 31. Changes in H3K27ac and transcription due to p53 binding. A) Differential H3K27ac between Nut 1h and Nut 0h (left) and Nut 10h and Nut 0h (right) either with p53 or without an overlapping p53 binding site. Two-sample Wilcoxon test was performed to test whether distributions differed significantly. **B)** As in panel A, but in this case analyzing the differential RNA-seq between Nut 1h and Nut 0h (left) and Nut 10h and Nut 0h (right).

Discussion

Influence of p53 activation on chromatin structure

Despite extensive investigations on recognizing p53's central function in coordinating cellular responses to stress, the potential involvement of p53 in shaping globally the 3D chromatin structure has remained largely unexamined. Considering the role of the genome's 3D arrangement in governing the dynamics of gene expression and recognizing p53's central function in transcriptionally coordinating cellular responses to stress, the potential involvement of p53 in shaping 3D chromatin structure underscores the innovative potential of this research. Therefore, this thesis conducted a comprehensive analysis of the impact of p53 activation on the dynamic 3D chromatin structure through a time-course omics analysis, including Hi-C, PCHi-C, ChIP-seq and RNA-seq.

One of the most remarkable unveilings of this thesis is a substantial reconfiguration of genome topology after p53 activation, particularly evident in the first hour of Nutlin-3a treatment. The activation of p53 after 1h of Nutlin-3a treatment triggers a discernible shift from the B compartment to the A compartment and a reorganization of a significant number of TAD borders. The compartment switch favoring the A compartment, raises the possibility that p53 induces changes in chromatin structure to potentially prime the genome for subsequent or simultaneous transcriptional activation. This orchestrated remodeling could potentially serve to unveil previously inaccessible genomic regions, triggering a cascade of events aimed at addressing the stress induced by the treatment. The reconfiguration of TAD borders, particularly the loss of these borders, potentially reflects a purposeful disruption of spatial constrains impacting interactions interrupted by a TAD border, thus facilitating the physical interaction between regulatory elements and target genes crucial for the execution of p53-mediated responses.

As the treatment duration progresses, the effects gradually wane at 4 and 7 hours, a resemblance toward the control time-point at 10 hours post-treatment, and an inclination toward terminal cell fate at the 24-hour mark manifest. One perspective to this phenomenon could be due to a transition from the initial acute stress response to a more sustained and adaptable cellular state. Chronic stress might potentially engender a compensatory mechanism leading to the attenuation of p53 function. This

attenuation might serve as a safeguard against the potential repercussions of persistent p53 activity. By mitigating the risks of unchecked p53-driven responses, the cell navigates towards terminal cell fate determinations. Therefore, the attenuation of p53 at 10 hours after Nutlin-3a treatment might result in the reversal of the early p53 activation events, leading to a re-establishment of certain TAD borders and a re-location of genomic regions to the B compartment finalizing into a terminal state after 24 hours of treatment. An alternative viewpoint to this phenomenon could be attributed to the indirect outcomes of p53's actions. The dynamism observed within the first hour of p53 activation is primarily attributed to p53's direct interaction with DNA, emphasizing its role as a transcriptional activator. However, when p53 activation advances to later stages, we begin to witness the emergence of indirect consequences, which are mediated by factors whose transcription is regulated by p53. These indirect effects tend to exhibit greater diversity and depend on whether the downstream effectors activate or repress genes.

When examining the distribution of p53 bindings within distinct categories of A/B compartments dynamics, a significant proportion of p53 bindings are located within regions classified as B compartments at basal conditions that switch to A compartment at 1 or 10 hours. This observation implies a potential causal relationship between p53 binding and changes in compartment dynamics.
Moreover, our findings unveil p53's notable propensity to bind near TAD borders, a phenomenon that surpasses statistical expectations rooted in random chance. One potential explanation for this event could be that p53 regulates genes situated at the junction of chromatin domains, potentially enabling cross-TAD communication. Therefore, the binding of p53 near a TAD border could initiate a cascade of molecular events that impact the overall chromatin architecture.

The question at hand is whether the observed loss of TAD borders and compartmental switches following p53 activation are causally connected or are simultaneous but independent events. Although these phenomena manifest concurrently in our results, studies have indicated that compartmentalization and TAD organization may be governed by distinct mechanisms (104,105). This raises the possibility that they might represent parallel outcomes arising from the intricate network of cellular responses to p53 activation inviting further investigation of these processes.

As chromatin organization orchestrates transcription by shaping the 3D-interactions between enhancers and promoters in the nucleus, considering the observed changes in chromatin organization after p53 activation within the first hour followed by a reestablishment effect at the 10-hour time point, the next focus within this thesis turned to the exploration of DNA loops. Given the inherent limitation of Hi-C data resolution unless ultradeep sequencing of the libraries in a highly cost manner, in accurately capturing looping interactions (106–108), this thesis used PCHi-C to map the promoter interactome after p53 activation at 0, 1 and 10 hours after Nutlin-3a treatment.
At 1 hour of p53 activation, the PCHi-C analysis shows a transition towards short-range interactions within the promoter interactome in comparison to control. This phenomenon likely mirrors a localized reconfiguration of the chromatin architecture, where short-range interactions assume a crucial role in the synchronized expression of genes residing within specific genomic regions enabling rapid and precise cellular responses to stress. Moreover, the increased prevalence of promoter-promoter interactions may signify the formation of transcriptional hubs, where promoters of stress-responsive genes closely interact to collectively regulate their expression. This spatial arrangement could enhance the efficiency of transcriptional activation under acute stress conditions. However, as stress signals persist and the cellular demands evolve, another shift in the promoter interactome becomes evident at the 10-hour mark after p53 activation: a greater proportion of Promoter-Other End interactions and longer genomic distances (>300Kb). In this context, sustained stress signals may necessitate the coordination of gene expression across larger genomic distances, driving the formation of long-range interactions to ensure comprehensive and coordinated stress responses at 10h.

In summary, this thesis has illuminated a substantial reorganization of genome topology triggered by p53 activation, particularly evident within the first hour of Nutlin-3a treatment.

As binding of transcription factors to chromatin has been reported to allow chromatin structure reorganization and stimulate gene expression through changes in histone PTM (109), as next step in this thesis a more extensive study using ChIP-seq in some histone modifications in the same time points was performed to define more precisely how chromatin responds to p53 activation.

Influence of p53 activation on epigenetics

The invariant binding of p53 to the genome across different cell types reported in several studies (72,110) seems to modulate gene expression by the regulatory context of enhancers and promoters. Thus, this thesis uncovered the promoters and enhancers that become active in HCT116 cell line influenced by p53 activation with Nutlin-3a by profiling histone hallmarks associated with primed enhancers (H3K4me1), active enhancers (H3K4me1 and H3K27ac) and active promoters (H3K27ac) at 0, 1 and 10 hours.

Upon p53 activation, a significant increase in the deposition of both histone modifications at 1 and 10 hours compared to the control group is observed. The sustained effect on H3K4me1 and H3K27ac deposition at 10 hours suggests a robust and enduring chromatin landscape, indicative of prolonged transcriptional activity beyond the immediate cellular stress.

To assess whether p53 binding is directly involved, we examined H3K4me1 and H3K27ac enrichment at functional p53-binding sites. The analysis revealed that p53 binding in active promoter regions, located within -1000 +200bp of any TSS with H3K27ac mark, account for the smallest group of p53 binding sites (6,5%). Genes involved in these bindings had some degree of activity at basal conditions (i.e., H3K27ac at promoters), and tended to gain H3K27ac deposition after Nutlin-3a treatment. Previous analyses of p53 genomic occupancy had suggested a correlation between p53 binding near gene promoters and changes in gene activation (111). Nevertheless, a significant number of p53-dependent gene targets are likely regulated by more distal p53-binding events, which might explain the relatively small representation of active promoter regions among p53 bindings. Among the non-promoter regions (distal regions from TSS), 42,8% of p53 bindings were observed at enhancers (*i.e.,* co-presence of H3K4me1 and H3K27ac at some of the 1- or 10-hour time-points). Interestingly, most of them were already active in basal conditions and increased their H3K27ac levels as consequence of p53 activation. This observation aligns with previous indications that p53's presence is not essential for basal enhancer RNA abundance (112). These findings suggest that p53-bound enhancers are pre-

established. Binding to preexisting enhancers could represent a mechanism for generating specific p53 responses tailored to cell type or lineage, potentially contributing to rapid and context-specific stress responses.

In summary, our findings indicate that p53's initial impact doesn't involve the initiation of new enhancer establishment. However, it remains viable that p53's activity serves as a crucial factor in augmenting preexisting enhancers that hold significant importance in p53-mediated responses. To thoroughly test this hypothesis, this thesis embarked on an examination of the transcriptional state.

Influence of p53 activation on transcription

By comparing the basal transcriptional state to the time-points at 1h and 10h, this thesis aimed to dissect the immediate and evolving transcriptional shifts invoked by p53 activation.

The results show that both at the 1-hour and 10-hour time-point, the cellular landscape witnesses both up-regulation and downregulation due to p53 activation.
Although p53 is well established as a transcriptional activator, it's important to note that within the p53 network, there exists a multitude of microRNAs that play a role in indirect repression. One of the pioneering discoveries in this context was the identification of miR-34a as a microRNA induced by p53. MiR-34a actively contributes to the arrest of the cell cycle by exerting post-transcriptional repression on genes essential for cell cycle progression (113–116). This highlights the regulatory mechanisms orchestrated by p53, involving not only direct actions but also the modulation of microRNAs to achieve its cellular effects.
Indeed, a study of genes downregulated in HCT116 cells following Nutlin-3a treatment found that 67% of these genes were confirmed as targets of miR-34a (76). Moreover, there are several other microRNAs that are directly activated by p53, which could contribute to the indirect repression of gene expression (117).

This thesis makes exclusive use of poly-A+ RNA-Seq data, which does not always reflects the transcriptional rates. Consequently, the conventional measurement of mRNA levels may not fully capture the subtle increases in transcriptional rates at 1h of p53 activation of regions harboring functional p53 binding sites. This discrepancy in our results arises because miRNAs and RNA binding proteins (RBPs), whose expression is modulated by p53, can exert significant influence on mRNA stability culminating in alterations on the levels of mature RNA.
A method that removes this limitation that poly-A+ RNA-Seq has is GRO-seq. One study performed a GRO-seq analysis after 1h of Nutlin-3a treatment (112) and illuminated the changes associated with detecting early p53-induced transcriptional responses when relying on the quantification of steady-state mRNA levels. In their research, they performed a comparative analysis that contrasted nascent RNA levels

with the RNA steady-state levels determined via microarray revealing that numerous p53 target genes, which exhibit low transcriptional activity, remained undetected when relying on microarray alone.

In summary, our perspective suggests that a significant portion of p53's immediate target genes undergo upregulation within one hour of activation. However, it's important to note that these changes in gene expression might not be readily discernible through conventional mRNA level measurements.

Future directions

Despite the significant advancements in this thesis in the understanding of p53's role in 3D chromatin structure, to further expand and refine this thesis, several future directions can be pursued:

1. **Diversify experiments:** The experimental design of this thesis has primarily been centered on the activation of p53 in human colorectal cancer cells using a specific drug. One way for diversification can be to investigate the effects of p53 in primary human cell types, such as fibroblasts, endothelial cells, or immune cells to understand how p53 operates in these non-cancerous contexts and shed light on its fundamental functions in normal cellular processes. Furthermore, it would be interesting to consider examining p53's function in various human cancer cells as different tissues may exhibit unique p53-dependent responses. Additionally, extending the research to different species, such as mice or other model organisms, can offer comparative insights enabling us to explore the evolutionary conservation of p53 function and identify conserved mechanisms as well as species-specific variations. Finally, to gain a more comprehensive understanding of p53's response under different stimuli, conducting experiments using alternative activation mechanisms, such as employing different drugs like etoposide as well as using different amount of drugs concentrations, can uncover the different aspects of p53's regulatory functions.
2. **Determine the exact time-point of the early changes**: The results of this thesis reveal changes at 1 hour following p53 activation. However, it will be interesting to initiate a new time-course study within the 0 to 1-hour timeframe to pinpoint the exact moment at which these changes occur.
3. **Further validate the p53 effects:** To confirm the observed alterations in chromatin compartments and provide additional evidence of changes in chromatin accessibility an ATAC-seq (Assay for Transposase-Accessible Chromatin using Sequencing) experiment could be performed. Furthermore, since cohesin removal leads to deletion of promoter-enhancer interactions, to explore whether p53 relies on chromatin loops to trigger transcriptional responses an experiment employing an auxin-inducible degron (AID) system

for RAD21 depletion can be conducted to later perform an RNA-seq and PCHi-C analysis and compare. Additionally, examining the reversibility of p53-driven changes can enhance the quality of this work. This can be achieved by removing Nutlin-3a treatment after specific time intervals (e.g., 1 hour, 10 hours, and 24 hours) and observing any persistence of the effects. If the effects persist beyond the removal of Nutlin-3a, it suggests that they are not solely due to p53 activation but may involve other factors. Finally, to differentiate the specific contributions of p53 from those of the drug, it is essential to perform experiments using HCT116 cells with p53 knockout. This will allow us to compare the observed effects in cells lacking functional p53, thus clarifying the unique role of p53 in the observed outcomes.

4. **Computational Methods Validation**: Create 3D models to illustrate specific chromatin regions from the time-course Hi-C datasets, focusing on regions where a TAD border is lost or gained and there is a compartment switch at early and late stages of p53 activation or specific direct target genes of p53.

Conclusions

As a concluding remarks, we can consider this thesis proved that:

- **There is a dynamic and rapid 3D genome reorganization due to p53 activation**: This thesis has shown pronounced changes in the spatial arrangement of specific genomic regions, with a notable shift towards active compartments and TAD border dynamism during the early stages of p53 activation.
- **There is a pre-established set of enhancers to perform a rapid action upon p53 activation:** By binding to a predefined set of enhancers, p53 initiates its regulatory function within minutes of activation. This rapid response highlights the efficiency and precision with which p53 orchestrates gene expression changes, emphasizing its pivotal role in cellular stress responses.
- **There are early-stage transcriptional changes, not discernible through RNA-seq:** This finding underscores the need for additional research to delve into the details of these early alterations.

References

1. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. Nature. 2003 May 8;423(6936):145-50.
2. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. Molecular biology: The transcriptional landscape of the mammalian genome. Science. 2005 Sep 2;309(5740):1559–63.
3. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011 Feb 4;144(3):327-39.
4. Spitz F, Furlong EEM. Transcription factors: From enhancer binding to developmental control. Nat Rev Genet. 2012 Aug 7, 613–626.
5. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012 Sep 6;489(7414):109–13.
6. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol. 2015 Mar;16(3):178-89.
7. Esteller M. Molecular Origins of Cancer Epigenetics in Cancer. N Engl J Med. 2008 Mar 13;358(11):1148-59
8. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007 Mar;39(3):311–8.
9. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3):144-54.
10. Meaburn KJ, Misteli T. Cell biology: chromosome territories. Nature. 2007 Jan 25;445(7126):379-781.
11. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010 Mar;2(3):a003889.
12. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, et al. Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol Cell. 2010 May 28;38(4):603–13.
13. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct 9;326(5950):289-93.

14. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 May 17;485(7398):376–80.
15. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665–80.
16. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 May 17;485(7398):381–5.
17. Hou C, Li L, Qin ZS, Corces VG. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. Mol Cell. 2012 Nov 9;48(3):471–84.
18. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012 Feb 3;148(3):458–72.
19. Matharu N, Ahituv N. Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. PLoS Genet. 2015 Dec 3;11(12):e1005640.
20. Lazaris C, Aifantis I, Tsirigos A. On Epigenetic Plasticity and Genome Topology. Trends Cancer. 2020 Mar;6(3):177-180.
21. Bouwman BA, de Laat W. Architectural hallmarks of the pluripotent genome. FEBS Lett. 2015 Oct 7;589(20 Pt A):2905-13.
22. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell. 2014 Jun 5;14(6):762-75.
23. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 30;161(5):1012–25.
24. Hnisz D, Weintrau AS, Day DS, Valton AL, Bak RO, Li CH, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016 Mar 25;351(6280):1454–8.
25. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538(7624):265–9.

26. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018 Oct 26;362(6413).
27. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016 Oct 14;17(11):661-678.
28. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019 Aug;20(8):437-455.
29. Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. 2017 Dec 15;36(24):3573–99.
30. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31-46.
31. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002 Feb 15;295(5558):1306-11.
32. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell. 2002 Dec;10(6):1453-65.
33. Comet I, Schuettengruber B, Sexton T, Cavalli G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. Proc Natl Acad Sci U S A. 2011 Feb 8;108(6):2294-9.
34. Liu N, Low WY, Alinejad-Rokny H, Pederson S, Sadlon T, Barry S, Breen J. Seeing the forest through the trees: prioritising potentially functional interactions from Hi-C. Epigenetics Chromatin. 2021 Aug 28;14(1):41.
35. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. J Vis Exp. 2018 Jun 28;(136):57320.
36. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009 Feb 11;27(2):182–9.
37. Johnson RB, Onwuegbuzie AJ. Toward a Definition of Mixed Methods Research. J Mix Methods Res. 2007;1(2):112–33.

38. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 2017 Jan 16;6:e21856.
39. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Dec 1;10(1).
40. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. Nat Protoc. 2020 Oct 1;15(10):3264–83.
41. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5(7):621–8.
42. Kukurba KR, Montgomery SB. RNA sequencing and analysis. Cold Spring Harb Protoc. 2015 Nov 1;2015(11):951–69.
43. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (1979). 2008 Dec 19;322(5909):1845–8.
44. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;339(6127):1546-58.
45. Levine AJ. p53, the cellular gatekeeper for growth and division. Cell. 1997 Feb 7;88(3):323-31.
46. Lane DP, Crawford LV. T antigen is bound to a host protein in SV40-transformed cells. Nature. 1979 Mar 15;278(5701):261-3.
47. DeLeo AB, Jay G, Appella E, Dubois GC, Law LW, Old LJ. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. Proc Natl Acad Sci U S A. 1979 May;76(5):2420-4.
48. Finlay CA, Hinds PW, Levine AJ. The p53 proto-oncogene can act as a suppressor of transformation. Cell. 1989 Jun 30;57(7):1083-93.
49. Malkin D, Li FP, Strong LC, Fraumeni JF Jr, Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990 Nov 30;250(4985):1233-8.
50. Matlashewski G, Lamb P, Pim D, Peacock J, Crawford L, Benchimol S. Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. EMBO J. 1984 Dec 20;3(13):3257-62.

51. Isobe M, Emanuel BS, Givol D, Oren M, Croce CM. Localization of gene for human p53 tumour antigen to band 17p13. Nature. 1986 Mar 6-12;320(6057):84-5.
52. Lamb P, Crawford L. Characterization of the human p53 gene. Mol Cell Biol. 1986 May;6(5):1379-85.
53. Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther. 2013 Dec 19;7:57-68.
54. Joerger AC, Fersht AR. Structural biology of the tumor suppressor p53. Vol. 77, Annual Review of Biochemistry. 2008. p. 557–82.
55. Candau R, Scolnick DM, Darpino P, Ying CY, Halazonetis TD, Berger SL. Two tandem and independent sub-activation domains in the amino terminus of p53 require the adaptor complex for activity. Oncogene. 1997 Aug 14;15(7):807-16.
56. Tanaka T, Watanabe M, Yamashita K. Potential therapeutic targets of *TP53* gene in the context of its classically canonical functions and its latest non-canonical functions in human cancer. Oncotarget. 2018 Mar 23;9(22):16234-16247.
57. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, et al. A global map of p53 transcription-factor binding sites in the human genome. Cell. 2006 Jan 13;124(1):207–19.
58. Hafner A, Kublo L, Tsabar M, Lahav G, Stewart-Ornstein J. Identification of universal and cell-Type specific p53 DNA binding. BMC Mol Cell Biol. 2020 Feb 18;21(1).
59. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ. 2018 Jan;25(1):154-160.
60. Bode AM, Dong Z. Post-translational modification of p53 in tumorigenesis. Vol. 4, Nature Reviews Cancer. 2004. p. 793–805.
61. Chehab NH, Malikzay A, Stavridi ES, Halazonetis TD. Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. Proc Natl Acad Sci U S A. 1999 Nov 23;96(24):13777-82.
62. Shieh SY, Ikeda M, Taya Y, Prives C. DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. Cell. 1997 Oct 31;91(3):325-34.

63. McDade SS, Patel D, Moran M, Campbell J, Fenwick K, Kozarewa I, Orr NJ, Lord CJ, Ashworth AA, McCance DJ. Genome-wide characterization reveals complex interplay between TP53 and TP63 in response to genotoxic stress. Nucleic Acids Res. 2014 Jun;42(10):6270-85.
64. Sammons MA, Zhu J, Drake AM, Berger SL. TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. Genome Res. 2015 Feb 1;25(2):179–88.
65. Su D, Wang X, Campbell MR, Song L, Safi A, Crawford GE, et al. Interactions of Chromatin Context, Binding Site Sequence Content, and Sequence Evolution in Stress-Induced p53 Occupancy and Transactivation. PLoS Genet. 2015;11(1).
66. Léveillé N, Melo CA, Rooijers K, Díaz-Lagares A, Melo SA, Korkmaz G, et al. Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. Nat Commun. 2015;6.
67. Younger ST, Rinn JL. P53 regulates enhancer accessibility and activity in response to DNA damage. Nucleic Acids Res. 2017 Sep 1;45(17):9889–900.
68. Jenkins LM, Durell SR, Mazur SJ, Appella E. p53 N-terminal phosphorylation: a defining layer of complex regulation. Carcinogenesis. 2012 Aug;33(8):1441-9.
69. Sablina AA, Budanov A V., Ilyinskaya G V., Agapova LS, Kravchenko JE, Chumakov PM. The antioxidant function of the p53 tumor suppressor. Nat Med. 2005 Dec;11(12):1306–13.
70. Kruse JP, Gu W. Modes of p53 regulation. Cell. 2009 May 15;137(4):609-22.
71. Horn HF, Vousden KH. Coping with stress: multiple ways to activate p53. Oncogene. 2007 Feb 26;26(9):1306-16.
72. Verfaillie A, Svetlichnyy D, Imrichova H, Davie K, Fiers M, Atak ZK, et al. Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. Genome Res. 2016 Jul 1;26(7):882–95.
73. Andrysik Z, Galbraith MD, Guarnieri AL, Zaccara S, Sullivan KD, Pandey A, et al. Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. Genome Res. 2017 Oct 1;27(10):1645–57.
74. Younger ST, Rinn JL. P53 regulates enhancer accessibility and activity in response to DNA damage. Nucleic Acids Res. 2017 Sep 1;45(17):9889–900.

75. Sammons MA, Zhu J, Drake AM, Berger SL. TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. Genome Res. 2015 Feb 1;25(2):179–88.
76. Sullivan KD, Galbraith MD, Andrysik Z, Espinosa JM. Mechanisms of transcriptional regulation by p53. Cell Death Differ. 2018 Jan;25(1):133-143.
77. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013 Mar 7;49(5):825-37.
78. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015 Mar;16(3):144-54.
79. Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, McMahon SB. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Mol Cell. 2006 Dec 28;24(6):841-51.
80. Vassilev LT. Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. Cell Cycle. 2004 Apr;3(4):419-21..
81. Tovar C, Rosinski J, Filipovic Z, Higgins B, Kolinsky K, Hilton H, Zhao X, Vu BT, Qing W, Packman K, Myklebost O, Heimbrook DC, Vassilev LT. Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. Proc Natl Acad Sci U S A. 2006 Feb 7;103(6):1888-93.
82. Huang B, Deo D, Xia M, Vassilev LT. Pharmacologic p53 activation blocks cell cycle progression but fails to induce senescence in epithelial cancer cells. Molecular Cancer Research. 2009 Sep;7(9):1497–509.
83. Hu B, Gilkes DM, Farooqi B, Sebti SM, Chen J. MDMX overexpression prevents p53 activation by the MDM2 inhibitor nutlin. Journal of Biological Chemistry. 2006 Nov 3;281(44):33030–5.
84. Yang A, Schweitzer R, Sun D, Kaghad M, Walker N, Bronson RT, Tabin C, Sharpe A, Caput D, Crum C, McKeon F. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature. 1999 Apr 22;398(6729):714-8.
85. Schmale H, Bamberger C. A novel protein with strong homology to the tumor suppressor p53. Oncogene. 1997 Sep;15(11):1363-7.
86. Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dötsch V, Andrews NC, Caput D, McKeon F. p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. Mol Cell. 1998 Sep;2(3):305-16.

87. Melino G, Lu X, Gasco M, Crook T, Knight RA. Functional regulation of p73 and p63: development and cancer. Trends Biochem Sci. 2003 Dec;28(12):663-70.
88. Javierre BM, Sewitz S, Cairns J, Wingett SW, Várnai C, Thiecke MJ, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016 Nov 17;167(5):1369-1384.e19.
89. Serra F, Baù D, Goodstadt M, Castillo D, Filion G, Marti-Renom MA. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol. 2017 Jul 1;13(7).
90. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: Fast, accurate and versatile alignment by filtration. Nat Methods. 2012 Dec;9(12):1185–8.
91. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods. 2012 Oct;9(10):999–1003.
92. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature. 2015 Jul 9;523(7559):240–4.
93. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP: Pipeline for mapping and processing Hi-C data. F1000Res. 2015;4.
94. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al. CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 2016 Jun 15;17(1).
95. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009 Mar 4;10(3).
96. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008 Sep 17;9(9).
97. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15–21.
98. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923–30.

99. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol. 2013;9(8).
100. Yang T, Zhang F, Yardımcı GG, Song F, Hardison RC, Noble WS, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Cold Spring Harbor Laboratory Press; 2017.
101. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. 2013 Sep 1; Available from: http://arxiv.org/abs/1309.0238
102. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020 Mar 1;17(3):261–72.
103. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020 Sep;585(7825):357-362.
104. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, et al. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017 Nov 2;551(7678):51–6.
105. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017 May 18;169(5):930-944.e22.
106. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, et al. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol Cell. 2020 May 7;78(3):539-553.e8.
107. Krietenstein N, Abraham S, Venev S V., Abdennur N, Gibcus J, Hsieh THS, et al. Ultrastructural Details of Mammalian Chromosome Architecture. Mol Cell. 2020 May 7;78(3):554-565.e7.
108. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, Zhang MD, et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell. 2019 Jan 10;176(1–2):377-390.e19.
109. Stadhouders R, Filion GJ, Graf T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature. 2019 May;569(7756):345-354.

110. Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. Cell. 2009 May 1;137(3):413-31.
111. Uzunbas GK, Ahmed F, Sammons MA. Control of p53-dependent transcription and enhancer activity by the p53 family member p63. Journal of Biological Chemistry. 2019 Jul 5;294(27):10720–36.
112. Allen MA, Andrysik Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, et al. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. Elife. 2014 May 27;3.
113. Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH, et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol Cell. 2007 Jun 8;26(5):745-52.
114. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, et al. Transcriptional Activation of miR-34a Contributes to p53-Mediated Apoptosis. Mol Cell. 2007 Jun 8;26(5):731–43.
115. Tarasov V, Jung P, Verdoodt B, Lodygin D, Epanchintsev A, Menssen A, et al. Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G 1-arrest. Cell Cycle. 2007 Jul 1;6(13):1586–93.
116. Tazawa H, Tsuchiya N, Izumiya M, Nakagama H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. Proc Natl Acad Sci U S A. 2007 Sep 25;104(39):15472-7.
117. Hoffman Y, Pilpel Y, Oren M. MicroRNAs and Alu elements in the p53-Mdm2-Mdm4 regulatory network. Vol. 6, Journal of Molecular Cell Biology. Oxford University Press; 2014. p. 192–7.
118. Yesbolatova A, Saito Y, Kitamoto N, Makino-Itou H, Ajima R, Nakano R, et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. Nat Commun. 2020 Dec 1;11(1).
119. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015 Dec 1;16(1).
120. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016 Jul;3(1):95-8.