Ollscoil na Gaillimhe
University of Galway

September 2022

An epigenomics approach to characterize alterations in diseases

Candidate:

Cristiano Petrini
ID: 18236794

Supervisor:

Kevin Sullivan

Professor of Cell Biology in the Centre for chromosome biology

University of Galway

External supervisor:

Dr. Francesco Ferrari

Principal Investigator

IFOM ETS - The AIRC Institute of Molecular Oncology

Table of Contents

1 Summary

Epigenetics is the term that describes all the mechanisms that regulate the genomic activities such as transcription, DNA repair, DNA replication, transposon duplication. There is a great variety of epigenetic mechanisms that act at different genomic scales: from a single nucleotide resolution (e.g. DNA modifications) to several megabases (chromatin-lamina interactions).

As epigenetics mechanisms are involved in many cellular processes, their alteration could lead to the development of pathologies including genetic diseases and cancer.

Thus, considering the importance and variety of epigenetics mechanisms numerous experimental techniques have been developed to study them. Despite the progresses so far, the genome-wide techniques to study epigenetics mechanisms are still affected by several technical biases and most of them are not suitable for investigation of tissue samples, especially in case of biological material scarcity.

In this thesis I am presenting a new technique: the Sequential Analysis of Macro-Molecule Accessibility by sequencing (SAMMY-seq). This method is not affected by most common technical biases of epigenomics methods and can be easily performed on as little as 10 thousand cells, as well as in tissue samples. With SAMMY-seq it is possible to extract and separately analyze distinct chromatin regions according to their different biochemical properties.

We analyzed different cells, organisms and samples, showing that the technique is highly reproducible and the results in line with known epigenetic profiles. Once applied on two different disease cellular models (progeroid fibroblast and MCF10DICIS cancer cells) SAMMY-seq has been able to detect chromatin alterations where other techniques failed.

In addition, thanks to its low starting material requirement, it has been possible to apply SAMMY-seq on prostate biopsies from patients with cancer, thus describing for the first time on tissue an association between epigenomic alterations and patients subgroups.

Finally, using SAMMY-seq we developed a data analysis technique to reconstruct chromatin compartments. We have been able to investigate the relationship between chromatin compartments and chromatin fractions separated by solubility.

Our results demonstrate that SAMMY-seq is a powerful method to study epigenomics providing new insights over common diseases such as prostate cancer.

Furthermore, altering chromatin solubility inside the cell nucleus we observe changes in compartments distribution, we have been able to suggest solubility as a new epigenetic mechanism.

2 Introduction

2.1 Epigenetics regulation of genome functionality

In 1944 it was first demonstrated that all the "instructions" (genes) necessary to make an organism are coded and distributed along its DNA sequences, the ensemble of these is called genome. Obviously not all the genes are expressed or copied at the same time and their activity should be finely regulated (Avery et al., 1944). This is the role of epigenetic mechanisms which modulate the genome activity by ensuring that genomic processes (e.g. transcription, replication and DNA repair) take place at the right moment and in the right genomic region. The ensemble of epigenetic modifications on the genome is defined as the epigenome.

Despite the fact that genome and epigenome are strictly connected (the prefix “epi” comes from ancient Greek and means “above”, so we are referring to all the events that happen “above” the genome), they are very different in their features. Indeed, while the genome remains almost the same in all the cells of an organism, the epigenome is different among each cell type. That makes it possible to respond to external stimuli and regulate inner cellular programs such as differentiation (Suzuki & Bird, 2008).

Epigenetics mechanisms are crucial for proper genome activity and their alteration may lead to pathologies such as cancer or genetic diseases. Studies on the field of epigenetics are therefore important to understand gene regulation in physiological and pathological conditions. This may increase our understanding of diseases, how to prevent them and, possibly, how to treat them (Atlasi & Stunnenberg, 2017; Cavalli & Heard, 2019; D. Li et al., 2021; L. Zhang et al., 2020).

In this first section I will briefly introduce the most studied epigenetics mechanisms starting from the ones regarding the genome at the finest scale (i.e. the modification of single DNA nucleotides), and going to the broadest ones (e.g. nuclear lamina interactions with chromatin that involve DNA regions of several megabases).

2.1.1 DNA modifications

The DNA modifications consist in covalent chemical changes of the Watson and Crick DNA molecule, in particular they are referring to the substitution of a “canonical” nucleotide (adenine, cytosine, thymine or guanine) with a chemical modified version of it (Yuan, 2020).

Currently, a great variety of them have been reported in different species. In mammals in particular, the most common are cytosine nucleotide modifications: 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Recently others have been described such as N6-methyladenine (6mA) and 5-hydroxymethyluracil (5hmU) (Sood et al., 2019). Their presence allows the recruitment of specific enzymes that regulate the genome activity (e.g. transcription) (Sepehri et al., 2019).

The most studied and known of these modifications is the 5mC. Chemically it is the addition of a methyl group on the fifth carbon of the cytosine ring. It has been found prevalently on contiguous sequences of cytosine and guanine regions (CpG islands), even though it has been identified in non-CpG islands as well (Breiling & Lyko, 2015). The 5mC has been associated to multiple biological mechanisms including: (i) gene expression regulation, where it exerts a repressive role if it localizes 1Kb near the transcription starting site (TSS), whereas it is associated to active transcription if it localizes on the gene body; (ii) splicing regulation; (iii) nucleosome positioning and (iv) recruitment of transcription factors. This DNA modification plays an important role from the first phases of development of a new organism and is involved in genomic imprinting and chromosome X inactivation in marsupial and eutherian mammals (Liyanage et al., 2014; Shevchenko et al., 2013; Tirado-Magallanes et al., 2017). Due to its involvement in various genomic activities, an alteration of 5mC deposition along the genome has been associated to several diseases including atherosclerosis (G. Lund et al., 2004) and cancer (Kulis & Esteller, 2010). In particular, for this last one, the characterization of 5mC per patient has been proposed as a prognostic analysis (Kandi & Vadakedath, 2015; Storebjerg et al., 2018).

The 5hmC is another DNA modification that chemically differs from 5mC with a hydroxylated methyl group at the fifth carbon, it is considered also as an intermediate modification in the pathway of DNA de-methylation. Recently it has been suggested to regulate tissue specific genes (Cui et al., 2020) and to be involved in cancer (Pfeifer et al., 2014). Hence, similarly to the 5mC, 5hmC has also been proposed as a marker to define patient prognosis (S. Chen et al., 2020; He et al., 2021; Storebjerg et al., 2018; Zhu et al., 2018).

Other modifications affecting the fifth cytosine carbon (5fC and 5caC), are much rarer compared to 5mC and 5hmC and their function is still subject of investigation. They

may structurally contribute to the genome flexibility and could be associated to specific enzymes involved in gene expression regulation (Zhu et al., 2018).

The 6mA has been prevalently found and studied in prokaryotes and it has been found in eukaryotes only at low percentages, with multiple roles suggested, yet highly debated in vertebrates (Zhu et al., 2018).

Finally, the 5hmU is a DNA modification that is debated if it has a specific role as epigenetics mark or it is just an oxidative lesion (Zhu et al., 2018).

2.1.2 Protein-DNA interactions

DNA-protein interactions are non-covalent bindings between the genome and peptides that modulate biological processes (such as transcription). Their role is to guide a specific enzymatic complex in a particular genomic sequence and mediate its activity (Mobley, 2019).

The binding takes place when an aminoacidic sequence within a protein recognizes a particular DNA sequence or a particular DNA structure. This event follows biophysical and biochemical rules depending on the structure of molecules involved. Thus, a lot of efforts have been done, especially at computational level, to predict when a DNA-protein interaction will occur starting from an aminoacidic and nucleotide sequence, and consequently apply that knowledge for biomedical research purposes as drug design (Emamjomeh et al., 2019; Rohs et al., 2010).

One of the most studied cases of DNA-protein interactions is the one concerning transcription factors (TF). TFs are proteins whose role is to recruit enzymatic complexes modulating gene expression, and their binding localizes in the gene regulatory sequences as promoters and enhancers (Neidle & Sanderson, 2022). Hundreds of TFs have been identified and characterized: it has been estimated that ~8% of the genes code for TFs in the human genome and, for some of them, a role in major diseases has been found (Lambert et al., 2018; Radaeva et al., 2021). In the next paragraphs I will describe two TFs in particular, i.e. CCCTC-binding factor (CTCF) and androgen receptor (AR), because they will be relevant for other sections of this thesis.

CTCF is a protein implicated in multiple mechanisms, one of these is as TF binding the genome on a CCCTC DNA sequence (through a zinc finger domain). In mammalian genomes, more than 80000 sites where CTCF could bind were found, but

its activity as TF, regulating gene expression, is performed only on TSS regions. In particular CTCF controls the expression of tumor suppressor genes, thus mutations affecting this transcription factor have been associated to cancer (e.g. adenocarcinoma) (Debaugny & Skok, 2020; Marshal et al., 2017).

AR is a protein responsible for the androgen hormones signaling through modulation of androgen response elements (ARE) regulation. It has been found playing a role in multiple tissues and cell lines, but mainly it is known for its activity in breast and prostate cancer, where it is treated as therapeutical target (Dai et al., 2017; Labbé & Brown, 2018; Leung & Sadar, 2017).

2.1.3 Non coding RNAs

Other epigenetic mechanisms involve non coding RNAs (ncRNAs). ncRNA are molecules that, after transcription from a DNA sequence, are not translated into peptides. They represent the majority of the transcribed genome (60%) (Anastasiadou et al., 2017). They have multiple molecular features and functions, and for this reason the cataloguing of these elements is still an open challenge (Panni et al., 2020; Seal et al., 2020). In this section I will focus on the most studied ncRNA and their role in regulating genome activity.

MicroRNAs (miRNA) are ncRNA molecules 22 nucleotides (nt) long, highly conserved across species; some authors refer to them as short interference RNA (siRNA) while others distinguish between miRNA and siRNA based on their mechanism of action. At the moment more than 2000 of them have been annotated and involved in many different processes (e.g. cell proliferation, differentiation, immunity). They perform their regulatory role silencing gene expression and facilitating mRNA degradation or preventing its translation. They have been associated to diseases as cancer (Martello et al., 2010) when their silencing mechanisms on oncogenic genes is altered (Budakoti et al., 2021; Kumar et al., 2020; Matullo et al., 2016).

The lncRNAs are generally defined as non-coding RNA molecules longer than 200 nt acting on multiple biological processes. Some of them are involved in gene expression regulation (Herzog et al., 2014), others in chromatin remodeling (Csorba et al., 2014), others act as scaffold for protein complexes (West et al., 2016)) through multiple mechanisms as: increasing or decreasing chromatin accessibility, repressing miRNA

silencing or modulating the mRNA stability (Statello et al., 2020). While the studies on these group of molecules is still ongoing, the most studied ones are XIST, HOTAIR and NEAT1.

XIST is a lncRNA longer than 15Kb with a crucial role to inactivate the chromosome X in eutherian and marsupial mammal females for gene dosage compensation on sex chromosomes. It plays this function by recruiting several repressing epigenetic factors (e.g. PRC1 and PRC2) and interacting with the lamin B receptor (LBR) (Kumar et al., 2020; McHugh et al., 2015).

HOTAIR is again a lncRNA regulating genome accessibility to the transcription machinery: increasing if it interacts with PRC1 complex (promote the methylation of histone H3 at lysine 4), decreasing if it interacts with PRC2 complex (promote the methylation of histone H3 at lysine 27)(Cai et al., 2014; Kumar et al., 2020).

NEAT1 shows different epigenetics features with respect to the previously described lncRNAs. Indeed, it may be involved in maintenance of structures proximal to speckles that have been found associated to gene transcription (paraspeckles) (West et al., 2016). An alteration of this lncRNA has been observed in prostate cancer (Kumar et al., 2020; Zeng et al., 2018).

As for miRNA and other epigenetic mechanisms already described, even the alteration of the mechanisms in which lncRNA are involved could lead to major diseases as cancer. For this reason they have been suggested as therapeutical targets in treatments of some cancer types (Statello et al., 2020).

eRNA are a class of ncRNAs transcribed from regulatory regions (enhancers) that have been more recently discovered and characterized. So far, they are defined in two major groups depending on size. The shorter eRNAs are non-polyadenylated, bidirectional, non-spliced and function in cis (2D-eRNAs) while the longer eRNAs (around 150bp) are unidirectional, polyadenylated and spliced (1D-eRNA) (Natoli & Andrau, 2014; Sartorelli & Lauberth, 2020). Expression of eRNAs is commonly considered as an index of enhancer activation, but it is not clear if the eRNA itself has a role in gene regulation. Actually, it seems they could have an indirect role in gene expression controlling genome accessibility. An example of this has been described for KLK3 eRNA (KLK3e), an eRNA that reinforces the AR expression (Hsieh et al., 2014; Sartorelli & Lauberth, 2020).

Despite the debate on the gene expression control by eRNAs, they have been associated to the genome stability control: a high level of eRNAs expression could

facilitate the formation of DNA:RNA interactions that lead to error while the DNA is duplicating (Sartorelli & Lauberth, 2020).

Other studies suggest that eRNAs could be involved in formation of biomolecular condensates, thus having a role in phase separation, in particular interacting with bromodomain contacting domain protein 4 (BRD4), a protein that regulates transcription through an acetyl lysine histone reading mechanism (Rahnamoun et al., 2018).

The circular RNAs (circRNAs) are single strand RNA molecules covalently closed. Their epigenetic function is still unclear, even though it seems they can act as miRNAs sponge: i.e. by sequestering miRNAs in the nucleus (Zhou et al., 2020).

2.1.4 Chromatin

The core of epigenetic regulation is the chromatin: an ensemble of nucleic acids (i.e. DNA and RNA) and associated proteins. Chromatin has first of all a structural role: packaging the genome to make it fit inside the nucleus. Chromatin, indeed, is an eukaryotic feature that is not shared with prokaryotic organisms whose genome is much smaller and does not need such extra compaction. Chromatin does not have the same density throughout the whole genome: some regions are less dense (euchromatin or "open" chromatin regions) and other denser (heterochromatin or "closed" chromatin regions). Euchromatin regions are more accessible to enzymatic complexes and are more transcribed, while the heterochromatic ones show low transcription levels (Coleman, 2018).

The nucleosomes are fundamental components of chromatin: octameric complexes around which the DNA is wrapped. The eight proteins that together form a nucleosome are called histones, they are characterized by long aminoacidic tails, that extend from the main core and are the substrate of covalent chemical modifications including methylation, acetylation, ubiquitination and phosphorylation (Coleman, 2018; S. Zhao et al., 2021). These modifications, together with DNA methylation, are responsible for the decreased or increased accessibility of chromatin. Indeed, according to the type of histone modifications in a specific genomic region we could predict if that region is in an open or closed chromatin status. The open regions are the ones associated to the following histone modifications: histone 3 lysine 4 monomethylation (H3K4me1, associated to enhancers and promoters), histone 3 lysine 4 trimethylation (H3K4me3,

associated to active promoters), histone 3 lysine 27 acetylation (H3K27ac, associated to active enhancer and promoter regions, coexisting with H3K4me1 or H3K4me3) and histone 3 lysine 36 trimethylation (H3K36me3, associated to the gene body of transcribed genes). Whereas histone modifications associated to closed chromatin are histone 3 lysine 9 trimethylation (H3K9me3, covering broad domains up to megabases in size along the genome, especially gene poor and repetitive regions) and histone 3 lysine 27 methylation 3 (H3K27me3, covering broad domains and modulating gene expression through Polycomb complex repression activity; it has associated to multiple processes as the organism development) (Nicetto & Zaret, 2019; Y. Zhang et al., 2021). Histone modifications are sometimes called "histone marks". The ensemble of epigenetic features associated to chromatin (e.g. histone modifications, DNA methylation, chromatin accessibility) are sometimes collectively called "chromatin marks".

The structure of the chromatin along the genome is dynamic, it can change according to external stimuli, in cell differentiation programs or during different phases of cell cycle. These remodeling events are due to specific enzymatic complexes such as the chromatin writers (Biswas & Rao, 2018; Hauer & Gasser, 2017). Polycomb group (PcG) of proteins are among the most studied and known chromatin writers. Mainly two complexes are in this group: Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2). Their role is manly to negatively regulate the expression of genes during development (e.g. HOX genes) through ubiquitylation of lysine 119 of histone H2A (done by PRC1) or methylation of the lysine 27 of histone 3 (done by PRC2). Among the target of these complexes there are the bivalent genes (Bernstein et al., 2006; Mikkelsen et al., 2007): genes that maintain a mixed set of chromatin marks associated to open and closed chromatin status, i.e. the active H3K4me3 and the inactive H3K27me3 marks (Blackledge & Klose, 2021). The activity of PcG is antagonized by another group of proteins: Trithorax G (TrxG), mainly represented by Switch/sucrose non-fermentable (SWI/SNF) and complex of proteins associated with Set1 (COMPASS) that methylate the H3K4 (Blanco et al., 2020).

Chromatin readers are a different group of proteins interacting with chromatin marks. These are enzymatic complexes “reading” the chromatin modifications and mediating their effects. Bromodomain family members are an example of chromatin readers. They recognize chromatin acetylation and modulate processes such as DNA repair,

transcription and replication. Their clinical relevance has been highly studied over the past few years and they have been selected as targets for cancer treatments (Biswas & Rao, 2018; Boyson et al., 2021).

Another group of proteins that increases the chromatin dynamics are the so-called "erasers", that constitute enzymatic complexes remove modifications from histones tails. Histone deacetylase (HDAC) complexes are among the most studied of this type of proteins, that, together with bromodomain family, have been selected as targets in tumor therapies (Biswas & Rao, 2018; Boyson et al., 2021).

Epigenetic regulation of chromatin is at the base of cell identity definition in physiological processes such as development and it is crucial for the balance between pluripotency and cell differentiation (Serrano et al., 2013). Indeed, aberrant chromatin epigenetic regulation affects the genome activity leading to severe diseases (e.g. cancer). It has been observed, especially in recent years, that the chromatin structural role is not limited to compacting the genome, but it also confers rigidity to the cell and its modulation allows cell migration and extravasation (Gerlitz, 2020; Maeshima et al., 2018; Nava et al., 2020; Stephens, 2020; Stephens et al., 2019).

Recently, chromatin has been studied for its role in genome compartmentalization and its connections to liquid-liquid phase separation dynamics (Larson et al., 2017), however this is still matter of debate. The liquid-liquid phase separation (LLPS) is an event that happens when, in a solution, there are two or more regions with different molecular concentrations, such difference enucleates areas with different properties in the solution; these area are called liquid phases (Alberti et al., 2019). Several known membraneless organelles (e.g. speckles and paraspeckles, polycomb bodies, nucleoli or transcription factories) are supposed to be formed through the LLPS mechanism (Rippe, 2022). In this context chromatin is associated to LLPS with an interdependent mechanism, where chromatin structure is able to influence LLPS and LLPS influence the chromatin structure (Rippe, 2022).

In this paragraph I focused the attention on eukaryotic cells as prokaryotes do not have a chromatin structure based on histone proteins binding DNA to pack it and interact with other molecules to regulate the genomic activities inside the nucleus. Anyway, it has been found that even in bacteria the genome is located in a specific subcellular region (nucleoid) (Berlatzky et al., 2008; Fisher et al., 2013; Gray et al., 2019; Hadizadeh Yazdi et al., 2012; Kellenberger et al., 1958; Kleckner et al., 2014; Mason & Powelson, 1956; Umbarger et al., 2011) and, similarly to eukaryotic cells, it is

organized and its activities are modulated by specific proteins called nucleoid associated proteins (or “histone-like” proteins) (Dillon & Dorman, 2010; Dorman & Deighan, 2003; Drlica & Rouviere-Yaniv, 1987).

2.1.4.1 Chromatin solubility

According to (Schmidl et al., 2015), “chromatin can be considered as a polyanion-polycation complex” with biochemical properties varying across different portions of the genome. These features have been investigated through the alteration of the surrounding chemical and physical conditions, using fractionation techniques to separate chromatin fractions on the basis of their solubility (Ausio et al., 1986; Ausio, Seger, et al., 1984; Chalkley & Jensen, 1968; Fulmer et al., 1981; McCarthy & Duerksen, 1971).

The separation of different chromatin portions is usually achieved firstly digesting the DNA with an enzyme (MNase in most of the cases) and then treating the sample with buffers containing high NaCl concentrations. In addition, some studies have been done using buffers containing divalent salts (Ausio, Borochov, et al., 1984; Borochov et al., 1984) and their effect on chromatin solubility has been reported.

One of the main differences among distinct chromatin regions it has been found in the ionic strength. Indeed, the use of different amount of salt for chromatin extraction led to a different material distribution across fractions (Ausio et al., 1986; X. W. Guo & Cole, 1989). In particular, it has been reported that on chicken erythrocytes at 0.15 mM of NaCl part of the chromatin is insoluble but its amount decreases with the increase of the NaCl concentration (Ausio et al., 1986). In addition to the salt, even the pH seems to have a major effect on chromatin solubility. Indeed, in a basic medium, a bigger portion of chromatin tend to be soluble than in an acid one (X. W. Guo & Cole, 1989), this seems particularly relevant as variation of pH inside the nucleus have been reported even among different phases of cell cycle (Gerson & Burton, 1977; Gillies & Deamer, 1979).

These studies on chromatin solubility allowed identifying histones as the major responsible of chromatin biochemical properties heterogeneity (Ausio et al., 1986; X. W. Guo & Cole, 1989) together with level of acetylation (Perry & Chalkley, 1981a) and the ratio between gene density and DNA fragment size (Beacon & Davie, 2021).

Interestingly, the differences in chromatin composition do not influence solubility at variable temperatures. In fact, it has been found that in a range between 0 and 37 Celsius degrees the amount of material extracted through chromatin fractionation has the same distribution across fractions (Ausio et al., 1986).

2.1.5 3D chromatin architecture

Another way for the cell to regulate the genome activity is through the arrangement of chromatin in the tree dimensional (3D) space. Indeed, during interphase the chromosomes are not in a mitotic condensed status, but instead they are spread all across the nuclear space arranged in a functional and well-defined architecture. This spatial arrangement of the genome can be observed at multiple scales: starting from few kilobases (i.e. nuclesome clutches (Ricci et al., 2015) and nanodomains (Szabo et al., 2020)) to several megabases (topological associated domains, TADs; compartments) (Jerkovic´ & Cavalli, 2021).

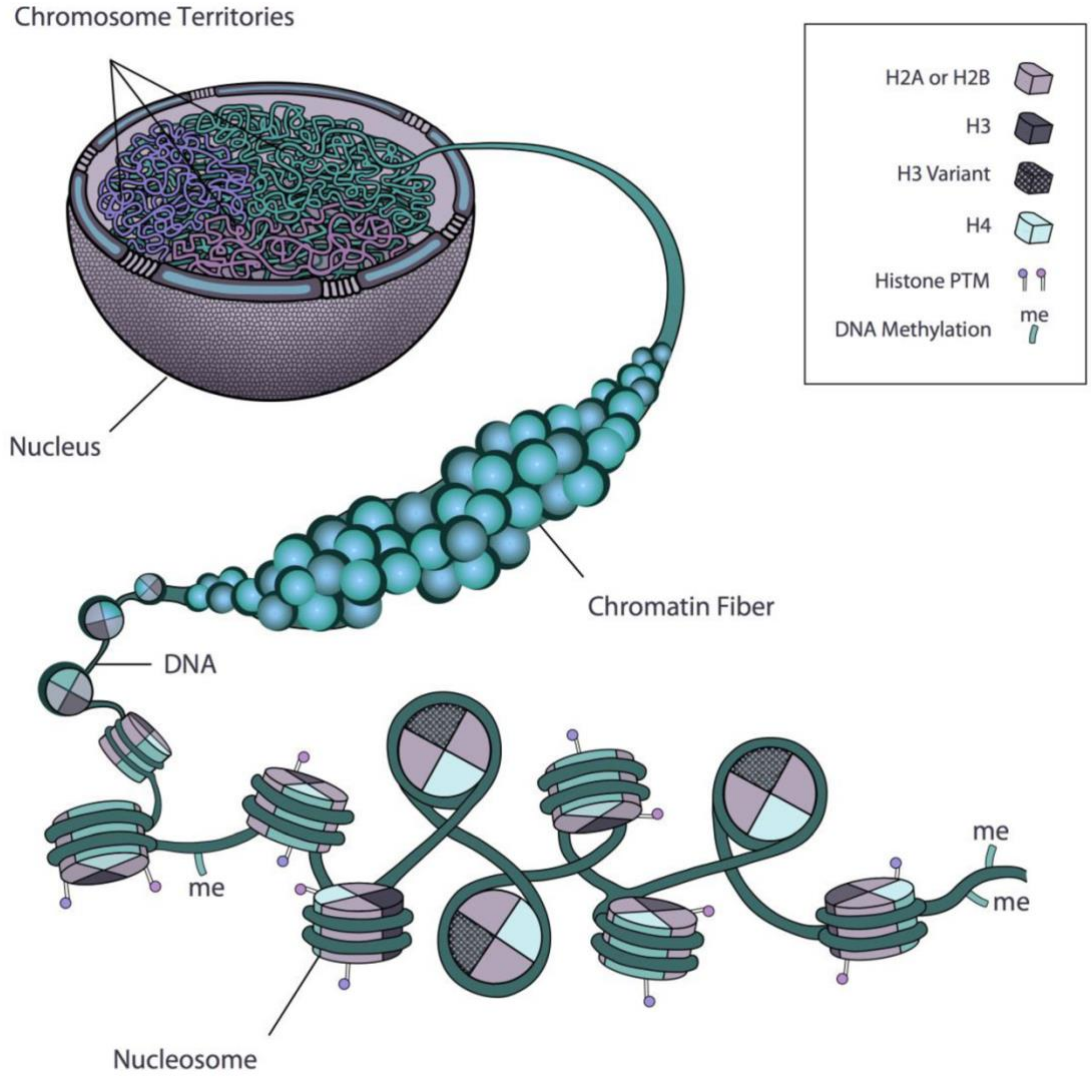


Fig1: *Schematic representation of how the genome is assembled in the nucleus during interphase, figure from (Rosa & Shaw, 2013).*

Chromosomes clutches (Jerkovic´ & Cavalli, 2021; Ricci et al., 2015) consist of a small group of nucleosomes spatially closed and separated from the rest of the genome by a nucleosome free DNA sequence and they span a genomic region of 1-2 Kb. Their name derives from the fact that they look like an "egg clutch". The number of nucleosomes composing the clutches determine the accessibility of RNA polymerase II to that region: less nucleosomes correspond to higher accessibility. According to (Ricci et al., 2015) chromosome clutches define the transition between the pluripotency to differentiated cells, indeed they observe that embryonic stem cells have prevalently low density clutches. Clutches density level increases according to the differentiated state of the cell (the more a cell is differentiated the denser the clutches will be).

Denser nucleosome clutches can cluster together in broad genomic domains (~10-100Kb) called nanodomains. Due to their level of compaction, these regions are poorly accessible by enzymatic complexes, thus they behave as heterochromatin regions with low accessibility. The mechanisms that keep in contact these chromatin regions are the nucleosome-nucleosome interactions (Kalashnikova et al., 2013; Nozaki et al., 2017). This mechanism is antagonized by the histone acetylation (Jerkovic´ & Cavalli, 2021; Kantidze & Razin, 2020; Nozaki et al., 2017).

Another genome 3D structure is the one represented by chromatin loops which results from folding of chromatin to allow physical contact in the 3D space between DNA sequences that would be distant along the linear genome sequence. These structures could be found at multiple scales and according to the dimension could have different roles. Loops bring into contact enhancers and promoters, allowing the transcription of the genes associated to these regulatory elements (Visel et al., 2009). There are several mechanisms that allow the formation of these structures and there is still not a consensus around it. The one around which there is the highest consensus and has been better characterized, is the loop extrusion through DNA-cohesin interaction. There are several models that describe the mechanism, in all of them a dimer of cohesin constituent proteins (shaped as a ring) contacts a specific region of DNA that start to be introduced in the hole between the two dimer subunits. The loop continues to elongate until it is stopped by a CTCF molecule bound to the DNA in a convergent configuration, with one convergent CTCF binding site per direction is required to stop

loop extrusion (Fudenberg et al., 2016; Sanborn et al., 2015). Other mechanisms proposed for the formation of smaller functional loops could be due to several factors as: local high torsion force on double helix (DNA supercoiling), liquid-liquid phase separation, alternance between nanodomains and H3K27ac regions and molecular crowding (Kantidze & Razin, 2020; Nozaki et al., 2017).

When the chromatin loops reach the size of hundreds kilobases with a genome-genome interaction frequency inside the loop higher than outside we talk about topologically associated domains (TADs). These structures do not directly put in proximity enhancer and promoters, but instead can force two distant sequences of DNA to get in contact (Gabriele et al., 2022) and, importantly, they insulate the domains between the two anchor points from the rest of the genome, drastically reducing their interactions. Usually, these regions are defined at the extremity by two directionally convergent CTCF binding sites (Rao et al., 2014; Vietri Rudan et al., 2015), but it is possible even with different CTCF dispositions (de Wit et al., 2015; Y. Guo et al., 2015) or, in some organisms as Drosophila, they could be delimited by other factors than CTCF (Sexton et al., 2012).

At a broader resolution compared to TADs there are the 3D architectures called compartments. They have been defined as genomic domains of megabases of size that can be characterized by transcription level inside the domain: it has been defined A compartments the ones with high transcription level and B compartments the ones with low/non transcribed regions (Lieberman-Aiden et al., 2009). With the application of the new technologies and deeper sequencing in chromatin 3D conformation studies the model has been further complicated. Indeed, it has been proposed that compartments could be divided in sub-compartments of smaller sizes (Rao et al., 2014) and could be characterized in more than just two states based on transcription levels. In this type of representation, the compartments have been defined on the basis of interchromosomal interactions (Johnstone et al., 2020; Rao et al., 2014, 2017; Xiong & Ma, 2019) or, more recently, a method based on intrachromosomal interactions has been proposed (Liu et al., 2021).

Despite the fact that usually compartments and TADs are represented on a different hierarchy level, it has been found that compartments could antagonize TADs formation and *vice versa* (L. Xie et al., 2022).

On a larger scale, there are the chromosome territories (CT), that come from the observation that each chromosome occupies a discrete area inside the nucleus (Cremer

& Cremer, 2001). The disposition of the chromosome has important implications for the activity of the chromosome itself: indeed, the nuclear center is more active than the rest (e.g. it has been observed that chromosomes in the center are replicated earlier than the others) (Fabiao de Lima et al., 2022).

Membraneless subnuclear structures like nucleoli or the nuclear lamina are also important for the chromosome/genome position inside the nucleus as further described in the next section.

2.1.6 Interactions with subnuclear structures

The nucleolus is a subnuclear structure where most of the ribosomal RNAs transcription take place and it is in general the nuclear compartment with the highest transcription rate (Belmont, 2022). Alteration of ribosomal gene expression in the nucleolus has been described as a marker of premature aging (Buchwalter & Hetzer, 2017). The nucleolus has been associated even to a role of chromatin architecture, indeed it has been found that chromatin contact the nucleolar periphery (Schöfer & Weipoltshammer, 2018).

Nuclear speckles are small bodies that are localized in the center of the nucleus for most cells. The speckles are enriched in RNA polymerase II (RNA pol II) and have been proposed as splicing factor compartments (Denker & De Laat, 2016; Galganski et al., 2017).

Cajal bodies are another type of elements present in the nucleus, usually they are between 1 and 10 per cell. They have been shown to have a high concentration of non-ribosomal protein and molecules involved in their maturation process. Similar to speckles they are connected to transcriptional activity (Belmont, 2022; Machyna et al., 2013; Mario Cioce, 2005; Ogg & Lamond, 2002).

The nuclear lamina, differently from the previous subnuclear structures described so far, is not a single compact feature, but, instead, a network of proteins mainly surrounding internally the nuclear membrane and, at low percentage, could be found spread inside the nucleus (de Leeuw et al., 2018; Fawcett, 1966; Paschal & Kelley, 2013).

Among the proteins composing the lamina the most prominent ones are Lamin A, Lamin B1, Lamin B2 (only present in mammals) and Lamin C (transcribed from the same gene of Lamin A). All together these proteins form the backbone of intermediate filaments on which other proteins are connected. Some of these proteins anchor

specific genomic portions to the nuclear periphery. Indeed, the concept of lamina associated domains (LADs) has been coined, referring to heterochromatic regions with sizes in the order of megabases (Belmont, 2022; Briand & Collas, 2020; de Leeuw et al., 2018; Shimia et al., 2015; R. Wu et al., 2006). The importance of the lamina for cell physiology is attested by the multitude of diseases associated to mutations of genes for lamina proteins (e.g. Hutchingson-Gilford progeria and Emery-Dreyfus dystrophy) that will be further described in the next sections. Knockout of LMNA/C and Lamin B receptor (one of the proteins acting as a bridge between lamina and the chromatin) has been shown to lead to major rearrangements in the nucleus with heterochromatin relocating from the periphery to the center, and euchromatin to periphery (Belmont, 2022; Falk et al., 2019; Kang et al., 2018; Osmanagic-Myers & Foisner, 2019).

2.2 Epigenetics alterations in diseases

After a general description of all the mechanisms that regulate the genomic processes, I am next describing two examples in which epigenetics alterations lead or contribute to a pathological phenotype.

2.2.1 Laminopathies

Laminopathies are a group of diseases that are caused by mutation of one of the genes coding for lamina proteins or for other proteins associated to the nuclear lamina. Among the pathologies directly associated to lamina proteins, two derive from *LMNB1* or *LMNB2* mutations, while all the others are connected to *LMNA* (Donnaloja et al., 2020).

From a clinical point of view the laminopathies are classified in four groups according to the type of disorders they are causing and the tissue they affect. We can then distinguish: myopathies, if they interest skeletal and cardiac muscles (e.g. Emery-Dreyfuss muscular dystrophy, EDMD); lipodystrophy diseases, if they affect the adipose tissue distribution in the body; neuropathies, if they affect neural tissues; and the ones connected to multisystemic disorders as premature aging (e.g. Hutchinson-Gilford progeria syndrome, HGPS) (Béroud et al., 2005; Worman & Bonne, 2007).

Despite the fact that the lamina deficiency could be due to a great variety of mutations (just for *LMNA* gene more than 500 mutations have been discovered), these are connected to a relatively small number of pathologies (around 20). These pathologies

show a similar set of symptoms, thus suggesting that alterations of lamina network affect the same mechanisms (Crasto & Di Pasquale, 2018).

From different studies, two different models have been developed to explain the role of lamina in laminopathies. Some of them show how the lamina is necessary to the cell to maintain the correct chromatin conformation. An example of this genome-lamina interaction has been observed in HGPS, where the disruption of LAD domains has been found correlated with a relocalization of H3K27me3 marked regions and, in the late passages of primary fibroblasts lines derived from patients, an altered compartmentalization (McCord et al., 2013). Thus, according to this finding, an altered epigenomic regulation by the lamina has been suggested as cause of the aberrant phenotype observed in lamina diseases (Sullivan et al., 1999; Zwerger et al., 2015).

Other studies show how tissues affected by lamina alterations become more sensitive to mechanical stress, justifying the fact that laminopathies principally affect muscle tissues (as the ones more subjected to mechanical stress). These works support the idea that the laminopathies are a consequence of increased cell fragility due to lamina mutations (Lammerding et al., 2006; Perovanovic et al., 2016).

To connect the two models, lamina has been proposed as the sensor of mechanical stress and then involved in mechano-transduction activity (Donnaloja et al., 2020). This has been supported by studies like (Dworak et al., 2019) where it has been demonstrated the role of GTPases (enzymes involved in mechano-transduction pathway) in HGPS (Kelley et al., 2011; Terada, 2019).

2.2.2 Cancer

Cancer is a disease that has been defined as one or more cells of an organism that have lost homeostatic regulation and have acquired a high proliferative rate. Starting from this definition and based on observations on different cancer types, multiple hallmarks have been identified. At first six cancer features have been described: self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan & Weinberg, 2000). Now, after some years of this first attempt to identify the hallmarks of cancer, other lists

have been produced with other elements included: epigenetic alteration is a common one across different lists (Senga & Grose, 2021).

Indeed, through epigenetic alterations, cancer can immediately respond to immune system and drug treatments. The alteration of epigenetic has been considered so important for cancer development that has been suggested that it could take place without any genomic mutation (Hanahan, 2022; Senga & Grose, 2021). Known epigenetics alterations in cancer mostly concern methylation (e.g. hyper- and hypo-methylation), histone modifications (e.g. H3K27me3 deregulation) and DNA-ncRNA interactions (e.g. microRNA deregulation). These have been associated to specific activities that characterize cancer progression, as the invasion of other tissues (metastasis) (Hanahan, 2022; J. W. Park & Han, 2019).

All these statements are based on observations of common features in different cancer types, but, while some features are shared across tumours, others are more cancer type-specific. In the next section I will present findings of specific epigenomic alterations on prostate cancer, one of the tumour types with the highest incidence in the population and subject of our studies, as described later.

2.2.2.1 Prostate cancer

Prostate cancer is the most frequent diagnosed tumour in men and the second most lethal (Siegel et al., 2022). Its outcome and development is hormone driven. Indeed, prostate cancer has been found mostly associated to androgen receptor misregulation. Intra-tumour heterogeneity is another feature associated to prostate cancer: it has been found that multiple genetically different cancer foci could co-exist in the same prostate. Another feature of this disease, that adds lethality potential, is its propensity to metastasize: different tissues have been reported as prostate cancer metastasis target (e.g. bones, brain, liver), although the mechanism remains unclear. Hints of a connection with epithelial mesenchymal transition (EMT) has been proposed (G. Wang et al., 2018; Yamada & Beltran, 2021).

From a clinical point of view the tumour diagnosis requires several biopsies collected from the patient's prostate that are then visually inspected by a pathologist. The doctor determines then the level of cancer aggressiveness according to several tissue features (described at first by Dr. Gleason in 1974) and associates a score to the patient that represents the cancer grade. It is called Gleason score and its range goes from 1 (less

severe) up to 5 (most severe). For each biopsy two scores are produced: the first taking into account the most evident cancer features and the second taking into account the less evident ones (Carlsson & Vickers, 2020; G. Wang et al., 2018).

At the moment, several treatments are available for this type of cancer and even prostate surgical removal could be an option, but tumour plasticity and metastasis can make these solutions ineffective (Bluemn et al., 2017; Hu et al., 2015; G. Wang et al., 2018).

Then alternative approaches have been suggested both to perform an earlier diagnosis and more precise grading, as well as to treat the tumour before metastasis. In this context, epigenetic mechanisms have been considered as alternative approaches to improve diagnosis, as they can play important roles in prostate cancer. This is the case of the alteration of gene methylation that has been found causing activation of EMT genes and promoting metastasis (Pistore et al., 2017). Even histone modifications have been supposed to be important for prostate cancer development (Ke et al., 2009; Q. Wang et al., 2009). Indeed, several recurrent mutations have been found in genes coding for chromatin remodelers (Huang et al., 2012). In addition, direct evidences of association between histone remodeling and prostate cancer aggressiveness have been found for H3K9me2 and H3Ac, whereas H3K4me1 has been found as predictor of PSA recurrence (Ellinger et al., 2010). Even H3K27me3 has been included in the list of markers related to prostate cancer (Ke et al., 2009). Indeed it has been proposed as therapeutical through PRC2 silencing in castration resistance prostate cancer (Sparmann & Van Lohuizen, 2006; Xu et al., 2012).

Besides studies focusing on specific alterations of histones or methylated/demethylated regions, it has even been observed that prostate cancer progression correlates with a general increase of chromatin accessibility (Urbanucci et al., 2017). It was speculated that chromatin accessibility alterations could actually affect the nuclear conformation. This is important because nuclear conformation is a parameter to determine the cancer grade. One of the drivers of this mechanism of "chromatin relaxation" has been identified in BRD4 (an H3K27ac reader), thus it has been proposed to use JQ1 (a BRD4 signaling pathway inhibitor) as a possible prostate cancer treatment (Rustøen Braadland & Urbanucci, 2019; X. Shi et al., 2018; Urbanucci et al., 2017).

While these studies on chromatin rearrangements of prostate cancer have a high potential to better understand and treat the disease, they are complicated by the scarcity

of *in vivo* material and they are often performed on cell lines (Adli & Bernstein, 2011; Pomerantz et al., 2015).

2.3 Genome-wide technologies to study epigenetics

In line with the high variety of epigenetic mechanisms, a large number of experimental techniques have been developed to study epigenetics features. In this third section of the introduction, I will go through some of the most commonly used genome-wide high-throughput sequencing based techniques to study epigenomic profiles and their limits.

2.3.1 Main techniques applied on epigenomics studies

2.3.1.1 Whole genome bisulfite sequencing

The whole genome bisulfite sequencing (WGBS) is the most comprehensive technique to study DNA methylation of cytosines.

It consists in the treatment of the DNA with sodium bisulfite, a reagent that converts cytosines unmethylated into uracil, which is then converted into thymine when sequences go through polymerase chain reaction (PCR) amplification. Methylated cytosines are instead unchanged by the treatment. After this operation the genome is sequenced. From the comparison between the bisulfite treated sample and a reference it is possible to determine if there was a methylated cytosine every time there is a mismatch. I.e., every time cytosine is expected according to the reference, but a thymine is instead detected in the WGBS sequence, we can assume that in that point there was no methylation in our sample.

This method is considered the most comprehensive one to detect 5mC genome-wide at single nucleotide resolution, and it has been adopted as standard for 5mC detection by many different consortia. Unfortunately, it is expensive and could not be used to discriminate between 5mC and 5hmC, furthermore it cannot be used to detect other DNA modifications (Rauluseviciute et al., 2019).

2.3.1.2 DNase sequencing

WGBS probes cytosine methylation without providing directly any information about the level of genome activity. Other methods, instead, focus on the chromatin accessibility and from their results it is possible to understand if a region is silenced

or not. One group of these techniques is the one leveraging endonucleases (enzymes that cleavage the DNA), as DNase-sequencing (DNase-seq).

DNase-seq is a method based on the partial digestion of chromatin by DNase I. The peculiarity of this enzyme is that it is more efficient in sequences where the chromatin is accessible (e.g. transcriptionally active gene regions). Thus, performing an experiment where the entire genome is digested and sequenced by standard next generation sequencing (NGS) techniques, it will be possible to determine the more accessible regions by distinguishing the regions with higher coverage from the regions with lower coverage.

DNase-seq is a powerful technique for studying regulatory elements in multiple cell types and states, but it requires, at least in the first developed protocols, a high number of cells that limits the applicability in case of small samples. In addition, dealing with DNase I is often problematic due to its efficiency varying in different cell types, different manufacturers and even different lots (A. Chen et al., 2018; Liu et al., 2019).

2.3.1.3 Assay for Transposase-Accessible Chromatin sequencing

To overcome the limitations affecting DNase-seq, another technique has been developed: the Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) (Buenrostro et al., 2013). With this method, a modified version of the Tn5 transposase is used to study the high accessible regions of chromatin. The mechanism is similar to the one explained for DNase-seq, with the main difference that the Tn5 cuts the chromatin and inserts a specific sequence that is used as primer for high-throughput sequencing libraries preparation.

This technique is easier to perform than DNase-seq, and importantly it can be applied on a smaller number of cells and avoid the variability problems due to DNase (A. Chen et al., 2018; Grandi et al., 2022).

2.3.1.4 Chromatin Immuno-Precipitation sequencing

Another genome-wide technique commonly used to study chromatin accessibility or specific epigenomic mechanisms (i.e. the ones involving DNA-protein interactions) is the Chromatin Immuno-Precipitation sequencing (ChIP-seq) assay. This method is based on the use of antibodies to target elements anchored to chromatin. Then the DNA is either digested by endonucleases or sonicated, and the fragments associated

to the targeted protein are enriched by immunoprecipitation by binding to a specific antibody. The enriched fragments are then sequenced. Analysing the data, enrichment of reads sequenced could be found in the genomic regions where the target protein was anchored; in other words, it is possible to identify regions of target protein anchoring through detection of reads enrichments (P. J. Park, 2009).

This technique could be used to study either specific transcription factors or histone modifications, making possible investigation of specific epigenetic mechanisms and, through histone distribution analysis, even the chromatin accessibility (Nakato & Sakata, 2021).

Indeed, combining multiple ChIP-seq experiments targeting different histone modifications, it is possible to obtain information about chromatin status in each genomic region. To reach this aim, the ChIP-seq data collected by ENCODE (Feingold et al., 2004) or ROADMAP (Roadmap Epigenomics Consortium et al., 2015) projects are commonly used. This operation of combining different ChIP-seq experiments is done at bioinformatic level with software like ChromHMM (Ernst & Kellis, 2012), that segment the genome and assign to each bin a status according to the combination of histone marks ChIP-seq enrichment on it (Nakato & Sakata, 2021).

While ChIP-seq is a broadly used method, it is affected by several limitations: in the case of the study of a specific transcription factor its mechanism should be known *a priori*, more in general the standard protocol requires a high level of starting biological material and the immune precipitation step is known to be bias prone due to the antibodies (Grandi et al., 2022).

2.3.1.5 Chromosome conformation capture derived techniques

Until now I described techniques developed to investigate protein binding and chromatin accessibility. Instead, in this section I am focusing on a group of techniques developed to study the genome-genome interactions: the chromosome conformation capture techniques. Among them four of the most known and used techniques are chromosome conformation capture (3C), circularized 3C (4C), carbon copy 3C (5C) and chromosome capture followed by high-throughput sequencing (Hi-C). All of these share the same initial steps where the chromatin is cross-linked, then digested and re-ligated followed by reverse cross-linking. After that, for the 3C protocol the fragments are sonicated and amplified with a semi-quantitative PCR (Dekker et al., 2002). While

for the 4C, fragments are circularized after a secondary digestion and sequenced (Z. Zhao et al., 2006). Instead 5C fragments are hybridized with a mix of oligonucleotides that are then ligated, amplified with a reverse PCR and then sequenced (Dostie et al., 2007). Hi-C fragments ends are instead biotinylated and then ligated, finally the ligated fragments are pulled down and sequenced (Denker & De Laat, 2016; Lieberman-Aiden et al., 2009).

The different procedures lead to a different level of investigation of chromatin contacts along the genome. With the 3C technique we have information over interaction of two DNA sequence that we targeted *a priori*. Using 4C is possible to know which are the interacting sequences of a specific target region. Instead 5C investigates on the interaction among multiple sequences. Finally with Hi-C we have information about all the contacts inside the genome (Denker & De Laat, 2016; Kempfer & Pombo, 2019).

Among these techniques Hi-C is the most used one. Since its first publication it led to important discoveries about the 3D genomic architecture, e.g. TADs (Dixon et al., 2012) and compartments (Lieberman-Aiden et al., 2009).

However, it is very expensive and requires a high amount of starting biological material for experiments intended to have a high resolution of interaction frequency (10kb resolution and below) (Denker & De Laat, 2016). Recently new derived techniques have been developed such as as Micro-C (Krietenstein et al., 2020) and capture Hi-C (Mifsud et al., 2015) that allow increasing data resolution starting from similar amount of cells. Plus, it is important to remark that it gives information just about interaction frequency, but not if this interaction led to an increase or decrease of genome accessibility.

2.3.1.6 Other genome-wide techniques to study chromatin architecture

The ones described until now are some of the most used and established techniques to study epigenetics, but in the last few years new techniques have been developed to address specific problems and overcome several existing common limitations (Kempfer & Pombo, 2019).

For example, an alternative to ChIP-seq to study DNA-proteins interaction is the DNA adenine methyltransferase identification (DamID), a technique that is based on the comparison of the methylated Adenine in the genome of two cell samples, one

transfected with DNA adenine methyltransferase (Dam) fused with a target protein and one just with Dam. As results of the experiment the sample transfected with the Dam plus the fusion protein will show an enrichment of methylated Adenine in the genomic regions where the peptide was localized compared to the sample where only the Dam was transfected (Van Steensel et al., 2001; Van Steensel & Henikoff, 2000; F. Wu et al., 2016). This technique has the advantage that it does not require the use of antibodies and reduces the related artifact generation problem: indeed, if not enough specific, antibodies can bound to non-target elements. It has been successfully applied to investigate for example genome-LaminB1 interactions (Aughey et al., 2019). Another interesting method to study DNA-protein interactions is the Tyramide signal amplification sequencing (TSA-seq), it consists in the introduction in the cell nucleus of an antibody combined with the horseradish peroxidase (HRP) (Finn et al., 2019; G. Wang et al., 1999), an enzyme that catalyses biotin-free radicals reaction. This implies that, around the area targeted by the antibody, a gradient of biotin radicals is formed. It could be detected through staining or through analysis of DNA biotinylated fragments (Y. Chen et al., 2018). Thus, this technique allows combining information coming from sequencing and microscopy. It provides information about spatial distance between genomic regions and subnuclear structures (e.g. speckles) (Y. Chen et al., 2018; L. Zhang et al., 2021).

For what concerns techniques that are complementary or alternative to Hi-C for studying the 3D conformation of the genome, some of the most recent ones are split-pool recognition of interactions by tag extension (SPRITE) (Quinodoz et al., 2018) and genome architecture mapping (GAM) (Beagrie et al., 2017) two methods to identify multi-contacts (i.e. involving more than two loci) while avoiding the ligation biases of Hi-C; and GPSeq, a method that provides a radial map of chromatin disposition (Girelli et al., 2020).

2.3.2 Current technologies limits

All the described technologies to study epigenetics have been successfully applied in multiple works increasing our knowledge of cell biology and mechanisms that lead to disease. They all are affected by common limitations that reduce their applicability.

One major problem of many techniques is the amount of starting biological material required, as it is the case for ChIP-seq or Hi-C. This is a relevant problem especially

if we are working with human tissue samples, where the material available is often scarce as their collection requires invasive procedures for the patient. In addition, not all methods could be applied to tissues or primary cell lines even if the input is abundant. Indeed, some techniques require cells transfection (as DamID).

From a technical point of view, further limitations exist when use of biochemical modifications of samples are required during the analysis that could produce artifacts. The use of antibodies could be problematic as well, namely polyclonal antibodies that have different affinity for different targets that complicates the comparison of independent experiments even if targeting the same protein.

Another type of problems affecting epigenomic methods are practical ones. Some protocol could be very expensive, thus a great number of replicates is harder to afford, especially for small laboratories. e.g. WGBS and Hi-C require a high sequencing depth to reach high resolution. Other elements to take into account in evaluating the applicability of a method is its protocol complexity and time requirement that could compromise the reproducibility or complicate the scalability of the approach to many samples, respectively.

Finally, it is important to remark that most of the techniques can focus on one epigenetic feature: C-technologies can just provide information on genome interaction without specifying the chromatin status, DNase-seq and ATAC-seq only focus on open chromatin regions and ChIP-seq can study one target at a time.

2.4 SAMMY-seq: a new method to study epigenomics status by chromatin biochemical properties

Considering the limitations affecting the common epigenomics study techniques, in collaboration with Chiara Lanzuolo's lab we developed a new method: the Sequential Analysis of Macro-Molecules Accessibility sequencing (SAMMY-seq) (Sebestyén et al., 2020). It is based on the concept that different chromatin regions have different biochemical properties and levels of solubility, according to their epigenetic status. Different chromatin fractions are isolated based on their distinct biochemical properties and sequenced to identify the corresponding regions along the genome. We developed three distinct protocols based on the same mechanism that could be summarized in four main steps. At first, the nuclear membrane is permeabilized and soluble proteins are collected. Then the DNA is digested using one or more

endonucleases (according to the version of the protocol we are applying) and we collect the material exiting from the nuclei. Then we dissolve ionic interactions between the DNA and histones through NaCl and again we collect the material in another fraction. Finally, we treat what remains with urea to dissolve the remaining protein and membrane pellets and collect the supernatant. Then we separately sequence the collected DNA and proceed to bioinformatic analysis. Comparing the number of reads we got in each genomic region across fractions we assess the epigenetic properties affecting solubility of each region along the genome.

This method can be easily (and cheaply) performed on virtually any type of sample, even tissue, it does not require the use of antibodies or cross-linking and could be applied even with scarce material, allowing the characterization of both open and closed chromatin regions.

3 Materials and methods

3.1 Samples and data sources

3.1.1 Fibroblasts primary cells culture and treatments

Primary fibroblast cell lines were cultured in DMEM High glucose with glutamax supplemented with 15% FBS and 1% Pen/Strep. HGADFN167 (HGPS167)-8 years old, HGADFN169 (HGPS169)-8 years old, HGADFN188 (HGPS188)-2 years old human dermal fibroblasts derived from HGPS patients were provided by the Progeria Research Foundation (PRF). AG08498 (CTRL001) -1-year-old and AG07095 (CTRL002)-2-years-old human dermal fibroblasts were obtained from the Coriell Institute. Foreskin fibroblast strain #2294 (CTRL004)-4 years old was a gift to Chiara Lanzuolo from the Laboratory of Molecular and Cell Biology, Istituto Dermopatico dell'Immacolata (IDI)-IRCCS, Rome, Italy, while control dermal fibroblast CTRL013-13 years old was kindly provided by the Italian Laminopathies Network.
For JQ1 vs DMSO experiments, cells were treated 48h with Brd-4 inhibitor JQ1 (Sigma, SML0974), resuspended at 5 mM concentration in DMSO and then diluted in culture medium to 0.5 µM.

3.1.2 C2C12 culture

Mouse myoblast C2C12 cell line (ATCC) was cultured in DMEM High glucose supplemented with 10% FBS.

3.1.3 Prostate biopsies collection and processing

Our analyses involved cohorts of chemo-naïve patients followed-up in the Urology's Division of Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico (Milan). The patients were selected according to PSA analysis and digital rectal examination. The biopsies have been collected throught the transrectal ultrasound-guided systematic sampling of the prostate (TRUS biopsy). Fourteen multiregional cores TRUS prostate biopsy were collected from all enrolled patients for the diagnosis and one more core was taken for this study from the same surgery. The institutional ethics committee board authorized this study (authorization n. 1063). All specimens were obtained after the patients had provided written informed consent following the ethical principles of biomedical researches on biospecimens. To ensure optimal tissue recovery in terms of quality, fresh surgical prostate biopsy specimens were placed in

ice-cold saline buffer and then directly transported to the laboratory within 1 hour from sample collection. Tissues were then immediately processed to preserve the intranuclear genomic and protein architecture of prostate cells nuclei. Samples selected for next generation sequencing (NGS) consist of 18 fresh biopsies divided on the basis of the histology and the spatial distribution of the positive cores in two different groups: 10 prostate cancer (PCa) biopsies with histologically confirmed prostate cancer and 8 histologically negative biopsies from patients who had no cancer in any biopsy core. The enrolled patients were men aged 62-87, with a median age of 72 years.

All the clinical data were provided by the Urology division maintaining the patient confidentiality.

The biopsy specimen was stored at 4-8°C for maximum 6 hours before dissociation, avoiding its freezing. For the digestion step, given the little size of the needle biopsy tissues (typically 20-30 mg) we determined the dissociation condition to ensure optimal cell populations recovery. Briefly, tissue was cut into small pieces (~1 mm) in ice-cold PBS with autoclaved surgical scissors and resuspended in 1 ml HBSS (Gibco, 14025) containing 200 units of collagenase type I (Life Technologies, 17018-029) plus 67 μg DNase I (Sigma-Aldrich, 10104159001) each ~10 mg of minced tissue. Digestion was carried out in a water bath at 37 °C shaking vigorously for 10 sec every 5 minutes; at its completion (usually after 1 hour of digestion) single cells were washed once by topping up to 2 ml with RPMI + 10% FBS and centrifugated at 300g. The cell pellet was resuspended in RPMI + 10% FBS and dispersed by passing through a 75μm cells strainer, followed by an additional wash of the filter with RPMI + 10% FBS. Finally, the cells were centrifugated and resuspended in 1 ml of ice-cold PBS for counting with a hemacytometer. The entire procedure of digestion takes approximately 3 hours.

3.1.4 Literature data

3.1.4.1 Public genomics datasets for skin fibroblasts

We collected publicly available ChIP-seq datasets from the following sources:

Lamin A/C ChIP-seq from (McCord et al., 2013) (SRR605493, SRR605494, SRR605495 and SRR605496), Lamin B1 ChIP-seq from (Sadaie et al., 2013) and H3K27ac, H3K9me3, H3K36me3, H3K4me3 from Roadmap

Epigenomics (fibroblast sample: E055) (Roadmap Epigenomics Consortium et al., 2015).

Hi-C data used has been produced in (Finn et al., 2019), we downloaded the already count normalized matrix (4DNFIMDOXUT8) in .mcool format from 4DN data portal (Dekker et al., 2017).

3.1.4.2 Public genomics datasets for C2C12 cells

We collected publicly available datasets from the following sources:

H3K9me3 ChIP-seq from (Singh et al., 2015), H3K36me3, H3K79me2 and H3K79me3 from ENCODE/Caltech series GSE36023 (Dunham et al., 2012) and Lamin B1 Dam-ID-seq from (F. Wu & Yao, 2013) (SRX195288, SRX195289, SRX195290 and SRX195291).

3.1.4.3 Public genomics datasets for prostate cells and tissue

The ChIP-seq data of prostate gland have been downloaded from the following datasets available on ENCODE (Dunham et al., 2012): ENCSR763IDK (H3K27ac), ENCSR642CSX (H3K4me1), ENCSR748RBT (H3K4me3), ENCSR690CSD (H3K27me3) and ENCSR133QBG (H3K9me3). These data have been downloaded as raw data and analyzed as described in the next sections.

3.2 Experimental protocols

3.2.1 SAMMY-seq

3.2.1.1 SAMMY-seq 3f protocol performed on fibroblast cells

Chromatin fractionation was carried out as described in (Marasca et al., 2016; Sebestyén et al., 2020) with minor adaptions. Briefly, 4 million fibroblasts were washed in PBS 1×, and resuspended in cytoskeleton buffer (CSK: 10 mM PIPES pH 6.8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM $MgCl_2$; 1× protease Inhibitors by Roche Diagnostics; 1 mM PMSF) supplemented with 1 mM DTT and 0.5% Triton X-100. After 5 min at 4 °C the cytoskeletal structure was separated from soluble proteins by centrifugation at 900×g for 3 min, and the supernatant was labeled as S1 fraction. The pellets were washed with an additional volume of cytoskeleton buffer. Chromatin was solubilized by DNA digestion with 10U of RNase–free DNase (Turbo DNAse; Invitrogen AM2238) in CSK buffer for 60 min at 37 °C. To stop

digestion, ammonium sulfate was added in CSK buffer to a final concentration of 250 mM and, after 5 min at RT samples were pelleted at 2350×g for 3 min at 4 °C and the supernatant was labeled as S2 fraction. After a wash in CSK buffer, the pellet was further extracted with 2 M NaCl in CSK buffer for 5 min at 4 °C, cen- trifuged at 2350×g 3 min at 4 °C and the supernatant was labeled as S3 fraction. This treatment removed the majority of histones from chromatin. After 2 washing in NaCl 2 M CSK, the pellets were solubilized in 8 M urea buffer to remove any remaining protein component by applying highly denaturing conditions. This remaining fraction was labeled as S4. For the scaled-down experiment, samples of 250,000, 50,000, or 10,000 cells were treated analogously, except for a reduction of buffers volumes to half of those used for 10 million cells and a decrease of DNase to 8U.

3.2.1.2 SAMMY-seq 3f protocol performed on MCF10DICIS.com cells

Three distinct biological replicas of control and RAB5A expressing MCF10.DCIS.com monolayers were obtained in the context of a collaboration with Dr Giorgio Scita as detailed in (Frittoli et al., 2021) and processed for chromatin fractionation as described in (Marasca et al., 2016; Sebestyén et al., 2020), with minor adaptions. Briefly, 3 million cells were washed in PBS 1X, and extracted in 600 µl of cytoskeleton buffer (CSK: 10 mM PIPES pH 6,8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1X protease Inhibitors by Roche Diagnostics; 1 mM PMSF) supplemented with 1 mM DTT and 0,5% Triton X-100. After 10 min on wheel at 4°C the cytoskeletal structure was separated from soluble proteins by centrifugation at 900g for 3’ at 4°C, and the supernatant was labeled as S1 fraction. The pellets were resuspended with 600 µl of cytoskeleton buffer, put 10 min on wheel at 4°C followed by centrifugation at 900g for 3’ at 4°C. Chromatin was solubilized by DNA digestion with 25U of RNase–free DNase (Turbo DNAse; Invitrogen AM2238) in 100 µl of CSK buffer for 60 min at 37°C. To stop digestion, ammonium sulphate was added in CSK buffer to a final concentration of 250 mM and, after 5’ in ice samples were pelleted at 900g for 3 min at 4°C and the supernatant was labeled as S2 fraction. The pellets were resuspended with 200 µl of CSK buffer, put 10 min on wheel at 4°C followed by centrifugation at 3000g for 3 min at 4°C. The pellet was further extracted with 100 µl of CSK buffer with 2M NaCl for 5 min at 4°C, centrifuged at 2300g 3 min at 4°C and the supernatant was labeled as S3 fraction. This treatment removed the

majority of histones from chromatin. The pellets were washed twice with 200 µl of CSK buffer with 2M NaCl, put 10 min on wheel at 4°C followed by centrifugation at 3000g for 3 min at 4°C. The pellets were solubilized in 100 µl of 8M urea buffer for 10 min at room temperature to remove any remaining protein component by applying highly denaturing conditions. This fraction was labeled as S4. DNA was extracted from S2, S3 and S4 fractions. All fractions were quantified and analyzed by SDS-PAGE and immunoblotting. Anti-tubulin alpha (Sigma T5168, mouse 1:10000), H3 (Abcam ab1791, rabbit 1:4000), Beta-Actin (Santa-Cruz sc1616, rabbit 1:4000), were used as primary antibodies. HRP-conjugated secondary antibodies were revealed with the ECL chemiluminescence kit (Thermo Fisher Scientific). After incubation 90 min at 37°C with 6 µl of RNAse cocktail (Invitrogen AM2286) followed by 150 min at 55°C with 40 mg of Proteinase K (Invitrogen, AM2548), DNA was extracted by standard phenol/chloroform extraction, precipitated and resuspended in 15µl milliQ H2O. The S2 fraction was further purified with PCR DNA Purification Kit (Qiagen, 28106). After Qubit HS DNA quantification then samples were evaluated by capillary electrophoresis (Agilent 2100 Bioanalyzer) and then sonicated with a Covaris M220 Focused-ultrasonicator using screw cap microTUBEs with the parameters: water bath 20°C, peak power 30.0, duty factor 20.0, cycle/burst 50, duration: 125 seconds for S2 and S3 fractions, 150 seconds for S4 fraction. The DNA profiles were checked again by capillary electrophoresis (Agilent 2100 Bioanalyzer).

Libraries for SAMMY-Seq: for fractions S2, S3 and S4 obtained from chromatin fractionation procedure, at least 2.5 ng DNA were used to generate an indexed library (Kapa HyperPrep kit; Roche KK8504, KK8727). Indexed DNA libraries were purified (AmpureXP, Beckman, A63881), quantitated (Qubit dsDNA HS Assay, Q32851), checked for size distribution on Agilent Bioanalyzer 2100 (DNA HS kit, Agilent, 5067-4626) and normalized for pooling. 1% PhiX control was added to the sequencing pool, to serve as a positive run control. Sequencing was performed in SR mode (1x75nt) on an Illumina NextSeq550 platform, generating at least 30 million SR reads per sample. Experiment was performed using biological triplicates.

3.2.1.3 SAMMY-seq 4f protocol performed on fibroblast cells

Chromatin fractionation was carried out as described in (Marasca et al., 2016) with minor adaptions. Briefly, 3 million fibroblasts were washed in PBS 2×, and

resuspended in cytoskeleton buffer (CSK: 10mM PIPES pH 6.8; 100 mM NaCl; 1 mM EGTA; 300mM Sucrose; 3 mM MgCl2; 1× protease Inhibitors by Roche Diagnostics; 1 mM PMSF) supplemented with 1 mM DTT and 0.5% Triton X-100. After 10 min on a wheel at 4 °C the cytoskeletal structure was separated from soluble proteins by centrifugation at 900×g for 3 min, and the supernatant was labelled as S1 fraction. The pellet was washed with an additional volume of CSK buffer and then resuspended in 100ul of fresh CSK buffer. Chromatin was solubilized by DNA digestion with 25U of RNase–free DNase (DNAseI; Invitrogen, AM2222) for 60 min at 37 °C. To stop digestion, ammonium sulfate was added in CSK buffer to a final concentration of 250mM and, after 5 min on ice, the sample was pelleted at 900×g for 3 min at 4 °C and the supernatant was labelled as S2 fraction. After a wash in CSK buffer, the pellet was further extracted in 100ul of CSK buffer supplemented with 2M NaCl. The sample was incubated on a wheel for 10 min at 4 °C and then centrifuged at 2350×g 3 min at 4 °C, and the supernatant was labelled as S3 fraction. This treatment removed the majority of histones from chromatin. After two washings in 2M NaCl CSK, the pellet was solubilized in 100ul of 8M urea to denaturate any remaining protein component, and this fraction was labeled as S4. For the scale-down experiment, a sample of 10,000 cells was treated analogously, except for a reduction of buffer volumes to half of those used for 3 million cells and a decrease of DNase to 12.5U.

For DNA analysis, each fraction was diluted in TE to 200ul, incubated with 6ul of RNAse cocktail (Invitrogen, AM2288) (90 min at 37 °C) and 40ug of Proteinase K (Invitrogen, EO0491) (150 min at 55 °C). DNA was extracted by standard phenol/chloroform extraction, precipitated and resuspended in nuclease-free water (50ul for S2 and S2.2, 15µl for S1, S3 and S4). On the next day, S2 was additionally purified using PCR DNA Purification Kit (Qiagen, 28106) and suspend in 50µL of nuclease-free water. DNA fragments in S2 were separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) with the ratio 0,95: smaller fragments were conserved as S2S fraction and larger fragments as S2L fraction. Both were suspended in 20 ul of nuclease-free water and then S2L was brought to 15ul using a centrifugal vacuum concentrator. Samples were then quantified with Qubit 1X dsDNA HS (Invitrogen, Q33231), transferred to microTUBE-15 AFA Beads Screw-Cap (Covaris, 004078) and sonicated in a Covaris M220 focused-ultrasonicator (peak power 30, duty factor 20, cycles/burst 50 – S2L and S2L.2 for 125sec, S3 and S4 for 175sec). The DNA profiles were finally checked by capillary electrophoresis (Agilent

2100 Bioanalyzer with 2100 Expert Software). For sequencing, DNA libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB, E7645L) with Unique Dual Index NEBNext Multiplex Oligos for Illumina (NEB, E6440S) and then sequenced using an Illumina NovaSeq 6000 System according to manufacturer's instructions at the IEO Genomic Unit in Milan.

3.2.1.4 SAMMY-seq 4f protocol performed on prostate biopsies derived cells

Tissue-digested single cell suspension was performed with minor adaptations to the protocol described in (Sebestyén et al., 2020). Cells were counted, washed in cold PBS and resuspended in cold cytoskeleton buffer CSK: 10 mM PIPES pH 6,8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1X Protease Inhibitor Cocktail (Roche, 04693116001); 1 mM PMSF (Sigma-Aldrich, 93482) supplemented with 1 mM DTT and 0,5% Triton X-100. After 10 minutes on a wheel at 4°C, samples were centrifugated for 3 minutes at 900g at 4°C and cytoplasmic and nucleoplasmic components were collected as S1 fraction. Pellets were washed for 10 minutes on the wheel at 4°C with an additional volume of the same buffer. Chromatin was then digested by using DNase I (Invitrogen, AM2222) (25U for more than 100 thousand cells and 12.5U for less than 100 thousand cells) in CSK buffer (10 mM PIPES pH 6.8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1 mM PMSF, with protease inhibitors) for 60 min at 37°C. To stop digestion, ammonium sulphate was added to samples to a final concentration of 250 mM and, after 5 min on ice, samples were pelleted at 900g for 3 min at 4°C and the supernatant was collected as S2 fraction. After a wash in CSK buffer, the pellet was further extracted with 2M NaCl in CSK buffer for 10 min at 4°C, centrifuged at 2300 g 3 min at 4°C and the supernatant was conserved as S3 fraction. This treatment removed the majority of histones from chromatin. Pellets were washed twice for 10 minutes on the wheel at 4°C with double volume of high salt CSK buffer. Finally, after 3 minutes of 3000g centrifugation at 4°C, pellets were solubilized in in 8M urea for 10 minutes at RT to denature any remaining protein and to dissolve membranes, and labelled as S4. Factions were stored at – 80°C until DNA extraction.

Fractions were diluted 1:2 in TE buffer (10mM TrisHCl pH 8.0, 1 mM EDTA) and incubated with 61,5 U of RNAse cocktail (Ambion, AM2286) at 37° for 90 minutes, followed by 40µg of Proteinase K (Invitrogen, AM2548), at 55° for 150 minutes.

Genomic DNA was then isolated using phenol/chloroform (Sigma-Aldrich, 77617) extraction followed by a back extraction of phenol/chloroform with additional volume of TE1X. DNA was precipitated in 3 volumes of cold ethanol, 0.3M sodium acetate and 20ug glycogen (Ambion AM9510) for 1 hour on dry ice or overnight at -20°. Dry pellets were resuspended in 50 µl (S2) or 15 ul (S3 and S4) of nuclease-free water and incubated at 4°C overnight. On the next day, S2 was further purified using PCR DNA Purification Kit (Qiagen, 28106) and separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) with the ratio 0,90/0,95 to obtain smaller fragments conserved as S2S (< 300 bp) and larger fragments labelled as S2L (> 300bp) fractions. Both were suspended in 20 ul of nuclease-free water and then reduced to 15ul using a centrifugal vacuum concentrator. S2L, S3 and S4 fractions were sonicated in a Covaris M220 focused-ultrasonicator using screw cap microTUBEs (Covaris, 004078) to obtain a smear of DNA fragments peaking at 150-200 bp (waterbath 20°C, peak power 30.0, duty factor 20.0, cycles/burst 50). Prostate tissues: 150 seconds for S2L and 175 seconds for S3 and S4; Fibroblasts: 125 seconds for S2L and S3, 150 seconds for S4. Fractions were quantified using Qubit 4 fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and run on an Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were created from each sample using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and Unique Dual Index NEBNextMultiplex Oligos for Illumina (NEB, E6440S); libraries were then qualitatively and quantitatively checked on Bioanalyzer 2100. Libraries with distinct adapter indexes were then multiplexed and, after cluster generation on FlowCell, sequenced for 50 bases in paired-ends mode on an IlluminaNovaSeq 6000 instrument at the IEO Genomic Unit in Milan. A sequencing depth of at least 35 million reads was obtained for each sample.

3.2.1.5 SAMMY-seq 6f protocol performed on fibroblast cells

For the 6f protocol on fibroblasts, a sample of 3 million cells was treated as described for 4f on fibroblasts above, except for the insertion of a second step of digestion: after the recovery of the S2 fraction, the pellet was washed in CSK buffer, digested with 25U of TURBO DNAse (Invitrogen, AM2238) in 100ul of CSK buffer for another 60min at 37°C followed by stopping of the reaction with ammonium sulfate,

centrifuged at 900xg for 3min at 4°C and the supernatant was recovered as S2.2 fraction. After this step, the protocol was resumed as described above.

3.2.2 ChIP-seq

3.2.2.1 ChIP-seq performed on fibroblast cells

Cells were cross-linked with 1% HCHO for 12 min at room temperature, lysed and chromatin sheared. IP was performed overnight on a wheel at 4 °C with 2.4 μg of H3K9me3 antibody (ab8898, Abcam) or 2μg of H3K27me3 (07-449, Millipore). The following day, antibody-chromatin immunocomplexes were loaded onto Dynabeads Protein G (Invitrogen 10004D). The bound complexes were washed once in Low Salt Solution (0,1% SDS, 2 mM EDTA, 1% Triton X-100, 20 mM Tris pH 8, 150 mM NaCl), once in High Salt Solution (0,1% SDS, 2 mM EDTA, 1% Triton X-100, 20 mM Tris pH 8, 500 mM NaCl), once again in Low Salt Solution and once in Tris/EDTA 50 mM NaCl. Crosslinking was reversed at 65 °C overnight in Elution Buffer (50 mM Tris pH 8, 20 mM EDTA, 1%SDS), DNA was purified by standard phenol/chloroform extraction, precipitated and resuspended in 30 μl of 10 mM Tris pH 8. ChIP efficiency was tested by qPCR reactions, performed in triplicate using SYBR select master mix (Invitrogen, 4472908) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems).

For H3K9me3, ChIP-seq libraries for sequencing were created using the automation instrument Biomek FX (Beckman Coulter), while for H3K27me3 ChIP-seq they were created using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) with NEB-Next Multiplex Oligos for Illumina (NEB). Libraries were then qualitatively and quantitatively checked using dsDNA HS Assay kit (Invitrogen, Q32854) on a Qubit 2.0 fluorometer and High Sensitivity DNA Kit (Agilent Technologies, 5067–4627) on an Agilent Bioanalyzer 2100. Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, were sequenced for 50 bp in the single read mode on a HiSeq 2000 sequencer at the IEO Genomic Unit in Milan.

3.2.3 RNA-seq

3.2.3.1 RNA-seq performed on prostate biopsies derived cells

Ten thousand prostatic cells form the biopsy digestion step were stabilized in 200µl of 1Thioglycerol/Homogenization Solution of the Maxwell® RSC miRNA Tissue Kit (Promega, AS1460) and stored frozen at -80°C for later total RNA automated purification using Maxwell® RSC 48 Instrument (Promega, AS8500) according to manufacturer's instructions. The automated purification on the Maxwell® RSC Instrument ensures the higher RNA yield from this low number of cells. Total RNA was quantified by Qubit 4 fluorometer with Qubit RNA HS Assay Kit (Invitrogen, Q32852) and assessed by Agilent 2100 Bioanalyzer using Agilent RNA 6000 Pico Kit (Agilent, 5067-1513) to inspect RNA integration. For each sample, 1 ng of total RNA was used to construct strand specific RNAseq library with SMARTer Stranded Total RNA-Seq Kit - Pico Input (Takara, 634487). The yield and quality of the libraries were evaluated on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). RNAseq libraries were sequenced on the Illumina NextSeq™ 550 system at the sequencing facilities of Humanitas or Division of Pathology of Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinic in Milan to a minimal target of 40 million for 75 bases in paired-ends mode.

3.3 Prostate tissue evaluation

3.3.1 FACS analysis of prostate biopsies

To quantify the relative frequency of cell types for each prostate biopsy, ten thousand cells form the digestion step were stained, acquired on a BD FACS Canto™ Flow Cytometer and analysed with FlowJo software in the INGM FACS Facility. To avoid unspecific binding, antibodies were incubated with PBS-BSA 1% for 30 minutes at 4°C. TO-PRO®-3 stain is used to assess cell viability, tissue resident leukocytes are identified by CD45+/CD326-, epithelia are identified by CD326+/CD45-, and stroma are double negative (CD326-CD45-). The followed table displays information on antibodies used for flow cytometry.

Specificity	*Clone*	*Fluorochrome*	*Identifier*	*Specificity*
TO-PRO®-3	-	APC	Thermofisher, T3605	Viability
CD326	EBA-1	FITC	BD,	Epithelia

			347197	
CD45	H130	PB	Biolegend, 982306	Leukocyte

3.3.2 Histological evaluation of prostate biopsies

One third portion of each prostate tissue specimen was embedded in Killik (Bio-Optica, 05-9801), immediately frozen in precooled isopentane (MilliporeSigma, 277258) and stored at -80°. OCT embedded biopsy cores (one per patient) were serially sectioned, at 10 µm thick in a cryostat at -20°C. Ten glasses per patient, containing multiple sections representing distinct regions of the same tissue were prepared in parallel and stored at -80°, along with a corresponding hematoxylin and eosine stained slide. Hematoxylin and Eosin staining was performed using H&E Staining Kit (ab245880). The hematoxylin-eosin-stained slides were reviewed by an expert genitourinary pathologist to assign a Gleason score according to the International Society of Urinary Pathology grading system. The histopathological architectural pattern was also confirmed by comparing the Gleason score between our prostate biopsy tissue sample and the nearby biopsy used for clinical diagnosis of the patients. All the clinical data were provided by the Urology division maintaining the patient confidentiality.

3.4 Bioinformatics analysis

3.4.1 Trimming and sequencing

Sequencing reads have been trimmed using Trimmomatic (v0.39) (Bolger et al., 2014) using the following parameters for SAMMY-seq and literature data: 2 for seed_mismatch, 30 for palindrome_threshold, 10 for simple_threshold, 3 for leading, 3 for trailing and 4:15 for sliding window. The sequence minimum length threshold of 35 has been applied for all data except for H3K9me3, H3K4me3, H3K79me2, H3K79me3 datasets of C2C12 where 30 has been used. As clip file has been used the Trimmomatic provided dataset “TruSeq3-SE.fa” (for single end) and “TruSeq3-PE-2.fa” (for paired end). Only for MCF10DICIS.com SAMMY-seq data, all reads were cropped to 75 bp reads length (if longer) by setting the crop option of Trimmomatic (v0.39) to 75.

After trimmed reads have been aligned using BWA (v0.7.17-r1188) (H. Li & Durbin, 2009) setting –k parameter as 2 and using as reference genome the UCSC hg38 one for fibroblasts and the UCSC mm10 for the C2C12 (in both cases only canonical chromosomes have been taken into consideration).

The alignment duplicates have been marked with Picard (v2.22; https://github.com/broadinstitute/picard) MarkDuplicates option. And then filtered using Samtools (v1.9) (H. Li et al., 2009), in addition we filtered all the reads with mapping quality lower than 1.

Each lane has been analysed separately and then merged at the end of the process.

3.4.2 Genome coverage calculation

From the alignment of output a coverage analysis has been performed using Deeptools (v3.4.3) (Ramírez et al., 2016) bamCoverage function. For these analyses the genome has been binned at 50bp, the reads extended up to 250 bp and RPKM normalization method has been used.

For mouse data has been considered a genome size of 2308125349 bp (value suggested in the Deeptools manual https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html), from the analysis have been excluded regions known to be problematics in term of sequencing (the list has been downloaded from the ENCODE portal, https://www.encodeproject.org/files/ENCFF547MET/@@download/ENCFF547MET.bed.gz).

For human data has been considered a genome size of 2701495761 bp (value suggested in the Deeptools manual https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html), from the analysis have been excluded regions known to be problematics in term of sequencing (the list has been downloaded from the ENCODE portal, https://www.encodeproject.org/files/ENCFF356LFX/@@download/ENCFF356LFX.bed.gz).

3.4.3 SAMMY-seq fraction and ChIP-seq IP over INPUT comparison

The comparison between SAMMY-seq fractions or ChIP-seq IP over INPUT comparisons has been performed using the SPP (v1.16.0) (Kharchenko et al., 2008) R

(v3.5.2) library. The reads have been imported from the bam files using the "read.bam.tags" function, then they have been filtered using "remove.local.tag.anomalies" and finally the comparison has been performed using the function "get.smoothed.enrichment.mle" setting "tag.shift = 0" and "background.density.scaling = TRUE".

3.4.4 Identification of chromatin domains

3.4.4.1 Domains calling on SAMMY-seq 3f

We performed relative comparisons of SAMMY-seq fractions within each sample using EDD (version 1.1.15) (E. Lund et al., 2014), optimized to call very broad enrichment domains. EDD was originally designed for lamin ChIP-seq data, comparing IP and input samples. As SAMMY-seq data also shows broad enrichment regions, we used EDD to select significantly enriched SAMMY-seq domains by comparing less accessible fractions to more accessible ones (S4vsS2 comparison) in each sample, with the following options: –gap-penalty 25–bin-size 50–write-log-ratios–write-bin-scores and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions (H. Li & Wren, 2014). We also changed the required_fraction_of_informative_bins parameter to 0.98. We used the same set of parameters for the downsampled data, except changing –binsize to 100 for 50% down-sampling or 200 for 25% down-sampling. To account for the lower sequencing depth, in the SAMMY-seq scale-down experiment (results reported in Fig. 3), we run EDD (version 1.1.19) with –gap-penalty 25 and –bin-size 250 parameters, and in late-passage cells, we used –gap-penalty 45 and –bin-size 200.
We used bedtools (version 2.25.0) (Quinlan & Hall, 2010) and the bedtools jaccard to calculate Jaccard Index for the various overlap analyses.

3.4.4.2 Domains calling on fibroblast ChIP-seq data

Peaks for enrichment profile analysis shown in fig25a were calculated over H3K27ac ChIP-seq data downloaded from Roadmap (see 2.1.5.1) using MACS2 software (Y. Zhang et al., 2008) using the "callpeak --broad" algorithm and setting the "—broad-cutoff" parameter to 0.1.

3.4.4.3 Domains calling on MCF10DICIS.com ChIP-seq data

Heterochromatin (H3K9me3 enriched) domains were defined using the EDD (v1.1.19) software with parameters (binsize = 200 Kb and gap penalty = 25) processing the filtered bam files obtained as described above. The "required_fraction_of_informative_bins" parameter was set to 0.98. The unalignable regions were defined as for genome coverage (see 2.3.2).

3.4.4.4 Domains calling on prostate ChIP-seq data

EDD was used for calling domains even on prostate ChIP-seq, in this case the –binsize chosen was of 150 Kb and the –gap-penalty was setted at 25 while other parameters used for fibroblasts were kept, we excluded from this analysis the regions excluded for genome coverage analysis.

3.4.5 Metaprofile analysis

The metaprofile analyses shown multiple figures have been done using DeepTools (v3.4.3). The metaprofile matrix was calculated with the "computeMatrix" command of DeepTools, using as regions of interest the heterochromatin domains obtained from ChIP-seq experiment as signal the SAMMY-seq values (comparison S4vsS2 for fig 8b and S2S genome coverage for fig25a). In addition, the "skipZeros" option was added to remove regions with no coverage. The metaprofile plots have then been represented using the "plotProfile" tool of DeepTools using as input the previously created matrix.

3.4.6 RNA-seq data preprocessing and differential expression analysis of prostate data

The length reads used for sequencing is 75 bp. RNA samples were first trimmed with Trimmomatic (0.39) removing the specific adapters (ILLUMINACLIP:Picov2smart-PE.fa) used in the Stranded Total RNA-Seq Kit v2 Pico Input, primer dimers, and low-quality bases at the beginning and at the end of the reads (trimmomatic PE phred33 LEADING:3 TRAILING:3 SLIDINGWINDOWD:4:15 MINLEN:36). The reads quality was checked with Fastqc (Version 0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). STAR (V. 2.7.0f_0328) (Dobin et al., 2013) was used to index (STAR --runMode

genomeGenerate) the Human Genome (GENCODE Release 39, GRCh38 primary assembly genome) (Schneider et al., 2017) and to align sequenced reads in Paired-End mode (--readFiles R1.FASTQ R2.FASTQ) on the previous indexed reference. Multimapping reads and PCR duplicate unique mappers are marked in the final output (--bamRemoveDuplicatesType Unique) and unaligned reads stored in a different file (--outReadsUnmapped Fastx). To count reads that fall on genes, has been used as a reference a GTF file with RefSeq annotation downloaded from UCSC (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/genes/hg38.ncbiRefSeq.gtf.gz).

This file was further processed to remove non canonical and mitochondrial chromosome, selected only curated genes (NM, NR) and finally split in protein coding (NM) and noncoding (NR) files. Reads count was carried out with HTSeq-count (V. 0.13.5) on bam files (previously generated by STAR) using as a feature the union of all exons in a gene. Must be specified if the library is strand specific and if the first read (R1) is mapped on the same strand or the opposite strand of the gene. Using Takara kit, the second read (R2) yields sequences sense to the original RNA (htseq-count -s reverse). The reads that align to more than position in reference are discarded (htseq-count --non-unique none). The full matrix with raw count reads for each sample were loaded in R 3.6.1 and normalized using DESeq2 (Anders & Huber, 2010) median of ratios (https://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-10-r106).

Differential expression analyses are then performed with DESeq2 (V. 1.26) (Love et al., 2014) using Wald test and the Benjamini and Hochberg method, (correction for multiple tests) to compute p-values and adjusted p-values, respectively. These analyses were performed comparing one group of healthy samples against a group with all tumor samples. Subsequently, the group of tumors was split according to SAMMY-seq results. Thus, differential analysis was done for both comparisons (healthy group versus tumor group 1, healthy group versus tumor group 2). After differential expression we computed for every gene, a score based on log2 fold change and adjusted p-value (-log10(p-adjusted) * sign fold change). This score is representative of the differential expression significance and magnitude for each gene. A list was then made and ranked from significant and highest positive fold changes to significant and lowest negative fold changes.

3.4.7 Gene set enrichment analysis of prostate RNA-seq data

The ranked list was used as input to perform the Gene set enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) with following parameters: Number permutations: 1000; Collapse: No; Enrichment Statistic: classic; Max size: 500; Min size:15. For the GSEA the hallmark gene set (Fraser et al., 2017) has been used.

3.4.8 Gene ontology analysis of prostate differentially enriched genes

The gene ontology analysis (GO) has been performed loading the differentially expressed genes on enrichR (E. Y. Chen et al., 2013; Kuleshov et al., 2016; Z. Xie et al., 2021).

3.4.9 Genome coverage and comparison data visualization

The visual representation of tracks was performed thanks to the Gviz R library (v1.26.5) (Hahne & Ivanek, 2016). The track profile was calculated using the function "DataTrack" (the input file was imported using the function "import" of the rtracklayer library) and plotted using the function "plotTracks" setting the value "window = 1000". Line plots were drawn setting the parameter type as 'a' and overlayed using the function "OverlayTrack"; instead, mountain plots were obtained setting the parameter type as "polygon". Extra elements of these plot as chromosome ideogram (on top) and genome axis were plotted respectively using the functions "IdeogramTrack" and "GenomeAxisTrack". The analysis was performed identically for all the datasets.

3.4.10 Correlation analysis

The correlation analysis has been performed using "cor" R (v3.5.2) base function with "method = Spearman". The comparisons have been imported in the R session using the function "import" of the rtracklayer (v1.42.2) (Lawrence et al., 2009) library. Then the files have been binned using the function "tileGenome" and the correlation have been performed per chromosome across the two comparisons in analysis. The values of correlation obtained per each chromosome have been then summarized in one value describing the genome-wide samples correlations through a weighted mean, where the weight of each chromosome correspond to its length.

3.4.11 Chromatin compartmentalization analysis of SAMMY-seq 6f fibroblast data

3.4.11.1 Hi-C matrix loading into R environment

The Hi-C matrix from (Finn et al., 2019) has been imported in R (3.6.0) using the "contact_mat_processing" function from Calder 1.0 (Liu et al., 2021) setting the binsize to 50 Kb.

3.4.11.2 Computation of eigenvector for chromatin compartments analysis

Compartment calling process started with the combination of SAMMY-seq fractions in a "pair wise correlation matrix". This operation took, as input, the fractions (see genome coverage calulation, section 1.3.2) rebinned at 50 Kb (in the same way described for the correlation analysis), the analysis is done per chromosome. The data are imported in R (3.6.0) using the function "import" of the rtracklayer (v1.42.2) library. Then, fractions are assembled in an N x M matrix where N is the number of the genomic bins and M is the number of SAMMY fractions (i.e. six); so each value in the matrix corresponds to the coverage value of a fraction in a specific bin. Starting from this matrix the pair wise correlation matrix is built calculating the Spearman correlation among bins. The bins not present in the Hi-C matrix have not been taken into account to calculate this matrix.

On this matrix the principal component analysis (PCA) is performed and the first eigenvector has been used to define the compartments.

In particular compartment A have been defined as the the genomic regions associated to a positive eigenvector score while the opposite for the B ones.

3.4.12 Chromatin compartmentalization analysis of SAMMY-seq 4f prostate data

3.4.12.1 SAMMY-seq solubility distance matrix computation

Compartment calling process started with the combination of SAMMY-seq fractions in a "solubility distance matrix". This operation took, as input, the fractions (see genome coverage calulation, section 1.3.2) rebinned at 50 Kb (in the same way described for the correlation analysis), the analysis is done per chromosome. The data are imported in R (4.1.0) using the function "import" of the rtracklayer (v1.42.2)

library. Then, fractions are assembled in an N x M matrix where N is the number of the genomic bins and M is the number of SAMMY fractions (i.e. four); so each value in the matrix corresponds to the coverage value of a fraction in a specific bin. Starting from this matrix the solubility distance matrix is built calculating the Euclidean distance among bins, for this operation the values describing each bin across fractions are used as bin coordinates (the R base function "dist" have been used).

3.4.12.2 Computation of eigenvector for chromatin compartments analysis

Eigenvector for Hi-C matrix or SAMMY-seq solubility matrix is calculated following the Calder procedure (Liu et al., 2021). At first the pairwise correlation is performed two times consecutively on the matrix (for this operation Calder "fast_cor" function has been used) and then arctangent transformation is applied, producing a new matrix (A). Then on A the subcompartment boundaries are calculated using the function "generate_compartments_bed". A values are then summarized performing the mean of the values between boundaries previously calculated, producing a new matrix (B). B column values are then compared with the B values in X column before assessing 1 if the value is higher than the one in the column before, 0 in the other cases; this operation is repeted for X values 1, 2, 3 and 4. The four resulting matrix are then combined in the T matrix. On the T matrix twice the pairwise correlation ("fast_cor" Calder function) is applied and the on the values the arctangent is performed. On this matrix a principal component analysis is performed obtaining than the eigenvector of the first principal component that has been presented in results.

From the previously calculated eigenvector a binary tree is calculated performing recursively k-means analysis with "centers = 2" on it. From the tree two cluster are created: the one distributed subsequently the first branch on the left and the ones on the right. According to which cluster contain more bins overlapping genes A compartments are defined (the remaining instead are defined B compartments).

4 Results

In this section I am reporting the results obtained with SAMMY-seq in different experimental settings and with different protocol variants. Indeed, three different protocols of SAMMY-seq have been developed that have different peculiarity and have been applied in different projects. This section is then divided in three main subsections corresponding to different projects and SAMMY-seq protocols.

4.1 Detection of heterochromatin rearrangements in disease and physiological conditions by SAMMY-seq

Despite the vast landscape of techniques that study epigenetics, heterochromatin is still far to be fully characterized. In particular, reliable techniques to study lamina chromatin interactions in primary cell samples are few and affected by technical biases. To fill this gap, in collaboration with Dr. Lanzuolo's lab, we developed SAMMY-seq, a technique based on the sequential extraction of different chromatin fractions according to their biochemical properties that is not affected by common technical problems characterizing other related epigenomics techniques.

In this section I am going to present the major features of this method, how it has been validated on fibroblast and its contribution to characterize epigenetic alterations in diseases as laminopathies and cancer, with a focus on heterochromatic regions.

4.1.1 Description and validation of the SAMMY-seq method, pros and cons compared to previous reference methods to detect LAD and heterochromatin dynamics

The first version of SAMMY-seq is based on the sequencing of three chromatin fractions (3f), hereafter named the SAMMY-seq 3f protocol. In this technique cells are resuspended treated at first with Triton for nuclear permeabilization and elution of soluble elements inside the cell, collecting at the end of the treatment the S1 fraction. Then the DNase (Turbo) is added to fragment the DNA and collect the most soluble genomic regions after the digestion (S2 fraction collection). After that, high salt concentration is added to the sample to dissolve ionic interactions between DNA and histones (S3 fraction collection). Finally, the sample is treated with urea and all remaining proteins and membranes pellets are dissolved, then the material is collected

in the S4 fraction. Once performed the protocol, the 3 fractions containing genomic DNA (S2, S3 and S4), are sequenced and analysed (fig2).

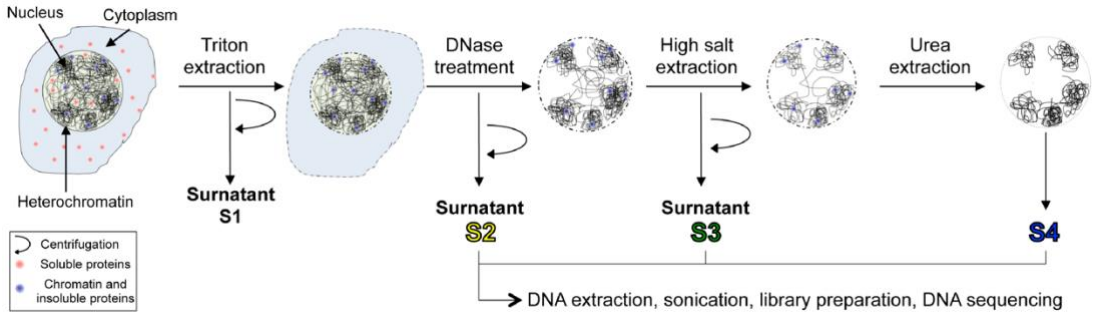


Figure 2: SAMMY-seq protocol. *Once cells have been isolated, they are treated with Triton and the surnatant is collected in the fraction S1. Then the sample is treated with the DNase and the mix of protein and DNA is collected in fraction S2. After that NaCl is added and again the output is collected in a fraction, S3. Finally, urea is added and the remaining material deriving from the cells is collected in the fraction S4. The DNA contained in S2, S3 and S4 is then extracted, sonicated, library is prepared and loaded on a flow-cell for sequencing (separately). Figure adapted from (Sebestyén et al., Nature Communications 2020).*

From the bioinformatic point of view, we performed explorative analyses of the sequencing data aligning them to a reference genome and looking at reads distribution along the genome. To validate that our results are in line with literature data, we performed correlation analyses between SAMMY-seq outputs and other epigenomics techniques data, such as ChIP-seq.

Our first protocol run test was done on human fibroblast primary cells. From the analysis of reads distribution along the genome we found that the enrichment signal of S3 and S4 fractions fall in close chromatin regions (fig3a). The enrichment of S2 instead is not immediately evident looking at the reads distribution, but correlating at genome wide level the S2 with the chromatin marks from ChIP-seq we obtain slightly negatively correlated values with closed chromatin marks (H3K9me3, Lamin A/C) and positively correlated with open (H3K36me3, H3K4me3) ones (fig3b). These observations support the fact that in SAMMY-seq we are collecting reads associated with closed chromatin domains in the fractions S3 and S4 and reads associated to open chromatin domains in fraction S2. This has been assessed comparing the reads distribution of these fractions against ChIP-seq data from literature (fig3).

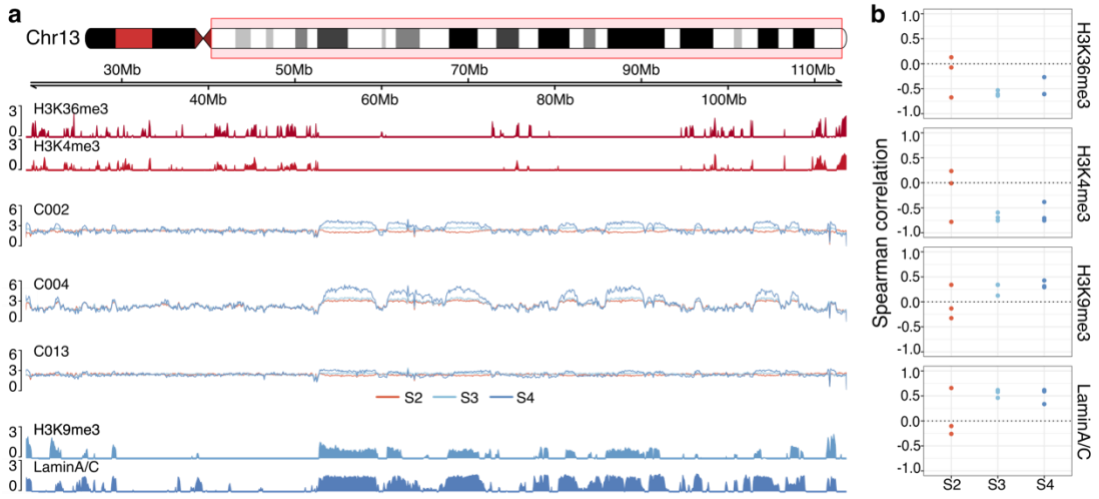


Figure 3: SAMMY-seq reads distribution along the genome. *(a) Visual representation along chr13:18,923,594-113,678,690 of two open chromatin marks (H3K36me3, H3K4me3; ChIP-seq experiments; on the Y axis: IP over INPUT reads distribution calculated with SPP) three SAMMY-seq experiments (samples: C002, C004, C013; for each sample all the three genomic fractions are represented: S2 in red, S3 in light blue, S4 in dark blue; reads coverage distribution) and closed chromatin marks (H3K9me3, Lamin A/C; ChIP-seq experiments; on the Y axis: IP over INPUT reads distribution calculated with SPP); (b) Genome-wide Spearman correlation (Y axis) between histone marks (rows) and each fraction (X axis) of each SAMMY-seq experiment; for this correlation analysis a bin size of 50Kb has been used.*

As SAMMY-seq was able to recapitulate open chromatin regions in S2 and closed chromatin principally in S4, we decided to combine the enrichments coming from the two fractions to improve our ability to capture both eu- and hetero-chromatin. We did it by performing the log ratio of the reads of each S4 bin against the reads of each S2 bin, producing in this way a new genomic track: the "S4vsS2" comparison (fig4a). As expected, the distribution of the comparison score along the genome showed a better correlation with both euchromatin and heterochromatin marks with respect to the single-track distributions separately (fig4b).

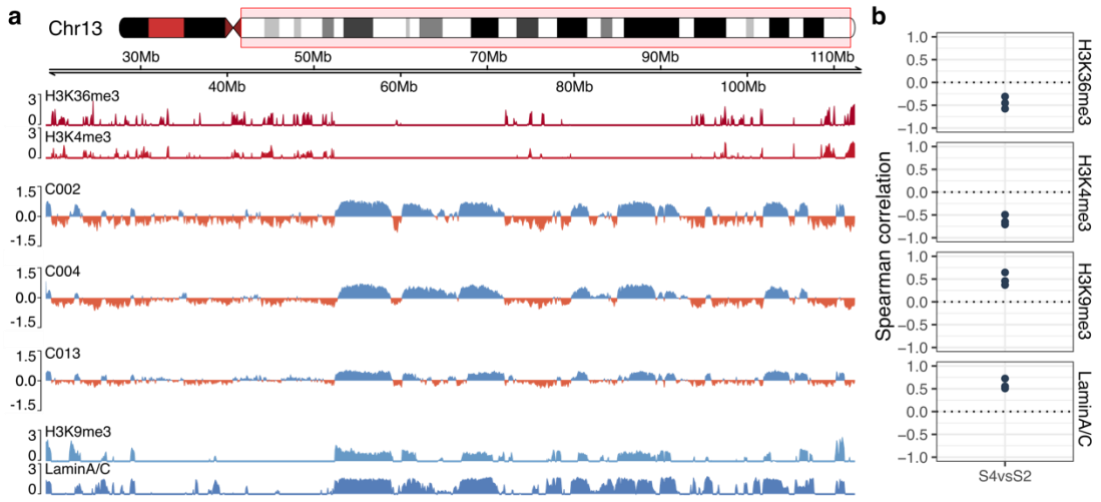


Figure 4: SAMMY-seq S4vsS2 ratio detects open and closed chromatin. *(a) Visual inspection along chr13:18,923,594-113,678,690 of two open chromatin marks (H3K36me3, H3K4me3; ChIP-seq experiments;* ChIP-seq experiments; on the Y axis: IP over INPUT reads coverage calculated with SPP*), three SAMMY-seq experiments (samples: C002, C004, C013; each sample is represented, Y axis, by S4vsS2 comparison calculated with SPP) and closed chromatin marks (H3K9me3, Lamin A/C;* ChIP-seq experiments; on the Y axis: IP over INPUT reads coverage calculated with SPP*). (b) Genome-wide Spearman correlation (Y axis) between histone marks (rows) and S4vsS2 comparison (X axis) for three samples; for this correlation analysis a bin size of 50Kb has been used.*

These experiments were performed on three samples of 4 million cells, showing a high level of reproducibility. Relying on these results we tested the performance of the method with a smaller number of cells.

We found that, even scaling down the amount of starting material to as little as 10 thousand cells, we are still able to get a correspondence in the reads enrichment profile between the SAMMY-seq experiments performed on different number of cells and with literature data (fig5a). As further confirmation, the genome-wide correlation of the down-sampled experiments vs 4 million cell SAMMY-seq 3f experiment is always greater than 0.5 (fig 5b).

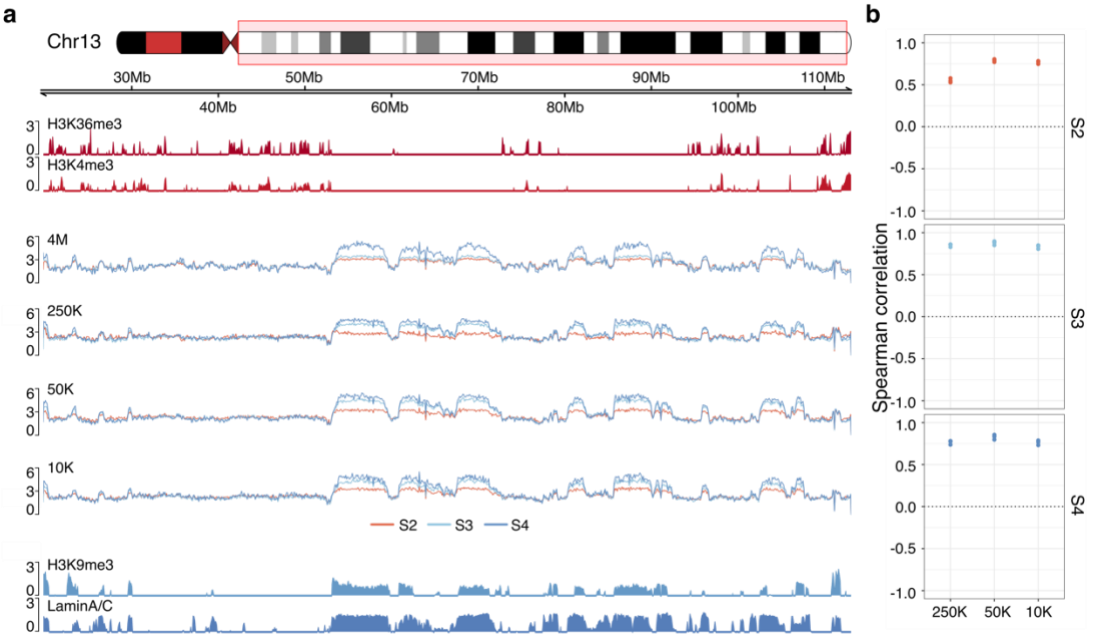


Figure 5: SAMMY-seq is reproducible at different cell amounts. *(a) Visual inspection along chr13:18,923,594-113,678,690 of two open chromatin marks (H3K36me3, H3K4me3; ChIP-seq experiments;* on the Y axis: IP over INPUT reads coverage calculated with SPP*), four SAMMY-seq replicates of sample C004 done using different cell amount (4M: 4 million cells, 250K: 250 thousand cells, 50K: 50 thousand cells; 10K: 10 thousand cells;* for each sample all the three genomic fractions are represented: S2 in red, S3 in light blue, S4 in blue; reads coverage on the Y axis) *and two closed chromatin marks (H3K9me3, Lamin A/C;* ChIP-seq experiments; on the Y axis: IP over INPUT reads coverage calculated with SPP*). (b) Genome-wide Spearman correlation (Y axis) between 4M replica and all the other lower cell number replicates (X axis) on which SAMMY-seq experiment has been performed; for this correlation a bin size of 50Kb has been used.*

Taken together these results show that SAMMY-seq 3f is a reliable method to study genome accessibility with a major focus on heterochromatin regions. In particular, the possibility to perform the protocol on a limited number of cells (as little as 10K) combined with the fact that it does not require chemical modifications and could be applied on primary cells make it highly suitable to study epigenetics on tissue derived samples.

4.1.2 SAMMY-seq detects LAD heterochromatic regions relocalization across nuclear compartments in different model systems

Encouraged by the results on the performances of SAMMY-seq we decided to apply it to investigate epigenetic mechanisms of two systems: 1) fibroblast cells derived by patients affected by progeria disease and 2) cancer engineered breast cancer cells subjected to mechanical stress. For both of the study cases an alteration of the lamina-genome interaction was expected based on previous literature but never mapped with

sequencing-based techniques. Therefore, our aim was to characterize the epigenetic landscape of these systems through the application of SAMMY-seq 3f and compare our results with other techniques that have been used to investigate on these topics such as ChIP-seq and Hi-C.

4.1.2.1 Application on progeria affected patients detects LADs alterations

Hutchinson-Gilford Progeria is a rare disease caused by an aberrant mutation on Lamina A coding gene. The mutation affects the correct interaction of chromatin with the nuclear lamina. This has been characterized through microscopy already at early stages of cell development, but not identified by sequencing techniques. Thus, the full comprehension of the role of the lamina mutation in the disease is still missing.

We decided to use SAMMY-seq to investigate this system as it does not address a specific component of chromatin (as genomic interactions in Hi-C or a specific molecule as in ChIP-seq), but it focuses on the biochemical properties of chromatin along the entire genome. We performed our analysis comparing primary fibroblast cells derived from healthy *versus* progeria affected donors.

From the study of differences among the S4vsS2 of the controls and the progeria patients, we found an alteration in heterochromatin domains. Indeed, while in control samples the S4vsS2 domains were reproducible across patients and overlapped the closed chromatin LADs from literature, in progeria samples they were much more heterogenous and the overlap with literature was largely lost (fig6).

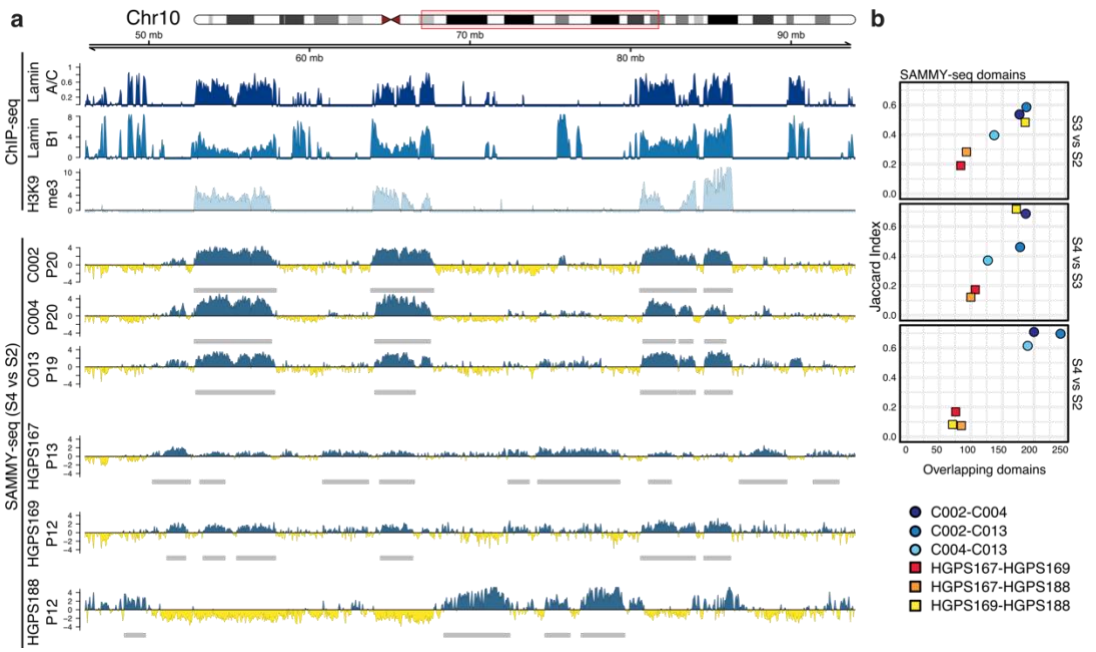


Figure 6: SAMMY-seq detects heterochromatin alterations in progeria samples. *(a) Visual inspection along chr10:46000000-94000000 of three closed chromatin marks (Lamin A/C, Lamin B1, H3K9me3;* ChIP-seq experiments; on Y axis: IP over INPUT reads coverage calculated with SPP*), three SAMMY-seq S4vsS2 reads coverage control sample tracks (C002 analysed at passage 20, C004 analysed at passage 20, C013 analysed at passage 19) and three SAMMY-seq S4vsS2 progeria sample tracks (HGPS167 analysed at passage 13, HGPS169 analysed at passage 12, HGPS188 analysed at passage 12). (b) Overlapping domains between controls and progeria samples, on X axis the number of domains (called by EDD) in common between two samples, while on Y axis there is represented the Jaccard index for the pair.*

While from SAMMY-seq point of view the changes between progeria and healthy samples could be appreciated, instead the same analysis done with ChIP-seq on H3K9me3 does not show any difference considering reads distribution (fig7a) and H3K9me3 domains detected remains almost identical across the six samples (fig7b).

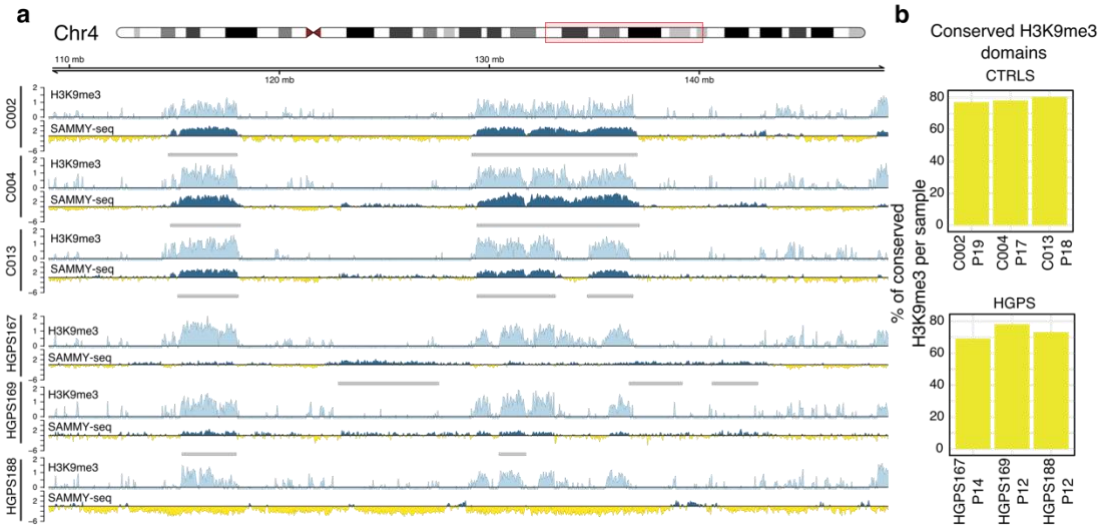


Figure 7: Progeria samples show heterochromatin accessibility alteration, keeping H3K9me3 profile unaltered. (a) *Visual inspection along chr4:109000000-149000000 of three control samples (C002, C004, C013) and three progeria samples (HGPS167, HGPS169, HGPS188); for each sample H3K9me3 ChIP-seq and SAMMY-seq S4vsS2 reads distribution has been represented.* (b) *Percentage of conserved domains across samples, calculated as the number of common domains over the sum of all H3K9me3 domains of the six samples.*

Confocal microscopy imaging of controls and progeria patients stained for H3K9me3 confirmed orthogonally the weakened association with lamina the nuclear periphery (Sebestyén et al., 2020).

Taken together these data suggest a mechanism in which the detachment of chromatin from the lamina does not immediately result in a redistribution of H3K9me3 in early passage fibroblasts, whereas it is expected to be more markedly changed at later passages (H. Zhang et al., 2016). This result could even explain why Hi-C has not been able to determine chromatin alteration after genome detachment from lamina, as it measures pairwise chromatin regions interactions, but it is not sensitive to a relocalization of chromatin domains across different subnuclear areas, as long as their local pairwise contact partners are preserved (Falk et al., 2019; Zheng et al., 2018).

Interestingly, while the LAD detachment does not alter H3K9me3 profile, we found a general decrease of H3K27me3 enrichment signal around the transcription starting site (TSS) (fig8a,b) and, in particular, for a set of 39 bivalent genes upregulated in progeria affected samples (fig8c).

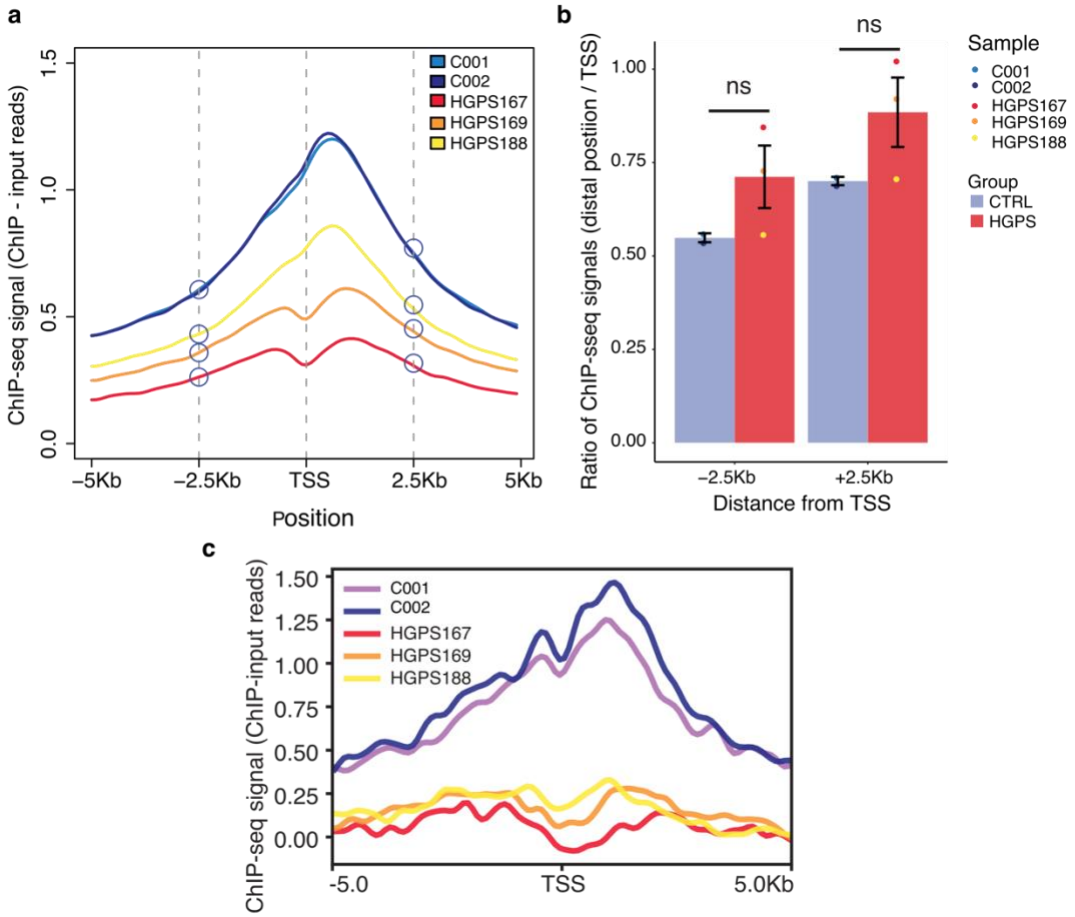


Figure 8: In progeria samples there is a redistribution of H3K27me3. *(a) Enrichment profile of H3K27me3 ChIP-seq reads distribution for control samples (C001, C002) and progeria samples (HGPS167, HGPS169, HGPS188) around TSS. (b) Ratio between the H3K27me3 value at TSS and the value 2,5 Kb before it (firsts two bars) and after it (last two bars) per each sample (controls: C001, C002; progeria: HGPS167, HGPS169, HGPS188). (c) Enrichment profile of H3K27me3 ChIP-seq 58 reads distribution for control samples (C001, C002) and progeria samples (HGPS167, HGPS169, HGPS188) around TSS for 39 bivalent genes upregulated between control and progeria samples.*

The presented data highlight that the alteration of LADs in progeria affected patients has an immediate effect on H3K27me3 deposition which is normally orchestrated by Polycomb complexes. This alteration leads to an aberrant transcription of Polycomb controlled genes, in line with and complementing other data reported in literature (Bianchi et al., 2020; Briand & Collas, 2018; Cesarini et al., 2015; Marullo et al., 2016; Oldenburg et al., 2017; Salvarani et al., 2019; Zheng et al., 2018).

The results described in this section were published in (Sebestyén et al., 2020) and part of the analyses presented in this last section was done in collaboration with Endre Sebestyén.

4.1.2.2 Application on MCF10A.DCIS Rab5+/-

Another application of SAMMY-seq 3f was performed on MCF10A ductal adeno carcinoma in situ cells (MCF10A.DCIS): a breast cancer precursor cell line. The control untreated MCF10A.DCIS cells were compared with cells treated to achieve overexpression of the RAB5A gene. This was part of a collaboration with Prof. Giorgio Scita's group at IFOM, and the details of the experimental model are described in the paper (Frittoli et al., 2022).

Rab5a is a GTPase protein mediating the cell motility, in cancer it seems to have the peculiar function of reducing the tissue stiffness. (Frittoli et al., 2014, 2021; Iannelli et al., 2021). As it has been observed that RAB5 activity correlates with a reduction of lamina levels (Iannelli et al., 2021), we wonder if this alteration that leads to major epigenetics rearrangements could be seen by SAMMY-seq.

As on fibroblasts, SAMMY-seq has been performed in parallel with ChIP-seq experiments targeting H3K9me3 and similarly to what happened on progeria sample patients, the overexpression of RAB5A leads to differences in S4vsS2 with respect to the controls; again, this alteration has not been observed in H3K9me3 ChIP-seq (fig9). This is in line with the hypothesis that enhanced mechanical stress tensions on the nuclear lamina, as induced by the cell motility following RAB5A overexpression, will induce detachment of heterochromatin from the lamina to facilitate nuclear deformation and reduce its stiffness. Thus, confirming that SAMMY-seq is able to detect genomic events associated to alterations in heterochromatin contacts with lamina, even if they are not immediately changing H3K9me3 distribution as measured by ChIP-seq.

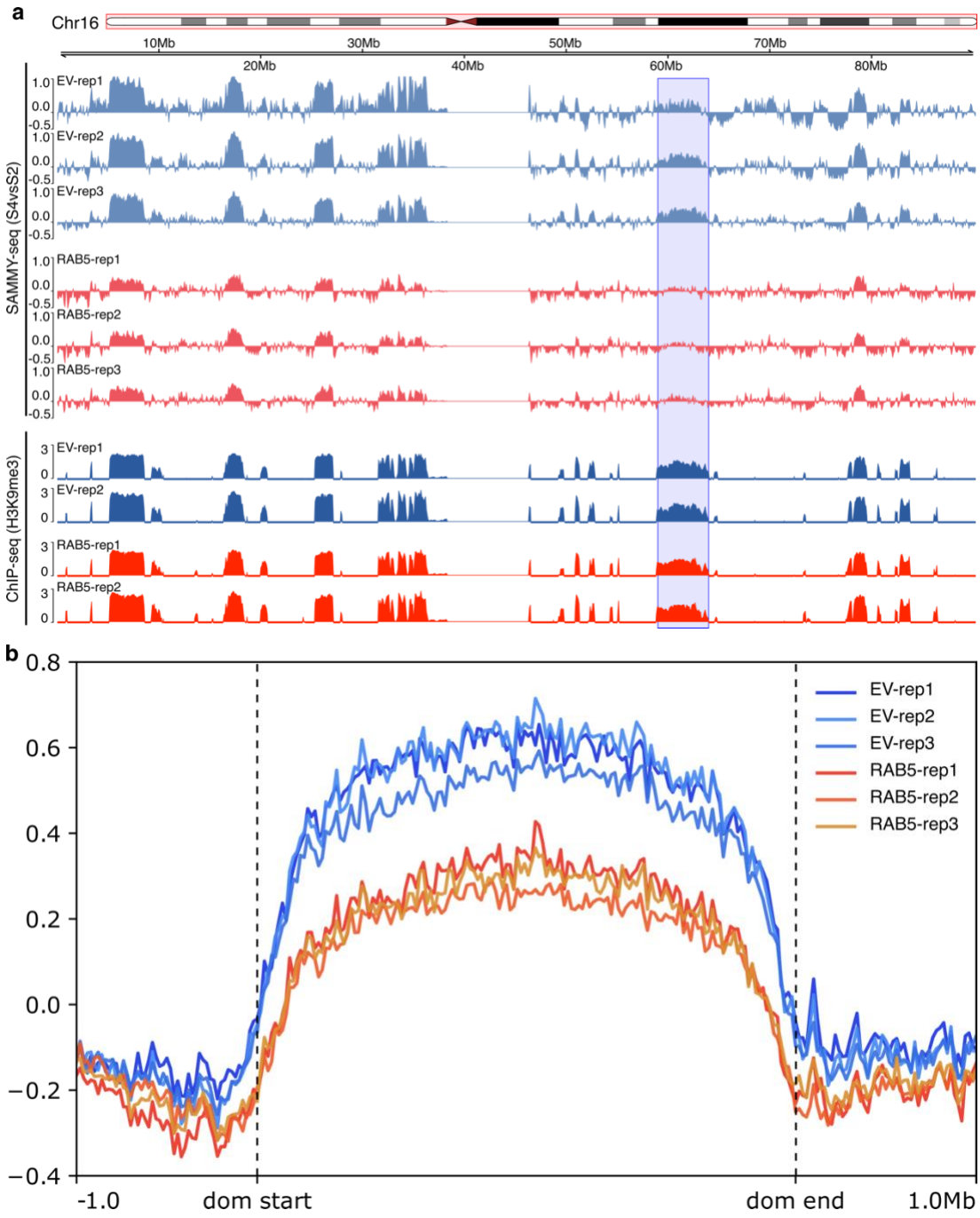


Figure 9: *In MCF10DICIS.com cells, overexpression of Rab5A leads to heterochromatin accessibility alterations keeping unaltered the H3K9me3 distribution profile. (a) Visual inspection of chr16 for six SAMMY-seq samples (three control samples: EV1, EV2, EV3; three RAB5A overexpressed samples: RAB5A-1, RAB5A-2, RAB5A-3; on Y axis: S4vsS3 reads coverage ratio calculated with SPP) and four H3K9me3 ChIP-seq samples performed on controls (dark blue profiles) and on RAB5A overexpressed samples (red profiles), on Y axis: IP over INPUT reads coverage calculated with SPP; highlighted in purple an example of H3K9me3 domain with enrichment in the SAMMY-seq reads coverage ratio S4vsS3 of controls and without enrichment in Rab5A overexpressed samples. (b) Enrichment profiles of the comparisons S4vsS2 of SAMMY-seq experiment over three controls (in blue) and three Rab5A overexpressed (in red/gold) samples over 78 H3K9me3 domains called with EDD on H3K9me3 ChIP-seq reads distribution.*

4.2 Simultaneous detection of both hetero-chromatin and eu-chromatin domains reorganization with a novel SAMMY-seq protocol (4f)

In this section I am presenting a new version of the SAMMY-seq based on the extraction and sequencing of four chromatin fractions. We set up this protocol to improve our ability to detect euchromatin regions, task addressed only on a large scale by the previous presented protocol.

According to our hypothesis, the DNA fragmentation of 3f SAMMY-seq was too strong and potentially leading to the complete loss of short and high accessible fragments characterizing the euchromatin. Thus, we reduced the strength of digestion to retain in our analysis the shorter DNA fragments in the S2 collection.

We tested this protocol on fibroblasts and we compared the results with the already presented SAMMY-seq version. Finally, we used this protocol to study prostate cancer. In particular, leveraging the high versatility of the method, we have been able to perform our analysis on prostate biopsies. We used our findings in prostate cancer epigenetics to identify a transcriptomic signature and identify two different tumour types among our samples. All work was performed in collaboration with Dr Chiara Lanzuolo's lab.

4.2.1 Limits of SAMMY-seq 3f protocol

The previously presented protocol (SAMMY-seq 3f) was shown to be a very reliable method to study heterochromatin and especially the dynamics of its connection with lamina, overcoming limitations affecting other techniques. As shown above, it could detect chromatin rearrangements connected to LADs dynamics. However, this technique is not equally able to dissect chromatin epigenetic status over smaller size euchromatic regions. Indeed, looking at the reads distribution along these domains the signal appears to be uniform all along the region instead of showing precisely localized spikes as it may happen for ChIP-seq targeting open chromatin markers such as H3K27ac, H3K4me1, H3K4me3 (fig10a). We addressed this problem by working on the fraction S2, that is the one that is supposed to represent the euchromatin, but that was actually variable across samples and mostly not informative in the original

SAMMY-seq 3f protocol (fig3b), and just slightly correlated with open chromatin marks (fig3b).

4.2.2 Description of the SAMMY-seq 4f protocol

We hypothesized that the reason why we have a poor sensitivity in detecting euchromatin regions was due to the DNase used to fragment the DNA: it was too efficient in digesting highly accessible DNA and that could lead to a complete loss of high accessible domains. This could explain why the S2 fraction of SAMMY-seq 3f protocol is not informative and we mostly use it as reference for computing relative enrichment in more condensed fractions (i.e. S4vsS2 and S3vsS2 comparisons).

We addressed this limitation by changing the enzyme used for the digestion. Instead of using the Turbo DNase, we moved to DNase I, an enzyme with lower relative efficiency of digestion (5 times lower than Turbo DNase) (https://www.thermofisher.com/order/catalog/product/AM2238). With this variation on the protocol, we expected to retain more open chromatin regions in the S2 fraction.

In addition, studying the DNA collected after the DNase I digestion, we found that we had genomic fragments of two different sizes. Thus, we decided to analyze them separately dividing the S2 fraction in two subfractions according to their DNA fragments length: the S2S, containing short fragments (that after separation are immediately sequenced), and the S2L, containing longer fragments (that after separation are sonicated and then sequenced). We initially thought that these two fractions could better characterize the level of accessibility of euchromatin.

In this way we created a new protocol composed by four sequenced chromatin fractions containing genomic DNA (S2S, S2L, S3, S4) that we called SAMMY-seq four fractions (4f).

4.2.3 Applicability of SAMMY-seq 4f protocol in several conditions:

4.2.3.1 On fibroblast SAMMY-seq 4f more precisely characterizes open chromatin regions

Once settled the protocol we tested the performance on fibroblasts and checked the differences comparing the results with the previously published SAMMY-seq 3f. Focusing in particular on the S2 derived fractions, we found a more precise detection

of open chromatin regions both in S2S and S2L with respect to S2 coming from the 3f experiment (fig10).

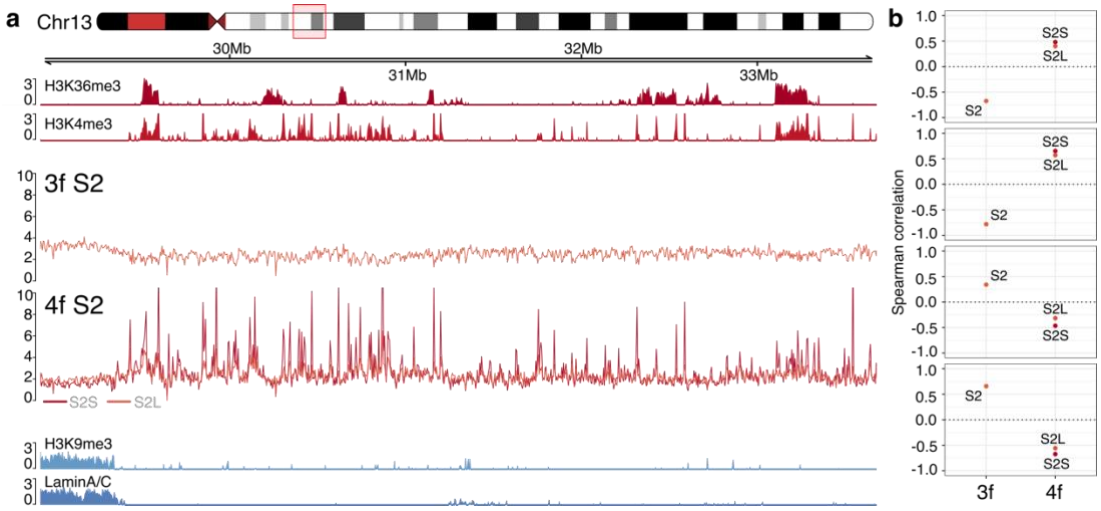


Figure 10: SAMMY-seq 4f detect open chromatin regions better than 3f. *(a) From top to bottom: distribution along chr13:28,923,594-33,678,690 of IP over INPUT profile of open chromatin marks (H3K36me3, H3K4me3), reads coverage of two SAMMY-seq experiments (3f_S2: S2 fraction deriving from SAMMY-seq 3f experiment; 4f_S2: in the plot are represented S2S, red line, and S2L, orange line), IP over INPUT profile of closed chromatin marks (H3K9me3, Lamin A/C). (b) Genome-wide Spearman correlation (Y axis) of S2 from 3f and S2S and S2L from 4f (X axis) against different chromatin marks (labelled on the left: H3K36me3, H3K4me3, H3K9me3, LaminA/C); for this correlation a bin size of 50Kb has been used.*

Unfortunately, the changes produced in the 4f protocol have a negative effect on our ability to detect closed chromatin regions. Indeed, performing a genome-wide Spearman correlation of the reads distribution of S3 and S4 fraction coming from the two protocols against chromatin marks, we found specific differences. While the S3 and S4 fractions of 3f are both negatively correlated with open chromatin marks and positively correlate with closed ones (fig11, left side); the S3 of the 4f protocol is not informative and the S4 does not show a clear enrichment towards eu- or hetero-chromatin either. We interpret this as a consequence of a lighter digestion in the step to extract the euchromatin, that results in not collecting all the open chromatin in S2 fraction and leaving a certain amount of euchromatic material in S3 and S4 fractions.

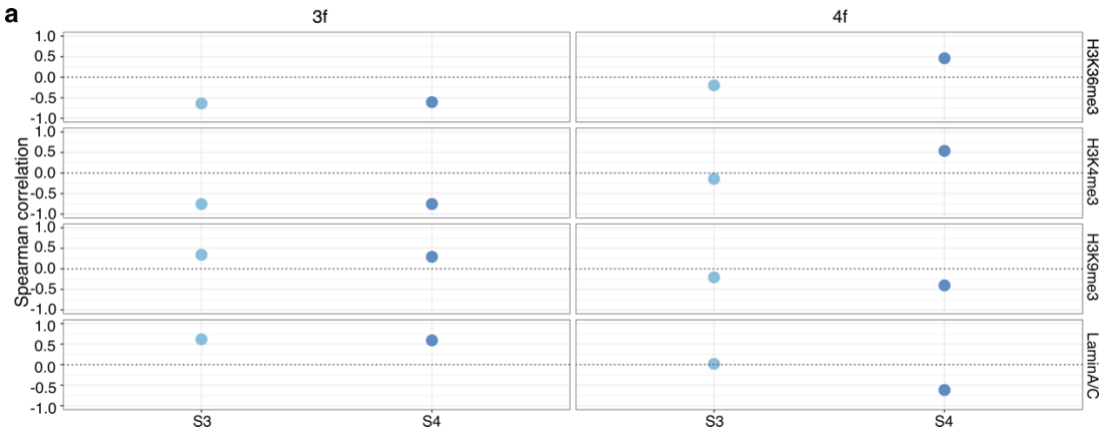


Figure 11: S3 and S4 SAMMY-seq 4f fractions do not correlate with close chromatin marks. *Genome-wide Spearman correlation (Y axis) for S3 and S4 fractions (X axis) coming from 3f and 4f protocols (column) against different chromatin marks (rows); for this correlation a bin size of 50Kb has been used.*

Looking at these results we expected the SAMMY-seq 4f protocol to be more reliable to study genome accessibility (euchromatin) rather than heterochromatin regions, thus we compared the S2vsS4 ratio coming from 3f experiment against the S2SvsS3 ratio coming from SAMMY-seq 4f.

We decided for these two ratios because they seem the best way to represent accessibility across the genome for each method. In these ratios, we expect the accessible regions as positive values and poorly accessible regions as negative.

From the comparison of the S2SvsS3 4f-SAMMY-seq and S2vsS4 3f-SAMMY-seq (fig12) we have been able to asses that the 4f protocol is in fact able to reliably detect both eu- and heterochromatin. In particular this is highlighted by genome-wide Spearman correlation analysis where S2SvsS3 shows a better correlation with euchromatin marks and a better anticorrelation with heterochromatin marks with respect to S2vsS4 (fig12b).

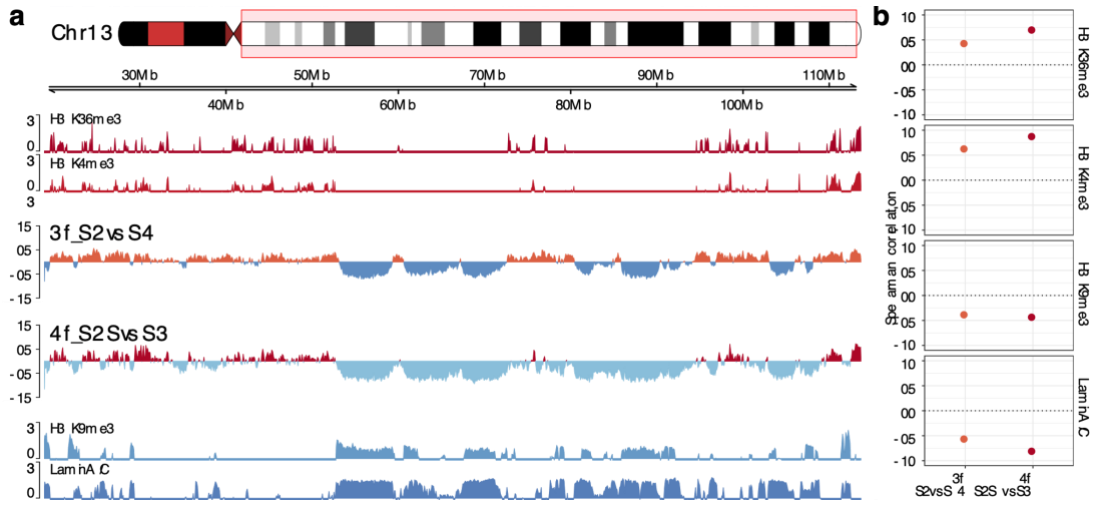


Figure 12: SAMMY-seq 4f S2SvsS3 recapitulates chromatin accessibility better than S2vsS4 3f. *(a) Distribution along chr13:18,923,594-113,678,690 of: IP over INPUT ChIP-seq ratio of open chromatin marks (H3K36me3, H3K4me3), S2 over S3 SAMMY-seq 3f experiment, S2S over S3 SAMMY-seq 4f experiment and IP over INPUT ChIP-seq ratio of closed chromatin marks (H3K9me3, Lamin A/C). (b) Genome-wide Spearman correlation (Y axis) of S2vsS4 3f and S2SvsS3 4f (X axis) against different chromatin marks (rows); for this correlation a bin size of 50Kb has been used.*

One of the main features of the 3f protocol was its reliability on a small number of cells (10K). This makes it suitable for applications characterized by scarcity of material. Considering that the differences between 3f and 4f protocols are minimal we expect that even this new version of SAMMY-seq could be performed on a limited number of cells. Thus, we perform a SAMMY-seq 4f experiment on a sample of fibroblast of 10K cells and then we compared the results with the previously presented data (based on 3M cells). Both the analysis of S2SvsS3 signal distribution along the genome of the two experiments (fig13a) and their genome-wide correlation (fig13b) confirm that the experiment on 10K cells is almost identical to the 3M cells experiment.

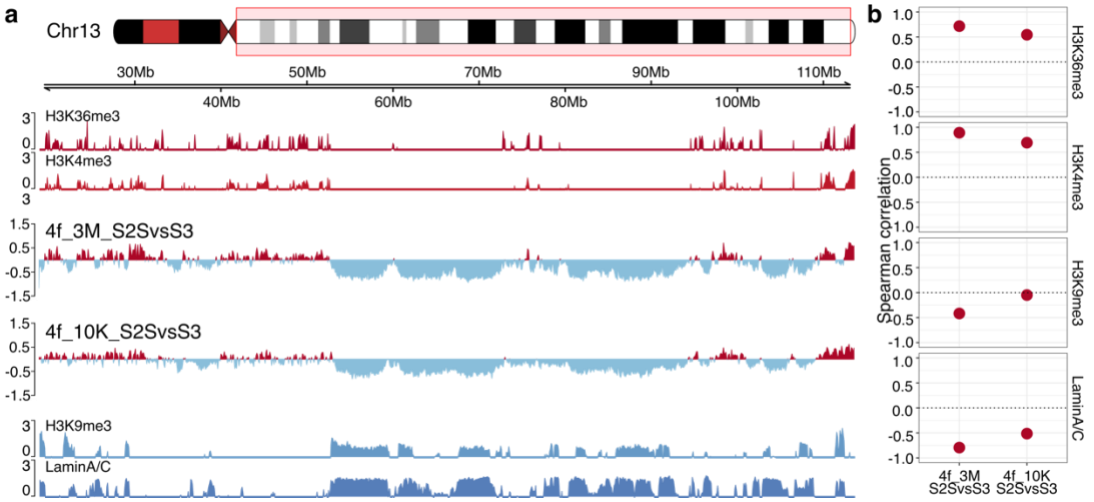


Figure 13: SAMMY-seq 4f is reproducible even performed on 10 thousand cells. *(a) Distribution along chr13:18,923,594-113,678,690 of: IP over INPUT ratio of open chromatin marks (H3K36me3, H3K4me3), S2S over S3 of SAMMY-seq experiments (4f_3M_S2SvsS3: experiment performed on 3 million cells; 4f_10K_S2SvsS3: experiment performed on 10 thousand cells) and IP over INPUT ratio of closed chromatin marks (H3K9me3, Lamin A/C). (b) Genome-wide Spearman correlation (Y axis) of S2SvsS3 respectively for a 3 million cells experiment and a 10 thousand cells one (X axis) against different chromatin marks (rows); for this correlation a bin size of 50Kb has been used.*

4.2.3.2 SAMMY-seq 4f characterizes open and closed chromatin regions applied on mouse C2C12 cells

The analyses presented until now for the 4f protocol were based mainly on the same cell type (fibroblast) and in general SAMMY-seq has been performed only on human samples. Thus, to understand if our method could be reliable even for studying other organisms and other cell lines, we performed several experiments on mouse C2C12 cells. In particular we did the tests on two 2 million cells samples and one 50 thousand cells sample to confirm the scalability of the technique. The data, as for other experiments, were compared with several ChIP-seq experiment from literature performed on the same cell line.

The results of these experiments were in line with all the other shown for 4f: the fractions S2S and S2L and, partially, the S4 were found positively correlated with open chromatin marks and negatively with closed chromatin marks; while the S3 didn't present any enrichment (fig14a,b). Again, performing the log ratio between S2S and S3 we obtain a genomic track with high correlation values both for open and closed chromatin (fig14c,d). Finally, even in this case, performing the analysis on different numbers of cells (2 million cells and 50 thousand cells) we obtain highly reproducible results (fig14).

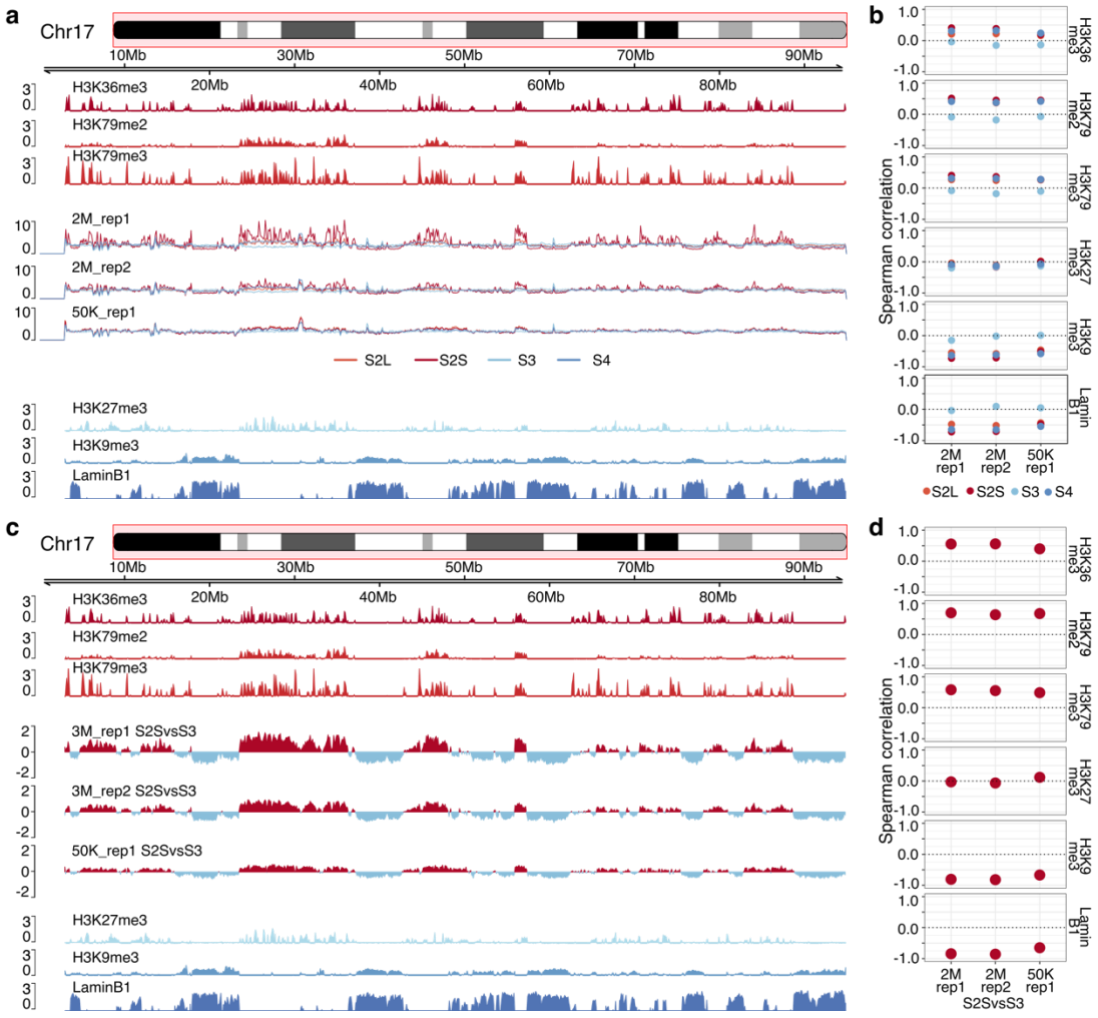


Figure 14: *SAMMY-seq 4f performed on C2C12 is reliable on detecting open and closed chromatin even using different number of cells. (a) Distribution along chr17 of: IP over INPUT ratio of open chromatin marks (H3K36me3, H3K79me2, H3K79me3), reads distribution of three SAMMY-seq 4f experiments (first and second performed on 2 million cells, while third has been performed on 10 thousand cells; in each plot all the four fraction per experiment are represented: S2S in red, S2L in orange, S3 in light blue and S4 in blue) and IP over INPUT ratio of closed chromatin marks (H3K27me3, H3K9me3, Lamin B1). (b) Genome-wide Spearman correlation (Y axis) of SAMMY-seq 4f experiments (two replicas performed on 2 million cells and one performed on 50 thousand cells; X axis) against different chromatin marks (rows), each dot represents the correlation of a fraction (S2S in red, S2L in orange, S3 in light blue, S4 in blue) with a chromatin mark; for this correlation analysis a bin size of 50Kb has been used. (c) Distribution along chr17 of: IP over INPUT ratio of open chromatin marks (H3K36me3, H3K79me2, H3K79me3), S2S over S3 of three SAMMY-seq experiments (first and second performed on 2 million cells, while third has been performed on 10 thousand cells and IP over INPUT of closed chromatin marks (H3K27me3, H3K9me3, Lamin B1). (d) Genome-wide Spearman correlation (Y axis) of SAMMY-seq 4f experiments (two replicas performed on 2 million cells and one performed on 50 thousand*

cells; X axis) against different chromatin marks (rows), each dot represents the correlation of the log ratio S2SvsS3 of an experiment with a chromatin mark; for this correlation analysis a bin size of 50Kb has been used.

4.3 Stratification of prostate cancer patients based on epigenomic profiles using SAMMY-seq 4f

The possibility of studying open and closed chromatin in a sample composed by a small number of cells make this SAMMY-seq protocol suitable for studies of diseases in which the role of epigenetics is still not completely understood. Especially in cases where the lack of reliable techniques to work on tissue samples is critical, such as in prostate cancer.

Indeed, prostate cancer is a type of tumour whose epigenetics alterations are mainly studied on cell lines, avoiding in this way several technical problems affecting analysis on tissues (e.g. tumour purity, cell composition heterogeneity and tissue scarcity) but losing reliability with the biology of real tumour tissue. In addition, work done on prostate cancer mainly focuses on euchromatin, but observations of prostate cancer nuclear atypia suggest that even large scale heterochromatin rearrangements could be involved as cause or consequence of the diseases.

The results obtained with SAMMY-seq 4f, especially considering the comparisons, were reliable to study both open and closed chromatin. So, we decided to apply this method to study prostate cancer directly on biopsies taking into account tumour purity, cell composition and gene expression for each sample.

4.3.1 Prostate biopsies data collection and cohort description

With the aim of taking into account as much as possible variables that could be of interest for our analysis on a tissue sample, we setup an *ad hoc* experimental design to get multiple types of information from our samples. So, for each prostate needle biopsies (composed, on average, by 50 thousand cells), we used one third for histology (done by a pathologist), while the remaining part was used for tissue composition analysis (done through fluorescence activated cell sorting, FACS), epigenetics (SAMMY-seq) and transcriptomic (RNA-seq) analysis (fig15). In addition, for each prostate biopsy processed we collected patient clinical information (i.e. age, PSA

level, percent of tumor detected) and the grade of the closest biopsies with respect to the one we processed for our experiments.

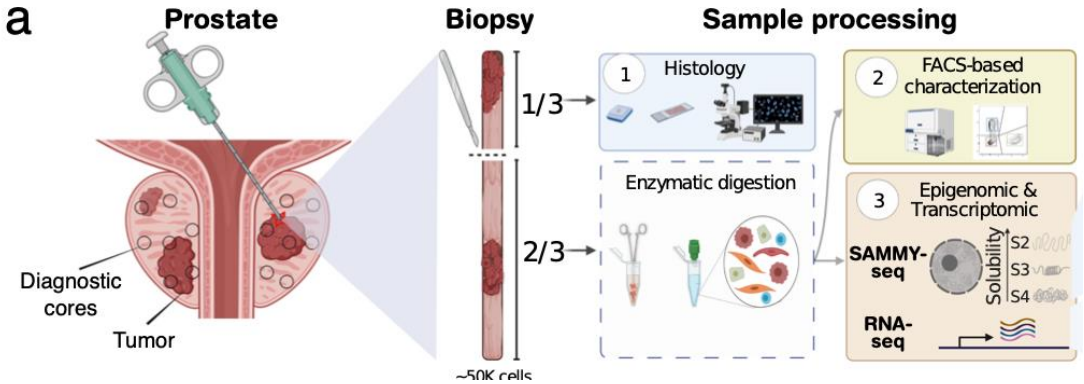


Figure 15: Prostate biopsies processing experimental design. *Schematic representation explaining our experimental design on prostate tissue. Once a prostate biopsy was obtained from a patient, we use one third for histology, while the remaining two thirds were digested enzymatically and the cells were used for FACS based characterization and for SAMMY-seq and RNA-seq. Cartoon created with BioRender.com.*

We collected data from 29 patients, but we focused our analyses only on 17 biopsies from 17 patients: we focused on patients with a clear classification of tumor (affected or non-affected), whereas we left out samples with prostatitis, high grade prostatic intraepithelial neoplasia, and samples with a tumor diagnosis but not detected in the specific needle biopsy donated for our experiments. Thus, among our analysis cohort 7 samples were tumour free and we used them as controls samples (CTR), the remaining 10 (PCa) were affected by tumour considering the grading performed either on the tissue we analysed (GS1) or on the closest biopsy (GS2). While considering the GS2 we had represented the Gleason score according to the standard grading, for the GS1, as we had just one third of biopsy, we represented the cancer grade using just one number (fig16).

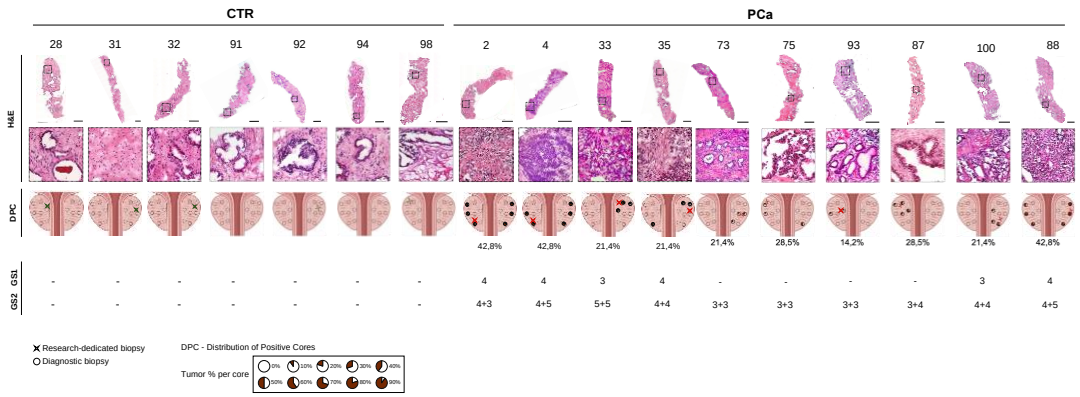


Figure 16: Histology and grading of sample cohort. *In the picture are represented the prostate biopsies we used in this work and their features. First row: representation of the entire biopsy; second row: zoom on a biopsy region representative of the status of the entire piece; third row: schematic representation of patient prostate from where our biopsy come from (the empty circles represent the regions where the biopsies are taken but not tumour has been detected, the black circles represent the biopsies found having tumour and the its percentage respect to the entire piece, the green stars for CTR patients and the red star for PCa patient represent the region where our biopsy was located, under each Pca patient is indicated the percentage of total tumoral cores respect to the entire biopsy); forth row: the Gleason score associated to our biopsy (GS1) and the Gleason score associated to the entire prostate (GS2).*

Unfortunately, the cell composition analysis was not possible for all samples. For the tissues that have been characterized, they appear to have a similar composition between PCa and CTR samples (fig18) for lymphocytes (~10,11% for heathy and ~12,93% for tumour patients), epithelial cells and (~18,21% for heathy and ~15,22% for tumour patients) stromal cells (~71,68% for heathy and ~71,85% for tumour patients).

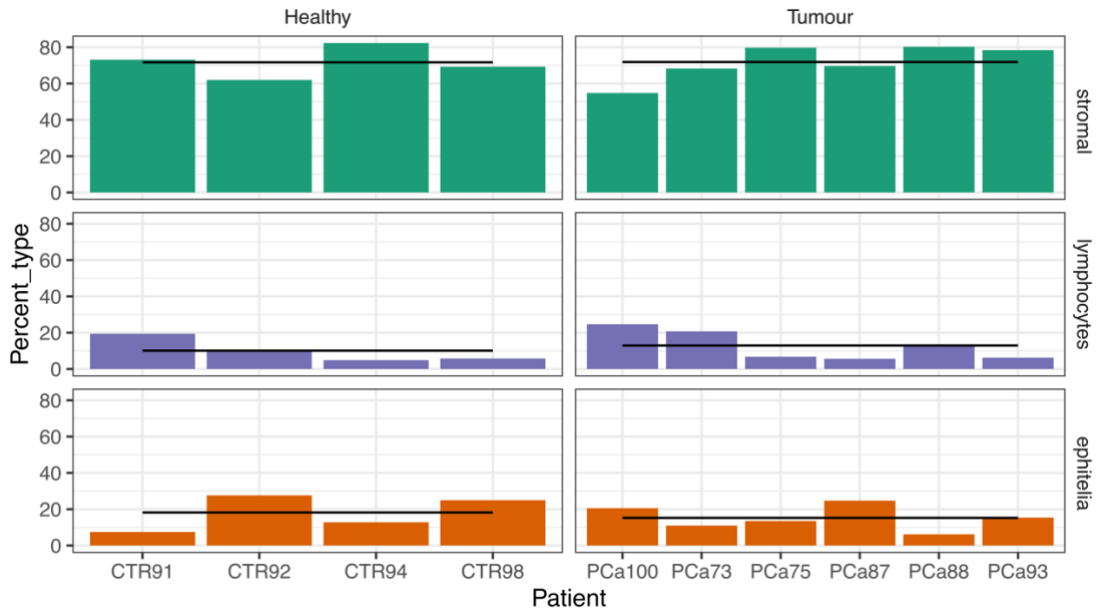


Figure 17: Prostate cancer does not affect biopsy cell composition. *The picture represents the percentage (Y axis) of each cell type (rows) detected using FACS for 10 biopsies (X axis) used in our analysis. The samples are grouped in CTR and PCa samples (column). The black line represents the mean of percent of cells in each group (from left to right and from top to bottom: 71,68%, 71,85%, 10,11%, 12,93%, 18,21% and 15,22%).*

4.3.2 SAMMY-seq 4f reliably detects epigenetics features in prostate biopsies samples

The analyses performed with SAMMY-seq were compared with two ChIP-seq datasets from ENCODE performed on an entire prostate.

Initially we focused on CTR samples to setup the bioinformatic analysis. Looking at the genomic profile of the single fractions we noticed, as expected, that the S2 subfractions and partially the S4 where enriched in euchromatin regions while, as already seen in the other experiments in human fibroblast and mouse C2C12, the S3 was not enriched neither for open neither for closed chromatin (fig18). Interestingly, the S2S fraction in this case is more variable across samples than the S2L fraction, thus we chose this last one to perform the log ratio with the S3 fraction.

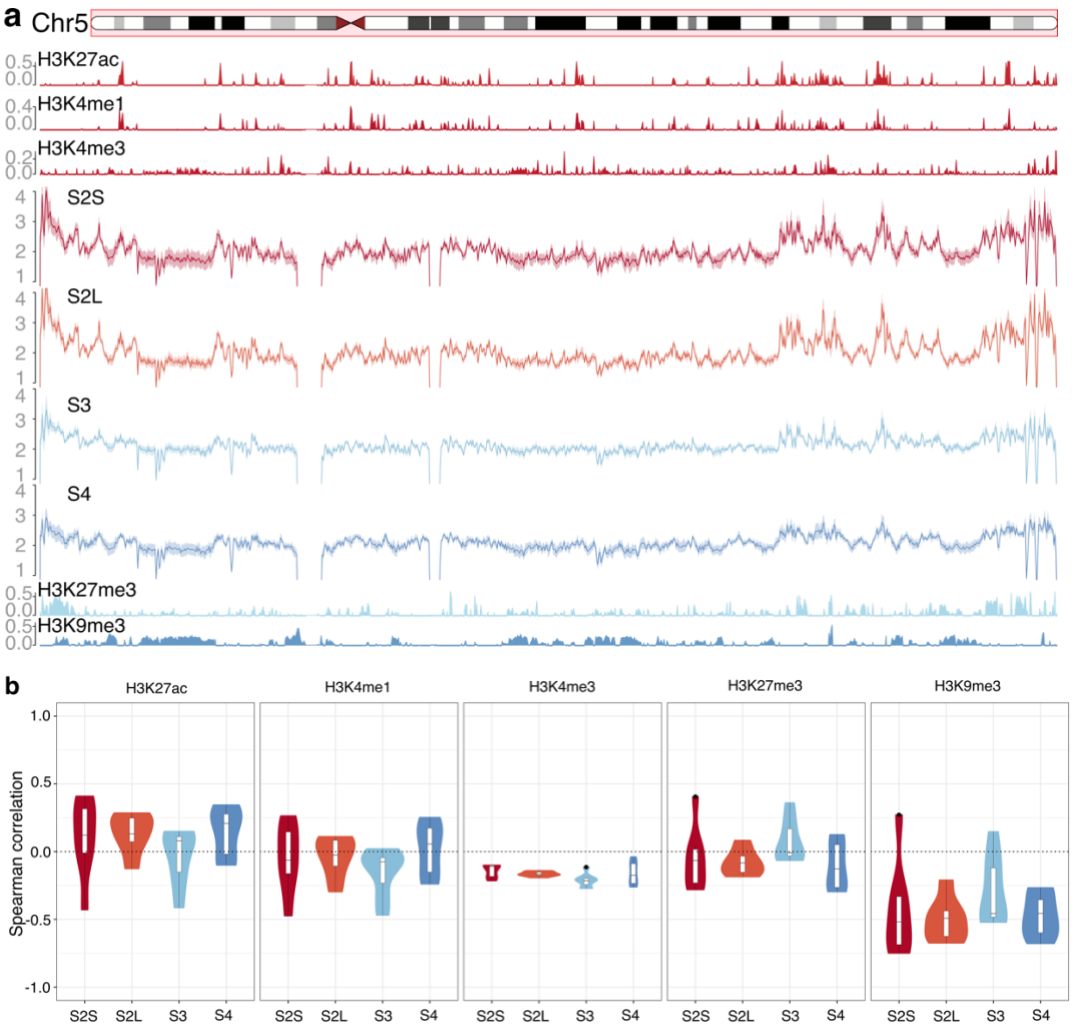


Figure 18: SAMMY-seq 4f fractions on prostates biopsy control samples are in line with literature information. *(a) Distribution along chr5 of: IP over INPUT for open chromatin marks (H3K27ac, H3K4me1, H3K4me3; ChIP-seq experiment), reads coverage for SAMMY-seq 4f fractions of 7 CTR patients' consensus per each fraction (S2S, S2L, S3, S4; the continuous line represents the mean across the patients and the shadow around represents the confidence interval) and IP over INPUT for closed chromatin marks (H3K27me3, H3K9me3; ChIP-seq experiment). (b) Spearman correlation values (Y axis) for each SAMMY-seq 4f fraction (X axis) for five different histone marks (rows). Each violin represents the distribution of correlation 7 CTR samples; for this correlation analysis the bin size of 150Kb has been used.*

Analysing the log ratio of the S2LvsS3 of CTR samples, again, SAMMY-seq is concordant with chromatin marks, in particular H3K27ac and H3K9me3 (fig19) detecting then euchromatin and heterochromatin domains.

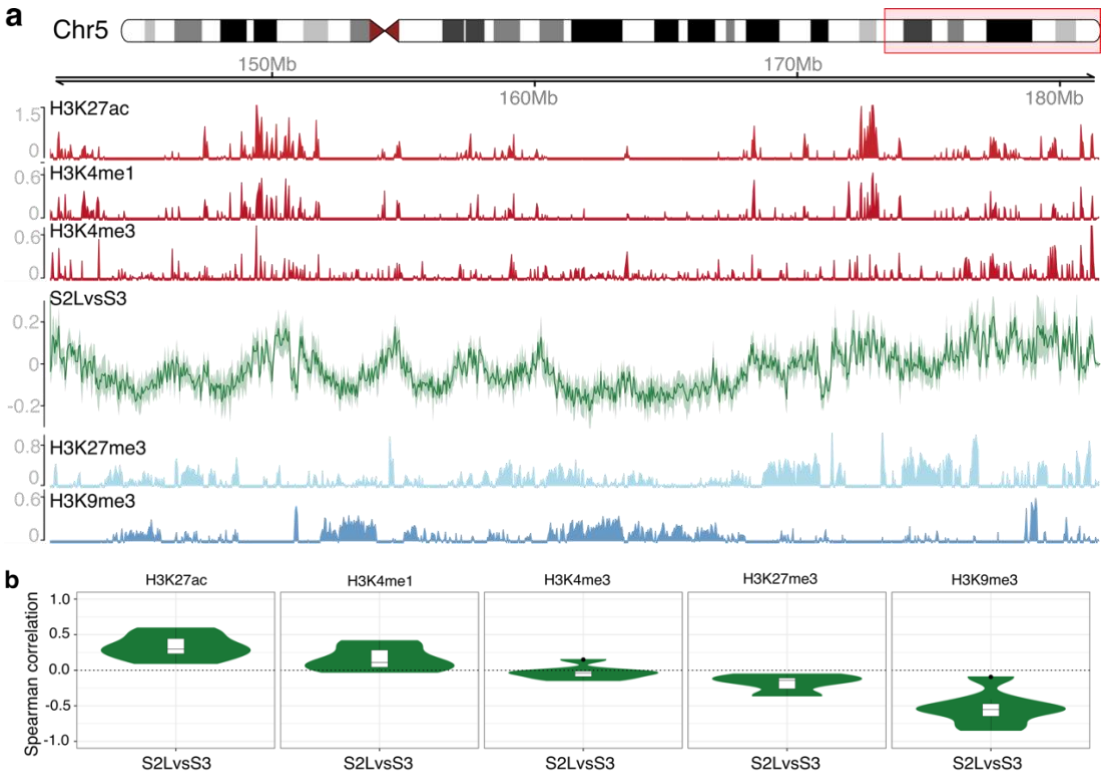


Fig19: SAMMY-seq 4f S2LvsS3 log ratio represent open and closed chromatin in prostate biopsies. (a) *Distribution along chr5:141,538,250-181,538,259 of: IP over INPUT of open chromatin marks (H3K27ac, H3K4me1, H3K4me3; ChIP-seq experiments), SAMMY-seq 4f S2LvsS3 log ratio consensus of 7 CTR samples (the continuous line represents the mean across the patients and the shadow around represents the confidence interval) and IP over INPUT ChIP-seq closed chromatin marks (H3K27me3, H3K9me3; ChIP-seq experiments).* (b) *Distribution of Spearman correlation values (Y axis) for SAMMY-seq 4f S2LvsS3 log ratio (X axis) for five different histone marks (rows); each violin represents the correlation values distribution for the 7 CTR samples; for this correlation analysis the bin size of 150 Kb has been used.*

4.3.3 SAMMY-seq 4f distinguishes two different groups of prostate cancer patients

Instead, the analysis on PCa patients revealed a more variable situation, indeed, just in some cases we have concordance with chromatin marks and, in other cases, we have a completely opposite trend with respect to the controls (fig20).

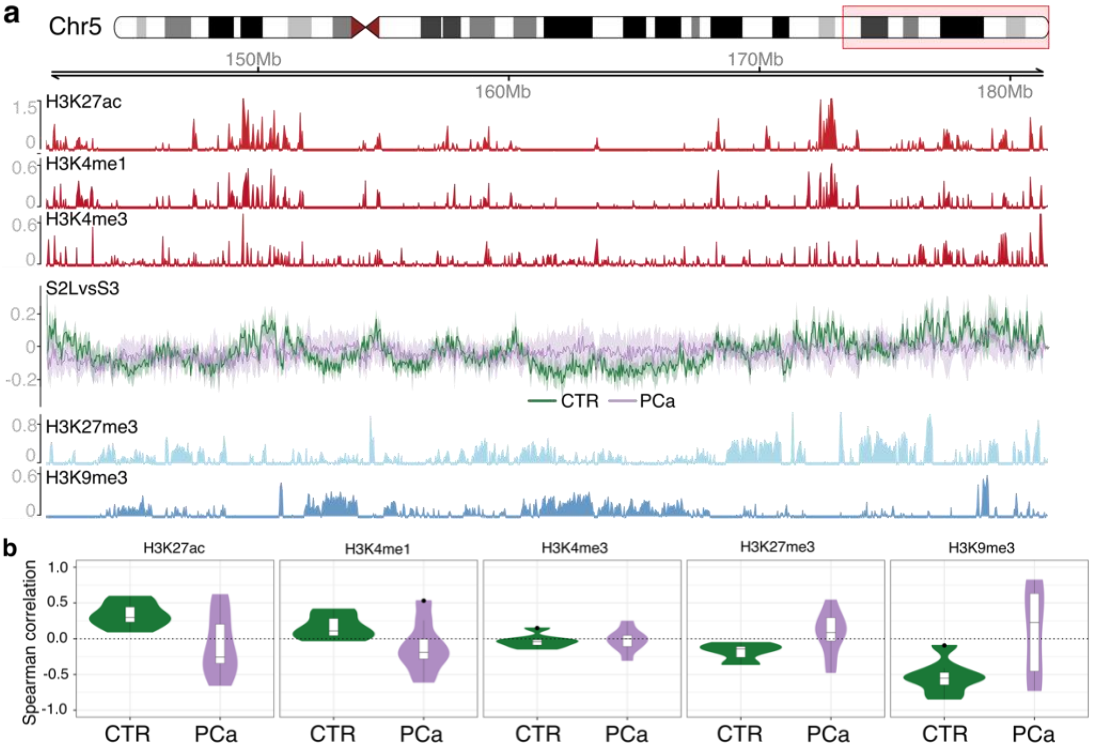


Figure 20: SAMMY-seq 4f S2LvsS3 log ratio for PCa samples does not represent a clear pattern among samples. (a) *Distribution along chr5:141,538,250-181,538,259 of: IP over INPUT open chromatin marks (H3K27ac, H3K4me1, H3K4me3; ChIP-seq experiment), SAMMY-seq 4f S2LvsS3 log ratio consensus of 7 CTR (green) and of 10 PCa (purple) samples (the continuous line represents the mean across the patients and the shadow around represents the confidence interval) and IP over INPUT closed chromatin marks (H3K27me3, H3K9me3; ChIP-seq experiments).* (b) *Distribution of Spearman correlation values (Y axis) for SAMMY-seq 4f S2LvsS3 log ratio for 7 CTR (green) and 10 PCa (purple) samples (X axis) for five different histone marks (columns); each violin represents the correlation values of patients; for this correlation analysis the bin size of 150 Kb has been used.*

Considering the heterogeneity across cancer samples, we decided to study the epigenetics alterations singularly for each PCa sample, looking for regions commonly altered in each patient (fig21a). We compared the S2LvsS3 value of each genomic bin of each PCa sample with the mean of the S2LvsS3 values of CTR samples in the same bin. We considered positively or negatively switched in accessibility each bin whose difference between PCa and CTR consensus was higher or lower two times the consensus standard deviation.

From this analysis we found that it is possible to cluster the samples in two groups (fig21b) with different features: one group with less differences with respect to CTR

patients (that we called low decompartmentalization degree LDD) and the other characterized by a higher number of accessibility regions switched (high decompartmentalization degree or HDD, fig21c).

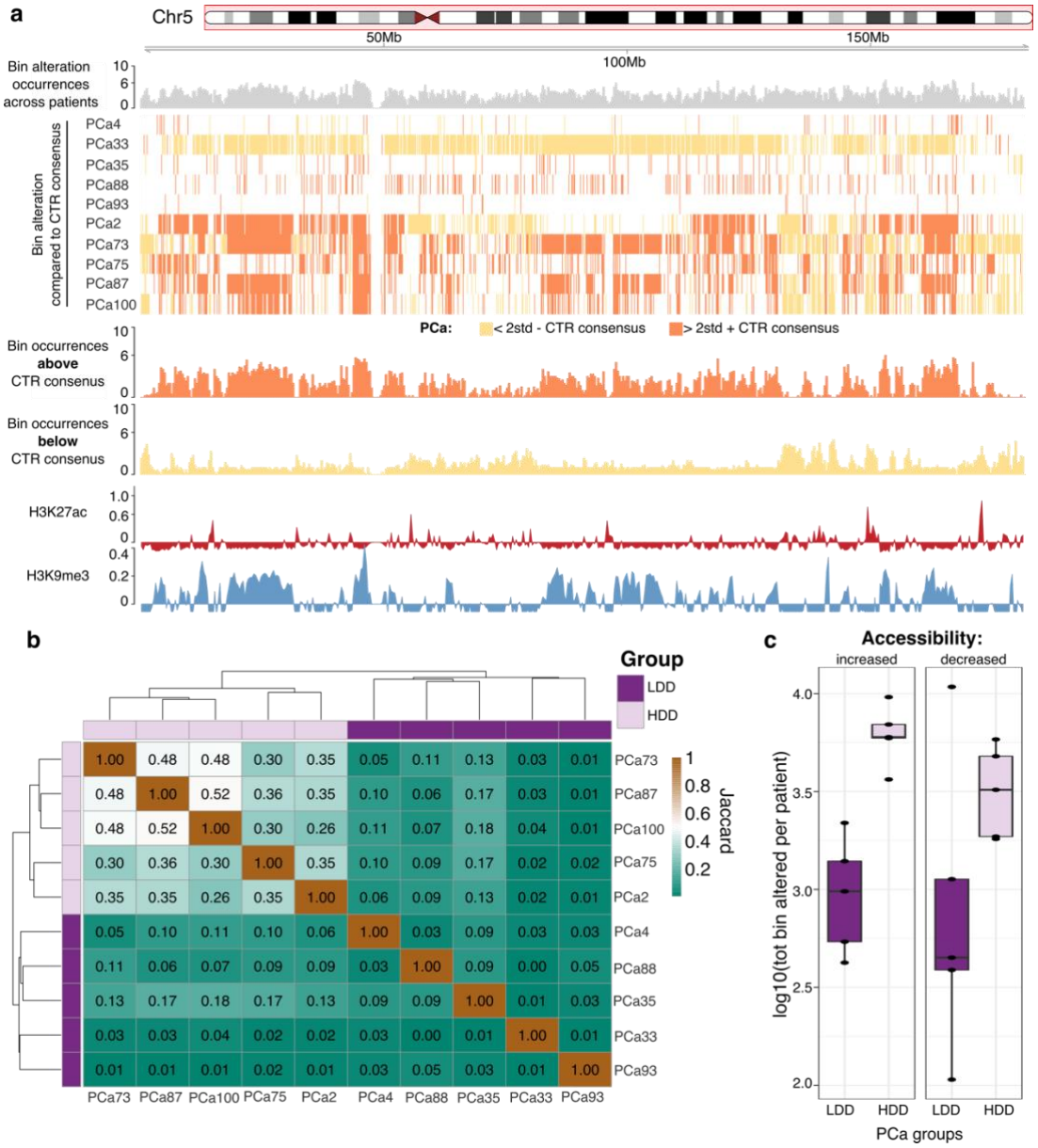
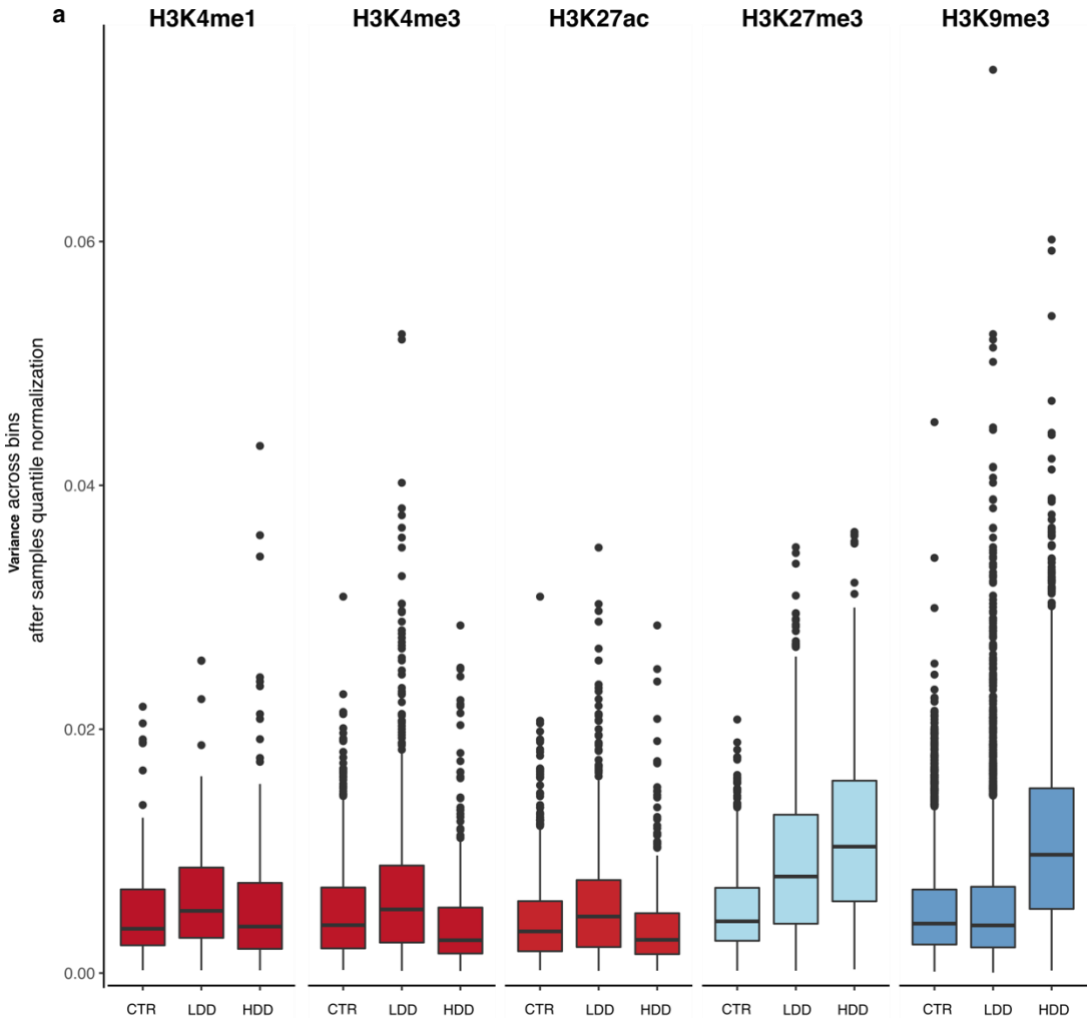


Figure 21: Accessibility changes in PCa patients allow clustering them in two groups. *(a) Distribution along chr5 of chromatin accessibility changes in tumours. Starting from the chromosome ideogram to the bottom: the number of PCa samples having a S2LvsS3 SAMMY-seq log ratio higher or lower two standard deviations from the consensus of CTRs (calculated as the mean of CTR values for each bin), a barcode for each sample where each bar represents a bin (the bar is coloured in orange if the PCa in that bin has a value two standard deviation higher than the value of the CTR consensus for that bin, yellow if lower and white if it falls in the range), the number of PCa samples having a S2LvsS3 SAMMY-seq log ratio higher two standard deviation from the consensus of CTRs, the number of PCa samples having a S2LvsS3 SAMMY-seq log ratio lower two standard deviation from the consensus of CTRs, the genomic track of the ChIP-seq for H3K27ac and H3K9me3 (represented as IP over INPUT ratio). (b) Heatmap representing the value of overlap (represented by Jaccard score) of pair of PCa samples; the patients are organised in row and in column on the base of a hierarchical cluster among them. (c) Number of bins*

increasing (left) or decreasing (right) the S2LvsS3 log ratio value of at least 2 standard deviations for LDD (dark purple) and HDD (pink).

Studying separately these two groups of patients and focusing on different histone mark domains, we confirmed that the values of S2LvsS3 are similar among CTR patients and LDD patients, while the HDD is substantially different, except for the H3K27me3 domains, where even for this second group we found similar values distribution across CTR samples (fig22a).

Looking instead at the variance of the values inside the groups we found that they are consistent in open chromatin regions, while there is a great variance in closed chromatin regions for the tumoral groups: in particular in H3K27me3 and, just for the HDD group, in H3K9me3 domains (fig22b).



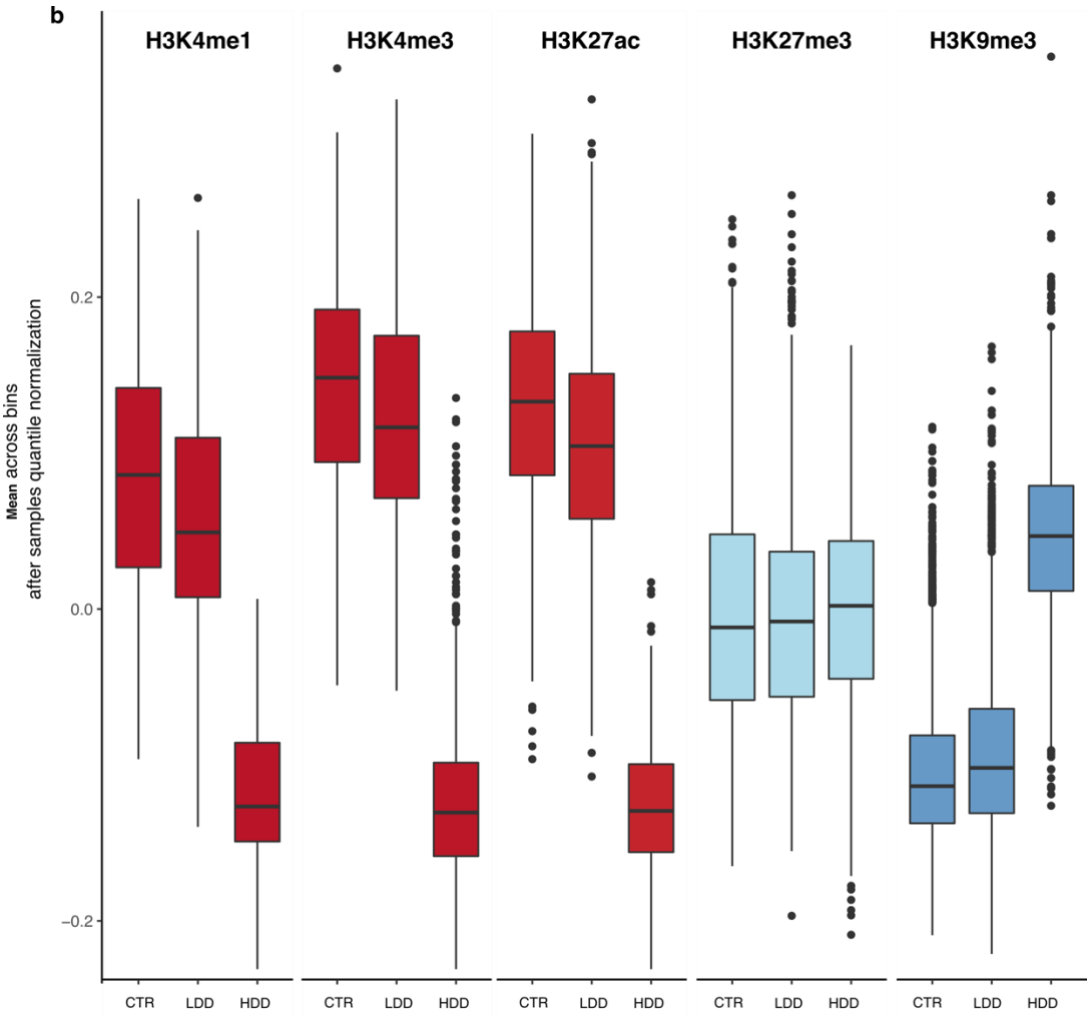


Figure 22: Accessibility alterations between two PCa groups focused on eu- and hetero-chromatin domains. *(a) Boxplots representing the mean of the S2LvsS3 log ratio values (Y axis) per sample group (X axis) per each bin falling on different ChIP-seq histone mark domains (columns). (b) Boxplots representing the variance of the S2LvsS3 log ratio values (Y axis) per sample group (X axis) per each bin falling on different ChIP-seq histone mark domains (columns). For both of panels the values are plotted after normalized the SPP sample score using a quantile normalization and then the values associated to the domains of different histone marks have been selected.*

Taken together, these results could assign a central role to closed chromatin in defining tumours subgroups. In particular, the only feature in common among the two PCa groups and different from CTR is the variability on H3K27me3 regions, suggesting that an alteration of genes controlled by H3K27me3 is needed for tumour development.

4.3.4 Epigenetics rearrangements distinguish biologically relevant subgroups of tumour-affected patients

The differential expression analysis between PCa and CTR patients does not yield many differentially expressed genes considering all the PCa patients together, probably due to the presence of two different subgroups of patient samples (fig23a). Instead, dividing the PCa patients in the two groups previously detected with SAMMY-seq epigenetic analysis of differential accessibility, we found that the number of differentially expressed genes is smaller for LDD samples (fig23b) and larger when considering only the HDD (fig23c). This is in accordance with the epigenetics results: LDD has been detected by SAMMY as being more similar to controls whereas HDD goes is more different (fig21c).

Considering this correspondence between epigenetics and expression alterations we try to correlate if genes that have a significantly high fold change results in a region whose accessibility appears altered according to SAMMY-seq. For HDD we found a large overlap between loss of accessibility and gene down-regulation in terms of gene expression, whereas the opposite is not equally evident for genes gaining accessibility (fig23d). However, we should consider that there is a larger number of genes down-regulated in terms of gene expression (84 number) as opposed to the up-regulated ones (29) in HDD vs controls comparison. This could be due to the fact that we need more than an event to activate transcription of a silent gene: increasing its accessibility is not enough, as it will also require its specific upstream regulators (transcription factors) to be active. Whereas a reduction in a gene epigenetic accessibility may be enough to inhibit it. For what concerns the LDD patients instead, the vast majority of altered genes are in regions that are not affected by switch in SAMMY-seq. This is somehow expected as the changes in chromatin accessibility are limited in the LDD group, as such we can’t expect the changes in gene expression to be determined by small epigenetic changes.

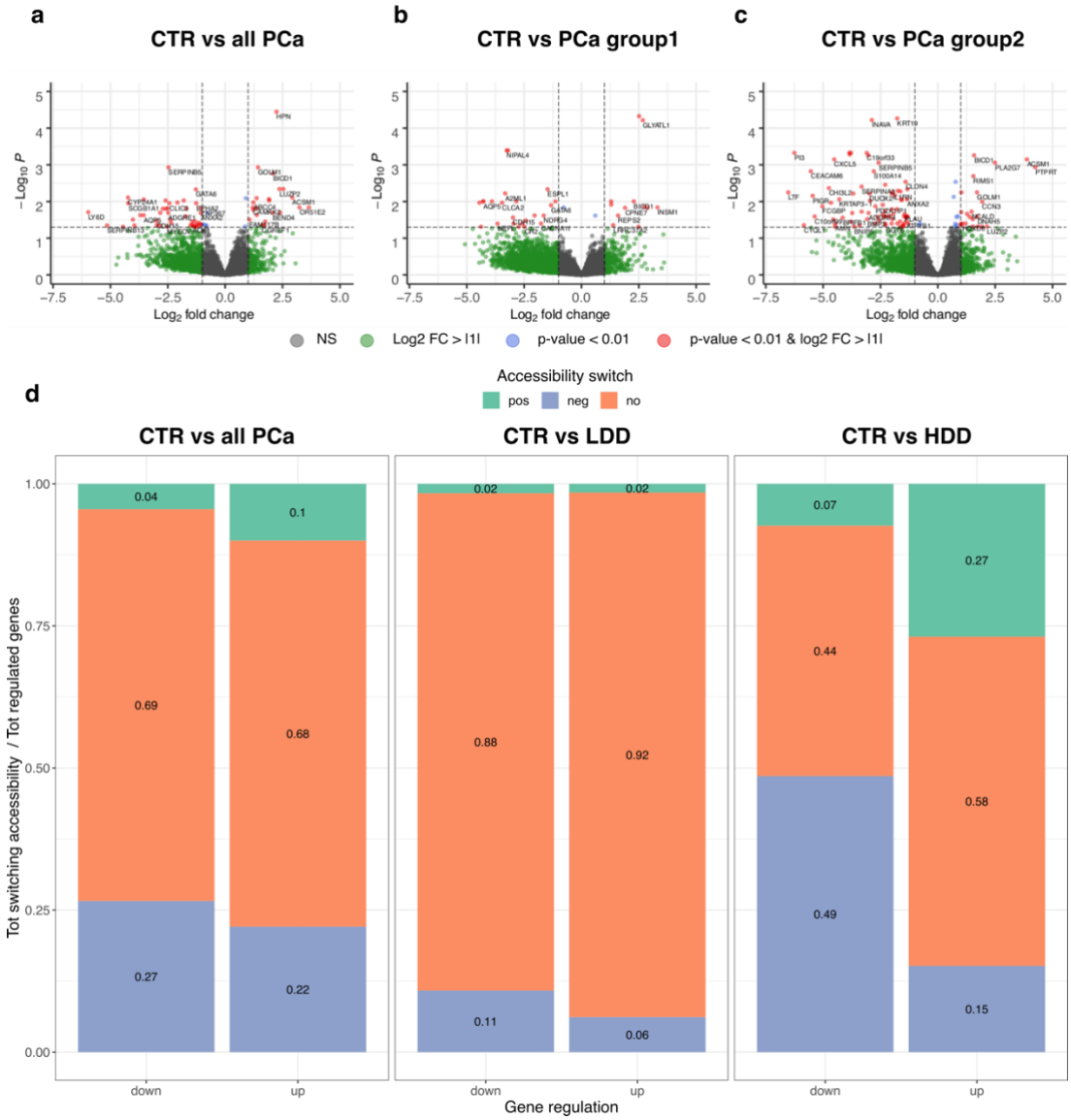


Figure 23: Epigenetic driven transcriptomic analyses determine specific PCa group features. *Volcano plot of differentially expressed genes between: (a) CTR and all PCa samples, (b) CTR and LDD samples and (c) CTR and HDD samples. Each dot represents a gene (total genes 19367): in green the genes whose log2 fold change (FC) is higher than 1 or lower than –1, while in red are represented all the genes having a log2 fold change higher than 1 or lower than –1 whose associated probability has been found statistically significant (< 0,01) and in gray the not significant ones (NS). (d) Barplot representing the ratio of up and down regulated genes among different differential expression analysis that increase (pos), decrease (neg) or do not change (no) accessibility a the TSS.*

Pathway analysis for this second group of genes reveal several pathways associated with cell migration, adhesion and epithelial mesenchymal transition (table2), often associated in literature with metastatic activity (Byles et al., 2012; Khan et al., 2015). Analysis done in collaboration with Giovanni Lembo in my group.

Table 1: *Pathway analysis on differentially expressed genes PCa group2 and CTR.*

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value
HALLMARK_ESTROGEN_RESPONSE_LATE [200]	Genes defining late response to estrogen.	5		$6.08\ e^{-5}$	$1.01\ e^{-3}$
HALLMARK_P53_PATHWAY [200]	Genes involved in p53 pathways and networks.	5		$6.08\ e^{-5}$	$1.01\ e^{-3}$
HALLMARK_TNFA_SIGNALING_VIA_NFKB [200]	Genes regulated by NF-kB in response to TNF [GeneID=7124].	5		$6.08\ e^{-5}$	$1.01\ e^{-3}$
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION [200]	Genes defining epithelial-mesenchymal transition, as in wound healing, fibrosis and metastasis.	4		$7.97\ e^{-4}$	$7.97\ e^{-3}$
HALLMARK_KRAS_SIGNALING_UP [200]	Genes up-regulated by KRAS activation.	4		$7.97\ e^{-4}$	$7.97\ e^{-3}$

In the previous sections I demonstrated that SAMMY-seq is reliable to work on primary cells and with scarce material (10 thousand cells). In our work on prostate cancer patients biopsies, we leveraged both these two features 1) to analyse the ensemble of tumour microenvironment and 2) to set up an experimental design that allows tissue characterization according to multiple aspects: cell composition, tumour grade, epigenetics and transcriptomics (fig15).

Our analysis was based on 17 samples divided in 7 CTRs and 10 PCas (fig16). Epigenetically the PCa patients could be distinguished in two groups according to their similarity with CTRs (fig21a,b). From the medical point of view (PSA levels, Gleason score, age, etc.) the two groups of patients were found heterogenous (fig16) while they all have a similar tissue composition (fig17). Inspecting instead the differential expression, we found that the PCas epigenetically different from CTR have a higher number of differentially expressed genes compared to the other group (fig23). Taking into account these data with the fact that the vast majority of the cells composing our tissue are stromal (~70%) we can assume that we are actually capturing chromatin profiles mostly determined by the tumour environment.

Thus, considering the consistency across epigenetics and transcriptomic pattern in the HDD (fig21b), we can hypothesize that the tumour microenvironment could have a role in prostate cancer development and/or progression, this is in line with recent

findings from literature (Berglund et al., 2018; Bonollo et al., 2020; González et al., 2022; Karkampouna et al., 2020; Levesque & Nelson, 2018; Mo et al., 2018; Tyekucheva et al., 2017).

4.4 Complete characterization of chromatin compartments by extending SAMMY-seq to dissect the biochemical properties of multiple chromatin fractions

In this final section of results, I am presenting a new SAMMY-seq protocol developed to further improve our chromatin characterization. In particular, we aimed to achieve, in a single experiment, the reliability of detecting closed chromatin as for the 3f protocol and the precision in detecting open chromatin as for the 4f protocol. This effort has produced the 6f SAMMY-seq protocol, a mix of 3f and 4f. With this new SAMMY-seq version we have been able to directly detect both eu- and hetero-chromatin and combine the information to call genomic compartments on the bases of chromatin biochemical properties. These compartments are similar to the ones detected with Hi-C and provide information on the influence of solubility on the compartmentalization. However, the double digestion setup required for the 6f protocol is difficult to optimize and to achieve stable performances, therefore for later work we focused on combinations of 3f and 4f protocols performed in parallel and combining their data.

4.4.1 Limits of SAMMY-seq 4f protocol: closed chromatin is not directly determined by low solubility fractions

As previously discussed, in the 4f protocol we are able to detect chromatin accessibility along the genome mainly focusing on euchromatin. While this method has always provided good results, characterizing directly with equal level of details both chromatin regions could be an advantage and in general could give us more precise information about the genomic domains we are studying.

4.4.2 Description of the SAMMY-seq 6f protocol: overcoming 3f and 4f limits

Thus, considering that in the previously presented protocols we have been able to detect with more detail closed chromatin (3f) or with more detail open chromatin (4f),

we decided to mix the best features of the two protocols in a new one (in collaboration with Dr Lanzuolo's group).

Again, as for the development of the 4f protocol, we considers the S2 collection step the critical one that could influence all the experiment results. In particular we focused our attention on the choice of the DNase: in the new protocol we maintain the DNA digestion with the DNase I (Ambion) to collect the S2, but before the S3 collection, we add a second digestion using the Turbo DNase. We expected that a first digestion with Ambion DNase could allow us to collect open chromatin while a second digestion with the Turbo DNase could digest a larger fraction of the remaining open chromatin DNA fragments so as to enrich more markedly for closed chromatin in the subsequent fractions (S3 and S4).

In this way we produced the SAMMY-seq 6 fractions protocol, where two S2 fractions are collected (S2-1 after Ambion DNase digestion, S2-2 after Turbo DNase digestion) at first and then the S3 and S4 as for the other protocols described. The two S2 are then divided according their fragment size in S2S-1 and S2L-1 for S2-1 and S2S-2 and S2L-2 for S2-2, obtaining in total 6 fractions containing genomic DNA (S2S-1, S2L-1, S2S-2, S2L-2, S3 and S4). All the three SAMMY-seq protocols are described in (table2).

Table2: Summary of SAMMY-seq three protocols.

	Permeabilize and isolate nuclei (Triton X-100)	Fragment DNA (DNase I)		Fragment DNA (Turbo DNase)		Dissolve bonds (NaCl)	Denature proteins (Urea)
3f-SAMMY	S1			S2		S3	S4
4f-SAMMY	S1	size sep. S2S	S2L			S3	S4
6f-SAMMY	S1	size sep. S2S	S2L	size sep. S2S-2	S2L-2	S3	S4

4.4.3 Applied on fibroblasts, the 6f protocol overcomes the limits affecting the 3f and 4f protocols

The DNA fragments collected in the different fractions of SAMMY-seq 6f have a size greater than 7000bp, similar to the fragments collected in the fractions of SAMMY-seq 3f and 4f (fig24a). These fragments are then sonicated to reach the size

of 200bp, which then extends due to adapter ligation during the library preparation (fig24b).

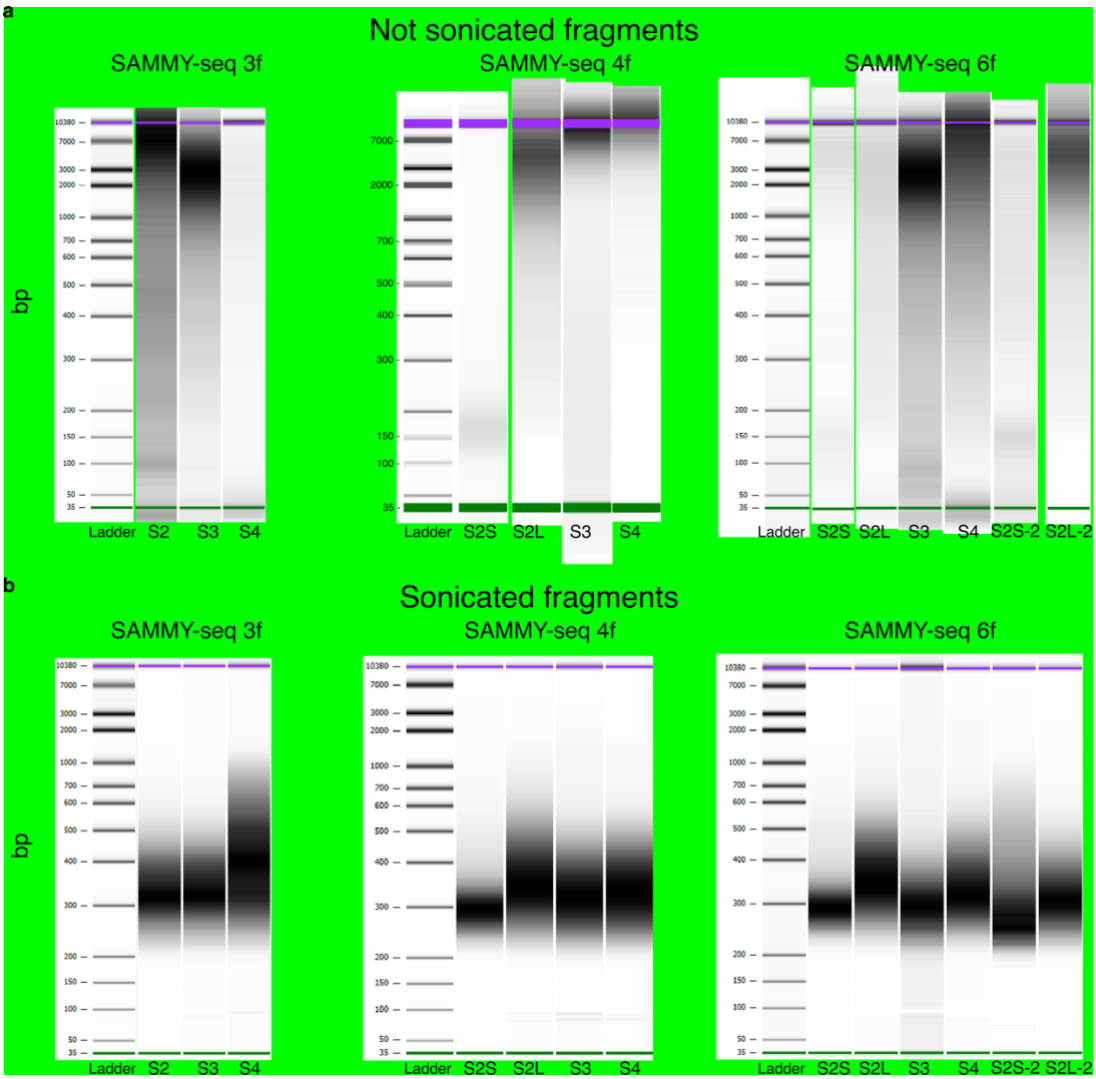


Figure 24: Bioanalyzer runs for the three SAMMY-seq protocols. *On the X axis the SAMMY-seq fractions for the protocol have been represented while in the Y axis there is the base pairs corresponding to several position of the run. (a) Fragments size distribution before sonication. (b) Fragments size distribution after sonication and adapter ligation.*

As for SAMMY-seq 4f protocol the S2 fractions are divided in two subfractions before sonication. The S2S and S2S-2 fragments, that are selected of a size smaller than 200bp, do not need sonication; while the S2L and S2L-2, that contain S2 fragments larger than 200bp, are sonicated as S3 and S4.

We tested this new protocol on fibroblast comparing its results with the ones from 3f and 4f. From our analysis of correlation, we found that the single fractions correlate directly with open and closed chromatin without performing a log ratio among fractions (fig25). It is important to remark that these should be considered preliminary results as we are still working on the protocol optimization.

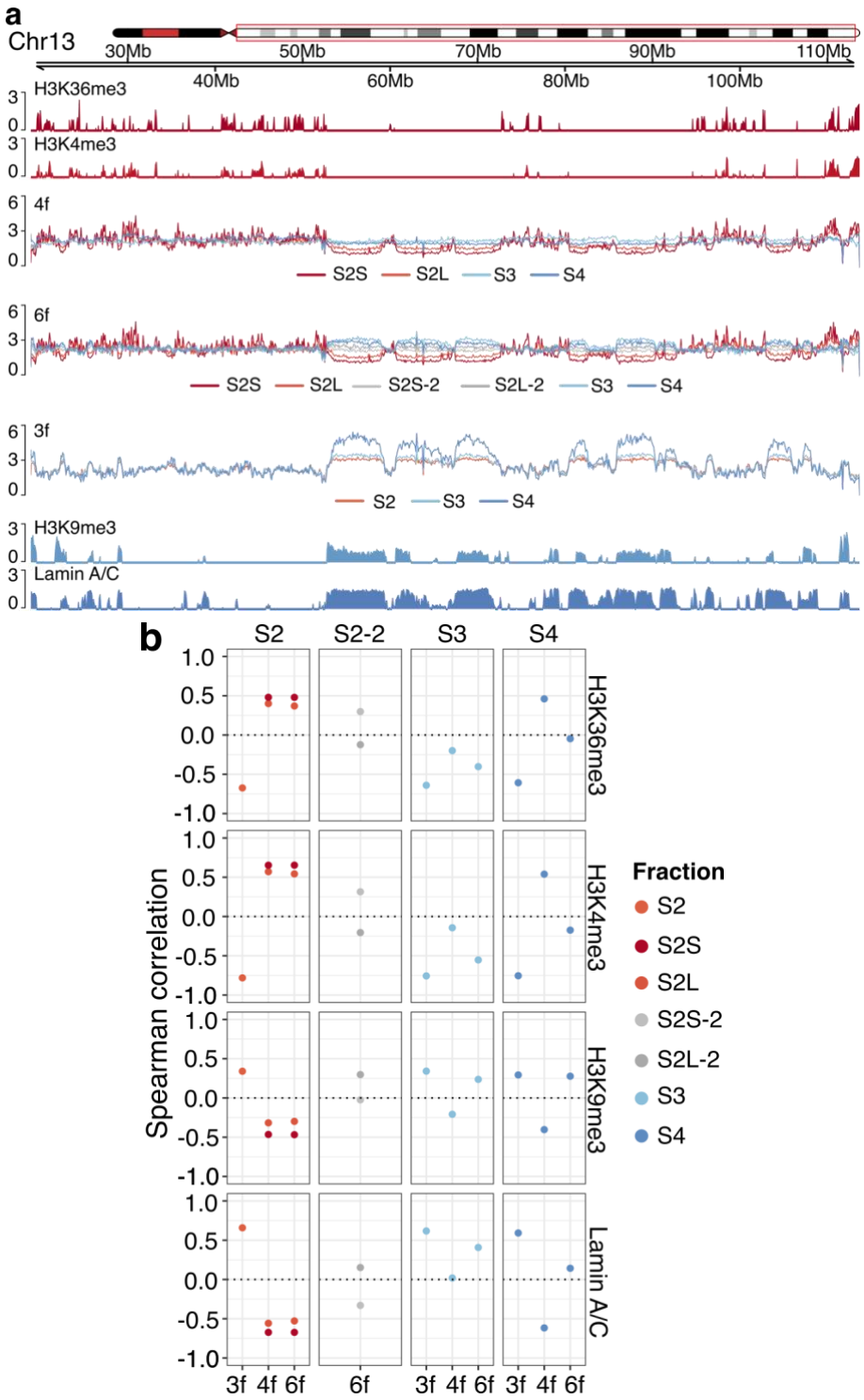


Figure 25: SAMMY-seq 6f protocol detect open chromatin in the S2 fractions and closed chromatin in S3 and S4 fractions. *(a) Distribution along chr13:18,923,594-113,678,690 of IP over INPUT of two open chromatin marks (H3K36me3, H3K4me3), reads coverage of three SAMMY-seq experiments on fibroblasts where different protocols have been applied (from top to bottom are represented the fractions overlapping per protocol of 4f, 6f and 3f) and IP over INPUT of two closed chromatin marks (H3K9me3, Lamin A/C). (b) Genome-wide Spearman correlation (Y axis) for all the fractions (column) of all protocols (X axis) against different chromatin marks (rows).*

This result shows that with this new protocol we reached the goal of directly collecting open and closed chromatin accessibility information.

4.4.4 SAMMY-seq allows characterizing chromatin compartments

With this new protocol we developed a new method to combine all the information coming from the different fractions, despite this method has been developed for the 6f protocol it could be used even for all the other version of SAMMY-seq. This approach consists in the computation of pair wise correlation matrix based on SAMMY-seq values (see methods). On this matrix we calculate the first eigenvector performing the principal component analysis as in (Lieberman-Aiden et al., 2009); this method has been method developed in collaboration with Elisa Salviato and Koustav Pal. The produced eigenvector is then the combination of all fractions information.

By summarizing the information in this way, we expected to have a clear summary of chromatin accessibility along the genome similarly to the one obtained from Hi-C analysis. We found indeed a similarity between the SAMMY-seq 6f eigenvector and thee Hi-C one (fig26). However, there's a crucial difference between the chromatin compartmentalization (eigenvector track) obtained from Hi-C, which is based on the interaction frequency profiles, vs the one obtained from SAMMY-seq. Indeed, SAMMY-seq compartmentalization is based on the similarity of biochemical properties shared by distant genomic regions, as opposed to the similarity in contact profiles. Thus, this result is even more remarkable and attesting to the fact that distant genomic regions may be located in similar biochemical environments and be close to each other at the same time.

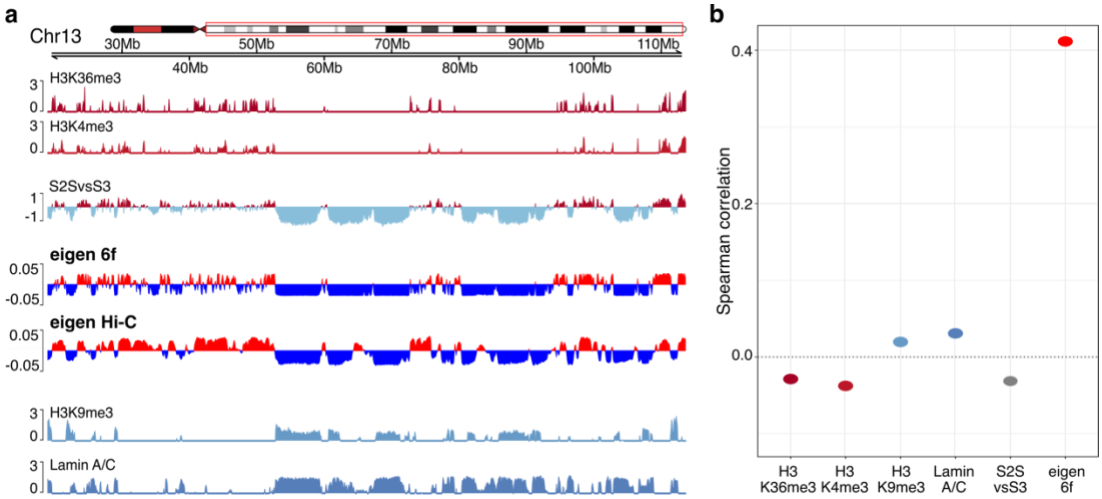


Figure 26: SAMMY-seq 6f partially recapitulate the compartments detected with Hi-C. *(a) Distribution alogn chr13:18,923,594-113,678,690 of: IP over INPUT for open chromatin marks (H3K36me3, H3K4me3), the log ratio S2SvsS3 of the SAMMY-seq 6f experiment on fibroblasts, the eigen vector values calculated from the SAMMY-seq 6f experiment on fibroblast, the eigen vector values calculated on Hi-C experiment on fibroblast and IP over INPUT of two closed chromatin marks (H3K9me3, Lamin A/C). (b) Genome-wide Spearman correlation (Y axis) of four histone marks, SAMMY-seq log ratio S2SvsS3 of 6f protocol, the eigen vector calculated from SAMMY-seq 6f experiment on fibroblasts (X axis) against SAMMY-seq 6f and Hi-C eigen vectors (rows).*

In literature has been hypothesized that compartments formation could be driven by phase separation (Erdel & Rippe, 2018; Nomoto et al., 2021; M. Shi et al., 2021; Shin et al., 2018), but there is not a general consensus about this, especially as the mechanisms enucleating compartments is subject of open discussion (Musacchio, 2022). Thus, considering that with SAMMY-seq 6f we are able to detect genomic compartments leveraging the solubility feature of the genomic regions, we decided to investigate over the effects of solubility alteration in compartment organization.

In the current model, molecular solubility has been proposed as the main driver of the phase separation (Banani et al., 2017; Mehta & Zhang, 2022). Therefore, we tried to alter the phase separation modifying the solubility of H3K27ac marked chromatin regions. As BRD4 has been shown to form aggregates in the nucleus (Sabari et al., 2018), we performed an experiment where a sample of cells were treated with the JQ1, a drug reducing the H3K27ac-BRD4 interactions (X. Shi et al., 2018). The smaller number of H3K27ac-BRD4 bounds is expected to lead to a lower number of condensates in euchromatin regions (Sabari et al., 2018, 2020) and an increase of free acetylated groups; both these elements could increase open chromatin solubility (Perry & Chalkley, 1981b) and consequently have an effect on phase separation. We performed on the treated cell an experiment of 6f SAMMY-seq, using as control an

experiment in which the cells were treated with the solvent DMSO instead of JQ1 drug.

Studying the results of this analysis we found that, as expected, JQ1 caused an alteration in the genome solubility, making more accessible the domains associated to H3K27ac (fig27a). We then investigate if this alteration has an effect on the genome compartmentalization and, indeed, we found that samples treated with JQ1 have compartments more defined considering the sample treated with DMSO (fig27b). This in line with similar results obtained by (L. Xie et al., 2022) where a similar experiment was performed based on Hi-C data. We obtained similar results in other replicate experiments (not shown), moreover we are in the process of confirming these preliminary results with other SAMMY protocol versions as well.

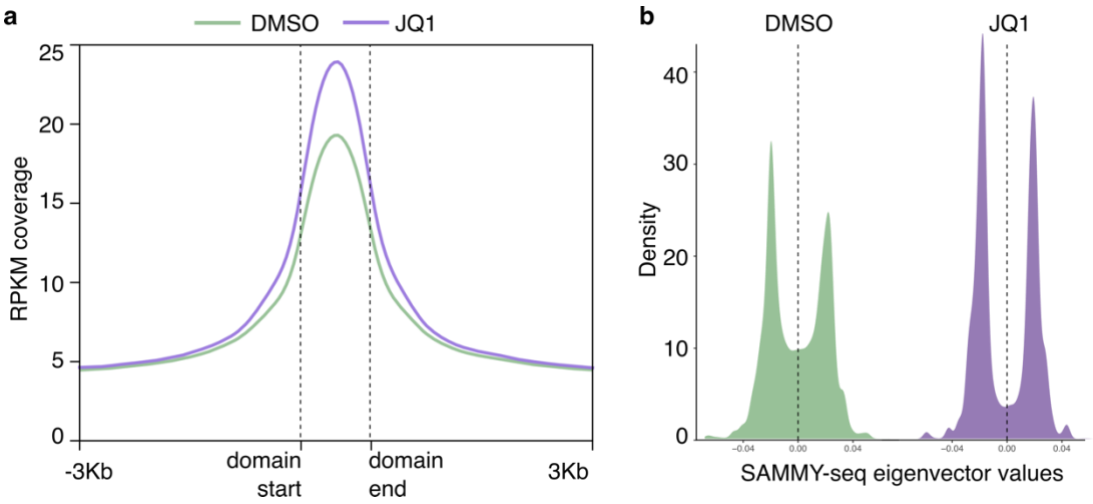


Figure 27: JQ1 cell treatment increase solubility in H3K27ac domains and lead to stronger compartmentalization than DMSO treatment. (a) Enrichment profiles of S2S fraction over H3K27ac domains (detected with MACS2) on a DMSO treated sample (purple line) and a JQ1 treated sample (green line). (b) Distribution of eigenvector values calculated on a SAMMY-seq 6f experiment performed over DMSO (green) and JQ1 (purple) treated samples.

These results suggest that SAMMY-seq is actually able to provide information over genomic compartments directly leveraging on the different solubility environment along the genome.

4.4.5 Studying compartments across prostate samples refines previous SAMMY-seq samples classification

The promising results obtained combining the SAMMY-seq 6f fractions to study compartments prompted us to perform a similar analysis on our prostate cancer datasets, even if using only the SAMMY-seq 4f protocol.

From the results we get (fig28) we have been able again to refine the samples classifications previously shown. Namely, two cancer samples are re-classified when taking into account the whole complexity of chromatin compartmentalization, as assessed by SAMMY-seq chromatin compartment analysis. Indeed, the new clustering over compartment distribution similarity show that the LDD gain the PCa100 patient and loose the PCa4 patient (obviously these changes affect the HDD consequently).

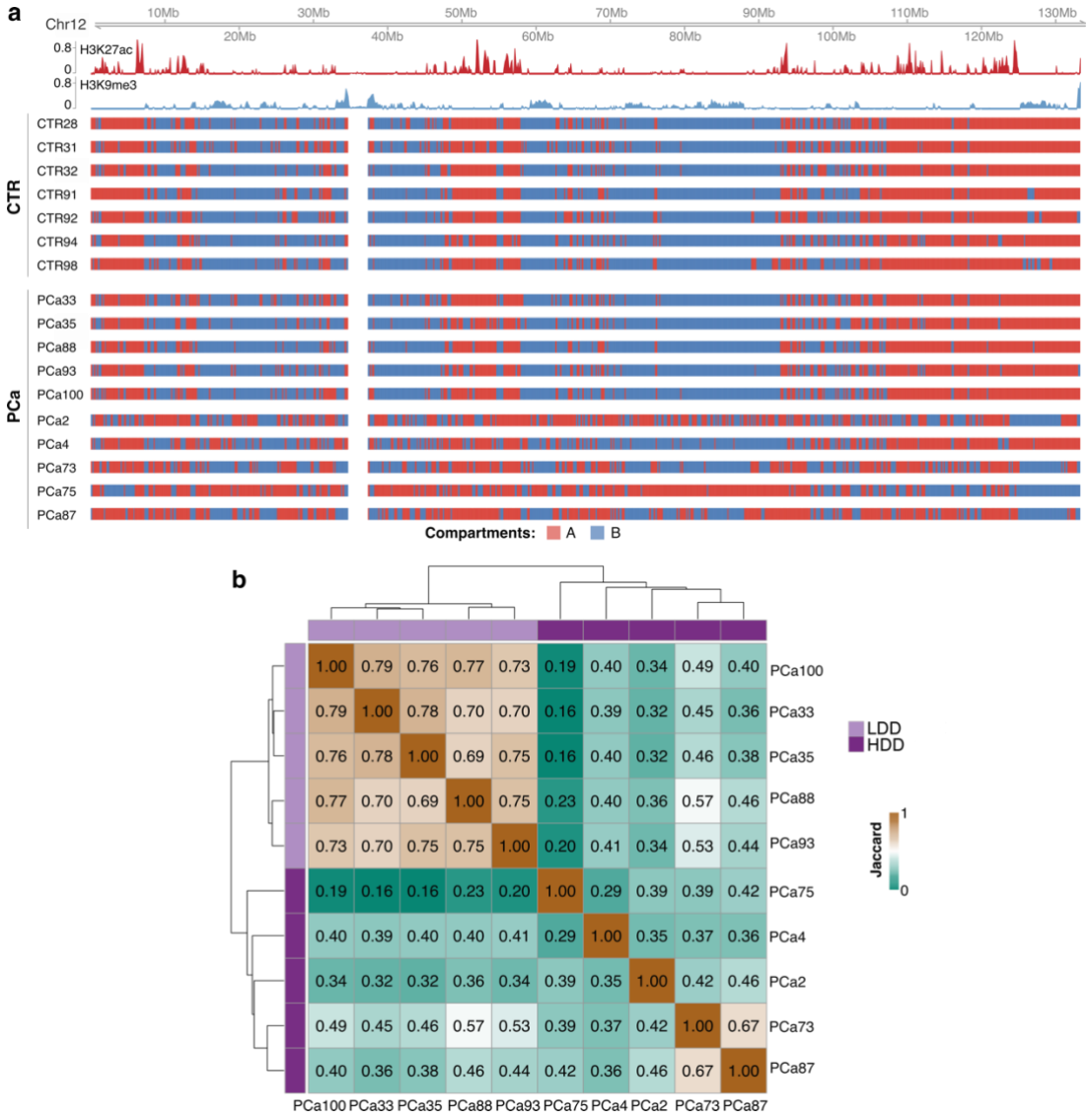


Figure 28: Compartments calling on prostate 4f experiments divide tumour samples in two groups. *(a) Visual representation on chromosome 12 of (from top to bottom) two ChIP-seq experiments (SPP reads coverage ratio), compartments distribution along the genome of CTR samples and PCa samples (in red A compartments are represented, while in blue the B ones). (b) Matrix representing similarity of compartments distribution between pair of patients (described through the jaccard index). Patient have been reordered according to a hierarchical cluster based on jaccard index.*

Performing the differential expression analysis among the groups (fig29) we actually found a much higher number of significant differentially expressed genes associated

to HDD, attesting to the fact that the two groups are more markedly different in terms of expression profiles.

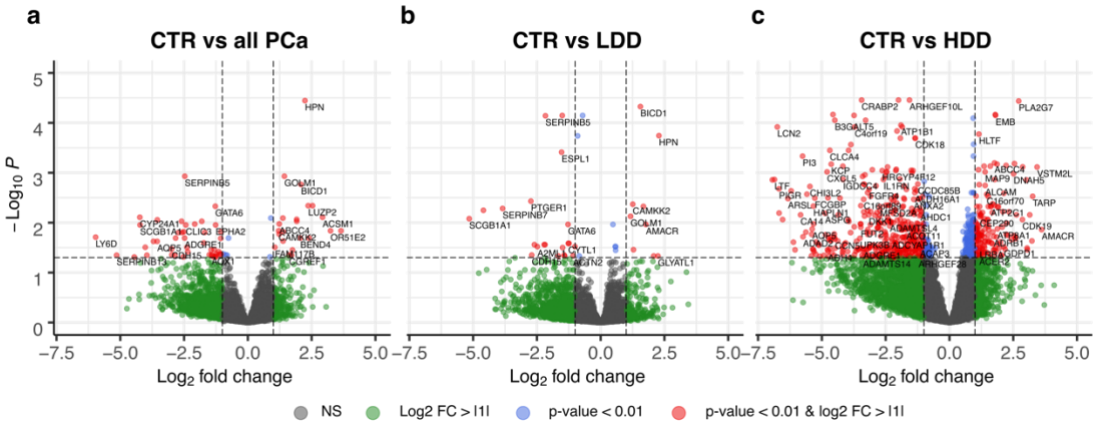


Fig29: Epigenetic driven transcriptomic analyses determine specific PCa group features. *Volcano plot of differentially expressed genes between: (a) CTR and all PCa samples, (b) CTR and LDD samples and (c) CTR and HDD samples. Each dot represents a gene (total genes 19367): in green the genes whose log2 fold change (FC) is higher than 1 or lower than –1, while in red are represented all the genes having a log2 fold change higher than 1 or lower than –1 whose associated probability has been found statistically significant (< 0,01) and in gray the not significant ones (NS).*

In particular, performing a gene ontology on that genes (fig30), we found that in the Cancer Cell Line Encyclopedia (CCLE) proteomics and in ChIP Enrichment Analysis (ChEA) a correspondence between our samples and vertebral cancer of the prostate (VCaP) cell line, a line derived from prostate cancer bone metastasis tumor (Korenchuk et al., 2001), is highlighted.

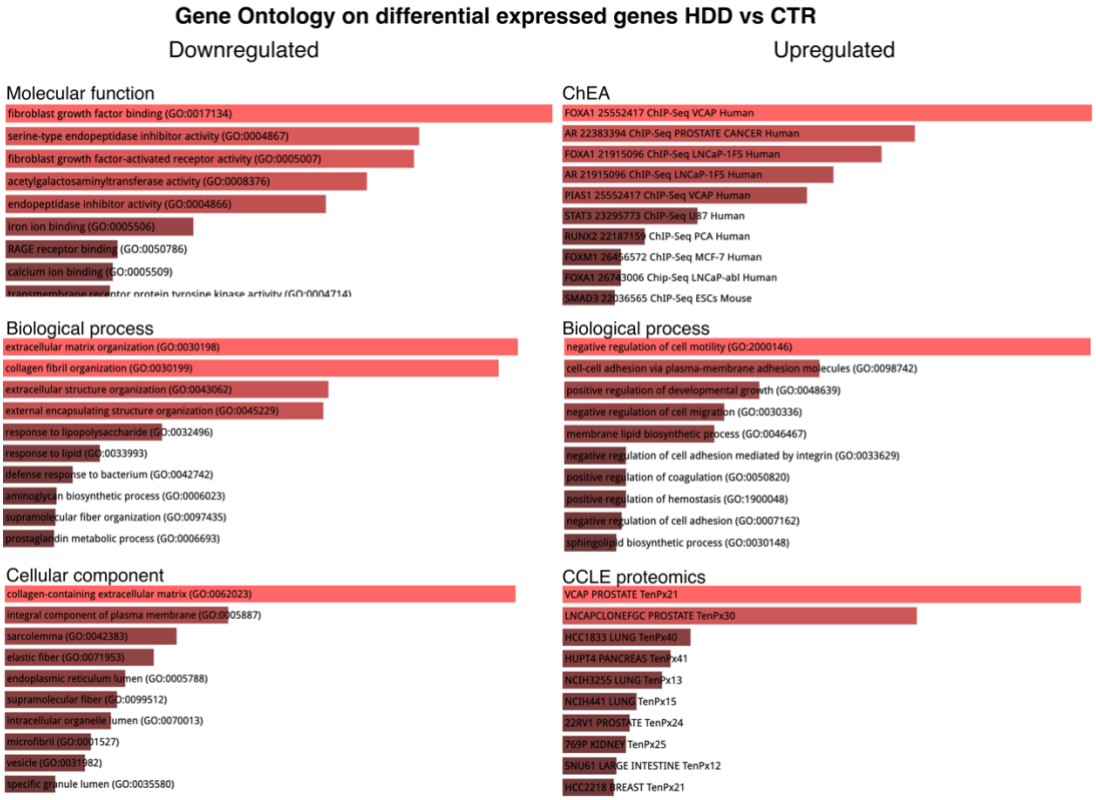


Figure 30: HDD patients show enrichment GO terms associated to EMT. *In the picture are represented the results of gene ontology term enrichment for gene found differentially expressed between HDD patients and CTR samples. The length of the bar and the brightness represent the level of significance of the GO associated.*

4.5 Author contribution

In this section I presented several results that have been obtained in collaboration with other colleagues and groups. Throughout these collaborative interdisciplinary projects my contribution has always been as the main or one of the two main researchers leading the bioinformatic data analyses. As such, I never performed personally the experiments in the lab even if I participated in their planning, troubleshooting and interpretation of results.

The first experiments that I presented were concerning the SAMMY-seq 3f protocol and its application on progeria, the work has been done in collaboration with Dr. Lanzuolo's group. I joined the project at the beginning of my PhD when it was already started. My involvement was principally on the scale-down analysis (fig5), the study on the redistribution of H3K27me3 signal around the TSS (fig8), I did analysis on progeria samples at late passages (data not shown), I contributed in the setting up of the protocol presented in the paper (data not shown). In the resulting article (cit Sebestyen et al., Nat Comm 2020) I am listed as second author.

The other test case I presented, the 3f SAMMY-seq protocol was on MCF10DICIS.com. This was a collaboration with Dr. Giorgio Scita's lab. My role has been to analyse the SAMMY-seq and ChIP-seq experiments produced in the lab, to identify additional public datasets to be used in comparatvie analyses that I did. In the article (in press) in which these analyses are present (Frittoli et al., 2021) I am a middle author.

The setting up of the 4f SAMMY-seq protocol, together with the 6f protocol variantare part of the same project on which I am working as co-first author of the manuscript in preparation, in collaboration with Dr. Lanzuolo's group. All the analyses presented here are performed by me, except for the calling of compartments on fibroblasts. In addition, I contributed to the experiments planning and troubleshooting .

Finally, for what concerns the prostate cancer application, I will be the co-first author in collaboration with Dr. Lanzuolo's lab on the manuscript in preparation arising from this project. I did all the analyses concerning SAMMY-seq (including chromatin compartments calling), while the analyses on differential gene expression and pathway analyses have been performed by Giovanni Lembo, a colleague in the lab.

Throughout these interdisciplinary collaborations, I always participated in the discussions and interpretation of results, I presented the work done in multiple conferences and in formal joint lab meetings with other groups.

5 Discussion

Since its first definition epigenetics has continued to expand into a complex scientific field, multiple mechanisms at multiple resolution are now been described (from DNA modification to chromosome territories). Comprehension of epigenetics mechanisms has been fundamental even for the understanding of diseases such as laminopathies and cancer. Such relevance of epigenetics in disease led to suggest regulatory mechanisms as diagnostic or therapeutical targets.

In parallel with the interest for epigenetics, novel experimental techniques to study these regulatory mechanisms have been developed and continuously improved. However, most genome-wide methods are affected by practical and technical limitations that reduce their applicability (i.e. high costs, requirement of abundant starting material, impossibility to work on primary cells and chemical modifications that may lead to artifacts generation).

In this context, our lab, in collaboration with Dr Chiara Lanzuolo's lab, developed SAMMY-seq a new technique to overcome the previously cited limitations (table3) and give new insight in epigenetics. This method leverages on biochemical features of different chromatin regions allowing their separate collection and analysis.

Table 3: SAMMY-seq compared with other genome-wide methods to map heterochromatin and LADs (adapted from Sebestyèn et al.).

	DamID	ChIP-seq H3K9me3	ChIP-seq Lamin A or B	SAMMY -seq
Can be applied on primary cells and tissue		YES	YES	YES
Can be applied on single cell	YES			
Can be applied on small number of cells (10K)	YES	(*)		YES
Avoid the use of antibody	YES			YES
Avoid the use of chemical modification of chromatin	YES	(**)		YES
Avoid transfection of constructs		YES	YES	YES
Time to perform the protocol	**(***) 2 days**	**2 days**	**2 days**	**3 hours**

Notes in the table:

(*) Possible with non-standard protocol (ChIPmentation) (Schmidl et al., 2015)

(**) Possible with a non-standard protocol (N-ChIP) (O'Neill & Turner, 2003)

(***) Construct cloning, production and transfection not included

From our tests on methods reliability, we found that it is highly reproducible (fig3.4,5) and it is scalable up to 10K cells (fig5). Furthermore, the results are in line with literature data coming from ChIP-seq experiments performed on histone marks (fig3,4).

At first the main focus of this technique was the heterochromatin regions, with particular focus on lamina associated domains (LADs). Thus, we applied it on model systems where LADs alteration would be expected according to literature: Hutchinson-Gilford progeria syndrome and MCF10DICIS Rab5 over-expressing cells. Our results in both systems demonstrate that SAMMY-seq is able to detect heterochromatin rearrangements where other techniques did not i.e. ChIP-seq (fig7,8) and Hi-C (Falk et al., 2019). In particular, in progeria we found that the lamina alterations do not have an immediate effect on H3K9me3 distribution, but instead affect the H3K27me3 deposition by Polycomb (fig9) with a consequent deregulation of Polycomb associated genes.

While SAMMY-seq 3f was mainly developed to investigate over heterochromatin related mechanisms, we produced a new version of the protocol keeping all the advantages of the old one but greatly increasing our ability to describe euchromatin regions (fig10).

Our most relevant application of this protocol has been on prostate cancer. The peculiar features of SAMMY-seq not only allowed to investigate, for the very first time, over epigenetics of open and closed chromatin in prostate cancer biopsies; but they even allow an experimental setup to completely characterize a patient from its tissue taking into account: morphology, biopsy cell composition, epigenetics and transcriptomic.

Our analysis demonstrates that SAMMY-seq is a reliable method to study chromatin solubility and provides information about epigenetic mechanisms in conditions that are challenging if not impossible for most genome-wide analysis methods commonly used in literature. Our analysis highlights how epigenetics is fundamental to understand the cancer progression and its association to tumor hallmark mechanisms (i.e. metastases). Indeed here we found a potential connection between epigenetics alterations in prostate cancer and epithelial mesenchymal transition (EMT), a topic already discussed in literature but never tested on tissue samples (Chaves et al., 2021). These results may even suggest that the experimental design applied in this work has

a clinical potential for better characterization of patients and to determine the more appropriate treatment to cure the tumor.

With a further version of the protocol, where we characterized more precisely both eu- and hetero-chromatin, we have been able to retrieve information of genomic compartments with similar results obtained with Hi-C. This finding has been particularly interesting cause, while Hi-C study 3D genomic contacts, SAMMY-seq leverages on genome accessibility; suggesting, in line with other models described in literature, that the solubility is the mechanisms promoting the compartment formation. Considering that, we could suggest the solubility itself as a new element to be included among the epigenetic mechanisms.

In conclusion the approach of chromatin fractionation adopted by SAMMY-seq has revealed to be a powerful method to study epigenetics. Its potential is its ability to focus on all biochemical chromatin features along the genome instead of focusing on target elements as for DamID, TSA-seq and ChIP-seq. Respect to other similar approaches as DNase-seq and ATAC-seq it captures both hetero- and eu-chromatin information.

The three protocols here presented are complementary and could be used for different scopes. 3f protocol is more suitable for studies focusing on heterochromatin, while it is more advisable use 4f for works focusing mostly on euchromatin. 6f instead provides a good insight on both open and closed chromatin, feature that makes it more suitable for compartment calling based on fraction combining. Unfortunately, it has never been tested on a limited number of cells and in general the use of two different DNase sequentially makes the protocol more complicated to be performed.

Based on our experience each analysis on epigenetics done with SAMMY-seq should be combined with analysis of gene expression (e.g. RNA-seq) to be able to connect the effect of epigenetics on the transcription and give additional insight over the mechanisms under investigation, as presented in the study of prostate cancer or progeria. Researcher should take advantage of SAMMY-seq for all those epigenomics studies involving multiple mechanisms that interest broadly the genome (e.g. perturbation of the nuclear environment through JQ1).

6 Bibliography

Adli, M., & Bernstein, B. E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature Protocols 2011 6:10*, *6*(10), 1656–1668. https://doi.org/10.1038/nprot.2011.402

Alberti, S., Gladfelter, A., & Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell*, *176*(3), 419–434. https://doi.org/10.1016/J.CELL.2018.12.035

Anastasiadou, E., Jacob, L. S., & Slack, F. J. (2017). Non-coding RNA networks in cancer. *Nature Reviews Cancer 2017 18:1*, *18*(1), 5–18. https://doi.org/10.1038/nrc.2017.99

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), 1–12. https://doi.org/10.1186/GB-2010-11-10-R106/COMMENTS

Atlasi, Y., & Stunnenberg, H. G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews Genetics 2017 18:11*, *18*(11), 643–658. https://doi.org/10.1038/nrg.2017.57

Aughey, G. N., Cheetham, S. W., & Southall, T. D. (2019). DamID as a versatile tool for understanding gene regulation. *Development (Cambridge)*, *146*(6). https://doi.org/10.1242/DEV.173666/49075

Ausio, J., Borochov, N., Seger, D., & Eisenberg, H. (1984). Interaction of chromatin with NaCl and MgCl2. Solubility and binding studies, transition to and characterization of the higher-order structure. *Journal of Molecular Biology*, *177*(3), 373–398. https://doi.org/10.1016/0022-2836(84)90291-2

Ausio, J., Sasi, R., & Fasman, G. D. (1986). Biochemical and physiochemical characterization of chromatin fractions with different degrees of solubility isolated from chicken erythrocyte nuclei. *Biochemistry*, *25*(8), 1981–1988. https://doi.org/10.1021/BI00356A022

Ausio, J., Seger, D., & Eisenberg, H. (1984). Nucleosome core particle stability and conformational change: Effect of temperature, particle and NaCl concentrations, and crosslinking of histone H3 sulfhydryl groups. *Journal of Molecular Biology*, *176*(1), 77–104. https://doi.org/10.1016/0022-2836(84)90383-8

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *Journal of Experimental Medicine*, *79*(2), 137–158. https://doi.org/10.1084/JEM.79.2.137

Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology 2017 18:5*, *18*(5), 285–298. https://doi.org/10.1038/nrm.2017.7

Beacon, T. H., & Davie, J. R. (2021). Transcriptionally Active Chromatin—Lessons Learned from the Chicken Erythrocyte Chromatin Fractionation. *Cells*, *10*(6). https://doi.org/10.3390/CELLS10061354

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., Barbieri, M., De Santiago, I., Lavitas, L. M., Branco, M. R., Fraser, J., Dostie, J., Game, L., Dillon, N., Edwards, P. A. W., Nicodemi, M., & Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature 2017 543:7646*, *543*(7646), 519–524. https://doi.org/10.1038/nature21411

Belmont, A. S. (2022). Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. *Cold Spring Harbor Perspectives in Biology*, *14*(7), a041268. https://doi.org/10.1101/CSHPERSPECT.A041268

Berglund, E., Maaskola, J., Schultz, N., Friedrich, S., Marklund, M., Bergenstråhle, J., Tarish, F., Tanoglidi, A., Vickovic, S., Larsson, L., Salmén, F., Ogris, C., Wallenborg, K., Lagergren, J., Ståhl, P., Sonnhammer, E., Helleday, T., & Lundeberg, J. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-04724-5

Berlatzky, I. A., Rouvinski, A., & Ben-Yehuda, S. (2008). Spatial organization of a replicating bacterial chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14136–14140. https://doi.org/10.1073/PNAS.0804982105/SUPPL_FILE/SM2.AVI

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315–326. https://doi.org/10.1016/J.CELL.2006.02.041

Béroud, C., Hamroun, D., Collod-Béroud, G., Boileau, C., Soussi, T., & Claustres, M. (2005). UMD (Universal Mutation Database): 2005 Update. *Human Mutation*, *26*(3), 184–191. https://doi.org/10.1002/humu.20210

Bianchi, A., Mozzetta, C., Pegoli, G., Lucini, F., Valsoni, S., Rosti, V., Petrini, C., Cortesi, A., Gregoretti, F., Antonelli, L., Oliva, G., de Bardi, M., Rizzi, R., Bodega, B., Pasini, D., Ferrari, F., Bearzi, C., & Lanzuolo, C. (2020). Dysfunctional polycomb transcriptional repression contributes to lamin A/C-dependent muscular dystrophy. *The Journal of Clinical Investigation*, *130*(5), 2408–2421. https://doi.org/10.1172/JCI128161

Biswas, S., & Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *European Journal of Pharmacology*, *837*, 8–24. https://doi.org/10.1016/J.EJPHAR.2018.08.021

Blackledge, N. P., & Klose, R. J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. *Nature Reviews Molecular Cell Biology 2021 22:12*, *22*(12), 815–833. https://doi.org/10.1038/s41580-021-00398-y

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S., & Di Croce, L. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. *Trends in Genetics*, *36*(2), 118–131. https://doi.org/10.1016/J.TIG.2019.11.004

Bluemn, E. G., Coleman, I. M., Lucas, J. M., Coleman, R. T., Hernandez-Lopez, S., Tharakan, R., Bianchi-Frias, D., Dumpit, R. F., Kaipainen, A., Corella, A. N., Yang, Y. C., Nyquist, M. D., Mostaghel, E., Hsieh, A. C., Zhang, X., Corey, E., Brown, L. G., Nguyen, H. M., Pienta, K., … Nelson, P. S. (2017). Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell*, *32*(4), 474-489.e6. https://doi.org/10.1016/J.CCELL.2017.09.003

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Bonollo, F., Thalmann, G. N., Julio, M. K. De, & Karkampouna, S. (2020). The Role of Cancer-Associated Fibroblasts in Prostate Cancer Tumorigenesis. *Cancers*, *12*(7), 1–28. https://doi.org/10.3390/CANCERS12071887

Borochov, N., Ausio, J., & Eisenberg, H. (1984). Interaction and conformational changes of chromatin with divalent ions. *Nucleic Acids Research*, *12*. https://academic.oup.com/nar/article/12/7/3089/1088044

Boyson, S. P., Gao, C., Quinn, K., Boyd, J., Paculova, H., Frietze, S., & Glass, K. C. (2021). Functional Roles of Bromodomain Proteins in Cancer. *Cancers*, *13*(14). https://doi.org/10.3390/CANCERS13143606

Breiling, A., & Lyko, F. (2015). *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond*. https://doi.org/10.1186/s13072-015-0016-6

Briand, N., & Collas, P. (2018). Laminopathy-causing lamin A mutations reconfigure lamina-associated domains and local spatial chromatin conformation. *Https://Doi.Org/10.1080/19491034.2018.1449498*, *9*(1), 216–226. https://doi.org/10.1080/19491034.2018.1449498

Briand, N., & Collas, P. (2020). Lamina-associated domains: peripheral matters and internal affairs. *Genome Biology 2020 21:1*, *21*(1), 1–25. https://doi.org/10.1186/S13059-020-02003-5

Buchwalter, A., & Hetzer, M. W. (2017). Nucleolar expansion and elevated protein translation in premature aging. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00322-z

Budakoti, M., Panwar, A. S., Molpa, D., Singh, R. K., Büsselberg, D., Mishra, A. P., Coutinho, H. D. M., & Nigam, M. (2021). Micro-RNA: The darkhorse of cancer. *Cellular Signalling*, *83*, 109995. https://doi.org/10.1016/J.CELLSIG.2021.109995

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods 2013 10:12*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Byles, V., Zhu, L., Lovaas, J. D., Chmilewski, L. K., Wang, J., Faller, D. V., & Dai, Y. (2012). SIRT1 induces EMT by cooperating with EMT transcription factors and enhances prostate cancer cell migration and metastasis. *Oncogene*, *31*(43), 4619–4629. https://doi.org/10.1038/ONC.2011.612

Cai, B., Song, X. Q., Cai, J. P., & Zhang, S. (2014). HOTAIR: A cancer-related long non-coding RNA. *Neoplasma*, *61*(4), 379–391. https://doi.org/10.4149/NEO_2014_075

Carlsson, S. V., & Vickers, A. J. (2020). Screening for Prostate Cancer. *The Medical Clinics of North America*, *104*(6), 1051. https://doi.org/10.1016/J.MCNA.2020.08.007

Cavalli, G., & Heard, E. (2019). Advances in epigenetics link genetics to the environment and disease. *Nature 2019 571:7766*, *571*(7766), 489–499. https://doi.org/10.1038/s41586-019-1411-0

Cesarini, E., Mozzetta, C., Marullo, F., Gregoretti, F., Gargiulo, A., Columbaro, M., Cortesi, A., Antonelli, L., Di Pelino, S., Squarzoni, S., Palacios, D., Zippo, A., Bodega, B., Oliva, G., & Lanzuolo, C. (2015). Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *Journal of Cell Biology*, *211*(3), 533–551. https://doi.org/10.1083/JCB.201504035

Chalkley, R., & Jensen, R. H. (1968). Structure of isolated chromatin. *Biochemistry*, *7*(12), 4380–4388. https://doi.org/10.1021/bi00852a034

Chaves, L. P., Melo, C. M., Saggioro, F. P., Dos Reis, R. B., & Squire, J. A. (2021). Epithelial–Mesenchymal Transition Signaling and Prostate Cancer Stem Cells: Emerging Biomarkers and Opportunities for Precision Therapeutics. *Genes*,

12(12). https://doi.org/10.3390/GENES12121900

Chen, A., Chen, D., & Chen, Y. (2018). Advances of DNase-seq for mapping active gene regulatory elements across the genome in animals. *Gene*, *667*, 83–94. https://doi.org/10.1016/J.GENE.2018.05.033

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*. https://doi.org/10.1186/1471-2105-14-128

Chen, S., Zhou, Q., Liu, T., Zhang, W., Zeng, X. T., & Guo, Z. (2020). Prognostic value of downregulated 5-hydroxymethylcytosine expression in renal cell carcinoma: a 10 year follow-up retrospective study. *Journal of Cancer*, *11*(5), 1212. https://doi.org/10.7150/JCA.38283

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., Goldman, R., Van Steensel, B., Ma, J., & Belmont, A. S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of Cell Biology*, *217*(11), 4025. https://doi.org/10.1083/JCB.201807108

Coleman, W. B. (2018). The Human Genome: Understanding Human Disease in the Post-Genomic Era. *Molecular Pathology: The Molecular Basis of Human Disease*, 121–134. https://doi.org/10.1016/B978-0-12-802761-5.00006-7

Crasto, S., & Di Pasquale, E. (2018). Induced pluripotent stem cells to study mechanisms of laminopathies: Focus on epigenetics. *Frontiers in Cell and Developmental Biology*, *6*(DEC), 172. https://doi.org/10.3389/FCELL.2018.00172/XML/NLM

Cremer, T., & Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics 2001 2:4*, *2*(4), 292–301. https://doi.org/10.1038/35066075

Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 16160–16165. https://doi.org/10.1073/PNAS.1419030111/SUPPL_FILE/PNAS.1419030111.SAPP.PDF

Cui, X. L., Nie, J., Ku, J., Dougherty, U., West-Szymanski, D. C., Collin, F., Ellison, C. K., Sieh, L., Ning, Y., Deng, Z., Zhao, C. W. T., Bergamaschi, A., Pekow, J., Wei, J., Beadell, A. V., Zhang, Z., Sharma, G., Talwar, R., Arensdorf, P., … He, C. (2020). A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nature Communications 2020 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-20001-w

Dai, C., Heemers, H., & Sharifi, N. (2017). Androgen Signaling in Prostate Cancer. *Cold Spring Harbor Perspectives in Medicine*, *7*(9), a030452. https://doi.org/10.1101/CSHPERSPECT.A030452

de Leeuw, R., Gruenbaum, Y., & Medalia, O. (2018). Nuclear Lamins: Thin Filaments with Major Functions. *Trends in Cell Biology*, *28*(1), 34–45. https://doi.org/10.1016/J.TCB.2017.08.004

de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H. L., & de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell*, *60*(4), 676–684. https://doi.org/10.1016/J.MOLCEL.2015.09.023

Debaugny, R. E., & Skok, J. A. (2020). CTCF and CTCFL in cancer. *Current Opinion*

in Genetics & Development, *61*, 44–52. https://doi.org/10.1016/J.GDE.2020.02.021

Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., Mirny, L. A., O'Shea, C. C., Park, P. J., Ren, B., Ritland Politz, J. C., Shendure, J., & Zhong, S. (2017). The 4D nucleome project. *Nature 2017 549:7671*, *549*(7671), 219–226. https://doi.org/10.1038/nature23884

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*(5558), 1306–1311. https://doi.org/10.1126/SCIENCE.1067799/ASSET/CF16CFE4-EFDA-4A69-867D-220ACFB4397F/ASSETS/GRAPHIC/SE0720192003.JPEG

Denker, A., & De Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. *Genes & Development*, *30*(12), 1357–1382. https://doi.org/10.1101/GAD.281964.116

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology 2010 8:3*, *8*(3), 185–195. https://doi.org/10.1038/nrmicro2261

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature 2012 485:7398*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Donnaloja, F., Carnevali, F., Jacchetti, E., & Raimondi, M. T. (2020). Lamin A/C Mechanotransduction in Laminopathies. *Cells*, *9*(5). https://doi.org/10.3390/CELLS9051306

Dorman, C. J., & Deighan, P. (2003). Regulation of gene expression by histone-like proteins in bacteria. *Current Opinion in Genetics & Development*, *13*(2), 179–184. https://doi.org/10.1016/S0959-437X(03)00025-X

Dostie, J., Zhan, Y., & Dekker, J. (2007). Chromosome Conformation Capture Carbon Copy Technology. *Current Protocols in Molecular Biology*, *80*(1), 21.14.1-21.14.13. https://doi.org/10.1002/0471142727.MB2114S80

Drlica, K., & Rouviere-Yaniv, J. (1987). *Histonelike Proteins of Bacteria*. https://journals.asm.org/journal/mr

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature 2012 489:7414*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Dworak, N., Makosa, D., Chatterjee, M., Jividen, K., Yang, C. S., Snow, C., Simke, W. C., Johnson, I. G., Kelley, J. B., & Paschal, B. M. (2019). A nuclear lamina-chromatin-Ran GTPase axis modulates nuclear import and DNA damage signaling. *Aging Cell*, *18*(1), e12851. https://doi.org/10.1111/ACEL.12851

Ellinger, J., Kahl, P., Von Der Gathen, J., Rogenhofer, S., Heukamp, L. C., Gütgemann, I., Walter, B., Hofstädter, F., Büttner, R., Müller, S. C., Bastian, P. J., & Von Ruecker, A. (2010). Global levels of histone modifications predict prostate cancer recurrence. *The Prostate*, *70*(1), 61–69. https://doi.org/10.1002/PROS.21038

Emamjomeh, A., Choobineh, D., Hajieghrari, B., MahdiNezhad, N., & Khodavirdipour, A. (2019). DNA–protein interaction: identification, prediction and data analysis. *Molecular Biology Reports 2019 46:3*, *46*(3), 3571–3596. https://doi.org/10.1007/S11033-019-04763-1

Erdel, F., & Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal*, *114*(10), 2262–2270. https://doi.org/10.1016/J.BPJ.2018.03.011

Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods 2012 9:3*, *9*(3), 215–216. https://doi.org/10.1038/nmeth.1906

Fabiao de Lima, M., Lisboa, M. de O., Terceiro, L. E. L., Rangel-Pozzo, A., & Mai, S. (2022). Chromosome Territories in Hematological Malignancies. *Cells*, *11*(8), 1368. https://doi.org/10.3390/CELLS11081368

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., & Mirny, L. A. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature 2019 570:7761*, *570*(7761), 395–399. https://doi.org/10.1038/s41586-019-1275-3

Fawcett, D. W. (1966). On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates. *American Journal of Anatomy*, *119*(1), 129–145. https://doi.org/10.1002/AJA.1001190108

Feingold, E. A., Good, P. J., Guyer, M. S., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F. S., Gingeras, T. R., Kampa, D., Sekinger, E. A., Cheng, J., Hirsch, H., Ghosh, S., Zhu, Z., Patel, S., Piccolboni, A., Yang, A., Tammana, H., Bekiranov, S., … Harvey, S. C. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, *306*(5696), 636–640. https://doi.org/10.1126/SCIENCE.1105136

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., Mirny, L., & Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, *176*(6), 1502-1515.e10. https://doi.org/10.1016/J.CELL.2019.01.020

Fisher, J. K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., & Kleckner, N. (2013). Four-Dimensional Imaging of E. coli Nucleoid Organization and Dynamics in Living Cells. *Cell*, *153*(4), 882–895. https://doi.org/10.1016/J.CELL.2013.04.006

Fraser, M., Sabelnykova, V. Y., Yamaguchi, T. N., Heisler, L. E., Livingstone, J., Huang, V., Shiah, Y. J., Yousif, F., Lin, X., Masella, A. P., Fox, N. S., Xie, M., Prokopec, S. D., Berlin, A., Lalonde, E., Ahmed, M., Trudel, D., Luo, X., Beck, T. A., … Boutros, P. C. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. *Nature 2016 541:7637*, *541*(7637), 359–364. https://doi.org/10.1038/nature20788

Frittoli, E., Palamidessi, A., Iannelli, F., Zanardi, F., Villa, S., Barzaghi, L., Abdo, H., Cancila, V., Beznoussenko, G. V, Della Chiara, G., Pagani, M., Malinverno, C., Bhattacharya, D., Pisati, F., Yu, W., Galimberti, V., Bonizzi, G., Martini, E., Mironov, A. A., … Scita, G. (2022). Tissue fluidification promotes a cGAS–STING cytosolic DNA response in invasive breast cancer. *Nature Materials 2022*, 1–12. https://doi.org/10.1038/s41563-022-01431-x

Frittoli, E., Palamidessi, A., Iannelli, F., Zanardi, F., Villa, S., Barzaghi, L., Ando, H., Cancila, V., Beznuskenko, G., Della Chiara, G., M, P., C, M., D, B., F, P., W, Y., V, G., G, B., E, M., A, M., … G, S. (2021). *Tissue fluidification promotes a*

cGAS/STING-mediated cytosolic DNA response in invasive breast cancer. https://doi.org/10.21203/RS.3.RS-845173/V1

Frittoli, E., Palamidessi, A., Marighetti, P., Confalonieri, S., Bianchi, F., Malinverno, C., Mazzaro, G., Viale, G., Martin-Padura, G., Garré, M., Parazzoli, D., Mattei, V., Cortellino, S., Bertalot, G., Di Fiore, P. P., & Scita, G. (2014). A RAB5/RAB4 recycling circuitry induces a proteolytic invasive program and promotes tumor dissemination. *The Journal of Cell Biology*, *206*(2), 307. https://doi.org/10.1083/JCB.201403127

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–2049. https://doi.org/10.1016/J.CELREP.2016.04.085

Fulmer, A. W., Benbasat, J. A., & Bloomfield, V. A. (1981). Ionic strength effects on macroion diffusion and excess light-scattering intensities of short DNA rods. *Biopolymers*, *20*(6), 1147–1159. https://doi.org/10.1002/BIP.1981.360200606

Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., Hsieh, T. H. S., Mirny, L., Zechner, C., & Hansen, A. S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, *376*(6592), 476–501. https://doi.org/10.1126/SCIENCE.ABN6583/SUPPL_FILE/SCIENCE.ABN6583_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Galganski, L., Urbanek, M. O., & Krzyzosiak, W. J. (2017). Nuclear speckles: Molecular organization, biological function and role in disease. *Nucleic Acids Research*, *45*(18), 10350–10368. https://doi.org/10.1093/nar/gkx759

Gerlitz, G. (2020). The Emerging Roles of Heterochromatin in Cell Migration. *Frontiers in Cell and Developmental Biology*, *8*. https://doi.org/10.3389/FCELL.2020.00394

Gerson, D. F., & Burton, A. C. (1977). The relation of cycling of intracellular pH to mitosis in the acellular slime mould Physarum polycephalum. *Journal of Cellular Physiology*, *91*(2), 297–303. https://doi.org/10.1002/JCP.1040910214

Gillies, R. J., & Deamer, D. W. (1979). Intracellular pH changes during the cell cycle in Tetrahymena. *Journal of Cellular Physiology*, *100*(1), 23–31. https://doi.org/10.1002/JCP.1041000103

Girelli, G., Custodio, J., Kallas, T., Agostini, F., Wernersson, E., Spanjaard, B., Mota, A., Kolbeinsdottir, S., Gelali, E., Crosetto, N., & Bienko, M. (2020). GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nature Biotechnology 2020 38:10*, *38*(10), 1184–1193. https://doi.org/10.1038/s41587-020-0519-y

González, L. O., Eiro, N., Fraile, M., Beridze, N., Escaf, A. R., Escaf, S., Fernández-Gómez, J. M., & Vizoso, F. J. (2022). Prostate Cancer Tumor Stroma: Responsibility in Tumor Biology, Diagnosis and Treatment. *Cancers*, *14*(18). https://doi.org/10.3390/CANCERS14184412

Grandi, F. C., Modi, H., Kampman, L., & Corces, M. R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nature Protocols 2022 17:6*, *17*(6), 1518–1552. https://doi.org/10.1038/s41596-022-00692-9

Gray, W. T., Govers, S. K., Xiang, Y., Parry, B. R., Campos, M., Kim, S., & Jacobs-Wagner, C. (2019). Nucleoid Size Scaling and Intracellular Organization of Translation across Bacteria. *Cell*, *177*(6), 1632-1648.e20. https://doi.org/10.1016/J.CELL.2019.05.017

Guo, X. W., & Cole, R. D. (1989). Chromatin aggregation changes substantially as pH varies within the physiological range. *Journal of Biological Chemistry*,

264(20), 11653–11657. https://doi.org/10.1016/S0021-9258(18)80114-4

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T., & Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910. https://doi.org/10.1016/J.CELL.2015.07.038

Hadizadeh Yazdi, N., Guet, C. C., Johnson, R. C., & Marko, J. F. (2012). Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions. *Molecular Microbiology*, *86*(6), 1318–1333. https://doi.org/10.1111/MMI.12071

Hahne, F., & Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. *Methods in Molecular Biology*, *1418*, 335–351. https://doi.org/10.1007/978-1-4939-3578-9_16/FIGURES/11

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

Hauer, M. H., & Gasser, S. M. (2017). Chromatin and nucleosome dynamics in DNA damage and repair. *Genes & Development*, *31*(22), 2204. https://doi.org/10.1101/GAD.307702.117

He, B., Zhang, C., Zhang, X., Fan, Y., Zeng, H., Liu, J., Meng, H., Bai, D., Peng, J., Zhang, Q., Tao, W., & Yi, C. (2021). Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nature Communications 2021 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-24425-w

Herzog, V. A., Lempradl, A., Trupke, J., Okulski, H., Altmutter, C., Ruge, F., Boidol, B., Kubicek, S., Schmauss, G., Aumayr, K., Ruf, M., Pospisilik, A., Dimond, A., Senergin, H. B., Vargas, M. L., Simon, J. A., & Ringrose, L. (2014). A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nature Genetics 2014 46:9*, *46*(9), 973–981. https://doi.org/10.1038/ng.3058

Hsieh, C. L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C. J., Lee, G. S. M., Chen, S., Balk, S. P., Liu, X. S., Brown, M., & Kantoff, P. W. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7319–7324. https://doi.org/10.1073/PNAS.1324151111/SUPPL_FILE/PNAS.201324151SI.PDF

Hu, C. D., Choo, R., & Huang, J. (2015). Neuroendocrine differentiation in prostate cancer: A mechanism of radioresistance and treatment failure. *Frontiers in Oncology*, *5*(APR), 90. https://doi.org/10.3389/FONC.2015.00090/XML/NLM

Huang, S., Gulzar, Z. G., Salari, K., Lapointe, J., Brooks, J. D., & Pollack, J. R. (2012). Recurrent deletion of CHD1 in prostate cancer with relevance to cell invasiveness. *Oncogene*, *31*(37), 4164. https://doi.org/10.1038/ONC.2011.590

Iannelli, F., Zanardi, F., Villa, S., Barzaghi, L., Cancila, V., Beznuskenko, G., Chiara, G. Della, Pagani, M., Malinverno, C., Bhattacharya, D., Pisati, F., Yu, W., & Galimberti, V. (2021). *Tissue fluidification promotes a cGAS/STING-mediated cytosolic DNA response in invasive breast cancer*. https://doi.org/10.21203/RS.3.RS-845173/V1

Jerkovic´, I., & Cavalli, G. (2021). Understanding 3D genome organization by multidisciplinary methods. *Nature Reviews Molecular Cell Biology 2021 22:8*,

22(8), 511–528. https://doi.org/10.1038/s41580-021-00362-w

Johnstone, S. E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., Chen, J. H., Zou, L. S., Drier, Y., Hecht, V., Shoresh, N., Selig, M. K., Lareau, C. A., Iyer, S., Nguyen, S. C., Joyce, E. F., Hacohen, N., Irizarry, R. A., Zhang, B., … Bernstein, B. E. (2020). Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell*, *182*(6), 1474-1489.e23. https://doi.org/10.1016/J.CELL.2020.07.030

Kalashnikova, A. A., Porter-Goff, M. E., Muthurajan, U. M., Luger, K., & Hansen, J. C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. *Journal of The Royal Society Interface*, *10*(82). https://doi.org/10.1098/RSIF.2012.1022

Kandi, V., & Vadakedath, S. (2015). Effect of DNA Methylation in Various Diseases and the Probable Protective Role of Nutrition: A Mini-Review. *Cureus*, *7*(8). https://doi.org/10.7759/CUREUS.309

Kang, S. mi, Yoon, M. H., & Park, B. J. (2018). Laminopathies; Mutations on single gene and various human genetic diseases. *BMB Reports*, *51*(7), 327. https://doi.org/10.5483/BMBREP.2018.51.7.113

Kantidze, O. L., & Razin, S. V. (2020). Weak interactions in higher-order chromatin organization. *Nucleic Acids Research*, *48*(9), 4614–4626. https://doi.org/10.1093/NAR/GKAA261

Karkampouna, S., De Filippo, M. R., Ng, C. K. Y., Klima, I., Zoni, E., Spahn, M., Stein, F., Haberkant, P., Thalmann, G. N., & de Julio, M. K. (2020). Stroma Transcriptomic and Proteomic Profile of Prostate Cancer Metastasis Xenograft Models Reveals Prognostic Value of Stroma Signatures. *Cancers*, *12*(12), 1–29. https://doi.org/10.3390/CANCERS12123786

Ke, X. S., Qu, Y., Rostad, K., Li, W. C., Lin, B., Halvorsen, O. J., Haukaas, S. A., Jonassen, I., Petersen, K., Goldfinger, N., Rotter, V., Akslen, L. A., Oyan, A. M., & Kalland, K. H. (2009). Genome-Wide Profiling of Histone H3 Lysine 4 and Lysine 27 Trimethylation Reveals an Epigenetic Signature in Prostate Carcinogenesis. *PLOS ONE*, *4*(3), e4687. https://doi.org/10.1371/JOURNAL.PONE.0004687

Kellenberger, E., Ryter, A., & Séchaud, J. (1958). Electron Microscope Study of DNA-Containing Plasms II. Vegetative and Mature Phage DNA as Compared with Normal Bacterial Nucleoids in Different Physiological States. *The Journal of Biophysical and Biochemical Cytology*, *4*(6), 671–678. https://doi.org/10.1083/JCB.4.6.671

Kelley, J. B., Datta, S., Snow, C. J., Chatterjee, M., Ni, L., Spencer, A., Yang, C.-S., Cubeñas-Potts, C., Matunis, M. J., & Paschal, B. M. (2011). The Defective Nuclear Lamina in Hutchinson-Gilford Progeria Syndrome Disrupts the Nucleocytoplasmic Ran Gradient and Inhibits Nuclear Localization of Ubc9. *Molecular and Cellular Biology*, *31*(16), 3378–3395. https://doi.org/10.1128/MCB.05087-11/ASSET/F930BC57-6506-4432-BB54-2517C2959A8F/ASSETS/GRAPHIC/ZMB9991091680014.JPEG

Kempfer, R., & Pombo, A. (2019). Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics 2019 21:4*, *21*(4), 207–226. https://doi.org/10.1038/s41576-019-0195-2

Khan, M. I., Hamid, A., Adhami, V. M., Lall, R. K., & Mukhtar, H. (2015). Role of Epithelial Mesenchymal Transition in Prostate Tumorigenesis. *Current Pharmaceutical Design*, *21*(10), 1240. https://doi.org/10.2174/1381612821666141211120326

Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology 2008 26:12*, *26*(12), 1351–1359. https://doi.org/10.1038/nbt.1508

Kleckner, N., Fisher, J. K., Stouf, M., White, M. A., Bates, D., & Witz, G. (2014). The bacterial nucleoid: nature, dynamics and sister segregation. *Current Opinion in Microbiology*, *22*, 127–137. https://doi.org/10.1016/J.MIB.2014.10.001

Korenchuk, S., Lehr, J. E., McLean, L., Lee, Y. G., Whitney, S., Vessella, R., Lin, D. L., & Pienta, K. J. (2001). VCaP, a cell-based model system of human prostate cancer. *In Vivo (Athens, Greece)*, *15*(2), 163–168. https://europepmc.org/article/med/11317522

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. H. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, *78*(3), 554-565.e7. https://doi.org/10.1016/J.MOLCEL.2020.03.003

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Maayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90–W97. https://doi.org/10.1093/NAR/GKW377

Kulis, M., & Esteller, M. (2010). DNA Methylation and Cancer. *Advances in Genetics*, *70*(C), 27–56. https://doi.org/10.1016/B978-0-12-380866-0.60002-2

Kumar, S., Gonzalez, E. A., Rameshwar, P., & Etchegaray, J. P. (2020). Non-Coding RNAs as Mediators of Epigenetic Changes in Malignancies. *Cancers 2020, Vol. 12, Page 3657*, *12*(12), 3657. https://doi.org/10.3390/CANCERS12123657

Labbé, D. P., & Brown, M. (2018). Transcriptional Regulation in Prostate Cancer. *Cold Spring Harbor Perspectives in Medicine*, *8*(11), a030437. https://doi.org/10.1101/CSHPERSPECT.A030437

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. https://doi.org/10.1016/J.CELL.2018.01.029

Lammerding, J., Fong, L. G., Ji, J. Y., Reue, K., Stewart, C. L., Young, S. G., & Lee, R. T. (2006). Lamins a and C but not lamin B1 regulate nuclear mechanics. *Journal of Biological Chemistry*, *281*(35), 25768–25780. https://doi.org/10.1074/jbc.M513511200

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature 2017 547:7662*, *547*(7662), 236–240. https://doi.org/10.1038/nature22822

Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, *25*(14), 1841–1842. https://doi.org/10.1093/BIOINFORMATICS/BTP328

Leung, J. K., & Sadar, M. D. (2017). Non-Genomic Actions of the Androgen Receptor in Prostate Cancer. *Frontiers in Endocrinology*, *0*, 2. https://doi.org/10.3389/FENDO.2017.00002

Levesque, C., & Nelson, P. S. (2018). Cellular Constituents of the Prostate Stroma: Key Contributors to Prostate Cancer Progression and Therapy Resistance. *Cold Spring Harbor Perspectives in Medicine*, *8*(8). https://doi.org/10.1101/CSHPERSPECT.A030510

Li, D., Yang, Y., Li, Y., Zhu, X., & Li, Z. (2021). Epigenetic regulation of gene expression in response to environmental exposures: From bench to model. *Science of The Total Environment*, *776*, 145998. https://doi.org/10.1016/J.SCITOTENV.2021.145998

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, H., & Wren, J. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics (Oxford, England)*, *30*(20), 2843–2851. https://doi.org/10.1093/BIOINFORMATICS/BTU356

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/SCIENCE.1181369/SUPPL_FILE/LIEBERMAN-AIDEN.SOM.PDF

Liu, Y., Fu, L., Kaufmann, K., Chen, D., & Chen, M. (2019). A practical guide for DNase-seq data analysis: from data management to common applications. *Briefings in Bioinformatics*, *20*(5), 1865–1877. https://doi.org/10.1093/BIB/BBY057

Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., Ceri, S., Oricchio, E., & Ciriello, G. (2021). Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nature Communications 2021 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-22666-3

Liyanage, V. R. B., Jarmasz, J. S., Murugeshan, N., Del Bigio, M. R., Rastegar, M., & Davie, J. R. (2014). DNA Modifications: Function and Applications in Normal and Disease States. *Biology*, *3*, 670–723. https://doi.org/10.3390/biology3040670

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Lund, E., Oldenburg, A. R., & Collas, P. (2014). Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research*, *42*(11). https://doi.org/10.1093/NAR/GKU324

Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M. F., Villar-Garea, A., Ballestar, E., Esteller, M., & Zaina, S. (2004). DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. *Journal of Biological Chemistry*, *279*(28), 29147–29154. https://doi.org/10.1074/jbc.M403618200

Machyna, M., Heyn, P., & Neugebauer, K. M. (2013). Cajal bodies: where form meets function. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 17–34. https://doi.org/10.1002/WRNA.1139

Maeshima, K., Tamura, S., & Shimamoto, Y. (2018). Chromatin as a nuclear spring. *Biophysics and Physicobiology*, *15*(0), 189. https://doi.org/10.2142/BIOPHYSICO.15.0_189

Marasca, F., Marullo, F., & Lanzuolo, C. (2016). Determination of Polycomb Group of Protein Compartmentalization Through Chromatin Fractionation Procedure. *Methods in Molecular Biology (Clifton, N.J.)*, *1480*, 167–180. https://doi.org/10.1007/978-1-4939-6380-5_15

Mario Cioce, A. I. L. (2005). Cajal Bodies: A Long History of Discovery. *Annual Review of Cell and Developmental Biology*, *21*, 105–131. https://www.proquest.com/docview/217954842?pq-origsite=gscholar&fromopenview=true

Marshal, A. D., Bailey, C. G., Champ, K., Vellozzi, M., O'Young, P., Metierre, C., Feng, Y., Thoeng, A., Richards, A. M., Schmitz, U., Biro, M., Jayasinghe, R., Ding, L., Anderson, L., Mardis, E. R., & Rasko, J. E. J. (2017). CTCF genetic alterations in endometrial carcinoma are pro-tumorigenic. *Oncogene 2017 36:29*, *36*(29), 4100–4110. https://doi.org/10.1038/onc.2017.25

Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T., Parenti, A. R., Daidone, M. G., Bicciato, S., & Piccolo, S. (2010). A MicroRNA targeting dicer for metastasis control. *Cell*, *141*(7), 1195–1207. https://doi.org/10.1016/J.CELL.2010.05.017

Marullo, F., Cesarini, E., Antonelli, L., Gregoretti, F., Oliva, G., & Lanzuolo, C. (2016). Nucleoplasmic Lamin A/C and Polycomb group of proteins: An evolutionarily conserved interplay. *Nucleus*, *7*(2), 103. https://doi.org/10.1080/19491034.2016.1157675

Mason, D. J., & Powelson, D. M. (1956). *NUCLEAR DIVISION AS OBSERVED IN LIVE BACTERIA BY A NEW TECHNIQUE*. https://journals.asm.org/journal/jb

Matullo, G., Naccarati, A., & Pardini, B. (2016). MicroRNA expression profiling in bladder cancer: the challenge of next-generation sequencing in tissues and biofluids. *International Journal of Cancer*, *138*(10), 2334–2345. https://doi.org/10.1002/IJC.29895

McCarthy, B. J., & Duerksen, J. D. (1971). Distribution of deoxyribonucleic acid sequences in fractionated chromatin. *Biochemistry*, *10*(8), 1471–1478. https://doi.org/10.1021/bi00784a031

McCord, R. P., Nazario-Toole, A., Zhang, H., Chines, P. S., Zhan, Y., Erdos, M. R., Collins, F. S., Dekker, J., & Cao, K. (2013). Correlated alterations in genome organization, histone methylation, and DNA–lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Research*, *23*(2), 260–269. https://doi.org/10.1101/GR.138032.112

McHugh, C. A., Chen, C. K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature 2015 521:7551*, *521*(7551), 232–236. https://doi.org/10.1038/nature14443

Mehta, S., & Zhang, J. (2022). Liquid–liquid phase separation drives cellular function and dysfunction in cancer. *Nature Reviews Cancer 2022 22:4*, *22*(4), 239–252. https://doi.org/10.1038/s41568-022-00444-7

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., Leproust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne,

C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics 2015 47:6*, *47*(6), 598–606. https://doi.org/10.1038/ng.3286

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/NATURE06008

Mo, F., Lin, D., Takhar, M., Ramnarine, V. R., Dong, X., Bell, R. H., Volik, S. V., Wang, K., Xue, H., Wang, Y., Haegert, A., Anderson, S., Brahmbhatt, S., Erho, N., Wang, X., Gout, P. W., Morris, J., Karnes, R. J., Den, R. B., … Collins, C. C. (2018). Stromal Gene Expression is Predictive for Metastatic Primary Prostate Cancer. *European Urology*, *73*(4), 524–532. https://doi.org/10.1016/J.EURURO.2017.02.038

Mobley, A. S. (2019). Induced Pluripotent Stem Cells. *Neural Stem Cells and Adult Neurogenesis*, 67–94. https://doi.org/10.1016/B978-0-12-811014-0.00004-4

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., Tamayo, P., Spiegelman, B., Lander, E. S., Hirschhorn, J. N., … Groop, L. C. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics 2003 34:3*, *34*(3), 267–273. https://doi.org/10.1038/ng1180

Musacchio, A. (2022). On the role of phase separation in the biogenesis of membraneless compartments. *The EMBO Journal*, *41*(5), e109952. https://doi.org/10.15252/EMBJ.2021109952

Nakato, R., & Sakata, T. (2021). Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods*, *187*, 44–53. https://doi.org/10.1016/J.YMETH.2020.03.005

Natoli, G., & Andrau, J.-C. (2014). *Transcriptional and epigenetic mechanisms governing T-cell differentiation and activation View project Development of computational tools for high-throughput techniques View project*. https://doi.org/10.1146/annurev-genet-110711-155459

Nava, M. M., Miroshnikova, Y. A., Biggs, L. C., Whitefield, D. B., Metge, F., Boucas, J., Vihinen, H., Jokitalo, E., Li, X., García Arcos, J. M., Hoffmann, B., Merkel, R., Niessen, C. M., Dahl, K. N., & Wickström, S. A. (2020). Heterochromatin-Driven Nuclear Softening Protects the Genome against Mechanical Stress-Induced Damage. *Cell*, *181*(4), 800-817.e22. https://doi.org/10.1016/J.CELL.2020.03.052

Neidle, S., & Sanderson, M. (2022). Principles of protein–DNA recognition. *Principles of Nucleic Acid Structure*, 347–396. https://doi.org/10.1016/B978-0-12-819677-9.00001-9

Nicetto, D., & Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Current Opinion in Genetics & Development*, *55*, 1–10. https://doi.org/10.1016/J.GDE.2019.04.013

Nomoto, A., Nishinami, S., & Shiraki, K. (2021). Solubility Parameters of Amino Acids on Liquid–Liquid Phase Separation and Aggregation of Proteins. *Frontiers in Cell and Developmental Biology*, *9*, 1527. https://doi.org/10.3389/FCELL.2021.691052/BIBTEX

Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., Joti, Y., Tomita, M., Hibino, K., Kanemaki, M. T., Wendt, K. S., Okada, Y., Nagai, T., & Maeshima, K. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Molecular Cell*, *67*(2), 282-293.e7. https://doi.org/10.1016/J.MOLCEL.2017.06.018

O'Neill, L. P., & Turner, B. M. (2003). Immunoprecipitation of native chromatin: NChIP. *Methods*, *31*(1), 76–82. https://doi.org/10.1016/S1046-2023(03)00090-2

Ogg, S. C., & Lamond, A. I. (2002). Cajal bodies and coilin—moving towards function. *Journal of Cell Biology*, *159*(1), 17–21. https://doi.org/10.1083/JCB.200206111

Oldenburg, A., Briand, N., Sørensen, A. L., Cahyani, I., Shah, A., Moskaug, J. Ø., & Collas, P. (2017). A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic MIR335 locus. *Journal of Cell Biology*, *216*(9), 2731–2743. https://doi.org/10.1083/JCB.201701043

Osmanagic-Myers, S., & Foisner, R. (2019). The structural and gene expression hypotheses in laminopathic diseases - Not so different after all. *Molecular Biology of the Cell*, *30*(15), 1786–1790. https://doi.org/10.1091/MBC.E18-10-0672/ASSET/IMAGES/LARGE/MBC-30-1786-G001.JPEG

Panni, S., Lovering, R. C., Porras, P., & Orchard, S. (2020). Non-coding RNA regulatory networks. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1863*(6), 194417. https://doi.org/10.1016/J.BBAGRM.2019.194417

Park, J. W., & Han, J. W. (2019). Targeting epigenetics for cancer therapy. *Archives of Pharmacal Research*, *42*(2), 159. https://doi.org/10.1007/S12272-019-01126-Z

Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics 2009 10:10*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

Paschal, B. M., & Kelley, J. B. (2013). Nuclear Lamina. *Encyclopedia of Biological Chemistry: Second Edition*, 310–313. https://doi.org/10.1016/B978-0-12-378630-2.00477-1

Perovanovic, J., Dell'Orso, S., Gnochi, V. F., Jaiswal, J. K., Sartorelli, V., Vigouroux, C., Mamchaoui, K., Mouly, V., Bonne, G., & Hoffman, E. P. (2016). Laminopathies disrupt epigenomic developmental programs and cell fate. *Science Translational Medicine*, *8*(335). https://doi.org/10.1126/SCITRANSLMED.AAD4991/SUPPL_FILE/8-335RA58_SM.PDF

Perry, M., & Chalkley, R. (1981a). *The Effect of Histone Hyperacetylation on the Nuclease Sensitivity and the Solubility of Chromatin*. 256*(7), 33133318. https://doi.org/10.1016/S0021-9258(19)69608-0

Perry, M., & Chalkley, R. (1981b). The Effect of Histone Hyperacetylation on the Nuclease Sensitivity and the Solubility of Chromatin*. *Journal of Biological Chemistry*, *256*(7), 3313–3318. https://doi.org/10.1016/S0021-9258(19)69608-0

Pfeifer, G. P., Xiong, W., Hahn, M. A., & Jin, S. G. (2014). The role of 5-hydroxymethylcytosine in human cancer. *Cell and Tissue Research*, *356*(3), 631–641. https://doi.org/10.1007/S00441-014-1896-7/FIGURES/3

Pistore, C., Giannoni, E., Colangelo, T., Rizzo, F., Magnani, E., Muccillo, L., Giurato, G., Mancini, M., Rizzo, S., Riccardi, M., Sahnane, N., Del Vescovo, V., Kishore, K., Mandruzzato, M., MacChi, F., Pelizzola, M., Denti, M. A., Furlan, D., Weisz, A., … Bonapace, I. M. (2017). DNA methylation variations are required for

epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells. *Oncogene 2017 36:40*, *36*(40), 5551–5566. https://doi.org/10.1038/onc.2017.159

Pomerantz, M. M., Li, F., Takeda, D. Y., Lenci, R., Chonkar, A., Chabot, M., Cejas, P., Vazquez, F., Cook, J., Shivdasani, R. A., Bowden, M., Lis, R., Hahn, W. C., Kantoff, P. W., Brown, M., Loda, M., Long, H. W., & Freedman, M. L. (2015). The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature Genetics 2015 47:11*, *47*(11), 1346–1351. https://doi.org/10.1038/ng.3419

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033

Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, *174*(3), 744-757.e24. https://doi.org/10.1016/j.cell.2018.05.024

Radaeva, M., Ton, A. T., Hsing, M., Ban, F., & Cherkasov, A. (2021). Drugging the 'undruggable'. Therapeutic targeting of protein–DNA interactions with the use of computer-aided drug discovery methods. *Drug Discovery Today*, *26*(11), 2660–2679. https://doi.org/10.1016/J.DRUDIS.2021.07.018

Rahnamoun, H., Lee, J., Sun, Z., Lu, H., Ramsey, K. M., Komives, E. A., & Lauberth, S. M. (2018). RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nature Structural & Molecular Biology 2018 25:8*, *25*(8), 687–697. https://doi.org/10.1038/s41594-018-0102-0

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/NAR/GKW257

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., … Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305-320.e24. https://doi.org/10.1016/J.CELL.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/J.CELL.2014.11.021

Rauluseviciute, I., Drabløs, F., & Rye, M. B. (2019). DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis. *Clinical Epigenetics*, *11*(1). https://doi.org/10.1186/S13148-019-0795-X

Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M., & Cosma, M. P. (2015). Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell*, *160*(6), 1145–1158. https://doi.org/10.1016/J.CELL.2015.01.054

Rippe, K. (2022). Liquid–Liquid Phase Separation in Chromatin. *Cold Spring Harbor Perspectives in Biology*, *14*(2), a040683.

https://doi.org/10.1101/CSHPERSPECT.A040683
Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature 2015 518:7539*, *518*(7539), 317–330. https://doi.org/10.1038/nature14248
Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry*, *79*, 233. https://doi.org/10.1146/ANNUREV-BIOCHEM-060408-091030
Rosa, S., & Shaw, P. (2013). Insights into Chromatin Structure and Dynamics in Plants. *Biology 2013, Vol. 2, Pages 1378-1410*, *2*(4), 1378–1410. https://doi.org/10.3390/BIOLOGY2041378
Rustøen Braadland, P., & Urbanucci, A. (2019). Chromatin reprogramming as an adaptation mechanism in advanced prostate cancer. *Endocrine-Related Cancer*, *26*(4), R211–R235. https://doi.org/10.1530/ERC-18-0579
Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, *361*(6400). https://doi.org/10.1126/SCIENCE.AAR3958/SUPPL_FILE/AAR3958_SABARI_SM_TABLE_S3.XLSX
Sabari, B. R., Dall'Agnese, A., & Young, R. A. (2020). Biomolecular Condensates in the Nucleus. *Trends in Biochemical Sciences*, *45*(11), 961. https://doi.org/10.1016/J.TIBS.2020.06.007
Sadaie, M., Salama, R., Carroll, T., Tomimatsu, K., Chandra, T., Young, A. R. J., Narita, M., Pérez-Mancera, P. A., Bennett, D. C., Chong, H., Kimura, H., & Narita, M. (2013). Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes & Development*, *27*(16), 1800–1808. https://doi.org/10.1101/GAD.217281.113
Salvarani, N., Crasto, S., Miragoli, M., Bertero, A., Paulis, M., Kunderfranco, P., Serio, S., Forni, A., Lucarelli, C., Dal Ferro, M., Larcher, V., Sinagra, G., Vezzoni, P., Murry, C. E., Faggian, G., Condorelli, G., & Di Pasquale, E. (2019). The K219T-Lamin mutation induces conduction defects through epigenetic inhibition of SCN5A in human cardiac laminopathy. *Nature Communications 2019 10:1*, *10*(1), 1–16. https://doi.org/10.1038/s41467-019-09929-w
Sanborn, A. L., Rao, S. S. P., Huang, S. C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., & Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6456–E6465. https://doi.org/10.1073/PNAS.1518552112/SUPPL_FILE/PNAS.1518552112.ST01.XLSX
Sartorelli, V., & Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nature Structural & Molecular Biology 2020 27:6*, *27*(6), 521–528. https://doi.org/10.1038/s41594-020-0446-0
Schmidl, C., Rendeiro, A. F., Sheffield, N. C., & Bock, C. (2015). ChIPmentation:

fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature Methods*, *12*(10), 963–965. https://doi.org/10.1038/NMETH.3542

Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H. C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., … Church, D. M. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Research*, *27*(5), 849–864. https://doi.org/10.1101/GR.213611.116

Schöfer, C., & Weipoltshammer, K. (2018). Nucleolus and chromatin. *Histochemistry and Cell Biology 2018 150:3*, *150*(3), 209–225. https://doi.org/10.1007/S00418-018-1696-3

Seal, R. L., Chen, L.-L., Griffiths-Jones, S., Lowe, T. M., Mathews, M. B., O'Reilly, D., Pierce, A. J., Stadler, P. F., Ulitsky, I., Wolin, S. L., & Bruford, E. A. (2020). A guide to naming human non-coding RNA genes. *The EMBO Journal*, *39*(6), e103777. https://doi.org/10.15252/EMBJ.2019103777

Sebestyén, E., Marullo, F., Lucini, F., Petrini, C., Bianchi, A., Valsoni, S., Olivieri, I., Antonelli, L., Gregoretti, F., Oliva, G., Ferrari, F., & Lanzuolo, C. (2020). SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nature Communications 2020 11:1*, *11*(1), 1–16. https://doi.org/10.1038/s41467-020-20048-9

Senga, S. S., & Grose, R. P. (2021). Hallmarks of cancer—the new testament. *Open Biology*, *11*(1). https://doi.org/10.1098/RSOB.200358

Sepehri, Z., Beacon, T. H., Osman, F. D. S., Jahan, S., & Davie, J. R. (2019). DNA methylation and chromatin modifications. *Nutritional Epigenomics*, 13–36. https://doi.org/10.1016/B978-0-12-816843-1.00002-3

Serrano, L., Vazquez, B. N., & Tischfield, J. (2013). Chromatin structure, pluripotency and differentiation: *Https://Doi.Org/10.1177/1535370213480718*, *238*(3), 259–270. https://doi.org/10.1177/1535370213480718

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, *148*(3), 458–472. https://doi.org/10.1016/J.CELL.2012.01.010

Shevchenko, A. I., Zakharova, I. S., & Zakian, S. M. (2013). The Evolutionary Pathway of X Chromosome Inactivation in Mammals. *Acta Naturae*, *5*(2), 40. https://doi.org/10.32607/20758251-2013-5-2-40-53

Shi, M., You, K., Chen, T., Hou, C., Liang, Z., Liu, M., Wang, J., Wei, T., Qin, J., Chen, Y., Zhang, M. Q., & Li, T. (2021). Quantifying the phase separation property of chromatin-associated proteins under physiological conditions using an anti-1,6-hexanediol index. *Genome Biology*, *22*(1), 1–26. https://doi.org/10.1186/S13059-021-02456-2/FIGURES/7

Shi, X., Liu, C., Liu, B., Chen, J., Wu, X., & Gong, W. (2018). JQ1: a novel potential therapeutic target. *Die Pharmazie*, *73*(9), 491–493. https://doi.org/10.1691/PH.2018.8480

Shimia, T., Kittisopikul, M., Tran, J., Goldman, A. E., Adam, S. A., Zheng, Y., Jaqaman, K., & Goldman, R. D. (2015). Structural organization of nuclear lamins A, C, B1, and B2 revealed by superresolution microscopy. *Molecular Biology of the Cell*, *26*(22), 4075–4086. https://doi.org/10.1091/MBC.E15-07-0461/ASSET/IMAGES/LARGE/MBC-26-4075-G005.JPEG

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., Wingreen, N. S., Haataja, M., & Brangwynne, C. P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell*, *175*(6), 1481-1491.e13. https://doi.org/10.1016/J.CELL.2018.10.057

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(1), 7–33. https://doi.org/10.3322/CAAC.21708

Singh, K., Cassano, M., Planet, E., Sebastia, S., Min Jang, S., Sohi, G., Faralli, H., Choi, J., Youn, H. D., Dilworth, F. J., & Trono, D. (2015). A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. *Genes & Development*, *29*(5), 513. https://doi.org/10.1101/GAD.254532.114

Sood, A. J., Viner, C., & Hoffman, M. M. (2019). DNamod: The DNA modification database. *Journal of Cheminformatics*, *11*(1), 1–10. https://doi.org/10.1186/S13321-019-0349-4/FIGURES/3

Sparmann, A., & Van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nature Reviews Cancer 2006 6:11*, *6*(11), 846–856. https://doi.org/10.1038/nrc1991

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2020). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 2020 22:2*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stephens, A. D. (2020). Chromatin Rigidity Provides Mechanical and Genome Protection. *Mutation Research*, *821*, 111712. https://doi.org/10.1016/J.MRFMMM.2020.111712

Stephens, A. D., Banigan, E. J., & Marko, J. F. (2019). Chromatin's physical properties shape the nucleus and its functions. *Current Opinion in Cell Biology*, *58*, 76–84. https://doi.org/10.1016/J.CEB.2019.02.006

Storebjerg, T. M., Strand, S. H., Høyer, S., Lynnerup, A. S., Borre, M., Ørntoft, T. F., & Sørensen, K. D. (2018). Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. *Clinical Epigenetics*, *10*(1). https://doi.org/10.1186/S13148-018-0540-X

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/PNAS.0506580102/SUPPL_FILE/06580FIG7.JPG

Sullivan, T., Escalante-Alcalde, D., Bhatt, H., Anver, M., Bhat, N., Nagashima, K., Stewart, C. L., & Burke, B. (1999). Loss of a-Type Lamin Expression Compromises Nuclear Envelope Integrity Leading to Muscular Dystrophy. *Journal of Cell Biology*, *147*(5), 913–920. https://doi.org/10.1083/JCB.147.5.913

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics 2008 9:6*, *9*(6), 465–476. https://doi.org/10.1038/nrg2341

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., Nora, E. P., Bruneau, B. G., Bantignies, F., & Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics 2020 52:11*, *52*(11), 1151–1157. https://doi.org/10.1038/s41588-020-00716-8

Terada, L. S. (2019). Shc and the mechanotransduction of cellular anchorage and metastasis. *Small GTPases*, *10*(1), 64–71. https://doi.org/10.1080/21541248.2016.1273172

Tirado-Magallanes, R., Rebbani, K., Lim, R., Pradhan, S., & Benoukraf, T. (2017). Whole genome DNA methylation: beyond genes silencing. *Oncotarget*, *8*(3), 5629–5637. https://doi.org/10.18632/ONCOTARGET.13562

Tyekucheva, S., Bowden, M., Bango, C., Giunchi, F., Huang, Y., Zhou, C., Bondi, A., Lis, R., Van Hemelrijck, M., Andrén, O., Andersson, S. O., Watson, R. W., Pennington, S., Finn, S. P., Martin, N. E., Stampfer, M. J., Parmigiani, G., Penney, K. L., Fiorentino, M., … Loda, M. (2017). Stromal and epithelial transcriptional map of initiation progression and metastatic potential of human prostate cancer. *Nature Communications 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/s41467-017-00460-4

Umbarger, M. A., Toro, E., Wright, M. A., Porreca, G. J., Baù, D., Hong, S. H., Fero, M. J., Zhu, L. J., Marti-Renom, M. A., McAdams, H. H., Shapiro, L., Dekker, J., & Church, G. M. (2011). The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation. *Molecular Cell*, *44*(2), 252–264. https://doi.org/10.1016/J.MOLCEL.2011.09.010

Urbanucci, A., Barfeld, S. J., Kytölä, V., Itkonen, H. M., Coleman, I. M., Vodák, D., Sjöblom, L., Sheng, X., Tolonen, T., Minner, S., Burdelski, C., Kivinummi, K. K., Kohvakka, A., Kregel, S., Takhar, M., Alshalalfa, M., Davicioni, E., Erho, N., Lloyd, P., … Mills, I. G. (2017). Androgen Receptor Deregulation Drives Bromodomain-Mediated Chromatin Alterations in Prostate Cancer. *Cell Reports*, *19*(10), 2045–2059. https://doi.org/10.1016/J.CELREP.2017.05.049

Van Steensel, B., Delrow, J., & Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. *Nature Genetics 2001 27:3*, *27*(3), 304–308. https://doi.org/10.1038/85871

Van Steensel, B., & Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotechnology 2000 18:4*, *18*(4), 424–428. https://doi.org/10.1038/74487

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., & Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports*, *10*(8), 1297–1309. https://doi.org/10.1016/J.CELREP.2015.02.004

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature 2009 461:7261*, *461*(7261), 199–205. https://doi.org/10.1038/nature08451

Wang, G., Achim, C. L., Hamilton, R. L., Wiley, C. A., & Soontornniyomkij, V. (1999). Tyramide Signal Amplification Method in Multiple-Label Immunofluorescence Confocal Microscopy. *Methods*, *18*(4), 459–464. https://doi.org/10.1006/METH.1999.0813

Wang, G., Zhao, D., Spring, D. J., & Depinho, R. A. (2018). Genetics and biology of prostate cancer. *Genes & Development*, *32*(17–18), 1105–1140. https://doi.org/10.1101/GAD.315739.118

Wang, Q., Li, W., Zhang, Y., Yuan, X., Xu, K., Yu, J., Chen, Z., Beroukhim, R., Wang, H., Lupien, M., Wu, T., Regan, M. M., Meyer, C. A., Carroll, J. S., Manrai, A. K., Jänne, O. A., Balk, S. P., Mehra, R., Han, B., … Brown, M. (2009). Androgen Receptor Regulates a Distinct Transcription Program in Androgen-Independent Prostate Cancer. *Cell*, *138*(2), 245–256. https://doi.org/10.1016/J.CELL.2009.04.056

West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R. E., Hirose, T., Bond, C., Fox, A., & Nakagawa, S. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *Journal of Cell Biology*, *214*(7), 817–830. https://doi.org/10.1083/JCB.201601071

Worman, H. J., & Bonne, G. (2007). "Laminopathies": A wide spectrum of human diseases. *Experimental Cell Research*, *313*(10), 2121–2133. https://doi.org/10.1016/J.YEXCR.2007.03.028

Wu, F., Olson, B. G., & Yao, J. (2016). DamID-seq: Genome-wide Mapping of Protein-DNA Interactions by High Throughput Sequencing of Adenine-methylated DNA Fragments. *Journal of Visualized Experiments : JoVE*, *2016*(107), 53620. https://doi.org/10.3791/53620

Wu, F., & Yao, J. (2013). Spatial compartmentalization at the nuclear periphery characterized by genome-wide mapping. *BMC Genomics*, *14*(1), 1–14. https://doi.org/10.1186/1471-2164-14-591/FIGURES/7

Wu, R., Singh, P. B., & Gilbert, D. M. (2006). Uncoupling global and fine-tuning replication timing determinants for mouse pericentric heterochromatin. *Journal of Cell Biology*, *174*(2), 185–194. https://doi.org/10.1083/JCB.200601113

Xie, L., Dong, P., Qi, Y., Hsieh, T. H. S., English, B. P., Jung, S. K., Chen, X., De Marzio, M., Casellas, R., Chang, H. Y., Zhang, B., Tjian, R., & Liu, Z. (2022). BRD2 compartmentalizes the accessible genome. *Nature Genetics 2022 54:4*, *54*(4), 481–491. https://doi.org/10.1038/s41588-022-01044-9

Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current Protocols*, *1*(3), e90. https://doi.org/10.1002/CPZ1.90

Xiong, K., & Ma, J. (2019). Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. *Nature Communications 2019 10:1*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-12954-4

Xu, K., Wu, Z. J., Groner, A. C., He, H. H., Cai, C., Lis, R. T., Wu, X., Stack, E. C., Loda, M., Liu, T., Xu, H., Cato, L., Thornton, J. E., Gregory, R. I., Morrissey, C., Vessella, R. L., Montironi, R., Magi-Galluzzi, C., Kantoff, P. W., … Brown, M. (2012). EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. *Science*, *338*(6113), 1465–1469. https://doi.org/10.1126/SCIENCE.1227604/SUPPL_FILE/XU.SM.PDF

Yamada, Y., & Beltran, H. (2021). Clinical and Biological Features of Neuroendocrine Prostate Cancer. *Current Oncology Reports*, *23*(2), 1–10. https://doi.org/10.1007/S11912-020-01003-9/FIGURES/1

Yuan, B. F. (2020). Assessment of DNA Epigenetic Modifications. *Chemical Research in Toxicology*, *33*(3), 695–708. https://doi.org/10.1021/ACS.CHEMRESTOX.9B00372/SUPPL_FILE/TX9B00372_SI_001.PDF

Zeng, C., Liu, S., Lu, S., Yu, X., Lai, J., Wu, Y., Chen, S., Wang, L., Yu, Z., Luo, G., & Li, Y. (2018). The c-Myc-regulated lncRNA NEAT1 and paraspeckles modulate imatinib-induced apoptosis in CML cells. *Molecular Cancer*, *17*(1). https://doi.org/10.1186/S12943-018-0884-Z

Zhang, H., Sun, L., Wang, K., Wu, D., Trappio, M., Witting, C., & Cao, K. (2016). Loss of H3K9me3 Correlates with ATM Activation and Histone H2AX Phosphorylation Deficiencies in Hutchinson-Gilford Progeria Syndrome. *PLOS ONE*, *11*(12), e0167454. https://doi.org/10.1371/JOURNAL.PONE.0167454

Zhang, L., Lu, Q., & Chang, C. (2020). Epigenetics in Health and Disease. *Advances in Experimental Medicine and Biology*, *1253*, 3–55. https://doi.org/10.1007/978-981-15-3449-2_1/FIGURES/6

Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., & Belmont, A. S. (2021). TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Research*, *31*(2), 251–264. https://doi.org/10.1101/GR.266239.120/-/DC1

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), 1–9. https://doi.org/10.1186/GB-2008-9-9-R137/FIGURES/3

Zhang, Y., Sun, Z., Jia, J., Du, T., Zhang, N., Tang, Y., Fang, Y., & Fang, D. (2021). Overview of Histone Modification. *Advances in Experimental Medicine and Biology*, *1283*, 1–16. https://doi.org/10.1007/978-981-15-8104-5_1/FIGURES/2

Zhao, S., Allis, C. D., & Wang, G. G. (2021). The language of chromatin modification in human cancers. *Nature Reviews Cancer 2021 21:7*, *21*(7), 413–430. https://doi.org/10.1038/s41568-021-00357-x

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., & Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics 2006 38:11*, *38*(11), 1341–1347. https://doi.org/10.1038/ng1891

Zheng, X., Hu, J., Yue, S., Kristiani, L., Kim, M., Sauria, M., Taylor, J., Kim, Y., & Zheng, Y. (2018). Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Molecular Cell*, *71*(5), 802-815.e7. https://doi.org/10.1016/J.MOLCEL.2018.05.017

Zhou, W. Y., Cai, Z. R., Liu, J., Wang, D. S., Ju, H. Q., & Xu, R. H. (2020). Circular RNA: metabolism, functions and interactions with proteins. *Molecular Cancer 2020 19:1*, *19*(1), 1–19. https://doi.org/10.1186/S12943-020-01286-3

Zhu, Q., Stöger, R., & Alberio, R. (2018). A lexicon of DNA modifications: Their roles in embryo development and the germline. *Frontiers in Cell and Developmental Biology*, *6*(MAR), 24. https://doi.org/10.3389/FCELL.2018.00024/XML/NLM

Zwerger, M., Roschitzki-Voser, H., Zbinden, R., Denais, C., Herrmann, H., Lammerding, J., Grütter, M. G., & Medalia, O. (2015). Altering lamina assembly reveals lamina-dependent and -independent functions for A-type lamins. *Journal of Cell Science*, *128*(19), 3607–3620. https://doi.org/10.1242/JCS.171843/259936/AM/ALTERING-LAMINA-ASSEMBLY-IDENTIFIES-LAMINA

7 Acknowledgments

In these years of PhD I met, discuss, collaborate with many people, express properly my gratitude for each of them would produce at least as many pages as the one that are written in this thesis… but there is a limit in the number of words that are allowed in this manuscript!

As usual, I am starting with professional acknowledgments, but I want to remark that in many cases it has been hard to define where the line that distinguish professional and personal relationship is.

The first person that I want to thank is my supervisor Dr. Francesco Ferrari. You have been a wonderful guide, making me understand what does it means to produce high quality work, how to be professional and competitive in an international environment. I think that I am really changed as person after your knowledge.

I must also thank my Galway supervisor Prof. Kevin Sullivan. Every time I had opportunity to talk with you, I always found a very kind person and I always received very wise advices.

Going through the list of people that has teach me the most during this period definitely I have to thank Dr. Chiara Lanzuolo. I always appreciated your ambition in projects and pragmatism that often is extraneous to me. I thinks that it was with you that I started to change my way to read and extract information from papers.

Talking about teachers, I should mention Elisa as well. We talk as colleagues and we were joking about the fact that someone define you as “my inspiration”. Inside of me that was not a joke. I learned and I am still learning a lot from you and your rational approach that put inside all projects.

If I learnt many things from my mentors, I learnt many things from my colleagues and collaborators as well. Federica, Giovanni, Giulia, Judith, Raquel and Valentina interacting with you and sharing problems and complaints is an important part of the reason why I want to go to work every day.

Together with collaborators and colleagues I want to thank all the former members of the groups, the people from IFOM and INGM, and all the people I interacted with during these years, I get something from each of you. In particular I want to thanks Ilario for its numerous lessons on how to use the server and Koustav for contributing with suggestions and idea to my works.

In these growing group of people that I am thanking, I want to express my gratitude to: the “Physics of the next door” (Dr. Marco Cosentino-Lagormarsino Group). They always remember me how much fun we could be have conversation about science (at least as well as playing table tennis/padel/climbing/foosballe/etc). In particular among them I want to thank Ludovico, that since more time than the other have to deal with my crazy ideas and that went with me/organized the travel for my PhD VIVA.
PhD unfortunately is not only science, but also bureaucracy and technical problems. I very glad that Mio, Marina and Sara helped me, caring about everything about burocracy. As well I want to thanks all the members of the IT team for being very proactive in solving informatic issues I had during this period.
I want to thank Dr. Giorgio Scita and all the lab members for the fruitful collaboration we had.
Among all the people that I met thanks the PhD one in particular come with me through all the path and more: Clarisse “Bro” Reyes. For sure everything would be much much sadder without your unbreakable positivity and happiness.
In conclusion I want to thank all the people that are not directly connected with PhD but has always been of support. There are no words for Claudia and my family that I could say, you are more than I deserve.
To my friends of Viterbo and Milan, I am glad that I had the opportunity to share with most the most amazing experience of these more than four years of PhD.