How DNA methylation helps specify and maintain centromere identity and function

Catalina Salinas Luypaert

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie
UMR 144 – Biologie Cellulaire et Cancer
Laboratoire Mécanismes Moléculaires de la Dynamique des Chromosomes

How DNA methylation helps specify and maintain centromere identity and function

Le rôle de la méthylation de l'ADN dans la spécification et le maintien de l'identité et la fonction des centromères

Soutenue par
Catalina SALINAS LUYPAERT
Le 6 octobre 2023

École doctorale n° 515
Complexité du vivant

Spécialité
Biologie Cellulaire et Développement

Composition du jury :

Geneviève ALMOUZNI, PhD DRCE, CNRS – Institut Curie	*Présidente*
William C. EARNSHAW, PhD Professor – University of Edinburgh	*Rapporteur*
Sylvia ERHARDT, PhD Professor – Karlsruhe Institute of Technology	*Rapportrice*
Pierre-Antoine DEFOSSEZ, PhD DR1, CNRS – Université Paris Diderot	*Examinateur*
Ines Anna DRINNENBERG, PhD DR2, CNRS – Institut Curie	*Examinatrice*
Daniele FACHINETTI, PhD DR2, CNRS – Institut Curie	*Directeur de thèse*

institutCurie | PSL

To my family / A mi familia
Camila, Rolando, Anna, Jessica, Rebeca, Sebastián, Yangzela, Leonor & Luciana.

To the girl that dreamed of becoming a scientist.
To the woman that still dreams of making a difference.

ACKNOWLEDGEMENTS

Gracias a la vida, que me ha dado tanto
Me ha dado la risa y me ha dado el llanto
Así yo distingo dicha de quebranto
Los dos materiales que forman mi canto
Y el canto de ustedes que es el mismo canto
Y el canto de todos que es mi propio canto

— Violeta Parra —

I would like to begin by stating how grateful I am (to the Universe? to Life?) to have had the chance of doing this thesis. I truly think the planets aligned for me to get here. It is a marvelous thing, to feel that you are where you are supposed to be, doing what you are supposed to be doing.

THANK YOU

To Daniele, the boss. In the tiny cell culture room of the 3rd floor of the Burg building, back in 2017, you asked me what my career plans were and if I had any intentions of doing a PhD. I was working in another team then, across from your relatively newly opened lab, but we found ourselves passing our cells side by side more than once. I ended up going back to Chile, but that conversation was a little seed that I took with me, and that finally blossomed in 2019. Thank you for asking me that question and for following through with your (not-so-subtle) hint a couple years later, taking me in as a somewhat "unconventional" PhD student. Thank you for trusting me with this beautiful project, and for your never-ending enthusiasm and drive, for this and all projects. I have learned and grown tremendously, both scientifically and personally, in these past four years working with you. Special thanks also for giving me the gift of the best lab mates I could ever have wished for. How lucky am I to share most of my days with this amazing group of people?

Marie, María, bella; thank you for being the voice of order amid the daily chaos and for everything that you do to make sure the lab runs as smoothly as possible. We have been through a lot together –two pregnancies and what not! You were my first Fachinetti liaison and friend, back when we were just 3rd floor acquaintances. In many ways, you are at the core of this thesis, for without you I might have lost contact with the team when I went back to Chile. Ric, caro cuore; I saw you arrive, and I saw you leave, it feels like a lifetime ago. Your contribution to this work is tremendous; scientifically, yes, but most importantly on a personal level. *Fank you* for all the joy and laughter you brought me every day; you truly are a never-fading ray of sunshine. Nobody will ever sing along with me like you do. Appi, my darling tesoro; you arrived when I was at my absolute lowest and injected new life into me (and I dare say into all of us). Thank you for being so uniquely you, for organizing people's lives, for pushing me out of my comfort zone, for always having a smile, for staying positive, and for making me a #youngresearcher. Because "*Life is not a sacrifice*" should be everybody's motto. Leo, hermosito; life put us back together in the same lab to speak Esperanto and invent new words every other day. Thank you for always inviting me out on Friday nights, despite me declining often (specially at the beginning). I hope that you find your way, and that in the process you don't lose your passion for science –I am still rooting for you to find nickase activity in CENP-B! Andre(a), the last man standing (besides Marie) from the team I arrived into; thank you for teaching me the COBRA, for the countless tricks and tips, and for always being the person to go to with tough questions. Therese: thank you for the many talks we had about science and life. To my fellow PhD(C)s Camellia and Marco(lino); thank for your companionship and for and all the good times (and the bad) that we have had on this shared wild ride. Always remember that less is more, and that you too shall overcome. To the PhDs that came before me, Sebastian and Solène; thanks for paving the way. Marine, the

latest addition to the team; thank you for helping me finish the last experiments as I was writing this manuscript. Together we made it through when I thought it was impossible to get everything done in time. Finally, thanks to all the other past and new members of the team with whom I shared this adventure, even if (very) briefly.

To my reviewers and examinators: Professors Sylvia Erdhardt and William C. Earnshaw, thank you for your time and effort spent improving this manuscript. I am beyond grateful to Dr. Earnshaw for your commitment to carrying on this task in unthinkable circumstances. To my other examinators: Dr. Geneviève Almouzni, Dr. Ines Drinnenberg and Dr. Pierre-Antoine Defossez. Thank you all for not hesitating a second in saying yes to being part of my jury and for the rich discussion we had.

My gratitude goes to all the people at Institut Curie that make the clock tick. All the members of the laveries at Trouillet and Burg. The UMR144 admins, especially Céline Jaquet. Edwige Delage, for all the dry-ice smuggling and international shipment management, even after we moved buildings. The PICT-IBiSA imaging platform, in particular Anne-Sophie Mace and Olivier Leroy. The Cytometry platform. To all the members of the UMR144, and our "new" UMR3664, especially the teams we had shared lab meetings with during these years. To all those whose names I am forgetting.

Thank you to my collaborators, very particularly Guillaume Velasco for all the DNAme expertise. I also would like to thank my financing body, the French Agence Nationale de la Recherche (ANR) under project ANR-19-CE12-0022 – MEliCENDRe. To Claire Francastel and Florence Larminat for writing along with Daniele this grant which has supported me for the whole duration of this PhD.

While I was writing the introduction to this manuscript, I felt increasingly connected to the many generations of scientists that throughout the centuries have built the Science that we know and love today. It is a bit odd, but I would like to express my gratitude to all those past scientists –and very particularly to all the Women in science–, for without you, I would not be here today.

People say that doing a PhD is a transformative experience, and I could not agree more. Doing it ~11.638 Km away from your home country and during a pandemic is a whole other thing. Gracias a toda mi familia por ser una inagotable fuente de amor. Gracias por apoyarme siempre, sin condiciones. Al igual que mi padre vine a Europa a hacer el doctorado (no en barco eso sí). Gracias por heredarme tu oratoria y pedagogía. Tú desde la filosofía, yo desde la biología; son más nuestras similitudes que diferencias. Mami; las vueltas de la vida te llevaron a migrar a Chile y a mí me trajeron de vuelta. Llegué hasta aquí por y para ti. Gracias por ser mi todo. Hermanas, hermano –los adoro–; llevo en mí un poquito de cada uno de ustedes. Gracias por ser y dejarme ser, por estar siempre ahí cuando los necesito. Gracias totales por regalarme la dicha de ser tía. A mis sobrinas; Yangze, que desde el 2005 me enseñas lo que es el amor incondicional; y Leo y Lu, que vinieron al mundo mientras yo engendraba esta tesis y me hacen cada día aprender a querer a la distancia. Un grand merci aussi à toute ma famille en Belgique, qui ont toujours fait en sorte que je ne me sente pas loin de chez moi.

Gracias a Magdalena, Pablo, Lucía y Ernesto: mi familia latinoamericana en Paris. Son la mejor exportación de la República Oriental del Uruguay. Gracias a H & B por su apoyo emocional.

Finally, and actually first and foremost, I would like to thank you Camila. I will try my best to be succinct, as I am sure you know by now that no words will ever be enough to express it all. Without you, I would quite literally not be writing this today. Thank you for being my main supporter. Throughout this rite of passage, you have kept me well fed, both literally and metaphorically. Whenever I felt I was drowning, you pulled me afloat. Whenever I felt I wouldn't make it, you reminded me that I was already making it. Whenever I doubted myself, you believed in me for the two of us. I don't exaggerate when saying I would not have kept my sanity without your daily support, your infinitely patient ear, your kind, wise, encouraging words, and your warm hugs. Big thanks also for proofreading this manuscript; because yes, you can do it all. This is your accomplishment, as much as it is mine. I could not have dreamed to find a better partner to go through life (and a PhD) with. Gracias a la vida, que me ha dado a ti.

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

RESULTS

DISCUSSION AND PERSPECTIVES

RÉSUMÉ

Le maintien de l'intégrité du génome et la transmission correcte de l'information génétique à chaque division cellulaire sont des étapes clés pour la survie de tous les organismes. Chez les mammifères, la méthylation de l'ADN (ADNme) et le maintien de la fidélité de la ségrégation des chromosomes sont des déterminants majeurs de la stabilité du génome. L'ADNme est une marque épigénétique essentielle qui joue un rôle central dans la régulation du génome. En plus d'inhiber les transpositions, c'est le principal mécanisme de répression de l'expression des éléments répétés. La ségrégation des chromosomes est médiée par le centromère, un domaine chromosomique spécialisé où le kinétochore s'assemble pour attacher les chromosomes au fuseau mitotique. Le bon fonctionnement des centromères est essentiel à la préservation du caryotype dans les cellules en division. Les centromères sont entourés par des domaines péricentriques présentant des caractéristiques de l'hétérochromatine et sont assemblés sur de l'ADN satellite auquel se lie la protéine centromérique B (CENP-B). L'ADN centromérique et péricentromérique est fortement méthylé, à l'exception notable des domaines enrichis en CENP-A, la variante de l'histone H3 qui définit épigénétiquement la position des centromères et dont la régulation est déterminante pour leur fonction.

Des altérations pathologiques des patrons de méthylation de l'ADN et de la fonction du centromère ont été associées à l'instabilité chromosomique dans le cancer et dans le syndrome ICF (Immunodéficience, instabilité Centromérique et dysmorphie Faciale), une maladie génétique rare. Le lien fonctionnel entre l'hypométhylation de l'ADN centromérique et le dysfonctionnement du centromère reste cependant à être établi. Notre hypothèse est que l'ADNme est un déterminant majeur de l'identité du centromère, et que l'hypométhylation des centromères conduit à un assemblage défectueux du réseau protéique centromérique, ce qui impacte leur fonction.

En utilisant des outils cellulaires pour l'invalidation génique des ADN méthyltransférases DNMT1 et DNMT3B, j'ai démontré les mécanismes de régulation de l'ADNme au niveau des centromères humains et l'impact de la réduction de l'ADNme sur la physiologie des centromères. Pour aborder directement le lien entre l'ADNme et la fonction du centromère, j'ai généré des outils cellulaires pour induire une déméthylation ciblée de l'ADN centromérique. Le niveau d'ADNme centromérique peut être réduit jusqu'à 85% sans affecter la méthylation des péricentromères. Cette diminution de l'ADNme centromérique entraîne une instabilité du génome, avec formation de micronoyaux à l'issue d'une ségrégation incorrecte des chromosomes, et une létalité cellulaire. CENP-A et CENP-B, deux des protéines centrales de la chromatine centromérique, augmentent au niveau des centromères hypométhylés, ce qui explique en partie l'instabilité génomique observée dans ces cellules. La déméthylation de l'ADN centromérique n'est pas réversible et provoque à long terme un processus d'adaptation cellulaire. Certaines cellules parviennent à continuer à proliférer avec des niveaux d'ADNme centromérique plus faibles et des niveaux de CENP-A et de CENP-B stablement plus élevés au niveau des centromères. Enfin, j'ai identifié que l'ADNme peut servir de frontière pour délimiter la formation épigénétique des centromères médiée par CENP-A en préservant la position du centromère et en empêchant l'élargissement du domaine occupé par CENP-A. Dans l'ensemble, j'ai démontré qu'il existe un lien direct entre l'ADNme centromérique et l'identité et la fonction du centromère, tout en fournissant les premières preuves d'une relation fonctionnelle entre ces deux facteurs déterminants pour la stabilité du génome. Ce travail établit les bases de notre compréhension de la façon dont la perte d'ADNme centromérique peut être à l'origine de l'instabilité du génome observée dans le cancer et le syndrome ICF.

Mots Clés: centromère – méthylation de l'ADN – épigénétique – CENP-A – CENP-B – stabilité du génome

SUMMARY

In all living organisms, maintaining genome integrity and ensuring the proper transmission of genetic information in each cell division are key steps for a healthy development and survival. In mammals, both DNA methylation (DNAme) and the maintenance of chromosome segregation fidelity are known to be major determinants of genome stability. DNAme is an essential epigenetic mark that plays a central role in genome regulation as it is the major mechanism responsible for silencing the expression of repetitive elements, and it inhibits transpositions and recombination. Chromosome segregation is mediated by the centromere, a specialized chromosomal domain where the kinetochore assembles in mitosis to attach the chromosomes to the spindle microtubules. The proper function of the centromeres is essential for the preservation of a correct karyotype in all dividing cells. In most eukaryotes, the centromeres are often embedded into pericentric domains with hallmarks of heterochromatin, and they are assembled on large arrays of tandemly repeated satellite DNA bound by CENtromeric Protein B (CENP-B). Both the centromeric and the pericentromeric DNA are highly methylated, with the notable exception of the domains where CENP-A is enriched. CENP-A is the histone H3 variant that defines epigenetically the position of the centromeres, and its tight regulation is known to be key for centromere function.

Pathological alterations of DNA methylation patterns and centromere function have been associated to chromosomal instability in cancer and in the Immunodeficiency, Centromeric region instability and Facial anomalies (ICF) syndrome, a rare genetic disease. The functional link between centromeric DNA hypomethylation and centromere dysfunction remains however to be formally established. We hypothesize that DNAme is a major determinant of centromere identity, and that pathological hypomethylation of the centromeres leads to a deficient assembly of the centromeric protein network, which then impacts centromere function.

Using cellular tools to remove the DNA methyl transferases DNMT1 and DNMT3B I demonstrate the mechanisms regulating the DNAme at human centromeres and the impact of DNAme reduction on centromere physiology. To directly address the link between DNAme and centromere function, I have generated cellular tools to induce a targeted centromeric DNA demethylation and have demonstrated that they are centromere-specific and tunable. The centromeric DNAme level can be reduced up to 85% without affecting the methylation of the neighboring pericentromeres. This decrease of centromeric DNAme causes genome instability, with formation of micronuclei as a readout of chromosome mis-segregation, and cell lethality. CENP-A and CENP-B, two of the core proteins of the centromeric chromatin, increase at the centromeres upon hypomethylation, partially accounting for the genome instability observed in these cells. The demethylation of the centromeric DNA is not reversible, and in the long-term prompts a process of cellular adaptation at the cell population level: despite initially causing lethality, some cells manage to continue to proliferate with lower centromeric DNAme levels at and with stably higher CENP-A and CENP-B levels at the centromeres. Finally, I reveal that DNAme can act as a boundary for the epigenetic, CENP-A-mediated formation of the centromeres by preserving centromere position and preventing the enlargement of the domain occupied by CENP-A. Overall, I have demonstrated that there is a direct link between the centromeric DNAme and centromere identity and function, providing the first evidence of a functional relationship between these two determinant factors for genome stability. This work sets the foundation of our understanding on how loss of centromeric DNAme can be at the root of the genome instability observed in cancer and the ICF syndrome.

KEY WORDS: centromere – DNA methylation – epigenetics – CENP-A – CENP-B - genome stability

INTRODUCTION

I. A brief history of the Cell, and what that lies within its Nucleus

I.1. Origins and life of the Cell

I.1.1. What is a cell and where do they come from

The Cell is the common denominator and the basic functional and structural unit of all living organisms on Earth. In 1665 Robert Hooke published "Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon" (Hooke, 1665) where he drew with fine detail his observations made under one of the first microscopes ever built (**Figure 1**). Besides drawing ants and fleas with much detail, he used the term "cell" to describe the regular shapes or pores in cork, which to him resembled the honeycomb cells. Antonie van Leeuwenhoek continued improving the microscope and was the first to describe living bacteria and protozoa (which he called *animalcules*), not too long after Hooke's observations in 1675. van Leeuwenhoek, considered today as the forefather of microbiology, is also credited to have made, unknowingly, some of the first descriptions of the cell nucleus.

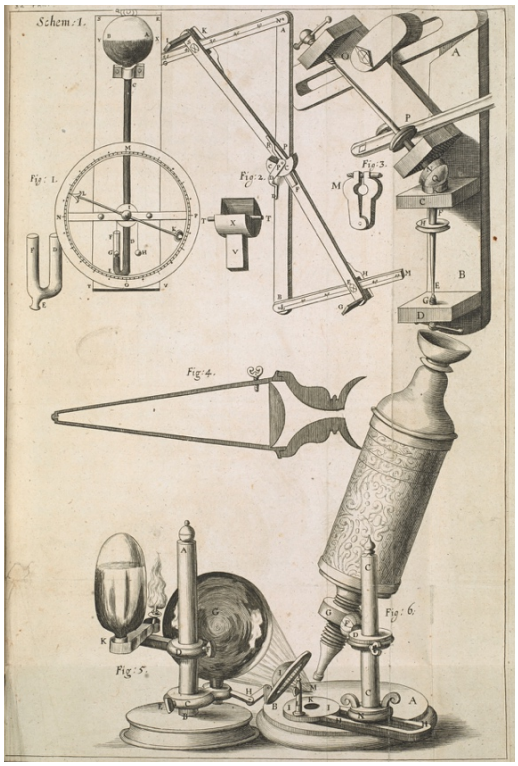

Figure 1. Scheme of Robert Hooke's microscope.
From *Micrographia, or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon*, Robert Hooke (1665). Accessible through The British Library online, of public domain.

For a very long time cells were believed to generate spontaneously from non-living matter after some kind of crystallization. Only at the beginning of the nineteenth century, would come the realization that *omnis cellula e cellula*; "when a cell arises, there a cell must have previously existed, just as an animal can spring only from an animal, a plant only from a plant". The cell theory is generally attributed to Rudolf Virchow (Virchow, 1859), but the expression seems to actually have been originated twenty-four years prior by François-Vincent Raspail (1825) (Wright and Poulsom, 2012).

I.1.2. The life of a cell (the cell cycle)

The life of a somatic cell is divided into four well-defined cell cycle stages: mitosis (M), gap 1 (G1), DNA Synthesis (S) and gap (G2). In certain circumstances, such as nutrient deprivation, they can exit the cell cycle and arrest in what is known as G0. Cells spend most of their lifespans in interphase, which collectively refers to G1, S and G2, and then divide during mitosis to give rise to two daughter cells (*omnis cellula e cellula).* After arising, each daughter cell is in the G1 phase of the cell cycle, where it grows and prepares for the duplication of its genetic material, a process that takes place during the subsequent DNA Synthesis (S) phase. Once the genome of the cell has been duplicated it enters the G2 phase where it grows some more and prepares for the last phase of its life; going through mitosis to give rise to two new daughter cells (O'Connor and Adams, 2010). A detailed recount on the processes that occur during mitosis will be provided in section II.1.

I.2. The nucleus and the *nuclein*

I.2.1. The nucleus

The term nucleus (from Latin, meaning "kernel" or "core") was coined by Robert Brown, a Scottish botanist studying the fertilization of orchids in 1831. The choice of the name nucleus reflected mostly its observed central position on the cells, but said nothing about its function, or what it was made of. Its mere existence in a cell —along with other membrane-bound organelles— would later become a keystone feature to classify said cell, or multicellular organism, as a eukaryote (from Greek, *eu* meaning "true" and *karyon*, meaning "nut" or "kernel"), and differentiate it from cells belonging to the other domains of life; archaea and bacteria, which are prokaryotes (from Greek, *pro* meaning "before" the nucleus).

We now know that the nucleus is the organelle that contains most of the eukaryotic cell's genetic material in the form of deoxyribonucleic acid (DNA) molecules. It is separated from the cytosol by the nuclear envelope, a complex structure composed of two nuclear membranes joined at nuclear pore complexes and the underlying nuclear lamina, a fibrous network that provides structural and functional support (Gary M. Cooper, 2000).

Cells are hypothesized to have arisen on earth 3.5 to 3.8 billion years ago as little more than a lipid membrane containing organic compounds and ribonucleic acid (RNA)-like molecules that carried both information and catalytic functions. The more stable DNA evolved later, to take over the information storage function and proteins, with a wider variety of structures, took over the catalytic processes. The formation of the nucleus, about 2.7 billion years ago, marked a major advance in the evolution of cells and branched a whole new domain of life; the eukaryotes. The appearance of membrane-bound organelles, including the nucleus, is believed to be the result of one prokaryotic cell that engulfed another, generating a profitable endosymbiotic relationship where the engulfed cell began functioning as an organelle inside the now larger early-eukaryotic cell (Martin et al., 2015). The fact that both chloroplasts and mitochondria retain their own genomes is thought of as proof of this hypothesis. About 1.7 billion years ago individual eukaryotic cells started aggregating and living in groups, or colonies, in what is considered the first transition towards the formation of multicellular organisms. Over a billion years later the process of cellular differentiation would begin, allowing animals to arise. The human body, for example, is composed of more than 200 different cell types, each with very specific functions (Geoffrey M. Cooper, 2000; O'Connor and Adams, 2010).

I.2.2. The *nuclein* (spoiler: it's the DNA)

When we think about the history of the DNA, two (or three) names immediately spring to mind: James Watson and Francis Crick –the name of Rosalind Franklin, for much too long forgotten, should be intertwined in there too. However, many crucial discoveries had taken place all along the 18th century, long before the double-helical structure of DNA was solved (Franklin and Gosling, 1953; Watson and Crick, 1953). The DNA's recorded history actually begun not too long after Brown's naming of the nucleus, when Swiss physician Friedrich Miescher obtained the first crude DNA purification in 1869 (Dahm, 2005; for a historical review see: Hall and Sankaran, 2021).

Not wanting to dedicate to patient care, Miescher went to Germany to work in the laboratory of biochemist Felix Hoppe-Seyler, where he focused on studying the proteins extracted from leucocytes obtained from pus-filled bandages. At that time, proteins were the most promising candidate molecules for understanding the cell's function, and he determined that the cytosol was indeed mainly composed of proteins and lipids. During a step his extraction protocol, he came across a viscous substance that precipitated when acid was added and was resuspended in the presence of alkali. The substance did not behave as proteins do; his analytical tests determined they were rich in phosphorous and, unlike proteins, lacked sulphur and were not digested by treatment with pepsin. He could only attribute it came from the cell's nucleus, and therefore called it *nuclein* (Miescher, 1871) (an integral English translation of Miescher's paper can be found in: Hall and Sankaran, 2021). After some initial doubts about the discovery of this completely new substance, Hoppe-Seyler reproduced Miescher's experiments by himself and had Pal Plósz, another of his students, repeat the protocol. Plósz

successfully isolated nuclein from the nucleated erythrocytes of birds and snakes, but not from bovine red blood cells, which much like human erythrocytes, lack nuclei. Their respective works were published back-to-back (Miescher, 1871; Plósz, 1871; Hoppe-Seyler, 1871). In his paper, Miescher stressed how the nucleus and the nuclein are separate entities from the protein-rich cytosol, and begun to ponder about the possible implications of this discovery. In an unpublished addendum, he even proposed that the nucleus should be defined by the presence of nuclein (implying, by its still unknown function), rather than by its structural or morphological characteristics.

A decade after Miescher's discovery, Albrecht Kossel -at that point, a research assistant to Hoppe-Seyler- also started working on the nuclein. He optimized the isolation, purification and analysis methods, which allowed him to reveal that it contained the five nitrogen bases adenine (A), cytosine (C), guanine (G), thymine (T) and uracil (U) (Kossel, 1879). A modern analysis of the product of these historical extraction protocols revealed that RNA was present in the nuclein, which explains the presence or uracil (Thess et al., 2021). Throughout this time, chemists and cytologists alike debated if the nuclein was real or just an experimental artifact, and both the concept and term were generally not well accepted. Kossel would eventually win the Nobel Prize in Medicine in 1910, "*in recognition of the contributions to our knowledge of cell chemistry made through his work on proteins, including the nucleic substances*". Contemporary to Kossel, Richard Altmann would further improve the nuclein purification protocol, completely removing all proteins by pepsin treatment and alkaline hydrolysis, and would rename his nuclein as nucleic acid (Altmann, 1889). His term would prevail.

During the first two decades of the twentieth century, Russian biochemist Phoebus Levene dedicated his work to trying to elucidate the structure of the deoxyribonucleic acid (DNA), then called "thymus nucleic acid" (Levene and Jacobs, 1912) and distinguish it from that of ribonucleic acid (RNA), then called "yeast nucleic acid" (Levene, 1919). It was not until Levene discovered that the sugar in "thymus nucleic acid" was deoxyribose that the name DNA actually made its debut (Levene and Tipson, 1935). Levene was also the first to propose that nucleic acids were composed of a series of four nucleotides, and that each nucleotide was in turn composed of just one of four nitrogen-containing bases, a sugar molecule, and a phosphate group. Along with these accurate statements regarding the chemical composition of the DNA, he also postulated the infamous "tetranucleotide hypothesis". His early measures suggested that the bases are found in equal ratios, and therefore he proposed that all DNA molecules were composed by repeating sequences of the four nucleotides. This was the paradigm of the 2D structure of DNA for several decades, and many believe that this oversimplification slowed down significantly the advances in the field. It certainly made scientists favour proteins as the molecules mediating heredity, and somewhat loose interest in studying the DNA. Nonetheless, Lavene's discoveries laid the groundwork of the field that we know today as molecular biology (Frixione and Ruiz-Zamarripa, 2019).

I.3. Genes, Genetics and the unexpected role of DNA

The roots of modern genetics can be traced back precisely to one man, and one date: Gregor Mendel and his work on the heredity of traits in the garden pea (Mendel, 1865). Mendel spoke of "cell elements" that would be stably inherited as the underlying determinants of visible characteristics in an organism; he was describing what we know now as genes. His work would remain mostly unknown for more than three decades until in 1900 three botanists (Hugo de Vries in the Netherlands, Carl Correns in Germany, and Erich von Tschermak in Austria) would independently rediscover Mendel's laws and bring them to the forefront of science (Gayon, 2016). The term "genetics" would be coined in 1906 by English biologist William Bateson to designate the science of heredity and variation. Bateson had been the main champion of Mendel's work for a few years, and he would play an instrumental role in the establishment and development of the new science of genetics, generalizing Mendel's findings to all sexually reproducing organisms and introducing the terms "allele", "homozygote" and "heterozygote" (Bateson, 1902). The terms "gene", "genotype" and "phenotype" would be coined later by Wilhelm Johannsen, a Danish botanist (Johannsen, 1909).

At this point in history, the concept of gene would refer to the mendelian factors of inheritance, but still nobody ventured to hypothesize about its nature or form. The general belief was that the genetic information had to be contained in proteins, since they are much more diverse and complex compared to the DNA and its simple four nucleotides. The inflexion point would come in 1944 with the Avery-MacLeod-McCarty experiment, which for the first time provided proof that DNA, and not proteins, is the substance that causes bacterial transformation (Avery et al., 1944). From there on, the DNA would become the object of more intense studies. One key finding emanated from the work of Erwin Chargaff, who, a couple of decades after Levene, analysed in detail the composition of nucleic acids from different species (human, ox, yeast and avian tubercles bacillus) and from different tissues (for human and ox: thymus, spleen, liver and sperm). He concluded that (i) the overall composition of the DNA (total number of A, T, C and G) varies from one species to another, but not within tissues of the same species and (ii) that regardless of the species, the number of Gs equals the number of Cs and the number of As equals the number of Ts; however, the number of GCs does not equal the number of ATs (Chargaff, 1950). These findings would put an end to the tetranucleotide theory era and would prove to be fundamental for Watson & Cricks' three-dimensional structural work, since it clearly hinted towards G-C and A-T interactions. Chargaff's conclusions, along with Rosalind Franklin's photograph 51 (Franklin and Gosling, 1953) (**Figure 2A**), led Watson and Crick to propose that the structure of the DNA corresponded to two antiparallel helical chains of nucleotides, coiled around the same axis, with opposing bases bound together by hydrogen bonds (Watson and Crick, 1953) (**Figure 2B**). In their model, A/T and G/C were physically considered most plausible pairings, matching Chargaff's measurements, and they went as far as proposing that "*if the sequence of bases of one chain is given, the sequence of the other is automatically determined*" (Watson and Crick, 1953). Their model would not take long to be proven right.

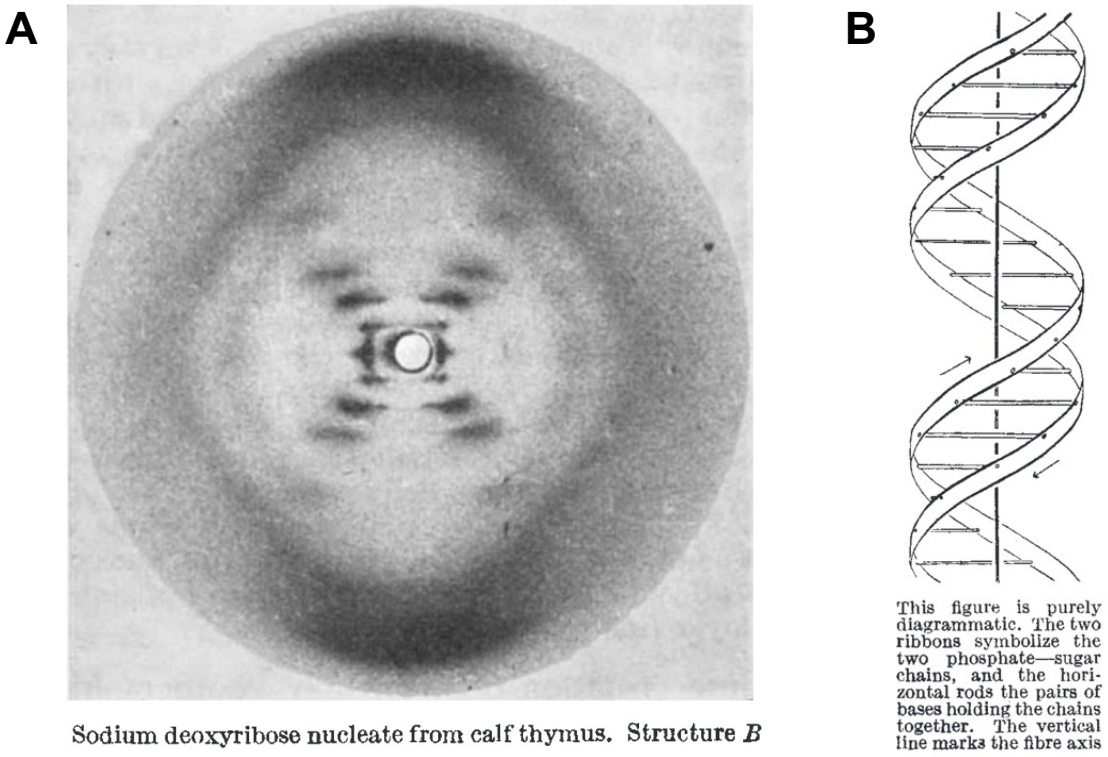

Figure 2. The structure of the DNA.
(**A**) Franklin & Gosling, 1953. Photograph 51, an X-ray diffraction of DNA at >75% humidity. From this image the authors concluded that the structure of the DNA "*is probably helical*" and "*the structural unit probably consists of two co-axial molecules*". (**B**) Watson & Crick, 1953. Proposed molecular structure of the DNA with "*two helical chains each coiled round the same axis*".

I.4. Chromosomes and chromatin

I.4.1. The Chromosomes

Walther Flemming was a German physician and is considered the pioneer of cytology, being one of the first to describe in detail the process of cell division (Flemming, 1879). He would publish in 1882 the book *Zellsubstanz, Kern und Zelltheilung* (Cell substance, Nucleus and Cell division), considered by many as the ground-breaking work on mitosis research (Flemming, 1882). In his book, Flemming would coin the terms mitosis and chromatin, to refer to "the stainable substance of the nucleus". The term "chromosomen" (in German, which translated to English became chromosome; from Greek *chroma*, "color" and *soma*, "body"-) would be introduced six years later by Flemming's colleague, Wilhelm von Waldeyer (von Waldeyer, 1888). Up until then, chromosomes were called *chromatinelemente*, chromatin elements, or just *Stäbchen*, rods.

Around the same time, Friedrich Miescher's nuclein was generating debate (see section I.2). Following-up his discovery on leucocytes, Miescher started working on the question of fertilization, isolating nuclein in spermatozoa. He realized that the nuclein was not found freely in the nucleus,

but rather in an "insoluble, salt-like compound" with proteins, which he called protamines (Miescher, 1874). The term protamines endured, and today it refers to proteins that replace up to 90% of histones in the final stages of spermiogenesis. Protamines bind the major groove of DNA, coiling and condensing it to a much higher degree than what can be achieved by wrapping the DNA around nucleosomes (see following section), effectively inactivating the whole genome (Balhorn, 2007). German botanist Eduard Zacharias combined Flemming's cytological observations with Miescher's analytical methods to investigate the chemical composition of the *chromatinelemente*. He was the first to report that the nuclein was an integral part of the chromosomes (Zacharias, 1881). Little over twenty years later, both Theodor Boveri and Walter Sutton connected Mendel's laws of inheritance to the chromosomes (Boveri, 1902; Sutton, 1902). In a follow up theoretical paper, Sutton summarized and discussed in detail the importance of his conclusions. He pointed out that both the Mendelian factors and the chromosomes exist in pairs and segregate on a one-to-one ratio, independently of all other pairs. He explicitly states that chromosomes "*may constitute the physical basis of the Mendelian law of heredity*" (Sutton, 1903). This theory remained controversial for several years, until one of its original detractors published the first decisive evidence to support it in a paper demonstrating that the eye colour in *Drosophila melanogaster* is determined by the inheritance of a gene carried on the X chromosome (Morgan, 1910). Thomas Morgan Hunt went from detractor to one of the main promoters of the mendelian theory of heredity (Morgan, 1913; Morgan et al., 1915), work that awarded him the Nobel Prize in 1933.

Structurally, chromosomes were believed for a long time to be formed as cables i.e., by several short molecules of DNA wound around each other. The development of pulsed field gradient gel electrophoresis, followed by Southern blotting with chromosomally assigned probes, allowed the first separation of intact *Saccharomyces cerevisiae* chromosomes (Schwartz and Cantor, 1984). This experiment allowed to conclude that each chromosome is formed by a single DNA molecule.

I.4.2. The Chromatin

The human genome contains twenty-two pairs of autosomes and two sex chromosomes; XX for females, XY for males. Each somatic cell in our body is diploid, inheriting one maternal and one paternal copy of each chromosome. It is calculated that in these 46 chromosomes there are about 6.32×10^9 nucleotide base pairs in total. Since each base pair is 0.34 nm (or 3.4 Å) long (Watson and Crick, 1953), this means that little over 2 meters of DNA are packed inside each cell's nucleus, which is on average only between 5 and 10 μm in diameter (Piovesan et al., 2019). Fitting this incredible length of DNA inside each cells' nucleus, while still allowing vital processes such as transcription and replication to occur, is achieved by a highly dynamic system of several layers of packaging, wrapping and coiling of the DNA (Annunziato, 2008) (**Figure 3**).

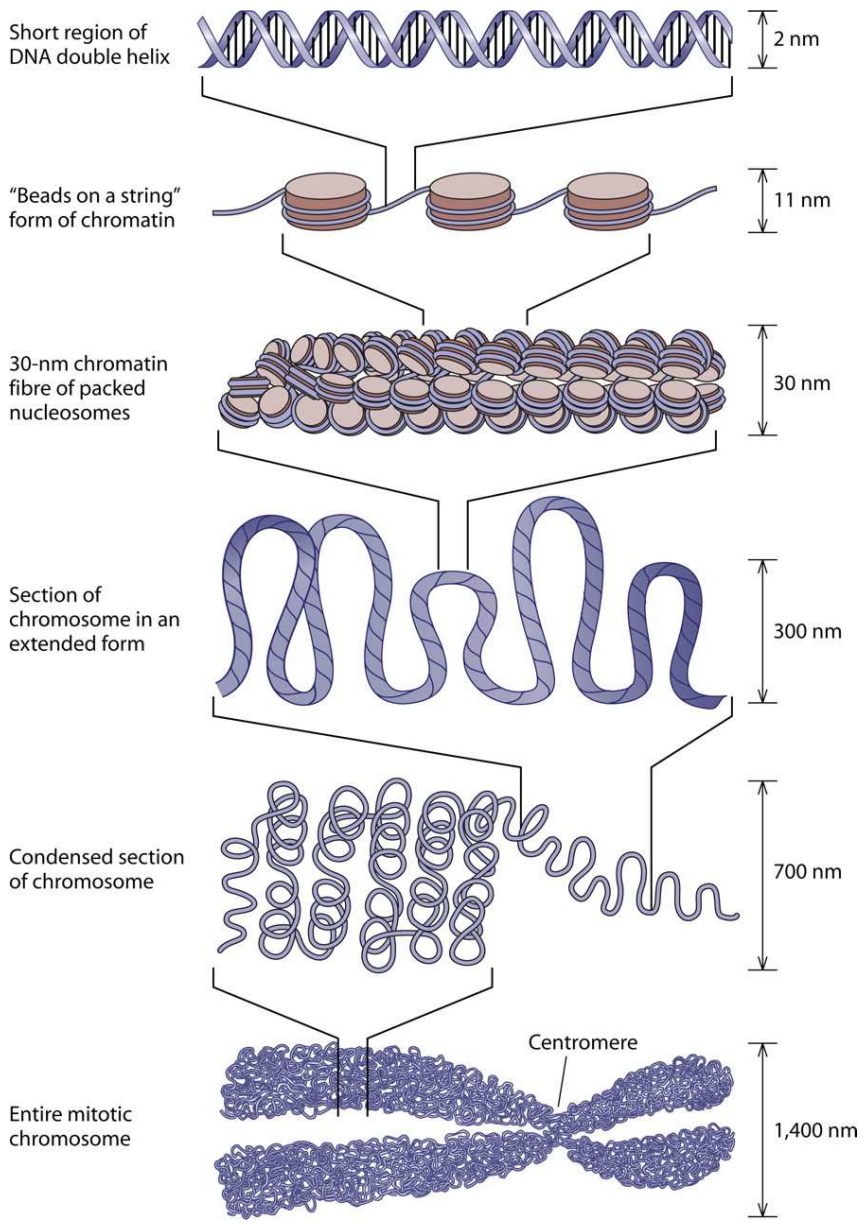

Figure 3. The structure of the chromatin.
(**A**) Jansen & Verstrepen, 2011 [*Figure 1*]. The DNA double helix (2 nm wide) is wrapped around a histone octamer to form nucleosomes which are connected by stretches of linker DNA and form an 11 nm thick "beads on a string" filament. This basic nucleosome structure is folded into a putative (see sub-section C) fiber-like structure of about 30 nm in diameter which is further looped and compacted into higher-order structures to culminate in the formation of condensed mitotic chromatids.

A) Histones

The first layer of higher-order packaging of chromosomal DNA is the wrapping of the negatively charged double-stranded helical molecule of DNA around histones, a family of small, positively charged proteins. The resulting DNA-protein complex is called chromatin. The discovery of histones is attributed to Albrecht Kossel, and was contemporary to his discovery of the nucleobases (see section I.2.2). While working on nuclear extracts of geese erythrocytes, Kossel found, besides the

acidic nuclein, a basic substance that was in composition like classical proteins yet somewhat apart from them. He proposed to name them histones (Kossel, 1884). Based on his observations, Kossel believed that histones were bound to the nuclein, similar to what Miescher had described about protamines a decade prior. It would be revealed later on that there is also a non-histone component to the nucleoproteins (Mirsky and Pollister, 1946). Histones were believed to be a homogenous type of protein until 1950, when, based on their differential alcohol solubilities, the existence of "main" (arginine-rich) and "subsidiary" (lysine-rich) histones was proposed (Stedman and Stedman, 1950). In this work, the authors also hypothesized that differences in the histone composition of the nuclei could perhaps account for the different physiological functions of the different cells in a same organism. With the exponential improvements in analytical methodologies in the following decades, these two groups would be better characterized, different histones individualized, and a unified nomenclature would be adopted: the core histones would be named H2A, H2B, H3, H4 and the linker histone named H1 (Bradbury, 1977). H2A and H3 histones have several variants, which diversify the possible histone combinations and contribute to a dynamic chromatin organization (Takizawa and Kurumizaka, 2022). One particular histone H3 variant, CENP-A, will be described in detail on the next chapter (section II.2.1).

Histones are small basic proteins consisting of a globular histone fold domain (HFD), and a flexible and charged N-terminal histone tail. The N-terminal tails of the core histones can undergo a range of at least eight different types of post-translational modifications (PTM) in multiple residues: acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines, ubiquitination of lysines, SUMOylation [conjugation of a small protein named SUMO, for Small Ubiquitin-like MOdifier]) of lysines, ADP ribosylation of glutamic acid, arginine deiminiation (or citrullination), and proline isomerization (**Figure 4**) . Histone PTMs regulate the access of chromatin-associated proteins to the DNA, and dictate dynamic transitions between transcriptionally active or silent chromatin states in what is known as "the histone code" (Jenuwein and Allis, 2001; Kouzarides, 2007).

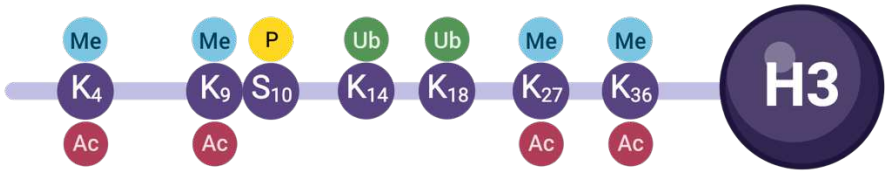

Figure 4. Histone H3 and some of its N-terminal tail post-translational modifications
Schematic representation of the histone H3 and its flexible N-terminal tail. Some lysine residues and their modifications are highlighted; they will be mentioned throughout the manuscript. Many other modifications exist (not depicted here) for H3 as well as for the other core histones and their variants. Ac: acetylation, Me: methylation (can be mono, di or tri), P: phosphorylation, Ub: ubiquitination. Acetylation and methylation are mutually exclusive. Created with BioRender.com.

B) Nucleosomes and the primary chromatin structure

The first model for chromatin structure was put forth not long after the DNA structure was resolved, and also emanated from X-ray diffraction studies. These early results were interpreted such as histones would cause DNA coiling and form bridges between different DNA molecules "*to link together generically similar DNA molecules to form a polytenic chromosome*" (Wilkins et al., 1959). This idea would evolve into the proposition of a "super-coil model", where histones would coat the DNA and induce a single, large, and uniformly supercoiled helix of chromatin (Paedon and Wilkins, 1972). Biochemical data, however, pointed towards a repetitive structure, since endonucleases would digest the chromatin into a regular series of DNA bands, all multiples of the smallest size unit, and no series would be observed when digesting deproteinized DNA (Hewish and Burgoyne, 1973). The definitive proof of the repetitive model would come with the first visualizations of chromatin fibres by electron microscopy (EM), which revealed a "beads on a string" structure with spherical particles —first named "nu (ν) bodies" (Olins and Olins, 1974), later re-baptized as nucleosomes (Oudet et al., 1975)— connected by thin filaments (of DNA) (**Figure 5A, B**). This repetitive architecture was further strengthened by the compilation of biochemical and analytical evidence that led Roger D. Kornberg to propose that chromatin structure is "*based on a repeating unit of two each of the four main types of histone and about 200 base pairs of DNA*" (Kornberg, 1974). The first EM observations were performed on chromatin isolates from rat, chicken and calf, and were quickly expanded to several other organisms (Woodcock et al., 1976). By 1977, the existence of nucleosomes in all eukaryotes was well accepted (Kornberg, 1977). In the following decades several efforts were made to obtain crystals of the nucleosome core particle and culminated with the publication of a 2.8Å resolution structure in 1997 (Luger et al., 1997) (**Figure 5C).**

The nucleosome is the primary repetitive structural unit of the chromatin, and the term refers collectively to the wrapping of a histone core particle by DNA. The core particle is composed by an $(H3–H4)_2$ tetramer that associates with two H2A–H2B dimers (Kornberg and Thomas, 1974), forming an octamer of core histones around which 146 base pairs (bp) of DNA wraps 1.65 turns (Luger et al., 1997). The flexible N-terminal tails of the histones face outwards of the core particle and are therefore accessible for their post-translational modification and for interacting with other proteins. DNA-wrapped nucleosomes form a 11 nm wide fibre (**Figure 3**), which corresponds to the "beads on a string" structure observed in electron microscopy. Importantly, this wrapping occurs throughout the genome, regardless of the underlying DNA sequence. The linker histone H1 binds outside of the nucleosome core, packaging another 20 bps of DNA to form a chromatosome. Two adjacent nucleosome core particles are joined by a linker DNA of varying length (20-90 bp), depending on species and tissue. This nucleoprotein complex exists every 200±40 bp throughout all eukaryotic genomes (Malik and Henikoff, 2003), a testament of its pivotal role in evolution of the Eukarya domain of life.

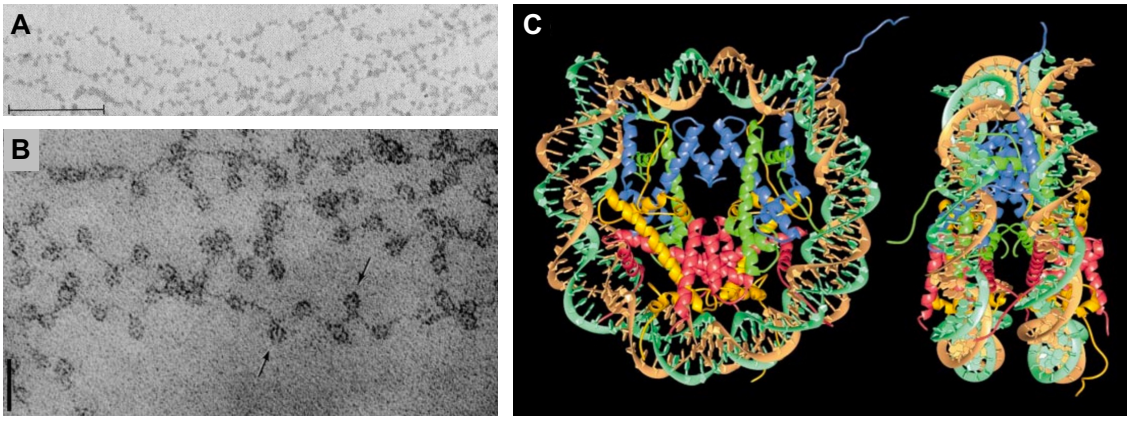

Figure 5. Chromatin fibres visualized by electron microscopy as "beads on a string", and the crystal structure of the nucleosome core particle (the beads).
(**A**) Olins & Olins, 1974 [*Figure 1A*]. Rat thymus chromatin fibers spilling out of a ruptured nucleus. Scale bar: 200 nm (**B**) Olins & Olins, 2003 [*Figure 3A*]. Low ionic-strength chromatin spread; the 'beads on a string'. Scale bar: 30 nm. (**C**) Lueger et al., 1997 [*Figure 1A*]. Nucleosome core particle at 2.8Å resolution. Ribbon traces of the 146-bp DNA backbones (brown and light green) and eight histone protein mainchains (blue: H3; green: H4; yellow: H2A; red: H2B). The views are down the DNA superhelix axis for the left representation and perpendicular to it for the right representation.

C) Chromatin fibres and higher-order structures

The second layer of higher-order packaging of chromosomal DNA is generally considered to be the folding and clustering of the nucleosomes to form a 30 nm chromatin fibre (**Figure 3**), which is readily visible in electron microscopy. The existence of such fibre in vivo, however, is still a matter of debate as it is considered by many to be an artifact of the low salt conditions used in the electron microscopy preparations (Maeshima et al., 2019). Early models proposed that the 30 nm fibre could be formed either as a solenoid (one-start model) or as a zigzag of nucleosomes (two-start model). Almost a decade ago the use of cryogenic electron microscopy (cryo-EM) revealed a new model for three-dimensional structure of the 30 nm fibre in a reconstitution of chromatin fibres with tandem repeats of nucleosome positioning DNA sequences and purified histones. The proposed structure showed a histone H1-dependent left-handed twist of repeating tetra-nucleosomal structural units, within which the four nucleosomes zigzag back and forth with a straight linker DNA in between (Song et al., 2014). These observations were performed on an artificial chromatin, and therefore the results were questioned. Recently, the combination of electron microscopy tomography with a DNA labelling method has allowed to visualize the chromatin inside the nucleus of cells in interphase and in mitosis. This study opposed the existence of an organized 30 nm fibre and concluded that the chromatin is a flexible and disordered granular chain, with a variable diameter between 5 and 24 nm which is packed at different densities in interphase nuclei and mitotic chromosomes (Ou et al., 2017). Whether an organized 30 nm fibre really exists or not, the accepted chromatin model indicates that additional higher-order structures of chromatin are formed to further compact the DNA into a

~250 nm fibre that will form the ~700 nm wide chromatid of mitotic chromosomes (**Figure 3**). It has been proposed that during mitosis (see section II.1), the condensin II protein complex forms large chromatin loops that are arranged as a spiral staircase helix around a central axis (Gibcus et al., 2018). As mitosis progresses, the loops become progressively larger and their number per helix turn increases. The very large condensin II-mediated loops are then sub-divided into smaller loops by condensin I, producing a nested loop arrangement that dramatically increases the linear density of the chromatin. By late prometaphase, the height of one helical turn was estimated to be ~200 nm (Gibcus et al., 2018), a value closely matching the text-book ~250 nm fibre.

D) Heterochromatin, Euchromatin, and all that's in between

Beyond its role in packing and compacting the DNA, the organization of the chromatin influences all nuclear functions that require a direct access to the DNA, in processes such as replication, transcription, and damage repair. Controlling of the chromatin dynamics, from the level of the nucleosome up to the higher order chromatin architecture is paramount for the proper function of the genome (Ray-Gallet and Almouzni, 2021). In 1928, German botanist Emil Heitz observed that in moss nuclei certain chromosomal regions were more densely stained than others and remained quite condensed and stained throughout interphase. He named these regions "heterochromatin" and the less-stained regions which readily decondensed in interphase, "euchromatin" (Heitz, 1928). Not long after, he proposed that euchromatin is related to gene activity, while heterochromatin corresponds to genetically inert regions. Heitz generalized his findings by documenting a similar compartmentalization of compaction (staining) in 115 species of plants and more than 70 species of insects (Passarge, 1979; Berger, 2019).

Specific histone PTMs are associated with each chromatin state, and profoundly influence transcription and gene expression (Talbert and Henikoff, 2021). For example, tri-methylation of lysine 9 and 27 in histone H3 (H3K9me3, H3K27me3) are heterochromatin marks par excellence, while H3K4 methylation and H3K36 acetylation are associated with active transcription. For a long time this strict bimodal chromatin organization was thought to separate the rigid, densely packed, gene poor and transcriptionally inactive heterochromatin from the open, lightly packed, gene rich, and often actively transcribed euchromatin. We know now that the situation is not black-or-white. The heterochromatin is actually dynamic, and plays extremely important structural and functional roles throughout the cell cycle (Allshire and Madhani, 2018). This is especially true in the case of facultative heterochromatin, which can be formed at previously euchromatic regions. This phenomenon has been observed, for example, to occur quite rapidly in the fission yeast *Schizosaccharomyces pombe* when facing glucose deprivation. In this scenario, facultative heterochromatin is assembled at the ribosomal gene cluster causing a transcriptional repression that is vital for the cellular adaptation to the starvation (Hirai et al., 2022).

II. The Centromere

The term centromere, from Latin *centrum* (center) and Greek *méros* (part) was coined by Cyril Dean Darlington, when he described the "*spindle fibre attachment* (to the) *chromosome*" (Darlington, 1936). Despite its name, the centromere is not always found at the centre of the chromosomes. Rather than defining a physical location, the term centromere refers to a region within each chromosome that carries out a particular function: the attachment of the chromosome to the microtubule spindle in mitosis (see section II.1) and meiosis. Failure to ensure this function has catastrophic consequences for the cells, causing chromosome mis-segregation which can lead to aneuploidy —an imbalance in the number of chromosomes—, structural alterations and tumorigenesis (reviewed in Holland and Cleveland, 2009). In human chromosomes, the centromere is also the primary constriction site where the two sister chromatids are bound. Depending on the position of their centromere, chromosomes can be called (sub)metacentric if both chromosome arms, p and q, are (roughly) the same length, acrocentric if one of the arms is considerably shorter than the other, or telocentric if the centromere is confounded with the tip of the chromosome, the telomere (**Figure 6**). Most eukaryotes are monocentric organisms, i.e., their chromosomes have a single centromere. They are generally (sub)metacentric or acrocentric, given the unstable nature of telocentric chromosomes. Certain species, such as the nematode and model organism *Caenorhabditis elegans*, several insects like the silk moth *Bombyx mori* and many plants are in contrast holocentric, meaning that spindle microtubules attach all along the chromosome arms (Melters et al., 2012; Senaratne et al., 2022) (**Figure 6**). All these organisms have regional centromeres, spread in relatively large loci. At the other end of the spectrum is *Saccharomyces cerevisiae* (budding yeast), which has point centromeres with a single nucleosome forming the minimal known unit of centromeric chromatin (Furuyama and Biggins, 2007). Bearing in mind this vast diversity, herein and unless stated otherwise when speaking about centromeres I will be referring specifically to the human monocentric, regional centromeres.

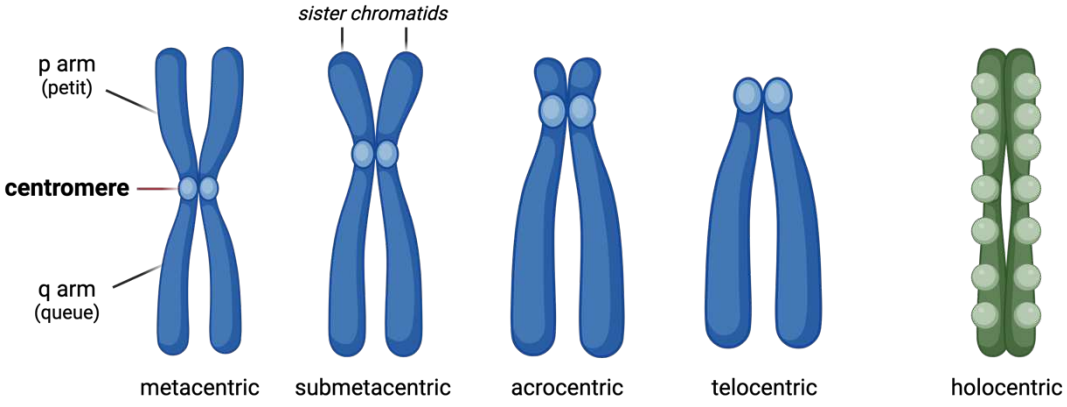

Figure 6. Chromosome types according to their centromere position
Monocentric chromosomes (in blue) compared to a holocentric chromosome (in green). The centromeres are highlighted as the lighter colour spheres in each chromosome. Created with BioRender.com

The centromere is composed by a network of proteins (see section II.2) associated to the centromeric DNA (see section II.3), and although centromere function and stability are essential for cell viability and proper cell division, both centromeric DNA and proteins evolve rapidly in what is known as the centromere paradox (Henikoff et al., 2001).

II.1. Centromeres in action throughout mitosis

Walther Flemming coined the term mitosis (from Greek *mitos*, thread), referring to the thread-like structures (chromosomes) he observed during the cell division process (Flemming, 1882). From his drawings, Flemming deducted correctly that chromosomes move in a sequential manner during mitosis, first getting aligned at the centre of the cell, and then being pulled apart to the extremities. Our modern understanding of the chromosome dynamics has allowed to differentiate five different phases of mitosis namely prophase, prometaphase, metaphase, anaphase and telophase, each determined by distinctive molecular events (O'Connor, 2008) (**Figure 7**).

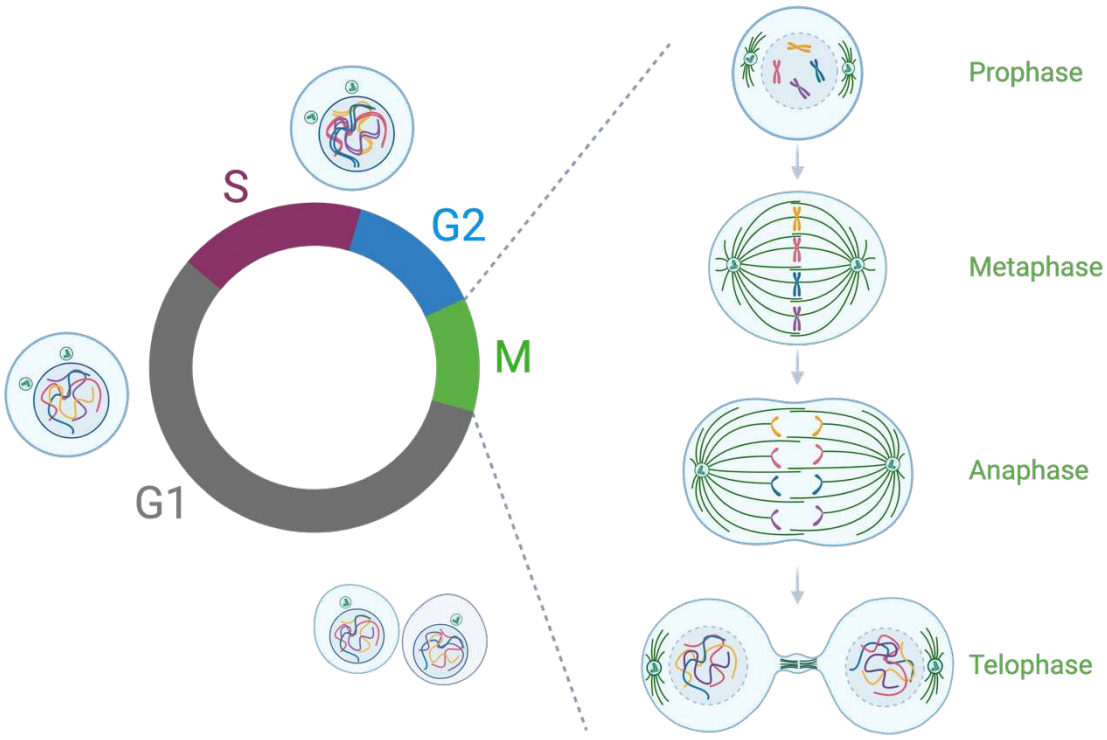

Figure 7. The cell cycle and the phases of mitosis.
A cell in G1 is represented with four chromosomes in yellow, red, blue, and purple. During S phase the DNA is replicated and in G2 the cell prepares to enter mitosis. In prophase each chromosome adopts its condensed, stereotypical form and the nuclear envelope breaks. In metaphase the chromosomes are attached at their centromere by the microtubule spindle (in green) and are aligned at the metaphase plate. In anaphase each sister chromatid is pulled by the microtubules towards opposing poles. In telophase the nuclear envelope begins to reform, the chromosomes begin to decondense, and cytokinesis ensues to separate the two daughter cells which enter G1 to repeat the cycle. Created with BioRender.com

To ensure the faithful propagation of the genetic material from mother to daughter, the genome is first duplicated during the S phase of the cell cycle, generating two identical copies of each chromosome named sister chromatids. As the copy is being synthesised, the sister chromatids are attached to each other through cohesin, a four proteins ring complex, that encircles both chromatids and ensures that the two remain in close proximity (Hagstrom and Meyer, 2003). At this stage, individual chromosomes are not distinguishable.

Prophase

Mitosis onset is triggered by the accumulation of cyclin B-CDK1 (cyclin-dependent kinase 1) and begins with prophase. In this phase of mitosis, the chromosomes start their condensation, a process that is mediated by the recruitment of condensin and that will continue until anaphase when chromosomes achieve reach their maximal compaction (Mora-Bermúdez et al., 2007). Condensin is a complex of five proteins that binds individual chromatids and promotes their condensation by generating and stabilizing large positively supercoiled DNA loops. As condensin is recruited, cohesin is removed from the arms of the chromosomes and retained exclusively at the centromeres. This first role of the centromere in mitosis allows the individual sister chromatids to be resolved while still keeping them tethered to each other (Hagstrom and Meyer, 2003). Here chromosomes begin to adopt their stereotypical form; two distinct rod-shaped sister chromatids joined at centromere, now clearly distinguishable as the primary constriction of the chromosome.

Prometaphase

The nuclear envelope rupture marks the beginning of prometaphase. Once the separation between cytosol and nucleus is no more, microtubules -long protein fibres that extend from the poles of the cell and form the mitotic spindle- can attach the chromosomes. The attachment is achieved through the interaction of the microtubules with the kinetochore, a complex of proteins that assembles for this purpose transiently in mitosis and specifically at the centromeres. This phase is also characterized by the congression of chromosomes towards the equator of the cell.

Metaphase

In metaphase the chromosomes assume their most compacted state and become aligned at the centre of the spindle in what is known as a metaphase plate. The high level of condensation achieved at this stage is crucial not only in reducing the volume of the individual chromosomes to facilitate their segregation to each daughter cell, but also in providing mechanical strength to withstand the pulling forces of the microtubules. In the metaphase plate all chromosomes are (should be) bi-oriented, with each sister kinetochore and therefore, sister chromatid, connected to microtubules that emanate from opposite spindle poles.

Anaphase

The spindle assembly checkpoint (SAC) (reviewed in McAinsh and Kops, 2023) monitors the attachment of chromosomes to the spindle and delays anaphase onset by inhibiting the anaphase-promoting complex/cyclosome (APC/C) until all chromosomes are successfully attached by microtubules, bi-oriented and under tension. Kinetochores that lack proper microtubule connections i) produce a diffusible anaphase inhibitor named mitotic checkpoint complex (MCC) that binds to and inhibits APC/C and ii) lack tension, a feature sensed by Aurora B, a kinase member of the chromosome passenger complex (CPC) that activates the SAC (reviewed in Carmena et al., 2012). Once all kinetochores are correctly attached, the SAC is relieved, and APC/C –which is an E3 ubiquitin ligase– triggers the irreversible mitotic exit via the polyubiquitylation and subsequent proteasomal degradation of cyclin B1 and securin. The degradation of securin activates separase, a protease that cleaves a subunit of the cohesin rings that remained bound at the centromeres (Hagstrom and Meyer, 2003) allowing both sister chromatids to be pulled towards opposite poles of the cell in a burst of movement generated by the depolymerization of the spindle microtubules.

Telophase

In the last phase of mitosis, the segregated chromosomes begin to decondense, the nuclear envelopes begin to re-form around them, and the process of cytokinesis divides the cytoplasm, generating two daughter cells, each one carrying a copy of the genetic material of the maternal cell.

For a long time, the difference between centromere and kinetochore (from the Greek *kineto*, 'move' and *chore*, meaning 'means for distribution') was not well-defined, and the terms were often used as synonyms. The difference is now clear; centromere refers to the stable chromatin foundation (DNA and proteins) upon which the kinetochore (protein network) is assembled transiently for mitosis (Fukagawa and Earnshaw, 2014).

II.2. Centromeric Proteins

Compared to the total size of the chromosomes, centromeres are very small. This characteristic historically made impossible the study of their specific protein composition by classical analytical methods. Serendipity led to the discovery of anti-centromere antibodies (ACA) in the sera of some patients with CREST syndrome (Calcinosis, Raynaud's phenomenon, Esophogeal dysmotility, Sclerodactyly, Telangiectasia), which was seen to stain the nuclei of histological sections of rat liver in a speckled pattern (Moroi et al., 1980; for an historical account see Earnshaw, 2015). The characterization of the sera containing ACA by affinity elution from gel led to the identification and designation of the first CENtromeric Proteins (CENPs) in human cells, namely CENP-A, CENP-B and CENP-C (Earnshaw and Rothfield, 1985). Over the years many other CENPs, that will be further discussed below, have been identified (Sugata et al., 2000; Nishihashi et al., 2002; Minoshima et al., 2005; Foltz et al., 2006; Okada et al., 2006; Izuta et al., 2006; Hori et al., 2008; Amano et al., 2009).

II.2.1. CENP-A

A) A Histone H3 variant at the core of the centromeric chromatin

CENP-A was first identified roughly at the same time and independently in Germany (Guldner et al., 1984) and in the USA (Earnshaw and Rothfield, 1985). The 1984 publication by Hans Guldner and colleagues gave no name to the targeted polypeptide, but already described it as not being one of the core histones, yet not soluble under conditions which favour the release of nuclear ribonucleoprotein particles. CENP-A behaved as a "histone-like" component of centromeric chromatin. It was found on nucleosome-like structures following micrococcal nuclease (MNase) digestion of nuclei from a human cell line (HeLa) and it co-eluted with core histones H3 and H4 under stringent isolation conditions (Palmer et al., 1987). Unlike the core histones, which mostly if not completely get replaced by protamines in spermatozoa, CENP-A remains present in comparable amounts to somatic cells (Palmer et al., 1990). This key observation allowed for the purification of CENP-A protein from bull sperm, without other contaminant histones. Peptide sequencing of most of this purified protein revealed that it had segments with high sequence similarity to bovine H3 (Palmer et al., 1991). These peptide sequences permitted in turn the synthesis of degenerate oligonucleotide primers for the amplification of bovine CENP-A cDNA, which thereafter was used for the isolation of full-length human CENP-A cDNA from cDNA pools, and its subsequent sequencing (Sullivan et al., 1994). The sequence revealed that the C-terminal ends of CENP-A and H3, where the histone fold domain (HFD) is located, share 60% identity and about 75% similarity, while their N-terminal domains are divergent (**Figure 8**). In the same work it was also determined that, surprisingly, the HFD and not the unique sequence of the NH2-terminal end dictated the centromeric specific localization of CENP-A. It was later understood that within the dissimilar amino acids found in the HFD, there is a sequence encompassing the loop 1 and α2 helix, relatively conserved in many CENP-A orthologues and divergent from H3, that is sufficient to direct CENP-A and a chimeric $H3^{CATD}$ to the centromeres, and therefore was named CENP-A targeting domain (CATD) (Black et al., 2004).

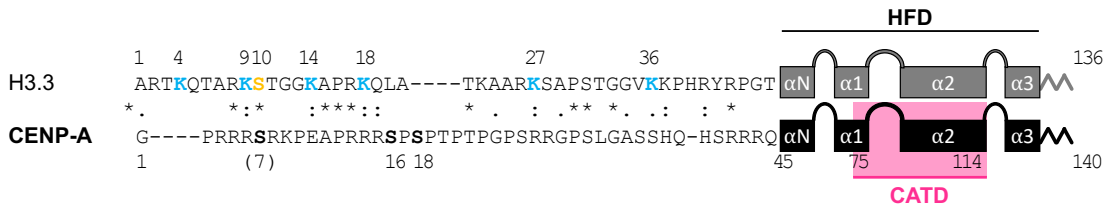

Figure 8. Comparison and partial alignment of CENP-A and H3.3
Alignment of the N-terminal tail of human histone H3.3 (NP_001365976.1) and human CENP-A (NP_001800.1) with ClustalOmega (.) small positive score, (:) similarity which scores more than 1.0, (*) conserved amino acid. The relatively conserved histone fold domain (HFD) of both proteins is schematically represented. The CENP-A targeting domain (CATD) is highlighted in pink. In H3.3 the residues subjected to PTMs are highlighted in blue and yellow (related to Figure 4); note the absence of corresponding lysine residues in CENP-A. Proposed residues of CENP-A PTMs are in bold; due to historical reasons on early works numbering the amino acids from the initial methionine, serine 6 is numbered here and widely referred to as serine 7.

The CATD not only targets CENP-A to the centromeres, but it also mediates its interaction with another centromeric protein, CENP-N (Carroll et al., 2009). Other crucial CENP-A interactions with different centromeric proteins depend on the specific sequences of both terminal tail regions, which differ from canonical histone H3. The N-tail of CENP-A interacts directly with CENP-B (Fachinetti et al., 2013; Fujita et al., 2015) and with CENP-T (Folco et al., 2015; Logsdon et al., 2015) while the C-tail, along with the CATD, are required for CENP-A interaction with CENP-C (Carroll et al., 2010; Fachinetti et al., 2013; Logsdon et al., 2015).

CENP-A has orthologs in most —but not all— eukaryotes and is known by several names: Cse4 in budding yeast *Saccharomyces cerevisiae*, Cnp1 in fission yeast *Schizosaccharomyces pombe*, CID in the fly *Drosophila melanogaster*, cenH3 in many plants and HCP3 in the nematode *Caenorhabditis elegans*. It is an essential protein; *Cenpa* null mice are not viable beyond 6.5 days and immunofluorescence analysis of 5.5 days embryos revealed that, in absence of CENP-A, there is also a total loss of CENP-C at the centromeres, and a diffuse CENP-B signal throughout the nucleus on top of a relocalization to larger foci due to loss of centromeric chromatin compaction (Howman et al., 2000). The knockdown of CENP-A leads to chromosomal mis-segregation, the formation of micronuclei, nuclear deformation and blebbing. CENP-A is required for the recruitment to the centromere of all kinetochore components analysed, in all species studied: mouse (Howman et al., 2000), *Drosophila* (Blower and Karpen, 2001), *C. elegans* -which is holocentric- (Oegema et al., 2001), *S. cerevisiae* (Collins et al., 2005), chicken cells (Régnier et al., 2005) and human cells (Goshima et al., 2003; Liu et al., 2006; Foltz et al., 2006). Moreover, CENP-A always is localized at the active centromeres, and is sufficient to promote kinetochore assembly when targeted artificially to ectopic loci (Van Hooser et al., 2001; Heun et al., 2006; Barnhart et al., 2011; Guse et al., 2011; Mendiburo et al., 2011; Hori et al., 2013; Logsdon et al., 2015). Taking all this evidence together, it was established that CENP-A containing chromatin is the epigenetic mark, both necessary and sufficient to define, maintain and propagate the centromere function, and this regardless of the underlying DNA sequence (apart from *S. cerevisiae,* see section II.3). This is achieved through a two-step mechanism; first by templating its own CATD-dependent replication and depositon (see sub-section D) and second by nucleating the kinetochore assembly (Fachinetti et al., 2013).

As all histones, CENP-A can be subjected to post-translational modifications (reviewed in García Del Arco and Erhardt, 2017; Srivastava and Foltz, 2018), but whether if and which are relevant for the centromeric function is still a matter of intense debate. The human N-terminal tail of CENP-A is enriched in arginines that do not appear to be frequently modified and lacks most of the lysines present in the histone H3 tail that are well-characterized sites of PTMs. Further, even among CENP-A orthologs there is an important size and sequence divergence that deters from thinking of a universal functional centromeric PTM code. The length of the tail alone ranges from 19 amino acids in *S. pombe*, 44 amino acids in humans, and up to 200 amino acids in *C. elegans* (Smith, 2002).

The role of the phosphorylation in serine residues 7, 16 and 18 of the human CENP-A tail and has been nonetheless interrogated. Chromosome condensation and subsequent segregation during mitosis requires the phosphorylation of canonical histone H3 at serine 10 (H3Ser10p), a process that initiates during G2 in pericentric foci prior to its expansion to the chromosome arms and persists throughout mitosis. CENP-A Ser7 is also phosphorylated and this was thought precede and perhaps initiate the H3Ser10 phosphorylation process (Zeitlin et al., 2001). It was demonstrated, however, that CENP-A Ser7 phosphorylation is short lived, starting in prophase –after H3Ser10p is well underway–, peaking in prometaphase, and lost during anaphase. The exact role of this modification is debated, with some evidence saying the prevention of Ser7 phosphorylation results in misalignment of chromosomes and impaired kinetochore attachment to microtubules (Kunitoku et al., 2003; Goutte-Gattat et al., 2013) while other points towards a totally dispensable role of this mark (Barra et al., 2019). On the other hand, there is clear evidence that phosphorylation of serines 16 and 18, which occurs in pre-nucleosomal CENP-A and forms a salt-bridged secondary structure that makes intra-nucleosomal CENP-A N-tails associations (Bailey et al., 2013), are two key modifications for proper centromere function (Barra et al., 2019; Takada et al., 2017).

In addition to the classical N-tail PTMs, CENP-A has been proposed to be phosphorylated at serine 68 and methylated, phosphorylated and/or ubiquitinated at lysine 124. Neither of these two residues are part of the CATD, yet both modifications were suggested to be crucial for CENP-A deposition (Niikura et al., 2015, 2017, 2019). All functional relevance of these two modifications have been contradicted by the Fachinetti and the Black laboratories, most recently in a work I carried out in parallel to my main thesis project and that was published in 2021 (Fachinetti et al., 2017; Salinas-Luypaert et al., 2021).

B) The CENP-A nucleosome

CENP-A purified from HeLa cells can replace histone H3 in nucleosome reconstitutions in vitro and the basic structure of CENP-A containing nucleosomes was observed to be the same as that of those formed by the core histones, with one CENP-A–H4 tetramer and two H2A–H2B dimers (Yoda et al., 2000). The crystal structure of CENP-A nucleosomes resolved at 3.6 Å confirmed these findings and revealed several interesting particularities. Compared to H3 nucleosomes which wrap 146 bp of DNA (Luger et al., 1997), the CENP-A nucleosomes only wrap tightly 121 bp, leaving the DNA segments at the entrance and the exit of the CENP-A nucleosome much more flexible and dynamic (**Figure 9**) (Tachiwana et al., 2011).

Over the years, conflicting evidence emerged regarding the histone composition and stoichiometry of the CENP-A nucleosomes in vivo. Several hypotheses were proposed such as the occurrence of tetrasomes (loss of H2A-H2B dimers), hemisomes (one CENP-A-H4 dimer + H2A-H2B) and

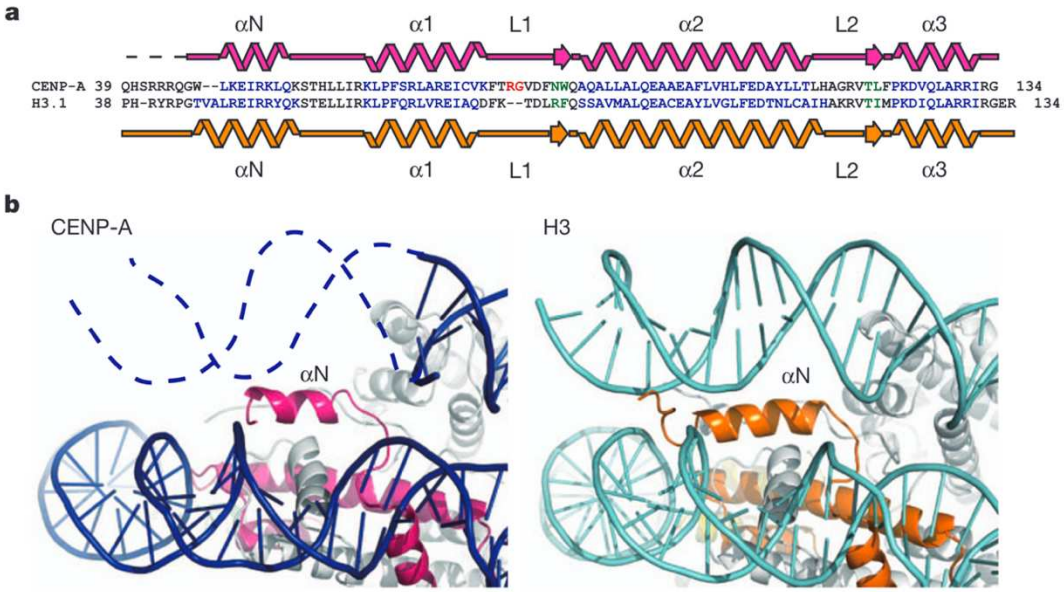

Figure 9. Structure of the DNA entrance and exit of the human CENP-A nucleosome. From Tachiwana et al., 2011 [*Figure 2*] (**a**) The sequences of human CENP-A and H3 are aligned, and secondary structure elements are shown (**b**) Close-up views of the αN helices and the DNA edge regions of the CENP-A (left panel) and H3 (right panel) nucleosomes. The dashed line in the left panel shows the DNA region that is not visible in the crystal structure (because it is too flexible to be crystalized). The CENP-A and H3 molecules are shown in magenta and orange, respectively.

hexasomes (reviewed in Black and Cleveland, 2011). Through native chromatin immunoprecipitation followed by sequencing (ChIP-seq) of CENP-A containing particles, it was demonstrated that the predominant form of CENP-A nucleosomes is octameric, and it was speculated that non-octameric species might be present either only at very low levels or transiently during the cell cycle (Hasson et al., 2013). These exceptions were subsequently fully discredited: CENP-A chromatin extracted from cells synchronized in G1, G2 and mitosis does not contain H3, co-purifies with H2A, H2B, and H4 at levels expected for a homotypic octamer and when fully digested by MNase, yields DNA fragments only of the expected length after an octamer protection (Nechemia-Arbely et al., 2017).

Human centromeres were estimated, though a combination of biochemical methods and imagining, to contain one CENP-A nucleosome for every 25 nucleosomes (Bodor et al., 2014). Despite being low, this number represents a 50 x enrichment compared to the overall genome. New insights have been provided with more precision through DiMeLo-seq, a novel sequencing technique that allows to map protein–DNA interactions with high resolution on native, long-read, single molecule DNA sequences, while simultaneously measuring endogenous DNA modifications and sequence variation (Altemose et al., 2022b). CENP-A-directed DiMeLo-seq has revealed that, at least for chromosome X, in the centromeric region where CENP-A is most abundant one in four nucleosomes is a CENP-A nucleosome, which is roughly six times more than the previous estimate. Regardless of the exact number, CENP-A nucleosomes are still a minority, and are surrounded by H3 nucleosomes.

The crystal structure of CENP-A nucleosomes also revealed that the loop 1 region, which is part of the CATD, has two extra amino acids (Arginine 80 and Glycine 81) that form an RG-loop which protrudes from the nucleosome core (Sekulic et al., 2010; Tachiwana et al., 2011). This means that, unlike its H3 counterpart, this region is accessible for interactions, notably with CENP-N (see sub-section II.2.2). A recent cryo-EM study on the 3D structure of groups of three consecutive nucleosomes revealed that the H3-CENP-A–H3 tri-nucleosome is less twisted than a purely H3 tri-nucleosome, therefore exposing the central CENP-A nucleosome more to the solvent (Takizawa et al., 2020). These two molecular traits that make CENP-A nucleosomes more accessible than their H3 counterparts may explain in part how centromeric proteins can specifically target the CENP-A nucleosomes, even when they are speckled in a sea of H3.

C) The centromeric chromatin and its organization

Despite historically being considered as heterochromatin, the centromere core chromatin (containing CENP-A nucleosomes) is clearly distinct and is therefore referred to as "centrochromatin". The staining of extended human centromeric chromatin fibres revealed that the centromere core has very particular histone post-translational modifications, which do not fall into the standard patterns of the heterochromatin/euchromatin dichotomy. The H3 domains interspersed in the CENP-A containing blocks show mono and di-methylation of lysine 4 (H3K4me1/2) and di- and tri-methylation of lysine 36 (H3K36me2/3), all marks permissive for transcription (see section II.3.2), and at the same time lack all the acetylation marks in both H3 and H4 that normally define the active chromatin state. They also lack the histone H3 di- and tri-methylation of lysine 9 (H3K9me2/3), which are characteristic marks of inactive/condensed chromatin and that are present in the pericentromeric heterochromatin (Sullivan and Karpen, 2004; Gopalakrishnan et al., 2009; Bergmann et al., 2011).
A recently published preprint provides evidence hinting towards the same notion with a radically different approach. Through state-of-the-art sequencing and mapping methods, the authors show that the centromeric chromatin presents highly accessible, clustered patches of chromatin embedded in extremely compacted regions. They go as far as saying that the centromere core has at the same time the most compacted chromatin and "*the most accessible domains within the entire human genome*", and propose to refer to it as "dichromatin" (Dubocanin et al., 2023).

The three-dimensional organization of the centromeric chromatin has been a topic of interest since early on, given its crucial role in forming nucleation point for the kinetochore assembly. Staining of stretched "kinetochores" (centromeres) with ACA showed back in 1991 a punctate linear array of fluorescent ACA positive subunits arranged in a repetitive pattern along a centromeric DNA fibre (Zinkowski et al., 1991). The staining of extended chromatin fibres showed that CENP-A and H3 nucleosomes are indeed interspersed (Blower et al., 2002). These observations led to the proposition of a looping chromatin model that would essentially cluster together in space these linearly scattered

CENP-A domains, therefore exposing the microtubule-binding segments to the outsides of the chromosomes (Zinkowski et al., 1991; Blower et al., 2002). Other proposed models for chromatin structure include solenoids (Blower et al., 2002) and a layered sinusoid boustrophedon (Ribeiro et al., 2010) (**Figure 10**). That CENP-A nucleosomes mostly face the poleward face of the metaphase chromosomes is a generally accepted fact. The best evidence of this was provided by serial-section transmission electron microscopy study of human mitotic chromosomes, where it was seen that at the centromeres, CENP-A occupies domains in the inner kinetochore plate, stretching across two thirds of the length of the constriction but encompassing only one third of the constriction width and height (Marshall et al., 2008b). However, how this exact 3D organization is achieved starting from few CENP-A molecules dotted between H3 nucleosomes is still unclear. A recent proposition has emanated from the combination of super resolution expansion microscopy (ExM), Capture-C analysis and polymer modelling. The authors propose that vertebrate centromeres are partitioned into two distinct subdomains during mitosis, each capable of binding distinct microtubule bundles. Condensin appears to be crucial for the formation of these subdomains, which then are linked and stabilized by cohesin (Sacristan et al., 2022).

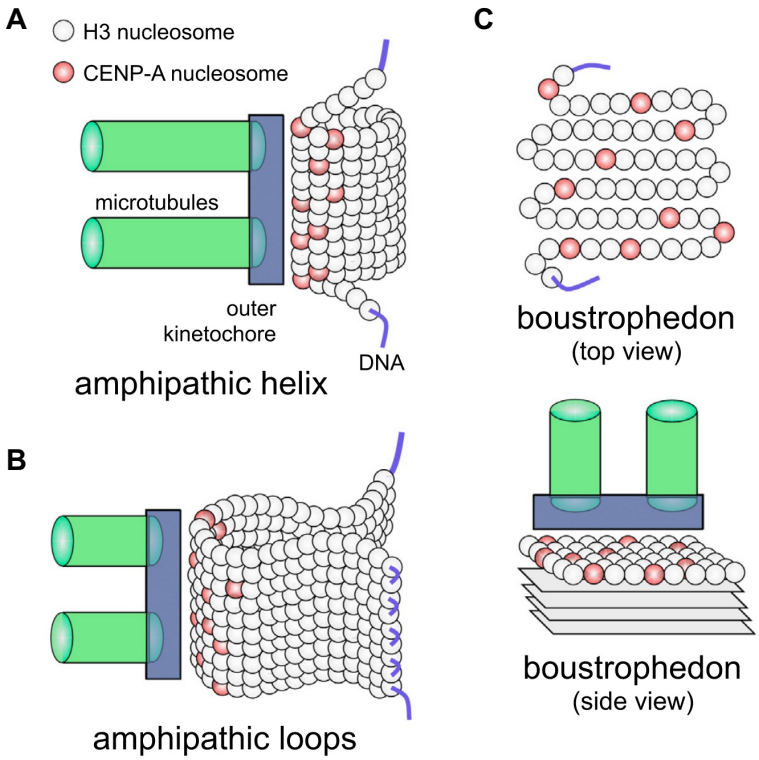

Figure 10. Models for CENP-A and H3 nucleosomes organization at the centromere.
Adapted from Fukagawa & Earnshaw, 2014 [*Figure 5*]. (A and B) The CENP-A chromatin was originally suggested to have an amphipathic organization (Blower et al., 2002), with CENP-A on the exterior facing the kinetochore, and H3 largely on the interior. This chromatin was proposed to form either a solenoid helix (**A**) or a loop structure (**B**). The boustrophedon model of centromeric CENP-A-containing chromatin was proposed based on super-resolution microscopy (Ribeiro et al., 2010).

D) CENP-A deposition at the centromere

Deposition of CENP-A at the centromeres is mediated by its chaperone, the Holiday Junction Recognition Protein (HJURP) (Dunleavy et al., 2009; Foltz et al., 2009). The CATD of CENP-A is recognized by the conserved N-terminal CENP-A binding domain (CBD) of HJURP, which directs the deposition of the pre-nucleosomal CENP-A/H4 tetramer on DNA (Shuaib et al., 2010; Hu et al., 2011). Super-resolution microscopy analyses have recently revealed that in early G1, HJURP seems to act as a nucleation point around which numerous CENP-A pre-nucleosomes arrange themselves in rosette-like clusters, forming large globular structures of ~250-300 nm diameter. As G1 progresses HJURP is lost from the centre of the rosette and CENP-A nucleosome clusters get increasingly compacted (Andronov et al., 2019).

The canonical histones are deposited in a DNA-replication dependent manner, and their expression peaks during S phase (MacAlpine and Almouzni, 2013). On the contrary, both CENP-A mRNA and protein levels peak during late G2 (Shelby et al., 1997, 2000), and their deposition is uncoupled to the replication of the centromeric DNA. New CENP-A nucleosomes are loaded into the centromeric chromatin starting in late telophase and though early G1 of the following cell cycle where they were synthesized; passage through mitosis is indispensable for the loading to occur (Jansen et al., 2007). The newly deposited CENP-A are then redistributed to, and retained by, daughter centromeres in the subsequent S phase. During the centromeric replication, both the canonical histone H3.1 and the variant H3.3 are deposited at centromeres, and while H3.1 remains as the H3 component of the centromeres, H3.3 is replaced by CENP-A in the following G1 (Dunleavy et al., 2011). All of this implies that , surprisingly, cells go through mitosis with half the amount of CENP-A they had loaded in G1 (Jansen et al., 2007).

The centromere position and identity are replicated and maintained between generations of cells in two steps; first by the substitution of H3.3 by CENP-A, and then by the nucleation of kinetochore assembly onto CENP-A containing chromatin (Fachinetti et al., 2013). Once the first steps of centromere assembly have been completed in G1 with the deposition of CENP-A, its presence at the centromeres is no longer required for maintaining kinetochore attachment, nor for ensuring centromere function during the following mitosis. However complete, rapid degradation of endogenous CENP-A using an auxin-inducible degron (AID) tag, prior to kinetochore assembly prevents the deposition of CENP-C and CENP-N (but not of CENP-T) and produces defective kinetochores which lead to chromosome segregation failure. In absence of CENP-A, the binding of CENP-B to the centromeric DNA (see section II.2.3) becomes essential to preserve CENP-C and kinetochore anchoring at the centromere (Hoffmann et al., 2020, 2016).

II.2.2. The CCAN

A group of 16 centromeric proteins (CENPs) stably associate with the centromeric chromatin specified by CENP-A throughout the cell cycle, and are therefore collectively referred to as the Constitutive Centromere Associated Network (CCAN) (reviewed in: Hara and Fukagawa, 2017). The CCAN proteins are organized into five sub-complexes composed respectively by CENP-C, CENP-L/N, CENP-T/W/S/X, CENP-H/I/K/M and CENP-O/P/Q/U/R (**Figure 11**). Only last year was the structure of this very complex scaffold of proteins been reconstituted by means of cryo-EM, both in presence (Yatskevich et al., 2022) and absence of a CENP-A nucleosome and DNA (Pesenti et al., 2022). While CENP-A epigenetically defines the centromere, its sole presence does not imply that the centromere is functional (i.e., capable of supporting the kinetochore assembly). Ectopically incorporated CENP-A results in the recruitment of only other three centromeric proteins and is insufficient for sustaining the kinetochore assembly; other key proteins of the CCAN are hence necessary to ensure the function of the centromere (Gascoigne et al., 2011). The CCAN is at the same time a non-histone component of the centromeric chromatin and a stable centromere-bound sub-complex of the kinetochore, often referred to as inner kinetochore. The central role of the CCAN is to recruit the core microtubule binding proteins of the kinetochore known as the KMN network (KNL1 complex/Mis12 complex/Ndc80 complex) to the centromere locus during mitosis (Cheeseman et al., 2006) (**Figure 11**). A clear interdependency between different CCAN sub-complexes for their centromeric localization has been reported. All CCAN proteins depend on CENP-C, the CENP-T/W/S/X, CENP-L/N and CENP-H/I/K/M complexes are interdependent, and the CENP-O/P/Q/U/R complex is downstream of the others (McKinley et al., 2015; Klare et al., 2015).

At the tip of the CCAN hierarchy is CENP-C, which as CENP-A, is essential for normal mouse embryonic development (Kalitsis et al., 1998). It is necessary for cells to progress into anaphase (Fukagawa, 1999) and was early on detected at the inner kinetochore plate (Saitoh et al., 1992). CENP-C plays a fundamental role in the incorporation and stabilization of CENP-A nucleosomes (Dambacher et al., 2012; Falk et al., 2015; Moree et al., 2011) and interacts directly not only with CENP-A, but also with CENP-B (Suzuki et al., 2004; Fachinetti et al., 2015), the CENP-H/I/K/M complex (Klare et al., 2015) and the CENP-L/N complex (McKinley et al., 2015). CENP-C knockout in human cell lines drastically reduces CENP-A and all other CCAN sub-complexes levels at the centromeres (McKinley et al., 2015). It is therefore widely considered as a blueprint centromeric protein for kinetochore assembly (Klare et al., 2015). Moreover, the N-terminal region of CENP-C has been demonstrated to be crucial for its direct interaction with the Mis12 complex, one of the microtubule binding complexes of the KMN network (Przewloka et al., 2011; Screpanti et al., 2011; McKinley et al., 2015); CENP-C therefore also directly bridges the centromeric chromatin to the outer kinetochore.

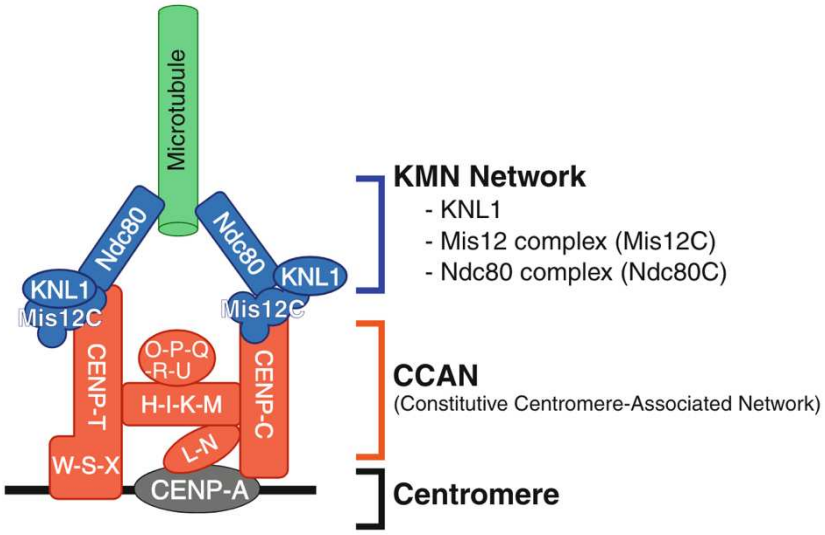

Figure 11. Schematic representation of the centromere, the CCAN (inner kinetochore) and the KMN network (outer kinetochore) bound to a microtubule.
From Hara & Fukagawa, 2017 [*Figure 6*]. The CCAN constitutively localizes to the centromere throughout the cell cycle and functions as a foundation to create link between the centromere and the microtubules. CENP-C and CENP-N (CENP-L/N subcomplex) make direct contact with CENP-A. The CENP-T/W/S/X subcomplex form a nucleosome-like structure that binds the DNA; therefore, the CCAN interacts with the centromere through three interfaces, CENP-C, the CENP-L/N and CENP-T/W/S/X. In mitosis, the KMN network, which binds to microtubules, is recruited onto the CCAN independently by CENP-C and CENP-T meaning that there are two parallel pathways from the centromere to the microtubules.

Another key component of the CCAN that can also directly bind kinetochore components is CENP-T (CENP-T/W/S/X complex). Ectopic targeting of CENP-C and/or CENP-T are enough to properly assemble the microtubule-binding KMN network (Gascoigne et al., 2011). While CENP-C preferentially recruits the Mis12 complex, CENP-T preferentially recruits the Ndc80 complex, indicating that the two provide parallel ways to ensure proper kinetochore assembly. CENP-T has a histone-like fold domain at the C-terminal region and forms a nucleosome-like complex with CENP-W that binds the centromeric DNA (Hori et al., 2008; Nishino et al., 2012). The extended N-terminal region of CENP-T spans to the outer kinetochore and interacts the NdC80 complex. Thus, as CENP-C, CENP-T interacts directly with both the centromeric chromatin and microtubule-binding kinetochore complexes.

CENP-N (CENP-L/N) interacts directly with CENP-A nucleosomes (Carroll et al., 2009), CENP-C, and the CENP-H-I-K-M complex (McKinley et al., 2015). It has been recently proposed that CENP-N favours the stacking of CENP-A nucleosomes, therefore promoting the formation of higher order chromatin structures at the centromere (Zhou et al., 2022). CENP-U (CENP-O/P/Q/U/R complex) was seen to be essential for cell viability during mouse early embryogenesis, while CENP-U-deficient mouse embryonic fibroblast and chicken DT40 cells are

viable but present important mitotic defects (Kagawa et al., 2014). CENP-Q was proposed to bind microtubules directly through its N-terminal tail (Pesenti et al., 2018). CENP-I (CENP-H/I/K/M complex) has been recently demonstrated to interact directly with the A/T rich centromeric DNA through a specific DNA-binding surface, which is required for its centromeric localization and for proper chromosome alignment during mitosis (Hu et al., 2023).

Extensive work by many groups constantly sheds new light on the complex interactions between the CCAN sub-complexes, their individual proteins, the centrochromatin and the kinetochore. Further details on each CCAN sub-complex are beyond the scope of this thesis. The take home message of this section is that the centromeric proteins from a complex network of interactions, and that a positive feedback loop between CCAN proteins and CENP-A determines that the CCAN is at the same time assembled into CENP-A chromatin and required for the incorporation of newly synthesized CENP-A into the centromeres (Hori et al., 2013; Okada et al., 2006).

II.2.3. The paradoxical CENP-B

CENP-B was the first of the human centromeric proteins to be cloned (Earnshaw et al., 1987), but the sequence did not reveal much of its function. A couple of years later it was demonstrated that a cloned centromeric DNA fragment incubated with a cell nuclear extract was selectively immunoprecipitated by the anticentromere sera from CREST patients. CENP-B was identified as the recognized antigen for the immunoprecipitation, specifically binding a non-palindromic A/T-rich 17-bp motif present on the centromeric DNA which was named "CENP-B box" (Masumoto et al., 1989). Immunoelectron microscopy experiments performed shortly after supported these findings, detecting CENP-B throughout the innermost heterochromatic region of the kinetochores of human chromosomes throughout the cell cycle (Cooke et al., 1990). Indeed, CENP-B staining is remarkably different from that of CENP-A (**Figure 12A**).

To the extent of our current knowledge, CENP-B is the only centromeric protein that binds the centromeric DNA in a sequence specific manner. The identification its N-terminal DNA binding domain (DBD) and its C-terminal dimerization domain (**Figure 12B**), came with the idea that CENP-B may organize the centrochromatin structure through its protein-DNA and protein-protein interactions (Yoda et al., 1992). Both terminal domains are fully conserved between human and mouse CENP-B, and the formation of homodimers, with each CENP-B molecule being able to bind DNA strand, was later confirmed (Kitagawa et al., 1995). Two-hybrid experiments would afterwards detect a direct interaction between CENP-B's acidic domain and CENP-C (**Figure 12B**), and suggested that their interaction may be involved in the correct assembly of CENP-C on the centromeric DNA (Suzuki et al., 2004). It was later demonstrated the binding of CENP-B to the centromeric DNA repeats stabilizes CENP-C and favours the nucleation of the kinetochore assembly

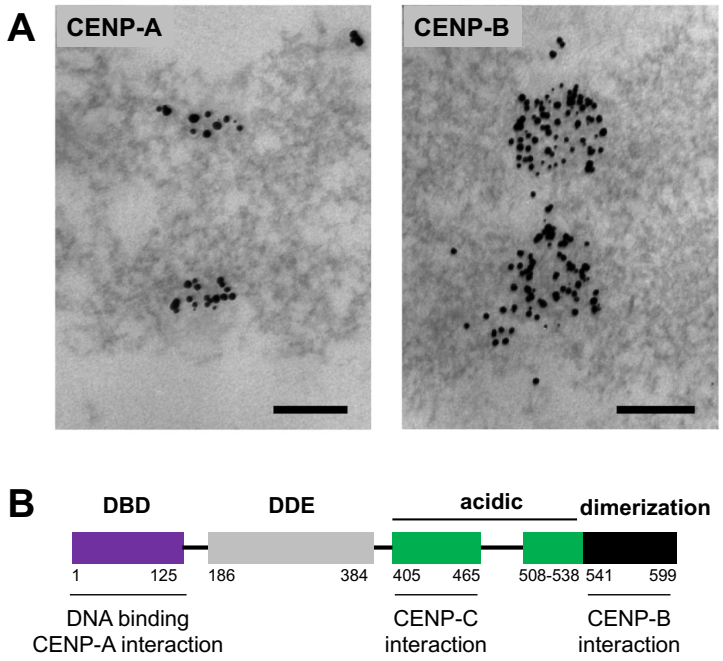

Figure 12. Comparison of electron micrographs of CENP-A and CENP-B and schematics of CENP-B domains and their interactions.
(**A**) Adapted from Marshall et al., 2008b [*Figures 1B-iii and 1D*]. Electron micrographs of acetone fixed chromosomes labeled for CENP-A (left) and CENP-B (right). Section thickness: 60 nm. Scale Bars: 200 nm. Electron dense regions correspond to the two sister chromatids. While CENP-A occupies a small proportion of the outermost centromeric chromatin, CENP-B staining is spread throughout the full thickness of the chromatids. (**B**) Schematic representation of the CENP-B protein, its domains and interactions. DBD (DNA binding domain), DDE (domain with homology to pogo-like transposases).

(Fachinetti et al., 2015). *In vivo*, binding of CENP-B to the CENP-B box enhances the retention of the preassembled CENP-A nucleosome on centromeric DNA, and *in vitro* it was demonstrated that CENP-B interacts, through its DBD, with the CENP-A-H4 complex, but not with the H3.1-H4 complex (Fujita et al., 2015).

CENP-B protein appears to have evolved from a member of the pogo-like transposase family which was domesticated by losing its transposase ability but retaining its DNA-binding capacity (Kipling and Warburton, 1997). While CENP-A has orthologs in all eukaryotes, and CENP-C in most of them, CENP-B is only present in some clades. Yet CENP-B, CENP-B-like proteins, CENP-B boxes and CENP-B box-like motifs have arisen independently multiple times in evolution, implying it has to provide some evolutionary advantage (Drinnenberg et al., 2016; Dumont and Fachinetti, 2017; Gamba and Fachinetti, 2020). CENP-B is not an essential protein for centromere or kinetochore function. It is absent from the human Y centromere (Earnshaw et al., 1989), and CENP-B knock-out mice are viable (Hudson et al., 1998; Kapoor et al., 1998; Perez-Castro et al., 1998). Yet we have extensively demonstrated that CENP-B bound to the centromeric DNA plays an important role in

the maintenance of chromosome segregation fidelity (Fachinetti et al., 2015; Hoffmann et al., 2016) by counteracting chromosome specific aneuploidies (Dumont, Gamba et al. 2020) and providing a memory for the de-novo CENP-A deposition through its interaction with CENP-C (Hoffmann et al., 2020). Moreover, human artificial chromosomes (HACs) are only efficiently established and maintained when the centromeric DNA contains CENP-B boxes, indicating that CENP-B plays an important role in the formation of centromeres (Ohzeki et al., 2002; Okada et al., 2007; reviewed in Ohzeki et al., 2019). Recently the combination of centromeric DNA repeats with and without CENP-B boxes significantly enhanced HAC formation (Okazaki et al., 2022), matching what is observed in natural chromosomes where some repeats have boxes and others do not.

CENP-B has also been implicated in the establishment and maintenance of the centrochromatin structure. The resolution of a crystal structure of CENP-B DBD bound to the CENP-B box revealed that CENP-B binding introduces a 59º bend in the DNA (Tanaka et al., 2001). CENP-B also was seen to contribute to the nucleosome positioning during the in vitro assembly of histones on centromeric DNA (Yoda et al., 1998). Moreover, by inducing a translational positioning of the nucleosome, *in vitro* CENP-B can bind the DNA even when the CENP-B box is wrapped on the nucleosome core particle (Tanaka et al., 2005). This is probably contributing to the observation that binding of CENP-B to the CENP-B box improves the naturally occurring phasing of nucleosomes along the centromeric DNA, generating an asymmetric unwrapping of CENP-A nucleosome terminal DNA, and placing the CENP-B box immediately 5' adjacent to the entry-exit site (Hasson et al., 2013). Since CENP-B can form homodimers, it was hypothesised early on, and some evidence of it was provided, that it could bring together two distant CENP-B boxes (Yoda et al., 1992, 1998; Tawaramoto et al., 2003). We have recently shown that indeed CENP-B favours the compaction of the centromeric chromatin by forming loops and clustering centromeres, and that these CENP-B generated structural features favour centromere position and stability in mitosis (**Figure 13**) (Chardon et al., 2022). CENP-B is therefore considered to have evolved possibly as a stabilizing factor of a pre-existing centromeric function, being dispensable yet at the same time playing an important role (Gamba and Fachinetti, 2020).

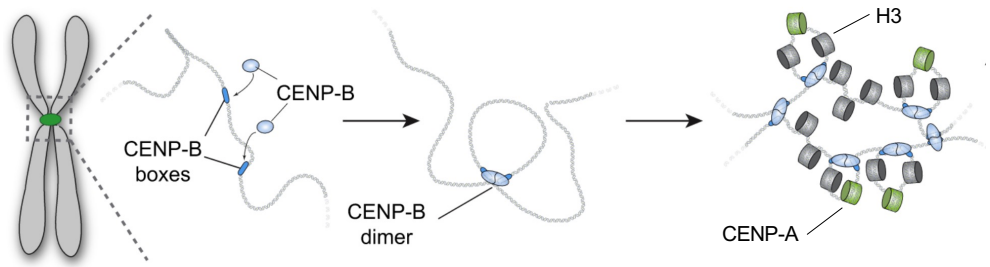

Figure 13. Model of centromeric DNA structure via CENP-B-mediated DNA loops.
Adapted from Chardon et al., 2022 [*Figure 7*]. CENP-B is represented as light blue spheres, CENP-B boxes are in blue, CENP-A nucleosomes in green and H3 nucleosomes in grey.

I.3. Centromeric DNA

II.3.1. The alpha satellite

The completion of the first human genome in 2001, after a decade of combined international efforts, came with the surprise that less than 5% of it coded for proteins (International Human Genome Sequencing Consortium et al., 2001). The recent telomere-to-telomere (T2T), the first complete human genome assembly (T2T-CHM13) filled almost all of the gaps from the first assembly, adding only a few more coding sequences that do not drastically alter this percentage (Nurk et al., 2022). At least half of the remaining 95% non-coding genome corresponds to repetitive DNA sequences (International Human Genome Sequencing Consortium et al., 2001). The existence and abundance of such repetitive DNA in the genomes of multiple higher organisms had been reported more than 30 years prior to the human genome assembly by Britten & Kohne. The premise of their work was that once denatured, the two strands from a unique genomic sequence would take very long to reanneal given how diluted they would be. By measuring the reassociation rate of denatured genomic DNA, they distinguished a fraction of highly repetitive DNA that reassociated up to 10^6 times faster than the unique sequences, and a fraction of moderately repetitive DNA reassociating at a moderate rate (Britten and Kohne, 1968). Combined, these two fractions made up about 50% of the total genomic DNA and at the time, it was also a surprising discovery. Repetitive DNA can be classified as interspersed (lacking higher order structure and scattered throughout the genome), or tandem, when clusters of individual sequence units are adjacent to one another and organized either as direct repeats (head-to-tail) or inverted repeats (head-to-head and tail-to-tail) (McNulty and Sullivan, 2018). One type of tandem repetitive elements is the satellite DNA.

The use of the term "satellite" stems from the original identification of these repeats as a satellite (secondary) band when genomic mouse DNA was separated along caesium chloride density gradients (Kit, 1961). According to its average base composition, AT-rich DNA migrates above and GC-rich below the main band of the bulk genome; therefore, sequences that have a different enough composition from the average bulk genome can be identified by this method. The generation of radioactive RNA probes translated in vitro from mouse satellite DNA and their in situ hybridization to spread chromosomes allowed to visualize for the first time in 1970 that the mouse satellite DNA was localized at the centromeres (Jones, 1970; Pardue and Gall, 1970). Human satellite DNA also began to be intensely studied, and the addition of silver or mercury to the caesium chloride gradients allowed to further separate several satellite bands, corresponding to human satellites I, II and III (Corneo et al., 1971, 1970, 1967). These three bands combined accounted for <2% of the total genome, while the estimate abundance of satellite DNA was closer to 10%. This led Laura Manuelidis, a physician from Yale School of Medicine, to try a different approach and use the recently discovered bacterial restriction endonucleases "*in order to investigate the possibility of other major repeated sequences in the human genome*", reasoning that any repeated sequence that contained restriction sites

would be digested periodically (Manuelidis, 1976). She found that the digestion of human genomic DNA with HaeIII generated a series of discrete bands, all multiples of the smaller size band, and which according to her estimates represented 2% of the genome. She further found a considerable enrichment of the 1-mer (170 bp) and 2-mer (340 bp) fragments in rapidly renaturing DNA, and found that digestion with EcoRI generated the same 340 bp 2-mer (Manuelidis, 1976). Later on, she would demonstrate through in situ hybridization that the radiolabelled EcoRI 2-mer would hybridize with the centromeres of some human chromosomes (Manuelidis, 1978). She would then sequence the 2-mer, revealing that it was composed by two tandem 171 bp A/T rich repeats that bared high sequence homology to the 172 bp AGMr(HindIII)-1 fragment from the African green monkey previously described and sequenced (Manuelidis and Wu, 1978). The analysis of the purified and rapidly reannealing fraction of total monkey DNA, named "α component", indicated that it is was largely composed of AGMr(Hind III)-1 and its multimers (Rosenberg et al., 1978). The primate-specific repeats would be therefore named α-DNA, or α-satellite monomers. They can be repeated thousands of times, and collectively are referred to as α-satellite DNA, or more loosely as centromeric DNA. Sequence variation between monomers led to the proposal, and posterior refinement, of a consensus human α-satellite monomer sequence (Waye and Willard, 1987; Vissel and Choo, 1987; Choo et al., 1991). It was quickly understood however that there is no real consensus α-satellite DNA and the α-satellite DNA from individual chromosomes widely differs from one to another (Willard, 1985). Only recently the T2T genome sequence revealed, with single-base resolution, the exact sequence of the centromeric DNA of each chromosome, and confirmed that α-satellites are the most abundant satellite repeat family of the genome, accounting for 2.8% of the total bases (Altemose et al., 2022a), a close value to the estimated 2% from Laura Manuelidis' 1976 paper.

The α-satellite DNA exhibits a higher order repeat (HOR) pattern of direct (head-to-tail) tandem repeats (Willard and Waye, 1987). The T2T genome has provided the most precise characterization of the centromere composition to date. Between 2 and 34 α-satellite monomers form each individual HOR, which can be up to 5.8 kbp long (Altemose et al., 2022a) (**Figure 14**). A given HOR is repeated hundreds to thousands of times to create homogenous HOR arrays spanning up to 5 Mbp and in which all the HORs are 97–100% identical. More than 1000 different α-satellite monomers, which are categorized in 20 different suprachromosomal families, form 80 different types of HOR arrays (Altemose et al., 2022a). While adjacent monomers on one HOR may share no more than 60%-80% homology, corresponding monomers from adjacent HORs within an array share close to 100% identity (Miga, 2017a) (**Figure 14**). A chromosome can have up to 9 different HOR arrays, but only one type of array, usually the largest, is enriched in CENP-A nucleosomes. It is therefore designated as the active array, as it is where the kinetochore machinery will be assembled (Altemose et al., 2022a). However, the same HOR will not necessarily be active on both chromosomes from a pair, giving rise to functionally heterozygous centromeric epialleles which further complexify the study of the centromeric function (Maloney et al., 2012).

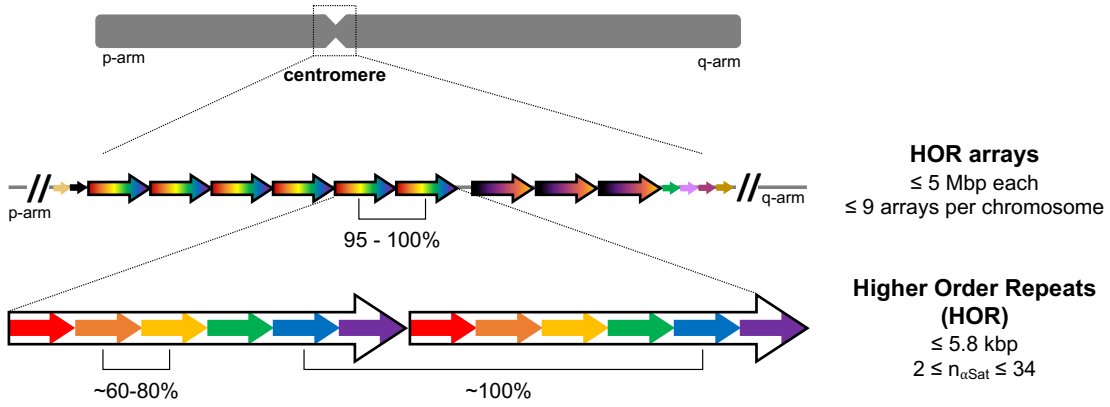

Figure 14. Simplified model of human centromeric DNA organization.
Human centromeric DNA regions are composed of up to 9 higher order repeat (HOR) arrays of α-satellite monomers. Two HORs are represented here, one in rainbow colours, the other in magma colours. The rainbow array is larger and each HOR in it is composed of six different monomers (red through purple), organized in direct tandem. Two adjacent 171 bp α-satellite monomers (orange and yellow in this example) can share as little as 60-80% homology, but the two blue monomers on adjacent HORs are close to 100% identical, making adjacent HORs from the same array more than ~95% to 100% identical. Monomeric and divergent α-satellite monomers are found directly adjacent to the HOR arrays. Not depicted are structural rearrangement, inversions, and/or insertions of transposable elements within the HOR arrays.

Although the sequence of the satellite DNA is poorly conserved between different chromosomes of the same organism, and more so across phylogeny, the centromere protein network is assembled in most eukaryotes preferentially onto this type of highly repetitive DNA sequences. A few notable exceptions to this rule are found in chickens, the *Equus* genus (horses, donkeys, zebras et al.) and in *Saccharomyces cerevisiae*. Chickens have 39 chromosomes, and 3 of them (5, 27, and Z, which happen to be evolutionary newer) are assembled into short interspersed repetitive sequences that do not contain tandem repetitive elements and span only 30 kbp (Shang et al., 2010). In the *Equus* genus several centromeres, including some described evolutionary new centromeres, have also been reported to be devoid of repetitive satellite DNA (Piras et al., 2010). In stark contrast to the regional centromeres of most higher eukaryotes, early on it was described that all 17 chromosomes of *Saccharomyces cerevisiae* have point centromeres, defined by a ~200-bp nuclease resistant region that encompasses ~125-bp of centromere determining elements (Bloom and Carbon, 1982). This sequence alone was seen to be enough to assemble a functional centromere.

The centromeres of humans and other higher eukaryotes were initially thought of as bigger, more complex versions of the yeast centromere, and were readily defined as the genomic loci with α-satellite DNA. However, in 1993 a supernumerary chromosome with a functional centromere that was not assembled onto α-satellite DNA was described in the context of human disease (Voullaire et al., 1993). This neocentromere had a positive staining with ACA, but not with anti-CENP- B antibodies. Following that first discovery, many more –rare but naturally occurring– neocentromeres have been

described (reviewed in Amor and Choo, 2002; Marshall et al., 2008a; Scott and Sullivan, 2014), challenging the idea that centromeres are genetically defined. Since centromere identity could be sequence-independent, then it had to be specified by epigenetic factors. With time, it became widely accepted that centromeres, defined as stated at the beginning of this chapter from a functional point of view as the sites of kinetochore assembly, are primarily defined epigenetically by CENP-A (Karpen and Allshire, 1997; Warburton et al., 1997). If repetitive DNA is not necessary for the centromere function, then why all human chromosomes and most eukaryotes have them? Hypotheses include the preservation of a particular chromosomal architecture and the promotion of the formation of the pericentromeric heterochromatin, both important factors from a structural standpoint to favour the centromeric function (Dumont and Fachinetti, 2017).

The chromosome segregation machinery is overall well conserved among all eukaryotes. However centromeric repeats are the most rapidly evolving DNA sequences in eukaryotic genomes, differing even between closely related species in what was enunciated more than two decades ago as the centromere paradox (Henikoff et al., 2001). The T2T sequence provided genome-wide evidence of recent layered expansions in centromeric α-satellite arrays, and confirmed that the most recently expanded repeats are more likely to interact with CENP-A (Altemose et al., 2022a), an observation that had been made prior to the T2T era (Henikoff et al., 2015). The genetic and epigenetic fates of centromeres seem therefore to be linked through evolution, and the centromere paradox stands stronger than ever.

II.3.2. The CENP-B box

Only one centromeric DNA sequence is conserved in the α-satellite DNA from primates and the murine centromeric satellite DNA: the CENP-B box, binding site of CENP-B (**Figure 15**). The 17-bp motif, and its 9 essential nucleotides, was first discovered on the human α-satellite DNA as the recognition site for CENP-B binding (Masumoto et al., 1989). With a single nucleotide difference, the whole motif was identified on the consensus sequence of *Mus musculus* minor satellite (Masumoto et al., 1989), and with further divergence on its distant relative the Asian mouse *Mus caroli*, which does not have minor satellites (Kipling et al., 1995). A variation of the CENP-B box was later found on the African green monkey (AGM) genome with the 9 essential nucleotides conserved despite having overall close to 3 orders of magnitude fewer CENP-B boxes than human (Goldberg et al., 1996; Yoda et al., 1996). As mentioned on section II.2.3, CENP-B, CENP-B-like proteins, CENP-B boxes and CENP-B box-like motifs have arisen independently multiple times in evolution, and these four species are only a few examples of it (Drinnenberg et al., 2016; Dumont and Fachinetti, 2017; Gamba and Fachinetti, 2020).

		satellite monomer size(bp)
Human	YTTCGTTGGAARCGGGA	171
African Green Monkey	TTTCGTTCAAAACGGGA	172
Mus musculus	ATTCGTTGGAAACGGGA	120
Mus caroli	TTTCGTCTAATGCGGGT	79
core	**** * ****	

Figure 15. Alignment of the CENP-B box sequence in two primates and two rodents.
The 10 conserved nucleotides are in blue/red and. The 9 core nucleotides, essential for CENP-B binding are highlighted with the consensus asterisks (*). Note that wo CpG dinucleotides (in red) are present among the essential nucleotides. Y: pyrimidine (C/T) R: purine (A/G).

In human, CENP-B boxes are found within type B but not type A α-satellite monomers. Therefore, depending on the monomeric composition of each HOR, CENP-B boxes may be regularly interspaced, clustered, abundant, scarce or even absent from any given HOR (McNulty and Sullivan, 2018). CENP-B boxes (and CENP-B) are detected, however, on all human chromosomes except for the Y (Masumoto et al., 1989), where CENP-B does not bind (Earnshaw et al., 1987). A higher density of CENP-B boxes within a given HOR has been correlated with stronger CENP-A enrichment and overall stabilization of CENP-A/B/C binding to the HOR (Thakur and Henikoff, 2018). The most recently expanded and CENP-A rich α-satellite dimeric arrays are the ones that contain the highest density of CENP-B boxes, with one CENP-B box per α-satellite dimer on average (Henikoff et al., 2015).

The consensus sequence of the A/T rich α-satellite monomer (Choo et al., 1991) presents only three CpG dinucleotides, with two of them being among the 9 conserved nucleotides of the CENP-B box sequence, highlighted in red on **Figure 15**. CpG dinucleotides are susceptible of being modified through methylation of the cytosine (see Chapter III), so this particularity of the CENP-B box did not go unnoticed and early on it was hypothesized that methylation could play a role in the regulation of CENP-B binding to the CENP-B box. Mouse cells grown in the presence of the DNA methyltransferase inhibitor, 5-aza-2'-deoxycytidine, exhibit a redistribution of CENP-B, indicating that the CpG methylation possibly affects the CENP-B binding to the CENP-B box (Mitchell et al., 1996). Further biochemical analyses demonstrated that CENP-B preferentially binds to the unmethylated CENP-B boxes, and that the affinity of CENP-B for the CENP-B box is reduced nearly to the level of nonspecific DNA binding by CpG methylation (Tanaka et al., 2004). These in vitro results have been recently supported by the single-molecule chromatin fibre sequencing of CHM13 cells. This technique employs a non-specific bacterial adenine methyltransferase to mark all accessible chromatin with N6-mdA, a non-endogenous DNA modification in human cells. The preferential protection from the mark of the unmethylated CENP-B boxes allowed to infer that they are robustly occupied by CENP-B, while 5mC methylation at even one of the two CpGs of the box made it readily accessible to the incorporation of N6-mdA (Dubocanin et al., 2023).

II.3.3. Transcription of the centromeric DNA

Historically the centromeric chromatin was considered to be strictly heterochromatic and therefore transcriptionally inactive. The fact that the centromeres were actually characterized by a complex mixture of chromatin marks that can support transcription (see section II.2.1.C on histone modifications), that they actually are transcribed, and that the transcripts seem to play important functional roles, was therefore initially surprising (reviews on the topic: Chan and Wong, 2012; Rošić and Erhardt, 2016; Duda et al., 2017; Perea-Resa and Blower, 2018; Corless et al., 2020).

Since the late 2000's it has become clear that, as it had been previously described in budding yeast and rice, the native human centromere repeats are transcribed (Wong et al., 2007; McNulty et al., 2017; Bury et al., 2020), as are the neocentromeres [centromeres that form at a chromosomal location other than on centromeric α-satellite DNA repeats] (Saffery et al., 2003; Chueh et al., 2009; Naughton et al., 2022) and the centromeres of human artificial chromosomes (HACs) (Bergmann et al., 2012, 2011; Molina et al., 2016). Both native satellite DNA transcripts and transcripts from retroelements have been described to emanate from mammalian centromeres (Carone et al., 2009; Chueh et al., 2009). In humans, the centromeric DNA is transcribed into non-coding RNAs (ncRNAs) that are chromosome-specific and alpha satellite array-specific (McNulty et al., 2017). Centromeric ncRNAs produced from active arrays were reported to be complexed with CENP-A and CENP-C, while transcripts arising from inactive-arrays associate preferentially with CENP-B and were generally less stable. Loss of CENP-A does not affect transcript abundance or stability, but the depletion of array-specific RNAs was reported to reduce CENP-A and CENP-C at the targeted centromere (McNulty et al., 2017). This evidence complements a long list of studies on human cells (Wong et al., 2007; Chan et al., 2012; Ideue et al., 2014), mouse cells (Ferri et al., 2009), *Xenopus* egg extract (Blower, 2016; Grenfell et al., 2016), *Drosophila* (Bobkov et al., 2018; Rošić et al., 2014) and *C. elegans* (J. Zhu et al., 2018), all stating that that transcription of the (peri)centromeric repeats is required for centromere formation and function.

Epigenetic editing on synthetic human artificial chromosomes (HACs) allowed to directly interrogate the role of transcription on centromere function. Depletion of H3K4me2, a characteristic histone modification mark of actively transcribed regions, reduced the levels of local transcription, and as a consequence, targeting of HJURP and CENP-A deposition were impaired and the HAC was gradually inactivated (Bergmann et al., 2011). A mild transcriptional activation did not have major effects on the HAC stability, while the over-activation of the transcription also rapidly inactivated the HAC centromere by causing a loss of pre-assembled CENP-A and abrogating de novo CENP-A deposition (Bergmann et al., 2012).

A recent study on the formation of neocentromeres has revealed that centromere repositioning is accompanied by RNAPII recruitment to the new locus, active transcription and spreading of repressive histone modifications to the neighbouring regions (Naughton et al., 2022). The authors propose that transcription opens the chromatin, and this is crucial to allow the centromere/kinetochore assembly. A key role of RNAPII had already been described at the native centromeres, where inhibition of RNAPII activity during mitosis led to an increase in lagging chromosomes and reduced CENP-C levels (Chan et al., 2012). Inhibition of RNAPII-dependent transcription in G1 was also seen to abrogate the recruitment of CENP-A and its chaperone HJURP to native human centromeres, in a study where a 1.3 kb long transcript was identified to physically interact with the CENP-A/HJURP pre-nucleosomal complex (Quénet and Dalal, 2014). In *Drosophila* a similar report said that the deposition of CENP-A is coupled with transcription and is dependent on the recruitment of RNAPII to the centromeres (Chen et al., 2015).

III. DNA methylation

The central dogma of molecular biology (Crick, 1957) states that the expression of the genome -all the genetic information of an organism- gives rise to the phenotype -the set of observable characteristics or traits of said organism- through the transcription of DNA into RNA and the subsequent translation of RNA into proteins. This gene-centric dogma is only partially accurate; one genome can be expressed in many ways, for example in different cell types or tissues from a same organism. Between genome and phenotype, is the epigenome, which refers to characteristics or gene expression patterns that do not result from changes in the DNA sequence itself, that can be inherited (but not necessarily) and that can be influenced by the environment. It was Conrad Hal Waddington, an English embryologist and geneticist who coined the term "epigenotype" in 1942 (Waddington 1942). Waddington is considered the father of epigenetics -the study of the epigenome- and is most remembered for his drawing of the epigenetic landscape (Waddington 1957), a visual metaphor representing the process of cellular development where different cell fates or outcomes (phenotypes) can be attained from the same starting point (genotype).

As of today, three main types of epigenetic information are known: DNA modifications (methylation and probably also hydroxymethylation), histone PTMs (see section I.4.2.A), and noncoding RNA action, which can often crosstalk and form intricate control networks. These epigenetic modifications are key regulators of the genome, influencing the association of proteins to the DNA and controlling important biological processes such as chromatin dynamics and gene expression patterns. The fact that the epigenetic information can and in many cases is heritable highlights its essential role. On the next sections I will focus on DNA methylation, the major form of DNA modification and one of the most studied epigenetic marks.

III.1. 5mC, "the fifth base"

DNA methylation (DNAme) refers to the covalent addition of a methyl group to the fifth carbon on the cytosine ring to form 5-methyl cytosine (5mC). The discovery of the natural existence of a modified cytosine occurred through the microscopic observation of hydrolysed nucleic acids crystals from *Mycobacterium tuberculosis* (Johnson and Coghill, 1925). Only twenty-three years later was this modification also described in a preparation of calf thymus resolved through paper chromatography (Hotchkiss, 1948). Hotchkiss observed in his chromatograms a fraction that separated from cytosine in a similar way as thymine (another name for 5-methyl uracil) separates from uracil. He therefore hypothesized that this fraction corresponded to 5mC. A couple of years later, this finding was generalized to other mammals (beef and ram), fish (herring sperm) and plants (wheat germ) (Wyatt, 1950). Given the historical context of these findings, and that the first reported amounts of 5mC

were so minute, it doesn't come as a surprise that almost thirty years had to pass before the role of 5mC as a master epigenetic factor was uncovered.

The origins of 5mC can be traced back to bacteria and therefore it was probably present in the first eukaryote. It is believed to have been present in the last common ancestor of plants, animals and fungi (Zemach et al., 2010), despite being lost nowadays in several eukaryotic lineages, including the common model organisms *C. elegans*, *S. cerevisiae* and *S. pombe*. DNAme was detected in the early seventies in different cell types of many animals such as echinoderms, cartilaginous and bony fish, amphibians, reptiles, birds, and mammals (Vanyushin et al., 1973). While both GC and 5mC content were seen to differ between the different specimens, they were often similar between closely related species. Genome wide DNAme levels range from very low in arthropods (and more recently also detected in *D. melanogaster,* which for a long time was thought to be devoid of DNAme), intermediate levels in other invertebrates, to high and very high levels in vertebrates (Bird, 1980).

5mC is the major form of DNA modification which is why it is often referred to as "the fifth base", representing 1% of the total nucleic acids in the human genome (Ehrlich et al., 1982). The methylation of cytosines was found to occur in animals almost exclusively in the context of the symmetrical CpG dinucleotides (Doskočil and Šorm, 1962; Sinsheimer et al., 1954). In mammals, and depending on the species, between 50% and 80% of all the CpGs are methylated (Gruenbaum et al., 1981). Early on it was noted that CpG dinucleotides are generally found in vertebrate DNA at one fifth of the expected frequency (Swartz et al., 1962). This is most likely explained by the hyper-mutability of the 5mCpG dinucleotide. Cytosines can be deaminated into uracils through the action of the activation-induced cytidine deaminase (AID; not to confuse the acronym with that of the auxin inducible degron). The resulting uracil sites are readily recognised as non-DNA and reliably corrected. On the other hand, the deamination of 5mC through the action of the apolipoprotein B RNA-editing catalytic component (APOBEC3A) generates a thymine, causing G/T mismatches that have to be mended by base-excision repair (BER) pathway and can often lead to transitions (Cooper et al., 2010). Organisms with the most extreme CpG deficiencies (vertebrates) also have the highest levels of DNA methylation and conversely, poorly methylated genomes display no significant CpG deficiencies (Bird, 1980).

III.2. DNA methylation (noun): a key epigenetic factor

"We are aware that no direct evidence exists for specific modification enzymes in eukaryotes, let alone that such enzymes might exercise control of gene activities. Nevertheless, in view of our almost complete ignorance of the mechanism for the unfolding of the genetic program during development, it seems justifiable to suggest speculative hypotheses that may lead to meaningful experimental approaches, particularly when these hypotheses are based on some of the known features of modification systems in bacteria." (Holliday and Pugh, 1975)

Throughout the sixties, DNAme was mostly studied in bacteria, where three different types of DNA methylation exist: N6-methyladenine (6mA), which is the most prevalent form of bacterial DNA modification, N4-methylcytosine (4mC), and 5-methylcytosine (5mC). These modifications were seen to play a central role in the Restriction and Modification (R-M) system of defence against foreign DNA, inhibiting the action of the restriction endonucleases in the methylated self-DNA (Arber, 1965; reviewed in: Boyer, 1971). With this evidence at the forefront, two groups in parallel speculated by the mid-seventies that similar mechanisms could exist in eukaryotes, where proteins could distinguish covalently modified nucleotides (5mC) from unmodified ones, and, for instance, regulate gene expression, act as developmental clocks (Holliday and Pugh, 1975), or be the key for the inactivation of the X chromosome (Riggs, 1975). Both groups also reasoned that since the CpG dinucleotide is symmetrical, patterns of methylated an unmethylated DNA could be copied semi-conservatively with replication, just as the DNA sequence itself. This was demonstrated three years later with the help of none other than restriction enzymes (Bird, 1978). DNAme is an essential mark that plays diverse and important roles throughout the genome, some of which I will discuss in the following sub-sections.

III.2.1 Regulation of transcription

Similar to the experiments that led to the identification of the centromeric α-satellite DNA (see section II.3.1), the use of restriction enzymes led to the discovery of one of the most striking features of the DNAme distribution in vertebrates: the existance of CpG islands (CGIs). The digestion of mouse genomic DNA with HpaII (recognition site: C^CGG) and HhaI (recognition site: GCG^C), two enzymes that are blocked by DNAme, uncovered "islands of DNA" that were heavily digested in fragments ranging from few bp up to 600 bp, with an average size of 120 bp. The mapping of some of the larger fragments revealed that the CGIs are on average ~1000 bp long (500 – 2000 bp range) and that, unlike the bulk of the genome, they are not deficient on CpG dinucleotides, a feature directly related to their lack of methylation. Importantly, in spite of their similarities CGIs do not constitute a family of defined sequence (Bird et al., 1985). The distribution of 5mC across the genome of vertebrates is therefore bimodal: the bulk of the genome is CpG poor and heavily methylated, whereas CGIs have expected CpG densities and low methylation levels. This is in sharp contrast to the methylation landscape of invertebrates, where methylation is globally low, and present as a mosaic of alternating methylated and nonmethylated regions. The transition from mosaic to global

methylation is actually considered to be evolutionary close to the origin of the vertebrates, and may have played a role or be a consequence of the separation of both subphyli (Tweedie et al., 1997).

CGIs were soon localized the 5' end of genes, and evidence that CGI methylation inhibited transcription, quickly piled up (reviewed in: Bird, 1986). The transcriptional repression mediated by DNAme can be due to a direct inhibition of transcription factor binding or through the recruitment of methyl-binding proteins (see section II.3.2) which can attract chromatin remodelling complexes that, depending on the context, can cause transcriptional repression (Nan et al., 1998). More than half of all CGIs contain transcription start sites and coincide with promoters of annotated genes which are therefore referred to as CGI promoters. The rest are either within (intragenic) or between (intergenic) known transctiptional units, and evidence shows that many are also sites of transcriptional initiation or regulation (reviewed in: Deaton and Bird, 2011). CGI promoters are considered transcriptionally permissive, and display dispersed transcription initiation patterns. Studies on isolated CGI chromatin revealed high levels of histones H3 and H4 acetylation, low amounts of linker histone H1 and some nucleosome-free regions (Tazi and Bird, 1990), all features of active chromatin. The tri-methylation of lysine 4 of histone H3 (H3K4me3) is a signature histone mark of the CGI promoters (Mikkelsen et al., 2007) and the deposition of this mark has been demonstrated to be dependent on the recognition of unmethylated CpGs by Cfp1, a member of the Setd1A and Setd1B H3K4 methyltransferase complexes (Thomson et al., 2010). This is a prime example of the epigenetic crosstalk between DNA methylation and histone modifications.

III.2.2. Silencing of transposable elements

In the early nineties, the exponential progress made in DNA sequencing technologies and the development of computational tools to analyse these sequences unveiled that transposable elements (TEs) represent about 35% of the human genome, a strikingly high number compared to the mere 5% of gene coding sequences (reviewed in Smit, 1996). The insertion of these DNA sequences in the genome can occur either via reverse transcription of an RNA intermediate (retrotransposition), or by excision and reintegration of the DNA itself (transposition). Retrotransposed repeats are classified as short interspersed nuclear elements (SINEs, 80 – 300 bp), long interspersed nuclear elements (LINEs 6 – 8 kbp), both of which are well defined elements present in all eukaryotes and that account for the biggest fraction of human interspersed repeats, or retrovirus-like elements containing long terminal repeats (1.5 – 10 kbp) (Smit, 1996). The long interspersed element 1 (LINE-1) is the only protein-coding retrotransposon in humans. It is transcribed as a bicistronic poly(A) RNA that encodes an RNA-binding protein [open reading frame 1 protein (ORF1p)] and an endonuclease and reverse transcriptase [ORF2p] (Mathias et al., 1991; Feng et al., 1996). LINE-1 elements therefore encode all the necessary proteins for their independent retrotransposition, while SINEs require factors produced by other retrotransposons for their integration. The human genome contains

approximately 10^6 Alu repeats –the most abundant of all SINEs–, meaning that on average there is one Alu element every 3 kbp of DNA. There are about 10^5 copies of LINE-1, equivalent one element every 30 kbp. The (retro)transposition of these parasitic elements needs to be tightly controlled since their integration can disrupt the expression genes (insertional mutagenesis), induce homology-directed genomic rearrangements, or form deleterious chimeric mRNAs. Despite their abundance, these sequences generally cause surprisingly little harm. Many are degenerate enough to be inactive, and the rest are controlled by means of epigenetic silencing, mainly through DNA methylation (reviewed in Yoder et al., 1997). The introduction of Moloney murine leukaemia retrovirus (M-MLV) into mouse zygotes revealed that at the pre-implantation stage, the retroviral genome became methylated de novo, and its expression was blocked (Jähner et al., 1982). In somatic cells, Alu and LINE-1 elements were seen early on to be methylated at virtually all testable Hpall sites (Sanford et al., 1987). A more recent study revealed that ~73% of Alu CpGs are deaminated, and 87% of the remaining CpG sites are methylated (Xiang et al., 2010). The promoter region of LINE-1 is a heavily methylated atypical CpG island, and it was reported that inhibition of DNA methylation increased the expression of LINE-1 (Woodcock et al., 1997). DNAme, however, is not the sole mechanism of TE repression. As for the human genome, 35% of the *Bombyx mori* genome is composed of TEs (Osanai-Futahashi et al., 2008); yet this lepidopteran also has very low 5mC levels, indicating that other mechanisms must exist to ensure TE control. One such mechanism in humans is the HUSH (human silencing hub) complex of proteins, which is recruited to LINE-1s located in transcriptionally permissive euchromatic environments to promote the deposition of H3K9me3 for targeted transcriptional silencing (Tchasovnikarova et al., 2015; Liu et al., 2018). Overall the silencing of transposable and repetitive elements is of great importance for the genome integrity, and their de-repression by DNA hypomethylation is associated with a countless human diseases (reviewed in Pappalardo and Barra, 2021).

III.2.3. Genomic Imprinting and X-Chromosome inactivation

Imprint (verb): to mark a surface by pressing something hard into it / to fix an event or experience so firmly in the memory that it cannot be forgotten.

Genomic imprinting, or "parent-of-origin gene expression", refers to the epigenetic modification that distinguishes maternal and paternal copies (alleles) of a given gene, leading to the differential expression of one specific parental allele in the somatic cells of the offspring (reviewed in Barlow and Bartolomei, 2014; Tucci et al., 2019). Genomic imprinting has only been reported in placental mammals and marsupials. Most of the genes subject to genomic imprinting code for factors regulating embryonic and neonatal growth so it is therefore likely that genomic imprinting evolved to play a specific role in mammalian reproduction (Kaneko-Ishino and Ishino, 2019).

For the vast majority of genes, both the maternal and the paternal alleles are equally expressed. In humans, 100 out of the estimated ~25.000 genes are imprinted (Court et al., 2014; Baran et al., 2015), meaning that they are expressed from only one parental copy while the other is silenced. Imprinted genes are usually found in clusters, allowing for their coordinated control by shared regulatory elements of the imprint control regions. The main cis-acting regulatory elements of the clusters are CpG rich regions which are methylated in only one of the two parental alleles and are therefore called differentially methylated regions (DMRs). While other epigenetic mechanisms such as histone modifications and lncRNA action also act as imprints, DNAme is the main mechanism by which the expression of imprinted genes is regulated (Delaval and Feil, 2004). The establishment of the differential imprinting patterns (i.e., differential methylation of the imprint control regions) occurs during the germline development in a sex-specific manner.

Mammalian females have two copies of the sex chromosome X, while males have only one. A dosage compensation is therefore required to ensure equal levels of X-linked gene expression in all individuals of the same species. This dosage compensation is achieved by randomly repressing one of the two X chromosomes in early female embryogenesis. The coat of the tortoiseshell female cats, a mosaic of the black and yellow homozygous colours, is the best visual exemplification of this phenomenon, where different cells will have one (black fur allele) or the other (yellow fur allele) X chromosome inactivated (Xi) (Lyon, 1961). The process of X chromosome inactivation (XCI) begins with the expression of the X inactive specific transcript (XIST), a lncRNA that specifically binds to and fully coats the Xi. XIST coating leads to the recruitment of chromatin modifiers, including the Polycomb group of proteins, which initiate the process of inactivation that is, at this stage, reversible. Throughout the development, the Xi accumulates a series of chromatin modifications that seal its irreversible inactivation, such as the deposition of the histone H2A variant macroH2A, the enrichment of heterochromatin histone modifications, and the de novo DNAme of promoters of repressed genes (Paro et al., 2021). DNAme is essential for a complete XCI, as its inhibition can lead to the reactivation of Xi silenced genes (Sado et al., 2000; Csankovszki et al., 2001). In the context of XCI, DNAme acts as a "lock" of the gene silencing that was initiated by other epigenetic factors.

III.2.4. Cellular differentiation and Genome Organization

DNAme is implicated in restricting the differentiation potential of stem cells and the loss of pluripotency upon differentiation involves the acquisition of DNA methylation at specific genetic loci (reviewed in Ehrlich and Lacey, 2013). The reprograming of somatic cells into induced pluripotent stem cells (iPSCs) using the Yamanaka factors (Takahashi and Yamanaka, 2006) can yield fully reprogrammed cells, with gene expression and epigenetic states that are highly similar to embryonic stem cells. The process can also yield partially reprogrammed cells that exhibit, among other features, DNA hypermethylation at pluripotency-related loci. DNAme acts as an important

barrier to the de-differentiation and its inhibition has been seen to improve the overall efficiency of the reprograming process (Mikkelsen et al., 2008). How DNAme is reset during the reprogramming process remains poorly understood. It has been observed that early passage male XY iPSCs have DNAme levels similar to the somatic cells they are derived from, while the reprogramming of female cells induces the reactivation of the silent X chromosome, and early passage female iPSCs display global DNA hypomethylation possibly for dosage-compensation (Janiszewski et al., 2018).

DNAme has also been reported to play an important role in the establishment and maintenance of genome architecture. The methylation of the DNA can change the dynamics and structure of mononucleosomes in vitro, leading to a more compact and rigid structure (Choy et al., 2010). In a recent publication, a molecular dynamics simulation revealed that DNAme causes pronounced changes in linker and nucleosomal DNA geometry, and that these changes were associated with enhanced interactions between the methylated DNA and the histone octamer (Li et al., 2022). The disruption of DNA methylation pathways has been seen to affect the replication timing and 3D genome organization (Du et al., 2021), to significantly remodel the genome compartmentalization (Spracklin et al., 2023), and cause changes in heterochromatin abundance and localization (Scelfo et al., 2023). It has also been well established that DNAme is important for chromatin condensation, as treatment of cells with demethylating agents was early on observed to induce uncoiling, decondensation and even "pulverisation" of chromosomes (Viegas-Péquignot and Dutrillaux, 1976). Finally, DNAme has also been implicated in the correct chromosomal compaction during mitosis, possibly by regulating the interactions between condensin, cohesin and the DNA (Flagiello et al., 2002) .

III.3. DNA methylation writers, readers and erasers

III.3.1. DNA methylation (verb) and the three writers of the fifth base

The DNA methylation patterns are established during development and cell differentiation, are mostly maintained thereafter (reviewed in Greenberg and Bourc'his, 2019). Early on it was observed that during the mouse embryo development, an initial period of loss of methylation is followed by a period of gain of methylation in what is referred to as DNA methylation reprogramming (Monk et al., 1987). The egg genome has low methylation levels while the sperm genome is normally methylated. After fertilization, there is a sharp global decrease in methylation that affects both genomes but more so the paternal, and following implantation, embryonic and extraembryonic lineages are progressively and independently re-methylated to different final extents (**Figure 16**). Data on human embryo development is scarce, but globally the epigenetic reprogramming seems to be equivalent to that in mouse, with some species-specificity. Recently, the single-cell methylome sequencing of human pre-implantation embryos revealed that human global DNA methylation

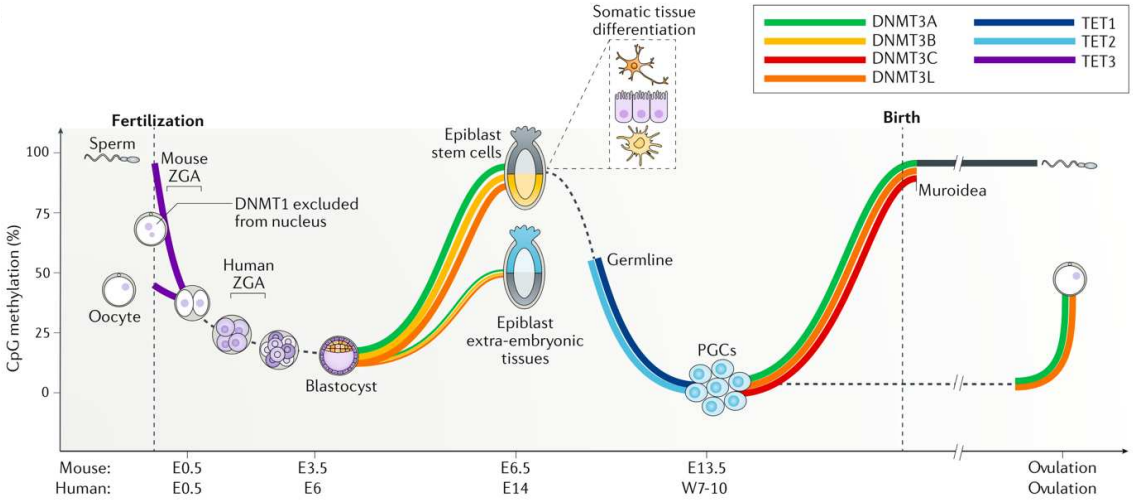

Figure 16. DNA methylation reprogramming during development.
Adapted from Greenberg & Bourc'his, 2019 [*Figure 4a*]. Embryonic and germline DNA methylation erasure and establishment in mouse and human. The methylcytosine dioxygenase TET3 is active in the fertilized zygote, leading to hydroxymethylation and active DNA demethylation. Following passive demethylation (dashed line), DNA methylation reaches a low point at the blastocyst stage, which is followed by DNMT3A and DNMT3B-mediated de novo DNA methylation after blastocyst implantation. DNMT3L is also expressed in this time window but is not absolutely required to methylate the embryonic genome. In extra-embryonic tissues, DNMT3A, DNMT3B and DNMT3L are expressed, but to a lesser extent than in the embryo and correlates with relative DNA hypomethylation. Post implantation, in the epiblast, a subset of stem cells is specified for the germline, where they undergo two waves of DNA demethylation: one passive and one mediated by TET1 and TET2. Male gametes become highly methylated before birth through the activity of DNMT3A and DNMT3L, and, in the case of Muroidea (the superfamily that includes mice and rats), DNMT3C. The oocyte gains methylation after meiosis and prior to ovulation through the activity of DNMT3A and DNMT3L in mice, and likely through DNMT3A in humans.

reprogramming is in fact a dynamic balance between strong genome-wide demethylation and focused milder re-methylation, with three waves of demethylation (sperm to early pronucleus (PN); late zygote to 2-cell; 8-cell to morula) interspersed by two waves of de novo DNA methylation (early male PN – mid PN; 4-cell to 8-cell) (P. Zhu et al., 2018). Imprint control regions robustly retain their parent-specific DNA methylation throughout the early development reprogramming. They are demethylated later in the primordial germ cells, allowing for a sex-specific re-establishment of the imprinting (Tucci et al., 2019). SINE and LINE repeats have been seen to retain (Guo et al., 2014) or gain de novo-methylation during the reprogramming (P. Zhu et al., 2018), likely as a means to keep them in a repressed state and avoid genome instability in the embryo. Overall methylation levels increase post-implantation along with the establishment of specific methylation patterns that define the cellular differentiation (Jones and Taylor, 1980).

The cloning and sequencing of bacterial methyltransferases allowed to elucidate the chemistry of the methylation reaction (Wu and Santi, 1987), where 5mC is formed after the covalent transfer of a

methyl moiety from S-adenosyl-L-methionine (SAM) to the fifth carbon on the cytosine ring. The search for the mammalian DNA methyltransferases (DNMTs) culminated with mouse *Dnmt1* being the first DNA methyl transferase cloned and sequenced (Bestor et al., 1988). The subsequent mutation of *Dnmt1* in mouse embryonic stem cells (ES) by homology recombination revealed a global loss of two-thirds of the original DNA methylation level (Li et al., 1992). DNMT1 is essential for development, and knock-out mice display growth delays, the disruption of the monoallelic expression of imprinted genes and lethality before embryonic day 11 (Li et al., 1992). The Cre-mediated conditional deletion of *Dnmt1* in primary embryonic mouse fibroblasts causes global demethylation, p53-dependent cell death, changes in gene expression of imprinted genes and the mobilization of retroelements (Jackson-Grusby et al., 2001). A few years later, the Cre recombinase–mediated disruption of DNMT1 in human colorectal carcinoma cell line HCT116 was described to lead to G2 arrest and subsequent slippage and mitotic catastrophe (Chen et al., 2007).

DNMT1 was early-on seen to have preferential affinity for hemimethylated CpGs (Pradhan et al., 1999), such as those generated after DNA replication, which is why it was immediately considered as the maintenance DNA methyl transferase. The function of DNMT1 was later seen to be tightly linked to that of UHRF1 (ubiquitin-like, with plant homeodomain (PHD) and really interesting new gene (RING) finger domain 1), a multifaceted protein that also shows preferential binding to hemimethylated CpGs and associates directly with DNMT1 (Bostick et al., 2007). Our current understanding of the molecular mechanisms for the maintenance of DNA methylation is that it can be achieved either through rapid replication-coupled events, or in a slower, replication-uncoupled manner in already re-chromatinized DNA (reviewed in Petryk et al., 2021). UHRF1 is an E3 ubiquitin ligase, and it is recruited to the replication fork by interacting with a methylated histone-like motif of LIG1 (LIG1K126me3) (Ferry et al., 2017), the DNA ligase responsible for joining the Okazaki fragments on the lagging strand. Alternatively, UHRF1 can bind H3K9me2/3 to ensure the post-replication DNAme maintenance, a process that is dependent on UHRF1's interaction with the chromatin remodeler HELLS/LSH (Han et al., 2020). Once recruited to the replication fork, UHRF1 mono-ubiquitinates both the tail of histone H3 (dual mono-ubiquitination H3Ub2 in lysines 14 and 18) and an H3-like motif present on PAF15 (PCNA [Proliferating Cell Nuclear Antigen] interacting factor 15), promoting its interaction with PCNA and the recruitment of the replicative DNA polymerase to the fork (Nishiyama et al., 2020). These UHRF1-mediated ubiquitinations are crucial for the recruitment of DNMT1 and for its activity. DNMT1 is conformationally inhibited, and its interactions with both H3Ub2 and PAF15Ub2 are required for relieving the auto-inhibition, therefore allowing it to exert its catalytic activity (Petryk et al., 2021).

The fact that *Dnmt1* KO ES cells retain about one third of residual CpG methylation (Li et al., 1992) and that, when infected with M-MLV, they are as capable as wild-type cells of methylating the newly integrated proviral DNA (Lei et al., 1996), prompted the informatic search for other protein(s) with

conserved methyltransferase domains that could be responsible for this de novo methylation activity. This search led to the identification of mouse and human DNMT3A and DNMT3B (Okano et al., 1998). Both proteins were seen to be highly expressed in ES cells, and at very low levels in differentiated embryoid bodies and adult tissues. It was also demonstrated that both proteins could methylate in vitro the cDNA of M-MLV and that they could equally methylate naked and hemi-methylated DNA substrates (Okano et al., 1998): the de novo DNA methyl transferases had been found. The subsequent gene inactivation of *Dnmt3a* and *Dnmt3b* in mice and ES cells revealed that both proteins are essential for mammalian development and are required for genome-wide de novo methylation (Okano et al., 1999) (**Figure 16**). *Dnmt3a*$^{-/-}$ mice develop to term but die at post-natal week 4. *Dnmt3b*$^{-/-}$ embryos develop normally until E9.5, when they start to accumulate developmental defects that make them unviable (Okano et al., 1999). Despite their overlapping functions in the de novo methylation of many genomic loci, there are also some functional differences between both DNMT3s. In particular, it was reported that only *Dnmt3b*$^{-/-}$ ES cells exhibit hypomethylation of the centromeric minor satellite repeats (Okano et al., 1999). Recently, a detailed structure-function characterization of both enzymes revealed that DNMT3A has a higher preference for methylating CpGs in the context CG(C/T) motif, whereas DNMT3B preferentially methylates in the context of CG(G/A) motifs (Gao et al., 2020); these substrate preferences extend to the 3 bp around the CpG dinucleotide, and explain why DNMT3B also preferentially methylates the human pericentromeric satellite 2 (HSAT2) at its consensus sequence (TCCATT**CG**ATGATG). A third member of the DNMT3 family, DNMT3L, was identified shortly after DNMT3A/B (Aapola et al., 2000). DNMT3L is catalytically dead, and mainly acts as a substrate exchange factor, accelerating the binding of DNMT3A/B to the DNA and to SAM, and overall stimulating their catalytic activities by ~15 fold (Gowher et al., 2005). The fourth and final member of the family known to date is DNMT3C, a male germ line specific protein responsible for methylating the promoters of evolutionarily young retrotransposons (Barau et al., 2016) (**Figure 16**). As abovementioned, the three catalytically active DNMTs (DNMT1, DNMT3A and DNMT3B) are essential for mouse development, yet they are dispensable in undifferentiated ES cells where the triple knock-out only causes minimal phenotypic changes (Tsumura et al., 2006).

The rigid separation between maintenance (DNMT1) and de novo (DNMT3) methyl transferase activities is a generalization of the function of these proteins. As mentioned, DNMT3A/B have equal affinity for naked and hemi-methylated DNA, and it was observed that DNMT1 deficient HCT116 were able to retain high methylation levels while replicating, implicating the DNMT3s in methylation maintenance (Rhee et al., 2002). The conditional deletion of DNMT3B, but not of DNMT3A, in mouse embryonic fibroblasts results in partial loss of DNAme throughout the genome, directly implicating DNMT3B in the DNAme maintenance (Dodge et al., 2005). The opposite is also true, with DNMT1 exerting de novo methyl transferase activity when overexpressed (Vertino et al., 1996) or in compensation to the absence of DNMT3B (Scelfo et al., 2023). It has also been reported that

DNMT1 can catalyze DNA methylation both de novo and in maintenance context with specific retrotransposon targeting, and this was proposed to provide additional stability for long-term repression throughout development (Haggerty et al., 2021).

III.3.2. DNA methylation readers

In eukaryotic cells, methyl-CpG binding proteins (MBPs) can recognize ("read") and specifically bind to methylated DNA. They often interact with other proteins and serve as scaffolds to organize effector recruitment at particular loci. MBPs are divided into three major families according to their methylated DNA recognition domain: (1) methyl-CpG binding domain (MBD), (2) C2H2 zinc finger domain, or (3) SET-and RING finger-associated (SRA) domain (Shimbo and Wade, 2016). Interestingly, some transcription factors seem to also have a sequence-dependent 5mCpG affinity. They possibly bind to methylated regulatory elements to activate gene expression, challenging the traditional view that DNA methylation equals transcriptional silencing (Zhu et al., 2016). These transcription factors can be considered as a new group of readers of DNAme without having dedicated methylated-DNA recognition domains.

Each MBP has its own DNA-binding preference, expression pattern and protein interactors, making them a diverse toolbox capable of reading the methylated DNA in different contexts and initiating diverse signaling pathways (reviewed in Fournier et al., 2012). MBPs are the bridge linking DNAme and histone modifications. For example, UHRF1 (a member of the SET and SRA domain family), binds to hemimethylated DNA and directly ubiquitinates the histone H3 tail. Another classical example is MeCP2, one of the best studied members of the MBD family, which has been seen to promote transcriptional repression through the recruitment of histone deacetylases (Nan et al., 1998). MeCP2 can also promote transcriptional activation through the recruitment of the transcriptional activator CREB1 (Chahrour et al., 2008). MBPs are involved in local and global chromatin organization, heterochromatin formation and maintenance, DNA repair, and are deregulated in disease (Fournier et al., 2012). Many MBPs have seen to be altered in cancer, and are considered as potential therapeutic targets (reviewd in Mahmood and Rabbani, 2019).

III.3.3. DNA demethylation (verb)

The passive loss of DNA methylation coupled with replication was envisioned as a possibility even before knowing exactly how methylation was established and maintained (Holliday and Pugh, 1975). Indeed, if the DNAme maintenance pathway is repressed, by absence, inhibition or mislocalization of DNMT1 or UHRF1, the methyl mark will not be copied in the newly replicated strand and therefore will be passively diluted in half with each cell cycle. This is the case when DNMT1 is inhibited by treatment with 5-azacytidine (5-AC) and 5-aza-2'-deoxycytidine (DAC or decitabine).

These drugs were developed as chemotherapeutic agents (Šorm et al., 1964) and only later were seen to inhibit DNA methylation and promote cellular differentiation (Jones and Taylor, 1980). These cytosine analogues are incorporated into the DNA (and RNA) during replication and are recognized as substrates by DNMT1. The attempt at methylating these modified bases covalently binds DNMT1 to the DNA, sequestering the protein and preventing the methylation of the newly formed DNA strand. The trapped DNMT1 then triggers DNA damage signalling which eventually leads to the degradation of the enzyme (Stresemann and Lyko, 2008). Despite their massive cytotoxic side effects, these drugs have been instrumental in the study of DNA methylation and its cellular functions, and up until recently were the only tools available for inhibiting the DNAme maintenance.

In the early 2000s, the immunostaining of 5mC in pre-implantation mouse embryos at different timepoints revealed that 6 hours after fertilization both the maternal and paternal pronuclei are methylated, while 8 hours after fertilization, the paternal pronucleus is notoriously and drastically demethylated (Mayer et al., 2000). This rapid (≤ 2 hours) demethylation occurred even if replication was inhibited by means of aphidicolin treatment, which clearly indicated that an active demethylation mechanism, independent of replication, must exist. Almost a decade had to pass before this mechanism would be unveiled. While the chemical deposition of the methyl group is a straightforward enzymatic reaction catalysed by the DNMTs, active demethylation is an intricate process (**Figure 17**) that involves the ten-eleven translocation (TET) family of enzymes (TET1, 2, 3 and their isoforms). The name of this family stems from the first identification of one of its members in a patient suffering from acute myeloid leukaemia, who cytogenetically presented a translocation between chromosomes ten and eleven. The Mixed Lineage Leukaemia (MLL) gene was seen to be fused to the then called "LCX" (leukaemia-associated protein with a CXXC domain) gene (Ono et al., 2002). The partner gene of MLL was later re-named TET1 in honour of the translocation that led to its identification, and was found to be a member of a novel, well-conserved protein family with at least two other members of unknown biological function (Lorsbach et al., 2003). Their function would remain undetermined for another six years.

A potential 5mC demethylation mechanism with similar chemistry as for thymine (5-methyl uracil) to uracil conversion had been envisioned for some time. This idea prompted the computational search for homologs of trypanosome JBP1 and JBP2, two proteins known to oxidize the 5-methyl group of thymine in the protozoan. The human TET1, TET2 and TET3 and their metazoan orthologs were all hits of this computational search (Tahiliani et al., 2009). It was then confirmed experimentally that the TET enzymes catalyse the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) in cultured cells and in vitro (Tahiliani et al., 2009). 5hmC was later demonstrated to be further oxidized, first into 5-formylcytosine (5fC) and then into 5-carboxylcytosine (5caC) (Ito et al., 2011). 5hmc, 5fC and 5caC can then be passively converted back to unmethylated C by dilution with

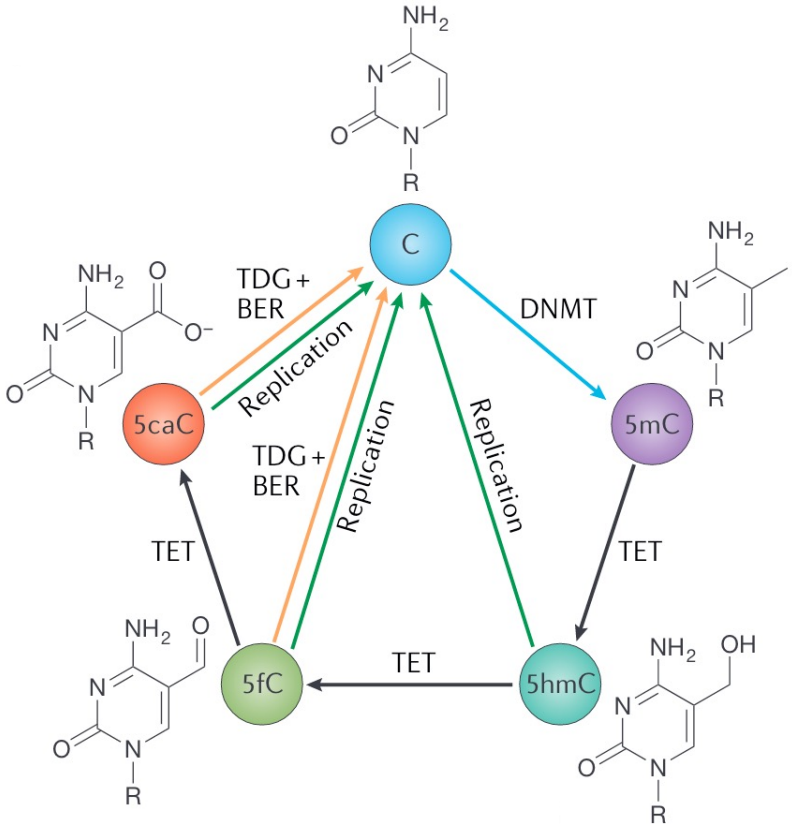

Figure 17. The cycle of DNA methylation and TET-mediated DNA demethylation.
Adapted from X. Wu & Zhang, 2017 [*Figure 1A*]. DNA methyltransferases (DNMTs) convert unmodified cytosine (C) to 5-methylcytosine (5mC). 5mC can be converted back to cytosine by TET-mediated oxidation, first to 5-hydroxymethylcytosine (5hmC), then to 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5caC). 5fC and 5caC can be actively converted to C (orange arrows) by action of thymine DNA glycosylase (TDG) coupled with base excision repair (BER). 5hmC, 5fC and 5caC can also be passively converted into C by replication-dependent dilution (green arrows).

the DNA replication (**Figure 17**), as they are not substrates of DNMT1. 5fC and 5caC resemble damaged bases and are substrates of thymine DNA-glycosylase (TDG) an enzyme that removes thymines from G/T mismatches (He et al., 2011). TDG action leaves an abasic site that is then corrected by the base-excision repair (BER) pathway, which restores the site with a cytosine, effectively completing the process of active demethylation (Jacobs and Schär, 2012). Unlike 5mC, which is always a symmetrical mark, the oxidized cytosines are deposited with a strong strand asymmetry and the TET proteins have been seen to have different affinities for their substrates, leading to the preferential accumulation of 5hmC over 5fC and 5caC (Wu et al., 2014). Evidence shows that 5hmc is not just an oxidation intermediate, but rather a stable epigenetic mark with regulatory functions of its own (Bachman et al., 2014). The specific distribution pattern of 5hmC in promoters, enhancers, and gene bodies of transcriptionally active genes, points towards specific biological functions (Wu et al., 2011). Mounting evidence is making 5hmC to be considered as the "sixth base" (Rusk, 2012).

High levels of TET1 and TET2, but not of TET3, are detected in mouse primordial germ cells and have been demonstrated to be crucial for the germ-line specific DNA demethylation, the sole process of imprint erasure (Hackett et al., 2013) (**Figure 17**). In somatic cells, TET1 is also considered a maintenance DNA demethylase: it has been seen to specifically localize at the edge of the CGIs where it accumulates 5hmC and forms a boundary, preventing methylation spreading from outside the CGI inwards (Jin et al., 2014). TET3 levels are elevated in oocytes and zygotes, and although replication-dependent dilution is the major contributor to demethylation in this developmental step, TET3 plays an important role in the first wave of methylation reprogramming after fecundation (Shen et al., 2014) (**Figure 16**). Mice lacking individual TET proteins have relatively normal pre and post-implantation development, attesting to the redundant function of these proteins; *Tet3*$^{-/-}$ mice, however, die perinatally for unknown reasons.

III.4. Epigenome editing

Historically it has been hard to distinguish if a given epigenetic modification and a biological phenotype have a relationship of correlation or one of causality (Stricker et al., 2017). The use of epigenetic drugs or the genetic disruption of chromatin-modifying enzymes have been widely employed to try and address these questions, but distinguishing between direct and indirect effects of these genome-wide perturbations remains challenging. This need for specificity propelled the development of epigenome editing tools. Epigenome editing refers to the modification of the local epigenetic marks at specific genomic loci by the action of targeted effectors (reviewed in Stricker et al., 2017; Gjaltema and Rots, 2020). To edit the epigenome, a DNA recognition moiety such as zinc-fingers (ZF), transcription activator-like effector (TALE) or the catalytically inactive (dead) CRISPR associated protein 9 (dCas9), is used to direct an "EpiEffector", usually the catalytic domain or a full-length chromatin-modifying enzyme, to exert its activity at a genomic locus of interest. I will focus briefly on the editing of DNA methylation, but transcriptional activators and repressors, histone acetyltransferases, histone methyltransferases and histone demethylases have all been used for targeted epigenetic modifications (reviewed in Policarpi et al., 2021).

The first reported DNAme editing tool dates from 1997. Based on the premise that the methylation of few CpGs within a promoter is sufficient to repress its transcription, a bacterial methyltransferase (M.SssI) was fused to a zinc-finger directed to the p53 binding site on the p21 gene, and was shown to have promising results in methylating a target DNA in vitro. "*Further refinements of the targeted methylation approach promise to yield a new type of gene-therapy agent for the selective inactivation of vertebrate genes*" (Xu and Bestor, 1997). In the following years, other prokaryotic DNA methyltransferases would also be used, but the lack of strong specificity and the challenges in designing artificial ZFs limited their use. TALEs would be reported in 2009 as second-generation DNA binding moieties, simplifying the design and construction of the editing tools. However the biggest burst in epigenome editing would

come in 2016 with the widespread use of dCas9 fusion proteins (reviewed in Lei et al., 2018) and in the last decade many efforts have been made to understand the precise role of DNAme in transcriptional regulation. A TALE fused to the catalytic domain of DNMT3A and to DNMT3L was used to methylate the promoter of p16 and resulted in decreased gene expression (Bernstein et al., 2015). Both TET1 and DNMT3A were later fused to dCas9 for the targeted editing of DNA methylation in multiple methylated or unmethylated promoter sequences, causing their respective activation or silencing both in vitro and in vivo (Liu et al., 2016). The targeted methylation only affected ~60% of the CpGs, so different combinations of full length DNMT3A, its catalytic domain and the DNMT3L "co-factor" have been tested since (Gjaltema and Rots, 2020). An alternative route to improve the methylation rate was undertaken by going back to bacterial methyltransferases M.SssI, its humanized derivative MQ1, and its slightly more specific counterpart MQ1^{Q147L} (Lei et al., 2017). In general, the activity of these enzymes is higher than that of DNMT3A, but it is also accompanied by an important increase in off-target effects.

Not long after the unveiling of the mechanisms of active DNA demethylation by the TET enzymes, the first TALE-TET1 fusions were generated and allowed to demonstrate that the demethylation of CpGs at promoters can lead to a substantial increase in gene expression (Maeder et al., 2013). Based on the observation that some CpGs were more efficiently demethylated than others, it was hypothesized that the extent of the demethylation observed may represent a steady state between demethylation and re-methylation (Maeder et al., 2013). Concomitantly ZF were used to target the catalytic domains (CD) of TET1, TET2, and TET3 to two methylated gene promoters, and showed that while they were both demethylated to varying degrees (depending on the CpG assessed), only one of them showcased increased transcription (Chen et al., 2014). This is the only work that has compared the demethylation activity of all three members of the TET family, and it showed that ZF-TET2$_{CD}$ induced the highest level of DNA demethylation followed by ZF–TET1$_{CD}$. ZF-TET3$_{CD}$ was seen to be ineffective. Most other TET-associated demethylation tools have utilized TET1$_{CD}$ ever since, likely because the catalytic domain of TET1 is smaller than that of TET2, a matter that facilitates the generation of the fusion proteins and their delivery, especially for in vivo applications. Since the first dCas9-TET1$_{CD}$ epigenome editing tools were generated (Liu et al., 2016), many variations aiming at improving the demethylation efficiency have followed, such as the use of the SunTag or MS2 coat protein to target multiple TET catalytic domains to the same locus. All these efforts have had varied degrees of demethylation efficiencies and generally resulted in a weak –if any– reactivation of the transcription, likely due to other chromatin modifications that reinforce the repression (Gjaltema and Rots, 2020). To address this issue, the simultaneous use of dCas9-TET1$_{CD}$ and dCas9-VP64 (a transcriptional activator) has been successfully used recently (Baumann et al., 2019). In an interesting turn of events, it has been also recently been proposed that the use of dCas9 alone, without any EppiEfector, is sufficient to physically interfere with the DNMTs and cause a site-specific passive DNA demethylation (Sapozhnikov and Szyf, 2021).

Overall, the editing of the epigenome is not only powerful tool to unveil the relevance of a given epigenetic modification in a specific genomic locus; it is also a potential therapeutic alternative for the treatment of diseases with known epigenetic components, and many efforts are underway for its application in cancer therapeutics (Ansari et al., 2022; Wang et al., 2021) .

III.5. DNA methylation alterations and disease

The genome wide CpG deficiency and the hypermutability of the dinucleotide prompted early investigations on the possible relationship between DNAme and human disease. A metanalysis of the until then reported single base pair mutations that were associated with genetic diseases revealed that 35% of all mutations occurred within CpG dinucleotides. Over 90% of these mutations were C>T or G>A transitions, which thus occurred within these coding regions at a frequency 42-fold higher than that predicted from random mutation (Cooper and Youssoufian, 1988). Disturbances in the expression of imprinted genes cause a handful of well-known congenital disorders, such as Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome and Silver-Russell syndrome (recently reviewed in Eggermann et al., 2023). Alterations in DNA methylation have also been linked to metabolic diseases –type 2 diabetes in particular– (Ahmed et al., 2020; Barres and Zierath, 2011), autoimmune disorders (Richardson, 2003; Zouali, 2021), ageing (Horvath and Raj, 2018; Jones et al., 2015), and neurological disorders such as the autism spectrum disorder (Tremblay and Jiang, 2019) and Parkinson's disease (Wüllner et al., 2016) (for a comprehensive review on DNAme and human disease see Robertson, 2005). In the following sections I will discuss the role of DNAme in cancer and in a rare disease of particular interest for this thesis; the Immunodeficiency, Centromeric heterochromatin instability and Facial anomalies (ICF) syndrome.

III.5.1. Cancer

The DNA methylation landscape of cancer is one of global hypomethylation with loci-specific hypermethylation (Jones and Baylin, 2002) and alterations of the DNAme patterns are considered today as one of the hallmarks of cancer cells (for a recent review, see Nishiyama and Nakanishi, 2021). It has even been proposed that epigenetic alterations of 'tumour-progenitor genes' are the initiating events leading to the malignant transformation of cancer stem cells (Feinberg et al., 2006).

The global hypomethylation of tumours was the first epigenetic abnormality observed in human cancer (Gama-Sosa et al., 1983; Feinberg and Vogelstein, 1983). In one of these first reports, the percentage of hypomethylated primary tumours was seen to be intermediate between benign neoplasms and metastases, hinting towards the fact that tumour progression could possibly be

accompanied by a gradual demethylation (Gama-Sosa et al., 1983). Later, the DNA from several benign (polyps) and malignant colonic neoplasms was analysed by restriction digestion and probed in four genomic loci: compared to matched normal tissues, all neoplasms were seen to be hypomethylated and the fact that polyps already presented methylation changes prior to malignancy onset suggested that alterations in DNAme could be a key event in the initiation of neoplasia (Goelz et al., 1985). Several subsequent reports will confirm the frequent overall genomic hypomethylation in cancers (reviewed in Ehrlich, 2009). The global hypomethylation phenotype has been reproduced in mice carrying one null and one hypomorphic *Dnmt1* allele. These mice express ~10% of the normal DNMT1 level and present a global hypomethylation, which is sufficient for 80% of them to develop at early age aggressive T-cell lymphomas (Gaudet et al., 2003). The loss of methylation was later directly linked to chromosomal instability with increased whole chromosome loss of heterozygosity (aneuploidy), a feature that is suggestive of a specific effect of hypomethylation on the stability of the centromeric or pericentric regions (Eden et al., 2003). Indeed, global hypomethylation is detected because it specially affects repetitive elements which are, as mentioned in the previous chapter, the most represented sequences in the genome. Frequent hypomethylation of the pericentromeric repeats of chromosome 1 (and 16) have been described in breast adenocarcinomas (Narayan et al., 1998), Wilms tumours (Qu et al., 1999) and hepatocellular carcinomas (Wong et al., 2001), to name a few, and have been directly related to chromosomal instability. The analysis of a database of human carcinomas revealed that up to 60% of certain types of tumours present (peri)centromeric rearrangements (Barra and Fachinetti, 2018). In ovarian cancer, the pericentromeric satellite hypomethylation has been seen to increase from normal tissue towards neoplasm, and further progress as the tumour advances grade and stage, and has been proposed as a useful marker of poor prognosis (Widschwendter et al., 2004). The global hypomethylation and chromosomal instability have been also linked to the activation of endogenous retroviral elements and their pro-oncogenic insertion in gene-coding regions (Howard et al., 2008).

Since early on, 5mC has been considered as an "*endogenous mutagen and carcinogen in humans*" because methylation seems to increase the potential for mutation at cytosine residues at least by a factor of 10 (Rideout et al., 1990). The study of the tumour suppressor p53, often mutated in cancer, revealed that in colon carcinomas in particular, 47% of all p53 mutations were transitions in the context of CpG dinucleotides (Greenblatt et al., 1994). Other than causing an increased mutagenesis rate, the hypermethylation of CGI promoters of tumour suppressor genes, and the consequent silencing of their expression, has been directly linked to the oncogenic process (reviewed in Herman and Baylin, 2003). One of the first reports of hypermethylation of CGI promoters was on the tumour suppressor p16, which was associated to a complete transcriptional repression and observed in many primary neoplasms (Merlo et al., 1995). Several other similar reports followed on a gene-to-gene basis until the analysis of 1,184 random CGIs in 98 primary human tumours revealed definite tumour-type specific CGI hypermethylation patterns (Costello et al., 2000). The association of DNA

hypermethylation with tumour suppressor silencing was conceptually easy with the traditional premise of associating DNAme to transcriptional repression. Recently, however, DNA hypermethylation of gene bodies has also been associated to an increased expression of Homeobox oncogenes (Su et al., 2018), therefore linking the cancer-associated hypermethylation to the dysregulation of both tumour suppressors and oncogenes.

III.5.2. The ICF syndrome

At the 1978 Symposium of the European Society of Human Genetics two cases of immunodeficiency associated with chromosome multibranching were reported. The first was only briefly depicted: a 5-year-old patient with variable combined immunodeficiency had been referred for a cytogenetic investigation because of a peculiar facial appearance and "*exceptional cytogenetic observations were made in PHA-stimulated peripheral blood cells*, *but were not found in direct bone marrow or cultured fibroblast preparations*" (Hulten, 1978). [Phytohemagglutinin (PHA) is a mitogen used for stimulation of lymphocyte proliferation in cell cultures]. No images were provided, but the cytogenetic observations were described as "*strikingly aberrant complicated figures with "branching" distal to 1q12*". The second case, also briefly reported at the Symposium, was published with a detailed recount of the findings one year later. A 12-year old patient, who likewise suffered from combined immunodeficiency, presented complex combinations of multibranched chromosomes that were identified as involving exclusively whole arms of chromosomes 1, 9 and 16 (Tiepolo et al., 1979). More than 50 different combinations of anomalies were described in this work and have since been described in many other patients (**Figure 18A, B**). Serendipity would play its role, as these structures can only be observed lymphocytes that have been mitogen stimulated, a common procedure for karyotyping from blood samples. These rearrangements may occur in vivo but certainly at very low rates since they do not seem to be compatible with life. It is likely that the in vitro conditions of mitogen stimulation are revealing an inherent susceptibility of lymphocytes to these rearrangements (Ehrlich, 2003). A decade later, four more cases had been reported, all sharing the same three components: Immunodeficiency, Centromeric heterochromatin instability, and Facial anomalies. The acronym ICF was proposed to describe this new syndrome (Maraschio et al., 1988). For comprehensive reviews on the disease see (Ehrlich et al., 2006; Vukic and Daxinger, 2019).

The distinction between centromere and pericentromere was not well defined at the time, and it was not until 1998 that it was clarified that it is the pericentromeric satellite 2 DNA that becomes decondensed and fused in the multibranched chromosomes, and not the centromeric α-satellite DNA (Sumner et al., 1998). We now know precisely that the 1q12 locus in chromosome 1 corresponds to the largest array of pericentromeric human satellite 2 in the genome (HSat2A2, 13.2 Mbp), followed closely by that of chromosome 16 (HSat2B2, 12.7 Mbp), while chromosome 9 has the largest array of human satellite 3 in the genome (HSat3B5, 27.6 Mbp) (Altemose et al., 2022a).

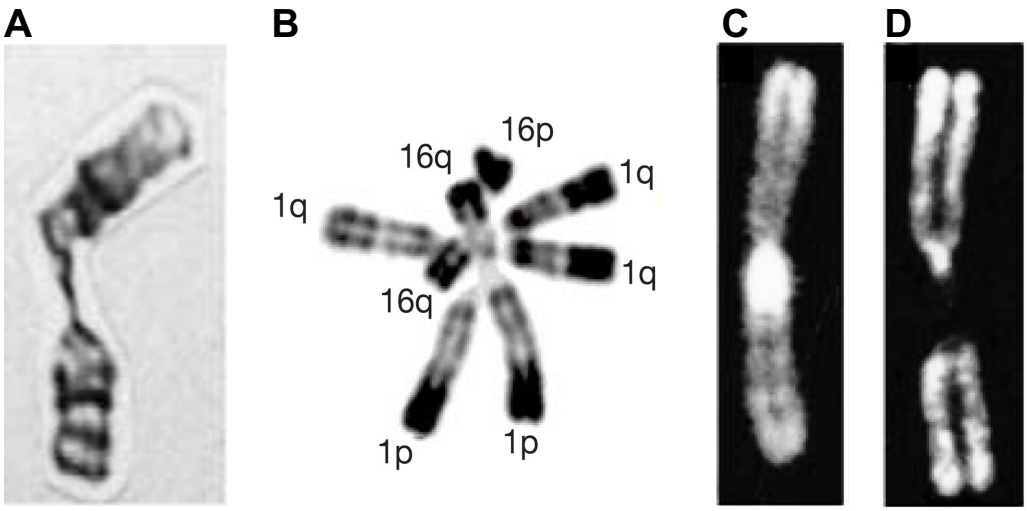

Figure 18. Chromosomal abnormalities in ICF syndrome.
(**A**) Adapted from Ehrlich, 2003 [*Figure 2B*]. Chromosome 1 with decondensation in the qh region. (**B-D**) Adapted from Xu et al., 1999 (**B**) [*Figure 1A*] Multiradiate chromosome from ICF lymphocyte with multiple p and q arms derived from chromosomes 1 and 16. Prepared for microscopy by R banding. (**C**) [*Figure 1Cb*] 5mC immunofluorescence of chromosome 1 from unaffected individual showing intense staining in highly condensed juxtacentromeric satellite 2 DNA in proximal q arm (**D**) [*Figure 1Cc*] decondensed juxtacentromeric region of chromosome 1 from ICF patient shows little to no staining for 5mC

Before the description of the first ICF cases, it had been observed that the treatment of human lymphocytes with the DNA demethylating drug 5-AC resulted in the decondensation of the juxtacentromeric heterochromatin and in the formation of "rosettes", which affected chromosomes 1, 9 and 16 particularly (Viegas-Péquignot and Dutrillaux, 1976). This outstanding resemblance to the cytogenetic observations from ICF patients prompted the investigation of the methylation status of the satellite DNA in four ICF patients and led to the discovery that they all had reduced DNA methylation at satellites I, II and III, with normal α-satellite DNA methylation levels (Jeanpierre et al., 1993). Follow up studies revealed that the hypomethylation of the classical human satellites II and III is a molecular signature of the disease and is found in all patients, while there is heterogeneity regarding the methylation status of the α-satellite DNA and other repeats, with some patients presenting with normal methylation levels while other also exhibit hypomethylation (Miniou et al., 1997). A phenotype that once was thought to be the result of prolonged antibiotic treatment (Tiepolo et al., 1979) and later was attributed to defects in the structure of heterochromatin (Maraschio et al., 1988) was now believed to be the result of constitutive demethylation of satellite DNA (Miniou et al., 1997) (**Figure 18D**). Following this line of thought, and twenty years after the report of the first cases, it was found that ICF patients have mutations in the -back then recently discovered- DNA methyl transferase DNMT3B (Hansen et al., 1999; Xu et al., 1999). At least 19 different DNMT3B mutations have been reported, most of them being missense mutations in or near the catalytic domain, that give rise to hypomorphic proteins. No ICF patient has been found to be homozygous

for a nonsense allele, reinforcing the notion that DNMT3B absence is not compatible with life (Wijmenga et al., 2000). The sequencing of DNMT3B in larger cohorts of ICF patients revealed that only some of them have DNMT3B mutations and allowed to establish a correlation with normal α-satellite methylation levels (Wijmenga et al., 2000; Jiang et al., 2005) and with sub-telomeric hypomethylation (Toubiana et al., 2018). A classification of two different types of ICF syndrome was proposed (Jiang et al., 2005): ICF type 1 (ICF1) affects patients with DNMT3B mutations, and presents with normal α-satellite methylation and sub-telomeric hypomethylation. ICF type 2 patients, on the other hand, do not have mutations in DNMT3B, exhibit hypomethylation of the α-satellites and normal sub-telomeric methylation levels.

Subsequent work aimed at deciphering what was the genetic origin of the disease in patients that did not have DNMT3B mutations led to the identification of mutations in the ZBTB24 (zinc-finger and BTB [bric-a-bric, tramtrack, broad complex]-domain-containing 24) gene in some, but not all, patients (de Greef et al., 2011). ZBTB24 belongs to a family of more than forty ZBTB transcription factors, some of which are known to have a regulatory role in hematopoietic differentiation. The mutations in ZBTB24 retained the ICF type 2 (ICF2) designation, and the remainder of patients were classified as ICFX. A few years later a study focused on ICFX patients identified mutations in the gene of the CXXC-type zinc finger protein CDCA7 (cell division cycle associated 7), referred to as ICF3, and in the gene of the chromatin remodeler of the SWI/SNF family HELLS (helicase lymphoid specific, also known as LSH), referred to as ICF4 (Thijssen et al., 2015). The latest systematic review of the literature says that there have been 118 cases of ICF reported worldwide since the late 70s, with ~60% of them being ICF1, ~30% ICF2, ~4% ICF3 and ~6% ICF4 (Kiaee et al., 2021). Two ICFX cases of unknown aetiology remain, while a third one has recently been found to have been caused by compound heterozygous mutations in UHRF1. This patient exhibits the classical hypomethylation of centromeric and pericentromeric repeats but also a genome-wide CpG hypomethylation not seen before in ICF (Unoki et al., 2023).

The identification of mutations in ZBTB24, CDCA7 and HELLS –three genes devoid of DNA methyltransferase activity or known relationship to DNA methylation pathways– was surprising and raised many questions about a possible functional link between them and DNMT3B. A comparative methylome analysis of 15 ICF patients representing the four possible genotypes was performed to evaluate the similarities and differences in their methylation landscapes and identify all genomic regions that rely on each factor for their methylation status (Velasco et al., 2018). Overall, the methylome from ICF1 patients was distinct from all others, while the ICF2, 3 and 4 methylomes clustered together, suggesting that these three factors are somehow connected or act on the same pathway. The methylome of ICF1 patients revealed the prominent role played by DNMT3B in establishing the methylation status of germ line genes, the inactive X chromosome, CpG-rich intergenic loci, and CGI promoters. The ICF2, 3 and 4 methylomes display reduced methylation at

open seas, in CpG-poor genomic regions that have hallmarks of heterochromatin and late-replicating signatures, and in several genes that are suspected to be relevant for neurodevelopment, while CGIs are barely affected.

How ZBTB24, CDCA7, and HELLS contribute to DNA methylation and possibly heterochromatin formation at pericentromeric repeats and open sea regions remains an open question. The deletion of the BTB domain of ZBTB24 causes early embryonic lethality in mice (Wu et al., 2016), and ZBTB24 was seen to associate with the CDCA7 promoter and control its expression in mouse ES cells, demonstrating a convergence of these two ICF genes at the level of transcription (Wu et al., 2016). Using *Xenopus* egg extracts it was proposed that CDCA7 is required to recruit HELLS to the chromatin, and that the two proteins associate to form a complex with nucleosome remodelling activity which allows DNMT3B to access the DNA (Jenness et al., 2018). In this model the four ICF factors act on the same pathway, ZBTB24 being the transcription factor regulating CDCA7 expression. This proposition is therefore incompatible with the clear differences observed in the methylomes of ICF1 versus ICF2-3-4 patients (Velasco et al., 2018). A subsequent study in human and mouse cell lines identified HELLS as an essential factor for DNMT1 mediated DNA methylation maintenance by directly interacting with UHRF1, facilitating its association to the chromatin and promoting its H3 ubiquitination activity (Han et al., 2020). UHRF1 was also identified as an interactor of CDCA7 (Unoki et al., 2020). Combined, these observations led to the hypothesis that DNA hypomethylation due to DNMT3B mutations in ICF1 is likely the result of a defect in the de novo DNA methylation during development, while that mediated by ZBTB24, CDCA7, and HELLS mutations in ICF2, 3 and 4 is likely a result of a defect in maintenance DNA methylation in a replication-uncoupled manner after the establishment of DNA methylation patterns (Unoki, 2021). The newly proposed model is that first UHRF1 interacts with and recruits the CDCA7/HELLS complex to hemi-methylated DNA of late-replicating regions, where they can remodel the chromatin, further facilitating the UHRF1 interaction with the DNA, its H3 ubiquitination activity and the recruitment of DNMT1 (Unoki, 2021). This model still needs to be experimentally confirmed, but at least theoretically it fits well the different features observed in the sub-types of ICF syndrome.

The dysregulation of the DNA methylation is likely involved in the causation of the disease, but precisely how the changes in DNA methylation patterns lead to the phenotypic aspects of the disorder remains an open question. Of particular interest to us, the ICF syndrome is a unique pathophysiological context in which DNA methylation of the centromeric α-satellites is decreased. Given that the unifying feature of all types of ICF is the loss of pericentromeric methylation, it remains unclear if the loss of centromeric methylation specifically observed on ICF2, 3, and 4 has a direct influence on the observed chromosomal fragility. Likewise, it is still poorly understood what is the particular role that ZBTB24, CDCA7 and HELLS are playing for the maintenance (or establishment) of the centromeric methylation.

PhD PROJECT

The maintenance of centromeric integrity and function is of the upmost importance for the cell, as centromere aberrations are a source of genome instability and have been linked to human disease (reviewed in Barra and Fachinetti, 2018). Several pieces of evidence point towards a role of DNA methylation in ensuring the centromeric function (Scelfo and Fachinetti, 2019), but little is known about how this is precisely occurring and if and how different centromeric components might also contribute to establish and maintain the centromeric methylation landscape.

Centromeric DNAme appears to play a direct role in the regulation of at least two core centromeric proteins: CENP-B and CENP-A. The fact that the CENP-B box contains two CpGs on its nine conserved base pairs (Masumoto et al., 1989) and that the affinity of CENP-B binding is decreased if the CENP-B box is methylated (Tanaka et al., 2004; Dubocanin et al., 2023) are clear indicators of a possible role of DNAme in regulating CENP-B binding to the centromeric DNA. A redistribution of CENP-B has indeed been observed after the demethylation of mouse cells by treatment with the hypomethylating drug DAC (Mitchell et al., 1996). In the recent telomere-to-telomere genome assembly a marked dip in DNA methylation was observed to correlate precisely with the CENP-A enriched regions of the active HORs in all chromosomes (Logsdon et al., 2021; Altemose et al., 2022a). It seems doubtful that this anticorrelation is a coincidence and it is likely an indication that DNAme is acting as a signal to define the domains where CENP-A can be deposited at the active centromeres. The opposite could also be true; CENP-A enrichment could be determining factor of a methylation-free domain in an overall highly methylated centromere.

Regarding how centromeric specific components might contribute to establish and maintain the centromeric methylation landscape, a two-hybrid screening identified an interaction between DNMT3B and CENP-C and proposed that CENP-C recruits DNMT3B to both centromeres and pericentromeres (Gopalakrishnan et al., 2009). The same study reported that CENP-C knock-down caused a ~20% reduction DNA methylation at α-satellite and pericentromeric satellite 2, and that chromosome segregation errors ensued following both CENP-C and DNMT3B reduction. However, no follow-up studies have been conducted.

The study of neocentromeres also has provided some interesting clues towards a possible role of DNAme in the de novo establishment of the centromere. The neocentric chromatin was found to have overall high levels of methylation, with specific sites of active transcription correlating with pockets of hypomethylated DNA (Wong et al., 2006). It is unclear if the hypermethylation is an epigenetic pre-requisite for the neocentromere formation at any particular locus, or a consequence of the neocentromere formation. In any case, these observations clearly point towards an intrinsic

relationship between the methylation status of the underlying DNA and the (neo)centromeric function.

Centromeric DNAme also seems to play an important role in the centromere function, as the treatment of human cells with DAC was shown to particularly reduce the methylation of the (peri)centromeres and was correlated with increased mitotic defects and aneuploidy (Costa et al., 2016). Mouse cells lacking both DNMT3A and DNMT3B presented increased mitotic centromeric recombination accompanied by shortening of the centromere repeats, suggesting that DNAme might also be important for keeping centromeric recombination at bay (Jaco et al., 2008), and/or possibly also play a role in the evolution of the centromeric DNA. It has also been reported that cells treated with a hypomethylating drug, 5-AC, have increased chromosome segregation into micronuclei, a known readout of centromeric dysfunction (Guttenbach and Schmid, 1994; Suzuki et al., 2002). Finally, the study of (peri)centromeric demethylation in cancer and ICF syndrome are two clear pathophysiological contexts in which the relevance of the (peri)centromeric DNAme is highlighted. It is however unclear if all the effects abovementioned are directly linked to the centromeric DNAme loss, and if the centromeres are indeed dysfunctional in all these contexts. The genomic instability observed could be the consequence of the global demethylation, or a combination of the methylation loss of centromeres and pericentromeres. No formal investigation on the specific and isolated role of DNAme at the centromeres has been conducted in a manner that precludes the genome-wide confounding effects ensuing from the genetic inactivation of the DNA methyltransferases or the treatment with DNA demethylating drugs, which are also cytotoxic.

HYPOTHESES

Based on our current understanding of the roles of DNA methylation throughout the genome, we hypothesise that upon centromeric demethylation: (1) the chromosome segregation fidelity is decreased, (2) the abundance and/or the distribution of centromeric proteins is altered, (3) the centromeric transcription is increased and (4) the centromeric chromatin is decompacted.

I conducted my experimental research to address each of these hypotheses.

AIMS and METHODOLOGY

I. Explore the role of DNMT1 and DNMT3B in the maintenance of the centromeric methylation.

To interrogate which DNA methyltransferase(s) (DNMTs) play a role in regulation of the centromeric methylation, I took advantage of two cellular tools recently developed in the Fachinetti laboratory. Through CRISPR-Cas9 genome editing, an auxin inducible degron tag was added to DNMT1(DNMT1^{AID}), for its rapid, inducible degradation, either in a DNMT3B wild type (WT) or knockout (KO) background (Scelfo et al., 2023). Using these two cell lines I could interrogate the effects of the depletion of DNMT1 alone, DNMT3B loss alone, or the two combined on the centromeric DNA methylation, and establish the relative contribution of each DNMT to the maintenance of centromeric methylation.

II. Study the cellular and molecular consequences of a centromere-specific demethylation.

The main objective of my thesis was to reveal how DNA methylation participates in the specification and maintenance of the centromere identity and function. Given all the evidence that indicates that loss of DNAme at the centromeres is deleterious, I developed a centromere-targeted epigenome editing method by fusing the DNA binding domain of CENP-B as a targeting moiety to the catalytic domain of TET1. With this tool I was able to specifically demethylate the centromeres in human cell lines, and study the consequences of the centromeric methylation loss, without any confounding genome-wide side effects.

III. Generate an ICF model cell line to study the consequences of the loss of HELLS in the maintenance of centromeric methylation.

Finally, I generated an ICF4 model cell line by tagging HELLS with an AID tag through CRISPR-Cas9 genome editing. This cell line allowed me to degrade the helicase in a controlled manner to assess kinetics of demethylation upon its loss. This cell line allows to study the effects of the centromeric hypomethylation in a pathophysiological context, and to interrogate the role of HELLS in the maintenance of the DNAme at the centromeres.

RESULTS

1. Regulation of the centromeric DNA methylation

1.1. DNMT1 and DNMT3B regulate the centromeric DNA methylation

To investigate which DNA methylation pathways are important for the regulation of the centromeric methylation, I took advantage of two DLD-1 colorectal cancer cell lines recently established in the Fachinetti laboratory. Through CRISPR-Cas9 genome editing, an auxin inducible degron tag was added to DNMT1(DNMT1^{AID}), for its rapid, inducible degradation, either in a DNMT3B wild type (WT) or knockout (KO) background (Scelfo et al., 2023). With these two cell lines we can interrogate the effects on the centromeres of the depletion of DNMT1 alone, DNMT3B loss alone, or the two combined. These two cell lines were treated for 6 days with auxin (IAA: indole-3-acetic acid) to deplete DNMT1 and subjected to CenRICH, a method developed in the Fachinetti laboratory for enrichment of centromeric DNA based on selective restriction digestion and size fractionation (Gamba et al., 2022). The enriched DNA fractions were sequenced with Nanopore (Oxford Nanopore Technologies), a long-read sequencing method that on top of the DNA sequence, can distinguish modified bases such as methylated cytosines. Since the centromeric sequence of DLD-1 cells is not assembled, the reads from the Nanopore sequencing could not be accurately mapped to specific chromosomic loci. We therefore applied a motif-based analysis, which allowed the identification of the reads containing centromeric α-satellite (α-sat) sequences –not containing CENP-B boxes–, reads specifically containing CENP-B boxes, or reads containing pericentromeric human satellite 2 (HSAT2) repeats. The methylation levels in these specific repetitive regions were therefore accurately quantified in a mapping-independent manner (**Figure 19A**). As expected, the wildtype cells (DNMT1$^{+/+}$, DNMT3B$^{+/+}$) had overall very high (~80%) CpG methylation levels at both the centromeric and pericentromeric repeats. The methylation level of the CENP-B boxes is significantly (18%) lower than that of the generic α-sat sequences. This difference possibly reflects the necessity of the boxes present at the active centromeres to be unmethylated, since CpG methylation of the CENP-B boxes was suggested to reduce the binding of CENP-B (Tanaka et al., 2004; Dubocanin et al., 2023). We measured by microscale thermophoresis the in vitro affinity of CENP-B to DNA fragments with methylated or unmethylated CENP-B boxes and quantified that indeed CENP-B binds to the unmethylated boxes with eight times stronger affinity (Kd~0.3 μM) than to the methylated ones (Kd~2.5 μM) (**Figure 19B**).

The DNMT3B^{KO} cells present less than 2% decrease in methylation at centromeres (α-sat) and pericentromeres compared to their wild type counterparts (**Figure 19A**). The CENP-B boxes seem slightly more affected by the loss of DNMT3B, exhibiting nearly a 9% reduction in their methylation level. The degradation of DNMT1^{AID} alone has marked effects in all three regions assessed, with up to 20% reduction in the methylation levels of both HSAT2 and CENP-B boxes (**Figure 19A**). The

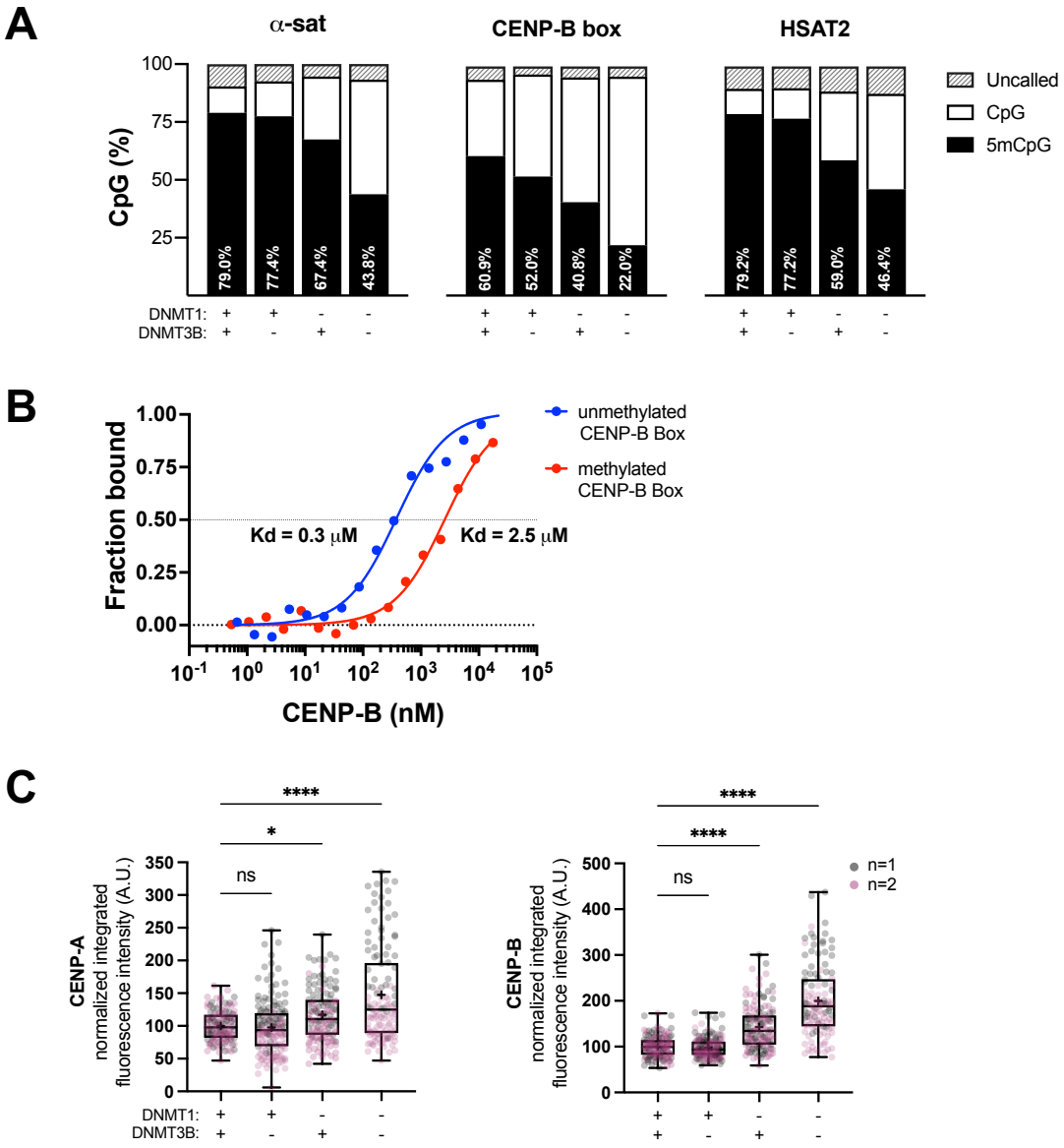

Figure 19. DNMT1, and to some degree DNMT3B, regulate the centromeric and pericentromeric DNA methylation.
(A) Mapping-independent motif-based analysis of Nanopore reads with methylation calls. **(B)** Microscale thermophoresis affinity curves of purified CENP-B binding to methylated or unmethylated DNA. **(C)** Quantification of the normalized integrated fluorescence intensity of CENP-A and CENP-B at the centromeres of cells in interphase. DNMT1 degradation was achieved by treatment with IAA for 10 days. Each dot represents the average intensity of all the centromeres of a single cell. Two experimental replicates with N>49 cells and >1646 centromeres per replicate and condition. Statistics is defined by Kruskal-Wallis test with Dunn's multiple comparisons test. ns: non-significant differences.

degradation of DNMT1^{AID} in the DNMT3B^{KO} background has the most profound effect on the methylation levels of all regions, with methylation reductions of 35.2% in α-sat, 32,8% in HSAT2 and 38.9% in CENP-B boxes. These reductions are larger than the sum of those observed in DNMT1$^{-/-}$ alone and DNMT3B^{KO} alone, and are a reflection the cooperativity between DNMT1 and DNMT3B (Scelfo et al., 2023).

I next investigated the effect of DNMT1 and/or DNMT3B loss on the abundance at the centromeres of two of the main centromeric proteins, CENP-A and CENP-B, though indirect immunofluorescence microscopy (**Figure 19C**). Compared to the wild-type cell line, and concomitant with the methylation measurements, the levels of the two proteins remained stable in the DNMT3B^{KO} cell line. DNMT1 degradation alone has a significant effect on CENP-B levels at the centromeres, which increased by 43%, and led to a 17% increase of CENP-A levels. The degradation of DNMT1 in the DNMT3B^{KO} background had the most significant effects in all three proteins, with a 48% mean increase in CENP-A levels, a doubling of mean CENP-B levels and a 22% increase in CENP-C levels at the centromeres. These results indicate that DNA methylation plays a role in regulating the levels of these three centromeric proteins at the centromeres, mainly affecting CENP-B. The remarkable increase in CENP-B levels is probably a direct reflection of the higher number of unmethylated CENP-B boxes available for its binding. However, given the genome-wide demethylation that occurs with this system, we cannot rule out that other factors, such as altered protein expression, might influence these results.

1.2 The replacement of CENP-A N-terminal tail by H3 N-terminal tail is not sufficient to increase centromeric methylation.

If the DNA methylation maintenance is not completed in a replication-coupled manner, then the methylation maintenance must occur in an already chromatinized DNA. This back-up mechanism depends on the dual mono ubiquitination of Histone H3 (H3Ub2) in lysines 14 and 18 by UHRF1, which mediate the recruitment of DNMT1 (Nishiyama et al., 2020). These lysine residues are absent in the CENP-A N-terminal tail; CENP-A therefore cannot be ubiquitinated and should not be able to recruit the methylation maintenance machinery. The centromeric methylation dip detected in the active HORs (Logsdon et al., 2021; Altemose et al., 2022a) could perhaps be a consequence and not a cause of the enrichment in CENP-A. To test this hypothesis, I took advantage of a set of cell lines published in Fachinetti et al. 2013 (**Figure 20A**). Briefly, several H3/CENP-A chimeric rescue variants were stably integrated by retroviral transduction into CENP-A$^{-/F}$ RPE-1 cells, and then the floxed CENP-A allele was inactivated by addition of Ad-Cre recombinase. These cells have therefore been expressing only the chimeric versions of CENP-A for many generations. On one chimera of interest here, the CENP-A NH2-tail has been replaced by the one of histone H3 ($^{H3-NH2}$CENP-A). On another the CATD and the C terminal domains of CENP-A were swapped with the corresponding regions of H3 (H3^{CATD+C}). Both these chimeras localize to the centromeres and sustain the kinetochore function (Fachinetti et al., 2013). We hypothesized that addition of the NH2-tail from H3 could increase the recruitment of the methylation maintenance machinery at the centromeres. I measured the DNA methylation levels at the centromeres in these cells by means of a Combined Bisulfite Restriction Analysis (COBRA). Briefly, the genomic DNA is treated with

sodium bisulfite in a reaction that deaminates all cytosines and converts them to uracils (C>U). Methylated cytosines are protected from the deamination. The subsequent amplification of the bisulfite converted DNA by PCR further converts the uracils into thymines (U>T). The epigenetic DNA methylation information is therefore transformed into a genetic sequence change, where all unmethylated cytosines become thymines (C>T) and all methylated cytosines remain as cytosines (5mC>C). These sequence changes can then be visualized as differences in the restriction digestion patterns of the PCR products (COBRA), or the PCR products can be pyruvate sequenced to accurately quantify the percentage of methylation at each CpG. I measured by COBRA the centromeric methylation levels of the centromeres using α-satellite specific primers reported in (Velasco et al., 2018). I compared the global centromeric methylation levels of the parental cell line (CENP-A$^{-/F}$) to that of floxed cells that carry a full-length CENP-A rescue construct, or the chimeric rescue constructs abovementioned (**Figure 20B**). Cells with the $^{H3-NH2}$CENP-A rescue exhibit the same centromeric methylation level as the parental and the full-length CENP-A rescue cells. Cells with the H3^{CATD+C} rescue exhibit a 10% lower methylation. The replacement of the CENP-A NH2-terminal tail for the H3 N-tail is therefore not sufficient to significantly alter the centromeric methylation, at least to levels detectable by COBRA.

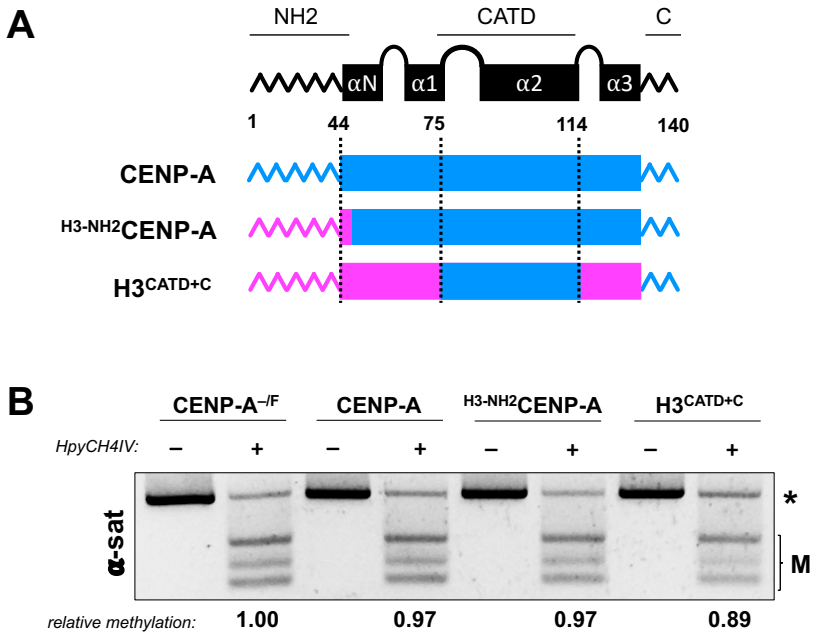

Figure 20. Cells with CENP-A chimeras with H3 N-terminal tail do not exhibit higher centromeric methylation levels.
(A) Schematic representation of the rescue constructs stably expressed on the floxed cells. CENP-A domains and amino-acid positions are indicated. CENP-A domains are in blue; H3 domains in magenta. Each construct is tagged with an N-terminal EYFP (enhanced yellow fluorescent protein) (not drawn) **(B)** α-satellite COBRA agarose gel. The absence (–) or presence (+) of the methylation-sensitive enzyme HpyCH4IV is indicated. The asterisk (*) demarks the position of the unmethylated (undigested) fragment, M indicates the methylated (digested) fragments. The methylation levels relative to the parental (CENP-A$^{-/F}$) sample are indicated.

2. Generation of a cellular toolbox to study the centromeric specific demethylation

2.1. Centromere-targeted demethylation constructs

To isolate the effects of the centromeric demethylation from the genome-wide demethylation that results from global DNMT3B/DNMT1 loss, I generated a cellular toolbox to specifically demethylate the centromeres in an inducible and degradable manner. To achieve this, I assembled a doxycycline inducible construct under the Tet operator (TetO) fusing the DNA binding domain of CENP-B (CENP-B_{DBD}) to the catalytic domain of the ten-eleven translocase 1 ($TET1_{CD}$), one of the enzymes that can catalyse the DNA demethylation. The construct is under the control of the cytomegalovirus (CMV) promoter and has two nuclear localization signals (NLS), a triple FLAG tag, and a micro-AID C-terminal tag to be degraded upon auxin addition (**Figure 21A**). I generated a second construct harbouring two point mutations that render the catalytic domain of TET1 inactive (or dead; herein referred to as $dTET1_{CD}$) to be used as control. For simplicity, the constructs will be referred to as (d)$TET1_{CD}$ herein, omitting the common CENP-B_{DBD} moiety. The addition of doxycycline (DOX) to the culture media induces the expression of the constructs, which should localize specifically to the centromeres through the binding of the CENP-B_{DBD} to the CENP-B boxes (**Figure 21B**). The $TET1_{CD}$ is expected to then catalyse the demethylation of the neighbouring.

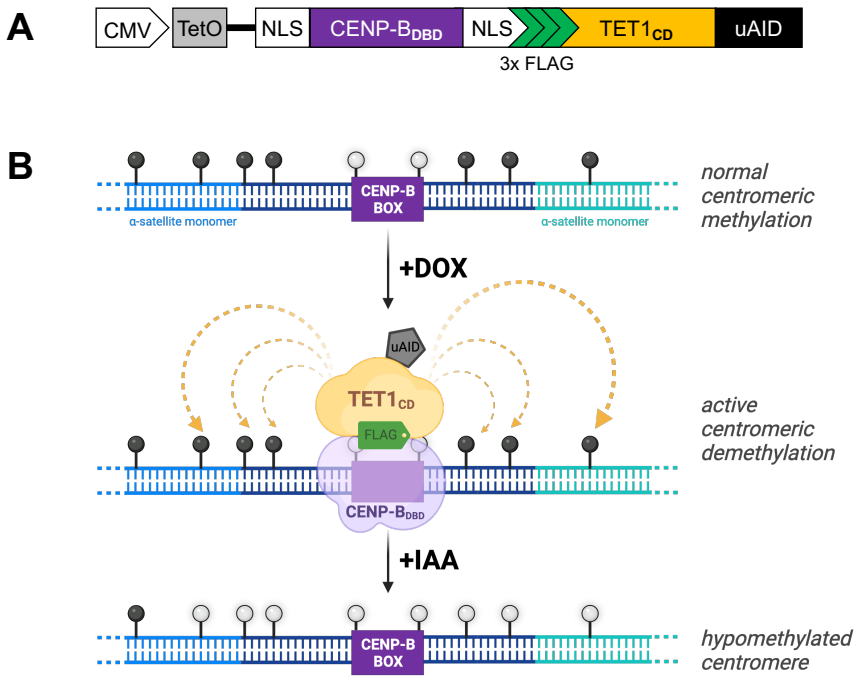

Figure 21. A method for targeted centromeric demethylation.
(A) Schematic representation of the doxycycline inducible, auxin degradable construct **(B)** Schematic representation of the targeted demethylation process. α-satellite monomers represented in shades of blue. 5mCpGs: black lollipops, CpGs: white lollipops. The addition of doxycycline (+DOX) induces the expression of the construct which localizes specifically to the centromeres through the binding of the CENP-B_{DBD} to the CENP-B boxe (in purple). The $TET1_{CD}$ catalyses the demethylation of the neighbouring 5mCpGs. The addition of auxin (IAA) directs the construct to the proteasome for its degradation. Created with BioRender.com

5mCpGs. The addition of auxin (IAA) allows to turn off the demethylation system by directing the construct to the proteasome for its degradation, effectively leaving the hypomethylated centromeres free from any exogenous protein and unhindered for binding centromeric proteins. The demethylating TET1 constructs can be introduced into cells using lentiviral particles or in a site-specific manner through Flp-In™ (Flp-FRT) recombination. The main advantage of the lentiviral method is that it is faster and can be applied to a wide variety of cell lines. However, the insertion of often several copies of the construct occurs randomly in the genome and can potentially have deleterious secondary effects. To bypass this issue, the Flp-In recombination method allows to integrate a single copy of the construct in a site-specific locus, the FRT (Flp recognition target) site. This method, however, can only be used in previously engineered cell lines with an FRT site.

I first introduced the active and inactive $TET1_{CD}$ constructs via Flp-In recombination into Flp-In T-Rex DLD-1 cells that have an FRT site and express Os-TIR1-9xMyc (Holland et al., 2012). I characterized individual clones for each construct and demonstrated that they only express the constructs after induction with doxycycline and that both constructs specifically locate to the centromeres as seen by FLAG and CENP-C colocalization in indirect immunofluorescence (**Figure 22**). I next measured by COBRA the α satellite methylation after 4, 8, 12 and 16 days of induction of the constructs with doxycycline. Only for cells expressing the active $TET1_{CD}$ I detected a progressive decrease in the methylation of their centromeres (**Figure 23A, C**). The kinetics of demethylation are relatively slow, with a 26% reduction in methylation of the centromeres after 4 days, 35% at 8 days and close to 60% after 16 days. In the cells that express the $dTET1_{CD}$ construct the α-sat methylation remained unaltered over time (**Figure 23B, C**). This result confirms that the inactive construct is indeed catalytically dead and proves that the binding of the constructs to the centromeres does not intrinsically alter the methylation status of the underlying α-satellites. To determine the specificity of the demethylation on cells expressing the active $TET1_{CD}$, I also measured by COBRA the methylation level of the pericentromeric HSAT2 repeats (**Figure 23D**) and saw no changes in the methylation over time. If the construct had some unspecific activity or if the demethylation were to spread outside the α-satellites, we would expect the adjacent pericentromeric regions to be the first to be affected. All the evidence indicates therefore that the expression of the active $CENP\text{-}B_{DBD}\text{-}TET1_{CD}$ construct leads to a centromeric specific hypomethylation.

The same α-sat and HSAT2 bisulfite converted PCR products from the active $TET1_{CD}$ cell line that were analysed by COBRA were then subjected to pyruvate sequencing to determine the exact methylation level of each CpG within the amplicons after 8 and 16 days of expression of the construct. For α-sat, we could further distinguish between non-CENP-B box (α-sat) and CENP-B box CpGs, which were sequenced from the same amplicon but using two distinct sequencing primers (**Figure 24A, B**). Each CpG has a distinct absolute methylation level (**Figure 24A**). It is of interest

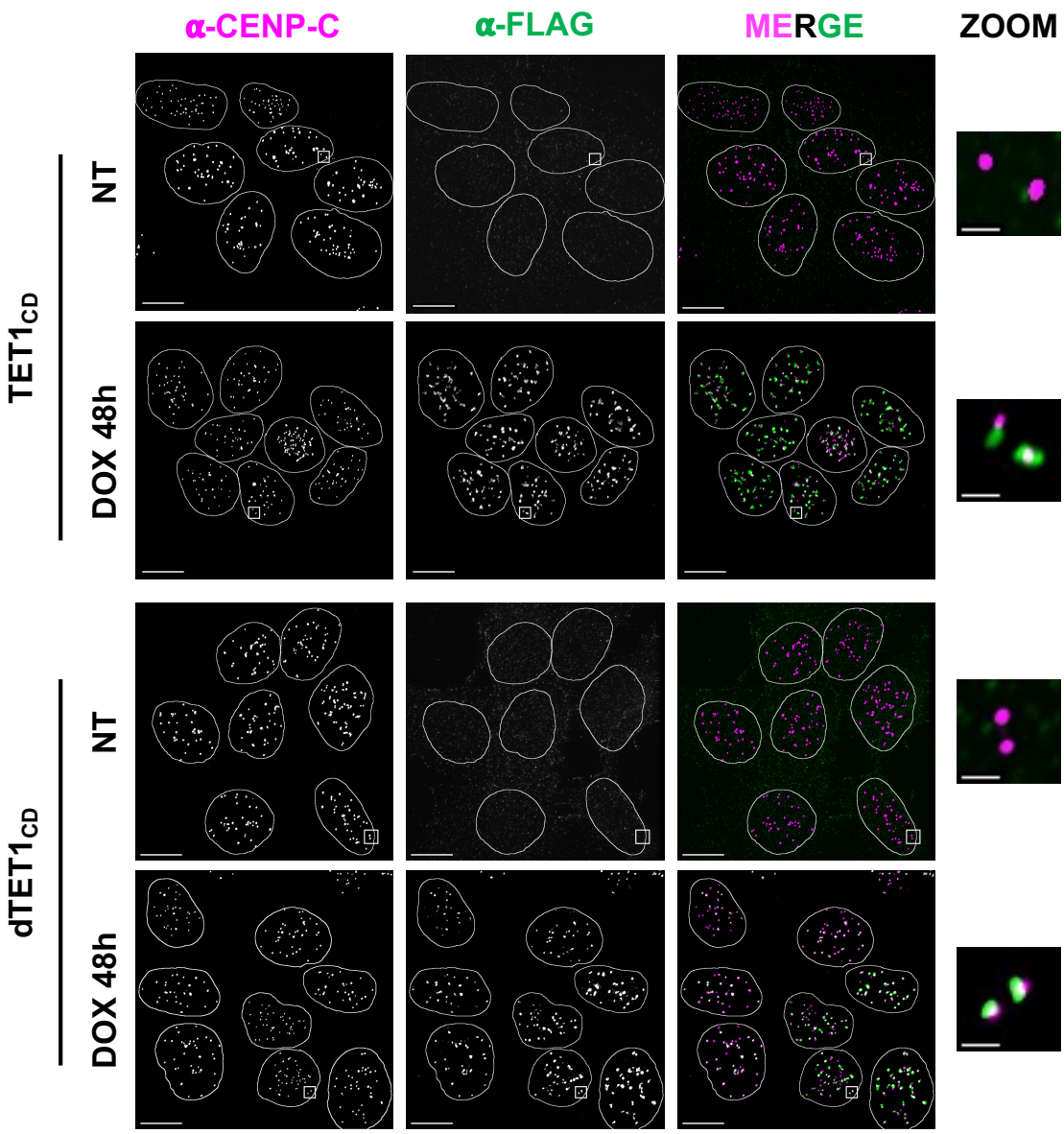

Figure 22. The TET1 constructs are expressed upon doxycycline addition and localize specifically to the centromeres.
Representative immunofluorescence images of untreated cells or after 48h doxycycline treatment. CENP-C staining (magenta) marks the centromere position and FLAG (green) marks the constructs. Scale bar: 10 μm. Zoomed squares scale bar: 1 μm.

to note that the CENP-B box CpGs are overall far less methylated than the rest of the α-sat CpGs. Normalization of each CpG methylation level to 100% at day 0, before the induction of the expression of the $TET1_{CD}$ construct (**Figure 24B**) allows to better visualize that both the CpGs from α-sat and CENP-B box specific sequences show a consistent and progressive decrease in their methylation level over time, diminishing by 61.8% and 53% respectively at 16 days. All HSAT2 CpGs maintained constant methylation levels over time.

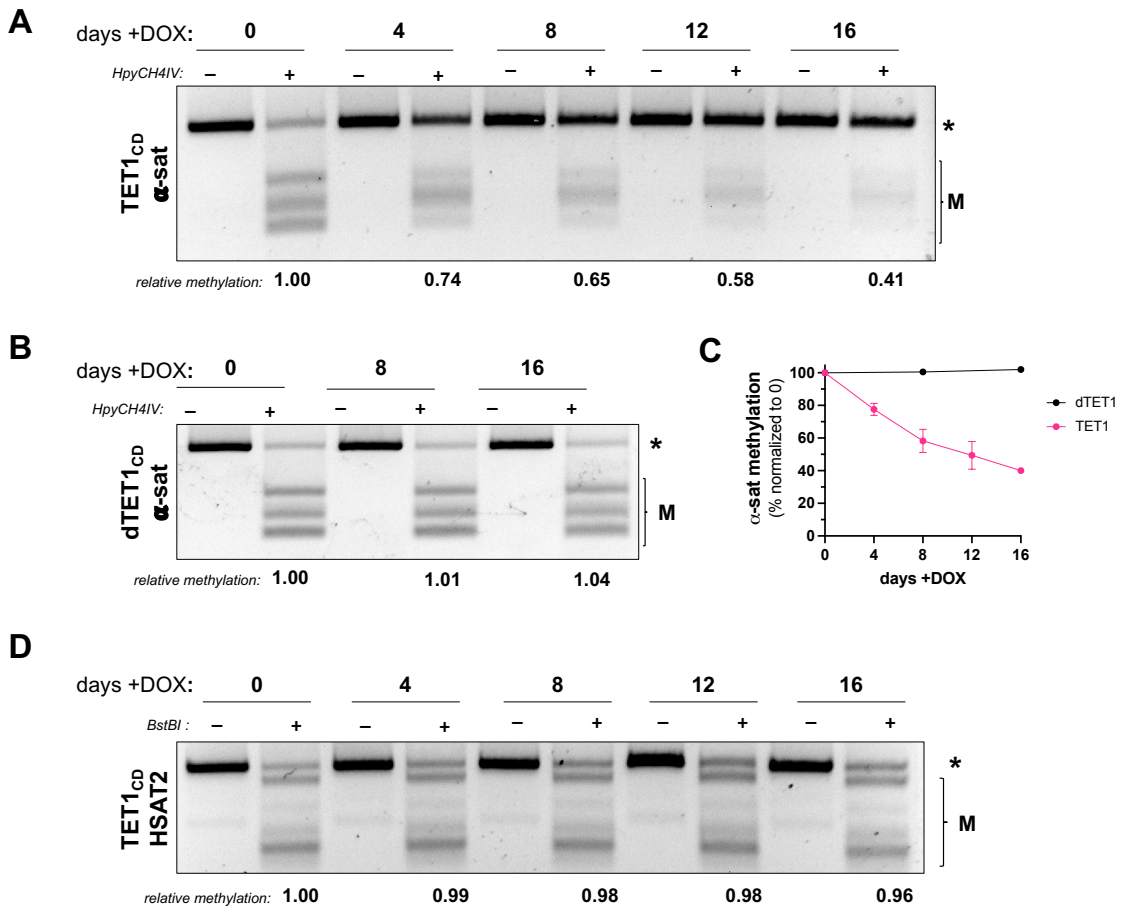

Figure 23. Centromeres are specifically demethylated by the TET1$_{CD}$ construct.
(**A**) Representative agarose gel of the α-satellite COBRA analysis. The demethylation kinetics of cells expressing the active TET1$_{CD}$ construct for 4, 8, 12 and 16 days of induction with doxycycline. The absence (–) or presence (+) of the methylation-sensitive enzyme HpyCH4IV enzyme is indicated. The asterisk (*) demarks the position of the unmethylated (undigested) PCR fragment, M indicates the methylated (digested) fragments. The methylation levels relative to the untreated (NT) sample are indicated **(B)** Representative agarose gel of the α-satellite COBRA analysis of cells expressing the inactive dTET1$_{CD}$ construct for 8 or 16 days. **(C)** Quantification of the methylation levels measured by COBRA on (A-B), normalized to untreated (day 0). **(D)** Representative agarose gel of an HSAT2 COBRA analysis showing the pericentromeric methylation level of cells expressing the active TET1$_{CD}$ construct for 4, 8, 12 and 16 days of induction with doxycycline.

While providing relatively fast and accurate information on the DNA methylation levels, both COBRA and pyruvate sequencing are based on the PCR amplification after bisulfite conversion of a specific sub-set of α-satellites which map almost exclusively to chromosome 1 (based on the alignment of the primers to the T2TCHM13v1.0 reference genome, not shown). This is an understandable bias considering the technical need to amplify a discrete-size PCR product from highly repetitive sequences. To generalize and strengthen our conclusions, we sequenced the genomic DNA of the active TET1$_{CD}$ cells, untreated and after 16 days of expression of the construct, using Nanopore as described above (**Figure 24C**). The motif-based analysis allowed the identification of

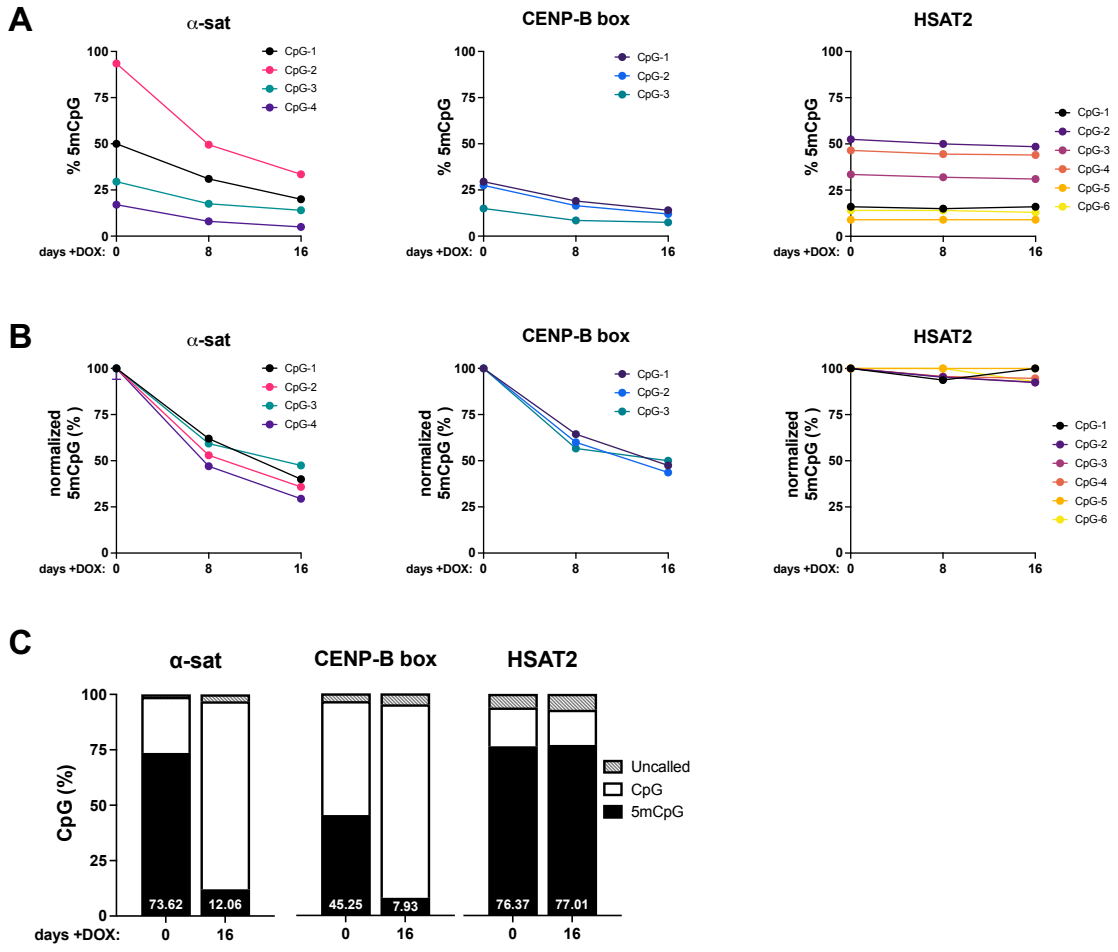

Figure 24. The TET1$_{CD}$ induced demethylation by pyruvate sequencing and mapping independent methylation calling with Nanopore.
(**A**) α-sat, CENP-B box and HSAT2 methylation level of individual CpGs from the active TET1$_{CD}$ cell line quantified by pyruvate sequencing. On the CENP-B box sequence CpG-1 and CpG-2 are located on the consensus box while CpG-3 is adjacent. **(B)** Same data as (A) normalized to day 0 (untreated cells) (**C**) Mapping-independent motif-based analysis of Nanopore sequences methylation calls.

thousands of reads containing centromeric α-sat sequences, CENP-B boxes, or HSAT2 repeats. The mapping-independent methylation calling of these reads allowed to determine that the cell-wide demethylation is more pronounced than what was measured by COBRA and pyruvate sequencing. An average 83.6% decrease of the α-sat methylation and an 82.5% decrease for CENP-B boxes was observed after 16 days of induction of the construct. The numerical discrepancy between Nanopore sequencing and pyruvate sequencing results can be explained by the inclusion of all chromosomes in the analysis. Depending on the density of CENP-B boxes, each chromosome will be demethylated to a different degree and the inclusion of reads from all chromosomes takes into account chromosomes that are probably more demethylated than the chromosome 1. The discrepancy can also be attributed, at least to some degree to a technical difference between the two methods. The methylation calling of Nanopore reads only considers 5mC as methylated; all oxidation intermediates

will be considered as unmethylated or uncalled by the Nanopore analysis. The bisulfite conversion, on the other hand, converts not only unmodified cytosines, but also 5fC and 5caC to uracils (Neri et al., 2016). Both 5mC and 5hmC remain protected and therefore will be considered as methylated bases. This means that, even though the methylation process might be underway, the accumulation of 5hmC after the first oxidation step will be still considered as methylated DNA. The HSAT2 methylation levels was again unchanged by the expression of the construct, reinforcing the conclusion that the high levels of demethylation achieved with this system is centromere specific.

2.2. TET1$_{CD}$ synergizes with the degradation of DNMT1^{AID}

The kinetics of demethylation observed with the active TET1$_{CD}$ are relatively slow (**Figures 23 and 24**). One explanation for this is that the cells have intact DNA methylation pathways, which can antagonize the effects of TET1. The demethylation process may generate hemimethylated sites that should be re-methylated by DNMT1 and the DNA methylation maintenance machinery. I hypothesized therefore that combining active TET1$_{CD}$ indued demethylation system with the passive DNMT1^{AID} degradation should yield faster centromeric demethylation rates. To have at the same time TET1$_{CD}$ expression and DNMT1 degradation I needed to remove the degron tags from the (d)TET1$_{CD}$ constructs. I introduced these new constructs via lentiviral transduction into the DNMT1^{AID}/DNMT3B^{WT} DLD-1 cell line (**Figure 25A**). After antibiotic selection and expansion of the transduced population, I treated the cells for 2, 4 and 8 days with doxycycline to induce the expression of the (d)TET1$_{CD}$ constructs, with auxin to degrade DNMT1, or with both drugs to assess for their combined effect. I measured the α-satellite methylation levels by COBRA at each timepoint (**Figure 25B**). After two days of induction both the active TET1$_{CD}$ expression and the DNMT1^{AID} degradation have a similar effect, reducing the α-satellite methylation by ~25%. At the same timepoint there is an additive effect between the concomitant expression of TET1$_{CD}$ and degradation of DNMT1^{AID}, which leads to a ~56% reduction of methylation. At the four days mark the relative trend is maintained, but the TET1$_{CD}$ effect alone is less pronounced, possibly due to counterselection of the transduced cell populations used in these experiments. These results clearly indicate that TET1$_{CD}$ construct expression synergizes with the degradation of DNMT1^{AID} to potently demethylate the centromeres. The levels of demethylation achieved here in two days are comparable to the levels achieved after 16 days of TET1$_{CD}$ expression with the Flp-In system.

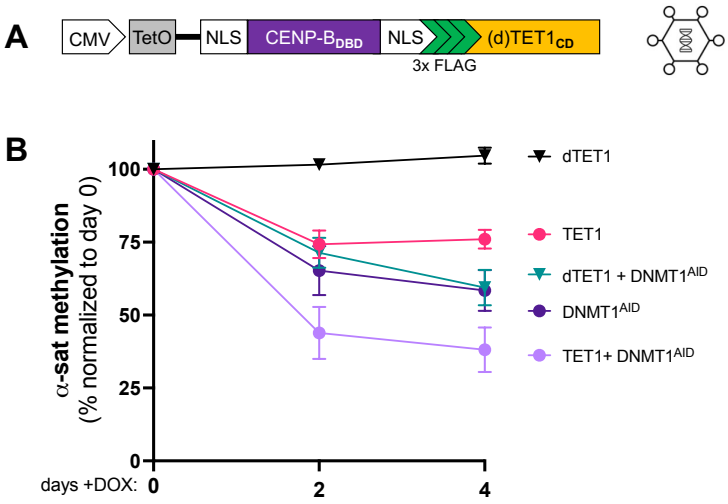

Figure 25. Additive effect of the TET1$_{CD}$ centromeric demethylation and the DNMT1^{AID} degradation.
(**A**) Schematic of the lentiviral construct integrated into DNMT1^{AID}/DNMT3B^{WT} cells.
(B) Quantification of α-satellite methylation levels over time by COBRA. (d)TET1 indicates expression of the constructs with doxycycline and DNMT1^{AID} indicates degradation of DNMT1 by auxin.

3. Cellular effects of the centromeric demethylation

3.1. The centromeric demethylation causes micronucleation and reduces cell viability

To first assess if centromeric function is perturbed by the loss of centromeric DNA methylation, I measured cell division considering the key role of centromeres in this process. I characterized the cell viability of DLD-1 cells in which I had inserted the (d)TET1$_{CD}$ constructs through Flp-In recombination by 14-days long clonogenic assays, with or without doxycycline pre-treatments (**Figure 26A**). I seeded naïve cells (**Figure 26B-I, IV**), or expressed the constructs for 8 days (**Figure 26B-II, V**) or for 16 days (**Figure 26B-III, VI**) prior to the seeding of the clonogenic assays. Cells were seeded in absence (NT or WO, top rows) or presence of doxycycline (+DOX, bottom rows). The expression of the active TET1$_{CD}$ construct, and consequent demethylation of the centromeres, affects the cell viability and decreases number of colonies formed (**Figure 26B-I, II, III**). The effect is stronger for naïve cells (**Figure 26B-I)**, with an average ratio of colonies formed in doxycycline versus colonies formed without treatment (+DOX/NT) of 0.138 (**Figure 26C**)**.** For both pre-treatment conditions cells were able to recover and form colonies when seeded in absence of doxycycline (washout, WO), and still exhibited reduced viability in doxycycline only when expressing the active TET1$_{CD}$ (**Figure 26B-II, III**). Based on COBRA measurements (**Figure 23A, C**), the centromeres of the cells seeded on (b) should have ~35% less methylation, and those on (c) ~60% less methylation than the cells on (a). The pre-treatments, however, do not have an increasingly stronger negative impact on the relative cell viability (**Figure 26C**)**.** On the contrary, pre-treated cells grow slightly better than naïve cells, with average colony formation ratios (+DOX/NT) of 0.271

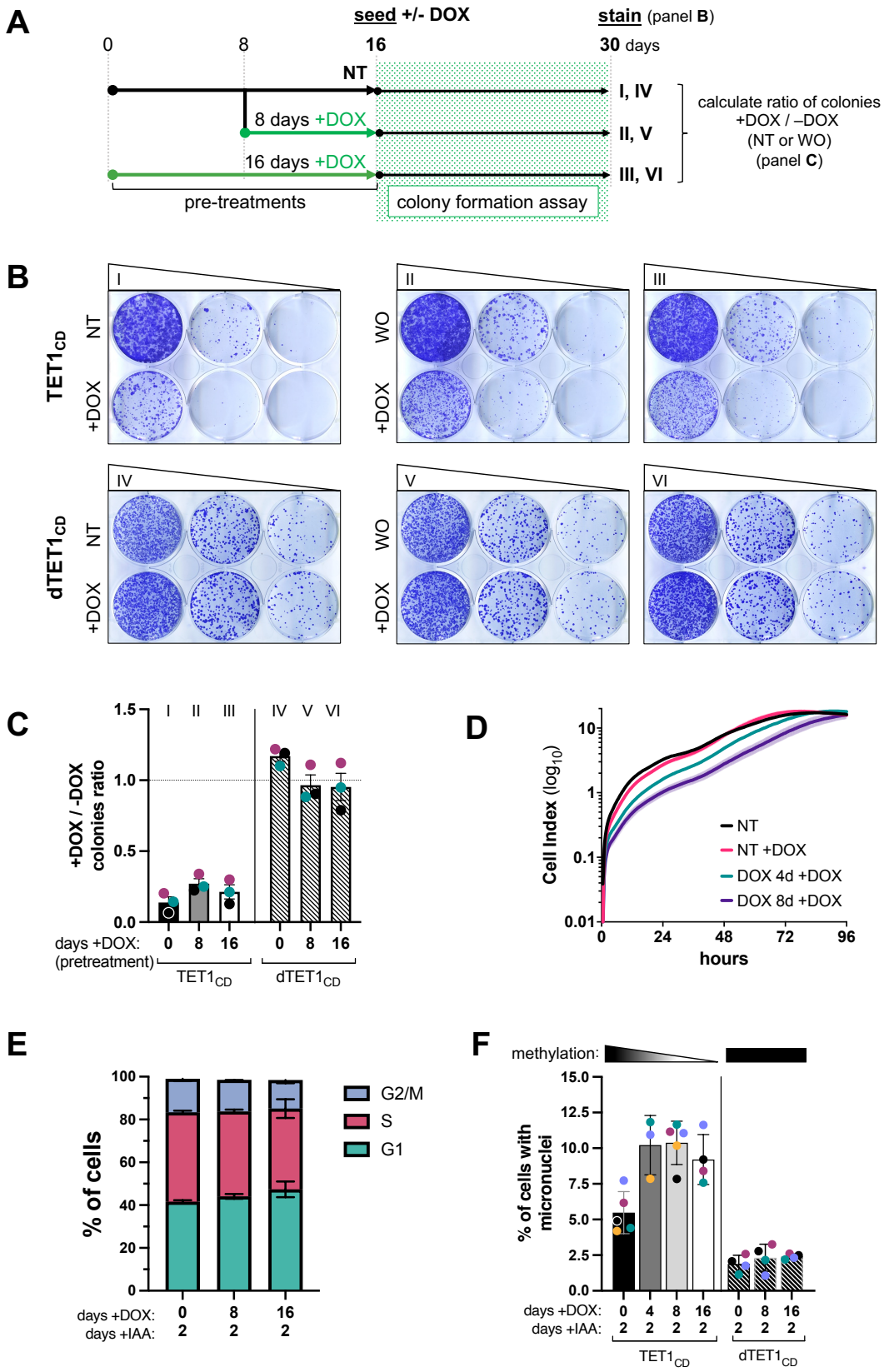

Figure 26. Chromosome mis-segregation causes decreased viability in cells with centromeric demethylation.

(A) Schematic of the clonogenic assays presented on (B) and quantified on (C). Cells were seeded without doxycycline (untreated [NT] or washout [WO], top rows) or with 100 ng/mL doxycycline (+DOX, bottom rows) on the first wells of 6 well plates and diluted laterally 1:5 (dilution represented

by the triangle). Cells were seeded without any pre-treatment (I, IV), after 8 days pre-treatment with doxycycline (II, V) or after 16 days pre-treatment with doxycycline (III, VI) **(B)** Representative images of colony formation assays stained with crystal violet 14 days after seeding. DLD-1 Flp-In $TET1_{CD}$ construct (I, II, III) or $dTET1_{CD}$ construct (IV, V, VI). **(C)** Quantification of the ratio of colonies formed with or without doxycycline from 3 experimental replicates **(D)** Real-time Cell Index measurement of DLD-1 Flp-In $TET1_{CD}$ naïve cells (NT), and naïve cells and cells pre-treated cells for 4 and 8 days seeded in presence of 1 μg/mL doxycycline (+DOX) for the duration of the assay (4 days). Curves represent the mean Cell Index value from triplicates ± SEM. **(E)** Cell cycle profiling by BrdU and PI incorporation. Two experimental replicates, error bars represent SEM. 100 ng/mL doxycycline. **(F)** Percentage of cells presenting micronuclei, quantified by immunofluorescence with DAPI staining. 3 to 5 colour coded experimental replicates per condition.

after 8 days of pre-treatment and 0.213 after 16 days pre-treatment. When partially-demethylated cells are seeded in absence of doxycycline (**Figure 26B-II, III,** top rows WO) they form comparable colonies to fully methylated cells. This result indicates that demethylation initially causes an important cell lethality but the cells that survive adapt to their new centromeric methylation status. If partially-demethylated cells are seeded again in presence of doxycycline (**Figure 26B-II, III,** bottom rows +DOX), there is additional cell lethality, possibly attributed to the continued demethylation of the surviving cells.

The expression of the CENP-B_{DBD} directed constructs will compete with the endogenous CENP-B for the binding of the CENP-B boxes. The possible reduced CENP-B occupancy of the centromeres and/or the binding of the ectopic construct at the CENP-B boxes could have deleterious effects on the centromeric function. Naïve and pre-treated cells expressing the $dTET1_{CD}$ construct have a similar colony formation capacity than untreated cells (**Figure 26B-IV, V, VI**) with colony formation ratios (+DOX/NT or +DOX/WO) close to 1 (**Figure 26C**). Given that the $dTET1_{CD}$ construct localizes to the centromeres (**Figure 22**) but has no effect on α-satellite methylation (**Figure 23B, C**) and does not affect cell viability, it can be concluded that the expression of CENP-B_{DBD} directed constructs by doxycycline induction does not have per-se a negative impact on the cells' viability. The reduced viability observed with the expression of the active $TET1_{CD}$ construct can therefore be considered a direct consequence of the centromeric demethylation.

I measured the proliferation rate of the DLD-1 Flp-In $TET1_{CD}$ cells in a real-time cell analyser (**Figure 26D**). Naïve cells seeded in presence of doxycycline (NT +DOX) exhibit a similar growth rate as their untreated counterparts (NT) during the four days of measurements. Cell pre-treated with doxycycline for four and eight days, on the other hand, present a mild and progressive reduction the proliferation rates. This observation supports the colony formation results. After 8 and 16 days of expression of the construct, the cell cycle profile remains almost normal, with only minor accumulation of cells in G1 to be reported (**Figure 26E**). Finally, I quantified by immunofluorescence the percentage of DLD-1 Flp-In (d)$TET1_{CD}$ cells with micronuclei as a readout of chromosome mis-segregation. DLD-1 cells are known to present a basal 3% to 5 % of cells with

micronuclei, values that matched my observations of both cell lines before induction of the constructs (**Figure 26F**). The percentage of cells expressing the active TET1$_{CD}$ construct that have micronuclei approximatively doubled (~10%) after 4 days and 8 days of induction, indicating that the demethylated centromeres fail to ensure a proper chromosomal segregation. After 16 days of induction there seems to be a mild decrease in the percentage of cells with micronuclei, possibly also a testament of the cellular adaptation observed on the clonogenic assays. Cells expressing the inactive dTET1$_{CD}$ construct had overall less micronuclei than their active counterparts at all timepoints assessed, and the percentage remained constant over time (**Figure 26F**), indicating that the binding of the constructs to the centromeres does not per se cause an increase in micronuclei formation. Overall, these results show that cells expressing the active the TET1$_{CD}$ construct (i.e., with centromeric demethylation), have progressively slower growth rates with almost normal cell cycle profiles, increased micronuclei formation and marked long-term cell lethality.

3.2. Increased construct expression through lentiviral delivery

We hypothesized that the slow rate of demethylation of the Flp-In cells could play a role in the cellular adaptation observed on the clonogenic assays and the overall mild cellular effects of the demethylation. The soft and progressive demethylation might be necessary to allow the cells enough time to adapt to having hypomethylated centromeres. Therefore, I sought to achieve faster demethylation rates to determine if cells would fail to adapt in this context. First, I tried to increase the expression of the constructs in the Flp-In cells by increasing the concentration of doxycycline. I titrated the doxycycline by measuring the α-satellite methylation levels by COBRA, using up to 50 times higher doses than the usual 100 ng/mL (**Figure 27A**). The demethylation rate remained constant with all the doses used and at all timepoints, indicating that the induction system was already at maximum capacity with the minimal doxycycline dose I was using (0.1 μg/mL ≡ 100 ng/mL).

To push the system, I had to turn to lentiviral transductions. I generated a new bicistronic TET1$_{CD}$ construct, adding to its C-terminus two tandem 2A self-cleaving peptides (tPT2A) followed by a fluorescent green protein (EGFP). I introduced this construct by lentiviral transduction into the same parental DLD-1 cell line used for Flp-In recombination (**Figure 27B**). The 2A peptides induce ribosome skipping, separating the demethylation construct from the fluorescent protein during translation. Only 17 additional amino acids should remain attached to the TET1$_{CD}$ construct, making the construct that will bind to the centromeres comparable to the one from the Flp-In cells. After transduction and a short antibiotic selection, I induced the expression of the construct overnight by adding doxycycline in presence of IAA to degrade the TET1$_{CD}$ moiety. In this manner I could detect the expression of the second cistron (EGFP) while not subjecting the cells to untimely centromeric demethylation. I sorted two populations of cells according to their EGFP expression level: the

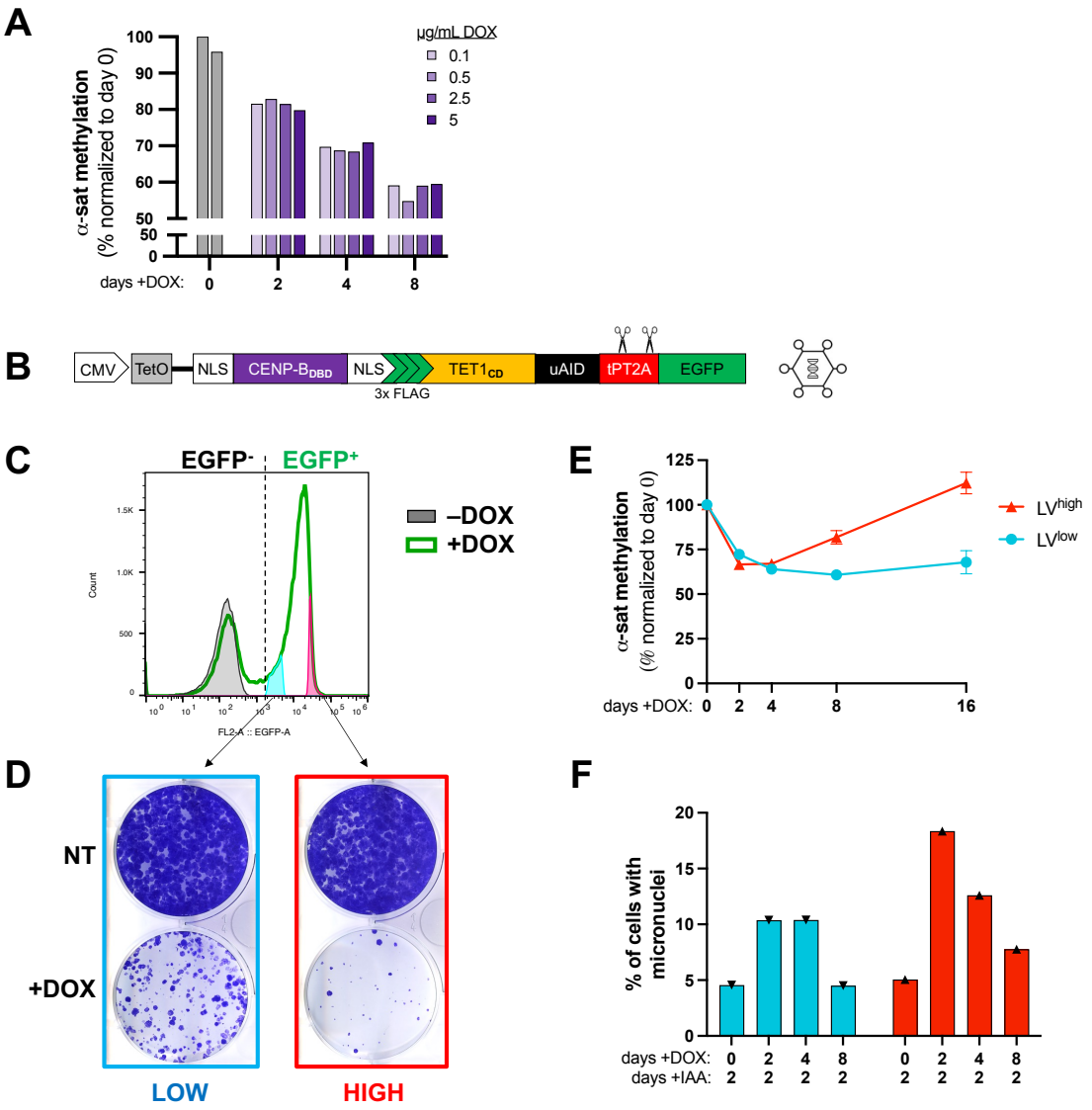

Figure 27. Centromeric demethylation rate-dependent effect on cell viability.
(A) Doxycycline titration by COBRA. DLD-1 Flp-In TET1$_{CD}$ cells were induced with 0.1, 0.5, 2.5 or 5 µg/mL of DOX for 2, 4 and 8 days. Quantification of the α-satellite methylation levels measured by COBRA and normalized to day 0 (no induction) **(B)** Schematic of the bicistronic lentiviral construct integrated into DLD-1 cells **(C)** Gating strategy for the sorting of DLD-1 LV TET1$_{CD}$ cells expressing low (LV^{low}) or high (LV^{high}) levels of the construct **(D)** Representative images of colony formation assays (14 days) with crystal violet staining. 100 ng/mL DOX **(E)** Quantification of the α-satellite methylation levels by COBRA after 2, 4, 8 and 16 days of induction with 100 ng/mL DOX and normalized to day 0 (no induction) **(F)** Percentage of cells presenting micronuclei at each timepoint quantified by immunofluorescence with DAPI staining.

bottom ~10% of the EGFP^{+} cells (LV^{low}), and the brightest ~10% of cells (LV^{high}) (**Figure 27C**). In a bicistronic construct the second protein (after the 2A peptide) can only be translated as well as the first protein, so it is safe to consider the level of EGFP expression as a readout of the level of the TET1$_{CD}$ construct. After recovery from the sorting and expansion (with constant IAA on the culture media to ensure the degradation of the demethylating moiety) I seeded the cells for clonogenic assays

in absence (NT, top rows) or presence of doxycycline (+DOX 100 ng/mL, bottom rows) (**Figure 27D**). The LV^{high} cells formed far fewer colonies than the LV^{low}. I next treated the two populations of cells with doxycycline for 2, 4, 8 and 16 days and quantified the centromeric methylation levels by COBRA (**Figure 27E**). The demethylation kinetics of the two populations are quite different. The LV^{low} population exhibits a slow constant demethylation over the first few days (0-4 days), that stagnates (4-8 days) and mildly recovers in the long term (16 days). The demethylation of the LV^{high} population, on the other hand, seems to peak at 2 days, and gradually and completely recovers over time. The levels of demethylation reached by the two populations within the first four days are surprisingly similar. The apparent recovery of the centromeric methylation of the LV^{high} population probably is also a reflection of the cell lethality observed on the clonogenic assay. The cells that had the highest levels of demethylation probably died, and therefore are lost from the pool of cells in which the methylation is measured; this biases the results towards higher methylation levels.

Finally, I also quantified the percentage of cells that have micronuclei at each timepoint by immunofluorescence microscopy (**Figure 27F**). The LV^{low} population doubles the percentage of cells presenting micronuclei after two and four days of demethylation and have a total recovery at eight days. The population expressing high levels of the construct, on the other hand, exhibits almost four times more micronuclei after two days demethylation, more than double at day four, and remain with elevated but closer-to-normal micronuclei levels at eight days. Overall, these results are in line with the idea that cells that have their centromeres demethylated too fast (LV^{high}) exhibit high levels of genome instability, as seen by the burst in micronuclei formation after only 2 days of expression of the construct, do not manage to adapt and die. A milder, more gradual centromeric demethylation, on the other hand (LV^{low}), gives the cells enough time to adjust to this new condition and survive. Further experiments are required to confirm these observations, with the need of including the catalytically inactive $dTET1_{CD}$ expressed at high levels to rule out that the phenotype observed on the (LV^{high}) is due to secondary effects from the overexpression. I am also isolating single-cell clones from each of these populations, to test these observations without the confounding effects of the population-recovery.

4. Molecular effects of a centromere specific demethylation

4.1. Centromeric DNA hypomethylation increases CENP-A and CENP-B levels at the centromeres.

I investigated the effect of the centromeric specific, gradual hypomethylation on the abundance of CENP-A, CENP-B and CENP-C at the centromeres though indirect immunofluorescence microscopy of the DLD-1 Flp-In (d)$TET1_{CD}$ cells (**Figure 28**). The constructs were expressed for 8 or 16 days and degraded for 48h by auxin treatment before cell fixation; the centromeres were hence

free of exogenous construct for roughly two cell cycles before proceeding with the immunofluorescence, removing the possibility of competition or steric hindrance from the construct to the centromeric proteins. The analysis of cells in interphase (**Figure 28A, B**) revealed that increasing levels of demethylation lead to a significant rise of CENP-A levels, by 9% on average at 8 days and reaching a 20% increase at 16 days. The effect of the demethylation is strongest on CENP-B, increasing at the centromeres by 44% at 8 days and 60% at 16 days. CENP-C, on the contrary, remains surprisingly unaffected by the changes in centromeric DNA methylation (**Figure 28B**). All three protein levels remain constant over time on the cells expressing the (d)TET1$_{CD}$; the observed protein increases at the centromeres in the active cell lines can be therefore attributed to DNA methylation changes, and any side effect of the prolonged binding of the exogenous constructs to the CENP-B boxes can be cast-off.

CENP-B staining was peculiar upon hypomethylation, with large round foci appearing over time (**Figure 28A**). To assess if the protein increase measured in interphase was due to clustering of the centromeres, as we have previously reported (Chardon et al., 2022), I measured the fluorescence intensity of the three proteins in metaphase chromosomes after 16 days induction of the constructs (**Figure 28C**). I confirmed that only when the centromeres are demethylated (i.e., when the active TET1$_{CD}$ is expressed), CENP-A intensity at the centromeres is increased by 14% on average, CENP-B by 50% and CENP-C remains constant, therefore refuting the clustering hypothesis. Given its known higher affinity for unmethylated CENP-B boxes (**Figure 28B**), an increase in CENP-B was anticipated in the hypomethylated context. The fact that CENP-A levels also increase while CENP-C levels, which usually correlate tightly with CENP-A levels, do not, was unexpected.

The results obtained with the centromeric specific hypomethylation system match those obtained with the DNMT1^{AID} cell line. Regardless of the method employed to demethylate the centromeres, CENP-A and CENP-B levels at the centromere increase.

4.2. Once demethylated, the centromeres remain demethylated and CENP-A and CENP-B levels remain elevated.

Given that these cells have intact de novo and maintenance DNA methylation pathways, I investigated if a long-term degradation of the active TET1$_{CD}$ construct after 16 days expression and marked hypomethylation of the centromeres could result in a recovery of the methylation and the consequent rescue of the CENP-A and CENP-B levels. I measured by COBRA the α-satellite methylation levels after expressing the construct (+DOX) for 8 or 16 days, followed by 2, 8 and 16 days of degradation (+IAA) (**Figure 29A, B**). I detected a very minor recovery of methylation with the longer degradation times, which could perhaps be attributed to the methylation of sites that were

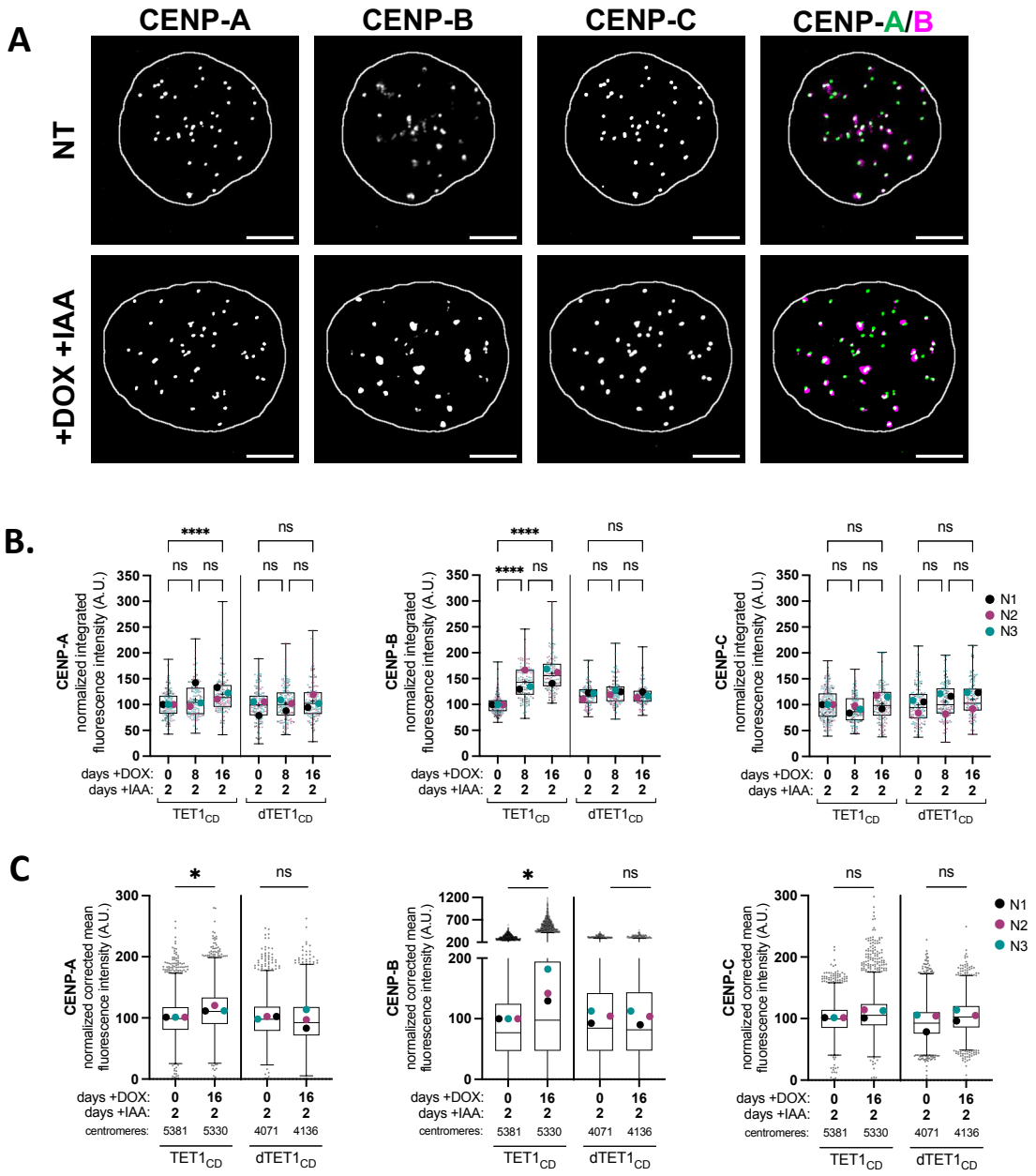

Figure 28. Centromeric demethylation increases CENP-A and CENP-B levels at the centromeres.

(**A**) Representative images of immunofluorescence in interphase Scale bar: 5 μm (**B**) Quantification of the normalized integrated fluorescence intensity of CENP-A, CENP-B and CENP-C at the centromeres of cells in interphase. Each dot represents the average intensity of all the centromeres of a single cell. The larger dots mark the average of one experimental replicate. N>110 cells and >2775 centromeres per condition. Kruskal-Wallis test with Dunn's multiple comparisons test. ns: non-significant differences. (**C**) Tukey bar plot of the normalized mean fluorescence intensity of CENP-A, CENP-B and CENP-C in metaphase centromeres. Each dot represents a centromere, N>4071 centromeres per condition. The larger coloured dots mark the average of each experimental replicate. Unpaired t-test defines the statistics between replicate averages. ns: non-significant differences

hemi-methylated after the active demethylation. Despite this minor increase, the centromeric methylation remains reduced by more than 70% after 8 and 16 days of degradation of the construct, indicating that there is essentially de novo methylation activity in these cells capable to recover the centromeric methylation level once it is lost. I next measured the abundance of CENP-A, CENP-B and CENP-C at the centromeres though indirect immunofluorescence microscopy in interphase after the prolonged degradation of the construct. To consider the possible effects of the extended time in culture, the treatments were staggered to collect and fix all samples at the same endpoint (**Figure 29C**). I also measured the abundance of the proteins after 30 days of continuous expression of the construct. Based on the more than 85% demethylation rate achieved at 16 days (as quantified by Nanopore sequencing, **Figure 24C**), I did not expect a longer time expressing the constructs would have further impact on centromeric methylation levels and therefore on the abundance of the centromeric proteins. In this set of experiments, after 16 days of expression of the $TET1_{CD}$ construct and 2 days of degradation, CENP-A increased by 25% on average. The prolonged expression of the construct for 30 days followed by 2 days of degradation led to a milder increase of 13%, which is close to the 10% increase in CENP-A levels detected after 16 days of expression of the construct and 16 days of degradation (**Figure 29D,** first graph). CENP-B exhibits a 64% increase after 16 days of expression and 2 days of degradation of the construct, and remains stable with a 68% increase measured after 30 days of continuous expression of the $TET1_{CD}$ followed by 2 days of degradation. The long term, 16 days degradation of the construct after 16 days of demethylation, led to a milder 44% increase of CENP-B levels (**Figure 29D,** second graph). As before, no matter the time of expression or degradation of the $TET1_{CD}$ construct, CENP-C levels remained stably unchanged (**Figure 29D,** third graph). Collectively these experiments demonstrate that once the centromeres are hypomethylated, there is very little recovery or de novo DNA methylation, and that this new, stable epigenetic status is translated into an equally stable increase of CENP-A and CENP-B levels. These results further support the notion that, provided that the hypomethylation occurs gradually, a process of cellular adaptation and remodelling of the centromeric protein landscape takes place, ensuring the cell survival in these new conditions.

4.3. A mild overexpression of CENP-B, without affecting the centromeric methylation, mostly recapitulates the cellular demethylation phenotype.

The increase in centromeric CENP-B levels could directly lead to centromere dysfunction. Through its dimerization, CENP-B has been seen to form loops and promote inter-chromosomal contacts, increasing compaction and clustering (Chardon et al., 2022). It is plausible that excessive CENP-B could be therefore detrimental for the centromeric function by generating disproportionate secondary structures that could be hard to resolve upon replication, and that could derive in centromeric fragility. To test if CENP-B is the direct responsible of the chromosome segregation

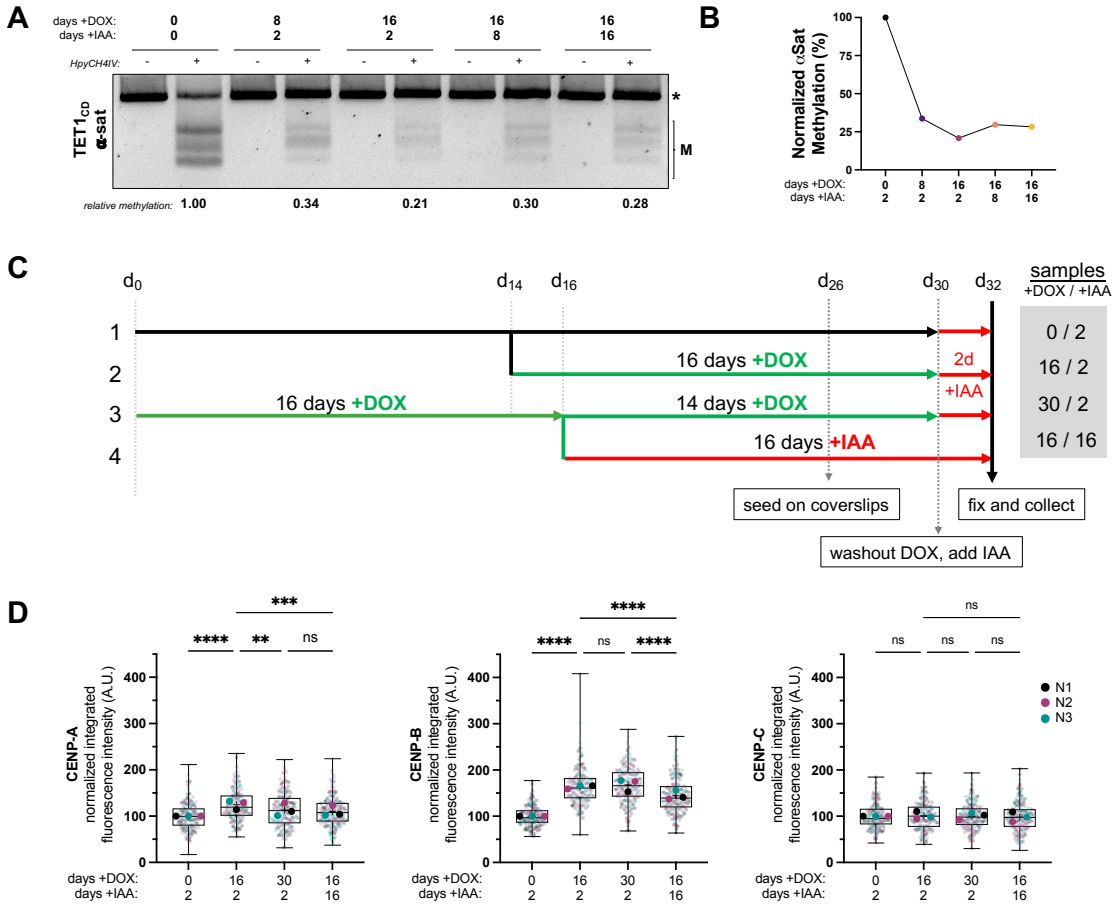

Figure 29. The centromeric demethylation is not reversible and leads to a persistent increase in CENP-A and CENP-B levels at the centromeres.
(A) α-satellite COBRA agarose gel at different times of demethylation (+DOX) and degradation of the construct (+IAA). The absence (–) or presence (+) of the methylation-sensitive enzyme is indicated. The asterisk (*) demarks the position of the unmethylated (undigested) PCR fragment, M indicates the methylated (digested) fragments. (**B**) Graphical representation of the normalized methylation from (A) (**C**) Schematic of the experimental procedure for the immunofluorescence microscopy measurements. (**D**) Box and whiskers plot (min to max) of the normalized integrated fluorescence intensity of CENP-A, CENP-B and CENP-C at the centromeres of cells in interphase. Each small dot represents the average intensity of all the centromeres of a single cell, color-coded per experimental replicate. The larger dots mark the average of all the cells from one experimental replicate. N>53 cells per replicate and condition. Statistics defined by Kruskal-Wallis test with Dunn's multiple comparisons test. ns: non-significant differences.

defects and cell lethality observed in the hypomethylated cells, I overexpressed full length CENP-B in a doxycycline inducible manner to levels similar of those detected on the hypomethylated cells but without altering the methylation of the centromeres. By titrating the doxycycline, I achieved an average 53 ± 7% increase in CENP-B levels after 4 days (**Figure 30A**), a value similar to the 60% increase obtained with the $TET1_{CD}$ demethylation system (results sections 4.1 and 4.2). After 4 days I observed a slight increase of the percentage of cells with micronuclei, going from ~4% in normal CENP-B levels to 7% in the cells with CENP-B overexpression (**Figure 30B**). Colony formation

assays revealed that cells that overexpress CENP-B form on average 70% less colonies than their wild-type counterparts, something easy to visualize at lower cellular densities (**Figure 30C, D**). This experiment revealed that a CENP-B increase, in the range of what is attained upon hypomethylation, is sufficient to partially recapitulate the two main aspects of the cellular phenotype of demethylated cells: the increased micronucleation and the cell lethality. Upon hypomethylation, however, the percentage of cells with micronuclei is higher (**Figure 26F**) and the cell viability is lower (**Figure 26C**), indicating that another component must also be playing a role in the centromeric dysfunction.

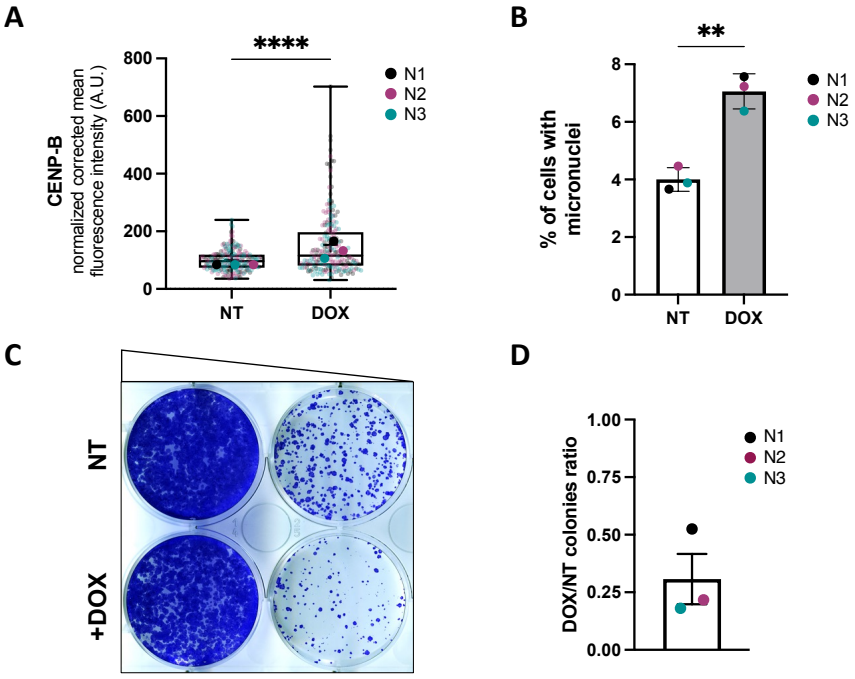

Figure 30. CENP-B overexpression without affecting centromeric methylation recapitulates the hypomethylation phenotype.
(**A**) Box and whiskers plot (min to max) of the normalized mean fluorescence intensity of CENP-B at the centromeres of cells in interphase, untreated (NT) or treated for 4 days with 2.5 ng/mL doxycycline. Each small dot represents the average intensity of all the centromeres of a single cell, color-coded per experimental replicate. The larger dots mark the average of all the cells from one experimental replicate. N>60 cells and >1390 centromeres per replicate and condition. Mann-Whitney test defines the statistics. (**B**) Percentage of cells with micronuclei, quantified by immunofluorescence with DAPI staining. Welch's t-test defines the statistics. (**C**) Representative image of a colony formation assay of untreated cells (NT) and cells treated with 2.5 ng/mL doxycycline to overexpress CENP-B for 14 days. Crystal violet staining. (**D**) Quantification of the ratio of colonies formed with or without doxycycline from 3 experimental replicates.

4.4. A mild overexpression of CENP-A, without affecting the centromeric methylation, does not recapitulate the demethylation phenotype.

As for CENP-B, the increase in centromeric CENP-A levels upon hypomethylation could be the direct responsible of the centromere dysfunction, increased formation of micronuclei and cell lethality observed in the hypomethylated cells. To address this hypothesis, I overexpressed full length CENP-A on DLD-1 cells to levels similar of those detected on the hypomethylated cells but without altering the methylation of the centromeres. After titration of doxycycline, I achieved a 31 ± 4% mean increase in CENP-A levels in interphase after 4 days of induction, a value comparable to the 20% increase occurring upon hypomethylation (**Figure 31A**). I did not observe a significant increase in micronuclei after four days (**Figure 31B**) and cell viability is maintained with prolonged overexpression of CENP-A to these levels (**Figure 31C, D**). Overall, an increase of CENP-A to levels comparable of those achieved upon hypomethylation, is not in itself sufficient to recapitulate the cellular phenotype of observed when centromeres are demethylated.

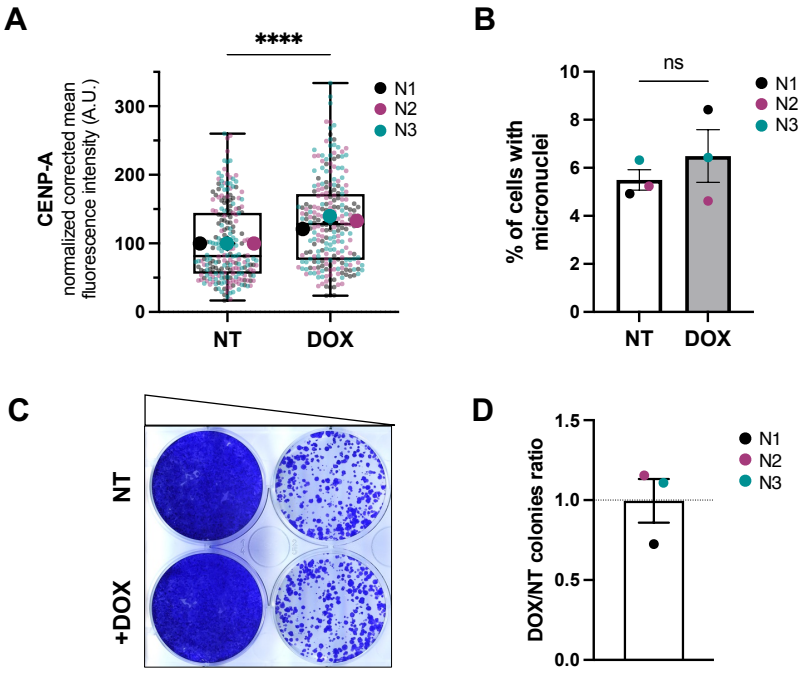

Figure 31. CENP-A overexpression without affecting the centromeric methylation does not recapitulate the hypomethylation phenotype.
(A) Box and whiskers plot (min to max) of the normalized mean fluorescence intensity of CENP-A at the centromeres of cells in interphase, untreated (NT) or treated for 4 days with 0.5 ng/mL doxycycline. Each small dot represents the average intensity of all the centromeres of a single cell, color-coded per experimental replicate. The larger dots mark the average of all the cells from one experimental replicate. N>71 cells and >2667 centromeres replicate and condition. Mann-Whitney test defines the statistics. **(B)** Percentage of cells with micronuclei, quantified by immunofluorescence with DAPI staining. Welch's t-test defines the statistics. ns: non-significant differences **(C)** Representative image of a colony formation assay of untreated cells (NT) and cells treated with 0.5 ng/mL doxycycline to overexpress CENP-A for 14 days. Crystal violet staining. **(D)** Quantification of the ratio of colonies formed with or without doxycycline from 3 experimental replicates.

4.5. The hypomethylation of the centromeres does not alter centromeric transcription.

One of the best known and most studied roles of DNA methylation is to repress transcription. We hypothesized that the cellular effects observed on the hypomethylated cells could be at least partially the result of an increase in centromeric transcription, which could be deleterious by causing transcription-replication conflicts or the accumulation of DNA-RNA hybrids. It was recently reported that CENP-B promotes the centromeric localization of a transcriptional regulator called ZFAT, claimed to be required for the centromeric ncRNA transcription, and that the ectopic expression of CENP-B induces the accumulation of ZFAT at the centromere (Ishikura et al., 2021). The increase in CENP-B levels measured in these cells could therefore also be promoting an excessive centromeric transcription. I performed quantitative polymerase chain reactions after reverse transcription (RT-qPCR) to measure the level of α-satellite transcripts after doxycycline induction of the (d)TET1 constructs in DLD-1 Flp-In cell lines (**Figure 32**). In five biological replicates and using two primer pairs that detect transcripts of centromeres 1, 5, 19, and centromeres 1, 3, 10 respectively, I found a minor but statistically significant increase in centromeric transcripts over time for both the active (1.45-fold increase) and the inactive (1.35-fold increase) $TET1_{CD}$ cell lines. The transcriptional increase is independent of the demethylation of the centromeres, and hence cannot be at the root of the phenotype observed on the hypomethylated cells. Further experiments are required to understand what the cause of this increase is.

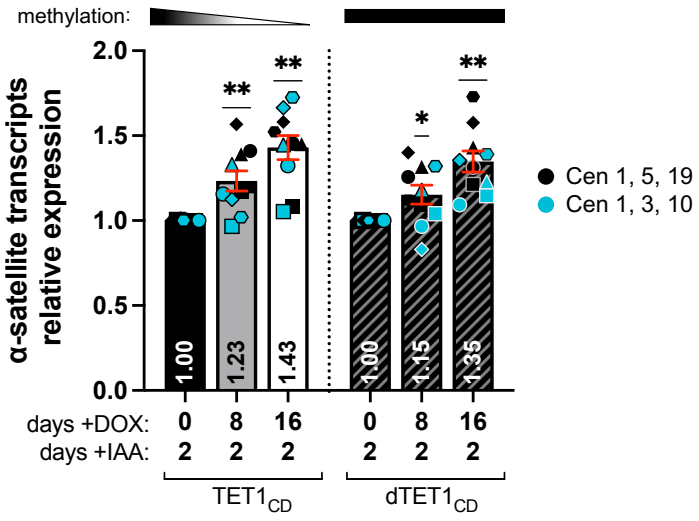

Figure 32. Hypomethylation-independent increase of the α-satellite transcription.
Quantification by qRT-PCR of the α-satellite transcript levels using primers targeting centromeres 1,5,19 (in black) and centromeres 1,3,10 (in cyan). The data is normalized to the corresponding day 0 (fully methylated). Each symbol represents the average of the normalization of three technical replicates against three reference genes (GUSB, PPIA and RPLP0) for each one of five biological replicates (represented by the different symbol shapes). The bars represent the SEM. Statistics defined by one-sample Wilcoxon signed-rank test.

4.6. The hypomethylation of the centromeres does not alter the centromeric compaction.

The precise role of DNA methylation in the structure of the chromatin is still a matter of debate, but most studies point towards an increased compaction of the nucleosomal and linker DNA when the DNA is methylated. It is therefore reasonable to hypothesize that the targeted centromeric hypomethylation could decrease centromere compaction. Indeed, one of the cytological hallmarks of the ICF syndrome is the decompaction of chromosome 1 at the juxta centromeric 1qh region (Ehrlich, 2003). This phenotype can mostly be attributed to a pericentromeric decompaction, but a centromeric component has not precisely been tested. On the other hand, and as mentioned in section 4.3, we have recently reported that through its dimerization CENP-B can form DNA loops that increase centromere compaction and clustering (Chardon et al., 2022). I have extensively demonstrated that CENP-B is increased upon centromeric hypomethylation; it is therefore also reasonable to hypothesize that the centromeric compaction can be increased in this context.

To address these hypotheses, I compared the compaction of the centromeric DNA of the DLD-1 Flp-In cell line, before and after demethylation through doxycycline induction of the active $TET1_{CD}$ construct, by performing fluorescent in situ hybridizations (FISH) with probes against α-satellite as described in Chardon et al., 2022. The measure of the circularity and area of the α-satellite DNA are two parameters that can be used to describe the level of compaction of the centromeres (**Figure 33A**). I analysed thousands of centromeres in three independent replicates. While the area presented a minor tendency to increase (**Figure 33B**) (0.301 μm^2 at day 0; 0.308 μm^2 at day 8 and 0.314 μm^2 at day 16), which could indicate decompaction, the average circularity also increased marginally (0.855 at day 0; 0.862 at day 16), which would indicate an enhanced compaction (**Figure 33C**). The average number of α-satellite foci detected per nucleus remained constant (21.5 at day 0; 22.2 at day 8 and 21.9 at day 16), indicating that there are no changes in the clustering of the centromeres (**Figure 33D**). Biologically speaking there does not seem to be any relevant changes in centromeric compaction upon hypomethylation. This could be explained by the theoretical decompaction induced by the DNA demethylation opposing the compaction induced by CENP-B increase. It could also be that there is no significant change induced by either, which would imply that the ~60% CENP-B increase by is not sufficient to increase the compaction the centromeric DNA. Two independent pieces of evidence support this last hypothesis. First, the report that mouse cells demethylated by means of treatment with 5-aza-2'-deoxycytidine present a redistribution of CENP-B without any changes to the degree of condensation of the centromeric chromatin (Mitchell et al., 1996). Second, the fact that the previous experiments from our team, showing that CENP-B overexpression increases compaction, were all performed in much higher overexpression levels. To provide a definitive answer and isolate the direct effects of the DNA demethylation on the centrochromatin structural properties, I envision to assess the centromeric compaction upon hypomethylation in a CENP-B KO background.

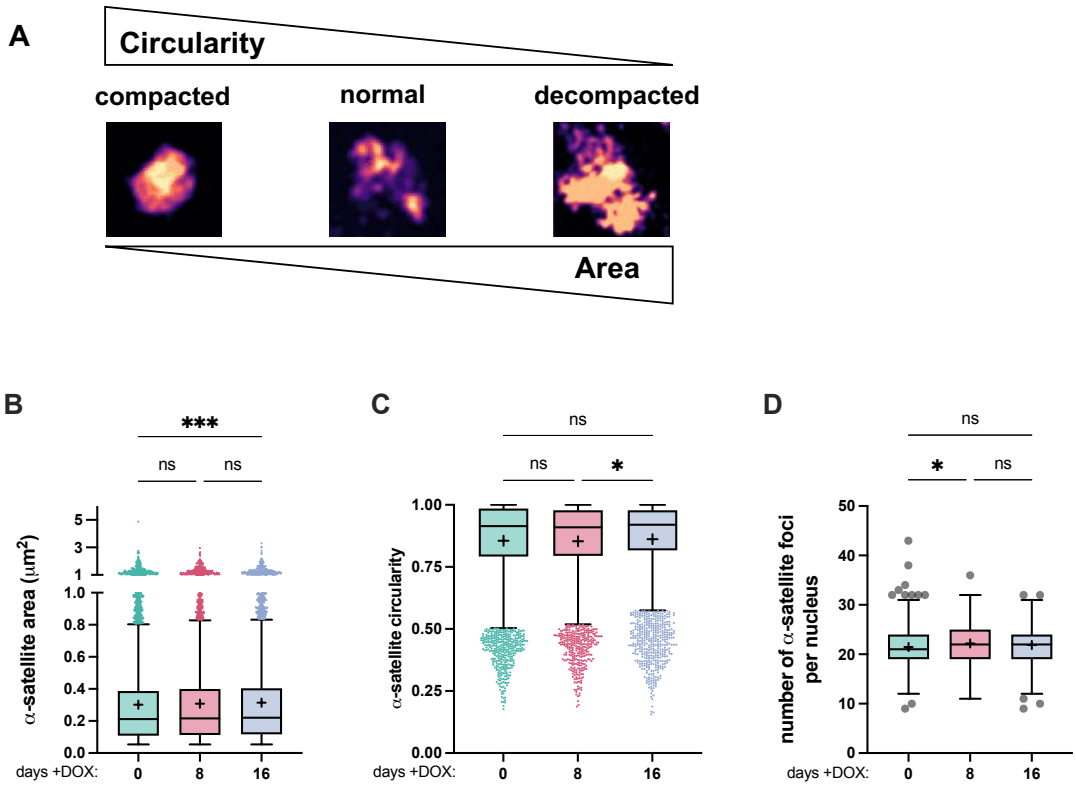

Figure 33. The DNA hypomethylation does not alter the centromere compaction.
(A) Illustration of the relationship between the physical parameters of circularity and area used to describe the α-satellite DNA compaction ranging from: compacted (high circularity, low area) to decompacted (low circularity, high area). Tukey box plots showing area **(B)**, circularity **(C)**, and number **(D)** of α-satellite foci in DLD-1 Flp-In TET1$_{CD}$ cells at the indicated times of induction of the construct with 100 ng/mL doxycycline. The central line of the box marks the median, the (+) marks the mean. Three experimental replicates, with >7097 centromeres and >320 cells analyzed in total per timepoint. Statistics defined by Kruskal-Wallis test with Dunn's multiple comparisons test. ns: non-significant differences.

4.7. CENP-A reloading is impaired upon centromere hypomethylation.

A key event in centromere maintenance is the correct loading of CENP-A in early G1 through the self-assembly epigenetic loop (Jansen et al., 2007).We hypothesize that the DNA methylation central dip region observed at the centromeres on the T2T sequencing can be an additional layer of epigenetic signalling, demarking the boundaries of the domain where newly CENP-A is to be deposited. Were this hypothesis correct, we anticipate CENP-A deposition to be affected by the loss of the boundary upon centromeric demethylation. To test this theory, I combined the centromere-specific demethylation strategy (TET1$_{CD}$ construct) with the CENP-A$^{OFF/ON}$ technology, previously developed in our laboratory (Hoffmann et al., 2020) (**Figure 34A**). Upon auxin addition, CENP-A is degraded, and after auxin washout, cells re-express and reload CENP-A. I introduced the TET1$_{CD}$ construct, through Flp-In recombination or lentiviral (LV) transduction, into a DLD-1 cell line that has an N-terminal EGFP and AID tagged CENP-A ($^{EYFP-AID/-}$ CENP-A), and isolated single cell

clones from each. Upon doxycycline addition for 4 days, the TET1 construct demethylates the centromeres by 27.3% for the best Flp-In clone and by 49% for the best LV clone (**Figure 34B**). Through immunofluorescence staining of ^{EYFP}CENP-A, ACA and CENP-C (**Figure 34C**), I counted the number of EYFP (CENP-A) positive centromeres per cell and calculated the percentage of cells that were able to fully or partially reload CENP-A after 48h of auxin washout (**Figure 34D**). In the normally methylated condition, more than 96.7% of cells fully reload CENP-A, as was previously reported (Hoffmann et al., 2020). On the contrary, between 25% (Flp-In) and 36% (LV) of the cells fail to reload CENP-A to at least 15 of their centromeres when the experiment is performed in cells that have hypomethylated centromeres (+DOX for 4 days) (**Figure 34D**). After the combined treatments, the cells have similar cycle profiles as the ones with IAA washout alone, as seen by flow cytometry after 24h BrdU incorporation (**Figure 34E**). Therefore, failure to reload CENP-A is not caused by the cells not going through G1. The lower the centromeric DNA methylation level is (LV vs. Flp-In), the stronger the impact in CENP-A reloading is. These results demonstrate that centromeric methylation is important for the loading of CENP-A to the centromeres, possibly by signalling the correct position for new CENP-A to be deposited.

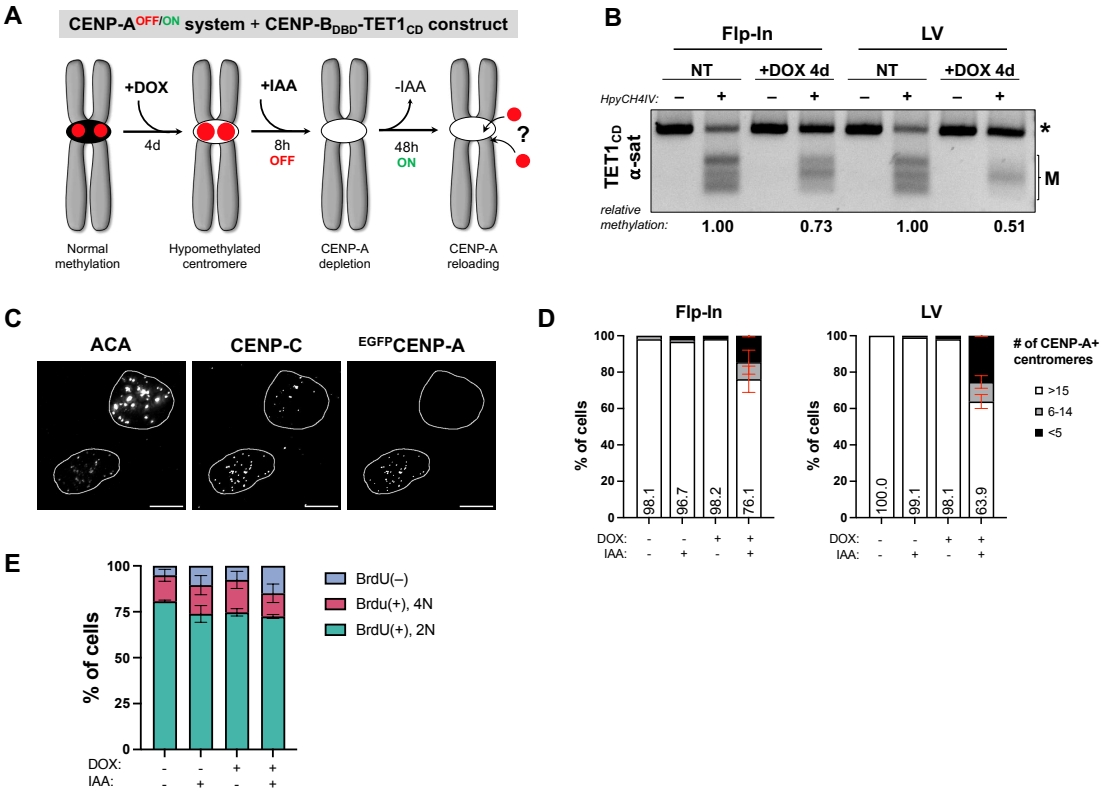

Figure 34. CENP-A reloading is impaired by centromeric DNA demethylation.
(A) Schematic of the combined TET1$_{CD}$ targeted demethylation and CENP-A$^{OFF/ON}$ system. Doxycycline (+DOX) addition leads to centromeric demethylation, auxin addition (+IAA) causes rapid CENP-A degradation and auxin washout (-IAA) allows reloading of new CENP-A into the hypomethylated centromeres. **(B)** α-satellite COBRA agarose gel of two TET1$_{CD}$ clones, one obtained

by Flp-In recombination and another after lentiviral (LV) transduction, in untreated conditions (NT) and after 4 days of expression of the construct with doxycycline. The absence (–) or presence (+) of the methylation-sensitive enzyme is indicated. The asterisk (*) demarks the position of the unmethylated (undigested) PCR fragment, M indicates the methylated (digested) fragments. The relative methylation of each sample compared to untreated is indicated. **(C)** Representative immunofluorescence image of after the targeted demethylation and CENP-A degradation and re-expression. Staining with ACA, anti-CENP-C and anti-GFP antibodies. Scale bar: 10 μm. The top-right cell failed to reload CENP-A and is likely more demethylated than the bottom-left cell, as evidenced by the intense ACA staining -a reflection of the important CENP-B increase upon hypomethylation. **(D)** Percentage of cells that have less than 5, between 6 and 14, or more than 15 centromeres with CENP-A positive staining. **(E)** Cell cycle profiling after 24h BrdU and PI incorporation. Two Flp-In clones and one LV clone combined. Error bars represent SEM.

4.8. Upon centromeric hypomethylation, CENP-A domains expand.

The increase in CENP-A levels detected by immunofluorescence at the hypomethylated centromeres (**Figure 28**) can be the reflection of a higher density of CENP-A nucleosomes within the same region or could indicate that CENP-A expands beyond its canonical domain, with CENP-A nucleosomes being deposited in newly hypomethylated positions. The results from the CENP-A overexpression experiments (**Figure 31**) indicate that the molecular phenotype of the hypomethylated cells is not just a matter of having more CENP-A nucleosomes deposited in the same region. We hypothesize that the methylation dip observed on the T2T assembly of centromeres at the CDR (centromere dip region) exists because DNA methylation is there to act as a boundary to delimit the centromeric domain where CENP-A can be/is localized. In line with this idea, we also hypothesize that the centromeric specific demethylation can possibly enlarge the CDR, and that this could in turn allow for an expansion of the domains occupied by CENP-A. This expansion could potentially weaken the centromere to kinetochore attachments and be upstream of the chromosome segregation errors, micronuclei formation, and cell lethality observed in the cells with centromeric hypomethylation. The microscopy-based immunofluorescence assays performed do not provide sufficient resolution to resolve this question; we therefore turned to sequencing methods. As extensively mentioned throughout this work, the full centromeric sequences of DLD-1 cells are not assembled, posing major constraints to mapping-based analyses. Nevertheless, mapping DLD-1 centromeric reads –in particular those of centromere 15– to the T2T-CHM13v1.0 reference genome, allows for an unprecedented mapping accuracy for this cell line, and despite not being perfect, can still provide valuable information.

I performed a CUT&RUN experiment on DLD-1 Flp-In TET1$_{CD}$ cells, before (NT) and after 16 days of doxycycline induction of the construct (+DOX) followed by two days degradation with auxin, to map the CENP-A interactions with the centromeric DNA. The DNA was sequenced with Illumina NovaSeq 6000. We started by performing a global counting of all the reads aligning within any of the genomic coordinates that contain both homogeneous HORs and monomeric/divergent α-sat and that are herein considered as centromeric reads. The demethylated cells present more than a three-

fold enrichment in centromeric reads compared to the untreated, fully methylated cells (**Figure 35A**). This is complementary evidence of the global CENP-A increase at centromeres upon hypomethylation but does not provide any information on if and how the enrichment is distributed along the α-sat array. We therefore focused on the alignment of reads at the centromere of chromosome 15 and determined the CENP-A enrichment profile over IgG (**Figure 35B**). Further, a peak calling analysis allowed to identify significantly enriched domains within the 3 Mb window encompassing the full centromere 15. In the untreated sample, the significantly CENP-A enriched domain coincides, with only one of the three HORs (S4C15H1L, in green), which is defined on the T2T-CHM13v1.0 reference genome as the active HOR of CHM13 cells. In the demethylated sample, the CENP-A enrichment in that same HOR is slightly higher, and more importantly, the significantly enriched domain stretched to a neighbouring HOR (S4C15H2), unoccupied by CENP-A in the normally methylated cells and annotated as inactive on the reference genome.

These results indicate that upon centromeric hypomethylation, and at least for centromere 15, CENP-A not only increases its levels at the same domain, but also expands to a previously unoccupied and inactive HOR. This strengthens the notion that the centromeric DNA methylation acts as a signal to delimit CENP-A position. Due to mapping limitations, we cannot generalize these observations to other centromeres on this cell line with confidence; further proofs are therefore needed to strengthen these conclusions.

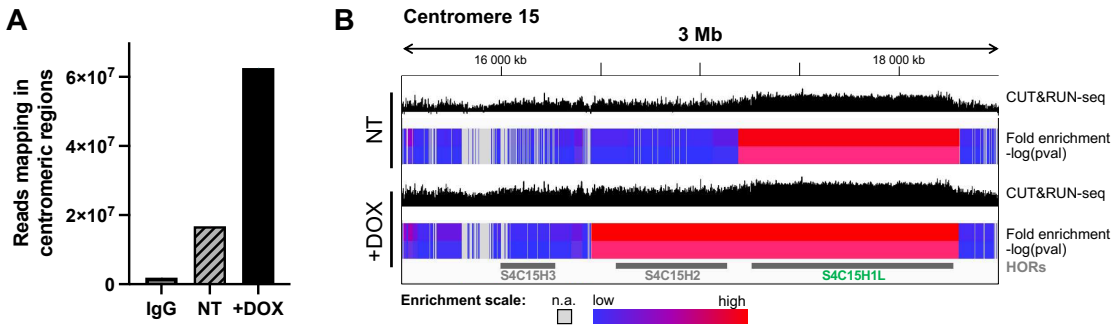

Figure 35. CENP-A domains expand upon hypomethylation.
(A) Total reads mapping to centromeric regions on T2T-CHM13v1.0 reference genome. IgG control and CENP-A CUT&RUN on DLD-1 Flp-In TET1$_{CD}$ before (NT) and after 16 days of doxycycline induction of the construct (+DOX) followed by two days degradation with auxin. The otal number of reads per sample was normalized. **(B)** Enrichment profile and identification of enriched domains on centromere 15. The CENP-A enrichment (CUT&RUN-seq, black tracks) is plotted as log2 ratio compared to IgG in 2-Kb wide genomic bins. Enriched domains are identified by peak calling and are presented as fold-enrichment and as -log(pval) on a scale from blue (low enrichment significance) to red (high enrichment significance); grey lines indicate no detected peak. HOR boundaries on centromere 15 are depicted as grey bars.

5. Slow demethylation kinetics and cellular adaptation of an ICF model cell line

The ICF syndrome presents a unique physiological context in which centromeres are naturally demethylated. Patients affected by the disease present a myriad of symptoms and have a poor life expectancy, mainly attributed to recurrent infections. However, they can survive up to adulthood, proving that centromeric methylation is important but dispensable for cell survival. Research on the ICF syndrome is limited and has mostly revolved around the study of patient derived cells or the use of cell lines with full knockouts of the ICF proteins. Neither of these cellular models allows to understand the early consequences of the ICF mutations, nor the demethylation kinetics upon the loss of these proteins. There was therefore in the field a need into develop cellular models in which inducible degradation of the ICF proteins could be achieved.

To this aim, I generated an ICF model cell line by CRISPR-Cas9 gene editing on HCT116 cells with the AID2 system (a kind gift of M. Kanemaki) which employs an OsTIR1(F74G) mutant and 5Ph-IAA as a ligand. This system showed no detectable leaky degradation, requires a 670-times lower ligand concentration, and achieves even quicker degradation than the conventional AID (Yesbolatova et al., 2020). I introduced an mNeonGreen fluorescent protein and an AID tag (mNG-AID) to the 3' end of the endogenous HELLS locus (**Figure 36A**). The biallelic targeting, herein referred to as HELLS$^{mNG\text{-}AID}$ for simplicity, was confirmed by PCR (**Figure 36B**) and by sequencing of the junctions of the inserted tags and the endogenous locus (**Figure 36C**). A rapid and total protein degradation upon addition of 5Ph-IAA was observed by immunoblot (**Figure 36D**).

Patients with ICF2, 3 and 4, i.e. with mutations in ZBTB24, CDCA7 and HELLS, respectively, have characteristic methylation profiles, with certain heterochromatic loci being particularly affected (Velasco et al., 2018). I measured by COBRA the methylation status of six genomic regions: centromeric α-satellites, pericentromeric HSAT2, short interspersed Alu repeats, long interspersed nuclear element LINE-1, and the hypermethylated CGI promoters of MAEL (Maelstrom Spermatogenic Transposon Silencer) and TDRD6 (Tudor Domain Containing 6), which are known to be more affected in ICF1 patients (i.e., with mutations in DNMT3B) than ICF2-3-4. Kinetics of demethylation of the first days of HELLS depletion revealed no significant changes in methylation levels at any locus, therefore I extended the measurements to a weekly sample for up to a month. The demethylation kinetics were extremely slow for all the regions probed, with HSAT2 presenting the highest demethylation levels after 28 days, and α-sat only being demethylated by 28% (**Figure 37A**). Since the focus of my work is in understanding the effects of the loss of centromeric methylation, I extended the treatment with 5Ph-IAA to two months in an attempt to achieve higher levels of α-satellite demethylation. I measured by COBRA the methylation of all six loci after depletion of HELLS for two months, and used for comparison HCT116 cells with a double DNMT1/DNMT3B knock-out (DKO) (Rhee et al., 2002) (**Figure 37B,C**)**.** The HELLS depleted cells exhibit a 52.7% decrease in α-sat methylation and a 64.1% decrease in HSAT2 methylation.

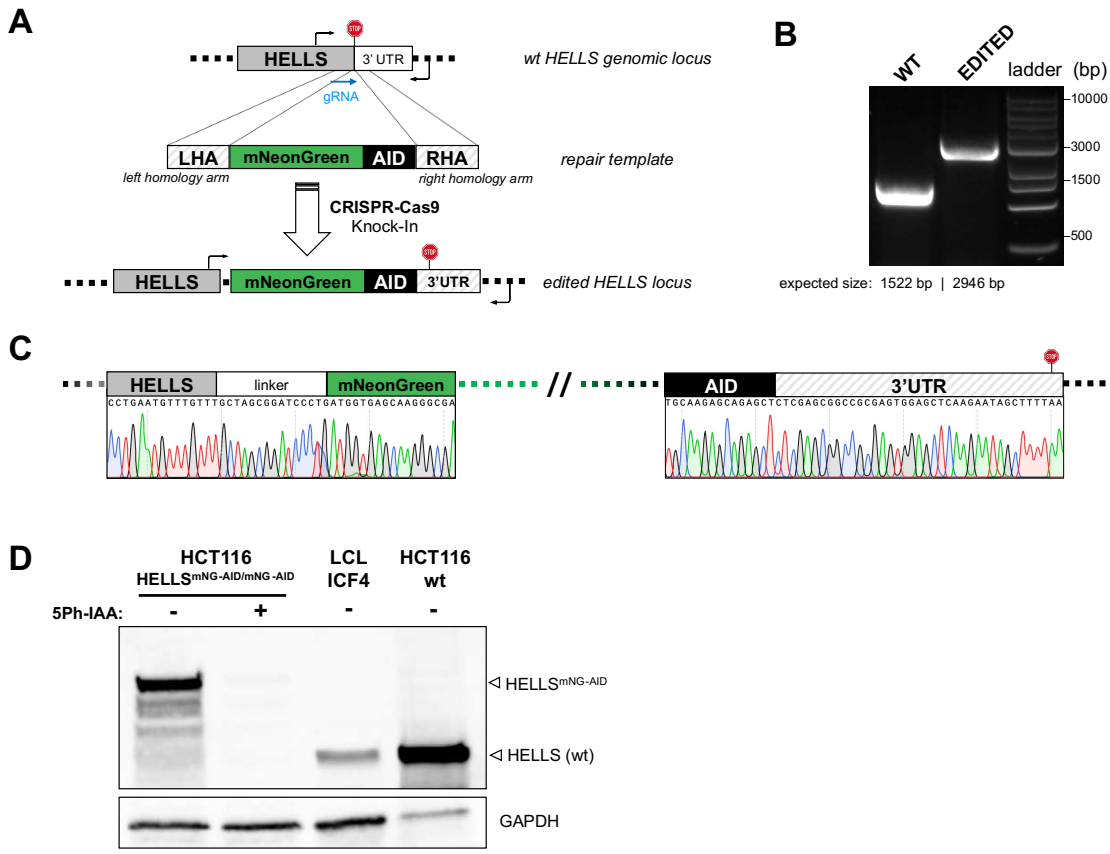

Figure 36. Generation of an ICF4 model cell line by CRISPR-Cas9 genome editing.
(**A**) Schematic of CRISPR-Cas9 mediated gene editing for the biallelic introduction of mNeonGreen (mNG) and AID tags at the C-terminus of HELLS on HCT116 cells. The blue arrow indicates the position of the gRNA used for guiding the Cas9 to the 3' end of the HELLS gene. The black arrows indicate the position of the primers used for PCR genotyping shown in (B) and sequenced in (C) **(B)** PCRs on genomic DNA extracted from wild type (WT) HCT116 cells and from one of edited clones **(C)** The PCR product from (B) was purified by agarose gel extraction and Sanger sequenced. Raw DNA sequencing data of the junctions between the tags and the endogenous sequence is shown, demonstrating that the tag insertion is in frame **(D)** Immunoblot of whole cell lysates from the indicated cells with or without treatment with 1 μM of 5Ph-IAA for 24 hours. The protein extract from a lymphoblastoid cell line (LCL) of an ICF4 patient (lane 4) was a gift of Guillaume Velasco and showcases a reduced protein level resulting from an hypomorphic mutation of HELLS in this patient.

All other loci, apart from LINE-1 which is demethylated by 37.7%, seem unaffected by HELLS depletion. The measure of the exact methylation levels of each CpG by pyruvate sequencing of the bisulfite converted PCR amplicons for α-sat and HSAT2 confirmed that the pericentromeric HSAT2 are demethylated faster and to a larger extent than the α-satellites in this cellular context (**Figure 37D**). On average all HSAT2 CpGs saw their methylation reduced by 79.6% after two months of HELLS depletion. The centromeric α-satellite DNA instead was demethylated by ~60% after two months of HELLS depletion, an equivalent level -measured by COBRA- to the one achieved with the active $TET1_{CD}$ demethylation system at 16 days expression (compare to **Figure 24B,** 61.8% of

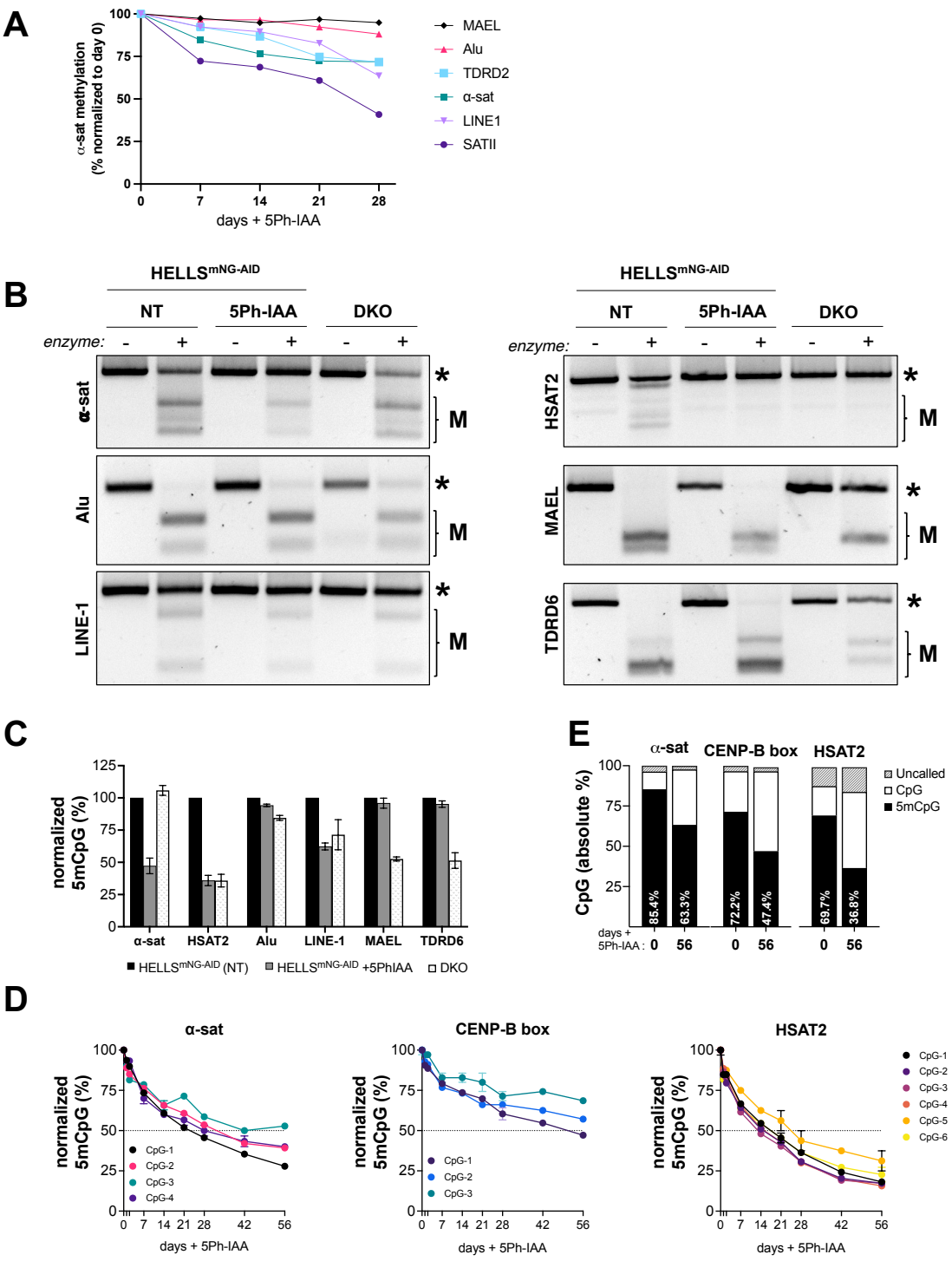

Figure 37. Methylation phenotype of cells with induced HELLS degradation.
(**A**) Quantification of the COBRA kinetics of HCT116 HELLS$^{mNG-AID}$ cells on six genomic loci, normalized to day 0. Average of two experimental replicates (**B**) Representative agarose gels of the COBRA analysis of HELLS$^{mNG-AID}$ cells untreated (NT), after 56 days of treatment with 5Ph-IAA and HCT116 cells with DNMT1/DNMT3B double knock-out (DKO). The absence (–) or presence (+) of the methylation-sensitive enzyme is indicated. The asterisk (*) demarks the position of the unmethylated (undigested) PCR fragment, M indicates the methylated (digested) fragments. (**C**) Quantification of the COBRA analysis shown on (B), normalized to HELLS$^{mNG-AID}$ untreated cells (black bars). Average of two replicates. (**D**) Individual CpG methylation kinetics quantified by pyruvate sequencing (**E**) Mapping-independent motif-based analysis of Nanopore sequences with methylation calls.

reduction). Finally, we sequenced with Nanopore the genomic DNA of untreated HELLS$^{mNG\text{-}AID}$ cells and after two months of HELLS depletion. As for DLD-1, (peri)centromeric reads from HCT116 cannot be mapped precisely to the reference genome, therefore the same motif-based analysis as before was applied to identify the reads containing active centromeric α-satellite sequences, CENP-B boxes, and HSAT2 repeats (**Figure 37E**). This analysis confirmed that the HSAT2 satellite repeats are the most affected by the loss of HELLS, with methylation reduction of 47.2%. The centromeric α-satellite sequences exhibit a 25.9% methylation reduction, and the CENP-B boxes saw their methylation reduced by 34.3%. In this cellular model, it appears that the chromosome 1 amplicons probed by COBRA and pyruvate sequencing are more affected than the average of all chromosomes, giving rise to the numerical discrepancy between both types of quantification.

I assessed the long-term cell viability at increasing times of HELLS degradation by colony formation assays (**Figure 38A**). The HELLS depleted cells remained equally viable as their HELLS proficient counterparts, even at later timepoints, where the centromeres are demethylated to similar levels as the DLD-1 cells expressing the TET1$_{CD}$ constructs are. I also quantified the centromeric abundance of CENP-A, CENP-B and CENP-C though indirect immunofluorescence microscopy (**Figure 38B**). Two months of HELLS degradation led to a significant 19% increase in centromeric CENP-A levels, a minor yet statistically significant 6% increase in CENP-B levels and no significant changes in CENP-C levels. The demethylation of the CENP-B boxes is less pronounced in these cells than on the DNMT1^{AID} or TET1$_{CD}$ cells, but nonetheless their methylation levels are reduced, so an increased CENP-B binding was expected.

Overall, the generation of this HELLS$^{mNG\text{-}AID}$ cell line allowed to determine that the role of HELLS in the maintenance of the centromeric methylation is minor. The prolonged degradation time required to achieve a moderate centromeric demethylation deterred me from further using this model to study of the cellular and molecular effects of the loss of centromeric methylation. The results presented here, however, support the idea that a mild and gradual centromeric demethylation gives cells time to adjust and thrive in this new epigenetic landscape. Further studies should be conducted to dissect in finer detail the precise role of HELLS in the maintenance, and especially in the establishment, of the centromeric methylation.

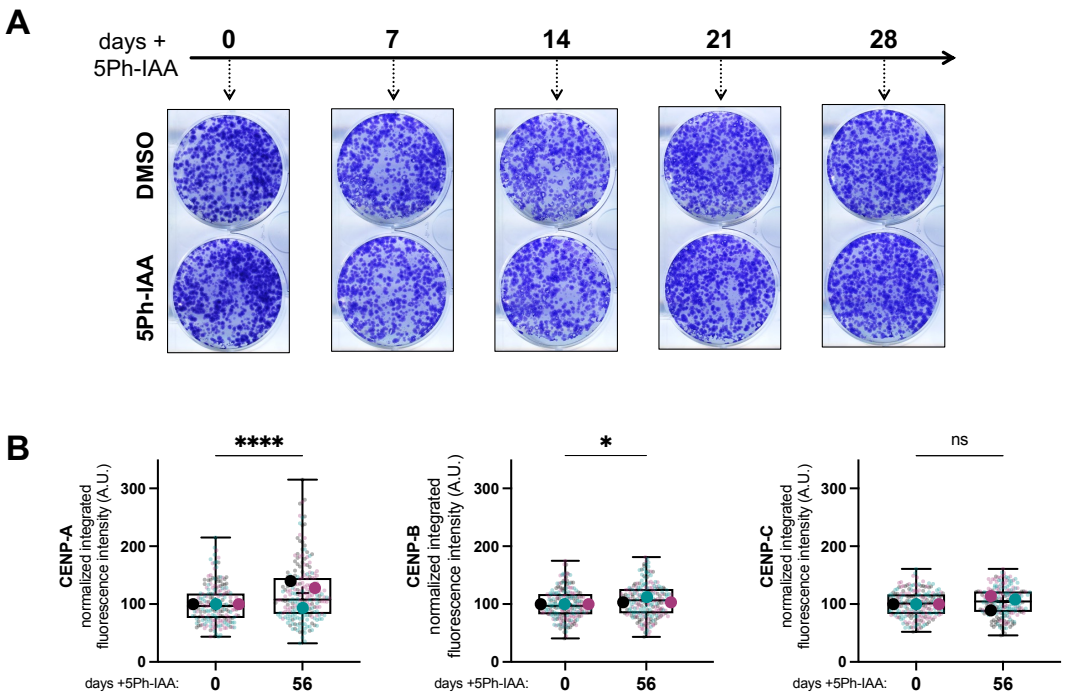

Figure 38. HELLS depletion has no impact on cell viability despite increased centromeric CENP-A levels.

(**A**) Schematic and representative images of colony formation assays stained with crystal violet 10 days after seeding. HELLS$^{mNG\text{-}AID}$ cells were seeded in DMSO (HELLS$^{+/+}$) or in presence of 5Ph-IAA (HELLS$^{-/-}$), without pre-treatment and every week for a month of continued HELLS depletion. (**B**) Quantification of the normalized integrated fluorescence intensity of CENP-A, CENP-B and CENP-C at the centromeres of cells in interphase. Each dot represents the average intensity of all the centromeres of a single cell. The larger dots mark the average of one experimental replicate. N>65 cells and >2507 centromeres per replicate and condition. Statistics defined by Mann-Withney test. ns: non-significant differences.

CONTRIBUTIONS

- **Riccardo Gamba**, PhD., (former post-doctoral researcher at the Fachinetti Lab). R.G. performed the CenRICH experiments, Nanopore sequencing, and bioinformatic analyses of both Nanopore sequences (Figures 19A, 24C and 37E) and Illumina sequences (Figure 35). The Illumina sequencing of the CENP-A CUT&RUN was performed at the NGS core facility of Institut Curie under the supervision of R.G.

- **Leonid Velikovsky,** MSc. (Biochemist, engineer at the Fachinetti Lab). L.V. purified CENP-B and performed the microscale thermophoresis experiment (Figure 19B).

- **Guillaume Velasco**, PhD. The pyruvate sequencings (Figure 24A, B; Figure 37D) were performed in collaboration with G.V. at the Plateforme Epigénomique Fonctionnelle, Epigenetic and Cell Fate Unit, Université Paris-Cité. G.V. also provided me with a protein extract from a lymphoblastoid cell line (LCL) of an ICF4 patient (HELLS mutation), and with an aliquot of anti-HELLS antibody (Figure 36D).

DISCUSSION AND PERSPECTIVES

The present paper [section] ***is devoted to a more detailed discussion of these aspects, the speculative character of which may be justified by the attempt to indicate certain lines of work calculated to test the validity of the conclusions drawn.***
Walter S. Sutton – The Chromosomes in Heredity (Sutton, 1903)

On the regulation of centromeric methylation maintenance

Taking advantage of two DLD-1 cell lines engineered and characterized respectively by two post-doctoral researchers in the Fachinetti lab, I was able to answer to some basic questions about centromeric methylation maintenance. That the DNMT3B^{KO} DLD-1 cells present an almost normal methylation level at centromeres compared to their wild type counterparts (**Figure 19A**) is in accordance with the observation that i) patients with ICF1 syndrome, i.e. that have mutations in DNMT3B, do not present with hypomethylated α-satellites (Jiang et al., 2005; Velasco et al., 2014, 2018), and ii) DNMT3B disruption in HCT116 cells resulted in less than 3% global methylation loss (Rhee et al., 2002). ICF1 patients, however, do exhibit a decreased methylation of the pericentromeric HSAT2, a feature that is not observed in this cell line. A genome-wide compensatory DNMT1 de novo methylation activity was detected in absence of DNMT3B (Scelfo et al., 2023) and both proteins had been already been described to cooperate for the maintenance of DNAme and gene silencing in human cancer cells (Rhee et al., 2002). This could account for the mild effects observed in the DNMT3B^{KO} alone, with DNMT1 possibly filling-in for DNMT3B. Upon DNMT1 degradation, upregulation of DNMT3B was observed, while no changes in the overall DNMT3A levels were detected (Scelfo et al., 2023). Despite this observation, it would be of interest to assess the potential contribution of DNMT3A in the maintenance of the centromeric methylation, particularly in the light of a recent work showing that DNMT1 and DNMT3A have a much stronger impact than DNMT3B in shaping the methylome of HCT116 (Yamaguchi et al., 2023 – manuscript submitted for publication), which, as DLD-1, are colorectal cancer cells.

In **Figure 19A** the methylation levels after 6 days of DNMT1 depletion are presented. I have measured that the DLD-1 DNMT3B^{KO} DNMT1^{AID} cells undergo roughly 4 cell divisions in this time frame (Scelfo et al., 2023). Starting from the 79% of 5mCpGs in α-sat, the passive dilution of the 5mC after four rounds of DNA replication should theoretically yield centromeric methylation levels below 5%, which is more than eight times less than the observed 43.8%. One cell cycle is all that would be needed to go from 79% to 43.8% by passive dilution: it is therefore clear that in absence of both DNMT1 and DNMT3B some other factor, which may or may not be DNMT3A, is actively maintaining the centromeric methylation.

The measurement of the global α-satellite methylation levels by COBRA in the cells with the chimeric CENP-A/ H3 N-tail rescues was first attempt at probing the hypothesis that the lack of ubiquitination on CENP-A N-terminal tail could impair the centromeric methylation maintenance. Many avenues still need to be considered regarding this possibility. First and foremost, the lack of ubiquitination would be an issue only if centromeric methylation maintenance was not completed coupled with DNA replication. This could well be the case, with reported mid- to late-S phase replication timing for human centromeres (Massey and Koren, 2022), but the hypothesis must to be put to the test. If the need for replication-uncoupled methylation was prominent at the centromeric region, we must however also consider that CENP-A nucleosomes are interspersed with H3 nucleosomes, so even in the CENP-A enriched domains, the H3 present should be able to recruit the methylation machinery. Second, the lack of ubiquitination of CENP-A and the hypothetical failure to recruit DNMT1 thereof remain speculative and should be formally addressed. Until proven otherwise, the possibility that CENP-A recruits DNMT1 in a ubiquitination-independent manner exists. Lastly, that I did not detect a global α-satellite methylation increase does not mean that the local methylation level at the CDR was not increased. The detection of methylation by COBRA is not sensitive enough to detect minor, local changes. This level of precise methylation assessment at the centromeres can only be achieved through Nanopore sequencing in a T2T-sequenced cell line for perfect centromeric mapping.

On the regulation of centromeric methylation establishment: a perspective.

Throughout this thesis, my focus has been mainly on understanding the consequences of the centromeric hypomethylation, and tangentially on the regulation of centromeric methylation maintenance in adult somatic cells. I consider that the specific regulation of the centromeric methylation establishment during development should be a topic of interest for future research. Human embryos take ~72 h to reach the blastocyst stage (equivalent to ~64 cells), so the mean cell division time during this period is of only ~14 h (Guo et al., 2014). In this critical stage of development, ensuring a proper chromosome segregation is paramount since mistakes occurring here could be determinant for the survival of the whole organism. During these first 72 hours, a total epigenetic reprogramming takes place, erasing most of the genome-wide DNA methylation. This made me question how the centromere function is safeguarded throughout this period of rapid cell division in a hypomethylated context.

Some interesting insights regarding this arose from the T2T sequencing of the human genome, in particular from the work of Gershman et al., 2022. The CHM13 cell line used for the new genome reference assembly was derived from a complete hydatidiform mole, and is considered trophoblastic i.e., it originated from the outer layer of cells from a pre-implantation blastocyst. As such, it represents the developmental stage where DNA methylation was erased genome-wide for reprogramming and

exhibits a median methylation level of 36.8%. This is roughly half of the overall methylation measured on the HG002 cell line (EBV-immortalized lymphoblastoid cell line [LCL]), which was derived from an adult 45-year-old male and is the first diploid genome to be sequenced T2T (Jarvis et al., 2022) (**Figure 39A**). Despite the general hypomethylation, which also attains the majority of the inactive centromeric HORs, the active HORs specifically present high methylation levels (**Figure 39B**), similar to those of HG002 active HORs. Based on this observation, it would appear that DNA methylation levels are either preserved at the active HORs throughout the genome-wide erasure of methylation, or that the active HORs are amongst the first genomic regions to regain methylation. Both hypothesis are plausible, given that: i) it was seen that early human embryos tend to retain higher residual methylation (Guo et al., 2014) or even gain de novo methylation (P. Zhu et al., 2018) at the evolutionarily younger and more active transposable elements, and ii) the active HOR on each chromosome is the youngest, most rapidly evolving HOR and, for the same reason, is the HOR with less but youngest transposable elements insertions (Altemose et al., 2022a). I believe these fascinating observations should be experimentally addressed in the future.

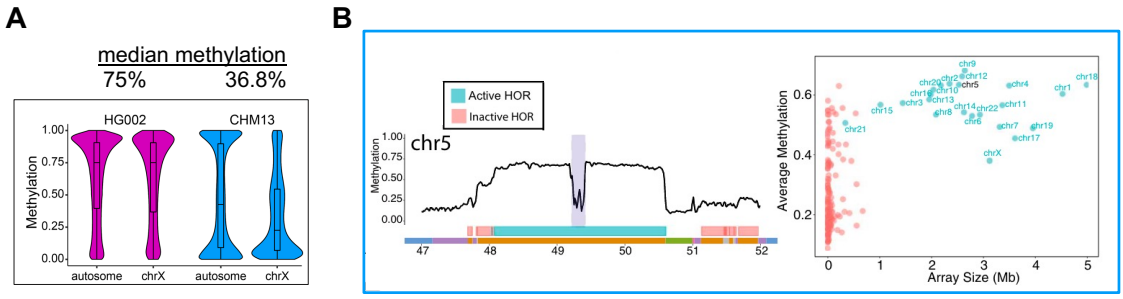

Figure 39. CHM13: a trophoblastic cell line with specifically high active HOR DNA methylation levels. Adapted from Gershman et al., 2022**.** [*Original figure numbers are indicated*] **(A)** [*Figure S10*] Whole genome methylation of CHM13 and HG002 cells. Distribution of methylation from Nanopore sequencing. **(B)** [*Figure 4B]* Left: CHM13 methylation in the centromeric region of chromosome 5. Smoothed methylation frequency is plotted in 10-kb bins. HOR arrays are annotated as blue ("active") and pink ("inactive"), the centromere dip region (CDR) is highlighted. Right: scatter plot of average methylation within each HOR array versus size in mega base-pairs.

On the centromeric transcription

Centromeric transcription has been proposed to play an important role in centromeric establishment and stability. One of most studied roles of DNAme is as a transcriptional repressor; therefore, we hypothesized that upon centromeric hypomethylation, centromeric transcription would increase and that this could possibly explain the cellular phenotype observed. Surprisingly, even though I did detect a minor but statistically significant 1.45-fold increase in centromeric transcription, it is clearly not a consequence of the DNA demethylation since the control dTET1$_{CD}$ cells exhibit a similar fold change (1.35) (**Figure 32**). Technically speaking, the RT-qPCR measurements were performed in five

biological replicates, using two sets of primers targeting different centromeres, and for each, the results were normalized each against three different reference genes. I am confident that the minor transcriptional increase observed is therefore a consequence of some factor shared between the two cell lines, and not an experimental artifact.

One common factor between the two cell lines is the constructs' occupancy of the CENP-B boxes, directed by the CENP-B_{DBD}. A recent publication reported that the overexpression of the CENP-B_{DBD} alone (by doxycycline induction on HeLa Tet-On cells) increased 20-fold the transcription of the centromere of chromosome 1 (Chen et al., 2021). The authors state that the "enhanced centromeric transcription is likely attributed to the formation of more transcription-permissible chromatin, but the exact underlying mechanism still remains unknown". The order of magnitude of this report is incompatible with my observations, but it is a possibility to consider. Another common point between the active and inactive cell lines is the prolonged doxycycline treatment. In the last decade several groups have described that the induction of constructs with doxycycline, even at low concentrations, induce mitochondrial proteotoxic stress that leads to changes in nuclear gene expression, altered mitochondrial dynamics and function, and metabolic changes -in cultured cells as well as in whole organisms- (Moullan et al., 2015; Ahler et al., 2013; De Boeck and Verfaillie, 2021). It is therefore possible that the minor increase in centromeric transcripts detected is a direct consequence of unspecific side-effects of the prolonged doxycycline treatment. Aware of this issue, I titrated the doxycycline as to use the smallest possible concentration to induce the expression of the constructs (100 ng/mL, 10 times less than the usual recommended concentration of 1 μg/ml). To test these two hypotheses and unravel the mystery of the increased transcription, I envision to measure the centromeric transcription levels of the parental cell line used for the Flp-In recombination, devoid of any construct or expressing the CENP-B_{DBD} alone, before and after a prolonged treatment with doxycycline.

On the generation of an ICF4 model cell line: HELLS$^{mNG\text{-}AID/mNG\text{-}AID}$

Given the α-satellite demethylation detected in ICF2, 3 and 4 patients (with mutations in ZBTB24, CDCA7 and HELLS respectively) specifically distinguishes them from ICF1 patients, we anticipated the role of HELLS to be prominent in the maintenance of the centromeric methylation. This prompted me to generate the auxin inducible degron tagged cell line HCT116 HELLS$^{mNG\text{-}AID/mNG\text{-}AID}$ to study the role of the centromeric methylation in this pathophysiological context. Upon HELLS degradation, however, I detected an extremely slow decrease in centromeric methylation and no cell lethality, which indicated that centromeric function was not affected significantly. This deterred me from going further in the characterisation of the cell line and convinced me that future research on the ICF syndrome should rather be focused on the loss of methylation at the pericentromeric repeats,

which are the common ground between all subtypes of the disease and are clearly more affected by the loss of HELLS than their centromeric counterparts. The pericentromeric heterochromatin has a direct influence in the centromeric function, as it ensures the recruitment of cohesin to the (peri)centromeric region. Cohesin plays key roles by establishing a centromere architecture capable of withstanding the mitotic spindle forces and also preventing merotelic spindle attachments (Pidoux and Allshire, 2005; Sacristan et al., 2022).

Previously, loss of function experiments performed in human primary fibroblasts demonstrated that the knock-out of HELLS (and ZBTB24 and CDCA7 also) was not sufficient to alter DNA methylation levels at unique genes, imprinted genes nor at centromeric or pericentromeric DNA repeats (Velasco et al., 2018). This supports the idea that HELLS, ZBTB24 and CDCA7 have likely prominent roles in the establishment of the centromeric DNA methylation profiles in the early steps of development, a hypothesis that should be formally addressed. Only 8 years ago it was demonstrated that HELLS mutations are associated with the ICF syndrome (Thijssen et al., 2015). Since then, efforts have been put in place to elucidate the precise mechanism by which HELLS participates in the maintenance (or establishment) of DNA methylation. The cell line I have generated can be a tool to further investigate these questions with a unique temporal resolution.

On the centromere-targeted demethylation

During my thesis I developed a novel methodology for achieving a centromere specific demethylation through the CENP-B_{DBD}-$TET1_{CD}$ construct. The slow demethylation kinetics (**Figure 23C**) was at first unsettling but can probably be explained by the fact that the TET1 mediated demethylation does not occur in a single enzymatic reaction, but rather after three sequential oxidation steps, a fourth reaction mediated by TDG and is finally resolved by the BER pathway (see Introduction, section III3.2). Nonetheless, out of the three demethylation systems employed and developed in my thesis ($DNMT1^{AID}$; CENP-B_{DBD}-$TET1_{CD}$; $HELLS^{mNG\text{-}AID/mNG\text{-}AID}$) this is still the fastest, the one that allows to achieve the deepest demethylation levels (at the α-satellites in general, and the CENP-B boxes in particular), and the only one that is centromere specific, without altering the pericentromeric HSAT2 methylation. A side-by-side comparison of the methylation levels before and after demethylation for all three systems is presented in **Table 1**.

The basal methylation level of the three repetitive sequences interrogated is, as expected, elevated in all cell lines, but with varying absolute levels for each. This can probably be accounted for by the natural cell-line variability. A point must be made for the particularly low methylation (45.2%) of the CENP-B boxes from the DLD-1 CENP-B_{DBD}-$TET1_{CD}$ cell line. Throughout the prolonged time in culture during the Flp-In recombination process and single cell clone selection, the construct may

have had a leaky expression, which could explain the decreased basal methylation in this particular region. This unfortunate possibility does not invalidate the current results, given that a comparison between untreated and +DOX was always performed, but could explain some of the results, such as the higher starting micronuclei percentage (~5%) compared to the dTET1$_{CD}$ cell line (~2.5%) (**Figure 26F**).

	α -satellite	CENP-B box	HSAT2
DNMT1^{AID}, DNMT3B$^{+/+}$ (NT)	79.0	60.9	79.2
DNMT1^{AID}, DNMT3B^{KO} +IAA 6 days	43.8	22.0	46.4
decrease (%)	**44.6**	**63.9**	**41.1**
CENP-B$_{DBD}$-TET1$_{CD}$ (NT)	73.6	45.2	76.4
CENP-B$_{DBD}$-TET1$_{CD}$ +DOX 16 days	12.1	7.9	77.0
decrease (%)	**83.6**	**82.5**	**-0.8**
HELLS$^{mNG-AID}$ (NT)	85.4	72.2	69.7
HELLS$^{mNG-AID}$ +5Ph-IAA 56 days	63.3	47.4	36.8
decrease (%)	**25.9**	**34.3**	**47.2**

Table 1. Comparison of 5mCpG (%) between three demethylation systems.
Mapping-independent motif-based analysis of Nanopore reads with 5mCpG methylation calling, expressed as percentage of 5mC. Data presented in Figure 19A for the DLD-1 DNMT1^{AID} cell lines, Figure 24C for the DLD-1 CENP-B$_{DBD}$-TET1$_{CD}$ cell line, and Figure 37E for the HCT116 HELLS$^{mNG-AID}$ cell line.

The use of the CENP-B$_{DBD}$ as a targeting domain has the big advantage that all centromeres, except the Y, can be aimed at once. This allowed me to draw general conclusions about the effect of overall centromeric methylation loss. Nevertheless, for that same reason the demethylation is not homogeneous for all chromosomes, and centromeres with higher densities of CENP-B boxes will be more demethylated. As a parallel alternative to these cell lines, I assembled and integrated into DLD-1 through Flp-In recombination the (d)TET1$_{CD}$ fused to dCas9 as targeting moiety. As a future direction, the use of centromere-specific gRNAs recently published (Bosco et al., 2023) could allow to interrogate the effects of the centromeric demethylation at a single chromosome level, and could help unravel a more detailed mechanistic understanding of the consequences of the centromere demethylation on chromosome segregation fidelity. Many questions still remain unanswered at the time of submission of this thesis, the most important one being what is the precise mechanism linking the increase of CENP-A and CENP-B to the increased micronuclei and cell lethality.

The increase of CENP-B levels at the centromere detected by immunofluorescence upon hypomethylation was anticipated based on the previously known higher affinity of CENP-B to unmethylated CENP-B boxes. The remarkable reproduction of the hypomethylated cellular phenotype upon overexpression of CENP-B (**Figure 30**) clearly indicates that centromeric DNA

methylation is an important CENP-B regulatory mechanism. Indeed data from T2T sequencing of HG002 cells (GM24385 (RRID:CVCL_1C78) EBV-immortalized LCL) shows that CENP-B boxes have extremely low methylation levels compared to non-CDR boxes (**Figure 40A**, from Gershman et al., 2022). This observation is complemented by single-molecule chromatin fiber sequencing of CHM13 cells (**Figure 40B**, from Dubocanin et al., 2023) which shows that CENP-B occupies preferentially the unmethylated CENP-B boxes present at the centromere core (defined as the CENP-A enriched region) and does not significantly bind to CENP-B boxes outside the core, even if they are unmethylated. This last observation implies that CENP-B binding to the CENP-B box is dependent on the chromatin context of the box, and that methylation is not the main factor determining the binding. Collectively these observations make me wonder if the increased CENP-B at the centromeres that I detect by immunofluorescence is limited to CENP-B binding to the newly hypomethylated boxes withing the CDR (which should be a minority of all the newly hypomethylated boxes) or if CENP-B is expanding beyond the CDR. Possibly there is an interdependency between the centromeric hypomethylation, putative CDR expansion, putative CENP-A domains expansion and CENP-B increase. The fact that two CDRs have also been detected on the Y chromosome, which lacks CENP-B boxes and CENP-B binding (**Figure 40C**, from Rhie et al., 2022), is a clear indication that the role of DNA methylation at the centromere is not reduced to the regulation of CENP-B binding. In line with these ideas, I hypothesize that on a CENP-B^{KO} background, the centromeric specific hypomethylation would still cause an increase and a putative expansion of CENP-A, and therefore have an impact in centromere function.

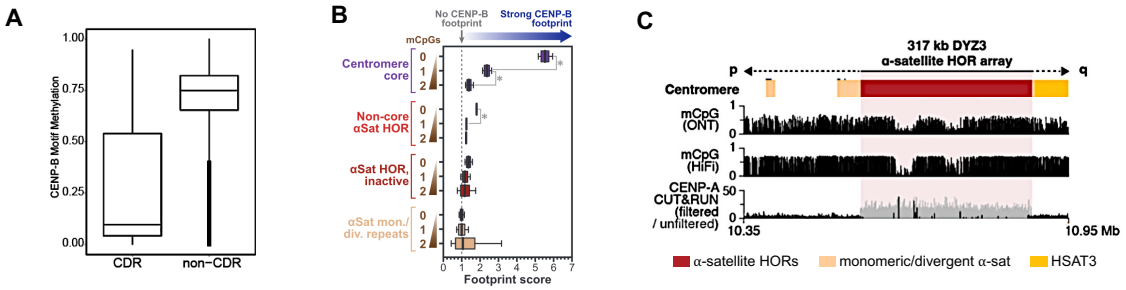

Figure 40. CDR CENP-B boxes are less methylated than non-CDR. CENP-B preferentially binds to the centromere core, regardless of the methylation status of the CENP-B boxes outside the core. [*Original figure numbers are indicated*]
(A) Gershman et al., 2022 *[Figure S28]*. Methylation frequency of the CpG sites within the CENP-B motif inside and outside the CDR ($p < 1e\text{-}15$, Kruskal-Wallis) on HG002 chromosome X **(B)** Adapted from Dubocanin et al. 2023, preprint *[Figure 3D]*. CHM13 single-molecule chromatin fiber sequencing. Box-and-whisker plots of CENP-B footprint scores at CENP-B boxes within various satellite regions as a function of mCpG status at the CENP-B box. Scores higher than 1 quantitatively indicate greater CENP-B occupancy. **(C)** Adapted from Rhie et al., 2022, preprint *[Figure 1B]*. The structure of the HG002 T2T-Y centromere. Histograms show the fraction of methylated CpG sites called by both ONT and HiFi, and CENP-A binding signal from CUT&RUN.

On the CENP-A increase and its putative expansion

CENP-A levels can vary up to 50-fold among healthy tissues and be increased by up to 1000-fold in patient-derived tumours (Jeffery et al., 2021). The ~1.2-fold increase in CENP-A levels I have detected upon hypomethylation does not seem like much by comparison, and the results from the overexpression experiments (**Figure 31**), where a ~34% increase in CENP-A did not cause any cellular phenotype, support this idea.

The key difference between the hypomethylation induced increase and the direct overexpression, to me, resides on the effects of the hypomethylation in the localization of CENP-A. The CENP-A reloading experiments using the combined CENP-A$^{OFF/ON}$ and centromeric hypomethylation systems (**Figure 34**) demonstrate that centromeric methylation is important for the correct loading of CENP-A to the centromeres, possibly by acting as a boundary to signal the correct position for new CENP-A to be deposited. The CENP-A increase can be therefore a consequence of its mis-localization, and not the other way around. In the overexpression experiments what we have is probably more CENP-A nucleosomes deposited in the same active region. In these experiments I did not observe CENP-A deposition at chromosome arms, a feature that is observed when CENP-A was increased by ≥4 fold in HeLa cells (Lacoste et al., 2014; Jeffery et al., 2021).

Further supporting this hypothesis, and likely the most promising result of my thesis, is the CENP-A CUT&RUN (**Figure 35**), which allowed to observe an expansion of CENP-A to a previously unoccupied HOR when the centromeres are hypomethylated. This could have massive implications for the centromeric function, perhaps by weakening the kinetochore assembly. Taking as example the centromere 15 of DLD-1 cells, CENP-A seems to have expanded to a second HOR, making the total CENP-A enriched region at least 50% larger than in untreated cells. However, CENP-A protein levels only increased by 20%, meaning that the density of CENP-A nucleosomes is diminished all along this larger region. It is therefore possible that the local CENP- A concentration in this context is not sufficient to form stable higher-order chromatin structures compatible with kinetochore assembly, much like what it is proposed for CENP-T in a recent work by the Cheeseman group (preprint: Sissoko et al., 2023). All this remains speculative, since this result is reduced to a single DLD-1 centromere that can be mapped with relative accuracy to the T2T-CHM13v1.0 reference genome, and we do not have concrete evidence of a weakened kinetochore. To provide further proof of the spreading of CENP-A we have started two collaborations. First, with the Kops group (Hubrecht Institute, Oncode Institute, Utrecht, The Netherlands), to perform expansion microscopy experiments and assess in detail the 3D structure of CENP-A at the centromeres. They have recently reported that CENP-A is generally observed as a defined bipartite structure in mitotic chromosomes (Sacristan et al., 2022). We wonder if this structure is altered when the centromeres are hypomethylated. Comparing the CENP-A domains number, size, and shape between untreated

DLD-1 Flp-In TET1$_{CD}$ cells and after demethylation might provide insights into changes in CENP-A localization.

A second and crucial collaboration was established with the Altemose group (Department of Genetics, School of Medicine, Stanford University, USA) to perform a CENP-A directed DiMeLo-Seq experiment. This experiment will provide with up to nucleosome resolution simultaneous information about the DNA methylation and CENP-A position at a single molecule level. To have perfect centromeric mappability and circumvent the limitations we have encountered with the DLD-1 cells, I introduced by lentiviral transduction the CENP-B$_{DBD}$-TET1$_{CD}$ demethylation construct into CHM13, the cell line that gave rise to the T2T-CHM13v1.0 reference genome. After antibiotic selection of the positively transduced cells, I obtained a population with more than 80% positive cells for the construct and achieved close to 60% reduction in centromeric methylation after 10 days of induction. The analysis of the DiMeLo-seq is currently ongoing; we hope that these results will provide definitive answers regarding the hypothesized CENP-A expansion.

MATERIALS AND METHODS

Cell Culture and Drugs

Cells were cultured at 37°C in a humidified incubator at 5% CO_2. For all cell lines with doxycycline inducible constructs, tetracycline-free foetal bovine serum (FBS) was used (biowest, S140T), otherwise, regular FBS was used (BioSera, FB-1001). DLD-1 cells and HEK 293-T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) GlutaMAX medium (Gibco, 61965-059) supplemented with 10% FBS. hTERT RPE-1 cells were cultured in DMEM/F12 GlutaMAX medium (Gibco, 31331-028) containing 10% FBS and 0.123% sodium bicarbonate. HCT116 cells were cultured in McCoy's 5A modified medium (Gibco, 26600-080), 10% FBS, and 2 mM L-glutamine (Invitrogen, 25030024). Doxycycline (Sigma, D9891) was dissolved in water and used at 100 ng/mL unless stated otherwise. IAA (Sigma, I5148;) was dissolved in water and used at 500 μM. 5Ph-IAA was a gift of Masato Kanemaki, (currently commercially available), it is dissolved in DMSO and used at 1 μM.

Constructs

[In brackets are indicated the numbers of the constructs in the Fachinetti Laboratory database]

The NLS-CENP-B_{DBD}-NLS-3xFLAG-$TET1_{CD}$-uAID construct [p532] was generated by Gibson assembly of NLS-CENP-B_{DBD}-NLS-3xFLAG, $TET1_{CD}$ and "micro"AID (uAID) fragments in a pcDNA5/FRT backbone. The uAID is a codon-optimized, 46-amino acid segment that corresponds to amino acids 63–108 of *A. thaliana* IAA17 (AtIAA17, NP_171921), and is one amino acid shorter than the tag reported in (Brosh et al., 2016). To generate the inactive version of the catalytic domain of TET1, $dTET1_{CD}$ [p672] two point-mutations, namely C>T H1672Y (in catalytic domain (CD) only: H256Y) and A>C D1674A (in CD only: D258A) were introduced by Gibson assembly following the principle of the QuickChange site-directed mutagenesis using primers fwd: 5'-TGGACTTCTGTGCTCATCCC**T**ACAGGG**C**CATTCACAACATGAATAATGG-3' and rev: 5'-CCATTATTCATGTTGTGAATG**G**CCCTGT**A**GGGATGAGCACAGAAGTCCA-3'. Doxycycline-inducible lentiviral plasmids were generated by sub-cloning the abovementioned constructs, after either removing the uAID tag [p739/p740] or adding a tPT2A and EGFP [p792], into a pInducer-20 backbone using BsmBI and BstXI sites added by PCR amplification of the fragments. For the CRISPR-Cas9 genome editing of HELLS, sgRNAs targeting the 3' end of HELLS were selected using CRISPOR (http://crispor.tefor.net/) (Concordet and Haeussler, 2018). No suitable guide RNAs with high efficiency scores were found for SpCas9 (from *Streptococcus pyogenes*, NGG PAM sites), but two good candidates for SaCas9 (from *Staphylococcus aureus*, NNGRRT PAM sites) were found. The sgRNA 5'- GAAAATTCTGAAGATTCCAGCC -3' was cloned using the BsaI restriction sites into pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA (Addgene, #61591) [p618]. The repair template [p628] was assembled by amplifying the genomic DNA of untransformed human RPE-1 cells with primers fwd: 5'-CTCCTCACACTGCACTTACAA

ACT-3' and rev: 5'-TGGCTTATGGAGTATGTCCCCT-3', yielding a 2.5 kb fragment around the 3' end of the HELLS gene. Nested PCRs were performed to amplify the two homology arms, introducing suitable restriction sites for cloning (in lowercase). Left homology arm (LHA) primers fwd: 5'-cgatggtaccAACGAATCGTGATTGGGCCTC-3' and rev: 5'-gtcagctagcAAACAAACATTC AGG**G**CTGGAATCTTCAG-3'. The reverse primer also introduces a silent mutation (in bold), two nucleotides before the PAM site (underscored) to block SaCas9 from re-cutting once the allele is successfully edited. Right homology arm (RHA) primers fwd: 5'-tcgagcggccgcgAGTGGAGCTCAA GAATAGCT-3' and rev: 5'- gagctcgaattCTACCGACAGCCAAGTCTTTT-3'. The LHA was cloned using KpnI and NheI and the RHA using EcoRI and NotI into a pUC57 vector carrying a mNeonGreen-AID fragment (full length AID, 227 amino acids long *A. thaliana* IAA17) assembled by GENEWIZ.

Cell line generation

For Flp-In recombination, three wells of a 6 well-plate with 60% confluent Flp-In TRex DLD-1 OsTIR1-9xMyc cells (Holland et al., 2012) were co-transfected in a 1:9 ratio with the constructs cloned into pcDNA5/FRT backbone (300 ng) and pOG44 expressing the Flp-recombinase (2.7 μg), 3 μg total DNA per well. The non-liposomal transfection reagent FuGENE®HD (Promega, E2311) was used in a 1:6 μg of DNA : μl of FuGENE ratio, in 150 μl final volume of OptiMEM (Gibco, #31985062). The transfections were carried out in 1.5 mL culture media for 48 hours. Cells were trypsinized and the three 6-wells pooled into a 15 cm ∅ dish for selection and isolation of clones. Correct integration of the construct at the isogenic FRT site gives rise to hygromycin resistance; after selection for 10-14 days with 0.4 mg/mL hygromycin-B (Invitrogen, 10687010), clones were tested for successful integration by immunofluorescence microscopy and immunoblot.

For the generation of 2nd-generation lentiviral particles, HEK 293-T in 10 cm ∅ dishes cells were co-transfected overnight with 6 μg of the constructs in pInducer20 backbone, 4 μg of psPAX2 (gag, pol, rev) and 2 μg pMD2.G (VSV-G env) (12 μg total DNA), using a 1:3 ratio of DNA:FuGeneHD in 5 mL of OptiMEM. After 12 -16h the transfection media was replaced by 6 mL normal growth media. The supernatant containing lentiviral particles was collected after 48 hours, filtered through a Minisart® 0.45 μm polyethersulfone syringe filter (Sartorius, 16537-K). An appropriate volume of lentiviral supernatant was added to the target cells in culture in the presence of 8 μg/mL polybrene (Santa Cruz, sc-134220) for 48 hours. After removal of the lentiviral supernatant the cells were selected with 0.4 mg/mL geneticin (G418 sulphate, Gibco, 10131-035) for up to 14 days. When a fluorescent marker was used, the cells were placed for 2-4 days on antibiotic selection, with frequent media changes to ensure no lentiviral particles remained, and then the construct expression was induced overnight with 100 ng/mL in the presence of 500 μM IAA before sorting the positive cells on a SONY SH800 cell sorter.

For the CRISPR/Cas9 biallelic tagging of HELLS, $1x10^6$ HCT116 AAVS1::CMV-OsTIR1(F74G) cells (Yesbolatova et al., 2020, a kind gift of M. Kanemaki) were electroporated using the Amaxa IIb electroporator (nucleofection solution V, program D-032) with 400 ng of pAAV-CMV-SaCas9;U6-HELLS-C-ter_gRNA-2 [p618] and 4.6 μg of the repair template [p628] linearized by digestion with EcoO109I (molar ratio 1:17 of sgRNA-Cas9 to repair template). 4 days after the transfection, mNeonGreen positive cells (0,07% of the total population) were sorted in bulk (SONY SH800 cell sorter) and expanded for 14 days. Next, single cells were sorted into 96 well plates. Individual clones were screened by PCR using primers fwd: 5'-CTCTCCTCACACTGCACTTACAAA-3' and rev: 5'-CTCTCCTCACACTGCACTTACAAA-3'. The PCR products were Sanger sequenced to cover both insertion sites. The biallelic tagging and protein degradation were confirmed by immunoblot in three clones.

Colony Formation Assays

5000 cells were plated on the first column of 6-well plates and two 1:5 serial dilutions were performed laterally. After 10 days for HCT116, 14 days for DLD-1, colonies were fixed for 30 minutes in methanol, washed with PBS, stained for 15 minutes using a 1% crystal violet in 20% ethanol stain, thoroughly washed with PBS and finally with ddH2O. The plates were air dried, scanned, and the images analysed with Fiji (Schindelin et al., 2012). A suitable threshold was determined manually, and the images were converted into a mask. Equivalent circles were drawn in control and treated wells, excluding the edges where cells tend to accumulate. The raw integrated density of pixels on each circle was measured on the masks as a readout of the area covered by the colonies. The value of the treated well was divided by the value of its corresponding control well to determine the fraction of colonies formed upon treatment.

Indirect immunofluorescence

For immunofluorescence of cells in interphase, cells were grown to 60-80% confluence in 12 mm ∅ 1.5H glass coverslips (Marienfeld Superior™, 0117520). For immunofluorescence of mitotic cells, cells were grown in 18 mm ∅ 1.3-1.6 glass coverslips (Marienfeld, 0111580) to ~60% confluence, blocked in mitosis for 3 hours with 0.1 μg/mL colcemid (Roche, 10 295 892 001) and subjected to a 7-minute-long hypotonic shock in a 60% media, 40% water solution. The coverslips, almost completely dry, were then centrifuged for 3 minutes at 320 rcf to spread the chromosomes.

Cells were fixed with 4% formaldehyde, 0.5% Triton X-100 in PBS for 10 minutes (fixation with extraction), then washed three times with 0,1% Triton X-100 in PBS and blocked for 30 minutes at room temperature or overnight at 4ºC in blocking buffer (0,2 M glycine, 2,5% FBS and 0,1% Triton X-100). Primary antibody incubation was performed for 2 hours at room temperature in 2% BSA, 0,1% Triton X-100 in PBS, with: mouse-anti-CENP-A (1:500; clone 3-19, ENZO, ADI-KAM-

CC006-E), rabbit anti-CENP-B (1 μg/mL polyclonal, Abcam, ab25734 or 0.539 μg/mL, clone EPR24047-64, Abcam, ab259855), guinea pig anti-CENP-C (1:1000, MBL, PD030), rabbit anti-GFP (1 μg/mL, ChromoTek, PABG1) or mouse anti-FLAG (4 μg/mL, Sigma, F3165). Cells were washed three times for three minutes with 0,1% Triton X-100 in PBS before incubation with species specific secondary antibodies for 1 hour at room temperature in 2% BSA, 0,1% Triton X-100 in PBS. Alexa Fluor® 488, Cy3 and Alexa Fluor® 647 conjugated secondary antibodies were used at 1:500, all from Jackson ImmunoResearch Laboratories. Cells were washed three times with 0,1% Triton X-100 in PBS before DAPI staining (1 μg/mL in PBS) for 10 minutes and subsequent mounting with ProLong Gold Antifade Mountant (Invitrogen # P36934).

Fluorescent In Situ Hybridization (FISH)

Standalone protocol or following an immunofluorescence. Cells grown on coverslips were fixed for 15 min at room temperature in Carnoy's fixative (methanol/acetic acid 3:1), rinsed with 80% ethanol, and air-dried for 5 min. An mixture of biotinylated RNA probes (α-sat probes) was designed to match the consensus sequence of 21 Higher Order Repeats that exhibit CENP-A binding properties and that are distributed across all human centromeres (sequences reported in Giunta et al., 2021). A 1 μM equimolar mixture of the α-sat probes, diluted in 2x saline sodium citrate buffer (SSC) + 10% dextran + 1% Tween-20 + 50% formamide, was applied and the coverslip was inverted onto a slide and sealed with rubber cement. The slides were denatured at 75ºC for 2 min and hybridization was carried at 37ºC overnight in a ThermoBrite system (Abbott). The following morning the coverslips were unmounted, washed for 2 min with 0.4x SSC at 72ºC then for 30 s with 4x SSC + 0.1% Tween 20 at room temperature and rinsed with PBS. 2 μg/mL of Alexa Fluor™ 647 conjugated streptavidin (Invitrogen, S21374) diluted in BlockAid™ Blocking solution (Invitrogen, B10710) was added for 1h at room temperature. The excess streptavidin was washed-out two times with with 0,1% Triton X-100 in PBS and once with PBS before counterstaining with DAPI (1 μg/mL in PBS) for 10 minutes. The coverslips were mounted on slides with ProLong Gold Antifade Mountant (Invitrogen, P36934).

Image acquisition and quantification:

Fixed imaging was performed on a DeltaVision Core system (Applied Precision) consisting of an Olympus IX71 inverted microscope equipped with a CoolSNAP$_{HQ^2}$ camera (Phorometrics) and a 250-W xenon light source. Images ≤4 μm thick were captured in 0.2 μm z-sections at room temperature using a 100x Olympus UPlanSApo oil-immersion objective (numerical aperture 1.4), and then deconvolved and projected (3D maximum intensity) using DeltaVision's Softworx software. Fiji was used to quantify the fluorescence intensity in the deconvolved three-dimensional images. For the quantification of fluorescence intensity at the centromeres of cells in interphase, an automatic analysis was performed by detecting the maximum intensity of CENP-C, drawing 0.3 μm radius circles (ROIs)

and measuring the fluorescence of each channel on each ROI (macro available upon request). For the quantification of fluorescence intensity in metaphase cells, circles of 0.84 μm radius were automatically drawn and manually adjusted to encompass the two centromeres of each chromosome. The fluorescence intensity on each ROI was measured for each channel.

Immunoblot

Whole cell lysates were prepared by resuspending cell pellets in 100 mM Tris-HCl pH 7.6, 4% SDS, 20% glycerol buffer and by sonication. The lysates were BCA-quantified and denatured at 95°C in 1X Laemmli sample buffer for 5 minutes. Proteins were separated by SDS-PAGE using 7.5% TGX gels (BioRad) and transferred onto nitrocellulose membranes (0.45 um) using the Trans-Blot Turbo transfer system (BioRad). The membrane was blocked with 5% milk in TBS-T (Tris-buffered saline with 0.1% Tween 20) for 2 h at room temperature and incubated overnight at 4ºC with mouse anti-FLAG (0.8 μg/mL, Sigma, F3165), rabbit anti-GAPDH (1:5000, clone 14C10, CellSignaling, 2118), rabbit anti-HELLS (0.75 μg/mL, Proteintech, 11955-1-AP).

Real-Time Cell Analysis (RTCA)

Cell proliferation was measured on xCELLigence® (Agilent) E-Plate VIEW (96 wells) in an xCELLigence® eSight real-time cell analyzer. Cells were passaged and allowed to reach 70-80% confluency before being trypsinized, counted and resuspended in standard growth media to $4x10^4$ cells/mL. Blanks were measured on the E-Plates with 50 μl of standard growth media before adding 50 μl of the cell suspension, equivalent to 2000 cells, in triplicate wells per condition. Measurements were performed every 15 minutes and pictures of each well were taken every hour for a total of 4 days. Cell index was calculated with the RTCA eSight Software (Agilent).

Cell cycle analysis by flow cytometry

About 1 million cells at <80% confluence were pulsed with 30 μM BrdU for 30 minutes or 24 hours. The cells were then collected and fixed in 3 mL of 75% ethanol with a 30-minute incubation on ice or overnight at -20ºC. The fixed cells were pelleted at 1200 rcf for 5 minutes, washed once in PBS and resuspended 1 mL 2 N HCl for a 15-minute denaturation at room temperature. The acid was neutralized with 3 mL 0,1 M sodium tetraborate for 2 minutes at RTº before pelleting the cells again. The cells were washed three times with 1 mL PBS-T 0.1%, 1% BSA and then resuspended in 100 μL of mouse anti-BrdU antibody (BD, 555627) 2.5 ng/uL (1:200 in PBS-T 0.1%, 1% BSA) for a 1-hour incubation at room temperature in a spinning wheel. The cells were pelleted and washed 2 times with 1 mL PBS-T 0.1%, 1% BSA before being resuspended in 100 μL of anti-mouse-FITC antibody (Invitrogen, 62-6511) 1:200 in PBS-T 0.1% + 1% BSA for a 1-hour incubation at room temperature

in the dark. The cells were then washed three times with 1 mL PBS-T 0.1%, 1% BSA and resuspended in 100 µL of PBS with 2 µg/mL of propidium iodide (Invitrogen, P3566) and 150 µg/mL RNAse A (Invitrogen PureLink™, 12091039) for an overnight incubation at 4°C in the dark. BrdU and PI fluorescence was measured on an LSR II flow cytometer (BD) and the data was analyzed using FlowJo software v10 (BD).

Combined Bisulfite Restriction Analysis (COBRA)

Gennomic DNA was extracted using the NucleoSpin Tissu kit (Macherey-Nagel, 740952.50). Between 0.5 a 2 µg of DNA (equalized quantity for parallel comparisons) was bisulfite converted using the EpiTect bisulfite kit (Qiagen, 59104) according to the manufacturer's protocol. The converted DNA was amplified by PCR using locus-specific primers described in Velasco et al., 2018 and listed on **Table 2**. The reactions were performed with 1.5 mM $MgCl_2$, 200 µM dNTPs, 200 µM of each primer and 0.5 U Platinum Taq DNA polymerase (Invitrogen, 15966005). Thermal cycling conditions: initial denaturation for 1 min at 94ºC, then 50 cycles of: 15 s at 94ºC, 30 s at 53ºC and 40 s at 68ºC (no final elongation). The PCR products were digested for 3 hours with 10 U of the appropriate enzyme at its digestion temperature (listed on **Table 2**). An equal amount of PCR product was incubated alongside the digestions without the addition of enzyme as undigested control. The full digestions were loaded in a 3% agarose gel. Images were acquired using a ChemiDoc (BioRad) and band intensities were quantified using Fiji. The absolute methylation percentage is calculated from the ratio of the intensity of the methylated bands divided by the sum of all bands on the same lane (unmethylated + methylated).

target		primer sequence (5'→3')	enzyme	digestion temp.
α-sat	F	GTGGATATTGGGATTTTTTTGAGAATT	HpyCH4IV	37°C
	R	CCACCCAAAAAAATATTCAACTCTATAA		
HSAT2	F	TTATTGAATGGAAATGAAAGGGGTTAT	BstBI	65°C
	R	CCAATAAATTATTCCATTCCATTCCATTAA		
Alu	F	AGTAGTAGATTTTATTTAGGGAGTATTAA	HpyCH4IV	37°C
	R	TTCCATCCCTTTAAAACCTCAATTTTC		
LINE1	F	GGGAAGAGTAAGGGGTTAGGGA	HpyCH4IV	37°C
	R	CCCTCCCCCAACCTTACTAC		
TDRD6	F	AGGAGATGTAGGTTGTGTTTAAAATTT	BstUI	60°C
	R	AATTCAAAAAACCCAATAAAAAAAA		
MAEL	F	AGTTAATTAGAGTATTTGGTATTTG	BstUI	60°C
	R	AAACTTCCTAACCTCAAACAAAACA		

Table 2. Primers and enzymes for COBRA.

Pyruvate sequencing

The pyruvate sequencing on bisulfite converted DNA was performed as described in Velasco et al., 2018. Briefly, an α-satellite amplicon (see below) is obtained by PCR with the forward and biotinylated reverse primers listed on **Table 3** (in blue). The sequencing of the four α-sat CpGs (in red) is performed using an internal sequencing primer (in green). The CENP-B box (highlighted in grey) CpGs and the third adjacent CpG (all in purple) were sequenced starting from the same α-satellite amplicon but using the forward primer of the PCR for sequencing (in blue, underlined).

α-sat amplicon (from GenBank: M38468)

GTGGATATTGGGACTTCTCTGAGAATTTCG$_1$TTGGAAACG$_2$GGATAAACCTCACG$_3$TAACTGAAG
AGGAACATTCTCAGAACTTCTTTGTGATGTTGAACTTCAACTGACAGAGGTGAACCTTCCCTTG
TGAGTTCAGGTTGAAACG$_1$CTCCTTTCG$_2$TGGCATCTGCAAGTGAAGATTTGGAACG$_3$CTATGAG
GCCTACG$_4$GTAGTAAAGGAAACAGCTTCATGTAAAAACTGAACAGAAGCATTCTCAGAAAATAC
TTTGGGATGATTGAGTTCAACTCACAGAGCTGAACATTCCTTTGGGTGA

The HSAT2 amplicon is obtained by PCR with the forward and biotinylated reverse primers listed on **Table 3** (in blue). The sequencing of the six CpGs in red is performed using the internal sequencing primer (in green). There are several other CpGs on the amplicon that were not sequenced (highlighted in light grey). The sequences of all the primers used for PRC amplification (F) and (R), and for sequencing (S) are listed in **Table 3.** Note that the sequences of the primers consider the bisulfite conversion (C>T in forward strand and G>A in reverse strand).

HSAT2 amplicon (from GenBank: X72623)

TCATTGAATGGAAATGAAAGGGGTCATCATCTAATGGAATCGCATGGAATCATCATCAAATGGA
ATCGAATGGAATCATCATCAAATGGCAATCTAATGGAATCATTGAACAGAATTGAATGGAATCG
TCATCGAATGAATTGAATGCAATCATCGAATGGTCTCGAATGGAATCAT**CTTCTAATGGAAAGG**
AATGGAATCATCG$_1$CATAGAATCG$_2$AATGGAATTATCATCG$_3$AATGGAATCG$_4$AATGGTATCAAA
CG$_5$GAAAAAAACG$_6$GAATTATCGAATGGAATCGAAGAGAATCTTCGAACGGACCCGAATGGAATC
AT**CTAATGGAATGGAATGGAATAATCCACTGG**

target		primer sequence (5'→3')
α-sat	F	GTGGATATTGGGATTTTTTTGAGAATT
	R	Biot-CCACCCAAAAAAATATTCAACTCTATAA
	S	TTTTTTTTGTGAGTTTAGGT
CENP-B box	S	GTGGATATTGGGATTTTTTTGAGAATT
HSAT2	F	TTATTGAATGGAAATGAAAGGGGTTAT
	R	Biot-CCAATAAATTATTCCATTCCATTCCATTAA
	S	TTTTAATGGAAAGGAATGG

Table 3. Primers for pyrosequencing.

Enrichment of centromeric DNA: CenRICH

The enrichment of centromeric DNA (CenRICH) was performed as described in (Gamba et al., 2022). Briefly, 300–400 million cells were collected, washed and lysed for 30 minutes at 37ºC in TNES buffer (10 mM Tris-HCl pH 7.4, 1mM EDTA pH8, 100 mM NaCl+1% SDS) supplemented with 100 µg/mL RNaseA (Invitrogen cat #12091021). An overnight, 100 µg/mL proteinase K (Invitrogen, #25530049) treatment at 37ºC was followed by a phenol:chloroform:isoamylalcol (25:24:1) (Sigma Aldrich cat#77617) extraction. The genomic DNA was then precipitated with 0.1 volume of sodium acetate 3M pH 5.2 and one volume of isopropanol, washed with 70% ethanol, and gently resuspended in 1 mL of Tris-HCl 10 mM pH 8.0. 2.5 mg of DNA were added to 20 mL of 1X CutSmart Buffer (NEB cat#B7204S) to be digested overnight at 37 ° with 400 units of each ScrFI, EcoO109I and BstUI (New England Biolabs) [referred to as SEB enzyme mixture], to enrich the pericentromeric HSAT2 repeats alongside the centromeric α-satellite. The digestion products were purified by phenol:chloroform:isoamylalcol (25:24:1) extraction, precipitated with isopropanol and sodium acetate, and resuspended in 4.5 mL of TE 1X before being fractionated on sucrose gradients (20% to 40% w/v). The high molecular weight fractions (F4-6) were collected and concentrated before sequencing.

Nanopore Sequencing

Nanopore long read sequencing was performed after CenRICH. The DNA was treated with Short Read Eliminator kit (cutoff <25 kb, Circulomics cat# SKUSS-100-101-01) to further remove contamination from shorter DNA fragments. Libraries were prepared using the Library Preparation by Sequencing kit (Oxford Nanopore Technology), quantified with Qubit dsDNA HS Assay Kit (Thermo Fisher) and checked by capillary electrophoresis with a TapeStation 4150 system (Agilent). The sequencing was performed on a Spot-ON Flow Cell (R9.4.1) on a MinION Mk1B device.

Mapping-independent methylation analysis of Nanopore Data

The methylation analysis was performed in a mapping independent manner, as to avoid biases due to variations between the centromeric arrays of the target cell lines and the ones assembled in the reference genome. Briefly, the raw nanopore data (representing the patterns of electrical current) were basecalled using the Oxford Nanopore software Guppy version 4.0 with a high accuracy modified bases model (na_r9.4.1_450bps_modbases_5mc_hac.cfg). This model takes into account the variations in electrical current across the nanopore occurring when a methylated vs. unmethylated nucleotide passes through. The resulting data was then converted with the fast5mod utility (see ref: https://github.com/nanoporetech/fast5mod), to produce as output the sequences of all nanopore reads and their per-base likelihood of methylation, expressed as a value ranging from 0 to 255. It is important to point out that this likelihood of methylation is not indicative of the proportion of reads

being methylated at a given genomic site, or of the percentage of CpG sites being methylated in a heterogeneous population of cells. This value is rather an expression of how confidently the basecalling algorithm is able to predict the methylation status of a single C nucleotide of a specific ssDNA molecule (0: most likely unmethylated. 255: most likely methylated).

A bash script* was used to identify the CENP-B box motif (TT**CG**NNNNANN**CG**GG and its reverse complement CC**CG**NNTNNNN**CG**AA) within all the reads and to extract the corresponding methylation likelihood information. Considering only the two CpG sites that are present in the CENP-B box, the following stringent thresholds were applied to assign a methylation status to the Cs: likelihood values <30 were assigned as "unmethylated"; likelihood values > 220 were assigned as "methylated"; other values were defined as "uncalled". Based on these calls, per each CpG of the CENP-box, the proportions of methylated, unmethylated, uncalled were computed. The α-satellite and HSAT2 motifs were chosen as to also contain two CpGs (in bold) separated by as close as possible to 7 bp, as to resemble the CENP-B box motif. Their analysis was carried out exactly as described for the CENP-B box. The chosen motifs were an α-sat 16-mer: **CG**TAGGCNTCAAAG**CG** (and its reverse complement) and a HSAT2 19-mer: TCCATT**CG**ANTSCAWT**CG**A (S: G or C, W: A or T) (and its reverse complement). The alignment of these motifs on the T2T-CHM13v1.0 reference genome (Nurk et al., 2022) revealed that there are about 160.000 copies of the CENP-B box 15-mer. There are about 5.000 copies of the α-sat 16-mer, and except for chromosome 8, there is at least one copy per chromosome. The motif is especially enriched in the centromeres 12, 13 and 15 where it is mostly found at <200 bp from closest CENP-B box (therefore in adjacent α-sat monomers). There are about 120.000 copies of the HSAT2 19-mer, which is especially enriched at centromeres 1, 9, 15 and 16 and usually found >500 kbp from the closest CENP-B box.

CUT&RUN-seq

CUT&RUN was performed according to the procedure previously reported in (Skene and Henikoff, 2017), starting from $3x10^5$ cells and using a rabbit anti-CENP-A antibody (a gift from Aaron Straight, 1:100) and a rabbit IgG isotype control antibody (Invitrogen, 10500C, 1:100) was used for background detection. The general analysis was performed as described in Gamba et al., 2022. Briefly, the Illumina reads were down-sampled to the same total read count. The estimate quantification of α-satellite-derived reads was performed by counting all the reads containing at least two of the previously identified unique alpha 18-mers representative of the alpha satellite DNA variation in the human genome (Miga, 2017b). Illumina reads were mapped using bwa-mem algorithm of the BWA software package on the Telomere-to-Telomere T2T-CHM13v1.0 reference genome (Nurk et al., 2022). Reads mapping on the centromeric region of chromosome 15 (range: 15683284 - 18327600) were counted. Enrichment and CUT&RUN-seq profiles were generated with deeptools 3.1.0 bamCompare with a bin size of 2 Kb.

RT-qPCR

RNA was extracted from ~80% confluent 6 wells (~1 million DLD-1 cells) using RNeasy mini Kit (Qiagen) following the manufacturer's instructions, with addition of 2-mercaptoethanol and with DNase treatment on column. To ensure the complete degradation of centromeric DNA, which would give rise to false positives, 10 µg of extracted RNA (quantified by Nanodrop) were further digested with 10 units of TURBO DNase (1U/µg of DNA) for 2 hours at 37°C and purified with the RNeasy MinElute Cleanup Kit (Qiagen) following the manufacturer's instructions. The RNA integrity (RIN) and concentration was assessed by capillary electrophoresis with a TapeStation 4150 system (Agilent), RIN >7.2 was always achieved under these conditions. Reverse transcription was performed on 1.5 - 2 µg RNA with the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher) following the manufacturer's instructions. Each sample was also incubated without the reverse transcriptase to control for centromeric DNA contamination. The recovered cDNA was quantified by qPCR using a LightCycler 480 (Roche) using the primer pairs listed on **Table 4**. α-satellite primers were mapped on the T2T-CHM13v1.0 reference genome to update their target centromeres. The reference genes PPIA (peptidylprolyl isomerase A), RPLP0 (ribosomal protein lateral stalk subunit P0) and GUSB (glucuronidase beta) were selected because they were reported to be stable in colon adenocarcinoma cell lines (including DLD-1) subjected to serum starvation (Krzystek-Korpacka et al., 2016). I validated all reference genes by performing a standard curve calibration (100 ng to 10 pg of gDNA template in 1:10 dilutions). The efficiencies of all reactions were in the 90-110% range, and I confirmed that the expression of all three genes is stable after 16 days of doxycycline treatment. The fold enrichment was calculated with the ΔΔCt method as enrichment of the target sequence over the reference.

target		primer sequence (5'→3')	reference
Cen 1, 5, 19	F	TCATTCCCACAAACTGCGTTG	(Hoffmann et al., 2016)
	R	TCCAACGAAGGCCACAAGA	
Cen 1, 3, 10	F	CTAGACAGAAGAATTCTCAG	(Nechemia-Arbely et al., 2019)
	R	CTGAAATCTCCACTTGC	
PPIA	F	GGCAAATGCTGGACCCAACACA	(Krzystek-Korpacka et al., 2016)
	R	TGCTGGTCTTGCCATTCCTGGA	
RPLP0	F	TGGTCATCCAGCAGGTGTTCGA	
	R	ACAGACACTGGCAACATTGCGG	
GUSB	F	CTGTACACGACACCCACCAC	
	R	ATTCGCCACGACTTTGTT	

Table 4. Primers for RT-qPCR

CENP-A reloading

The CENP-A reloading experiments were performed as generally described in (Hoffmann et al., 2020). Briefly, the expression of the $TET1_{CD}$ constructs was induced for 4 days with 100 ng/mL doxycycline to demethylate the centromeres. Cells were seeded in poly-L-Lysine (Sigma, P8920) coated coverslips. The doxycycline was washed-out using first PBS, then two times with culture media. CENP-A was degraded by addition of 500 μM IAA to fresh media for 8 hours. IAA was washed-out with 2 volumes of PBS followed by a ≥ 15 minutes incubation at 37ºC with 1 volume of culture media to allow excess compound to diffuse from the cells. This process was repeated a second time. A fifth PBS wash was followed by the addition of fresh culture media. The following morning the media was again changed. Cells were fixed after 48h and immunofluorescence was carried as described above.

Microscale thermophoresis (MST)

MST measurements of were performed on a Monolith NT.115 (Nanotemper Technologies) as described in (Chardon et al., 2022) but using a 5'-Alexa Fluor™ 546 labelled 57 bp DNA carrying either a fully methylated or a fully unmethylated CENP-B box sequence in the central region (IDT). Two-fold dilution series (16 in total) of MBP-CENP-B (produced in house) were performed in the interaction buffer (20 mM Hepes pH 7.5, 150 mM NaCl, 2 mM MgCl2, 1 mM DTT and 0.05% (v/v) Tween 20). The DNA was kept at a constant concentration of 100 nM. The DNA/CENP-B mixtures were loaded into premium capillaries (Nanotemper Technologies). Initial fluorescence at wavelength of 600 nm was measured with excitation at 80% of green LED (550 nm wavelength), at 25ºC. CENP-B concentration-dependent fluorescence quenching was measured. The affinity was quantified by analyzing fluorescence change of the labelled DNA as a function of the concentration of the titrated protein using the MO.AffinityAnalysis v2.1.5 software provided by the manufacturer.

Quantification and statistical analysis

Statistical details, including the value of cells and/or of centromeres measured, can be found in the figure legends. All statistical tests were performed using GraphPad Prism 10.0 for Mac (GraphPad Software, San Diego, California, USA, www.graphpad.com).

ABBREVIATIONS

α-sat	α-satellite DNA (centromeric)
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
5Ph-IAA	5-phenyl-indole-3-acetic acid
ACA	Anti-Centromere-Autoantibody
AID	Auxin Inducible Degron
APC/C	Anaphase Promoting Complex/Cyclosome
BER	Base-excision repair
bp	Base Pair
BSA	Bovine Serum Albumin
CATD	CENP-A Targeting Domain
CCAN	Constitutive Centromere Associated Network
CENP	CENtromere Protein
CENP-A^{EA}	CENP-AEYFP-AID
CGI	CpG dinuleotide island
CREST	Calcinosis, Raynaud's phenomeno, Esophageal dysmotility, Sclerodactyly, Telangiectasia
CRISPR-Cas9	Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR associated protein 9
(cryo-)EM	(cryogenic) electron microscopy
CUT&RUN	Cleavage Under Targets and Release Using Nuclease
CpG	Cytosine-phosphate-Guanine (dinucleotide)
DAPI	4',6-DiAmidino-2-PhenylIndole
DBD	DNA Binding Domain
DLD-1	cell line derived from human colorectal adenocarcinoma
DNA	Desoxyriboucleic Acid
DNAme	DNA methylation
DNMT	DNA methyltransferase
DOX	Doxycycline
ES	Embryonic Stem cells
et al.	and others
EYFP	Enhanced Yellow Fluorescent Protein
ExM	Expansion Microscopy
FBS	Fetal Bovine Serum
FISH	Fluorescence In Situ Hybridization
Flp-In	site-directed recombination at the FRT (Flp recognition target) by the Flp recombinase
GFP	Green Fluorescent Protein
HCT116	cell line derived from human colorectal adenocarcinoma
HJURP	Holliday Junction Recognizing Protein
HOR	Higher order repeat
HSAT2 (3)	Human Satellite 2 (3)
IAA	Indole-3-acetic acid
ICF	Immunodeficiency, Centromeric region instability, Facial anomalies
i.e.	stands for the Latin *id est*, meaning "that is"
IgG	Immunoglobulin G
kbp	kilo-base pairs
LV	lentivirus
KMN	KNL1, Mis12 complex, Ndc80 complex
Mbp	mega base pairs
MNase	Micrococcal Nuclease
mNG	mNeonGreen, fluorescent protein
NT	Non-Treated
Os-TIR1	*Oryza sativa*-Transport Inhibitor Response 1
PBS	Phosphate Buffered Saline
qPCR	quantitative Polymerase Chain Reaction
RNA	RiboNucleic Acid
RNAPII	RNA Polymerase II
RPE-1	Retinal Pigment Epithelia-1, cell line

RT-qPCR	RetroTranscription and quantitative Polymerase Chain Reaction
SAC	Spindle Assembly Checkpoint
SAM	S-adenosyl-L-Methionine
SD	Standard Deviation
SDS-(PAGE)	Sodium Dodecyl Sulfate- (Polyacrylamide Gel Electrophoresis)
SEM	Standard Error of the Mean
TALE	Transcription activator-like effector
TDG	thymine DNA-glycosylase
TET	ten-eleven translocation (family of enzymes)
WO	Wash-Out

BIBLIOGRAPHY

Aapola, U., Shibuya, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S.E., Shimizu, N., Kudoh, J., Peterson, P., 2000. Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5-Methyltransferase 3 Gene Family. Genomics 65, 293–298. https://doi.org/10.1006/geno.2000.6168

Ahler, E., Sullivan, W.J., Cass, A., Braas, D., York, A.G., Bensinger, S.J., Graeber, T.G., Christofk, H.R., 2013. Doxycycline Alters Metabolism and Proliferation of Human Cell Lines. PLoS ONE 8, e64561. https://doi.org/10.1371/journal.pone.0064561

Ahmed, S.A.H., Ansari, S.A., Mensah-Brown, E.P.K., Emerald, B.S., 2020. The role of DNA methylation in the pathogenesis of type 2 diabetes mellitus. Clin. Epigenetics 12, 104. https://doi.org/10.1186/s13148-020-00896-4

Allshire, R.C., Madhani, H.D., 2018. Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. 19, 229–244. https://doi.org/10.1038/nrm.2017.119

Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., Sauria, M.E.G., Borchers, M., Gershman, A., Mikheenko, A., Shepelev, V.A., Dvorkina, T., Kunyavskaya, O., Vollger, M.R., Rhie, A., McCartney, A.M., Asri, M., Lorig-Roach, R., Shafin, K., Lucas, J.K., Aganezov, S., Olson, D., de Lima, L.G., Potapova, T., Hartley, G.A., Haukness, M., Kerpedjiev, P., Gusev, F., Tigyi, K., Brooks, S., Young, A., Nurk, S., Koren, S., Salama, S.R., Paten, B., Rogaev, E.I., Streets, A., Karpen, G.H., Dernburg, A.F., Sullivan, B.A., Straight, A.F., Wheeler, T.J., Gerton, J.L., Eichler, E.E., Phillippy, A.M., Timp, W., Dennis, M.Y., O'Neill, R.J., Zook, J.M., Schatz, M.C., Pevzner, P.A., Diekhans, M., Langley, C.H., Alexandrov, I.A., Miga, K.H., 2022a. Complete genomic and epigenetic maps of human centromeres. Science 376, eabl4178. https://doi.org/10.1126/science.abl4178

Altemose, N., Maslan, A., Smith, O.K., Sundararajan, K., Brown, R.R., Mishra, R., Detweiler, A.M., Neff, N., Miga, K.H., Straight, A.F., Streets, A., 2022b. DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. Nat. Methods 19, 711–723. https://doi.org/10.1038/s41592-022-01475-6

Altmann, R., 1889. Über Nucleinsäuren. Arch F Anat Physiol Physiol. Abt.

Amano, M., Suzuki, A., Hori, T., Backer, C., Okawa, K., Cheeseman, I.M., Fukagawa, T., 2009. The CENP-S complex is essential for the stable assembly of outer kinetochore structure. J. Cell Biol. 186, 173–182. https://doi.org/10.1083/jcb.200903100

Amor, D.J., Choo, K.H.A., 2002. Neocentromeres: Role in Human Disease, Evolution, and Centromere Study. Am. J. Hum. Genet. 71, 695–714. https://doi.org/10.1086/342730

Andronov, L., Ouararhni, K., Stoll, I., Klaholz, B.P., Hamiche, A., 2019. CENP-A nucleosome clusters form rosette-like structures around HJURP during G1. Nat. Commun. 10, 4436. https://doi.org/10.1038/s41467-019-12383-3

Annunziato, A., 2008. DNA Packaging: Nucleosomes and Chromatin. Nature Education 1(1):26 [WWW Document]. URL http://www.nature.com/scitable/topicpage/dna-packaging-nucleosomes-and-chromatin-310

Ansari, I., Chaturvedi, A., Chitkara, D., Singh, S., 2022. CRISPR/Cas mediated epigenome editing for cancer therapy. Semin. Cancer Biol. 83, 570–583. https://doi.org/10.1016/j.semcancer.2020.12.018

Arber, W., 1965. Host-Controlled Modification of Bacteriophage. Annu. Rev. Microbiol. 19, 365–378. https://doi.org/10.1146/annurev.mi.19.100165.002053

Avery, O., 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. J. Exp. Med. 79, 137–158. https://doi.org/10.1084/jem.79.2.137

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., Balasubramanian, S., 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat. Chem. 6, 1049–1055. https://doi.org/10.1038/nchem.2064

Bailey, A.O., Panchenko, T., Sathyan, K.M., Petkowski, J.J., Pai, P.-J., Bai, D.L., Russell, D.H., Macara, I.G., Shabanowitz, J., Hunt, D.F., Black, B.E., Foltz, D.R., 2013. Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. Proc. Natl. Acad. Sci. 110, 11827–11832. https://doi.org/10.1073/pnas.1300325110

Balhorn, R., 2007. The protamine family of sperm nuclear proteins. Genome Biol. 8, 227. https://doi.org/10.1186/gb-2007-8-9-227

Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E.K., Rivas, M.A., Pirinen, M., Gutierrez-Arcelus, M., Smith, K.S., Kukurba, K.R., Zhang, R., Eng, C., Torgerson, D.G., Urbanek, C., the GTEx Consortium, Li, J.B., Rodriguez-Santana, J.R., Burchard, E.G., Seibold, M.A., MacArthur, D.G., Montgomery, S.B., Zaitlen, N.A., Lappalainen, T., 2015. The landscape of genomic imprinting across diverse adult human tissues. Genome Res. 25, 927–936. https://doi.org/10.1101/gr.192278.115

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., Bourc'his, D., 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912. https://doi.org/10.1126/science.aah5143

Barlow, D.P., Bartolomei, M.S., 2014. Genomic Imprinting in Mammals. Cold Spring Harb. Perspect. Biol. 6, a018382–a018382. https://doi.org/10.1101/cshperspect.a018382

Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., Foltz, D.R., 2011. HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J. Cell Biol. 194, 229–243. https://doi.org/10.1083/jcb.201012017

Barra, V., Fachinetti, D., 2018. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat. Commun. 9, 4340. https://doi.org/10.1038/s41467-018-06545-y

Barra, V., Logsdon, G.A., Scelfo, A., Hoffmann, S., Hervé, S., Aslanian, A., Nechemia-Arbely, Y., Cleveland, D.W., Black, B.E., Fachinetti, D., 2019. Phosphorylation of CENP-A on serine 7 does not control centromere function. Nat. Commun. 10, 1–10. https://doi.org/10.1038/s41467-018-08073-1

Barres, R., Zierath, J.R., 2011. DNA methylation in metabolic disorders. Am. J. Clin. Nutr. 93, 897S-900S. https://doi.org/10.3945/ajcn.110.001933

Bateson, W., 1902. Mendel's principles of heredity: a defence. Cambridge: At the University Press.

Baumann, V., Wiesbeck, M., Breunig, C.T., Braun, J.M., Köferle, A., Ninkovic, J., Götz, M., Stricker, S.H., 2019. Targeted removal of epigenetic barriers during transcriptional reprogramming. Nat. Commun. 10, 2119. https://doi.org/10.1038/s41467-019-10146-8

Berger, F., 2019. Emil Heitz, a true epigenetics pioneer. Nat. Rev. Mol. Cell Biol. 20, 572–572. https://doi.org/10.1038/s41580-019-0161-z

Bergmann, J.H., Jakubsche, J.N., Martins, N.M., Kagansky, A., Nakano, M., Kimura, H., Kelly, D.A., Turner, B.M., Masumoto, H., Larionov, V., Earnshaw, W.C., 2012. Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. J. Cell Sci. 125, 411–421. https://doi.org/10.1242/jcs.090639

Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., Earnshaw, W.C., 2011. Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore: H3K4me2 and kinetochore maintenance. EMBO J. 30, 328–340. https://doi.org/10.1038/emboj.2010.329

Bernstein, D.L., Le Lay, J.E., Ruano, E.G., Kaestner, K.H., 2015. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J. Clin. Invest. 125, 1998–2006. https://doi.org/10.1172/JCI77321

Bestor, T., Laudano, A., Mattaliano, R., Ingram, V., 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. J. Mol. Biol. 203, 971–983. https://doi.org/10.1016/0022-2836(88)90122-2

Bird, A.P., 1986. CpG-rich islands and the function of DNA methylation. Nature 321, 209–213. https://doi.org/10.1038/321209a0

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8, 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Bird, A.P., 1978. Use of restriction enzymes to study eukaryotic DNA methylation. J. Mol. Biol. 118, 49–60. https://doi.org/10.1016/0022-2836(78)90243-7

Bird, A.P., Taggart, M., Frommer, M., Miller, O.J., Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40, 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Black, B.E., Cleveland, D.W., 2011. Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. Cell 144, 471–479. https://doi.org/10.1016/j.cell.2011.02.002

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., Cleveland, D.W., 2004. Structural determinants for generating centromeric chromatin. Nature 430, 578–582. https://doi.org/10.1038/nature02766

Bloom, K.S., Carbon, J., 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell 29, 305–317. https://doi.org/10.1016/0092-8674(82)90147-7

Blower, M.D., 2016. Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep. 15, 1624–1633. https://doi.org/10.1016/j.celrep.2016.04.054

Blower, M.D., Karpen, G.H., 2001. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat. Cell Biol. 3, 730–739. https://doi.org/10.1038/35087045

Blower, M.D., Sullivan, B.A., Karpen, G.H., 2002. Conserved Organization of Centromeric Chromatin in Flies and Humans. Dev. Cell 2, 319–330.

Bobkov, G.O.M., Gilbert, N., Heun, P., 2018. Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J. Cell Biol. 217, 1957–1972. https://doi.org/10.1083/jcb.201611087

Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T., Cleveland, D.W., Black, B.E., Shah, J.V., Jansen, L.E., 2014. The quantitative architecture of centromeric chromatin. eLife 3, e02137. https://doi.org/10.7554/eLife.02137

Bosco, N., Goldberg, A., Zhao, X., Mays, J.C., Cheng, P., Johnson, A.F., Bianchi, J.J., Toscani, C., Di Tommaso, E., Katsnelson, L., Annuar, D., Mei, S., Faitelson, R.E., Pesselev, I.Y., Mohamed, K.S., Mermerian, A., Camacho-Hernandez, E.M., Gionco, C.A., Manikas, J., Tseng, Y.-S., Sun, Z., Fani, S., Keegan, S., Lippman, S.M., Fenyö, D., Giunta, S., Santaguida, S., Davoli, T., 2023. KaryoCreate: A CRISPR-based technology to study chromosome-specific aneuploidy by targeting human centromeres. Cell 186, 1985-2001.e19. https://doi.org/10.1016/j.cell.2023.03.029

Bostick, M., Kim, J.K., Esteve, P.-O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science 317, 1760–1764. https://doi.org/10.1126/science.1147939

Boveri, T., 1902. Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns, Verhandlungen der Physikalisch-Medicinischen Gesellschaft zu Würzburg. Stuber.

Boyer, H.W., 1971. DNA Restriction and Modification Mechanisms in Bacteria. Annu. Rev. Microbiol. 25, 153–176. https://doi.org/10.1146/annurev.mi.25.100171.001101

Bradbury, E.M., 1977. Chapter 10 Histone Nomenclature, in: Methods in Cell Biology. Elsevier, pp. 179–181. https://doi.org/10.1016/S0091-679X(08)60099-0

Britten, R.J., Kohne, D.E., 1968. Repeated Sequences in DNA: Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161, 529–540. https://doi.org/10.1126/science.161.3841.529

Brosh, R., Hrynyk, I., Shen, J., Waghray, A., Zheng, N., Lemischka, I.R., 2016. A dual molecular analogue tuner for dissecting protein function in mammalian cells. Nat. Commun. 7, 11742. https://doi.org/10.1038/ncomms11742

Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., Cheeseman, I.M., 2020. Alpha-satellite RNA transcripts are repressed by centromere–nucleolus associations. eLife 9, e59770. https://doi.org/10.7554/eLife.59770

Carmena, M., Wheelock, M., Funabiki, H., Earnshaw, W.C., 2012. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. Nat. Rev. Mol. Cell Biol. 13, 789–803. https://doi.org/10.1038/nrm3474

Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Obergfell, C., O'Neill, M.J., O'Neill, R.J., 2009. A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma 118, 113–125. https://doi.org/10.1007/s00412-008-0181-5

Carroll, C.W., Milks, K.J., Straight, A.F., 2010. Dual recognition of CENP-A nucleosomes is required for centromere assembly. J. Cell Biol. 189, 1143–1155. https://doi.org/10.1083/jcb.201001013

Carroll, C.W., Silva, M.C.C., Godek, K.M., Jansen, L.E.T., Straight, A.F., 2009. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat. Cell Biol. 11, 896–902. https://doi.org/10.1038/ncb1899

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., Zoghbi, H.Y., 2008. MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. Science 320, 1224–1229. https://doi.org/10.1126/science.1153252

Chan, F.L., Marshall, O.J., Saffery, R., Won Kim, B., Earle, E., Choo, K.H.A., Wong, L.H., 2012. Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc. Natl. Acad. Sci. 109, 1979–1984. https://doi.org/10.1073/pnas.1108705109

Chan, F.L., Wong, L.H., 2012. Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res. 40, 11178–11188. https://doi.org/10.1093/nar/gks921

Chardon, F., Japaridze, A., Witt, H., Velikovsky, L., Chakraborty, C., Wilhelm, T., Dumont, M., Yang, W., Kikuti, C., Gangnard, S., Mace, A.-S., Wuite, G., Dekker, C., Fachinetti, D., 2022. CENP-B-mediated DNA loops regulate activity and stability of human centromeres. Mol. Cell 82, 1751-1767.e8. https://doi.org/10.1016/j.molcel.2022.02.032

Chargaff, E., 1950. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia 6, 201–209. https://doi.org/10.1007/BF02173653

Cheeseman, I.M., Chappie, J.S., Wilson-Kubalek, E.M., Desai, A., 2006. The Conserved KMN Network Constitutes the Core Microtubule-Binding Site of the Kinetochore. Cell 127, 983–997. https://doi.org/10.1016/j.cell.2006.09.039

Chen, C.-C., Bowers, S., Lipinszki, Z., Palladino, J., Trusiak, S., Bettini, E., Rosin, L., Przewloka, M.R., Glover, D.M., O'Neill, R.J., Mellone, B.G., 2015. Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. Dev. Cell 34, 73–84. https://doi.org/10.1016/j.devcel.2015.05.012

Chen, H., Kazemier, H.G., De Groote, M.L., Ruiters, M.H.J., Xu, G.-L., Rots, M.G., 2014. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res. 42, 1563–1574. https://doi.org/10.1093/nar/gkt1019

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., Li, E., 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat. Genet. 39, 391–396. https://doi.org/10.1038/ng1982

Chen, Y., Zhang, Q., Teng, Z., Liu, H., 2021. Centromeric transcription maintains centromeric cohesion in human cells. J. Cell Biol. 220, e202008146. https://doi.org/10.1083/jcb.202008146

Choo, K.H., Vissel, B., Nagy, A., Earle, E., Kalitsis, P., 1991. A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. 19, 1179–1182.

Choy, J.S., Wei, S., Lee, J.Y., Tan, S., Chu, S., Lee, T.-H., 2010. DNA Methylation Increases Nucleosome Compaction and Rigidity. J. Am. Chem. Soc. 132, 1782–1783. https://doi.org/10.1021/ja910264z

Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H.A., Wong, L.H., 2009. LINE Retrotransposon RNA Is an Essential Structural and Functional Epigenetic Component of a Core Neocentromeric Chromatin. PLoS Genet. 5, e1000354. https://doi.org/10.1371/journal.pgen.1000354

Collins, K.A., Castillo, A.R., Tatsutani, S.Y., Biggins, S., 2005. De Novo Kinetochore Assembly Requires the Centromeric Histone H3 Variant. Mol. Biol. Cell 16, 5649–5660.

Concordet, J.-P., Haeussler, M., 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. 46, W242–W245. https://doi.org/10.1093/nar/gky354

Cooke, C.A., Bernat, R.L., Earnshaw, W.C., 1990. CENP-B: a major human centromere protein located beneath the kinetochore. J. Cell Biol. 110, 1475–1488. https://doi.org/10.1083/jcb.110.5.1475

Cooper, D.N., Mort, M., Stenson, P.D., Ball, E.V., Chuzhanova, N.A., 2010. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. Hum. Genomics 4, 406. https://doi.org/10.1186/1479-7364-4-6-406

Cooper, D.N., Youssoufian, H., 1988. The CpG dinucleotide and human genetic disease. Hum. Genet. 78, 151–155. https://doi.org/10.1007/BF00278187

Cooper, Gary M., 2000. The Nuclear Envelope and Traffic between the Nucleus and Cytoplasm, in: The Cell: A Molecular Approach. Sinauer Associates, Sunderland (MA).

Cooper, Geoffrey M., 2000. The Origin and Evolution of Cells, in: The Cell: A Molecular Approach. Sinauer Associates, Sunderland (MA).

Corless, S., Höcker, S., Erhardt, S., 2020. Centromeric RNA and Its Function at and Beyond Centromeric Chromatin. J. Mol. Biol. 432, 4257–4269. https://doi.org/10.1016/j.jmb.2020.03.027

Corneo, G., Ginelli, E., Polli, E., 1971. Renaturation properties and localization in heterochromatin of human satellite DNA's. Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth. 247, 528–534. https://doi.org/10.1016/0005-2787(71)90689-7

Corneo, G., Ginelli, E., Polli, E., 1970. Repeated sequences in human DNA. J. Mol. Biol. 48, 319–327. https://doi.org/10.1016/0022-2836(70)90163-4

Corneo, G., Ginelli, E., Polli, E., 1967. A satellite DNA isolated from human tissues. J. Mol. Biol. 23, 619–622. https://doi.org/10.1016/S0022-2836(67)80130-X

Costa, G., Barra, V., Lentini, L., Cilluffo, D., Di Leonardo, A., 2016. DNA demethylation caused by 5-Aza-2'-deoxycytidine induces mitotic alterations and aneuploidy. Oncotarget 7. https://doi.org/10.18632/oncotarget.6897

Costello, J.F., Frühwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomäki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H.-J., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A., Cavenee, W.K., Plass, C., 2000. Aberrant CpG-island methylation has non-random and tumour-type–specific patterns. Nat. Genet. 24, 132–138. https://doi.org/10.1038/72785

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simón, C., Moore, H., Harness, J.V., Keirstead, H., Sanchez-Mut, J.V., Kaneki, E., Lapunzina, P., Soejima, H., Wake, N., Esteller, M., Ogata, T., Hata, K., Nakabayashi, K., Monk, D., 2014. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. 24, 554–569. https://doi.org/10.1101/gr.164913.113

Crick, F.H.C., 1957. On Protein Synthesis.

Csankovszki, G., Nagy, A., Jaenisch, R., 2001. Synergism of Xist Rna, DNA Methylation, and Histone Hypoacetylation in Maintaining X Chromosome Inactivation. J. Cell Biol. 153, 773–784. https://doi.org/10.1083/jcb.153.4.773

Dahm, R., 2005. Friedrich Miescher and the discovery of DNA. Dev. Biol. 278, 274–288. https://doi.org/10.1016/j.ydbio.2004.11.028

Dambacher, S., Deng, W., Hahn, M., Sadic, D., Fröhlich, J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., Schotta, G., 2012. CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3, 101–110. https://doi.org/10.4161/nucl.18955

Darlington, C.D., 1936. The external mechanics of the chromosomes. Proc. R. Soc. Lond. Ser. B-Biol. Sci. 121, 264–273.

De Boeck, J., Verfaillie, C., 2021. Doxycycline inducible overexpression systems: how to induce your gene of interest without inducing misinterpretations. Mol. Biol. Cell 32, 1517–1522. https://doi.org/10.1091/mbc.E21-04-0177

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022. https://doi.org/10.1101/gad.2037511

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., Gennery, A.R., Gimelli, G., Reisli, I., Schuetz, C., Schulz, A., Smeets, D.F.C.M., Sznajer, Y., Wijmenga, C., van Eggermond, M.C., van Ostaijen-ten Dam, M.M., Lankester, A.C., van Tol, M.J.D., van den Elsen, P.J., Weemaes, C.M., van der Maarel, S.M., 2011. Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2. Am. J. Hum. Genet. 88, 796–804. https://doi.org/10.1016/j.ajhg.2011.04.018

Delaval, K., Feil, R., 2004. Epigenetic regulation of mammalian genomic imprinting. Curr. Opin. Genet. Dev. 14, 188–195. https://doi.org/10.1016/j.gde.2004.01.005

Dodge, J.E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., Li, E., 2005. Inactivation of Dnmt3b in Mouse Embryonic Fibroblasts Results in DNA Hypomethylation, Chromosomal Instability, and Spontaneous Immortalization. J. Biol. Chem. 280, 17986–17991. https://doi.org/10.1074/jbc.M413246200

Doskočil, J., Šorm, F., 1962. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochim. Biophys. Acta 55, 953–959. https://doi.org/10.1016/0006-3002(62)90909-5

Drinnenberg, I.A., Henikoff, S., Malik, H.S., 2016. Evolutionary Turnover of Kinetochore Proteins: A Ship of Theseus? Trends Cell Biol. 26, 498–510. https://doi.org/10.1016/j.tcb.2016.01.005

Du, Q., Smith, G.C., Luu, P.L., Ferguson, J.M., Armstrong, N.J., Caldon, C.E., Campbell, E.M., Nair, S.S., Zotenko, E., Gould, C.M., Buckley, M., Chia, K.-M., Portman, N., Lim, E., Kaczorowski, D., Chan, C.-L., Barton, K., Deveson, I.W., Smith, M.A., Powell, J.E., Skvortsova, K., Stirzaker, C., Achinger-Kawecka, J., Clark, S.J., 2021. DNA methylation is required to maintain both DNA replication timing precision and 3D genome organization integrity. Cell Rep. 36, 109722. https://doi.org/10.1016/j.celrep.2021.109722

Dubocanin, D., Cortes, A.E.S., Hartley, G.A., Ranchalis, J., Agarwal, A., Logsdon, G.A., Munson, K.M., Real, T., Mallory, B.J., Eichler, E.E., O'Neill, R.J., Stergachis, A.B., 2023. Conservation of chromatin organization within human and primate centromeres. bioRxiv. https://doi.org/10.1101/2023.04.20.537689

Duda, Z., Trusiak, S., O'Neill, R., 2017. Centromere Transcription: Means and Motive, in: Black, B.E. (Ed.), Centromeres and Kinetochores, Progress in Molecular and Subcellular Biology. Springer International Publishing, Cham, pp. 257–281. https://doi.org/10.1007/978-3-319-58592-5_11

Dumont, M., Fachinetti, D., 2017. DNA Sequences in Centromere Formation and Function, in: Black, B.E. (Ed.), Centromeres and Kinetochores, Progress in Molecular and Subcellular Biology. Springer International Publishing, Cham, pp. 305–336. https://doi.org/10.1007/978-3-319-58592-5_13

Dumont, M., Gamba, R., Gestraud, P., Klaasen, S., Worrall, J.T., De Vries, S.G., Boudreau, V., Salinas-Luypaert, C., Maddox, P.S., Lens, S.M., Kops, G.J., McClelland, S.E., Miga, K.H., Fachinetti, D., 2020. Human chromosome-specific aneuploidy is influenced by DNA-dependent centromeric features. EMBO J. 39. https://doi.org/10.15252/embj.2019102924

Dunleavy, E.M., Almouzni, G., Karpen, G.H., 2011. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. Nucleus 2, 146–157. https://doi.org/10.4161/nucl.2.2.15211

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., Almouzni-Pettinotti, G., 2009. HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. Cell 137, 485–497. https://doi.org/10.1016/j.cell.2009.02.040

Earnshaw, W.C., 2015. Discovering centromere proteins: from cold white hands to the A, B, C of CENPs. Nat. Rev. Mol. Cell Biol. 16, 443–449. https://doi.org/10.1038/nrm4001

Earnshaw, W.C., Ratrie, H., Stetten, G., 1989. Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma 98, 1–12. https://doi.org/10.1007/BF00293329

Earnshaw, W.C., Rothfield, N., 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma 91, 313–321. https://doi.org/10.1007/BF00328227

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F., Cleveland, D.W., 1987. Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J. Cell Biol. 104, 817–829.

Eden, A., Gaudet, F., Waghmare, A., Jaenisch, R., 2003. Chromosomal Instability and Tumors Promoted by DNA Hypomethylation. Science 300, 455–455. https://doi.org/10.1126/science.1083557

Eggermann, T., Monk, D., De Nanclares, G.P., Kagami, M., Giabicani, E., Riccio, A., Tümer, Z., Kalish, J.M., Tauber, M., Duis, J., Weksberg, R., Maher, E.R., Begemann, M., Elbracht, M., 2023. Imprinting disorders. Nat. Rev. Dis. Primer 9, 33. https://doi.org/10.1038/s41572-023-00443-4

Ehrlich, M., 2009. DNA hypomethylation in cancer cells. Epigenomics 1, 239–259. https://doi.org/10.2217/epi.09.33

Ehrlich, M., 2003. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin. Immunol. 109, 17–28. https://doi.org/10.1016/S1521-6616(03)00201-8

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A., Gehrke, C., 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. 10, 2709–2721. https://doi.org/10.1093/nar/10.8.2709

Ehrlich, M., Jackson, K., Weemaes, C., 2006. Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet J. Rare Dis. 1, 2. https://doi.org/10.1186/1750-1172-1-2

Ehrlich, M., Lacey, M., 2013. DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. Epigenomics 5, 553–568. https://doi.org/10.2217/epi.13.43

Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., Cleveland, D.W., 2013. A two-step mechanism for epigenetic specification of centromere identity and function. Nat. Cell Biol. 15, 1056–1066. https://doi.org/10.1038/ncb2805

Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., Cleveland, D.W., 2015. DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev. Cell 33, 314–327. https://doi.org/10.1016/j.devcel.2015.03.020

Fachinetti, D., Logsdon, G.A., Abdullah, A., Selzer, E.B., Cleveland, D.W., Black, B.E., 2017. CENP-A Modifications on Ser68 and Lys124 Are Dispensable for Establishment, Maintenance, and Long-Term Function of Human Centromeres. Dev. Cell 40, 104–113. https://doi.org/10.1016/j.devcel.2016.12.014

Falk, S.J., Guo, L.Y., Sekulic, N., Smoak, E.M., Mani, T., Logsdon, G.A., Gupta, K., Jansen, L.E.T., Van Duyne, G.D., Vinogradov, S.A., Lampson, M.A., Black, B.E., 2015. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. Science 348, 699–703. https://doi.org/10.1126/science.1259308

Feinberg, A.P., Ohlsson, R., Henikoff, S., 2006. The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. 7, 21–33. https://doi.org/10.1038/nrg1748

Feinberg, A.P., Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301, 89–92. https://doi.org/10.1038/301089a0

Feng, Q., Moran, J.V., Kazazian, H.H., Boeke, J.D., 1996. Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition. Cell 87, 905–916. https://doi.org/10.1016/S0092-8674(00)81997-2

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., Francastel, C., 2009. Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res. 37, 5071–5080. https://doi.org/10.1093/nar/gkp529

Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., Kirsh, O., Amouroux, R., Dohmae, N., Suzuki, T., Filion, G.J., Deng, W., de Dieuleveult, M., Fritsch, L., Kudithipudi, S., Jeltsch, A., Leonhardt, H., Hajkova, P., Marto, J.A., Arita, K., Shinkai, Y., Defossez, P.-A., 2017. Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. Mol. Cell 67, 550-565.e5. https://doi.org/10.1016/j.molcel.2017.07.012

Flagiello, D., Bernardino-Sgherri, J., Dutrillaux, B., 2002. Complex relationships between 5-aza-dC induced DNA demethylation and chromosome compaction at mitosis. Chromosoma 111, 37–44. https://doi.org/10.1007/s00412-001-0180-2

Flemming, W., 1882. Zellsubstanz, Kern und Zelltheilung. F.C.W. Vogel, Leipzig.

Flemming, W., 1879. Ueber das Verhalten des Kerns bei der Zelltheilung, und fiber die Bedeutung mehrkerniger Zellen. Arch. Für Pathol. Anat. Physiol. Für Klin. Med. 1–29.

Folco, H.D., Campbell, C.S., May, K.M., Espinoza, C.A., Oegema, K., Hardwick, K.G., Grewal, S.I.S., Desai, A., 2015. The CENP-A N-Tail Confers Epigenetic Stability to Centromeres via the CENP-T Branch of the CCAN in Fission Yeast. Curr. Biol. 25, 348–356. https://doi.org/10.1016/j.cub.2014.11.060

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., Cleveland, D.W., 2009. Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. Cell 137, 472–484. https://doi.org/10.1016/j.cell.2009.02.039

Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., Cleveland, D.W., 2006. The human CENP-A centromeric nucleosome-associated complex. Nat. Cell Biol. 8, 458–469. https://doi.org/10.1038/ncb1397

Fournier, A., Sasai, N., Nakao, M., Defossez, P.-A., 2012. The role of methyl-binding proteins in chromatin organization and epigenome maintenance. Brief. Funct. Genomics 11, 251–264. https://doi.org/10.1093/bfgp/elr040

Franklin, R.E., Gosling, R.G., 1953. Molecular Configuration in Sodium Thymonucleate. Nature 171, 740–741. https://doi.org/10.1038/171740a0

Frixione, E., Ruiz-Zamarripa, L., 2019. The "scientific catastrophe" in nucleic acids research that boosted molecular biology. J. Biol. Chem. 294, 2249–2255. https://doi.org/10.1074/jbc.CL119.007397

Fujita, R., Otake, K., Arimura, Y., Horikoshi, N., Miya, Y., Shiga, T., Osakabe, A., Tachiwana, H., Ohzeki, J., Larionov, V., Masumoto, H., Kurumizaka, H., 2015. Stable complex formation of CENP-B with the CENP-A nucleosome. Nucleic Acids Res. 43, 4909–4922. https://doi.org/10.1093/nar/gkv405

Fukagawa, T., 1999. CENP-C is necessary but not sufficient to induce formation of a functional centromere. EMBO J. 18, 4196–4209. https://doi.org/10.1093/emboj/18.15.4196

Fukagawa, T., Earnshaw, W.C., 2014. The Centromere: Chromatin Foundation for the Kinetochore Machinery. Dev. Cell 30, 496–508. https://doi.org/10.1016/j.devcel.2014.08.016

Furuyama, S., Biggins, S., 2007. Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc. Natl. Acad. Sci. 104, 14706–14711. https://doi.org/10.1073/pnas.0706985104

Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., Ehrlich, M., 1983. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 11, 6883–6894. https://doi.org/10.1093/nar/11.19.6883

Gamba, R., Fachinetti, D., 2020. From evolution to function: Two sides of the same CENP-B coin? Exp. Cell Res. 390, 111959. https://doi.org/10.1016/j.yexcr.2020.111959

Gamba, R., Mazzucco, G., Wilhelm, T., Velikovsky, L., Salinas-Luypaert, C., Chardon, F., Picotto, J., Bohec, M., Baulande, S., Doksani, Y., Fachinetti, D., 2022. Enrichment of centromeric DNA from human cells. PLOS Genet. 18, e1010306. https://doi.org/10.1371/journal.pgen.1010306

Gao, L., Emperle, M., Guo, Y., Grimm, S.A., Ren, W., Adam, S., Uryu, H., Zhang, Z.-M., Chen, D., Yin, J., Dukatz, M., Anteneh, H., Jurkowska, R.Z., Lu, J., Wang, Y., Bashtrykov, P., Wade, P.A., Wang, G.G., Jeltsch, A., Song, J., 2020. Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms. Nat. Commun. 11. https://doi.org/10.1038/s41467-020-17109-4

García Del Arco, A., Erhardt, S., 2017. Post-translational Modifications of Centromeric Chromatin, in: Black, B.E. (Ed.), Centromeres and Kinetochores, Progress in Molecular and Subcellular Biology. Springer International Publishing, Cham, pp. 213–231. https://doi.org/10.1007/978-3-319-58592-5_9

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., Cheeseman, I.M., 2011. Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. Cell 145, 410–422. https://doi.org/10.1016/j.cell.2011.03.031

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., Jaenisch, R., 2003. Induction of Tumors in Mice by Genomic Hypomethylation. Science 300, 489–492. https://doi.org/10.1126/science.1083558

Gayon, J., 2016. From Mendel to epigenetics: History of genetics. C. R. Biol. 339, 225–230. https://doi.org/10.1016/j.crvi.2016.05.009

Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., Altemose, N., Caldas, G.V., Logsdon, G.A., Rhie, A., Eichler, E.E., Schatz, M.C., O'Neill, R.J., Phillippy, A.M., Miga, K.H., Timp, W., 2022. Epigenetic patterns in a complete human genome. Science 376, eabj5089. https://doi.org/10.1126/science.abj5089

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., Mirny, L.A., Dekker, J., 2018. A pathway for mitotic chromosome formation. Science 359, eaao6135. https://doi.org/10.1126/science.aao6135

Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., Funabiki, H., Fachinetti, D., 2021. CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc. Natl. Acad. Sci. 118, e2015634118. https://doi.org/10.1073/pnas.2015634118

Gjaltema, R.A.F., Rots, M.G., 2020. Advances of epigenetic editing. Curr. Opin. Chem. Biol. 57, 75–81. https://doi.org/10.1016/j.cbpa.2020.04.020

Goelz, S.E., Vogelstein, B., Hamilton, S.R., Feinberg, A.P., 1985. Hypomethylation of DNA from Benign and Malignant Human Colon Neoplasms. Science 228, 187–190. https://doi.org/10.1126/science.2579435

Goldberg, I.G., Sawhney, H., Pluta, A.F., Warburton, P.E., Earnshaw, W.C., 1996. Surprising Deficiency of CENP-B Binding Sites in African Green Monkey a-Satellite DNA: Implications for CENP-B Function at Centromeres. Mol. Cell. Biol. 16, 5156–5168.

Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., Robertson, K.D., 2009. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum. Mol. Genet. 18, 3178–3193. https://doi.org/10.1093/hmg/ddp256

Goshima, G., Kiyomitsu, T., Yoda, K., Yanagida, M., 2003. Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. J. Cell Biol. 160, 25–39. https://doi.org/10.1083/jcb.200210005

Goutte-Gattat, D., Shuaib, M., Ouararhni, K., Gautier, T., Skoufias, D.A., Hamiche, A., Dimitrov, S., 2013. Phosphorylation of the CENP-A amino-terminus in mitotic centromeric chromatin is required for kinetochore function. Proc. Natl. Acad. Sci. 110, 8579–8584. https://doi.org/10.1073/pnas.1302955110

Gowher, H., Liebert, K., Hermann, A., Xu, G., Jeltsch, A., 2005. Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J. Biol. Chem. 280, 13341–13348. https://doi.org/10.1074/jbc.M413412200

Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat. Rev. Mol. Cell Biol. 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6

Greenblatt, M.S., Bennett, W.P., Hollstein, M., Harris, C.C., 1994. Mutations in the p53 Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular Pathogenesis. Cancer Res. 54, 4855–4878.

Grenfell, A.W., Heald, R., Strzelecka, M., 2016. Mitotic noncoding RNA processing promotes kinetochore and spindle assembly in Xenopus. J. Cell Biol. 214, 133–141. https://doi.org/10.1083/jcb.201604029

Gruenbaum, Y., Stein, R., Cedar, H., Razin, A., 1981. Methylation of CpG sequences in eukaryotic DNA. FEBS Lett. 124, 67–71. https://doi.org/10.1016/0014-5793(81)80055-5

Guldner, H.H., Lakomek, H.J., Bautz, F.A., 1984. Human anti-centromere sera recognise a 19-5 kD non-histone chromosomal protein from HeLa cells. Clin. Exp. Immunololgy 58, 13–20.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., Fan, X., Yong, J., Wen, L., Xie, S.X., Tang, F., Qiao, J., 2014. The DNA methylation landscape of human early embryos. Nature 511, 606–610. https://doi.org/10.1038/nature13544

Guse, A., Carroll, C.W., Moree, B., Fuller, C.J., Straight, A.F., 2011. In vitro centromere and kinetochore assembly on defined chromatin templates. Nature 477, 354–358. https://doi.org/10.1038/nature10379

Guttenbach, M., Schmid, M., 1994. Exclusion of Specific Human Chromosomes into Micronuclei by 5-Azacytidine Treatment of Lymphocyte Cultures. Exp. Cell Res. 211, 127–132. https://doi.org/10.1006/excr.1994.1068

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., Surani, M.A., 2013. Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine. Science 339, 448–452. https://doi.org/10.1126/science.1229277

Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A.S., Mattei, A.L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B., Giehr, P., Buschow, R., Galonska, C., Von Meyenn, F., Pappalardi, M.B., McCabe, M.T., Wittler, L., Giesecke-Thiel, C., Mielke, T., Meierhofer, D., Timmermann, B., Müller, F.-J., Walter, J., Meissner, A., 2021. Dnmt1 has de novo activity targeted to transposable elements. Nat. Struct. Mol. Biol. 28, 594–603. https://doi.org/10.1038/s41594-021-00603-8

Hagstrom, K.A., Meyer, B.J., 2003. Condensin and cohesin: more than chromosome compactor and glue. Nat. Rev. Genet. 4, 520–534. https://doi.org/10.1038/nrg1110

Hall, K., Sankaran, N., 2021. DNA translated: Friedrich Miescher's discovery of nuclein in its original context. Br. J. Hist. Sci. 54, 99–107. https://doi.org/10.1017/S000708742000062X

Han, M., Li, Jialun, Cao, Y., Huang, Y., Li, W., Zhu, H., Zhao, Q., Han, J.-D.J., Wu, Q., Li, Jiwen, Feng, J., Wong, J., 2020. A role for LSH in facilitating DNA methylation by DNMT1 through enhancing UHRF1 chromatin association. Nucleic Acids Res. 48, 12116–12134. https://doi.org/10.1093/nar/gkaa1003

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M.R., Gartler, S.M., 1999. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl. Acad. Sci. U. S. A. 96, 14412–14417.

Hara, M., Fukagawa, T., 2017. Critical Foundation of the Kinetochore: The Constitutive Centromere-Associated Network (CCAN), in: Black, B.E. (Ed.), Centromeres and Kinetochores. Springer International Publishing, pp. 29–57. https://doi.org/10.1007/978-3-319-58592-5_2

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., Black, B.E., 2013. The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. 20, 687–695. https://doi.org/10.1038/nsmb.2562

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., Xu, G.-L., 2011. Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science 333, 1303–1307. https://doi.org/10.1126/science.1210944

Heitz, E., 1928. Das heterochromatin der moose. Jahrb. Für Wiss. Bot. 69, 762–818.

Henikoff, J.G., Thakur, J., Kasinathan, S., Henikoff, S., 2015. A unique chromatin complex occupies young α-satellite arrays of human centromeres. Sci. Adv. 1, e1400234. https://doi.org/10.1126/sciadv.1400234

Henikoff, S., Ahmad, K., Malik, H.S., 2001. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science 293, 1098–1102. https://doi.org/10.1126/science.1062939

Herman, J.G., Baylin, S.B., 2003. Gene Silencing in Cancer in Association with Promoter Hypermethylation. N. Engl. J. Med. 349, 2042–2054. https://doi.org/10.1056/NEJMra023075

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., Karpen, G.H., 2006. Mislocalization of the Drosophila Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. Dev. Cell 10, 303–315. https://doi.org/10.1016/j.devcel.2006.01.014

Hewish, D.R., Burgoyne, L.A., 1973. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Commun. 52, 504–510. https://doi.org/10.1016/0006-291X(73)90740-7

Hirai, H., Takemata, N., Tamura, M., Ohta, K., 2022. Facultative heterochromatin formation in rDNA is essential for cell survival during nutritional starvation. Nucleic Acids Res. 50, 3727–3744. https://doi.org/10.1093/nar/gkac175

Hoffmann, S., Dumont, M., Barra, V., Ly, P., Nechemia-Arbely, Y., McMahon, M.A., Hervé, S., Cleveland, D.W., Fachinetti, D., 2016. CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. Cell Rep. 17, 2394–2404. https://doi.org/10.1016/j.celrep.2016.10.084

Hoffmann, S., Izquierdo, H.M., Gamba, R., Chardon, F., Dumont, M., Keizer, V., Hervé, S., McNulty, S.M., Sullivan, B.A., Manel, N., Fachinetti, D., 2020. A genetic memory initiates the epigenetic loop necessary to preserve centromere position. EMBO J. https://doi.org/10.15252/embj.2020105505

Holland, A.J., Cleveland, D.W., 2009. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat. Rev. Mol. Cell Biol. 10, 478–487. https://doi.org/10.1038/nrm2718

Holland, A.J., Fachinetti, D., Han, J.S., Cleveland, D.W., 2012. Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc. Natl. Acad. Sci. 109, E3350–E3357. https://doi.org/10.1073/pnas.1216880109

Holliday, R., Pugh, J.E., 1975. DNA Modification Mechanisms and Gene Activity During Development: Developmental clocks may depend on the enzymic modification of specific bases in repeated DNA sequences. Science 187, 226–232. https://doi.org/10.1126/science.187.4173.226

Hooke, R., 1665. Micrographia, or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon. The British Library - Public Domain, London.

Hoppe-Seyler, F., 1871. Ueber die chemische Zusammensetzung der Eiterzellen. Med.-Chem. Untersuchungen 4, 486–501.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., McEwen, B.F., Shang, W.-H., Suzuki, E., Okawa, K., Cheeseman, I.M., Fukagawa, T., 2008. CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. Cell 135, 1039–1052. https://doi.org/10.1016/j.cell.2008.10.019

Hori, T., Shang, W.-H., Takeuchi, K., Fukagawa, T., 2013. The CCAN recruits CENP-A to the centromere and forms the structural core for kinetochore assembly. J. Cell Biol. 200, 45–60. https://doi.org/10.1083/jcb.201210106

Horvath, S., Raj, K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. 19, 371–384. https://doi.org/10.1038/s41576-018-0004-3

Hotchkiss, R.D., 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J. Biol. Chem. 175, 315–332. https://doi.org/10.1016/S0021-9258(18)57261-6

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., Eden, A., 2008. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene 27, 404–408. https://doi.org/10.1038/sj.onc.1210631

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., Choo, K.H.A., 2000. Early disruption of centromeric chromatin organization in centromere protein A (*Cenpa*) null mice. Proc. Natl. Acad. Sci. 97, 1148–1153. https://doi.org/10.1073/pnas.97.3.1148

Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y., Li, G., Xu, R.-M., 2011. Structure of a CENP-A–histone H4 heterodimer in complex with chaperone HJURP. Genes Dev. 25, 901–906. https://doi.org/10.1101/gad.2045111

Hu, L., Zhao, C., Liu, M., Liu, S., Ye, J., Wang, K., Shi, J., Tian, W., He, X., 2023. CENP-I directly targets centromeric DNA to support CENP-A deposition and centromere maintenance. Proc. Natl. Acad. Sci. 120, e2219170120. https://doi.org/10.1073/pnas.2219170120

Hudson, D.F., Fowler, K.J., Earle, E., Saffery, R., Kalitsis, P., Trowell, H., Hill, J., Wreford, N.G., De Kretser, D.M., Cancilla, M.R., Howman, E., Hii, L., Cutts, S.M., Irvine, D.V., Choo, K.H.A., 1998. Centromere Protein B Null Mice are Mitotically and Meiotically Normal but Have Lower Body and Testis Weights. J. Cell Biol. 141, 309–319. https://doi.org/10.1083/jcb.141.2.309

Hulten, M., 1978. Selective somatic pairing and fragility at 1q12 in a boy with common variable immunodeficiency. 14, 294.

Ideue, T., Cho, Y., Nishimura, K., Tani, T., 2014. Involvement of satellite I noncoding RNA in regulation of chromosome segregation. Genes Cells 19, 528–538. https://doi.org/10.1111/gtc.12149

International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research:, Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., The Sanger Centre:, Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Washington University Genome Sequencing Center, Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., US DOE Joint Genome Institute:, Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Baylor College of Medicine Human Genome Sequencing Center:, Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., RIKEN Genomic Sciences Center:, Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Genoscope and CNRS UMR-8030:, Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Department of Genome Analysis, Institute of Molecular Biotechnology:, Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., GTC Sequencing Center:, Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Beijing Genomics Institute/Human Genome Center:, Yang, H.,

Yu, J., Wang, J., Huang, G., Gu, J., Multimegabase Sequencing Center, The Institute for Systems Biology:, Hood, L., Rowen, L., Madan, A., Qin, S., Stanford Genome Technology Center:, Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., University of Oklahoma's Advanced Center for Genome Technology:, Roe, B.A., Chen, F., Pan, H., Max Planck Institute for Molecular Genetics:, Ramser, J., Lehrach, H., Reinhardt, R., Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center:, McCombie, W.R., De La Bastide, M., Dedhia, N., GBF—German Research Centre for Biotechnology:, Blöcker, H., Hornischer, K., Nordsiek, G., *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):, Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Scientific management: National Human Genome Research Institute, US National Institutes of Health:, Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Stanford Human Genome Center:, Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., University of Washington Genome Center:, Olson, M.V., Kaul, R., Raymond, Christopher, Department of Molecular Biology, Keio University School of Medicine:, Shimizu, N., Kawasaki, K., Minoshima, S., University of Texas Southwestern Medical Center at Dallas:, Evans, G.A., Athanasiou, M., Schultz, R., Office of Science, US Department of Energy:, Patrinos, A., The Wellcome Trust:, Morgan, M.J., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Ishikura, S., Yoshida, K., Hashimoto, S., Nakabayashi, K., Tsunoda, T., Shirasawa, S., 2021. CENP-B promotes the centromeric localization of ZFAT to control transcription of noncoding RNA. J. Biol. Chem. 297, 101213. https://doi.org/10.1016/j.jbc.2021.101213

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., Zhang, Y., 2011. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science 333, 1300–1303. https://doi.org/10.1126/science.1210597

Izuta, H., Ikeno, M., Suzuki, N., Tomonaga, T., Nozaki, N., Obuse, C., Kisu, Y., Goshima, N., Nomura, F., Nomura, N., Yoda, K., 2006. Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. Genes Cells 11, 673–684. https://doi.org/10.1111/j.1365-2443.2006.00969.x

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., Jaenisch, R., 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. 27, 31–39. https://doi.org/10.1038/83730

Jaco, I., Canela, A., Vera, E., Blasco, M.A., 2008. Centromere mitotic recombination in mammalian cells. J. Cell Biol. 181, 885–892. https://doi.org/10.1083/jcb.200803042

Jacobs, A.L., Schär, P., 2012. DNA glycosylases: in DNA repair and beyond. Chromosoma 121, 1–20. https://doi.org/10.1007/s00412-011-0347-4

Jähner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Löhler, J., Simon, I., Jaenisch, R., 1982. De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature 298, 623–628. https://doi.org/10.1038/298623a0

Janiszewski, A., Song, J., Vanheer, L., De Geest, N., Pasque, V., 2018. Dynamics of DNA Methylation Reprogramming Influenced by X Chromosome Dosage in Induced Pluripotent Stem Cells. Epigenetics Insights 11, 251686571880293. https://doi.org/10.1177/2516865718802931

Jansen, L.E.T., Black, B.E., Foltz, D.R., Cleveland, D.W., 2007. Propagation of centromeric chromatin requires exit from mitosis. J. Cell Biol. 176, 795–805. https://doi.org/10.1083/jcb.200701066

Jarvis, E.D., Formenti, G., Rhie, A., Guarracino, A., Yang, C., Wood, J., Tracey, A., Thibaud-Nissen, F., Vollger, M.R., Porubsky, D., Cheng, H., Asri, M., Logsdon, G.A., Carnevali, P., Chaisson, M.J.P., Chin, C.-S., Cody, S., Collins, J., Ebert, P., Escalona, M., Fedrigo, O., Fulton, R.S., Fulton, L.L., Garg, S., Gerton, J.L., Ghurye, J., Granat, A., Green, R.E., Harvey, W., Hasenfeld, P., Hastie, A., Haukness, M., Jaeger, E.B., Jain, M., Kirsche, M., Kolmogorov, M., Korbel, J.O., Koren, S., Korlach, J., Lee, J., Li, D., Lindsay, T., Lucas, J., Luo, F., Marschall, T., Mitchell, M.W., McDaniel, J., Nie, F., Olsen, H.E., Olson, N.D., Pesout, T., Potapova, T., Puiu, D., Regier, A., Ruan, J., Salzberg, S.L., Sanders, A.D., Schatz, M.C., Schmitt, A., Schneider, V.A., Selvaraj, S., Shafin, K., Shumate, A., Stitziel, N.O., Stober, C., Torrance, J., Wagner, J., Wang, J., Wenger, A., Xiao, C., Zimin, A.V., Zhang, G., Wang, T., Li, H., Garrison, E., Haussler, D., Hall, I., Zook, J.M., Eichler, E.E., Phillippy, A.M., Paten, B., Howe, K., Miga, K.H., Human Pangenome Reference Consortium, 2022. Semi-automated assembly of high-quality diploid human reference genomes. Nature 611, 519–531. https://doi.org/10.1038/s41586-022-05325-5

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., Viegas-Pequignot, E., 1993. An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum. Mol. Genet. 2, 731–735. https://doi.org/10.1093/hmg/2.6.731

Jeffery, D., Gatto, A., Podsypanina, K., Renaud-Pageot, C., Ponce Landete, R., Bonneville, L., Dumont, M., Fachinetti, D., Almouzni, G., 2021. CENP-A overexpression promotes distinct fates in human cells, depending on p53 status. Commun. Biol. 4, 417. https://doi.org/10.1038/s42003-021-01941-5

Jenness, C., Giunta, S., Müller, M.M., Kimura, H., Muir, T.W., Funabiki, H., 2018. HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc. Natl. Acad. Sci. 115, E876–E885. https://doi.org/10.1073/pnas.1717509115

Jenuwein, T., Allis, C.D., 2001. Translating the Histone Code. Science 293, 1074–1080. https://doi.org/10.1126/science.1063127

Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hultén, M., Jonveaux, P., Maraschio, P., Mégarbané, A., Moncla, A., Viegas-Péquignot, E., 2005. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Hum. Mutat. 25, 56–63. https://doi.org/10.1002/humu.20113

Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M.R.H., Barton, M.C., Issa, J.-P.J., 2014. TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acids Res. 42, 6956–6971. https://doi.org/10.1093/nar/gku372

Johannsen, W., 1909. Elemente der exakten erblichkeitslehre. Jena. Gustav Fischer.

Johnson, T.B., Coghill, R.D., 1925. Researches on pyrimidines. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. J. Am. Chem. Soc. 47, 2838–2844. https://doi.org/10.1021/ja01688a030

Jones, K.W., 1970. Chromosomal and Nuclear Location of Mouse Satellite DNA in Individual Cells. Nature 225, 912–915. https://doi.org/10.1038/225912a0

Jones, M.J., Goodman, S.J., Kobor, M.S., 2015. DNA methylation and healthy human aging. Aging Cell 14, 924–932. https://doi.org/10.1111/acel.12349

Jones, P.A., Baylin, S.B., 2002. The fundamental role of epigenetic events in cancer. Nat. Rev. Genet. 3, 415–428. https://doi.org/10.1038/nrg816

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85–93. https://doi.org/10.1016/0092-8674(80)90237-8

Kagawa, N., Hori, T., Hoki, Y., Hosoya, O., Tsutsui, K., Saga, Y., Sado, T., Fukagawa, T., 2014. The CENP-O complex requirement varies among different cell types. Chromosome Res. 22, 293–303. https://doi.org/10.1007/s10577-014-9404-1

Kalitsis, P., Fowler, K.J., Earle, E., Hill, J., Choo, K.H.A., 1998. Targeted disruption of mouse centromere protein C gene leads to mitotic disarray and early embryo death. Proc. Natl. Acad. Sci. 95, 1136–1141. https://doi.org/10.1073/pnas.95.3.1136

Kaneko-Ishino, T., Ishino, F., 2019. Evolution of viviparity in mammals: what genomic imprinting tells us about mammalian placental evolution. Reprod. Fertil. Dev. 31, 1219. https://doi.org/10.1071/RD18127

Kapoor, M., Montes De Oca Luna, R., Liu, G., Lozano, G., Cummings, C., Mancini, M., Ouspenski, I., Brinkley, B.R., May, G.S., 1998. The cenpB gene is not essential in mice. Chromosoma 107, 570–576. https://doi.org/10.1007/s004120050343

Karpen, G.H., Allshire, R.C., 1997. The case for epigenetic effects on centromere identity and function. Trends Genet. 13, 489–496. https://doi.org/10.1016/S0168-9525(97)01298-5

Kiaee, F., Zaki-Dizaji, M., Hafezi, N., Almasi-Hashiani, A., Haleh, H., Araz, S., Shirkani, A., Zeineb, Z., Jadidi-Niaragh, F., Aghamahdi, F., Mahdi, G., Yazdani, R., Abolhassani, H., Aghamohammadi, A., Azizi, G., 2021. Clinical, Immunologic and Molecular Spectrum of Patients with Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome: A Systematic Review. Endocr. Metab. Immune Disord. - Drug Targets 21, 664–672. https://dx.doi.org/10.2174/1871530320666200613204426

Kipling, D., Mitchell, A.R., Masumoto, H., Wilson, H.E., Nicol, L., Cooke, H.J., 1995. CENP-B Binds a Novel Centromeric Sequence in the Asian Mouse *Mus caroli*. Mol. Cell. Biol. 15, 4009–4020. https://doi.org/10.1128/MCB.15.8.4009

Kipling, D., Warburton, P.E., 1997. Centromeres, CENP-B and Tigger too. Trends Genet. 13, 141–145. https://doi.org/10.1016/S0168-9525(97)01098-6

Kit, S., 1961. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. J. Mol. Biol. 3, 711–716. https://doi.org/10.1016/S0022-2836(61)80075-2

Kitagawa, K., Masumoto, H., Ikeda, M., Okazaki, T., 1995. Analysis of Protein-DNA and Protein-Protein Interactions of Centromere Protein B (CENP-B) and Properties of the DNA-CENP-B Complex in the Cell Cycle. Mol. Cell. Biol. 15, 1602–1612. https://doi.org/10.1128/MCB.15.3.1602

Klare, K., Weir, J.R., Basilico, F., Zimniak, T., Massimiliano, L., Ludwigs, N., Herzog, F., Musacchio, A., 2015. CENP-C is a blueprint for constitutive centromere–associated network assembly within human kinetochores. J. Cell Biol. 210, 11–22. https://doi.org/10.1083/jcb.201412028

Kornberg, R.D., 1977. Structure of Chromatin. Annu. Rev. Biochem. 46, 931–954. https://doi.org/doi:10.1146/annurev.bi.46.070177.004435

Kornberg, R.D., 1974. Chromatin Structure: A Repeating Unit of Histones and DNA: Chromatin structure is based on a repeating unit of eight histone molecules and about 200 DNA base pairs. Science 184, 868–871. https://doi.org/10.1126/science.184.4139.868

Kornberg, R.D., Thomas, J.O., 1974. Chromatin Structure: Oligomers of the Histones: The histones comprise an (F2A1) $_2$ (F3) $_2$ tetramer, a different oligomer of F2A2 and F2B, and monomer of F1. Science 184, 865–868. https://doi.org/10.1126/science.184.4139.865

Kossel, A., 1884. Ueber einen peptonartigen Bestandtheil des Zellkerns. Z. Für Physiol. Chem. 8, 511–515.

Kossel, A., 1879. Ueber das Nuclein in der Hefe. Z. Für Physiol. Chem. 3, 284–291.

Kouzarides, T., 2007. Chromatin Modifications and Their Function. Cell 128, 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Krzystek-Korpacka, M., Hotowy, K., Czapinska, E., Podkowik, M., Bania, J., Gamian, A., Bednarz-Misa, I., 2016. Serum availability affects expression of common house-keeping genes in colon adenocarcinoma cell lines: implications for quantitative real-time PCR studies. Cytotechnology 68, 2503–2517. https://doi.org/10.1007/s10616-016-9971-4

Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Honda, S., Kobayashi, O., Hatakeyama, K., Ushio, Y., Saya, H., Hirota, T., 2003. CENP-A Phosphorylation by Aurora-A in Prophase Is Required for Enrichment of Aurora-B at Inner Centromeres and for Kinetochore Function. Dev. Cell 5, 853–864. https://doi.org/10.1016/S1534-5807(03)00364-2

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A.V., Barth, T., Cantaloube, S., Kurumizaka, H., Imhof, A., Almouzni, G., 2014. Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. Mol. Cell 53, 631–644. https://doi.org/10.1016/j.molcel.2014.01.018

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R., Li, E., 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122, 3195–3205. https://doi.org/10.1242/dev.122.10.3195

Lei, Y., Huang, Y.-H., Goodell, M.A., 2018. DNA methylation and de-methylation using hybrid site-targeting proteins. Genome Biol. 19, 187. https://doi.org/10.1186/s13059-018-1566-2

Lei, Y., Zhang, X., Su, J., Jeong, M., Gundry, M.C., Huang, Y.-H., Zhou, Y., Li, W., Goodell, M.A., 2017. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat. Commun. 8, 16026. https://doi.org/10.1038/ncomms16026

Levene, P.A., 1919. The structure of yeast nucleic acid. J. Biol. Chem. 40, 415–424. https://doi.org/10.1016/S0021-9258(18)87254-4

Levene, P.A., Jacobs, W.A., 1912. On the structure of thymus nucleic acid. J. Biol. Chem. 12, 411–420. https://doi.org/10.1016/S0021-9258(18)88677-X

Levene, P.A., Tipson, R.S., 1935. The ring structure of thymidine. J. Biol. Chem. 109, 623–630. https://doi.org/10.1016/S0021-9258(18)75193-4

Li, E., Bestor, T.H., Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915–926. https://doi.org/10.1016/0092-8674(92)90611-F

Li, S., Peng, Y., Landsman, D., Panchenko, A.R., 2022. DNA methylation cues in nucleosome geometry, stability and unwrapping. Nucleic Acids Res. 50, 1864–1874. https://doi.org/10.1093/nar/gkac097

Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., Wysocka, J., 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature 553, 228–232. https://doi.org/10.1038/nature25179

Liu, S.-T., Rattner, J.B., Jablonski, S.A., Yen, T.J., 2006. Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. J. Cell Biol. 175, 41–53. https://doi.org/10.1083/jcb.200606020

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., Jaenisch, R., 2016. Editing DNA Methylation in the Mammalian Genome. Cell 167, 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056

Logsdon, G.A., Barrey, E.J., Bassett, E.A., DeNizio, J.E., Guo, L.Y., Panchenko, T., Dawicki-McKenna, J.M., Heun, P., Black, B.E., 2015. Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. J. Cell Biol. 208, 521–531. https://doi.org/10.1083/jcb.201412011

Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., de Lima, L.G., Dvorkina, T., Porubsky, D., Harvey, W.T., Mikheenko, A., Bzikadze, A.V., Kremitzki, M., Graves-Lindsay, T.A., Jain, C., Hoekzema, K., Murali, S.C., Munson, K.M., Baker, C., Sorensen, M., Lewis, A.M., Surti, U., Gerton, J.L., Larionov, V., Ventura, M., Miga, K.H., Phillippy, A.M., Eichler, E.E., 2021. The structure, function and evolution of a complete human chromosome 8. Nature. https://doi.org/10.1038/s41586-021-03420-7

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., Downing, J.R., 2003. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 17, 637–641. https://doi.org/10.1038/sj.leu.2402834

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 A° resolution. Nature 389, 251–260.

Lyon, M.F., 1961. Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature 190, 372–373. https://doi.org/10.1038/190372a0

MacAlpine, D.M., Almouzni, G., 2013. Chromatin and DNA Replication. Cold Spring Harb. Perspect. Biol. 5, 1–22. https://doi.org/10.1101/cshperspect.a010207

Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., Costello, J.F., Wilkinson, M.F., Joung, J.K., 2013. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 31, 1137–1142. https://doi.org/10.1038/nbt.2726

Maeshima, K., Ide, S., Babokhov, M., 2019. Dynamic chromatin organization without the 30-nm fiber. Curr. Opin. Cell Biol. 58, 95–104. https://doi.org/10.1016/j.ceb.2019.02.003

Mahmood, N., Rabbani, S.A., 2019. DNA Methylation Readers and Cancer: Mechanistic and Therapeutic Applications. Front. Oncol. 9, 489. https://doi.org/10.3389/fonc.2019.00489

Malik, H.S., Henikoff, S., 2003. Phylogenomics of the nucleosome. Nat. Struct. Mol. Biol. 10, 882–891. https://doi.org/10.1038/nsb996

Maloney, K.A., Sullivan, L.L., Matheny, J.E., Strome, E.D., Merrett, S.L., Ferris, A., Sullivan, B.A., 2012. Functional epialleles at an endogenous human centromere. Proc. Natl. Acad. Sci. 109, 13704–13709. https://doi.org/10.1073/pnas.1203126109

Manuelidis, L., 1978. Chromosomal localization of complex and simple repeated human DNAs. Chromosoma 66, 23–32. https://doi.org/10.1007/BF00285813

Manuelidis, L., 1976. Repeating restriction fragments of human DNA. Nucleic Acids Res. 3, 3063–3076. https://doi.org/10.1093/nar/3.11.3063

Manuelidis, L., Wu, J.C., 1978. Homology between human and simian repeated DNA. Nature 276, 92–94. https://doi.org/10.1038/276092a0

Maraschio, P., Zuffardi, O., Dalla Fior, T., Tiepolo, L., 1988. Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. J. Med. Genet. 25, 173–180. https://doi.org/10.1136/jmg.25.3.173

Marshall, O.J., Chueh, A.C., Wong, L.H., Choo, K.H.A., 2008a. Neocentromeres: New Insights into Centromere Structure, Disease Development, and Karyotype Evolution. Am. J. Hum. Genet. 82, 261–282. https://doi.org/10.1016/j.ajhg.2007.11.009

Marshall, O.J., Marshall, A.T., Choo, K.H.A., 2008b. Three-dimensional localization of CENP-A suggests a complex higher order structure of centromeric chromatin. J. Cell Biol. 183, 1193–1202. https://doi.org/10.1083/jcb.200804078

Martin, W.F., Garg, S., Zimorski, V., 2015. Endosymbiotic theories for eukaryote origin. Philos. Trans. R. Soc. B Biol. Sci. 370. https://doi.org/10.1098/rstb.2014.0330

Massey, D.J., Koren, A., 2022. Telomere-to-telomere human DNA replication timing profiles. Sci. Rep. 12, 9560. https://doi.org/10.1038/s41598-022-13638-8

Masumoto, H., Masukata, H., Muro, Yoshinao, Naohito, Nozaki, Okazaki,Tuneko, 1989. A Human Centromere Antigen (CENP-B) Interacts with a Short Specific Sequence an Alphoid DNA, a Human Centromeric Satellite. J. Cell Biol. 109, 1963–1973.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Boeke, J.D., Gabriel, A., 1991. Reverse Transcriptase Encoded by a Human Transposable Element. Science 254, 1808–1810. https://doi.org/10.1126/science.1722352

Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Demethylation of the zygotic paternal genome. Nature 403, 501–502. https://doi.org/10.1038/35000656

McAinsh, A.D., Kops, G.J.P.L., 2023. Principles and dynamics of spindle assembly checkpoint signalling. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/s41580-023-00593-z

McKinley, K.L., Sekulic, N., Guo, L.Y., Tsinman, T., Black, B.E., Cheeseman, I.M., 2015. The CENP-L-N Complex Forms a Critical Node in an Integrated Meshwork of Interactions at the Centromere-Kinetochore Interface. Mol. Cell 60, 886–898. https://doi.org/10.1016/j.molcel.2015.10.027

McNulty, S.M., Sullivan, B.A., 2018. Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosome Res. 26, 115–138. https://doi.org/10.1007/s10577-018-9582-3

McNulty, S.M., Sullivan, L.L., Sullivan, B.A., 2017. Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Dev. Cell 42, 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001

Melters, D.P., Paliulis, L.V., Korf, I.F., Chan, S.W.L., 2012. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Res. 20, 579–593. https://doi.org/10.1007/s10577-012-9292-1

Mendel, G., 1865. Versuche über Pflanzen-Hybriden. Verhandlungen Naturforschenden Vereines Brünn Bd.4, 3–47.

Mendiburo, M.J., Padeken, J., Fülöp, S., Schepers, A., Heun, P., 2011. *Drosophila* CENH3 Is Sufficient for Centromere Formation. Science 334, 686–690. https://doi.org/10.1126/science.1206880

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., Sidransky, D., 1995. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat. Med. 1, 686–692. https://doi.org/10.1038/nm0795-686

Miescher, F., 1874. Das Protamin, eine neue organische Base aus den Samenfäden des Rheinlachses. Berichte Dtsch. Chem. Ges. 7, 376–379. https://doi.org/10.1002/cber.187400701119

Miescher, F., 1871. Ueber die chemische Zusammensetzung der Eiterzellen. Med.-Chem. Untersuchungen 4, 441–460.

Miga, K.H., 2017a. The Promises and Challenges of Genomic Studies of Human Centromeres, in: Black, B.E. (Ed.), Centromeres and Kinetochores, Progress in Molecular and Subcellular Biology. Springer International Publishing, Cham, pp. 285–304. https://doi.org/10.1007/978-3-319-58592-5_12

Miga, K.H., 2017b. Chromosome-Specific Centromere Sequences Provide an Estimate of the Ancestral Chromosome 2 Fusion Event in Hominin Genomes. J. Hered. 108, 45–52. https://doi.org/10.1093/jhered/esw039

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., Meissner, A., 2008. Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49–55. https://doi.org/10.1038/nature07056

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., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560. https://doi.org/10.1038/nature06008

Miniou, P., Jeanpierre, M., Bourc'his, D., Barbosa, A.C.C., Blanquet, V., Viegas-Péquignot, E., 1997. α-Satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. Hum. Genet. 99, 738–745. https://doi.org/10.1007/s004390050441

Minoshima, Y., Hori, T., Okada, M., Kimura, H., Haraguchi, T., Hiraoka, Y., Bao, Y.-C., Kawashima, T., Kitamura, T., Fukagawa, T., 2005. The Constitutive Centromere Component CENP-50 Is Required for Recovery from Spindle Damage. Mol. Cell. Biol. 25, 10315–10328. https://doi.org/10.1128/MCB.25.23.10315-10328.2005

Mirsky, A.E., Pollister, A.W., 1946. Chromosin, a desoxyribose nucleoprotein complex of the cell nucleus. J. Gen. Physiol. 30, 117–148. https://doi.org/10.1085/jgp.30.2.117

Mitchell, A.R., Jeppesen, P., Nicol, L., Morrison, H., Kipling, D., 1996. Epigenetic control of mammalian centromere protein binding: does DNA methylation have a role? J. Cell Sci. 109, 2199–2206.

Molina, O., Vargiu, G., Abad, M.A., Zhiteneva, A., Jeyaprakash, A.A., Masumoto, H., Kouprina, N., Larionov, V., Earnshaw, W.C., 2016. Epigenetic engineering reveals a balance between histone modifications and transcription in kinetochore maintenance. Nat. Commun. 7, 13334. https://doi.org/10.1038/ncomms13334

Monk, M., Boubelik, M., Lehnert, S., 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 99, 371–382. https://doi.org/10.1242/dev.99.3.371

Mora-Bermúdez, F., Gerlich, D., Ellenberg, J., 2007. Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. Nat. Cell Biol. 9, 822–831. https://doi.org/10.1038/ncb1606

Moree, B., Meyer, C.B., Fuller, C.J., Straight, A.F., 2011. CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J. Cell Biol. 194, 855–871. https://doi.org/10.1083/jcb.201106079

Morgan, T.H., 1913. Heredity and sex. Columbia University Press, New York. https://doi.org/10.5962/bhl.title.46224

Morgan, T.H., Sturtevant, A. h., Muller, H.J., Bridges, C.B., 1915. The mechanism of Mendelian heredity, Henry Hold and Company. ed. Columbia University. Libraries. Digital Program Division, New York.

Moroi, Y., Peebles, C., Fritzler, M.J., Steigerwald, J., Tan, E.M., 1980. Autoantibody to centromere (kinetochore) in scleroderma sera. Proc. Natl. Acad. Sci. 77, 1627–1631. https://doi.org/10.1073/pnas.77.3.1627

Moullan, N., Mouchiroud, L., Wang, X., Ryu, D., Williams, E.G., Mottis, A., Jovaisaite, V., Frochaux, M.V., Quiros, P.M., Deplancke, B., Houtkooper, R.H., Auwerx, J., 2015. Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research. Cell Rep. 10, 1681–1691. https://doi.org/10.1016/j.celrep.2015.02.034

Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393, 386–389. https://doi.org/10.1038/30764

Narayan, A., Ji, W., Zhang, X.-Y., Marrogi, A., Graff, J.R., Baylin, S.B., Ehrlich, M., 1998. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. Int. J. Cancer 77, 833–838. https://doi.org/10.1002/(SICI)1097-0215(19980911)77:6<833::AID-IJC6>3.0.CO;2-V

Naughton, C., Huidobro, C., Catacchio, C.R., Buckle, A., Grimes, G.R., Nozawa, R.-S., Purgato, S., Rocchi, M., Gilbert, N., 2022. Human centromere repositioning activates transcription and opens chromatin fibre structure. Nat. Commun. 13, 5609. https://doi.org/10.1038/s41467-022-33426-2

Nechemia-Arbely, Y., Fachinetti, D., Miga, K.H., Sekulic, N., Soni, G.V., Kim, D.H., Wong, A.K., Lee, A.Y., Nguyen, K., Dekker, C., Ren, B., Black, B.E., Cleveland, D.W., 2017. Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. J. Cell Biol. 216, 607–621. https://doi.org/10.1083/jcb.201608083

Nechemia-Arbely, Y., Miga, K.H., Shoshani, O., Aslanian, A., McMahon, M.A., Lee, A.Y., Fachinetti, D., Yates, J.R., Ren, B., Cleveland, D.W., 2019. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. Nat. Cell Biol. 21, 743–754. https://doi.org/10.1038/s41556-019-0331-4

Neri, F., Incarnato, D., Krepelova, A., Parlato, C., Oliviero, S., 2016. Methylation-assisted bisulfite sequencing to simultaneously map 5fC and 5caC on a genome-wide scale for DNA demethylation analysis. Nat. Protoc. 11, 1191–1205. https://doi.org/10.1038/nprot.2016.063

Niikura, Y., Kitagawa, R., Fang, L., Kitagawa, K., 2019. CENP-A Ubiquitylation Is Indispensable to Cell Viability. Dev. Cell 50, 683-689.e6. https://doi.org/10.1016/j.devcel.2019.07.015

Niikura, Y., Kitagawa, R., Ogi, H., Abdulle, R., Pagala, V., Kitagawa, K., 2015. CENP-A K124 Ubiquitylation Is Required for CENP-A Deposition at the Centromere. Dev. Cell 32, 589–603. https://doi.org/10.1016/j.devcel.2015.01.024

Niikura, Y., Kitagawa, R., Ogi, H., Kitagawa, K., 2017. SGT1-HSP90 complex is required for CENP-A deposition at centromeres. Cell Cycle 16, 1683–1694. https://doi.org/10.1080/15384101.2017.1325039

Nishihashi, A., Haraguchi, T., Hiraoka, Y., Ikemura, T., Regnier, V., Dodson, H., Earnshaw, W.C., Fukagawa, T., 2002. CENP-I Is Essential for Centromere Function in Vertebrate Cells. Dev. Cell 2, 463–476. https://doi.org/10.1016/S1534-5807(02)00144-2

Nishino, T., Takeuchi, K., Gascoigne, K.E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I.M., Fukagawa, T., 2012. CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. Cell 148, 487–501. https://doi.org/10.1016/j.cell.2011.11.061

Nishiyama, A., Mulholland, C.B., Bultmann, S., Kori, S., Endo, A., Saeki, Y., Qin, W., Trummer, C., Chiba, Y., Yokoyama, H., Kumamoto, S., Kawakami, T., Hojo, H., Nagae, G., Aburatani, H., Tanaka, K., Arita, K., Leonhardt, H., Nakanishi, M., 2020. Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation. Nat. Commun. 11, 1222. https://doi.org/10.1038/s41467-020-15006-4

Nishiyama, A., Nakanishi, M., 2021. Navigating the DNA methylation landscape of cancer. Trends Genet. 37, 1012–1027. https://doi.org/10.1016/j.tig.2021.05.002

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims, Y., Smit, A.F.A., Soto, D.C., Sovic, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., Schatz, M.C., Eichler, E.E., Miga, K.H., Phillippy, A.M., 2022. The complete sequence of a human genome. Science 376, 44–53.

O'Connor, C., 2008. Cell Division: Stages of Mitosis. Nature Education 1(1):188 [WWW Document]. URL http://www.nature.com/scitable/topicpage/mitosis-and-cell-division-205 (accessed 5.1.23).

O'Connor, C., Adams, J.U., 2010. Essentials of Cell Biology. Cambridge, MA: NPG Education [WWW Document]. URL https://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/ (accessed 5.1.23).

Oegema, K., Desai, A., Rybina, S., Kirkham, M., Hyman, A.A., 2001. Functional Analysis of Kinetochore Assembly in Caenorhabditis elegans. J. Cell Biol. 153, 1209–1226. https://doi.org/10.1083/jcb.153.6.1209

Ohzeki, J., Larionov, V., Earnshaw, W.C., Masumoto, H., 2019. De novo formation and epigenetic maintenance of centromere chromatin. Curr. Opin. Cell Biol. 58, 15–25. https://doi.org/10.1016/j.ceb.2018.12.004

Ohzeki, J., Nakano, M., Okada, T., Masumoto, H., 2002. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J. Cell Biol. 159, 765–775. https://doi.org/10.1083/jcb.200207112

Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., Desai, A., Fukagawa, T., 2006. The CENP-H–I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat. Cell Biol. 8, 446–457. https://doi.org/10.1038/ncb1396

Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., Masumoto, H., 2007. CENP-B Controls Centromere Formation Depending on the Chromatin Context. Cell 131, 1287–1300. https://doi.org/10.1016/j.cell.2007.10.045

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell 99, 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Okano, M., Xie, S., Li, E., 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. 19, 219–220. https://doi.org/10.1038/890

Okazaki, K., Nakano, M., Ohzeki, J., Otake, K., Kugou, K., Larionov, V., Earnshaw, W.C., Masumoto, H., 2022. Combination of CENP-B Box Positive and Negative Synthetic Alpha Satellite Repeats Improves De Novo Human Artificial Chromosome Formation. Cells 11, 1378. https://doi.org/10.3390/cells11091378

Olins, A.L., Olins, D.E., 1974. Spheroid Chromatin Units (ν Bodies). Science 183, 330–332. https://doi.org/10.1126/science.183.4122.330

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., Hayashi, Y., 2002. LCX, Leukemia-associated Protein with a CXXC Domain, Is Fused to MLL in Acute Myeloid Leukemia with Trilineage Dysplasia Having t(10;11)(q22;q23). Cancer Res. 62, 4075–4080.

Osanai-Futahashi, M., Suetsugu, Y., Mita, K., Fujiwara, H., 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, Bombyx mori. Insect Biochem. Mol. Biol. 38, 1046–1057. https://doi.org/10.1016/j.ibmb.2008.05.012

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., O'Shea, C.C., 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 357, eaag0025. https://doi.org/10.1126/science.aag0025

Oudet, P., Gross-Bellard, M., Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4, 281–300. https://doi.org/10.1016/0092-8674(75)90149-X

Paedon, J.F., Wilkins, M.H.F., 1972. A super-coil model for nucleohistone. J. Mol. Biol. 68, 115–124. https://doi.org/10.1016/0022-2836(72)90267-7

Palmer, D.K., O'Day, K., Margolis, R.L., 1990. The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma 100, 32–36. https://doi.org/10.1007/BF00337600

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., Margolis, R.L., 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. 88, 3734–3738. https://doi.org/10.1073/pnas.88.9.3734

Palmer, D.K., O'Day, K., Wener, M., Andrews, B., Margolis, R., 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J. Cell Biol. 104, 805–815. https://doi.org/10.1083/jcb.104.4.805

Pappalardo, X.G., Barra, V., 2021. Losing DNA methylation at repetitive elements and breaking bad. Epigenetics Chromatin 14, 25. https://doi.org/10.1186/s13072-021-00400-z

Pardue, M.L., Gall, J.G., 1970. Chromosomal Localization of Mouse Satellite DNA. Science 168, 1356–1358.

Paro, R., Grossniklaus, U., Santoro, R., Wutz, A., 2021. Dosage Compensation Systems, in: Introduction to Epigenetics, Learning Materials in Biosciences. Springer International Publishing, Cham, pp. 67–89. https://doi.org/10.1007/978-3-030-68670-3_4

Passarge, E., 1979. Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation was Recognized Fifty Years Ago. Am. J. Hum. Genet. 31, 106–115.

Perea-Resa, C., Blower, M.D., 2018. Centromere Biology: Transcription Goes on Stage. Mol. Cell. Biol. 38, e00263-18. https://doi.org/10.1128/MCB.00263-18

Perez-Castro, A.V., Shamanski, F.L., Meneses, J.J., Lovato, T.L., Vogel, K.G., Moyzis, R.K., Pedersen, R., 1998. Centromeric Protein B Null Mice Are Viable with No Apparent Abnormalities. Dev. Biol. 201, 135–143. https://doi.org/10.1006/dbio.1998.9005

Pesenti, M.E., Prumbaum, D., Auckland, P., Smith, C.M., Faesen, A.C., Petrovic, A., Erent, M., Maffini, S., Pentakota, S., Weir, J.R., Lin, Y.-C., Raunser, S., McAinsh, A.D., Musacchio, A., 2018. Reconstitution of a 26-Subunit Human Kinetochore Reveals Cooperative Microtubule Binding by CENP-OPQUR and NDC80. Mol. Cell 71, 923-939.e10. https://doi.org/10.1016/j.molcel.2018.07.038

Pesenti, M.E., Raisch, T., Conti, D., Walstein, K., Hoffmann, I., Vogt, D., Prumbaum, D., Vetter, I.R., Raunser, S., Musacchio, A., 2022. Structure of the human inner kinetochore CCAN complex and its significance for human centromere organization. Mol. Cell 82, 2113-2131.e8. https://doi.org/10.1016/j.molcel.2022.04.027

Petryk, N., Bultmann, S., Bartke, T., Defossez, P.-A., 2021. Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucleic Acids Res. 49, 3020–3032. https://doi.org/10.1093/nar/gkaa1154

Pidoux, A.L., Allshire, R.C., 2005. The role of heterochromatin in centromere function. Philos. Trans. R. Soc. B Biol. Sci. 360, 569–579. https://doi.org/10.1098/rstb.2004.1611

Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M., Vitale, L., 2019. On the length, weight and GC content of the human genome. BMC Res. Notes 12, 106. https://doi.org/10.1186/s13104-019-4137-z

Piras, F.M., Nergadze, S.G., Magnani, E., Bertoni, L., Attolini, C., Khoriauli, L., Raimondi, E., Giulotto, E., 2010. Uncoupling of Satellite DNA and Centromeric Function in the Genus Equus. PLoS Genet. 6, e1000845. https://doi.org/10.1371/journal.pgen.1000845

Plósz, P., 1871. Ueber das chemische Verhalten der Kerne der Vogel-und Schlangenblutkörperchen. Med.-Chem. Untersuchungen 4, 461–462.

Policarpi, C., Dabin, J., Hackett, J.A., 2021. Epigenetic editing: Dissecting chromatin function in context. BioEssays 43, 2000316. https://doi.org/10.1002/bies.202000316

Pradhan, S., Bacolla, A., Wells, R.D., Roberts, R.J., 1999. Recombinant Human DNA (Cytosine-5) Methyltransferase. J. Biol. Chem. 274, 33002–33010. https://doi.org/10.1074/jbc.274.46.33002

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M., Glover, D.M., 2011. CENP-C Is a Structural Platform for Kinetochore Assembly. Curr. Biol. 21, 399–405. https://doi.org/10.1016/j.cub.2011.02.005

Qu, G., Grundy, P.E., Narayan, A., Ehrlich, M., 1999. Frequent Hypomethylation in Wilms Tumors of Pericentromeric DNA in Chromosomes 1 and 16. Cancer Genet. Cytogenet. 109, 34–39. https://doi.org/10.1016/S0165-4608(98)00143-5

Quénet, D., Dalal, Y., 2014. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. eLife 3, e26016. https://doi.org/10.7554/eLife.03254

Ray-Gallet, D., Almouzni, G., 2021. The Histone H3 Family and Its Deposition Pathways, in: Fang, D., Han, J. (Eds.), Histone Mutations and Cancer, Advances in Experimental Medicine and Biology. Springer Singapore, Singapore, pp. 17–42. https://doi.org/10.1007/978-981-15-8104-5_2

Régnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W., Brown, W., 2005. CENP-A Is Required for Accurate Chromosome Segregation and Sustained Kinetochore Association of BubR1. Mol. Cell. Biol. 25, 3967–3981. https://doi.org/10.1128/MCB.25.10.3967-3981.2005

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-W., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., Baylin, S.B., Vogelstein, B., 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 416, 552–556. https://doi.org/10.1038/416552a

Rhie, A., Nurk, S., Cechova, M., Hoyt, S.J., Taylor, D.J., Altemose, N., Hook, P.W., Koren, S., Rautiainen, M., Alexandrov, I.A., Allen, J., Asri, M., Bzikadze, A.V., Chen, N.-C., Chin, C.-S., Diekhans, M., Flicek, P., Formenti, G., Fungtammasan, A., Garcia Giron, C., Garrison, E., Gershman, A., Gerton, J., Grady, P.G., Guarracino, A., Haggerty, L., Halabian, R., Hansen, N.F., Harris, R., Hartley, G.A., Harvey, W.T., Haukness, M., Heinz, J., Hourlier, T., Hubley, R.M., Hunt, S.E., Hwang, S., Jain, M., Kesharwani, R.K., Lewis, A.P., Li, H., Logsdon, G.A., Lucas, J.K., Makalowski, W., Markovic, C., Martin, F.J., Mc Cartney, A.M., McCoy, R.C., McDaniel, J., McNulty, B.M., Medvedev, P., Mikheenko, A., Munson, K.M., Murphy, T.D., Olsen, H.E., Olson, N.D., Paulin, L.F., Porubsky, D., Potapova, T., Ryabov, F., Salzberg, S.L., Sauria, M.E., Sedlazeck, F.J., Shafin, K., Shepelev, V.A., Shumate, A., Storer, J.M., Surapaneni, L., Taravella Oill, A.M., Thibaud-Nissen, F., Timp, W., Tomaszkiewicz, M., Vollger, M.R., Walenz, B.P., Watwood, A.C., Weissensteiner, M.H., Wenger, A.M., Wilson, M.A., Zarate, S., Zhu, Y., Zook, J.M., Eichler, E.E., O'Neill, R., Schatz, M.C., Miga, K.H., Makova, K.D., Phillippy, A.M., 2022. The complete sequence of a human Y chromosome. bioRxiv 518724. https://doi.org/10.1101/2022.12.01.518724

Ribeiro, S.A., Vagnarelli, P., Dong, Y., Hori, T., McEwen, B.F., Fukagawa, T., Flors, C., Earnshaw, W.C., 2010. A super-resolution map of the vertebrate kinetochore. Proc. Natl. Acad. Sci. 107, 10484–10489. https://doi.org/10.1073/pnas.1002325107

Richardson, B., 2003. DNA methylation and autoimmune disease. Clin. Immunol. 109, 72–79. https://doi.org/10.1016/S1521-6616(03)00206-7

Rideout, W.M., Coetzee, G.A., Olumi, A.F., Jones, P.A., 1990. 5-Methylcytosine as an Endogenous Mutagen in the Human LDL Receptor and p53 Genes. Science 249, 1288–1290. https://doi.org/10.1126/science.1697983

Riggs, A.D., 1975. X inactivation, differentiation, and DNA methylation. Cytogenet. Genome Res. 14, 9–25. https://doi.org/10.1159/000130315

Robertson, K.D., 2005. DNA methylation and human disease. Nat. Rev. Genet. 6, 597–610. https://doi.org/10.1038/nrg1655

Rosenberg, H., Singer, M., Rosenberg, M., 1978. Highly Reiterated Sequences of SIMIANSIMIANSIMIANSIMIANSIMIAN. Science 200, 394–402. https://doi.org/10.1126/science.205944

Rošić, S., Erhardt, S., 2016. No longer a nuisance: long non-coding RNAs join CENP-A in epigenetic centromere regulation. Cell. Mol. Life Sci. 73, 1387–1398. https://doi.org/10.1007/s00018-015-2124-7

Rošić, S., Köhler, F., Erhardt, S., 2014. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J. Cell Biol. 207, 335–349. https://doi.org/10.1083/jcb.201404097

Rusk, N., 2012. The sixth base and counting. Nat. Methods 9, 646–646. https://doi.org/10.1038/nmeth.2095

Sacristan, C., Samejima, K., Ruiz, L.A., Lambers, M.L.A., Buckle, A., Brackley, C.A., Robertson, D., Hori, T., Webb, S., Fukagawa, T., Gilbert, N., Marenduzzo, D., Earnshaw, W.C., Kops, G.J.P.L., 2022. Condensin reorganizes centromeric chromatin during mitotic entry into a bipartite structure stabilized by cohesin (preprint). Cell Biology. https://doi.org/10.1101/2022.08.01.502248

Sado, T., Fenner, M.H., Tan, S.-S., Tam, P., Shioda, T., Li, E., 2000. X Inactivation in the Mouse Embryo Deficient for Dnmt1: Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation. Dev. Biol. 225, 294–303. https://doi.org/10.1006/dbio.2000.9823

Saffery, R., Sumer, H., Hassan, S., Wong, L.H., Craig, J.M., Todokoro, K., Anderson, M., Stafford, A., Choo, K.H.A., 2003. Transcription within a Functional Human Centromere. Mol. Cell 12, 509–516. https://doi.org/10.1016/S1097-2765(03)00279-X

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., Maurer, M., Rothfield, N.F., Earnshaw, W.C., 1992. CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. Cell 70, 115–125. https://doi.org/10.1016/0092-8674(92)90538-N

Salinas-Luypaert, C., Allu, P.K., Logsdon, G.A., Dawicki-McKenna, J.M., Gambogi, C.W., Fachinetti, D., Black, B.E., 2021. Gene replacement strategies validate the use of functional tags on centromeric chromatin and invalidate an essential role for CENP-AK124ub. Cell Rep. 37, 109924. https://doi.org/10.1016/j.celrep.2021.109924

Sanford, J.P., Clark, H.J., Chapman, V.M., Rossant, J., 1987. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. 1, 1039–1046. https://doi.org/10.1101/gad.1.10.1039

Sapozhnikov, D.M., Szyf, M., 2021. Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methyltransferase with CRISPR/dCas9. Nat. Commun. 12, 5711. https://doi.org/10.1038/s41467-021-25991-9

Scelfo, A., Barra, V., Abdennur, N., Spracklin, G., Busato, F., Salinas-Luypaert, C., Bonaiti, E., Velasco, G., Chipont, A., Guérin, C., Tijhuis, A.E., Spierings, D.C.J., Francastel, C., Foijer, F., Tost, J., Mirny, L., Fachinetti, D., 2023. Tunable DNMT1 degradation reveals cooperation of DNMT1 and DNMT3B in regulating DNA methylation dynamics and genome organization. bioRxiv. https://doi.org/10.1101/2023.05.04.539406

Scelfo, Fachinetti, 2019. Keeping the Centromere under Control: A Promising Role for DNA Methylation. Cells 8, 912. https://doi.org/10.3390/cells8080912

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schwartz, D.C., Cantor, C.R., 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37, 67–75. https://doi.org/10.1016/0092-8674(84)90301-5

Scott, K.C., Sullivan, B.A., 2014. Neocentromeres: a place for everything and everything in its place. Trends Genet. 30, 66–74. https://doi.org/10.1016/j.tig.2013.11.003

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E., Musacchio, A., 2011. Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochore. Curr. Biol. 21, 391–398. https://doi.org/10.1016/j.cub.2010.12.039

Sekulic, N., Bassett, E.A., Rogers, D.J., Black, B.E., 2010. The structure of (CENP-A–H4)2 reveals physical features that mark centromeres. Nature 467, 347–351. https://doi.org/10.1038/nature09323

Senaratne, A.P., Cortes-Silva, N., Drinnenberg, I.A., 2022. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. Semin. Cell Dev. Biol. 127, 90–99. https://doi.org/10.1016/j.semcdb.2022.01.003

Shang, W.-H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., Fukagawa, T., 2010. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. 20, 1219–1228. https://doi.org/10.1101/gr.106245.110

Shelby, R.D., Monier, K., Sullivan, K.F., 2000. Chromatin Assembly at Kinetochores Is Uncoupled from DNA Replication. J. Cell Biol. 151, 1113–1118. https://doi.org/10.1083/jcb.151.5.1113

Shelby, R.D., Vafa, O., Sullivan, K.F., 1997. Assembly of CENP-A into Centromeric Chromatin Requires a Cooperative Array of Nucleosomal DNA Contact Sites. J. Cell Biol. 136, 501–513. https://doi.org/10.1083/jcb.136.3.501

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., Zhang, Y., 2014. Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. Cell Stem Cell 15, 459–471. https://doi.org/10.1016/j.stem.2014.09.002

Shimbo, T., Wade, P.A., 2016. Proteins That Read DNA Methylation, in: Jeltsch, A., Jurkowska, R.Z. (Eds.), DNA Methyltransferases - Role and Function, Advances in Experimental Medicine and Biology. Springer International Publishing, Cham, pp. 303–320. https://doi.org/10.1007/978-3-319-43624-1_13

Shuaib, M., Ouararhni, K., Dimitrov, S., Hamiche, A., 2010. HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc. Natl. Acad. Sci. 107, 1349–1354. https://doi.org/10.1073/pnas.0913709107

Sinsheimer, R.L., Koerner, J.F., Vadla, J., Lunan, K., 1954. The action of pancreatic desoxyribonuclease. Isolation of mono- and dinucleotides. J. Biol. Chem. 208, 445–459. https://doi.org/10.1016/S0021-9258(18)65663-7

Skene, P.J., Henikoff, S., 2017. CUT&RUN: Targeted in situ genome-wide profiling with high efficiency for low cell numbers (preprint). Genomics. https://doi.org/10.1101/193219

Smit, A.F., 1996. The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. 6, 743–748. https://doi.org/10.1016/S0959-437X(96)80030-X

Smith, M.M., 2002. Centromeres and variant histones: what, where, when and why? Curr. Opin. Cell Biol. 14, 279–285. https://doi.org/10.1016/S0955-0674(02)00331-9

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., Li, G., 2014. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science 344, 376–380. https://doi.org/10.1126/science.1251413

Šorm, F., Píškala, A., Čihák, A., Veselý, J., 1964. 5-Azacytidine, a new, highly effective cancerostatic. Experientia 20, 202–203. https://doi.org/10.1007/BF02135399

Spracklin, G., Abdennur, N., Imakaev, M., Chowdhury, N., Pradhan, S., Mirny, L.A., Dekker, J., 2023. Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. Nat. Struct. Mol. Biol. 30, 38–51. https://doi.org/10.1038/s41594-022-00892-7

Srivastava, S., Foltz, D.R., 2018. Posttranslational modifications of CENP-A: marks of distinction. Chromosoma 127, 279–290. https://doi.org/10.1007/s00412-018-0665-x

Stedman, Edgar, Stedman, Ellen, 1950. Cell Specificity of Histones. Nature 166, 780–781. https://doi.org/10.1038/166780a0

Stresemann, C., Lyko, F., 2008. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int. J. Cancer 123, 8–13. https://doi.org/10.1002/ijc.23607

Stricker, S.H., Köferle, A., Beck, S., 2017. From profiles to function in epigenomics. Nat. Rev. Genet. 18, 51–66. https://doi.org/10.1038/nrg.2016.138

Su, J., Huang, Y.-H., Cui, X., Wang, X., Zhang, X., Lei, Y., Xu, J., Lin, X., Chen, K., Lv, J., Goodell, M.A., Li, W., 2018. Homeobox oncogene activation by pan-cancer DNA hypermethylation. Genome Biol. 19, 108. https://doi.org/10.1186/s13059-018-1492-3

Sugata, N., Li, S., Earnshaw, W.C., Yen, T.J., Yoda, K., Masumoto, H., Munekata, E., Warburton, P.E., Todokoro, K., 2000. Human CENP-H multimers colocalize with CENP-A and CENP-C at active centromere-kinetochore complexes. Hum. Mol. Genet. 9, 2919–2926. https://doi.org/10.1093/hmg/9.19.2919

Sullivan, B.A., Karpen, G.H., 2004. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat. Struct. Mol. Biol. 11, 1076–1083. https://doi.org/10.1038/nsmb845

Sullivan, K.F., Hechenberger, M., Masri, K., 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J. Cell Biol. 127, 581–592. https://doi.org/10.1083/jcb.127.3.581

Sumner, A.T., Mitchell, A.R., Ellis, P.M., 1998. A FISH study of chromosome fusion in the ICF syndrome: involvement of paracentric heterochromatin but not of the centromeres themselves. J. Med. Genet. 35, 833–835. https://doi.org/10.1136/jmg.35.10.833

Sutton, W.S., 1903. The chromosomes in heredity. Biol. Bull. 4, 231–250. https://doi.org/10.2307/1535741

Sutton, W.S., 1902. On the morphology of the chromoso group in Brachystola magna. Biol. Bull. 4, 24–39. https://doi.org/10.2307/1535510

Suzuki, N., Nakano, M., Nozaki, N., Egashira, S., Okazaki, T., Masumoto, H., 2004. CENP-B Interacts with CENP-C Domains Containing Mif2 Regions Responsible for Centromere Localization. J. Biol. Chem. 279, 5934–5946. https://doi.org/10.1074/jbc.M306477200

Suzuki, T., Fujii, M., Ayusawa, D., 2002. Demethylation of classical satellite 2 and 3 DNA with chromosomal instability in senescent human fibroblasts. Exp. Gerontol. 37, 1005–1014. https://doi.org/10.1016/S0531-5565(02)00061-X

Swartz, M.N., Trautner, T.A., Kornberg, A., 1962. Enzymatic Synthesis of Deoxyribonucleic Acid. J. Biol. Chem. 237, 1961–1967. https://doi.org/10.1016/S0021-9258(19)73967-2

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., Kimura, H., Kurumizaka, H., 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476, 232–235. https://doi.org/10.1038/nature10258

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science 324, 930–935. https://doi.org/10.1126/science.1170116

Takada, M., Zhang, W., Suzuki, A., Kuroda, T.S., Yu, Z., Inuzuka, H., Gao, D., Wan, L., Zhuang, M., Hu, L., Zhai, B., Fry, C.J., Bloom, K., Li, G., Karpen, G.H., Wei, W., Zhang, Q., 2017. FBW7 Loss Promotes Chromosomal Instability and Tumorigenesis via Cyclin E1/CDK2–Mediated Phosphorylation of CENP-A. Cancer Res. 77, 4881–4893. https://doi.org/10.1158/0008-5472.CAN-17-1240

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takizawa, Y., Ho, C.-H., Tachiwana, H., Matsunami, H., Kobayashi, W., Suzuki, M., Arimura, Y., Hori, T., Fukagawa, T., Ohi, M.D., Wolf, M., Kurumizaka, H., 2020. Cryo-EM Structures of Centromeric Tri-nucleosomes Containing a Central CENP-A Nucleosome. Structure 28, 44-53.e4. https://doi.org/10.1016/j.str.2019.10.016

Takizawa, Y., Kurumizaka, H., 2022. Chromatin structure meets cryo-EM: Dynamic building blocks of the functional architecture. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1865, 194851. https://doi.org/10.1016/j.bbagrm.2022.194851

Talbert, P.B., Henikoff, S., 2021. The Yin and Yang of Histone Marks in Transcription. Annu. Rev. Genomics Hum. Genet. 22, 147–170. https://doi.org/10.1146/annurev-genom-120220-085159

Tanaka, Y., Kurumizaka, H., Yokoyama, S., 2004. CpG methylation of the CENP-B box reduces human CENP-B binding. FEBS J. 272, 282–289. https://doi.org/10.1111/j.1432-1033.2004.04406.x

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, Iwahara, J., Okazaki, T., Yokoyama, S., 2001. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. EMBO J. 20, 6612–6618. https://doi.org/10.1093/emboj/20.23.6612

Tanaka, Y., Tachiwana, H., Yoda, K., Masumoto, H., Okazaki, T., Kurumizaka, H., Yokoyama, S., 2005. Human Centromere Protein B Induces Translational Positioning of Nucleosomes on α-Satellite Sequences. J. Biol. Chem. 280, 41609–41618. https://doi.org/10.1074/jbc.M509666200

Tawaramoto, M.S., Park, S.-Y., Tanaka, Y., Nureki, O., Kurumizaka, H., Yokoyama, S., 2003. Crystal Structure of the Human Centromere Protein B (CENP-B) Dimerization Domain at 1.65-Å Resolution. J. Biol. Chem. 278, 51454–51461. https://doi.org/10.1074/jbc.M310388200

Tazi, J., Bird, A.P., 1990. Alternative chromatin structure at CpG islands. Cell 60, 909–920. https://doi.org/10.1016/0092-8674(90)90339-G

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., Lehner, P.J., 2015. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science 348, 1481–1485. https://doi.org/10.1126/science.aaa7227

Thakur, J., Henikoff, S., 2018. Unexpected conformational variations of the human centromeric chromatin complex. Genes Dev. 32, 20–25. https://doi.org/10.1101/gad.307736.117

Thess, A., Hoerr, I., Panah, B.Y., Jung, G., Dahm, R., 2021. Historic nucleic acids isolated by Friedrich Miescher contain RNA besides DNA. Biol. Chem. 402, 1179–1185. https://doi.org/10.1515/hsz-2021-0226

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., Lemmers, R.J.L.F., de Greef, J.C., Gennery, A., Picco, P., Kloeckener-Gruissem, B., Güngör, T., Reisli, I., Picard, C., Kebaili, K., Roquelaure, B., Iwai, T., Kondo, I., Kubota, T., van Ostaijen-Ten Dam, M.M., van Tol, M.J.D., Weemaes, C., Francastel, C., van der Maarel, S.M., Sasaki, H., 2015. Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome. Nat. Commun. 6, 7870. https://doi.org/10.1038/ncomms8870

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., Turner, D.J., Illingworth, R., Bird, A., 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082–1086. https://doi.org/10.1038/nature08924

Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G.F., Romano, C., 1979. Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. Hum. Genet. 51, 127–137. https://doi.org/10.1007/BF00287166

Toubiana, S., Velasco, G., Chityat, A., Kaindl, A.M., Hershtig, N., Tzur-Gilat, A., Francastel, C., Selig, S., 2018. Subtelomeric methylation distinguishes between subtypes of Immunodeficiency, Centromeric instability and Facial anomalies syndrome. Hum. Mol. Genet. 27, 3568–3581. https://doi.org/10.1093/hmg/ddy265

Tremblay, M.W., Jiang, Y., 2019. DNA Methylation and Susceptibility to Autism Spectrum Disorder. Annu. Rev. Med. 70, 151–166. https://doi.org/10.1146/annurev-med-120417-091431

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R., Nakayama, J., Okano, M., 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells 11, 805–814. https://doi.org/10.1111/j.1365-2443.2006.00984.x

Tucci, V., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., Tucci, V., Bartolomei, M.S., Benvenisty, N., Bourc'his, D., Charalambous, M., Dulac, C., Feil, R., Glaser, J., Huelsmann, L., John, R.M., McNamara, G.I., Moorwood, K., Muscatelli, F., Sasaki, H., Strassmann, B.I., Vincenz, C., Wilkins, J., Isles, A.R., Kelsey, G., Ferguson-Smith, A.C., 2019. Genomic Imprinting and Physiological Processes in Mammals. Cell 176, 952–965. https://doi.org/10.1016/j.cell.2019.01.043

Tweedie, S., Charlton, J., Clark, V., Bird, A., 1997. Methylation of Genomes and Genes at the Invertebrate-Vertebrate Boundary. Mol. Cell. Biol. 17, 1469–1475. https://doi.org/10.1128/MCB.17.3.1469

Unoki, M., 2021. Chromatin remodeling in replication-uncoupled maintenance DNA methylation and chromosome stability: Insights from ICF syndrome studies. Genes Cells 26, 349–359. https://doi.org/10.1111/gtc.12850

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., Sasaki, H., 2020. CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. Sci. Rep. 10, 17865. https://doi.org/10.1038/s41598-020-74636-2

Unoki, M., Velasco, G., Kori, S., Arita, K., Daigaku, Y., Yeung, W.K.A., Fujimoto, A., Ohashi, H., Kubota, T., Miyake, K., Sasaki, H., 2023. Novel compound heterozygous mutations in *UHRF1* are associated with atypical immunodeficiency, centromeric instability and facial anomalies syndrome with distinctive genome-wide DNA hypomethylation. Hum. Mol. Genet. 32, 1439–1456. https://doi.org/10.1093/hmg/ddac291

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., Brinkley, B.R., 2001. Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J. Cell Sci. 114, 3529–3542. https://doi.org/10.1242/jcs.114.19.3529

Vanyushin, B.F., Mazin, A.L., Vasilyev, V.K., Belozersky, A.N., 1973. The content of 5-methylcytosine in animal DNA: The species and tissue specificity. Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth. 299, 397–403. https://doi.org/10.1016/0005-2787(73)90264-5

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., Francastel, C., 2018. Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Hum. Mol. Genet. 27, 2409–2424. https://doi.org/10.1093/hmg/ddy130

Velasco, G., Walton, E.L., Sterlin, D., Hédouin, S., Nitta, H., Ito, Y., Fouyssac, F., Mégarbané, A., Sasaki, H., Picard, C., Francastel, C., 2014. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. Orphanet J. Rare Dis. 9, 56. https://doi.org/10.1186/1750-1172-9-56

Vertino, P.M., Yen, R.-W.C., Gao, J., Baylin, S.B., 1996. De Novo Methylation of CpG Island Sequences in Human Fibroblasts Overexpressing DNA (Cytosine-5-)-Methyltransferase. Mol. Cell. Biol. 16, 4555–4565. https://doi.org/10.1128/MCB.16.8.4555

Viegas-Péquignot, E., Dutrillaux, B., 1976. Segmentation of human chromosomes induced by 5-ACR (5-azacytidine). Hum. Genet. 34, 247–254. https://doi.org/10.1007/BF00295287

Virchow, R., 1859. Cellular Pathology as based upon Physiological and Pathological Histology. Translated from the Second Edition of the Original, by Frank Chance (1861). Med. Crit. Psychol. J. 311–319.

Vissel, B., Choo, K.H., 1987. Human alpha satellite DNA - consensus sequence and conserved regions. Nucleic Acids Res. 15, 6751–6752. https://doi.org/10.1093/nar/15.16.6751

von Waldeyer, W., 1888. Ueber Karyokinese und ihre Beziehungen zu den Befruchtungsvorgängen. Arch. Für Mikrosk. Anat. 32, 1–122.

Voullaire, L.E., Slater, H.R., Petrovic, V., Choo, K.H.A., 1993. A Functional Marker Centromere with No Detectable Alpha-Satellite, Satellite Ill, or CENP-B Protein: Activation of a Latent Centromere? Am. J. Hum. Genet. 52.

Vukic, M., Daxinger, L., 2019. DNA methylation in disease: Immunodeficiency, Centromeric instability, Facial anomalies syndrome. Essays Biochem. 63, 773–783. https://doi.org/10.1042/EBC20190035

Waddington, C.H., 1957. The strategy of the genes : a discussion of some aspects of theoretical biology. Allen & Unwin.

Waddington, C.H., 1942. The Epigenotype. Endeavour 18–20. https://doi.org/10.1093/ije/dyr184. Rpt. in: International Journal of Epidemiology 41(1). 10-13. 2012.

Wang, J., Yang, J., Li, D., Li, J., 2021. Technologies for targeting DNA methylation modifications: Basic mechanism and potential application in cancer. Biochim. Biophys. Acta BBA - Rev. Cancer 1875, 188454. https://doi.org/10.1016/j.bbcan.2020.188454

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., Poirier, G.G., Earnshaw, W.C., 1997. Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr. Biol. 7, 901–904. https://doi-org.insb.bib.cnrs.fr/10.1016/S0960-9822(06)00382-4

Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737–738. https://doi.org/10.1038/171737a0

Waye, J.S., Willard, H.F., 1987. Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucleic Acids Res. 15, 7549–7569. https://doi.org/10.1093/nar/15.18.7549

Widschwendter, M., Jiang, G., Woods, C., Müller, H.M., Fiegl, H., Goebel, G., Marth, C., Müller-Holzner, E., Zeimet, A.G., Laird, P.W., Ehrlich, M., 2004. DNA Hypomethylation and Ovarian Cancer Biology. Cancer Res. 64, 4472–4480. https://doi.org/10.1158/0008-5472.CAN-04-0238

Wijmenga, C., Hansen, R.S., Gimelli, G., Bj□ rck, E.J., Davies, E.G., Valentine, D., Belohradsky, B.H., Van Dongen, J.J., Smeets, D.F.C.M., Van Den Heuvel, L.P.W.J., Luyten, J.A.F.M., Strengman, E., Weemaes, C., Pearson, P.L., 2000. Genetic variation in ICF syndrome: Evidence for genetic heterogeneity. Hum. Mutat. 16, 509–517. https://doi.org/10.1002/1098-1004(200012)16:6<509::AID-HUMU8>3.0.CO;2-V

Wilkins, M.H.F., Zubay, G., Wilson, H.R., 1959. X-ray diffraction studies of the molecular structure of nucleohistone and chromosomes. J. Mol. Biol. 1, 179-IN10. https://doi.org/10.1016/S0022-2836(59)80046-2

Willard, H.F., 1985. Chromosome-Specific Organization of Human Alpha Satellite DNA. Am. J. Hum. Genet. 37, 524–532.

Willard, H.F., Waye, J.S., 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet. 3, 192–198. https://doi.org/10.1016/0168-9525(87)90232-0

Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., Choo, K.H.A., 2007. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res. 17, 1146–1160. https://doi.org/10.1101/gr.6022807

Wong, N., Lam, W.-C., Lai, P.B.-S., Pang, E., Lau, W.-Y., Johnson, P.J., 2001. Hypomethylation of Chromosome 1 Heterochromatin DNA Correlates with q-Arm Copy Gain in Human Hepatocellular Carcinoma. Am. J. Pathol. 159, 465–471. https://doi.org/10.1016/S0002-9440(10)61718-X

Wong, N.C., Wong, L.H., Quach, J.M., Canham, P., Craig, J.M., Song, J.Z., Clark, S.J., Choo, K.H.A., 2006. Permissive Transcriptional Activity at the Centromere through Pockets of DNA Hypomethylation. PLoS Genet. 2, e17. https://doi.org/10.1371/journal.pgen.0020017

Woodcock, C.L.F., Safer, J.P., Stanchfield, J.E., 1976. Structural repeating units in chromatin. I. Evidence for their general occurrence. Exp. Cell Res. 97, 101–110. https://doi.org/10.1016/0014-4827(76)90659-5

Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., Warren, W.D., 1997. Asymmetric Methylation in the Hypermethylated CpG Promoter Region of the Human L1 Retrotransposon. J. Biol. Chem. 272, 7810–7816. https://doi.org/10.1074/jbc.272.12.7810

Wright, N.A., Poulsom, R., 2012. Omnis cellula e cellula revisited: cell biology as the foundation of pathology. J. Pathol. 226, 145–147. https://doi.org/10.1002/path.3030

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., Zhang, Y., 2011. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 25, 679–684. https://doi.org/10.1101/gad.2036011

Wu, H., Thijssen, P.E., De Klerk, E., Vonk, K.K.D., Wang, J., Den Hamer, B., Aytekin, C., Van Der Maarel, S.M., Daxinger, L., 2016. Converging disease genes in ICF syndrome: *ZBTB24* controls expression of *CDCA7* in mammals. Hum. Mol. Genet. 25, 4041–4051. https://doi.org/10.1093/hmg/ddw243

Wu, H., Wu, X., Shen, L., Zhang, Y., 2014. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. Nat. Biotechnol. 32, 1231–1240. https://doi.org/10.1038/nbt.3073

Wu, J.C., Santi, D.V., 1987. Kinetic and catalytic mechanism of HhaI methyltransferase. J. Biol. Chem. 262, 4778–4786. https://doi.org/10.1016/S0021-9258(18)61263-3

Wüllner, U., Kaut, O., deBoni, L., Piston, D., Schmitt, I., 2016. DNA methylation in Parkinson's disease. J. Neurochem. 139, 108–120. https://doi.org/10.1111/jnc.13646

Wyatt, GR., 1950. Occurrence of 5-Methyl-Cytosine in Nucleic Acids. Nature 166, 237–238. https://doi.org/10.1038/166237b0

Xiang, S., Liu, Z., Zhang, B., Zhou, J., Zhu, B.-D., Ji, J., Deng, D., 2010. Methylation status of individual CpG sites within Alu elements in the human genome and Alu hypomethylation in gastric carcinomas. BMC Cancer 10, 44. https://doi.org/10.1186/1471-2407-10-44

Xu, G.-L., Bestor, T.H., 1997. Cytosine methylation targetted to pre-determined sequences. Nat. Genet. 17, 376–378. https://doi.org/10.1038/ng1297-376

Xu, G.-L., Bestor, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., Viegas-Péquignot, E., 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187–191.

Yatskevich, S., Muir, K.W., Bellini, D., Zhang, Z., Yang, J., Tischer, T., Predin, M., Dendooven, T., McLaughlin, S.H., Barford, D., 2022. Structure of the human inner kinetochore bound to a centromeric CENP-A nucleosome.

Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Itou, H., Ajima, R., Nakano, R., Nakaoka, H., Fukui, K., Gamo, K., Tominari, Y., Takeuchi, H., Saga, Y., Hayashi, K., Kanemaki, M.T., 2020. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. Nat. Commun. 11, 5701. https://doi.org/10.1038/s41467-020-19532-z

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., Okazaki, T., 2000. Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. Proc. Natl. Acad. Sci. 97, 7266–7271. https://doi.org/10.1073/pnas.130189697

Yoda, K., Ando, S., Okuda, A., Kikuchi, A., Okazaki, T., 1998. *In vitro* assembly of the CENP-B/α-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. Genes Cells 3, 533–548. https://doi.org/10.1046/j.1365-2443.1998.00210.x

Yoda, K., Kitagawa, K., Masumoto, H., Muro, Y., Okazaki, T., 1992. A Human Centromere Protein, CENP-B, Has a DNA Binding Domain Containing Four Potential Helices at the NH2 Terminus, Which Is Separable from Dimerizing Activity. J. Cell Biol. 119, 1413–1427. https://doi.org/10.1083/jcb.119.6.1413

Yoda, K., Nakamura, T., Masumoto, H., Suzuki, N., Kitagawa, K., Nakano, M., Shinjo, A., Okazaki, T., 1996. Centromere Protein B of African Green Monkey Cells: Gene Structure, Cellular Expression, and Centromeric Localization. Mol. Cell. Biol. 16, 5169–5177. https://doi.org/10.1128/MCB.16.9.5169

Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335–340. https://doi.org/10.1016/S0168-9525(97)01181-5

Zacharias, E., 1881. Ueber die chemische Beschaffenheit des Zellkerns. Bot. Ztg. 39, 169–176.

Zeitlin, S.G., Barber, C.M., Allis, C.D., Sullivan, K.E., 2001. Differential regulation of CENP-A and histone H3 phosphorylation in G2/M. J. Cell Sci. 114, 653–661. https://doi.org/10.1242/jcs.114.4.653

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science 328, 916–919. https://doi.org/10.1126/science.1186366

Zhou, K., Gebala, M., Woods, D., Sundararajan, K., Edwards, G., Krzizike, D., Wereszczynski, J., Straight, A.F., Luger, K., 2022. CENP-N promotes the compaction of centromeric chromatin. Nat. Struct. Mol. Biol. 29, 403–413. https://doi.org/10.1038/s41594-022-00758-y

Zhu, H., Wang, G., Qian, J., 2016. Transcription factors as readers and effectors of DNA methylation. Nat. Rev. Genet. 17, 551–565. https://doi.org/10.1038/nrg.2016.83

Zhu, J., Cheng, K.C.L., Yuen, K.W.Y., 2018. Histone H3K9 and H4 Acetylations and Transcription Facilitate the Initial CENP-AHCP−3 Deposition and De Novo Centromere Establishment in Caenorhabditis elegans Artificial Chromosomes. Epigenetics Chromatin 11, 16. https://doi.org/10.1186/s13072-018-0185-1

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., Wang, X., Wei, Y., Liu, P., Yan, J., Ren, X., Yuan, P., Yuan, Y., Yan, Z., Wen, L., Yan, L., Qiao, J., Tang, F., 2018. Single-cell DNA methylome sequencing of human preimplantation embryos. Nat. Genet. 50, 12–19. https://doi.org/10.1038/s41588-017-0007-6

Zinkowski, R.P., Meyne, J., Brinkley, B.R., 1991. The centromere-kinetochore complex: a repeat subunit model. J. Cell Biol. 113, 1091–1110. https://doi.org/10.1083/jcb.113.5.1091

Zouali, M., 2021. DNA methylation signatures of autoimmune diseases in human B lymphocytes. Clin. Immunol. 222, 108622. https://doi.org/10.1016/j.clim.2020.108622

RÉSUMÉ

Lorsque les cellules se divisent, chaque cellule fille doit recevoir un jeu complet de chromosomes. Cela se produit grâce à que les chromosomes sont attachés au niveau des centromères pour leur séparation. Dans le cancer, ce processus peut échouer, générant des cellules avec un nombre anormal de chromosomes. L'ADN centromérique présente hauts niveaux de méthylation, une modification de l'ADN qui peut être réduite en cas de maladie. J'ai étudié les mécanismes de régulation de la méthylation de l'ADN centromérique et j'ai observé que cette modification est une frontière qui préserve la position et la fonction du centromère. J'ai généré des outils cellulaires pour réduire significative et spécifiquement les niveaux de méthylation de l'ADN des centromères. Lorsque la méthylation diminue, la ségrégation des chromosomes est altérée. Ce travail établit les bases permettant de comprendre comment la perte de méthylation de l'ADN centromérique peut être à l'origine de l'instabilité du génome.

MOTS CLÉS

centromère – méthylation de l'ADN – épigénétique – CENP-A – CENP-B – stabilité du génome

ABSTRACT

When cells divide each daughter cell should get a full set of chromosomes. This occurs because all chromosomes are attached at their centromeres to the mitotic spindle, which pulls them apart during division. In some diseases, as in cancer, this process can fail resulting in daughter cells with an abnormal chromosome number. The centromeric DNA has high levels of methylation, a DNA modification that can be reduced in disease. During my PhD I have studied the mechanisms regulating the centromeric DNA methylation and observed that this modification acts as a boundary, preserving the position and function of the centromere. I have generated cellular tools to significantly decrease the DNA methylation levels specifically at the centromeres. When the methylation is lowered, chromosome segregation is impaired due to the alteration of core centromeric proteins. This work is the foundation for understanding how the loss of centromeric DNA methylation can be at the root of genome instability.

KEYWORDS

centromere – DNA methylation – epigenetics – CENP-A – CENP-B – genome stability