Dissertation zur Erlangung des Doktorgrades der
Fakultät für Chemie und Pharmazie der
Ludwig-Maximilians-Universität München

Characterization of Novel Human Centromeric Interactors and their Role in Human Centromere Transcription

Chandni Natalia Kumar

aus

München, Deutschland

2022

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von

Herrn Dr. Franz Herzog

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 13.12.2022

_____

Chandni Natalia Kumar

Dissertation eingereicht am   13.12.2022

  1. Gutachter: Dr. Franz Herzog

  2. Gutachter: Prof. Dr. Axel Imhof

Mündliche Prüfung am   28.02.2023

Table of contents

1 Preface

The projects described in this thesis were performed in the laboratory of Dr. Franz Herzog from January 2019 to December 2022 at the Gene Center Munich, LMU. In the second half of my PhD (June 2021 to December 2022) I was additionally supervised by Prof. Dr. Axel Imhof (Faculty of Medicine, LMU).

I have worked on different topics throughout my PhD. In this thesis I am going to present two topics and start with a general introduction followed by the results section and the discussion. My overall aim in this thesis was to characterize the role of novel centromere interactors and how these interactors might have an impact on centromere identity. The characterization of novel centromere interactors is based on a previous unpublished work of our group by Dr. Götz Norman Hagemann (doctoral thesis, LMU, 2020). Key findings of the thesis were presented at the Chromatin Day on 21st of July 2022 under the title "*ZBTB9 – a new centromeric component*".

In addition, I was interested in investigating the influence of O-linked glycans attached to nuclear proteins and how this modification alters the chromatin environment including centromere chromatin. Here I collaborated with Prof. Anja Hoffman-Röder (Chemistry and Pharmacy Department, LMU) along with the PhD student Stefan Marchner to synthesize a O-glycan enrichment matrix. Moreover, in a second collaborative project with the group I performed cytotoxicity assays and *in vitro* enzymatic assays to characterize novel synthetic glycolabels. The latter project is not part of this thesis.

Notably, I worked on a third project in Dr. Herzogs group to investigate the general applicability of our novel quantitative cross-linking coupled to mass spectrometry method to measure apparent binding affinity constants of complexes. The method was originally developed by the former colleagues Dr. Victor Solis-Mezarino and Dr. Götz Norman Hagemann. In a current reviewing process, I performed experiments to demonstrate if the method can be applied to different biological systems to underline its general application, which earned me a co-first authorship. The results are part of the revised manuscript entitled "*Quantitative Crosslinking and Mass Spectrometry Detects Phospho-Induced Kinetochore Stabilization*", which is under review and got pre-published in bioRxiv (01.04.2022).

Publications (under revision or manuscript in preparation)

1) Hagemann G*., Solis-Mezarino V.*, **Kumar C.N.***, Singh S., Potocnjak M., Herzog F. (2020) Quantitative Crosslinking and Mass Spectrometry Detect Phospho-Induced Kinetochore Stabilization. (under Revision, first version published on bioRxiv 01.04.2022)

2) **Chandni Natalia Kumar***, Götz Normen Hagemann*, Victor Solis-Mezarino*, Sylvia Singh, Jeannette Koch, Hartmann Harz, Thomas Fröhlich, Gregor Witte, Nikolina Sekulic, Heinrich Leonhardt, Axel Imhof and Franz Herzog. (2022) A novel insight into human centromeric transcription by mass spectrometry (not finished title, manuscript in preparation).

2 Summary

The eukaryotic cell cycle is a fundamental process in life, which ensures the genomic integrity by passing on the replicated DNA to the daughter cells. One major regulator of the cell cycle is the accurate and timely-regulated protein phosphorylation performed by various kinases:cyclin complexes at individual cell cycle stages. In mitosis, the faithful segregation of sister chromatids depends on the correct order of events from chromosome compaction, alignment and sister chromatid separation which are tightly controlled through regulatory feedback mechanisms. The centromere is a specialized chromatin region that promotes the assembly of the kinetochore that links the microtubules of the mitotic spindle to chromosomes and has a central role in chromosome alignment and segregation.

The centromere is harbouring a histone H3 variant called CENP-A and is the primary mark for kinetochore recruitment. Moreover, the centromere consists of higher order repeats of 171 bp long alpha-satellite sequences in a highly repetitive manner, which can lead up to several mega bases in length. The link between the centromere and the microtubules is established by the kinetochore. The kinetochore is a macromolecular complex and composed of more than 80 proteins, which assembles in stable subcomplexes and follows a conserved hierarchy of assembly: the complexes associated with the inner centromere organization are called the Constitutive Centromere Associated Network (CCAN) and builds the foundation for the outer kinetochore microtubule building block, the KMN network (KNL/Mis12/NDC80) followed by the fibrous corona. Among eukaryotes, several kinetochore proteins are conserved. In addition, a multitude of stable complexes as well as regulatory post-translational modifications are involved in the process of dividing sister chromatids.

Researchers are interested in understanding the underlying mechanisms of faithful chromosome segregation, since aberrant or incorrect attachments of microtubules to centromeres lead to segregation defects and aberrant chromosome numbers, known as aneuploidy. These severe genomic alterations are associated with tumorigenesis and inherited genetic diseases. One field of chromosome research focuses on understanding the specific centromeric marks and their interplay with other complexes to ensure the timely and correct assembly of stable kinetochore complexes. Such marks include known modifications such as phosphorylation or ubiquitination on histones or other chromatin associated proteins.

Yet another important modification gained importance in recent years: O-Glycosylation. Glycosylation is the covalent attachment of a saccharide moiety to the amino acid side chain of serine or threonine by glycosyltransferases and is one of the most abundant post-translational modifications. In recent years this specific modification gained attention, since the modification alters fundamental cellular signalling processes such as the regulation of transcription, cell proliferation and differentiation. Glycosylation has been reported to modulate the function of key cell cycle regulators. However, a comprehensive molecular understanding of its role in cell cycle regulation or its contribution to centromere identity remains elusive. Notably, recent research has shown that glycosylation and phosphorylation are interlinked in a cellular "cross-talk", since the same serine or threonine can be glycosylated or phosphorylated. Therefore, it is believed that the interchangeable modifications lead to alterations in signalling cascades depending on the cellular situation.

Besides modifications found on histones or other chromatin associated proteins, latest research in this field observed that CENP-A incorporation and replenishment are additionally dependent on the transcription of centromeric sequences. The transcriptional process is important since the resulting transcripts are required for stability of the centromere. Deletions or alteration in transcriptional levels of this highly regulated and discrete process leads to segregation defects. In the last decade studies have uncovered important mechanisms on how the centromere turnover is orchestrated by several complexes and how transcription interplays with these complexes. Nonetheless, it remains elusive how the centromere changes over the course of the cell cycle and which complexes are associated specifically at certain stages. A previous study conducted in the Herzog group applied a ChIP-MS approach and highlighted the CENP-A proteome changes over the course of the cell cycle. Here novel interactors were identified which have not been linked to the centromere before. Besides the enrichment of known centromere associated proteins, other complexes involved in splicing or transcriptional regulation were identified. In addition, some of the identified interactors have been suggested to be O-glycosylated. It was of interest whether the identified proteins are indeed novel centromere interactors and if so, what is their potential role mediating identity and maintenance of the centromere.

The work presented in this thesis aimed at characterizing these novel centromere-associated proteins and establishing the path of direct interactions that mediate their recruitment to the centromere by binding to the close interactor of the CENP-A Nucleosomes, CENP-B. CENP-B binds to the CENP-B box at the entry side of the alpha-

satellite DNA sequences wrapped around CENP-A containing histone octamers and may directly links the novel interactors to the centromere. In addition, potential binding partners of these proteins could be determined by mass spectrometry and resulted in the identification of complexes involved in transcription, RNA-binding and mRNA splicing events. One of the novel centromere-associated proteins was further implicated in heterochromatin formation, chromatin condensation and nuclear import.

The combination of biochemical, biophysical, and mass spectrometric experimental approaches paved the way to propose a model, where the novel transcription factor might be involved in centromere maintenance and stability by interacting with proteins associated in heterochromatin formation, transcriptional processes and splicing or chromosome compaction at mitosis onset. These processes might ensure centromere identity over the course of the cell cycle and have to be further studied to confirm their importance in centromere integrity. Furthermore, another novel centromere associated chromatin remodeler complex have been implicated in CENP-A replenishment process at G1-Phase, since mass spectrometry analysis revealed the novel interaction with other classes of chromatin remodeler families and transcription to facilitate CENP-A incorporation by possible eviction of H3 placeholder by the possible formation of a novel chromatin remodeling complex. The third chromatin remodeler complex might play a role in late S-Phase centromere replication process and facilitates CENP-A incorporated at the newly replicated centromere of the duplicated DNA strand.

Notably, the preliminary data obtained by the O-glycan enrichment experiments show the tendency that kinetochore and centromere associated proteins can be O-glycosylated as well. Moreover, the initial analysis of this data further provided information on how the O-glycosylation modification might be associated with processes involved in ribosomal biogenesis, mRNA metabolism, gene silencing or cell cycle regulation. The O-glycosylation pulldown shows the identification of proteins and complexes which have been identified in the ChIP-MS CENP-A Pulldowns or in the Nuclear Fishing pulldowns of the novel centromere associated transcription factor. These identified proteins were associated with chromatin remodelling, DNA methylation or splicing factors. How this modification alters specifically the centromeric chromatin remains to be answered in future studies.

In summary, based on protein interaction data, *in vitro* reconstitution of wild-type and mutant complexes and mass spectrometric analysis, the findings in this thesis suggests a model on how novel centromere interactors shape the centromeric environment by establishing

important structural features such as heterochromatin marks, regulating transcriptional processes and being involved in splicing events of centromeric transcripts. Testing the physiological relevance of the proposed pathway in the given model will contribute to the understanding of mechanism that establish and maintain centromere identity and build up stable kinetochore structures for faithful chromosome segregation.

3 Main objective of the thesis

In the last decade many efforts have been made to study the centromere proteome and its complexes to understand the mechanism driving the accurate segregation of chromosomes to the progeny. The stability and architecture of the centromere is an integral part ensuring the correct inheritance of the replicated genome to the daughter cells. Although recent studies shed light on specific complexes known to interact with the centromeric chromatin to maintain and regulate centromere function, a global quantitative and time resolved survey of the CENP-A proteome as well as modifications guiding this process remains elusive.

The main objective of my PhD work was to investigate and characterize a subset of novel centromere interactors identified in a previous protein interaction study, which systematically analysed the CENP-A proteome by mass spectrometry from different cell cycle stages. In particular, the initial goal was to identify protein complexes that are involved in distributing CENP-A between the leading and lagging strand after DNA replication.

By using mass spectrometry, the novel interactors and their binding partners can be identified leading to a better understanding of the centromeric chromatin formation/ maintenance. A second layer of information on how centromere maintenance occurs is provided by studying post translational modifications, since these alter the chromatin state. An additional task of my PhD project was the development of an O-glycosylation enrichment matrix coupled to mass spectrometric analysis to identify possible post-translational modified centromere associated proteins, which might have been modified by O-glycosylation.

The main objectives of this PhD thesis aimed at addressing the following questions:

- What is the direct interaction partner of ZBTB9 responsible for its recruitment to the centromere?
- What is the possible role of ZBTB9 at centromeres?
- How could ZBTB9 contribute to the stabilization of centromere formation?
- Are the ISWI chromatin remodeler WICH-5 or the CHRAC1-5 complex specifically recruited to the centromere?
- What is the direct interaction partner of the WICH-5 complex at centromeres?
- What could be the possible role of WICH-5 and CHRAC1-5 at the centromere?

- What are the possible roles of WICH-5 and ZBTB9 in centromere function? How could mechanisms mediated by WICH-5 and ZBTB9 be integrated in a model for maintaining centromere identity?
- Does the enrichment-matrix specifically enrich O-GlcNAc/ O-GalNAc specific peptides? If so, which nuclear proteins could be identified?

4 General Introduction

4.1 The mammalian cell cycle

An organism can only live and fully develop by passing on its genetic inheritance to the next generation over an extraordinary amount of time. One key element in eukaryotes helping to pass on the genetic material is the cell cycle, a highly regulated process enabling cell growth, duplication of genetic material and subsequently cell division resulting in an identical daughter cell[1]. The progression of one full circled cell cycle to the next is driven by the core cell-cycle machinery, having its 'headquarter' in the nucleus. This machinery is composed of cyclins and their catalytic interacting partner cyclin-dependent kinases (CDKs)[1,2]. The activity of these cyclin/CDK complexes is cell cycle specific and their major task is to phosphorylate other target proteins contributing to the progression of the next cell cycle stage[1]. Cells with a normal cell cycle regulation tightly regulate the activity of cell-cycle specific proteins by controlling the transcription processes and protein degradation of these complexes as well as by activating CDK inhibitor proteins[3,4]. A closer look in cancer cells reveal an aberrant regulation of this machinery, where cell-cycle proteins are constantly activated leading to genetic lesions and the development of most tumor types[1,4]. The human genome contains at least 30 different genes encoding for a plethora of different cyclins whose product contain an amino acid domain called the cyclin box[4,5]. Further studies characterized some of the subfamilies and their defining features where cyclins oscillate in their abundance level during the cell cycle and their activities are tight to the steps in cell cycle progression. These few characterized subfamilies are A-, B-, C-, D- and E-cyclins, whereas their alphabetical name does not imply their order of appearance in the cell cycle stage, rather represent their time of discovery[4].

The cell cycle is subdivided into four stages, the Gap 1 (G1-), DNA synthesis (S-), Gap 2 (G2-) and mitosis (M-) phase. An overview is provided in Figure 1.
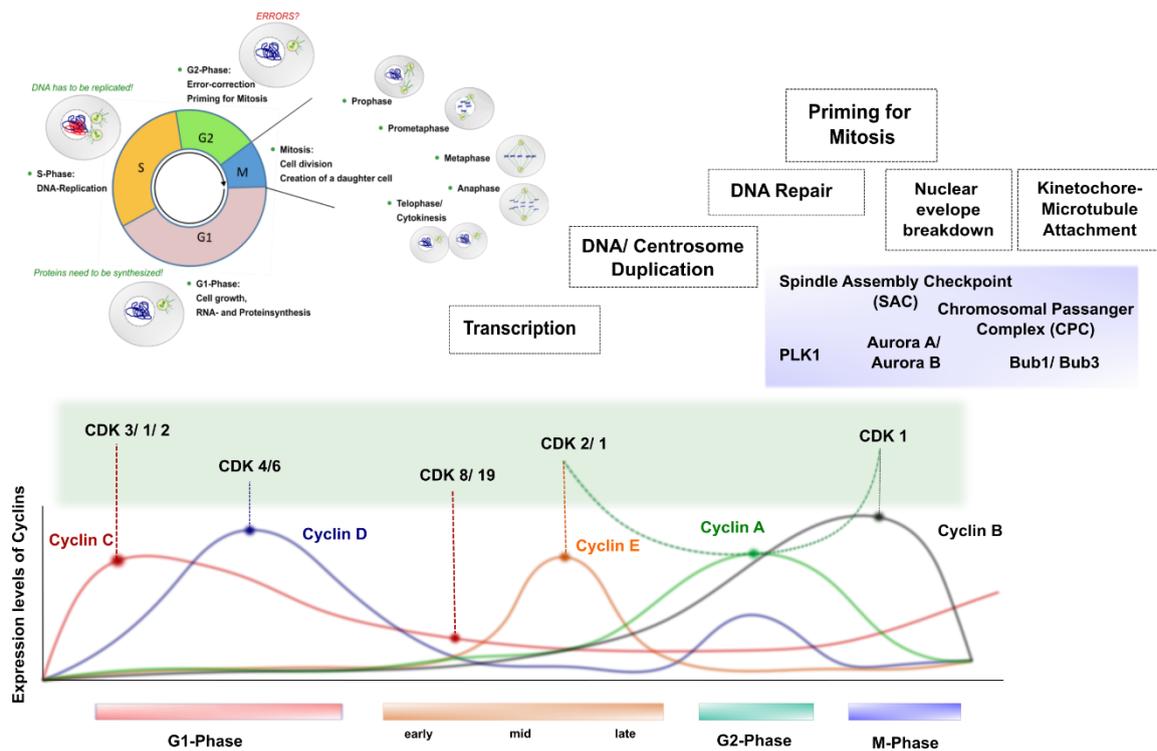
Figure 1: **General schematic representation of the cell cycle and its CDK-Cyclin partners linked to major cellular processes.** The overview was recreated from Diego Martinez-Alonso, Marcos Malumbres, (2020) and current knowledge in this field. The expression levels of each Cyclin Family members are shown with their respective CDK complex partner. The start is initiated by Cyclin C and Cyclin D upon binding to CDK 1-3 and CDK 4 and 6. The cell synthesises the necessary proteins to enter the next stage (S-Phase). At S-Phase, levels of Cyclin C/D decreases through degradation or inactivation mechanisms and Cyclin E has its peak and binding to CDK1 or 2 and promoting Transcription and DNA/ Centromere duplication. In late S-Phase the levels of Cyclin A rise and reach their optimum in G2-Phase, where the cell checks for any DNA errors. If there are duplication errors, DNA repair pathways are activated, and the cell is halt in this specific phase until all checkpoints have been passed. Lastly, in mitosis CDK1 binds to elevated levels of Cyclin B and with other mitotic checkpoint complexes (SAC, CPC, PLK1) the cell breaks down the nuclear envelope and facilitates correct bioriented (amphitelic) attachment of microtubules to the centromere region through the correctly built kinetochore complex. In late mitosis (cytokinesis) the process of Cyclin B degradation is executed, and levels of Cyclin C rise again, initiating another round of cell cycle by starting at G1-Phase. In the blue highlighted box, some other kinases/ checkpoint complexes are present, important for the mitotic progression and having an influence on the strength/ connectivity between the Centromere-Kinetochore-Microtubule axis.

In G1 the cell increases in size, translates proteins, and actively transcribes genes for cell cycle control. In addition, a series of DNA checkpoints are executed before entering S-

phase[6]. The progression into G1-Phase is mainly guided by cyclin C- and D-type, where C-type cyclins are still in debate of their importance in G1-Phase progression. In mouse models it has been shown that depletion of C-type cyclins resulted in embryonic lethality or severe developmental retardation[4,7]. The role of these specific type of cyclin has to be further investigated, since alteration of cyclin C were associated with the development of acute lymphoblastic leukaemia, prostate cancer and osteosarcoma[4,8]. Upon receiving mitogenic signals during G1-phase, D-type cyclins (D1-3) are upregulated and bind to CDK4 and CDK6. Furthermore, this complex is accompanied by KIP/CIP proteins (p21, p27, p57), which serve as inhibitor proteins for other CDK complexes (CDK2, CDK1)[3]. The task of CDK-cyclin complexes is the phosphorylation of their target proteins, in this case CDK4/6-cyclin D complex targets various cellular proteins, one among of them RB1 (retinoblastoma protein), p107 and p130. RB1 represses E2F transcription factor in its dephosphorylated form. Upon phosphorylation by CDK4/6-cyclin D complex, RB1 gets partially inactivated leading to the transcription of E2F and leading to a cascade of other transcriptional expressions e.g. E-type cyclins (E1, E2), which later on bind to CDK2. The CDK2-cyclin E complex completes the phosphorylation status of RB1 and leads to the transcriptional cascade of E2F, resulting in the synthesis of proteins required for S-Phase initiation[9].

The major task of the S-phase is the replication of the entire genome[6]. In the late G1-Phase, transcriptional target proteins by EF2 assemble on the origins of DNA replication and form pre-replication complexes. These complexes compose of ORC1-6, CDC6, CDT1 and the MCM2-7 DNA Helicase. As soon as the S-Phase is reached, DNA replication is triggered by a series of sequential phosphorylation events performed by DBF4-CDC7 and CDK2-cyclin E complex[1,10]. These phosphorylation events mediating the binding of DCD45 and GINS to the MCM2-7 DNA Helicase complex resulting in the formation of the CMG complex (CDC45-MCM2-7-GINS) and helps to activate the DNA Helicase[10]. Shortly after CDK2 switches its cyclin E to cyclin A and fires the DNA replication origins. Cyclin A binds to CDK1 and CDK2 and activates them to ensure proper S-Phase execution and has a role in G2-M transition too[1]. The complex of CDK-Cyclin E promotes the phosphorylation of DNA replication factors, factors involved in centrosome duplication, DNA repair proteins and regulators of histone synthesis [9,11].

In G2, the cell prepares for the final stage of cell division and examines the size and DNA duplication errors. In M-Phase, consisting of Pro-, Meta-, Ana- and Telophase, the chromatin is condensed and compacted into chromosomes before the nuclear envelope

breaks down[6]. The function of cyclin A in G2/M transition is the activation of CDK1-Cyclin B complex, where Cyclin B accumulates in the nucleus at mitosis onset and gets prior phosphorylated by the kinase PLK1. The complex then phosphorylates proteins intertwined with the processes in mitotic progression, cytoskeleton reorganization, nuclear envelope breakdown, chromosome condensation/ segregation, mitotic spindle assembly/ function and cytokinesis[4]. One of the major drivers of the cell-cycle progression are the so called CDK-activating kinases (CAK), which consists of Cyclin H, CDK7 and RING containing protein MAT1. This CAK complex is capable of phosphorylating other CDKs therefore stabilizing CDK-Cyclin complexes (e.g., CDK1-CyclinB). Moreover, it has been shown that CAK can me a member of the TFIIH transcription factor family and recruit RNA Polymerase II to gene promotors and activates it by phosphorylating the C-terminal domain by the subunit CDK7 and facilitate transcription[1,12].

Cytokinesis generates two diploid daughter cells. After complete division, cells enter G1 again for another cell division cycle or arrest in the quiescent state G0[6].

Besides their timely regulated appearance during the different cell cycle stages, the proteasomal degradation of key cell-cycle regulators is an important aspect in the regulation of the cell cycle. This process is supported by multi-protein E3 ubiquitin ligase complexes (e.g., APC/C, BRC, Skp-Cullin-F box-containing complex)[13,14]. Target proteins containing specific degron motifs are recognized by E3 ligase complexes, for instance CDC20, CHD1, cyclin F or SKP2 leading to ubiquitination and subsequently proteasomal degradation. The anaphase-promoting complex or cyclosome (APC/C) with CDH1 stabilizes G1-Phase where it maintains mitotic and S-Phase specific Cyclins at low levels to prevent premature S-Phase initiation[2]. The complex gets inactivated upon entering S-Phase by CDK2-Cyclin E complex and is further maintained by EMI1. Cyclin E is degraded by BCR (BTB-Cul3-Rbx1 complex) or can be degraded by FBXW17, mediating its ubiquitination during S-Phase[1,15,16]. Cyclin F inactivates E2F upon G2-Phase to stop/ prevent the synthesis of replicative histones therefore contributing to the regulation of centrosomal duplication and mediating the degradation of CDH1[13]. After bypassing G2, the cell enters the mitotic phase, where Cyclin A is degraded by APC/C CDC20 complex so it cannot reach the metaphase but also helps to break down the nuclear envelope. After this event, Cyclin B and securin are degraded with the aforementioned complex to promote anaphase onset [4]. When the cell is ready for the next round, the APC/C CDH1 complex appears again, starting in anaphase, modulating

the stability of mitotic key players (Aurora A, PLK1) to subsequently continue its task in the next G1-Phase[13].

4.2 Accompanying the cell cycle: The rising importance of glycosylation as post translational modification in cellular processes

Carbohydrate moieties (glycans) play a key role in generating energy for the metabolism of the cell and serve as a nutritional storage in a structural and biophysical context. Glycans provide a platform for symbionts but also for pathogens to interact in a highly specific manner with the host cell as seen in the antigen and antibody recognition pattern[17]. In the 1980s the discovery of glycosylated nuclear and cytoplasmic proteins, emphasised its importance as a specific post-translational modification (PTM)[18,19]. It has been shown, that altered glycosylation states in mammals causes the development of cancer[17].

Uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) serves as the universal substrate for the attachment of different N- and O-glycosylation modifications and is derived from the hexosamine biosynthetic pathway (HBP). Moreover, UDP-GlcNAc is the most abundant nucleotide-based structure after ATP and serves as a nutritional sensor. The HBP is involved in known metabolic pathways such as glucose, fatty acid or nucleotide metabolism[20].
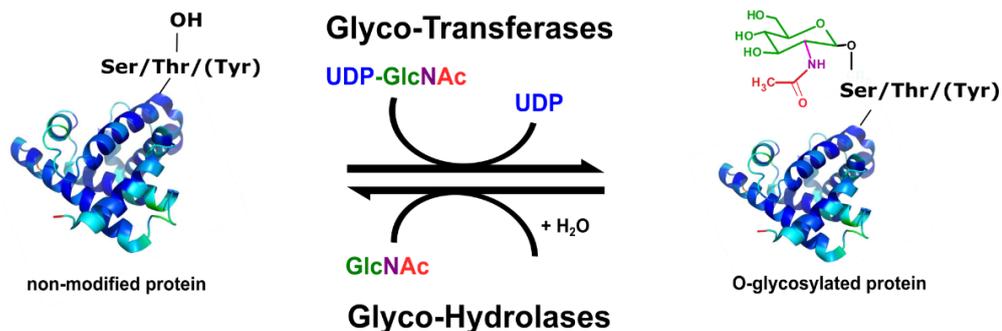
Figure 2: **The transfer of UDP-GlcNAc onto a serine/threonine/tyrosine side chain**. The glyotransferase transfers the GlcNAc onto a side chain of serine, threonine or tyrosine by using UDP-GlcNAc as substrate. The glycosylation mark is removed by glyco-hydrolases.

The role of glycans in the cell ranges from structural and modulatory functions to intrinsic and extrinsic recognition, including correct protein folding, protease protection and tissue

elasticity. Other pathways include intracellular signalling, fertilization, and reproduction[17,21]. There are two different types of glycosylation: N- and O-glycosylation. N-glycosylation has been studied extensively, whereas O-glycosylation still poses a challenge to analytical strategies and has not been comprehensively analysed so far[22]. In humans, O-glycosylation is built by six different monosaccharide attachments, which can be further extended to even more complex structures: O-Glucose, O-Fucose, O-Mannose, O-Galactose, O-N-acetylglucosamine (O-GlcNAc) and O-N-acetyl-galactosamine (O-GalNAc). The modification occurs on serine or threonine of nuclear, cytoplasmic, or mitochondrial proteins by an enzymatically catalysed glycosidic bond formation between the hydroxyl group of the sugar moiety and the hydroxyl group of the respective amino acid side chain (Figure 2). For O-glycosylation several glycosyltransferases and -hydrolases are involved in the catalysis of the reaction. In case of O-GlcNAcylation, this modification is regulated by only two enzymes: O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA)[21]. For mucin-O-type-glycosylation (O-GalNAc), 20 polypeptide N-acetylgalacto-saminyltransferases (ppGALNT) are known with their tissue- and substrate-specific activities. Up to date the consensus sequence for most O-glycosylation sites (except O-Fucosylation and O-Glucosylation) are unknown, impeding their site identification[6,23].

Glycosylation is implicated in the modification of transcription factors, RNA Polymerase II, histones, and histone deacetylase complexes, thus involved in the regulation of gene expression (e.g., C-terminal domain of RNAP II, TET, HDAC, MLL5, HIRA, HCF1 or Myc)[20,24]. Furthermore, previous studies indicated, that O-glycosylation, in particular O-GlcNAcylation, plays an important function in cell cycle regulation, in addition to the prominent role of protein phosphorylation[25].
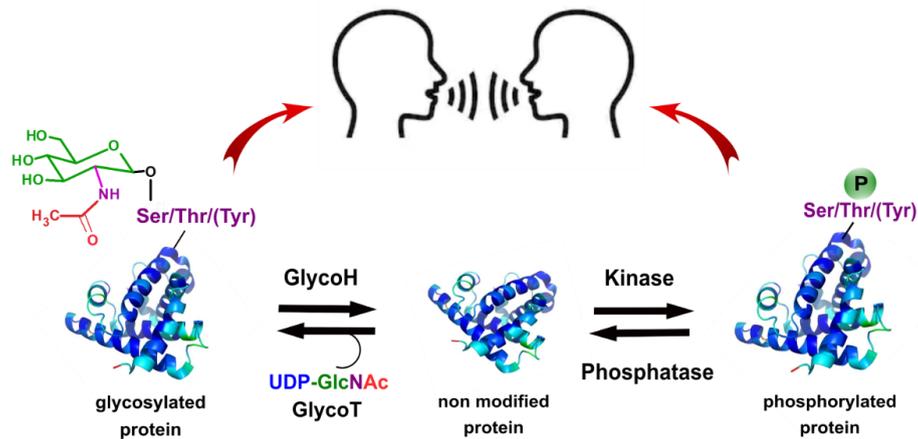
Figure 3: **The cross talk of glycosylation and phosphorylation alters the cellular environment.** The same serine or threonine side chain can be subjected to O-glycosylation modification or phosphorylation depending on the cellular status of the cell.

Both modifications are mutually exclusive, in targeting the same or neighbouring amino acid residues (Figure 3). The competitive occupancy through both modifications is derived from their distinct biophysical features. O-GlcNAcylation represents a bulky modification with a stroke radius four to five times larger than phosphorylation, whereas phosphorylation introduces negative charges. Either modification causes a conformational change of the target protein and therefore alters its structure and activity[6]. Moreover, it is known that O-GlcNAc regulates various protein kinases and phosphatases and hence, links the glycosylation pathway with the phosphorylation machinery[6]. It has been shown that the levels of O-glycosylated nuclear and cytoplasmic proteins rise, when the cell progresses into G1-phase but its role in S-Phase in unclear[26,27]. OGA inhibition in HeLa and 3T-LI cells delays the progression into S-Phase. O-GlcNAc modification of histone H3 is elevated in G2-Phase but decreases upon entry into M-Phase[27]. Moreover, a study described the increase of O-GlcNAcylation levels in mitosis, which decreases upon onset of cytokinesis. Overall, the reports on global changes of O-GlcNAcylation levels during the cell cycle events are controversial, since other studies found elevated O-GlcNAcylation during G2 or G2/M-Phase transition[28]. O-GlcNAcylation has been described to impose an important function in regulating mitotic progression as gain or loss of function mutants of OGT and OGA severely alters M-Phase progression leading to a delayed mitotic exit. Moreover, a recent report has indicated a direct interplay between OGT, OGA, aurora B and protein phosphatase 1[27,28].

Further evidence for regulation of gene expression through OGT and O-GlcNAcylation came through the investigation of TET (Ten-eleven translocation) proteins. The TET family

comprises of 5-methylcytosine hydroxylase members, which reverse the gene silencing mark mediated by DNMT enzymes (DNA-methyltransferases). The dialog between TET and OGT proteins lead to TET-promoted recruitment of OGT to CpG-rich transcription start sites, which corresponds to transcriptionally active genes[20,29,30]. Another interesting aspect of O-GlcNAcylation involves its assistance in protein translation. In several studies it has been shown that nearly half of ribosomal proteins are O-GlcNAcylated and both enzymes catalysing the reaction (OGT and OGA) are strongly associated with ribosome. Notably, overexpression of OGT or OGA (HepG2 cell line) affects the homeostasis of ribosomal subunits, thus maybe O-GlcNAc modification is important for the maturation and assembly of ribosomal subunits[20,31]. In conclusion, the modification by O-GlcNAc and their associated pathways have an impact on metabolic and cellular processes and provide an insight on the changes at the chromatin level.

4.3 Sharing neighbourhood in the Nucleus: The different types of Chromatin

Our genetic material is not a loose string; hence it is a compact fibre that occupies the nucleus. In fact, the DNA is in total 205 cm long[32], whereas the nucleus is only about 5 to 10 µm in diameter[33]. The question arises how this DNA fibre in its entire length can occupy such a small sphere? The answer is that our genome is compacted and consisting of different chromatin regions. The following paragraph provides an overview, how the human genome has three characteristic chromatin types and what makes them distinguishable from each other.

Around the 1880s Walter Flemming was one of the first scientists to describe the observed stainable threads in the nucleus during mitosis, which later lead to the term chromatin. chromatin adopts many functions such as packaging genetic material, regulating gene expression and consists of a complex formed of DNA and histone proteins[34,35]. After discovering DNA as "transforming principle" and therefore genes correlated with chromosomal bands, the interest in chromatin structure rose. From there, over one century later Kornberg proposed in 1974 the repeating unit of chromatin, which was made of two copies of each four canonical histone proteins known as H2A, H2B, H3 and H4 and about 200 bp long DNA[35,36]. Shortly after Olins and Olins visualized chromatin structure as a 'beads on a string' which is termed now as nucleosomes[37]. The first actual glimpse of a nucleosome structure was obtained in 1997, where X-ray crystal structure (2.8 Å) revealed

a 146 bp long DNA wrapped around a histone octamer in a 1.65 turn in a left-handed super helical fashion[38]. Depending on the higher-order packaging of nucleosomes, post-translational modification on histones and histone variants, chromatin can be divided into versatile categories[39,40]. The histone tails are subjected to modifications such as methyl or acetyl groups transferred by specific enzymes[35]. Other modifications include phosphorylation, ubiquitination, glycosylation, SUMOlyation as well as rare modifications [41,42]. For a comprehensive overview of these modifications, the review of Huiwen Xu *et.al*, 2021 is recommended [42] .

The silencing of genes is often associated with DNA methylation marks, whereas histone acetylation is often interlinked to actively transcribing DNA. The latter one reduces the affinity of histone octamer to the corresponding wrapped DNA, facilitating transcription processes[40,43,44]. At specific genomic locations histone variants are incorporated to replace canonical histones in nucleosomes and thus modifying chromatin structure and functions. For instance, H3 is commonly replaced by the variant H3.3 in active gene regions, whereas H3 is replaced by CENP-A (Centromeric Protein A) at the centromere region[45,46]. The major three categories of chromatin are heterochromatin, euchromatin and centromeric chromatin. Heterochromatin takes up the majority of chromatin and is characterized as condensed chromatin structure which is less transcriptionally active and present at pericentric regions, in general transcriptionally repressed regions and telomeres. In contrast, Euchromatin states a more open environment and are present in gene-rich genomic loci and marked by high transcription levels. The centromeric chromatin is marked by the histone variant CENP-A and has a more compact formation but not as seen in Heterochromatin[35,47,48]. Each chromatin type is associated and regulated by its specific chromatin modifying enzymes and architectural complexes, altering the chromatin compaction.

The chromatin can be further subdivided as shown by a publication in 2010. The investigators wanted to address if besides the known HP1 and PcG chromatin type finer subclassification can be found within the *Drosophila* genome[49]. Therefore, they generated genome-wide location maps of 53 broadly selected chromatin proteins and in addition four key histone marks (H3K9me2, K27me3, K4me2, K79me3) in embryonic *Drosophila melanogaster* cell line Kc167. In total they could define three more (in total five finer subclassifications of chromatin types) chromatin types with unique combinations of proteins by applying the DamID technology. The different regions that could be captured were

divided into five major different colour schemes: Blue and Green chromatin corresponds to the known heterochromatin types already depicted for *Drosophila*. The Black chromatin type exhibits repressive chromatin traits, whereas Red and Yellow subdivide the Euchromatin with a distinctive difference in histone mark H3K36me3 which has been specifically linked to the Yellow euchromatin type. Why this mark is absent in the Red euchromatin type has to be further investigated. The authors emphasize that his five subtypes are only a broader classification and one can even subdivide in even more finer classifications[49].

A simplified overview of the complex chromatin organization with its subtypes is depicted in Figure 4.
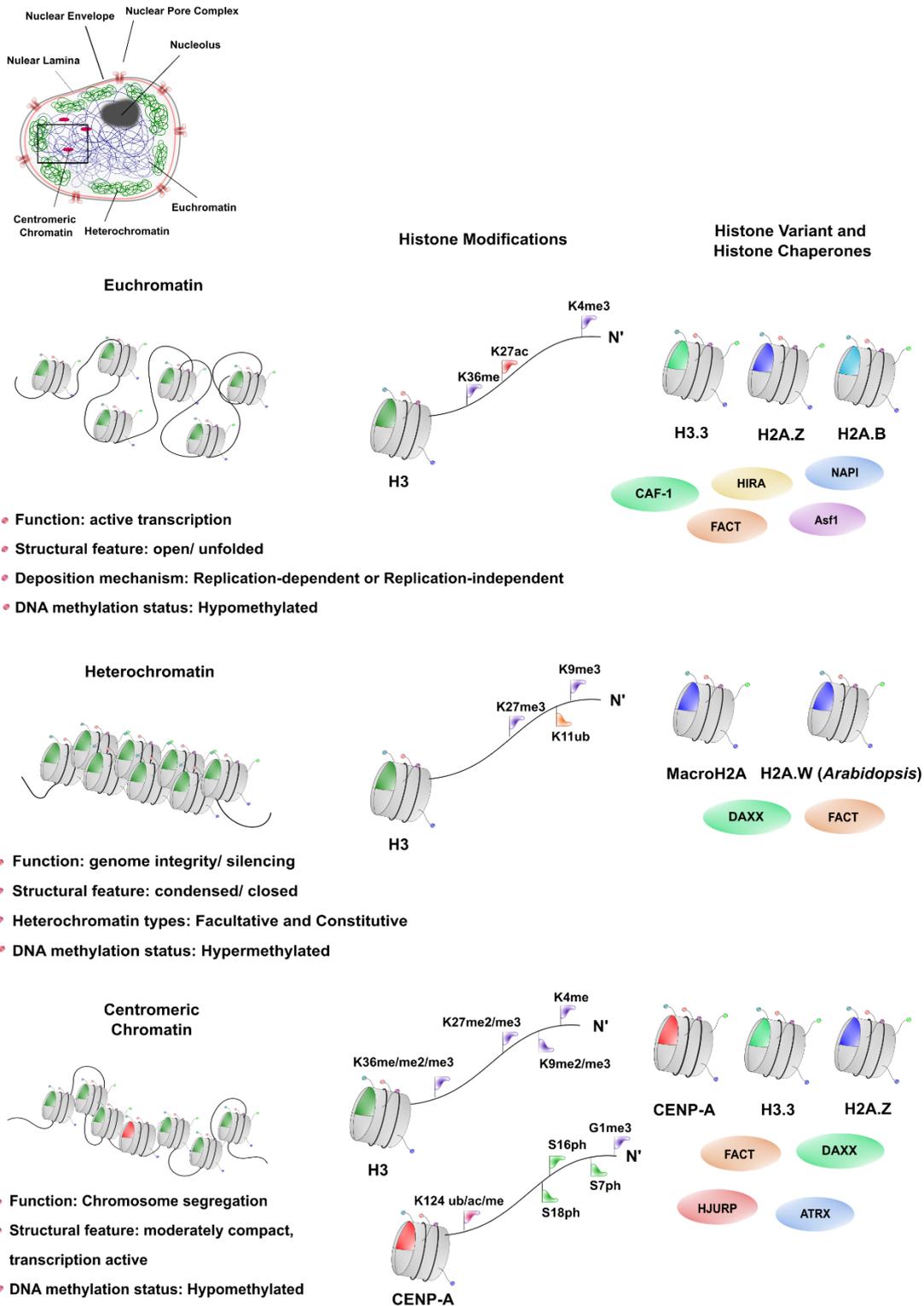
Figure 4: **Overview of the different chromatin found within the Nucleus**. Details for each chromatin type can be found in their respective section.

4.3.1 The Characteristics of Euchromatin

As described earlier, euchromatin has a characteristic open state by showing a wider space between nucleosomes and is present at active gene sites. This open conformation allows the transcription machinery a higher accessibility. In addition, histone variants and modifications present facilitate this process. The predominant histone PTM marks found at these sites are H3K4me3, H3K27ac, H3K36me.

A characteristic histone feature of chromatin present at promotor regions near transcription start sides (TSS) is Histone 3 lysine 4 trimethylation (H3K4me3). In mammals, this process is catalysed by six homologs of yeast histone lysine methyltransferases SET1 (SETD1A, SETD1B, MLL1, MLL2, MLL3 and MLL4). RNAPII associates with yeast SET1 and deposits SET1 in a co-transcriptionally manner[50–52]. The varying levels of H3K4me3 show enrichment for MLL2 and the SETD1 subunit CFP1. This subunit is capable of reading H3K4me3 mark by its PHD finger domain and thus mediates the interaction between SETD1 and the histone mark[53]. A closer look at Gene bodies of transcriptionally active genes shows an enrichment for H3K36me3, mediated by yeast SET2 (human SETD2). This mark is associated with RNAPII elongation process, where serine 2 of RNAPII is phosphorylated. Furthermore, this histone modification serves as a binding platform for histone deacetylases complexes (HDAC) and thus suppress excessive transcription initiation[54,55]. In yeast it has been shown that this mediated histone modification prevents the hyperacetylation and histone exchange and hence leading to well-spaced nucleosomes over coding regions. Another mark is the acetylation at lysine 27 on H3 (H3K27ac). It stands for active but not yet ready enhancers and are deposited by histone acetyltransferases complexes (HAT)[56,57]. It exists a competitive situation between H3K4me3 and H3K27ac, where the acetylation alone does not determine the enhancer activity. Moreover, H3K4me3 readers were identified in HAT complexes (e.g. SGF29) and it is suggested that at TSS H3K27ac bridges the transition between the mechanisms from transcriptional initiation to elongation[58,59].

Furthermore, histone variants of H2A.Z, H3.3 and H2A.B can be found. The incorporation of H2A.Z in yeast is mediated by the chromatin remodeler Swr1, a Snf2 family of ATPases. Moreover, this variant is replication-independent and in yeast its C-terminal region interacts with RNAPII and therefore promoting its recruitment to promoters [60,61]. The mechanism of incorporation has been shown in yeast, where Swr1 acts first as a histone chaperone and delivers H2A.Z-H2B dimer to the nucleosome and partially unwarps the DNA from the histone core[62,63]. After incorporation, the acetylation of H4 recruits' yeast Brd2, a double-

bromodomain containing protein, which is known to function in transcriptional activation. In addition, the yeast SAS complex (something about silencing) has acetyltransferase activity and can stimulate H2A.Z deposition and therefore suggesting a role of acetylation in facilitating H2A.Z deposition[64,65]. In contrast, histone variant H2A.B has a preference to H4 hyperacetylated regions[66,67]. In studies with ectopically expressed H2A.B it has been associated with active genes involved in transcriptional activation and mRNA processing and its appearance correlates with new DNA synthesis during replication and repair sites[68,69]. The incorporation of the conserved H3.3 histone variant happens throughout the cell cycle at euchromatic actively transcribed regions, unlike the deposition of canonical histones restricted to deposition during replication[70,71]. Its incorporation is mediated by HIRA (Histone regulator A) during the cell cycle stages but can also occur at telomeres and pericentric heterochromatin by the chaperone dimer ATRX/DAXX [72–74].

The associated proteins functioning as histone chaperones are Asf1, CAF-1, HIRA, NAP1 and FACT, which act in a replication-independent or replication-dependent nucleosome assembly fashion[35]. They bind to soluble histones and regulate nucleosome assemblies. When following the synthesis of newly H3-H4 molecules, they interact throughout the course of its assembly onto nucleosomes with different chaperones, which regulate free histone levels and nuclear import of these and thus impacting the deposition of H3-H4 onto DNA[35]. H3.1 is newly synthesized and interacts with Hsc70 and later assembled into a larger complex consisting of histone chaperone t-NASP, histone H4, chaperone Hsp90. When the H3/H4 dimer is assembled the complex interacts then with Hat1-RbAp46 (lysine acetyltransferase), histone chaperone Asf1 and importin-4[75]. This ensures the import of new H3/H4 into the nucleus. By interacting with CAF-1, which is a central replication-dependent chaperone, it deposits H3/H4 onto newly synthesised DNA due to its recruitment by PCNA (proliferating cell nuclear antigen) to the replication fork [76]. In non-dividing cells the predominant mechanism of histone replacement is the replication-independent nucleosome assembly. The histone chaperone HIRA is one of them. It assembles H3.3/H4 onto nucleosomes and HIRAs absence results in reduced localization of H3.3 on genic regions. Another chaperone is the complex formed by ATRX and DAXX, where absence of ATRX leads to defects in H3.3 at pericentric regions and telomeres[77–79]. As a complementary mechanism, H2A/H2B dimer have a chaperone called Nap1 and helps to incorporate these so the nucleosome can be completed. Nap1 facilitates the import of H2A-H2B, its incorporation by disrupting non-productive histone-DNA interactions[80]. Furthermore, H3/H4 and H2A/H2B can be incorporated into the nucleosome by Nap1s versatile association with

ATP-dependent chromatin and remodeler factors to regularly space and assemble nucleosomes. The chaperone FACT (facilitates chromatin transcription) is another species of histone chaperones and is capable to bind both oligomers of H2A/H2B and H3/H4 and facilitates their incorporation. Moreover, it is proposed that through partial DNA unwinding FACT can bind to both H2A/H2B dimers simultaneously during reorganization process[81–83].

The euchromatin contains the highest fraction of exact positioned nucleosomes relative to the underlying DNA sequence and adopts an important role in transcription regulation near the transcription initiation sites. The genomic regions associated with euchromatin are actively transcribing genes and regulatory elements like promotors and enhancers. Latter one regulates the access of transcription factors and chromatin remodelers to facilitate opening the chromatin and leading to active transcription [39,84–87].

A minor fraction of the genome is harbouring transcriptionally active genes and are located at the interior of interphase nucleus where euchromatin resides. The gene body displays the transcriptional region (introns, exons) in conjunction with activating histone modifications to ensure open chromatin state. Moreover, gene bodies are enriched for DNA methylation marks (80-90%). It is believed that DNA methylation on gene bodies prevents intragenic transcription initiation at highly expressed genes. The positioning of Nucleosomes becomes weaker within the gene bodies compared to the transcription initiation sites and decreases gradually along the gene bodies length[88–91].

Transcription is initiated at promotors, regions found typically upstream or at 5' end of a protein-coding or non-coding gene. RNA polymerase II (RNAPII) promotor contains a core promotor with minimal sequence motif such as TATA box and additional motifs to properly initiate transcription[92]. So called general transcription factors (GTFs) bind to the core promotor (TATA-binding protein, TFIIB/E/F/H). The RNAPII and these GTFs form to pre-initiation complex (PIC) and is sufficient to initiate transcription but with lower basal activity. Activating histone modifications in interplay with ~150-bp nucleosome depletion region (NDR) adjacent to the transcription start site are associated with promotors. Recent studies NDRs were shown to contain highly dynamic nucleosomes with histone variants H3.3 and H2A.Z, whereas NDRs are flanked by well positioned nucleosomes[87,93–95]. The localization of H3.3 to active promotors and transcribed gene bodies is mediated by HIRA. A characteristic trait of promotors is the presence of CpG islands, GC-rich DNA regions. The exact role of DNA methylation in transcription is not yet fully understood but studies in mice

showed that promotors need a specific methylation pattern and lack of DNA methyltransferases (DNMT) led to lethality in early development[96–98].

Transcription can be independently activated by a DNA element called Enhancer irrespective of their distance or orientation with regards to the promotors of transcribed genes. These Enhancer elements serve as a binding platform for a variety of transcription factors. Enhancers are frequently transcribed into non-coding RNAs (ncRNAs, regulating transcription) in a bidirectional fashion. The current model for long-range transcriptional regulation assumes a physical proximity of distal enhancers to their target promotor by chromatin looping mechanisms, but it has been shown in a differentiation experiment of embryonic stem cells to neural progenitors, that other chromatin structural conformation has to be encompassed to explain this phenomenon[99–103]. Enhancers are characterized by the inclusion of histone variants H3.3 and H2A.Z and activating histone modifications. The deposition of these histones is replication-independent, and it has been shown, that nucleosomes flanking the transcription factor binding sites of Enhancers are less flexible compared to H3.3/H2A.Z containing nucleosomes, which show hypermobility [56,77,95].

In summary, the euchromatin is transcriptionally active and poses an open conformation, which enables an easier access to promotors and enhancer regions facilitating the transcription of active genes. These processes are guided by histone marks and histone variants and the interaction with transcription factors and chromatin remodelers.

4.3.2 The Characteristics of Heterochromatin

Emil Heitz was the first to coin the term of 'heterochromatin' when he saw the compacted state within the nucleus spatially segregate from Euchromatin. When staining chromatin in cytological samples at different cell cycle stages, he observed regions strongly stained and others became invisible during interphase. It has been shown that heterochromatin is less accessible than compared to Euchromatin and associated with gene silencing due to decreased transcriptional activity. Moreover, heterochromatin localizes towards the nuclear periphery and regions around the nucleolus[104–107]. heterochromatin is mostly defined by sequence independent epigenetic mechanisms[35]. One of many known histone variants associated with heterochromatic regions is MacroH2A and is considered to be transcriptionally repressive and associates with inactive genes. It has been shown. Another interesting observation was made where heterochromatin demonstrated its ability to form phase-separated liquid condensates[108–112].

There are two categories of heterochromatin: facultative and constitutive heterochromatin. The latter one occurs at same genomic regions in every cell type and is propagated throughout the lifetime of an individual once it has been established. This type of heterochromatin is more static in structure, providing a condensed and transcriptionally less active chromatin conformation. Furthermore, constitutive heterochromatin is characterized by the histone mark H3K9me3, mostly found at highly repetitive and gene-poor regions such as pericentric, sub-telomeric regions or transposable elements. The histone mark is deposited by histone methyltransferases (HMTs) such as Suv39H[113–116]. It has been shown that pericentric regions consist of repetitive tandem satellite repeats and fulfil an important role in accurate chromosome segregation by preventing premature chromatid separation. Furthermore, satellite DNA underlying in constitutive heterochromatin are not conserved and greatly vary between different organisms or even within one organism, proposing an epigenetically control[35,117]. In *S.pombe* the formation of the constitutive heterochromatin has been well characterized and will be here briefly depicted. The RNAi pathway with its core machinery (Dicer, Argonaute and RNA-dependent RNA polymerase) are required for this process. The absence or loss of the RNAi machinery resulted in reporter gene silencing; accumulation of transcripts derived from heterochromatin repeats or loss of H3K9 methylation and defects in chromosome segregation[118–121]. The pericentric regions have specialized repeats (dg, dh, see section 4.4) which transcribe into ncRNAs, spliced into siRNAs (small interfering RNAs) which recruits the RNA-induced transcriptional silencing complex onto pericentric regions[122]. To prevent spreading of heterochromatin into Euchromatin (having their own histone modification marks), the pericentric regions are surrounded by boundary elements[35]. As referred earlier, heterochromatin is more compact compared to Euchromatin and the constitutive heterochromatin is densely packed and influences variety of functions such as heritable gene expression or maintenance of chromosome integrity. It has been shown that HP1 oligomerizes *in vitro* and compacts chromatin into phase-separated condensates but *in vivo* studies suggest the involvement of additional factors, which condense H3K9me heterochromatin in the nucleus [109,123–126]. Moreover, it has been shown that mammalian HP1 contains an RNA-binding domain and is therefore required to localize to pericentric regions residing in the heterochromatin[127]. Recent studies suggest a role of RNA and RNA-binding proteins in the compaction event of constitutive heterochromatin, and it cannot be ruled out that similar mechanisms can be found at facultative heterochromatin sites compacted by PRC1 complex [126–131].

In contrast to constitutive heterochromatin, the facultative heterochromatin assembles at regions regulating genes in development and differentiation by switching between transcriptionally active and inactive chromatin states[35]. Since cells differentiate differently the overall facultative heterochromatin will not display identical pattern. This specific region is regulated by polycomb group (PcG) proteins and mediate H3K27me3 marks. Moreover, this specific subtype of Heterochromatin can be found on the entire silenced X-chromosome or discrete domains that are distributed genome-wide. The histone marks present at the facultative heterochromatin are H3Kme2, H3K27me and DNA methylation pattern on inactive alleles found in imprinted autosomal genes. It has been shown that the CTCF protein (multiple zinc-finger containing architectural protein CCCTC-binding factor) serves as a gatekeeper between facultative heterochromatin and euchromatin is able to bind to RNA and therefore regulates chromatin looping[132–137].

4.3.3 The Characteristics of Centromeric Chromatin

The centromere serves as the nucleation site of spindle pole-microtubule binding to faithfully segregate chromosomes. The high evolution rate of DNA and centromeric components across eukaryotes may serve as explanation in the reproductive isolation of emerging species[138].

Since the centromeric sequences vary among the species (from ~100 bp point centromere in *S.cerevisiae* up to several megabases of repetitive centromeres in plants and mammals) this specific region consists of hypoacetylated and lacking H3K9me2/me3 Chromatin marks. The centromeric Chromatin is characterized by the histone variant CENP-A, which replaces canonical H3 at these sites. Most of the histones are similar and have common structural features but in case of CENP-A it is highly divergent, and CENP-A interacts with centromeric DNA unlike seen with H3. This specific histone variant is the foundation for the kinetochore assembly, which later attaches to the microtubules at the spindle pole. Upon applying tension, the sister chromatids can be separated to each opposite site and ensure faithful chromosome segregation. Till today the repetitive nature of centromeric sequences makes it difficult to study the mammalian centromere[35]. First studies by Furuyama and Henikoff (plasmid-based genetic assays) in *S.cerevisiae* showed that CENP-A containing nucleosomes undergo positive supercoiling by possible right-handed wrapping of centromeric DNA superhelix around the histone octamer [139]. The nature of this observation has to be determined in other species, since *S.cerevisiae* contains a so called "point" centromere and all higher eukaryotes are classified as "regional" or "holo" centromeres[35].

The histone marks observed at the centromere are diverse (Figure 4). Some of them are known to modulate the accessibility of CENP-A Nucleosome and thus facilitates the binding of centromere/ kinetochore associated proteins.

The human genome project was completed in 2003 and the information gained over the 13 years of research including the identification of ~20,500 genes and subsequently analysing the underlying sequence, helped to understand causes in cancer or heritage derived diseases[140]. However, it could not cover the centromeric regions due to its repetitive nature it still proposes a major challenge to modern sequencing methods. Latest research put an effort to overcome the limitations and just recently published the genomic and epigenetic landscape of the human (peri)centromere[141]. Before this paper got published, researchers used a bottom-up-functional genomics-based approach to address this specific question. By using CENP-As binding capacity to centromeric alpha satellites, the large majority of CENP-A bound human functional centromeres could be detected and found to be tandem arrays of dimers. These native alpha satellite dimeric arrays served as linear maps to profile CENP-A and other centromeric proteins. It has been shown that besides CENP-A also CENPB, CENP-T and CENP-C were tightly bound[142–144]. Results of these studies suggested a tight association of CENP-A containing nucleosomes and other centromeric proteins and are more tightly packed as compared to euchromatin[35].

A research article[141] publishing the complete sequence of the human centromere region stated out that satellite repeats constitute 6.2% of T2T-CHM13 genome assembly where alpha-satellite represents the single largest component of about 2.8% of the genome. By studying alpha-satellite repeats and their sequence relationship across each centromere, the researchers found that centromeres evolve through "*layered expansions*". There are distinct repetitive variants within each centromeric region. Moreover, these expand through a mechanism that resembles as successive tandem duplication. In contrast, older flanking sequences decrease and differ over time. The most recent expanding repeats within an alpha satellite array are more likely to interact with CENP-A, coinciding with regions of reduced CpG methylation. Taken together these findings point towards a strong relationship between local satellite repeat expansion, DNA hypomethylation and subsequently kinetochore positioning. In addition to these findings, they found unexpected large structural rearrangements affecting multiple satellite repeat types including active centromeric alpha satellite arrays. As a last finding, the comparison of 1600 individuals' X chromosome revealed that recent African ancestry possesses the greatest genetic diversity in the region

surrounding the centromere and lead to the evolution of African alpha satellite sequence variants[141].

Now the sequence of the human centromere is uncovered, researchers in the future can address the still open questions in the centromere biology field and draw conclusions on the molecular mechanisms between centromere, alpha satellite expansion, inner kinetochore architecture and DNA hypomethylation.

4.3.4 A new player at the Centromere: Centromeric Transcription and its impact on Centromeric Chromatin

The transcription of the centromere is a critical factor whether the centromere is a "point" or "regional" centromere. Most higher eukaryotes have a regional centromere, commonly comprised of tandem repeat arrays and transposable elements, whereas *Saccharomyces cerevisiae* has a point centromere. If holocentric species such as *Caenorhabditis elegans* and lepidopteran *Bombyx mori* have this kind of transcription is not known, since there is an anti-correlation between active transcription and localization of CENP-A and CENP-T chromatin[145–147]. It has been shown that RNA transcripts are strictly regulated and need maintenance in all cellular processes. lncRNAs have been implicated in cellular functions such as: chromatin architecture, chromatin remodeling, transcriptional and translational regulation, formation of nuclear bodies, formation of heterochromatin, inner centromere signalling and CENP-A recruitment[145]. Genes can be embedded in centromeric chromatin, but it is not known whether the repetitive centromeric DNA encode for proteins[145]. In order to obtain mature mRNA, the pre-mRNA is processed whereas other RNA species are actively degraded. Like mRNA the cenRNA species are processed in similar way, since they also undergo splicing events to produce mature transcripts[148,149]. It has been shown that centromeric RNA associates with splicing factors such as DHX38, which plays an important role in pre-mRNA splicing[150]. This provides a hint that cenRNAs interact with various RNA processing components.

It has been shown that the dynamic balance between euchromatin and heterochromatin at centromeric region is essential for establishment of an active kinetochore as well as the transcription of these regions. Latest publications described in *D. melanogaster* as well in human, the distinct set of histone modifications occurring at the centromeric region such as the demethylation of H3 Lysine 4, presenting an open chromatin yet not active. Moreover,

the centromere lacks the mark for heterochromatin (H3 Lysine 9 dimethylation) or euchromatin (H3 Lysine 9 acetylation)[151,152].

Recent work describes two centromeric transcripts (cenRNA): long non-coding RNAs (lncRNA) which are greater than 200 nucleotides whereas small RNAs comprise of smaller transcripts < 200 nucleotides[145]. The lncRNAs are further processed based on its nuclear localization, which can result in siRNAs (silencing RNAs) produced by the Dicer machinery[153]. The role of lncRNAs include modulating the mRNA cleavage, translational repression, and regulation of alternative splicing[153–155]. The rapid evolution of centromeric DNA leads to limited conservation among closely related species. Since the conservation is limited, the lncRNAs might be conserved based on their function rather than their sequence, therefore one aspect could be facilitating the centromere repositioning on chromosomes[156,157]. In human the transcription of the alpha satellite, which is the centromere core region, happens during early G1 and mitosis[145,158]. The length of the transcripts is between 0.5 to 2.45 kilo base pairs (kbp). It has been shown that centromere specific proteins such as CENP-A, CENP-B, CENP-C, HJURP, SGO1, Aurora B, DHX38 and Suv39H1 interact with centromeric transcripts[145]. Furthermore, the transcripts are polyadenylated as can act as *cis* or *trans*[145]. In contrast, in *Drosophila melanogaster* the satellite III (core region of centromere) is transcribed during mitosis and has a length of about 1.3 kbp[159,160]. CENP-C has been shown to interact with these transcripts, whether the transcripts are polyadenylated or act in a *cis* or *trans* way is not known yet[145]. Centromere transcription is also found in *Saccharomyces cerevisiae*, where the cenRNA is transcribed ruing S-Phase and has a transcript length of 462-1754 nucleotides. It is neither known which proteins bind to these shorter transcripts nor if these transcripts are polyadenylated or act in a *trans* or *cis* configuration[161–163].

One of the better understood model organisms for centromeric transcription is *Schizosaccharomyces pombe*. There the centromere structure is different compared to other plants or animals. The centromeric core is flanked by a pericentric inverted repeats and these repeats are transcribed. Moreover, it has been shown that these repeats and their transcription is essential for the pericentric heterochromatin formation. Here the distinct separation of the functional centromere and its neighbouring pericentromere is in higher eukaryotes difficult due to their repeat-rich regional centromeres[122,145,164].

4.4 The Evolution of the Centromere and how DNA methylation helps to shape the Chromatin

The centromere is characterized by the centromere-specific histone H3 variant CENP-A/ CENH3[145,151]. This distinct mark helps to recruit the kinetochore components to faithfully segregate the chromosomes during mitosis/ meiosis and ensure proper propagation of the genome to the daughter cell[152,165]. Centromeres vary in their occupation site across different organisms. In general centromeres are metacentric (in the middle of chromosomes) but there are cases where they are acrocentric (separate chromosome arms of different length) or telocentric (positioned at chromosome's end)[166]. Due to the fast evolving and rapid evolution of centromeric and kinetochore associated proteins, the fast evolution of centromeric DNA and its associated transcripts result out of this process. Moreover, the centromeric DNA is highly repetitive in its nature and is highly diverse even in closely related eukaryotes. This is summarized in the term "*centromere paradox*". The "*centromere paradox*"[138,145,151] describes the discrepancy between the essential function of the centromere but also its rapid evolution of its components[138]. Having a closer look, the variation in sequence and structure of the centromere has emerged as a common theme.

The rapid evolution of centromeres happened through homologous recombination between stretches of tandemly repeated sequences. These can vary within one organism and centromeric sequences differ significantly between its chromosomes[167].

Starting off in understanding centromere characteristics, the simplest model organism is *Saccharomyces cerevisiae*. It has a so called "point" centromere (Figure 5). It is made of 120 bp containing three DNA elements (CDEI, CDEII and CDEIII) wrapping around a single centromeric nucleosome. The centromeric histone H3 variant here is called Cse4. The CDEIII element consisting of alternate stretches of A and T nucleotide residues causes DNA bending. Moreover, this element is flanked by palindromic motifs of CDEI and CDEII. It has been shown that CDEI is not essential for kinetochore activity, but its mutations caused chromosome loss, therefore the centromeric sequence per se defines the centromeric identity[152,168–170].

In *Schizosaccharomyces pombe* the situation at the centromere becomes more complex. Here the centromere comprises of 40-110 kb in length and has been designates as "regional" centromere (Figure 5). The core (4-7 kb) named *cnt* encloses multiple centromeric nucleosomes with its centromeric histone Cnp1. The core is flanked by inverted

and 6 kb long innermost so called *imr* repeats and contains tRNA clusters. These three elements form the central domain which are further flanked (left and right) by outer repeats (*otr*) called *dg* and *dh*[152,168,171].

In *Drosophila melanogaster* the centromere with its centromeric histone CID comprises of 200-500 kb and made of up to 10 bp repeats (AATAACATAG)n followed by 11 to 12 bp tandem repeats (CCCGTACT[C]GGT)[152,172] (Figure 5). Nevertheless, in moths, butterflies, arachnids, and nematodes (*Caenorhabditis elegans*) the centromere is labelled as "holocentric" where the entire chromosome, except the telomere ends, are one centromere region. Moreover, in *C.elegans* only few tandem repeats are found and about 50% of the genome is with CENP-A associated in 20 centromeric domains varying in size[173–175].

Most eukaryotes are monocentric, containing only one centromere at each chromosome. In higher eukaryotes the centromeric repeats are about 150 bp in length[152,176].

In mice (*Mus musculus*) the alpha-satellite consist of homogenous arrays of 120-bp minor satellite repeats (MinSat), which are flanked by less ordered 234 bp major γ-satellite sequences (MajSat)[177].

In human (*Homo sapiens*) the centromere (with centromere histone CENP-A) is more complex (Figure 5). It contains alpha-satellite monomers of A and B-type[178]. Both differ in a 17-bp sequence (A-box or B-box), which has not been described for mice centromeres. The B-box is known for the binding of CENP-B[179]. It is not clear whether there are proteins binding to the A-box. Furthermore, the alpha-satellite sequences organize in a back-to-back fashion and forming higher-order repeats (HOR). Within these HOR the alpha-satellite monomers are 50-70% identical. The arrays are 2-5 Mb long which consists of 100-1000 times of repeated HOR[152,180,181].
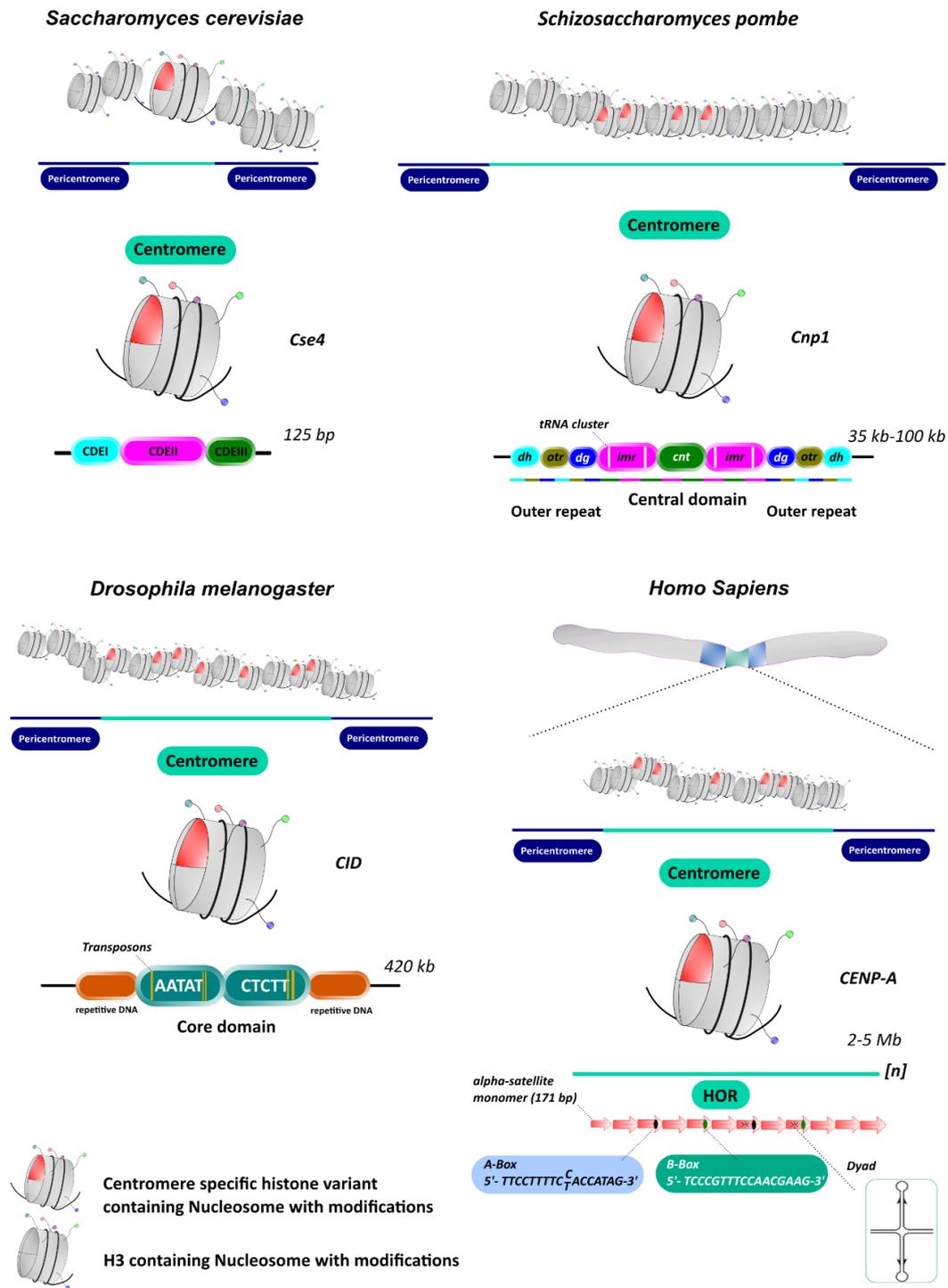
Figure 5: **The Evolution of the centromere**. Each model organism is depicted with its respective centromere organization. The centromeres were designed as described in the text based on current knowledge.

The repetitive elements in the genome are known for their DNA methylation status as seen for example at transposable elements, which are repressed by DNA methylation marks and therefore impede their mobility. Mammalian CpG dinucleotides (~70-80%) present at satellite/ repetitive and non-repetitive intergenic DNA are associated with transcriptional repression [182,183]. A closer look within the centromere, the alpha-satellite sequence contains three methylatable CpG di-nucleotides, where two of them are present in the CENP-B-box[179]. Since the nature of the centromere composes of these repetitive sequences it is tempting to hypothesize a role of DNA methylation in centromere biology. The mechanism of DNA methylation might be evolved as a protection mechanism as well as ensuring genomic integrity by defending the eukaryotic genome from random insertion by parasitic elements[184,185]. Moreover, DNA methylation might have a bi-functional role such as regulation of gene expression and chromatin architecture[186,187]. The DNA methylation epigenetic mark is associated with transcriptionally inactive region in the genome but is essential for homeostasis and development[185]. Possible consequences of aberrant DNA methylation at (peri)centromere could be a) increased rate of recombination processes at centromeric repeats; b) increased centromere breakages; c) changes in centromere-protein associated network due to displaced methyl-binding partners; d) increased rates of alpha-satellite transcripts leading to genomic instability[185]. CpG dinucleotides often occur as repetitive dinucleotides and the primary methyl modification happens at 5-methylcytosine (5mC) by a covalent attachment of a methyl group to the C-5 position of cytosine[182,187]. Specific methyltransferases enzymes are known to catalyse this reaction by cooperating to maintain the DNA methylation pattern: DNMT1, DNTM3A, DNTM3B. The enzymes DNMT3A and DNTM3B are responsible for *de novo* methylation by targeting newly integrated host sequences, which can be unmethylated or hemimethylated CpG substrates. Both enzymes have their prime time during early development specifically after global demethylation of the genome in the pre-implantation embryo[188,189]. In case of DNMT1, it acts on pre-existing DNA methylation marks and maintains these patterns over course of the cell division and has a high affinity towards hemimethylated DNA sequences. Moreover, it has been shown that DNMT1 might have a role in *de novo* methylation in the context of repetitive elements. In a DNA methylation pattern analysis by high-resolution sequencing of repetitive elements including major satellites unravelled a *de novo* methylation activity of DNMT1 in these specific regions[190]. Additionally, DNMT1 has been shown to be associated with heterochromatin in a replication-independent manner, therefore is maintains of the centromeric heterochromatin[191]. Nonetheless, DNMT3A and -3B might be additionally

required to preserve established methylation patterns[192–194]. Both enzymes have been shown to be involved in global methylation maintenance in mouse embryonic stem cells (mESC). It is possible that both enzymes might have an additional role in methylation pattern maintenance at the centromere[195]. Besides mastering gene transcription, DNA methylation is capable of modelling chromatin structure and therefore influencing the accessibility of DNA damage machinery and recruitment of DNA methyl-binding proteins. It has been shown that CpG within centromeric DNA sequence are strongly marked by DNA methylation and tightly regulated during the cell cycle stages[196,197]. Furthermore, it remains elusive if different DNMT enzymes act in a specific or more redundant way on the (peri)centromere to deposit the DNA methylation epigenetic marks[185]. In several studies the link between CENP-B and methylated DNA sequences have been established. CENP-B is currently the only known centromeric protein bearing a sequence specific DNA binding property to the 17 base pair long CENP-B (CBB) box present in ~30% of all human alpha-satellites[179,198,199]. A closer look at the sequence of the CENP-B reveals the presence of two out of the three CpG di-nucleotides in the whole alpha-satellite consensus sequence[179]. Moreover, CENP-B's association with the CBB is methylation-sensitive, since a steric clash between the DNA methyl group and the side chains of CENP-B (T44, R125) occurs[200]. Okada, T. *et.al* (2007) showed in their study, that once CENP-B is chromosomally integrated in human artificial chromosome carrying alphoid DNA, it enhances CpG methylation and H3K9 trimethylation deposition and thus preventing CENP-A incorporation. Nonetheless, the methylation mark at native CENP-B binding sites remained unaltered[201]. These observations lead to a possible role of CENP-B and its involvement in establishing precise DNA methylation patterns of centromeric sequences[185]. Some studies hint towards a functional link between CENP-C and DNMT3B but has to be further investigated if both share a role in overall maintenance of centromeric methylation. One study used Yeast two-hybrid screening and immune-precipitation assays and associated DNMT3B to CENP-C. However, knock-down of CENP-C induced a ~20% DNA methylation reduction at alpha-satellite and pericentromeric satellite 2. In addition, the reduction of CENP-C resulted in altered DNMT3B localization and increased centromeric loci transcription[202]. Other factors of centromeric chromatin assembly maybe involved in DNA methylation maintenance. One of these is Mis18α, which mediates the recruitment of de-novo synthesized CENP-A and has been shown to interact directly with DNMT3A and -3B in mouse cells. It might be possible that Mis18α tether DNMT3A/-3B to centromere and help to maintain DNA methylation levels, but further studies are required to investigate this relationship[185,203,204].

The DNA methylation process is influenced by the nucleosome modification pattern of pericentromeric chromatin. Evidence has been shown in a study where deletion of Suv39h1 and Suv39h2 resulted in altered methylation pattern of major satellites in mESCs. These methyltransferases are required for H3K9 di- and trimethylation. In addition, DNMT3B and Suv39h interact and Suv39h proteins are required for DNMT1 and UHRF1 recruitment to cooperatively maintain DNA methylation at major satellites in mouse, whereas *de-novo* DNMT3A and -3B recruitment was unaffected. Another interesting observation was the favoured tethering of DNA methylated major satellites to the nuclear lamina by chromatin reorganization upon depletion of all three DNMT enzymes simultaneously[205,206]. The interplay between DNA methylation and H3K9 trimethylation (H3Kme3) at (peri)centromere can be reinforced by the activity of the ISWI chromatin remodelling complex NoRC (nucleolar remodelling complex). It has been shown that a subunit (TIP5) of this complex is able to stimulate H3K9me3 deposition at minor and major satellite repeats in mouse and can interact with CENP-A[207]. A member of the SNF2 family of helicase/ATPases, the ATRX protein, has been linked to centromere chromatin composition and DNA methylation pattern. ATRX is involved in several cellular processes including H3.3 deposition at the (peri)centromere by interacting with DAXX. ATRX has been affected in ATRX syndrome (X-linked α-thalassaemia mental retardation). Patients with this disorder show aberrant methylation at satellite repeats at centromeric and sub-telomeric regions(ref58). ATRX is known to bind MECP2, which contains a methyl-l-CpG-binding domain (MBD), and this domain mediates MECP2 localization to the (peri)centromere as well as its interaction with ATRX[77,208,209]. The genomic stability relies on the maintenance of DNA methylation pattern since DNA hypomethylation results in the formation of micronuclei, which are formed from chromosome segregation errors, defective nuclear envelope formation or hypomethylation of (peri)centromere regions[210–213].

In some studies, the relation between DNA methylation and chromosome condensation have been addressed. Indeed, upon depletion of condensing complex resulted in chromatin decondensation which represents a similar effect as seen in AC-induced DNA demethylation[185,214,215]. DNMT3B and condensin complexes interact with each other leading to the assumption that condensin might be involved in DNA methylation maintenance and stability of (peri)centromeric regions[185]. Furthermore, a correlation between hypo- or demethlation of pericentromeric region and lack of centromeric cohesion has been assumed and propose that DNA methylation might interfere with chromatid cohesion and therefore regulating interaction between condensin, cohesion and DNA[216]. The precise mechanism

of action regarding DNA methylation and its influence altering the structure and accessibility of chromatin are not yet well described but it has been shown that DNA methylation is capable to induce closed chromatin conformation and a sequence-independent manner. This influences nucleosome stability and dynamics which are described in changes in nucleosome positioning and assembly[217–219] . In addition, DNA methylation might cooperate with histone post-translational modifications (PTMs), which differ at centromeric and pericentromeric chromatin. The abundance of histone marks H3K9me2/me3, H4K20me3 and H3S10ph resemble to heterochromatin landscape by low levels of histone acetylation marks. In contrast, centromeric chromatin is densely packed but devoid of H3K9me3 but enriched in H3K4me2, which is permissive for gene transcription. This specific mark is proposed to facilitate HJURPs task in CENP-A assembly as shown by assembly assays on synthetic human kinetochore. This specific transcription process at the centromere is required for proper functionality of the centromere and builds a link between DNA methylation, histone marks and centromere transcription but still is in its infancies[185,220–222].

In conclusion, DNA methylation has been shown to play a part in the (peri)centromere stability and several complexes involved in maintaining DNA methylation pattern as well as disease association, draws a link between centromere functionality, histone marks, centromere transcription and DNA methylation but have to be further studied to understand its implication and how the relationship between these biological processes ensures proper chromosome segregation and chromatin integrity.

4.5 Linking Centromere to the Microtubules: The Kinetochore as bridge for faithful Chromosome Segregation

For many decades, scientists have been dedicated unravelling the mechanisms of cancer development. One field of cancer research is interested in the mechanism of accurate chromosome segregation and its regulation. By understanding the process of chromosome segregation, novel regulatory proteins might serve as potential targets in drug development, leading to new therapeutic cues. One major aspect of chromosome segregation is the load-bearing attachment of microtubules to the centromere, mediated by the proteinaceous multi-subunit formation, called the Kinetochore. In the last decades biochemical reconstitution, immunofluorescence microscopy, structural biology and mass spectrometric approaches provided detailed insights into the kinetochore assembly and its mechanism to ensure proper chromosome segregation.[166,223,224]

The Kinetochore (Figure 6 C, 7) functions as a mechanosensory, controlling the stability of microtubule attachment to the centromere in a bi-orientated fashion. The incomplete or erroneous configurations seen in mono-orientation, merotelic attachment or syntelic attachment are corrected before the cell executes the segregation processes[225] (Figure 6 A). Before the cell can exit the mitotic state and ensuring the faithful segregation of sister chromatids to each side of the opposite spindle pole, the Kinetochore ensures proper attachment by applying tension and by regulating the spindle assembly checkpoint (SAC, metaphase checkpoint) as a feedback mechanism to favour the bi-orientation of sister-chromatids and initiate the mitotic exit[166,226–228]. An additional checkpoint complex, the Chromosome passenger complex (CPC) is actively involved in mitotic progression and error correction[229]. As described earlier the cell cycle is controlled by different Cyclin-CDK complexes, having their peak at specific cell cycle stages. By the proteasomal degradation of these specific mitotic associated complexes, the cell paves the path towards exiting mitosis and entering G1-Phase[166,230].

A recent study identified SENP6 as a critical factor for maintaining centromere and kinetochore integrity. SENP6 is a small ubiquitin-like modifier (SUMO)-protease[231]. The depletion of SENP6 leads to hyper-SUMOlyation of CENP-C and CENP-I leading to a reduction of centromere bound CENP-C, CENP-T and CENP-A. Moreover, a reduction of the KMN network (KNL1, Mis12 complex, Ndc80 complex) components NNF1, DSN1 and HEC1 was observed. The loss of SENP6 showed a phenotype in losing old and newly

deposited CENP-A and SENP6 is required throughout the entire cell cycle to prevent CENP-A from being removed from the centromere as well as for stabilizing the centromere and kinetochore[232]. This example shows how little is known till today, which proteins associate with the Kinetochore and shape their relationship towards the microtubule-spindle pole bodies. Moreover, it shows a prime example on the interplay between the CCAN and KMN network to achieve faithful chromosome segregation.

In the following paragraph the human CCAN as well as the KMN will be described in more details. In Figure 6 C, the CCAN bound to a centromeric CENP-A Nucleosome is depicted and provides a unique insight on how the centromere and the kinetochore modules interact with each other. A more summarized schematic model containing the latest findings is depicted in Figure 7. Recent structural studies uncovered the structural inheritance of the human CCAN bound to a centromeric CENP-A Nucleosome, providing an interesting insight on how the Kinetochore assembles on the centromeric CENP-A site. It has to be mentioned, that the human Kinetochore members have homologs in buddying yeast, which will not be elicited here in detail. For further details and a comprehensive summary of the buddying yeast kinetochore, the review of Kixmoeller K, Allu PK, Black BE, 2020[165] is recommended. Here the authors compare the known structure of the buddying yeast Ctf19 complex to the structural knowledge gained over the human CCAN and compare these.

A

- Monotelic attachment (Mono-orientation)
- Syntelic attachment

Kinetochore

Cohesin

Spindle pole and microtubule

- Amphitelic attachment (Bi-orientation, favoured)
- Merotelic attachment

B

H3.2 Tri-Nucleosome with H1, PDB:7pf0
CENP-A Tri-Nucleosome, PDB: 6l49

■ CENP-A (orange)	■ H3.2 (cyan)
□ H2A, H2B, H4	■ H1 (green)

CENP-A Nucleosome with CENP-N, PDB: 7u47

C

CCAN

CENP-A Nucleosome

front

side

CENP-C central motif

front

side

CENP-C PEST Domain

extranucleosomal DNA

bottom

top

Centromeric CENP-A Nucleosome with CCAN, PDB: 7ywx

■ CENP-A	
■ CENP-C central motif, CENP-C PEST	
■ CENP-N/L; CENP-N	
■ CENP-H/I/K/M	
■ CENP-O/P/Q/U/R	
■ CENP-T/W/S/X	

Figure 6: **The different attachment of the spindle poles to the centromere and the structural features of the CENP-A Nucleosome in comparison to H3**. **A)** The different attachment styles of the spindle poles favouring the amphitelic attachment which bi-orientates the sister chromatids. Other

attachment styles lead to aneuploidy and chromosome miss-segregation. **B)** The structural differences of the CENP-A containing Tri-Nucleosome in comparison to canonical H3 Tri-Nucleosome (here H3.2). The structure with H3 contains the histone linker H1. CENP-A Nucleosome do not bind H1 and depict the difference between both Nucleosome structures, the DNA-termini are in case of CENP-A to flexible/ apart and cannot be connected by H1. **C)** It is assumed that CENP-N takes over the role of H1 in CENP-A containing Nucleosomes, as seen in the structure of the CENP-A stacked Nucleosome with CENP-N (right; PDB: 7u47). The recent structure of the centromeric CENP-A Nucleosome in complex with the human CCAN has been published. Here the subcomplexes are coloured in similar colour to show the architecture of the CCAN. All structures were loaded into ChimeraX program and re-coloured for visualization purposes tailored to this doctoral thesis. For further information, the original publications are recommended.

4.5.1 The Constitutive Centromere Associated Network (CCAN)

4.5.1.1 Centromeric chromatin shaped by CENP-A containing Nucleosomes

By incorporating CENP-A into the nucleosome, it actively marks the centromere and neocentromeres. Moreover, most centromeric nucleosome contain histone H3 (~97%) and only three to four percent are assembled with CENP-A. Looking closer, per centromere 200 CENP-A containing nucleosomes are present. There have been several different models described for the structure of CENP-A containing nucleosomes[233]. The first model describes the CENP-A nucleosome as the known conventional octameric nucleosome with two molecules of CENP-A, H4, H2A and H2B, which resembles similar to the known H3 containing chromatin [234–236]. Another structure was described as a hemisome, where only one molecule of each histone is present. The model of a tetrasome where H2A and H2B are lacking was proposed, as well [237–239]. In some cases, a heterotypic nucleosome where one molecule of H3 and one molecule of CENP-A have been proposed [240,241]. In another model combining the previous described structures describes the oscillation between two forms. During G2, M and G1-Phase the CENP-A Nucleosome is present as a hemisome, whereas in S-Phase as an octameric complex[242]. Using different approaches over the past years, the majority of CENP-A molecules assemble as a homotypic, octameric nucleosome, which contains two molecules of CENP-A, H4 and H2A and H2B at all cell cycle points. Nevertheless, the heterotypic CENP-A/H3 containing nucleosomes comprised at most two percent of CENP-A containing chromatin [243,244]. Recent Cryo-EM structures (Figure 6 B) revealed the unique feature of CENP-A Nucleosomes compared to H3 containing Nucleosomes. The DNA ends are highly flexible at the entry and exit sites of CENP-A and helps to change the chromatin structure by transient DNA unwrapping at these specific sites at all cell cycle points[167,245–247]. The flexible linker at the entry/ exit side of the Nucleosome shows no cross above the dyad and therefore histone H1 cannot accommodate the binding at this region (Figure 6 B, left side compared to H3.2 bound to H1)[248]. In comparison to canonical nucleosomes containing H3 and H4 histones, CENP-A forms a more rigid contact with H4. Moreover, the alpha satellite DNA sequence providing another dimension of flexibility/ rigidity into the CENP-A Nucleosome-Kinetochore formation process[165].This specific feature is mediated by the CENP-A N-terminal tail as well from its targeting domain (CATD), which consists of a first loop (L1) and a second alpha helix (α2) and is sufficient for CENP-A centromere targeting[165]. This leads to a more global condensed chromatin state but allows accessibility for centromere binding proteins by loosening CENP-A Nucleosome

DNA superhelical termini. Sequence alignment of CENP-A with H3 has shown that CENP-As histone fold domain only shares 62% sequence identity whereas its N-terminal tail does not share sequence identity with H3 N-terminal tail. What is shared between both in the N-terminal tail is the basic character, where lysine residues of H3 have been replaced by arginine residues in CENP-A[165].

The more lose DNA superhelical termini at CENP-A Nucleosome entry/ exit side provides the Kinetochore complexes a platform to build on. Since the Kinetochore is a several Mega Dalton big complex, consisting of several subcomplexes, the Kinetochore can be divided into inner Kinetochore and outer Kinetochore layer, based on the localization of the subcomplexes in proximal distance to the centromere. The inner Kinetochore is called CCAN (Constitutive Centromere Associated Network) and building the very inner axis of Centromere Proteins (CENP) binding to the CENP-A Nucleosome directly, whereas the outer Kinetochore (KMN network) binds to the most outer layer of the CCAN and to the microtubules of the spindle pole bodies (Figure 7). Most of the kinetochore proteins have orthologs in *S.cerevisiae* but are lacking, with the exception of CENP-C, in certain lineages such as *D.melanogaster* or *C.elegans*. The subcomplexes making up the CCAN and rely on reciprocal interactions of their cognate complexes are CENP-B, CENP-T/W/S/X, CENP-C, CENP-N/L, CENP-H/I/K/M and CENP-O/P/Q/R/U[166].

4.5.1.2 CENP-A loading and recruitment are dependent on Centromere Transcription

CENP-A is loaded onto the centromeric chromatin in a cell-cycle specific manner[249]. During the cell division the total number of centromeric CENP-A is halved[250]. Hence, CENP-A levels have to be maintained to guarantee kinetochore formation and centromeric CENP-A pool must be replenished by new CENP-A. In recent studies, it has been shown that nascent CENP-A loading is dependent on centromeric transcription[145]. First clues were discovered by studying *Schizosaccharomyces pombe*, where mutation of RNAPII machinery in initiation and elongation factors reduced the overall level of CENP-A^{Cnp1} [251,252] . Following this observation other works in mammalian cell lines and *Xenopus* oocyte extract showed the knock-down of centromeric transcripts resulted in reduced CENP-A levels at the centromere[253–255]. By chemically inhibiting active RNAPII, centromeric CENP-A^{CID} was lost as shown in mammalian and *Drosophila* cell lines[256,257]. The elongation factor Spt6 facilitates the maintenance of centromeric CENP-A^{CID} [232]. Furthermore, inhibition of RNAPII during mitosis lead to decreased centromere transcription therefore resulting in anaphase

lagging and abnormal chromosome splitting[257,258]. This active transcription process is required to facilitate replacement of placeholder H3.3 histone by new CENP-A[259,260]. The recruitment of CENP-A/H4 is mediated through the specific CENP-A chaperone interaction (in humans HJURP and RbAp48[261–263]; in fruit flies CAL1[264] and fission/buddying yeast SCM3[265,266]). These chaperones have been shown to interact with centromeric transcripts. In fruit flies the knock-down of 359 bp satellite RNA reduced the levels of CAL1 at the centromere[267]. Similar in human, HJURP shows direct interaction with alpha satellite lncRNAs. The levels of CENP-A and HJURP reduced at the centromere when lncRNAs were knocked down[257].

Since the incorporation of ectopic CENP-A leads to chromosomal instability, the deposition of CENP-A has to be tightly regulated and its assembly onto centromeric chromatin is restricted once per cell cycle. Therefore, the replication of the centromeric chromatin is uncoupled from centromeric DNA replication[167]. In humans the process of replication happens in late S-Phase[268,269]. The active deposition and assembly of CENP-A occurs after the exit from mitosis, when HJURP activity rises (late anaphase/ telophase and is restricted to the early G1-Phase[270,271]. The presence of the Mis18 complex, consisting of Mis18α, Mis18β and M18BP1, is a prerequisite to target CENP-A and HJURP to the centromeres. In anaphase the complex localizes to the centromere and remains associated till G1[204,272]. Depletion studies on any of the Mis18 complex subunit resulted in chromosome segregation defects, misaligned chromosomes, and interphase micronuclei formation[204,273,274]. Besides the contribution of the Mis18 complex, the global regulation of CDK1/2 kinase activity is detrimental for CENP-A deposition and restricts this process for the G1 Phase. The highest activity of CDK1/2 is during S-, G2- and M-Phase until the mitotic exit. CDK1/2 phosphorylate M18BP1 and HJURP, leading to inactive complexes until at anaphase onset the CDK activity drops and M18BP1 gets dephosphorylated[275]. In parallel the phosphorylation of Mis18α, Mis18β and M18BP1 by PLK1 promotes its assembly and centromeric localization. This is the signal for the recruitment of HJURP and possible the incorporation of CENP-A in early G1[276–278]. After successful deposition, CENP-A incorporation is stabilized by the RSF complex and MgcRacGap[279,280]. A recent study[268] described how CENP-A inheritance at the centromeric DNA replication fork is maintained. CENP-A has been reproducibly localized to the same centromeric sequences before and after DNA replication. On one hand, the DNA replication machinery replicates the DNA but on the other hand it maintains the epigenetically defined centromere identity by mediating the precise reassembly of centromere-bound CENP-A chromatin onto the exact same

centromeric DNA sequence within the replicating daughter centromere. Moreover, CCAN and CENP-C were essential during DNA replication to retain CENP-A at the centromere. A rapid degradation experiment of endogenously tagged CENP-C at early S-Phase provided evidence that its importance during S-Phase for CENP-A is essential, since upon deletion, CENP-A gets lost in G2-Phase. During late S-Phase where centromeric DNA replication happens, CENP-A has been associated with MCM2, a core subunit of the DNA replicative helicase MCM2-7 complex. The complex recycles old histones at the replication fork[281]. A current model for CENP-A centromere during S-Phase could be described as followed: The retention of CENP-A depends on the presence and physical interaction of CENP-C and CCAN members with CENP-A but is additionally coupled with coordinated actions of MCM2, HJURP and CAF1. They enable the precise assembly of CENP-A into the chromatin within each daughter centromere and thereby maintaining epigenetically defined centromere identity[268]. It has been shown that the histone variant H3.3 serves as a "placeholder" for CENP-A during S-Phase[282]. In G1 this mark gets replaced by new CENP-A, which has been observed in fission yeast[283].

Recent studies in human and mice showed the importance of centromeric RNA in heterochromatin formation[284,285]. The influence of pericentromeric heterochromatin have been extensively studied in fission yeast[164]. In flies, CENP-A[CID] prefers to seed out at euchromatin-heterochromatin boundaries, therefore suggesting that CENP-A chromatin prefers to be right next to heterochromatin domain[286]. In paternal mouse pronucleus the pericentric transcripts created supresses SUV39H2 activity therefore resulting in reduced H3K9me3 levels, which might be caused by the formation of RNA-DNA triplexes near the nucleosomes[285]. In contrast, in human SUV39H1 needs to bind to alpha satellite RNA to establish constitutive heterochromatin in a dicer-mediated pathway[284].

This overexpression of centromeric transcripts caused genome instability leading to the development of various cancer types[287]. In another study in mouse and human breast tumours, lacking BRCA1, which facilitates monoubiquitination of histones associated with satellite DNA, lead to increased expression of satellite RNA[288,289]. In presence of BRCA1 overexpression of satellite transcripts is supressed. Since overexpression of cenRNA shows an influence on heterochromatin formation and causes segregation defects, these might function as good predictors for cancer prognosis[289]. A recent study investigated the role of ZFAT in centromere transcription[290]. ZFAT binds to a specific DNA sequence, which is present in each centromere. ZFAT's presence at the centromere increases the level of

centromeric histone acetylase KAT2B and produces the histone mark H4 Lysine 8 acetylation. This leads to the recruitment of RNAPII through BRD4 to the centromere, therefore providing evidence for the direct control of ZFAT in centromere transcription[290].

Since Transcription plays a role in centromere integrity, some factors associated with the process have been identified. The FACT (transcription-associated chromatin remodeling factors) and Mediator complex have been shown to be located at the centromere[152,159]. In *S.cerevisiae*, FACT promotes RNAPII elongation and is important to correct ectopic CENP-A loading by mediating its proteasomal degradation. FACT weakens the histone core-DNA contact and protect the nucleosome from falling apart before the remodelling of new CENP-A has been terminated[291]. Moreover, it has been shown FACT binds to CENP-T/W and might promote their deposition as well[292]. FACT can resolve R-loops as shown in yeast and humans. This removal of R-loops prior to mitotic entry could lead to a stimulation of centromeric remodelling and transcription[293]. The Mediator complex functions as a bridge between regulatory factors and general RNAPII machinery. In *S.pombe* the components of the Mediator complex form a submodule and moderating RNAPII to mediate the transcription, ensuring precise CENP-A loading[294,295]. It's not clear if this module is conserved in higher eukaryotes.
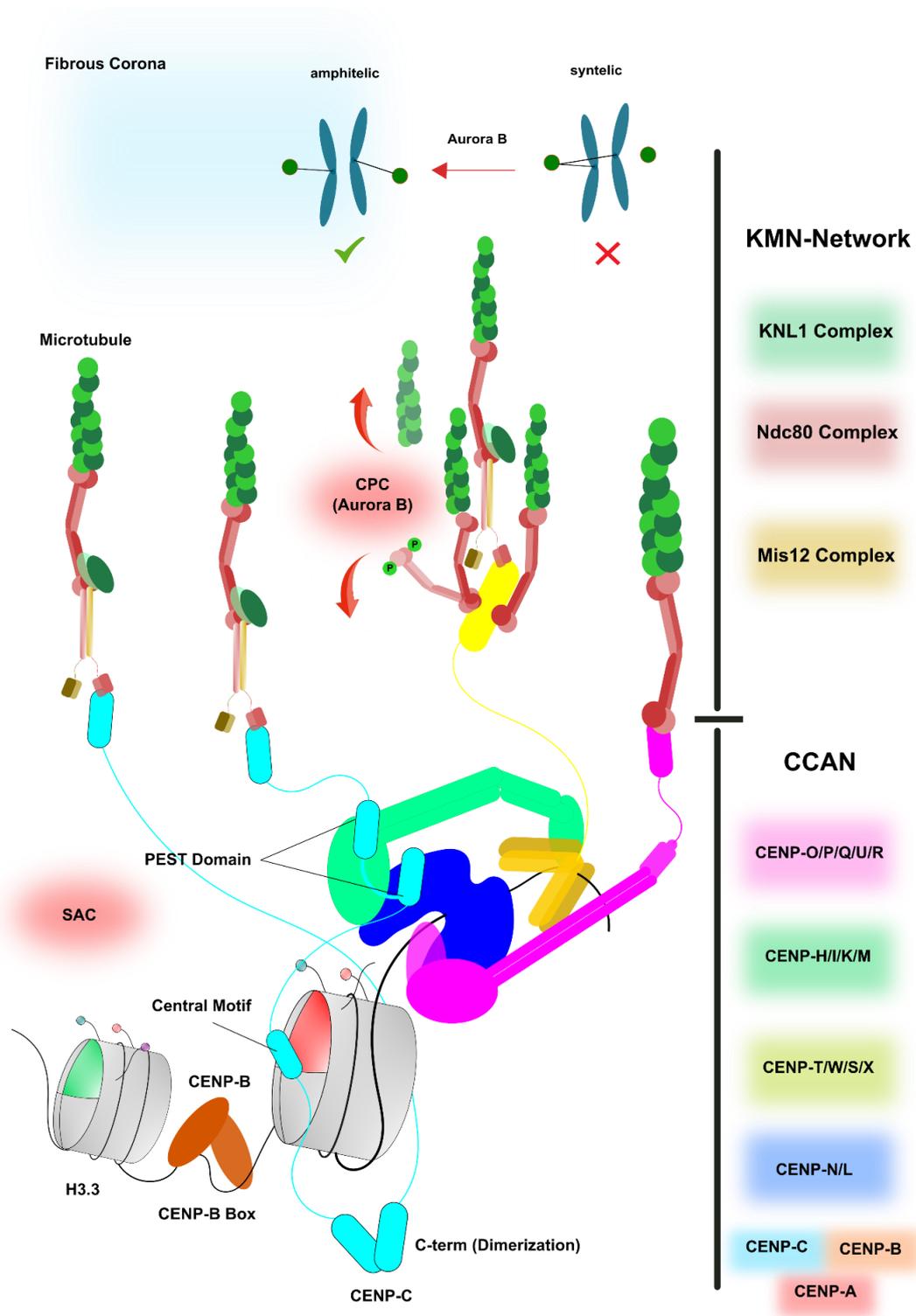
Figure 7: **Schematic representation of the human Kinetochore organization assembled on a centromeric CENP-A containing Nucleosome**. The schematic organization of the human Kinetochore is based on current knowledge in the field. For further understanding, details about each complex can be found within the text.

4.5.1.3   CENP-B as part of the Centromeric Organization

CENP-B is a 65 kDa big protein, consisting of a N-terminal DNA H-T-H (Helix-Turn-Helix) binding motif and a C-terminal dimerization domain. CENP-B is the only known CENP-protein to bind a 17 bp specific centromeric DNA sequence called CENP-B box in metazoans of which only 9 bp are essential for CENP-B binding as determined by *in vitro* assays. Moreover, CENP-B is capable to modify the DNA by inducing kinks and changes the shape of a given DNA sequence [179,296]. The CENP-B boxes are present at all centromeres in a ~340 bp periodicity except at Y-Chromosomes[297]. Through its ability to dimerize, it is believed that CENP-B might be capable of bringing two separate CENP-B Boxes at different alpha satellite monomers, therefore bringing them in proximity and influencing the higher order architecture of the centromeric region. CENP-B is localized to the centromere throughout the cell cycle[298] but little is known about its temporal regulation of loading at the centromere at a given time point[199]. It has been observed a fast turnover rate during G1- and S-Phase, while during G2- and M-Phase its stability increases. The fast turnover rate was linked to its dimerization ability. Therefore, it is believed that CENP-Bs dimerization might stabilize the centromere when CENP-A levels are halved at G2- and M-Phase to ensure Kinetochore stability and CENP-C maintenance[198,299]. The binding capabilities of CENP-B to its DNA substrate depends on the protein Nap1, where studies show that this additional factor improves the affinity to the CENP-B box, while it negatively effects CENP-Bs association with non-centromeric DNA [300]. Other proteins associated with CENP-B are INMAP and ATAC histone acetyltransferase complex and might regulate CENP-B[301,302]. Furthermore, CENP-B is post translationally modified and could have an additional impact on its regulation. The acidic domain in CENP-B has been shown to be phosphorylated *in vitro*[303] and its N-terminal Glycine is trimethylated by NRMT methyltransferase (also seen with CENP-As N-terminal tail at Glycine 1) positively influencing the binding to the CENP-B box[304]. SUMOylation on CENP-B has been linked to its stability at the centromere and prevents the interaction between DAXX and CENP-B and therefore the incorporation of H3.3. This proposes an additional role for CENP-B in heterochromatin architecture and pericentromeric formation[305,306]. Alteration of CENP-Bs binding capabilities towards the DNA is influence by DNA methylation marks. The CENP-B box contains two CpG methylation sites and in vitro studies showed decreased affinity of CENP-B towards methylated CENP-B box[200]. Recent studies proposed an inhibiting role of CENP-B in RNAi-dependent heterochromatin silencing, hence conferring it with a role of negative transcriptional and epigenetic modulator role[199,200]. Besides the binding of CENP-

B to the CENP-B box, it binds in addition to the N-terminal region of CENP-A and to CENP-C by its acidic domain[198,307]. The interaction towards CENP-A stabilizes the CENP-A Nucleosome and promotes incorporation of CENP-B into the Centromeric Chromatin, since *in vivo* depletion experiments of CENP-A's N-terminal domain led to decreased CENP-B levels[198]. Similarly, depletion of the acidic domain of CENP-B led to altered CENP-C assembly at the centromere as well as cell cycle delay. When CENP-B was completely depleted, the CENP-C level decreased 2-fold at the centromere [198,307]. Although CENP-Bs role in mice has been shown to be redundant, in human CENP-B shows that high levels of it at centromeres contribute to less mis-segregation events upon rapid removal of CENP-A[199]. Moreover, CENP-B has been implicated in replication initiation and stability of replication fork at centromeric sites[308]. As elevated in previous paragraphs, the formation of heterochromatin marks at the pericentric region is important for the stability of the centromere and subsequent chromosome segregation. CENP-B might play a role in the formation of this specific chromatin organization.

A recent study[309] provided an insight on how CENP-B is capable to mediate/alter between the pericentromeric heterochromatin mark and supresses its invasion into centromeric regions. The investigators set up a screen with tetR-EYFP-CENP-B as a bait protein to find specific partners of CENP-B at the centromere. Besides finding CENP-A and CENP-C, they found Suv39h1, a histone methyltransferase responsible for heterochromatin mark formation. Moreover, other factors such as ASH1L, NSD1, KMT2A, KDM4D, KDM68, KDM7A, PHF2, DAXX, Mis18BP1 and HP1 were identified. The authors chose to further investigate ASH1L and HP1 role in the binding to CENP-B. Both proteins bind at the same acidic region in CENP-B (aa 403-556). Moreover, the acidic domain was sufficient to recruit the other mentioned factors, hence leading to an important role for this domain on the overall centromeric chromatin formation. Assays showed ASH1L might act on CENP-B containing Chromatin by supressing excessive assembly of HP1, which maintains heterochromatin formation. Similarly, HP1 has been shown to act in a similar manner to supress the influence of ASH1L on chromatin. The authors conclude through CENP-Bs acidic domain ASH1L and/ or HP1 are recruited in a mutually exclusive fashion to 'bidirectionally' generate an alternative epigenetic chromatin, whose balance maintain centromere function. ASH1L function and localization at centromeres were not previously reported. ASH1L is known to methylate H3K36 and mutually exclusive with polycomb group-mediated H3K27-methylated facultative heterochromatin[310,311]. CENP-B might be able to alter chromatin formation by recruiting other centromere-related proteins to replace histones by e.g. H3.3, which have

been shown to be a placeholder or CENP-A replenishment but also important for heterochromatin maintenance[282,312,313]. Therefore, CENP-B can switch actively depending on the presence or absence of functional centromere. Since H3.3 localizes to transcriptionally active chromatin and centromeric transcription have been shown to maintain centromere integrity and histone marks such as H3K27 methylation inhibits the transcription process leading to CENP-A impaired assembly. Hence, transcriptional marks might provide further direction of CENP-B to "switch" between chromatin states. The authors conclude with a model, where binding of CENP-B's acidic domain with ASH1L promote CENP-A incorporation, whereas binding of HP1 promotes H3.3 incorporation and trimethylation at K9 by Suv39h1 and therefore the alternative chromatin state can be switched by CENP-B[309].

Another recent publication[314] shed light on how CENP-B can contribute to higher order chromatin architecture by performing DNA loops. CENP-B originates from the domestication of pogo-like transposases. It is the only known centromeric DNA-sequence-specific binding protein known till date, which associated with the alpha-satellite of all centromeres except at the Y-Chromosome. The centromere containing AT-rich DNA is known to have diverse structures, ranging from DNA dyad symmetries to non-B form DNA such as hairpins, Z-DNA or intercalated motifs. However, these secondary structures are not predicted in centromeres containing the CENP-B boxes. Therefore, this provides an implication, that CENP-B itself could promote secondary structures once it is bound to its target sequence. Since CENP-B is also able to introduce DNA-bending (kinks), leading to a translational position of the nucleosome, the positioning of centromeric nucleosomes and the overall chromatin fibre architecture might be influenced by CENP-B. By applying optical tweezer methods, the human centromeric DNA could be shown to self-organizes into non B-form secondary DNA structures. These preferentially form local double-stranded hairpins and complex secondary structures. It could be ruled out that the addition of Mg^{2+} ions present in the buffer induced the hairpin accumulation. The binding of CENP-B to the CENP-B box reshapes the centromere and forms submicron sized DNA loops between repeats. This leads to compacting and clustering of centromeric chromatin. In living cells, Chardon *et al.* (2022), showed the importance of these specific features, since these loopings preserve centromere positioning and DNA integrity. Upon binding to the CENP-B box, CENP-B and the DNA form a tetrameric complex, where a homodimeric CENP-B molecule binds two molecules of DNA. Further investigation showed the formation of compact centromeric DNA in secondary structures of defined size of 350-550 bp sized DNA loops.

The dimerization domain of CENP-B plays a crucial role, since a mutant unable to dimerize reduced the loop formation to 17-23%, whereas the full-length protein produced loops from 36-65%. Next the investigators examined the role of CENP-B *in vivo*. Here, depletion experiments with the auxin system showed, when CENP-A and CENP-C are depleted and subsequently activated by IAA removal, reloading of CENP-C failed, when the cells had downregulated CENP-B levels. The levels could be restored by induction of doxycycline inducible CENP-B wildtype but not in the case of CENP-B dimerization mutant. Therefore, the dimerization domain maintains centromere by promoting CENP-C association and aids to deposit CENP-A. In another experiment CENP-B has been shown to create a certain distance between sister chromatids. This observation seems to be significant, since disruption leads to centromere fragility due to chromosome breakage or decondensation of centromeric regions, which are further supported by the accumulation of DNA-damage marker yH2AX in mitosis and protein 35BP1 following in G1- Phase. When CENP-B was overexpressed, the accumulation of 35BP1 could be rescued, whereas the dimerization mutant was less efficient. Interestingly, CENP-B dimerization mutant was still able to compact the chromatin to a certain degree, therefore it could be that other domains of CENP-B are likely involved in heterochromatin formation and/ or transcriptional regulation. Taken all results together, CENP-B facilitates CENP-A deposition by forming these specific structures at the centromere and bridges centromere binding proteins (CENP-C and CENP-N) to the CENP-A Nucleosome. These specific structures also might be important for HJURP, which recognizes cruciform structures built by CENP-B. Another structure at the centromere has been observed: RNA-DNA triplexes such as R-loops, which were associated with ATR kinases, safeguarding the genome stability. The possibility exists, where these secondary structure formation by CENP-B and other centromere proteins lead to its organization[314].

Taken together current findings suggest a more prominent role of CENP-B at the centromeric chromatin organization. Future studies have to address the importance of CENP-B at the human centromere region and what structural features does CENP-B inherit to maintain or alter the (peri)centromeric chromatin.

4.5.1.4   The corner stone of the CCAN formation: CENP-C

The first CENP proteins binding to the CENP-A Nucleosome are CENP-C and CENP-N by a discrete binding fashion. The CENP-A Nucleosome can be seen as the root of a tree where the tree trunk is built upon CENP-C harbouring binding sites for several Kinetochore

components, serving as a 'blueprint' for CCAN assembly and organization. CENP-C is an elongated protein with several unstructured regions and two nucleosome binding domains, a central domain and the CENP-C motif followed at the C-terminus with a cupin fold, known for dimerization. The CENP-C motif is conversed from *S.cerevisiae*. In contrast, the central domain is only conserved in mammals and is characterized by a high specificity for CENP-A over H3, when compared to the less specific CENP-C motif inherited from previous ancestors[165,315]. The speciality of the central domain has been directly linked to DNA replication process, where CENP-C directs structural transition in the CENP-A Nucleosome and stabilizes it [316–318]. Within the last two years, incredible biochemical and structural biological insights could have been gained in understanding the assembly mechanism of the CCAN onto CENP-A containing Nucleosomes. A recent study[319] provided a closer look at CENP-Cs contribution to the very inner Kinetochore axis. Since CENP-C is intrinsically disordered and hard to express and purify recombinantly, the investigators used a SpyTag/SpyCatcher method, to covalently fuse two recombinantly expressed and purified polypeptide chains through the formation of an isopeptide bond and linking therefore CENP-C[1-600] with CENP-C[601-943]. After obtaining a functional fusion construct of full-length CENP-C, they subjected fused CENP-C for several biochemical assays to assess if it behaves as the wildtype form. Indeed, it showed CENP-A Nucleosome, CCAN and outer Kinetochore binding abilities. Next, the investigators were interested in the question, if human CENP-C would adopt a similar conformation as seen with S.*cerevisiae* ortholog Mif2, where Mif2 adopts an autoinhibited conformation at the N-terminal region, which supresses the binding of ortholog Mtw1 complex (in human Mis12 complex) before it binds to Cse4 (CENP-A) Nucleosomes. The investigators could not determine such a conformation for the human ortholog since binding of CENP-C to Mis12 complex or Ndc80 complex were not altered and additional phosphorylation by CDK1 did not show any phenotype in this direction. Therefore, it is assumed, that this mechanism observed in buddying yeast might be special to the point centromere. CENP-C has a dimerization domain at the C-terminal part of the protein and the fused CENP-C product showed stable formation of a dimer as judged by AUC and SEC analysis. Through its ability to form dimers, CENP-C is capable to bind at least two CENP-A Nucleosomes and simultaneously provide further binding sites for other Kinetochore members, thus stabilizing the CCAN at the centromeric site[319]. Another study biochemically and structurally investigated the behaviour of phosphorylated CENP-C central motif on CENP-A Nucleosomes[320]. By phosphorylating chicken CENP-C fragment [601-846] with CDK1 they observed a tighter binding to the CENP-A Nucleosome. Moreover,

they found out that extended regions around the central motif are required to stabilize the binding to CENP-A Nucleosomes but also specifies this region for binding only CENP-A. Furthermore, structural details highlighted how the interaction is coordinated by three elements of the CENP-A Nucleosome – C-terminal tail, RG loop in CENP-A as well as the acidic patch of H2A/H2B –. In addition, their structure could provide an explanation how CENP-C and CENP-N might bind at the same region without steric clashing, where they show that other studies used to short constructs to see a reach out to the RG loop of CENP-A. Since the used construct stretch is well conserved between chicken and human (including CDK1 phosphorylation site) it is likely, that this might reflect how the binding mode of the central motif of CENP-C in respect to CENP-N binding might be[319].

4.5.1.5 Building up the CCAN: CENP-N/L, CENP-H/I/K/M and CENP-O/P/Q/U/R

CENP-N/L complex

The N-terminal portion of CENP-N binds CENP-A by its CATD domain (centromere-targeting domain) at the flexible L1 loop and the α2 helix. The binding tights CENP-A to the nucleosomal DNA and stabilizes CENP-As surface bulge which might add additional stability to the centromere chromatin[165]. In addition, the CATD region is required as an interaction hub for HJURP deposition of CENP-A into centromeric chromatin(ref135,139)[166]. A recent study by Keda Zhou *et.al* (Nature structural & molecular biology, 2022)[321] shed light on how CENP-N might interact with CENP-A in a specific manner (Figure 4 C, right panel). The construct of CENP-N spanning form residue 1 to 289 promoted with centromeric CENP-A Nucleosome a stacking like behaviour. About 30% of the particles on a cryo-EM grid showed stacking ranging from 2 to 10 Nucleosomes. The same behaviour was seen with CENP-A Nucleosomes wrapped around a Widom601 sequence. To exclude cryo-EM artefacts, sedimentation-velocity analytical ultracentrifugation experiments were conducted. It showed that single CENP-A Nucleosomes behaved as canonical H3 containing Nucleosomes. But when incubated with CENP-N$^{1-289}$, the sedimentation coefficient was determined between 13S to 30S, similar values seen for dinucleosomes or 12mer nucleosome arrays. The same experiment was conducted with H3 Nucleosomes but there were no such species observed. In addition to exclude any artefacts of the used construct, the complete CENP-N/L module was used and showed the same outcome, therefore promotes CENP-A Nucleosome oligomerization. Another interesting observation was the dependency of salt concentration used to observe stacking behaviour. When the experiments were conducted with 200 mM salt, the stacking could be observed but with

increasing salt concentration the stacking disappeared although CENP-N remained bound to the CENP-A Nucleosome. Another aspect of this study was, if CENP-N can promote the stacking of CENP-A and H3 Nucleosomes. Therefore, they applied FRET experiments and obtained signals for this theory. Moreover, CENP-N seem to promote similar structural chromatin features as H1 linker histone but using a different mechanism[321]. Besides this interesting study, another study[322] promoted a big contribution on how the human interphase CCAN assembles onto a centromeric CENP-A Nucleosome (Figure 4 C, left panel). Structural details of the assembled CCAN with and without CENP-A Nucleosome revealed interesting insights, on how the CCAN interacts with each other, promote a stable conformation of the subunits and how they interact with the alpha satellite DNA. As a short notice aside, the presented structure might show how the interphase CCAN assembles onto the centromeric CENP-A Nucleosome. The structure does not include any mitotic phosphorylation states or any known modifications on the CENP-A Nucleosome. Similarly, another study[323] published a cryo-EM structure of the human interphase CCAN but without a CENP-A Nucleosome. Both studies had similar conclusion on how the CCAN assembles. As already described in previous studies, when CENP-A wraps around an alpha-satellite DNA, the DNA termini are flexible and unwarp about 22 bp at both ends mostly[324–326]. When assembly the CCAN onto the centromeric CENP-A Nucleosome the authors observed how a 24 bp DNA duplex channels through the CENP-N/L complex[322]. Here the first time it is reported that the contact is made by the extranucleosomal DNA to the inner axis of the CCAN. Since they did not include full length CENP-C into the structure, the structural features just mentioned for CENP-C could not be structurally determined here. Nonetheless, the DNA phosphate backbone was tightly gripped by the conserved basic residues in the CENP-N/L channel. The tight grip to the DNA is further extended by the raised head conformation of the CENP-H/I/K/M module[322].

It has to be mentioned, that the authors did not mention any stacking formation of CENP-N to the CENP-A Nucleosome used. By comparing both Cryo-EM structure preparations it became evident why this phenomenon could not be seen. Since stacking of CENP-N in the previous study[321] can only be observed at ≤ 200 mM salt concentration, the current structure of CCAN bound to centromeric CENP-A were conducted at 300 mM salt conditions, therefore the formation was perturbed. Taken both observations together, it could be, that maybe at similar salt concentrations a similar effect could have been seen or upon binding of CENP-N/L to the CCAN, its structural behaviour changes completely, which have to be addressed in further experiments.

CENP-H/I/K/M complex

The recruitment of the tetrameric CENP-H/I/K/M complex is mediated by CENP-C and CENP-N/L, which have been prebound to the centromeric nucleosomes[327]. It has been shown, that CENP-H/I/K/M contributes to the CENP-A nucleosome binding affinity of CENP-C and CENP-N/L but as a single unit it is not selective for CENP-A nucleosomes compared to canonical H3 Nucleosomes [328]. The subunit member CENP-M has been shown to contain a pseudo-GTPase unit, which is required to stabilize the binding to CENP-I and is essential for the stability of the entire complex. *In vivo* studies have shown that CENP-M:CENP-I interaction is disrupted, it leads to defective kinetochore assembly and chromosome alignment[329]. In an initial study[322], the investigators were interested in how CENP-C would fit in the CENP-N/L and -H/I/K/M binding, they tried to determine the structural features of CENP-C [PEST domain] upon binding to CENP-N/L and CENP-H/I/K/M. Former obtained biochemical data shows that his region is required to bind both modules. In their obtained structure, they can see an additional density but cannot assign any structural features. Therefore, they applied AlphaFold predictions to model how the PEST domain interacts with each module individually[322].

AlphaFold is a computational method capable predicting the 3D-protein structure of any given protein sequence near experimental accuracy in the majority of cases. This computational prediction by a neural network revolutionized the structural biological field. It does not do a template bases search, rather predicts the 3D coordinates of all heavy atoms given in a protein primary sequence and aligned sequences of homologues as an input. After training the network, predicted structures of different scores can be viewed. By comparing the prediction with the actual experimental dataset, the majority of the predicted structure fit near perfectly into the experimental determined one. Therefore, AlphaFold can provide valuable insights in a protein's domain architecture but also intrinsically disordered regions, allowing to study the protein of interest without doing actual structural experiments in the first place[330,331].

After obtaining the AlphaFold prediction for the PEST-Domain binding to CENP-N/L module they found the conserved DEFxIDE motif (aa 301-307) would from an edge ß-strand with the CENP-N ß-sheet, corresponding to the observed density and the amino acid residues Phe and Ile within this conserved motif dock into a hydrophobic pocket of the CENP-N/L interface, whereas these are flanked by acidic residues forming an electrostatic interaction

with the conserved Arg and Lys residues of the CENP-N/L module[322]. In agreement with previous studies, mutation in EFxID motif abolishes the interaction with CENP-N/L *in vitro* and shows *in vivo* impaired recruitment of CENP-N to the centromere, without perturbing the localization of CENP-C at the centromere[332]. Similarly, the FxxLFL motif (aa 262-267) is known to bind CENP-H/I/K/M. Here the prediction shows an alpha helix binding to a hydrophobic site in the combined CENP-H/K/M interface, which corresponds to the observed extra density. Previous studies investigating the binding to this motif described that the mutation of LFL residues resulted in loss of CENP-H/I/K/M *in vitro* and *in vivo*[333]. Moreover, the overall structure of the CENP-N/L and CENP-H/I/K/M module back-faces the CENP-A Nucleosome, suggesting how CENP-C might tethers the Kinetochore subcomplexes to the Nucleosome. In additional assays it became clear, that the extranucleosomal DNA is required for stable CCAN association with the CENP-A Nucleosome, since shorter construct resulted in impaired CCAN-CENP-A Nucleosome association. Comparing the results with previous structural studies of CENP-N alone with CENP-A Nucleosome, they found out, that the CENP-N/L module behaves differently when it is associated with the complete CCAN. As a standalone module, it has been shown to interact with the L1 loop of CENP-A and the adjacent DNA gyre but when it is in complex with the other members, this interaction could not be observed, rather it was linked to the extranucleosomal DNA and the interaction to the L1 loop of CENP-A was not established[322,334].

CENP-T/W/S/X complex

Another complex, the tetrameric CENP-T/W/S/X complex has been shown to be part of this Mega Dalton Kinetochore complex too. It consists of two-unit CENP-T/W and CENP-S/X and each component contains a histone fold domain[327,335]. The turnover of CENP-T/W happens frequently throughout the cell cycle, and its incorporation into the centromere occurs at late S-Phase and G2-Phase[336]. The N-terminal tail of CENP-A and the CENP-H/I/K/M complex tethers the recruitment of CENP-T to the Kinetochore[165,337]. CENP-S/X does not seem to be essential, since in vertebrates CENP-T/W are more important[338,339]. Since structural data of CENP protein complexes are rare, it has been proposed that CENP-T/W/S/X form a tetrameric structure and organizes the DNA in a nucleosome-like complex[166]. The ability to bind DNA relies on the histone fold domains of each protein member, but in case of CENP-T and -W it does not exhibit sequence specificity. Moreover, the binding of CENP-T/W/S/X induces a positive DNA supercoiling[340.] Another interesting

observation was made, when ablation of CENP-T/W or CENP-S/X lead to distinct effects on the outer kinetochore stability [338,339]. It seems that CENP-T/W/S/X might have a dual role at the kinetochore. In a very recent publication, the investigators addressed in their publication the role of CENP-T/W/S/X in the context to centromeric CENP-A Nucleosome binding[322]. The CENP-T/W/S/X module interacts as well with the extranucleosomal alpha satellite. In contrast to the structural findings, the module does not change the fundamental mechanism of the CCAN assembly. Electromobility Shift Assays revealed the module increases the overall affinity of the CCAN towards the Nucleosome but does not discriminate between CENP-A or H3 containing Nucleosomes. The histone folds present in each of the subunit members resemble in their structural feature as histone H3-H4 tetramer, which forms multiple interaction with neighbouring CCAN subunits and the DNA. Upon binding of the module, the H/I/K/M module gets slightly repositioned mediated by the CENP-T/W subcomplex and CENP-W contacts CENP-N. The extranucleosomal DNA gets topologically entrap upon binding of the CENP-T/W/S/X module, since they mediate the linkage of the H/I/K to CENP-N and CENP-Q/U forming an enclosed chamber. The extrusion of extranucleosomal DNA through the CENP-N/L channel is taken up by the CENP-T/W/S/X module, where the DNA wraps around the CENP-T/W histone fold domain, similar to H3-H4 wrapping mechanism, where the binding is augmented by the CENP-H/I/K/M module[322]. In previous studies the CENP-T member was proposed to have an additional role in recruiting the KMN network through a distinct pathway[341]. This additional function has been investigated in chicken DT40 cells and human RPE-1 cells[342]. They showed that in chicken DT40 cells the CENP-T-K/M/N recruitment is essential for cell viability, emphasizing that this could be crucial in chicken. CENP-T has been shown to bind one KMN complex and an additional copy of Ndc80 directly making up to 3 copies in total bound by CENP-T. Moreover, they show that CENP-T is crucial for KMN assembly, since its flexible N-terminal region binds two distinct Ndc80 molecules through a motif encompassing CDK1 phosphorylation sites (T11, T85, S201). In combination with another CDK1 phosphorylation at T195 mediates the binding an entire KMN complex through binding to Mis12 complex[342]. In contrast, CENP-C has been shown to bind only through Mis12 complex and therefore recruits only once the KMN at this site[343]. An interesting finding was the binding of CENP-T to the Mis12 complex but in future studies the regulatory mechanism of Mis12 complex recruitment onto the CENP-T pathway have to be investigated.

Taken together, the CENP-T/W/S/X module acts at both extreme ends: a) at the centromeric region, where it interacts with the extranucleosomal DNA of the CENP-A Nucleosome and

b) in the recruitment of the KMN network to the Centromere and thus reveals its multifaceted function at the centromere.

CENP-O/P/Q/U/R complex

As another important connector between CCAN and KMN network seems to be the five-subunit CENP-O/P/Q/U/R complex[344,345]. The recruitment is mediated by the CENP-C-CENP-N/L-CENP-H/I/K/M complex, preassembled on CENP-A Nucleosomes[166]. The binding occurs at a joint interface of CENP-N/L and CENP-H/I/K/M. The five units can be divided into three standalone subcomplexes: CENP-O/P, CENP-Q/U and CENP-R [346]. The recruitment of CENP-O/P/Q/U is interdependent and can localize to the Kinetochore without the need of CENP-R. CENP-R is interacting mainly with CENP-Q/U subcomplex and *in vivo* data showed that CENP-R might not be important for spindle damage recovery, as seen for the other components [346]. As seen in the study[322], the architecture of the CCAN resembles the buddying yeast one reported[325,347]. The CENP-O/P/Q/U/R module assembles on one flank, whereas the CENP-H/I/K/M lope assembles at the other flank of the central CENP-N/L module. The channel formed by CENP-N/L is further extended by one side with the N-terminal coiled coil domain of CENP-Q/U. In conjunction with the RWD domains of CENP-O/P, the C-terminal region of CENP-Q/U and with CENP-R, these structural features form into a cap domain above CENP-N. Moreover, a flexible linker within the CENP-O/P/Q/U/R module links the N-terminal alpha helix of CENP-O with the module but also interacts with a conserved pocket build by CENP-I/K. Upon depletion of the N-terminal 35 residues leads to destabilization of the entire CCAN. As the investigators studied further the role of CENP-O/P/Q/U/R in the context of CCAN assembly by *in vitro* assays, they concluded that the module contributes very little to the attachment of vertebrate kinetochores to the Centromere. These findings suggest that the module might have an additional functional task. Indeed, a previous study[346] and links the CENP-O/P/Q/U/R module to the binding of Ndc80 complex, which belongs to the outer layer of the Kinetochore, the KMN network. Specifically, CENP-Q plays an important role in binding to the KMN. By performing microtubule binding assays, they distinguished that the CENP-Q/U subcomplex is responsible for decorating microtubules, whereas the CENP-O/P does not have such a function and is more likely carried along by CENP-Q/U. The Ndc80 complex act at hight concentration cooperatively with the microtubule lattice and interacts along the protofilaments by alternating αβ and βα tubulin interface. At 100 nM Ndc80 decorated the microtubule lattice. By titrating in 400 nM CENP-O/P/Q/U, Ndc80 complex got removed and

the module decorated the microtubule lattice. The experiment was performed in a reverse order and when titrated in Ndc80 (400 nM) the same could be observed, where CENP-O/P/Q/U was displaced by Ndc80 complex. The authors found out that CENP-Q has an N-terminal 67 residue stretch of basic nature, which is required for binding to microtubules, explaining partially its competitive behaviour towards Ndc80 microtubule binding. *In vivo* studies showed strong accumulation of chromosomes near spindle poles, when CENP-Q was depleted, hinting towards congression errors. Moreover, this phenotype showed reduced levels of CENP-E, an additional factor in the fibrous corona. The phenotype could be rescued upon overexpression of CENP-Q to a wildtype scenario. Further studies showed that CENP-O/P/Q/U/R is required for chromosome alignment and microtubule interaction. The authors found two major differences in the binding mode of CENP-Q in comparison to Ndc80 towards the microtubules. The interaction in Ndc80 tail is accompanied by the presence of the calponin homology (CH) domain, contributing to microtubule binding. In contrast, CENP-Q/U do not have such a domain. Furthermore, the second difference is the regulation of Ndc80's tail by phosphorylation events mediated by Aurora A, Aurora B or CKD1, whereas CENP-Q's tail might have two possible phosphorylation sites for Aurora Kinases but not for CDK1[346].

In summary, the outermost CCAN module, CENP-O/P/Q/U/R is recruited by CENP-N/L and CENP-H/I/K/M and links the CCAN to the KMN network.

4.5.2 The KMN network connects the CCAN to the Microtubules

4.5.2.1 The KMN member Ndc80 links to the Microtubules

The primary microtubule receptor at the Kinetochore is the four-subunit Ndc80 complex, consisting of two heterodimers Ndc80, Nuf2 and Spc24 and Spc25 (Ndc80:Nuf2, Spc24:Spc25)[166]. The microtubule binding is mediated through a N-terminal globular domain as well as a C-terminal coiled coil domain of Ndc80:Nuf2. The basic N-terminal tail contains a calponin homology domain, facilitating the interaction with the acidic E-hook on both alpha- and beta-Tubulin. The calponin homology domain is required for its high-affinity binding of the complex to the microtubules but it has been shown that this domain does not provide a direct contact to the tubulin monomers [348–350]. The heterodimer Spc24:Spc25 complex consists of a C-terminal globular region containing paired RWD domains, as well as a N-terminal coiled coil domain. Each protein pair interacts via their coiled coil domain through the process of dimerization and forming the tetrameric complex[165]. The overall

dumbbell-like shape characterizes the Ndc80 complex where the outer parts have globular domains and the interior the dimerization by the coiled coil domains. Due to its elongated shape of Ndc80, a full complex formation could not be determined structurally, but modified constructs lead to the '*bonsai*' and the '*dwarf*' form of the complex, enabling important insights on how this complex executes its function as a part of the outer kinetochore[351]. One of the inner Kinetochore components, the N-terminal disordered part of CENP-T, has been shown to interact with the RWD domains of Spc24:Spc25 allowing to recruit two Ndc80 complexes for each copy of CENP-T[341,352,353].

4.5.2.2 The Mis12 complex as additional stabilizer for correct CCAN-Microtubule attachment

Another component of the KMN network is the four-subunit Mis12 complex, consisting of the proteins Dsn1, Nsl1, Mis12 and Pmf1, which come together as subcomplexes as Dsn1:Nsl1 and Mis12:Pmf1[343,354,355]. It has been shown that Mis12 complex interacts with CENP-C and CENP-T, hence bridging the KMN with the CCAN. Furthermore, the complex contains binding motifs for Ndc80 complex and Knl1 to form the KMN network. As previously described, CENP-T can bind two Ndc80 complexes. Now CENP-T is capable of binding Mis12, which in conclusion provides an additional Ndc80 molecule to bind through the Mis12 complex, making in total three Ndc80 complexes bound only by CENP-T[356,357]. Since also CENP-C with its N-terminal domain can bind the Mis12 complex, it has been proposed that with this pathway two Mis12 complexes can be recruited to a single CCAN[358]. In theory, one CCAN harbouring one CENP-T and one CENP-C sites (other side binding to CENP- N/L, CENP-H/I/K/M) can bind up to four Ndc80 complexes[165]. The N- and C-termini of all four subunits cluster at opposite ends of the rod and at the C-termini the linear motif of Nsl1:Dsn1 subunits provide the binding interface for the RWD domain in the Spc24:Spc25 subcomplex of the Ndc80 complex[359]. The Mi12 complex looks like a Y-rod-shaped complex with high helical content and linear motifs[343,359].

4.5.2.3 The KNL1 complex as additional proof-reader for correct Kinetochore-Microtubule Attachment

The KMN network is completed when the KNL1 complex, composed of Knl1 and ZWINT, joins the network. Knl1 is the largest subunit in the outer kinetochore sphere with 2316 amino acid residues in human. Knl1 is largely intrinsically disordered, except the last ~500 residues[165,359]. Moreover, the protein contains an array of protein docking motifs such as in

the N-terminus a canonical PP1 phosphatase binding site or multiple MELT (Met-Glu-Leu-Thr) repeats that after phosphorylation serve as docking sites for the SAC (Spindle Assembly Checkpoint) protein complex (Bub1:Bub3)[166]. The C-terminal part of Knl1 contains coiled-coils and tandem RWD domains, where the latter one function as interaction domain for Mis12 complex binding. ZWINT binds at an extended domain, comprising of coiled-coil regions. ZWINT has a central role in recruiting the SAC to ensure proper chromosome alignment and correct KT-MT attachment before mitosis onset[360–363].

4.5.3 The regulation of the Kinetochore-Microtubule Axis

Chromosome loss or gain at the end of mitosis is linked to erroneous Kinetochore-Microtubule attachments. These defects lead to aneuploid or micronuclei formation. The attachment state of chromosomes is monitored by SAC before anaphase onset[364]. The miotic progression is delayed as long as the incorrect attached Kinetochores are not resolved and correctly linked to spindle microtubules. If the correct KT-MT attachment have been created, the SAC is silenced, and chromosome segregation can be executed. Notably, the SAC cannot discriminate between correct and erroneous KT-MT attachment[365]. In humans, the microtubule turnover fastens during prometaphase and decreases in metaphase, modulated by CDK1:CyclinA complex levels. Following the degradation profile of this complex, reaching its peak in metaphase, it might contribute to indiscriminate regulation of KT-MT end-on attachments by maintaining this interaction in relatively unstable state. Furthermore, by applying gradual tension throughout the different mitotic states, the correct attachment (amphitelic) can be achieved and favourably stabilized[366].

The binding of microtubules through the Ndc80 complex only partially depicts the true situation at this interface. Many other proteins further mediate or physically connect or support the interaction. One of them is the Ska complex, consisting of the trimer Ska1, Ska2 and Ska3 forming a W-like shape formed from dimer association of helical bundles[367–369]. The C-terminal winged-helix domain in Ska1 interacts with tubulin and tracks both polymerizing and depolymerizing microtubules. Ska3's unstructured C-terminal region also facilitates the interaction between the complex and microtubules[370–372].

During mitosis an extended fibrous structure, known as the fibrous corona, radiates from the distal surface of the outer kinetochore plate without attached microtubules. Its morphology changes upon attaching kinetochore to microtubules during mitosis[373,374]. Proteins associated with this fibrous corona are Dynein, Kinesin, CENP-E, CENP-F, Rod-

Zw10-Zwilich complex (RZZ) and SAC proteins. The localization of RZZ depends on Knl1, Bub1 and Spindly. Depletion of Spindly or RZZ components resulted in defects in SAC signalling and prevents the formation of extended outer kinetochore suggesting a role in kinetochore expansion[165,366].

In order to ensure correct chromosome segregation, the correct attachment of microtubules to the kinetochore has to be carefully monitored. One of the surveillance systems is the CPC (Chromosome passenger complex)[165]. CPCs function correlates with the mitotic progression, where its localization alters dynamically from chromosome arms to spindle midzone via centromeres. The spatiotemporal activity of CPC lies in its localization during the cell division. First it is recruited to the kinetochore and centromeres in early mitosis and gradually shifts to an occupation state at the centromere at late prometa- and metaphase. The Aurora B kinase, a key regulator and catalytic subunit of the CPC phosphorylates sites at the N-terminal tail of Ndc80, leading to a neutralizing of its positive charge and thus decreasing its microtubule affinity. This way, a post-translational modification, here phosphorylation, serves as a negative regulator for kinetochore-microtubule attachment[373,375–378].

Another kinase, Aurora A, involved in spindle pole separation in early mitosis, shows also similar regulatory mechanisms by also phosphorylating sites in the N-terminal tail of Ndc80[379–381]. If the precise bi-polar attachment has been achieved, the Aurora B substrates are physically separated from the Aurora B activity gradient and are dephosphorylated for instance by protein phosphatase 1 (PP1)[382]. This dephosphorylation states is kept upon mitotic progression. Another mitotic kinase, Mps1, is a critical orchestrator of SAC signalling and is implicated in error-correction by phosphorylating key players involved in this process. Mps1 can directly phosphorylate Ndc80 to waken KT-MT interaction independently of Aurora B[382–386].

Moreover, another important kinase, PLK1 shows a positive and negative regulatory function towards KT-MT attachment[387]. How this is executed by PLK1 remains elusive. PLK1 is a crucial mitotic regulator with diverse functions like timely mitotic entry, centrosome maturation, mitotic spindle assembly, cohesion dynamics, KT-MT attachments, and cytokinesis[366].

Taken all together, the cell has found sophisticated mechanisms to ensure correct Kinetochore-Microtubule attachment and established feed-back mechanisms and

surveillance checkpoints to ensure correct chromosome segregation and subsequently genetic integrity to the next daughter cell.

4.6 The novel Transcription factor BTB/POZ zinc finger protein family (ZBTB)

The work of Jacob and Monod in 1961 paved the way of key principles in gene regulation. They postulated the existence of *trans*-acting factors controlling gene transcription by binding to specific DNA elements near genes[388]. In human cells, there are more than 100 so called DNA-binding transcription factors known, which modulate gene expression by influencing RNA polymerase II activity and most of these factors consists of distinct structural domains, necessary for their regulatory function. One domain is required for sequence-specific DNA-binding (e.g. leucine zipper, helix-loop-helix or zinc finger) and additional another domain required for transcription activation/ repression, which binds other cofactors enhancing/ decreasing transcriptional activation or repression. Such cofactors can directly activate or inhibit RNA polymerase II or influence local chromatin architecture to allow increased/ decreased gene transcription rate. Some of these transcription factors have dimerization domains in common[388].

One class of transcription factors are the Broad-complex, Tramtrack and Bric-à-brac (BTB) zinc-finger-domain-containing protein (ZBTB) family members. Each of the family members contain an N-terminal residing BTB domain, known for dimerization/ oligomerization[389] and mediating protein-protein-interaction, whereas at the C-terminal zinc fingers (ZFs, present in ~25% of all BTB-domain-containing proteins) often with the C2H2 type are found responsible for binding to specific DNA elements[390,391]. Both domains are joined by a flexible linker modulating the stability and versatile interaction mechanism of ZBTB protein complexes (Figure 8) [392,393]. Since zinc finger motifs exhibit less sequence conservation, they emerged as a specific class of DNA-binding proteins capable of binding unique DNA sequences. Moreover, it is believed that BTB with zinc fingers acquired their additional role as a transcription factor in higher eukaryotes in a late evolutionary phase[391,394].

There are 20 different zinc finger domain types known and categorized by their structure of their zinc ion stabilized by specific amino acids; the most common type is the C2H2 type (Cys2-His2). This type is mostly referred as the "classic" zinc finger, containing the consensus sequence $(F/Y)-x-C-x_{2-5}-C-x-(F/Y)-x_5-\psi-x_2-H-x_{3-4}-H$, where X can be any amino

acid and ψ represents any hydrophobic residue. This motif self-folds into a *ββα* structure, where the zinc ion is coordinated by two conserved cysteine and histidine residues[395,396]. Originally, C2H2 type zinc fingers were identified in *Xenopus laevis* TFIIIA transcription factor and it is estimated, that approximately 3% of genes in human encode for C2H2 proteins, marking them the second most prevalent protein motif within the human genome[397,398]. The zinc fingers are capable of binding to DNA or RNA but also mediating protein-protein interaction either by using the same DNA binding interface or more complex combination of zinc finger elements[395]. Consequently, these proteins serve in a variety of different biological functions such as differentiation/ development processes, metabolism, (post) transcriptional regulation or signal transduction[399]. As an example for a ZBTB protein binding to DNA but also mediating protein-protein interaction by the zinc finger region, PLZF or ZBTB16 is known to bind with its zinc finger 3-7 the DNA consensus sequence GTACAGTT(C/G)CAT but findings implicated that zinc finger 1-3 is the main binding interface for RARα (receptor for retinoic acid[395,400]. In combination with N-terminal BTB these ZBTB proteins have a versatile functional role within the cell.

Besides transcriptional regulation, the BTB containing proteins are involved in cytoskeleton dynamics or ubiquitin mediated protein degradation [391,401,402]. Other functions include DNA damage response, cell-cycle progression, and developmental processes such as gastrulation, organ formation and hematopoietic stem cell fate determination[392,403,404].

This novel class of transcription factors have been identified in species ranging from *Saccharomyces cerevisiae* and *Arabidopsis thaliana* to *Mus musculus* and *Homo sapiens*[391]. In human 49 ZBTB proteins are known and some of these members are listed in Table 1[394]. The name came from the initial discovery in *Drosophila melanogaster* and pox virus zinc finger proteins[391]. Through structural studies the BTB core could be structurally defined of five alpha helices, where helix α1 and α2 in conjunction with helix α4 and α5 form an alpha-helical hairpin and three beta strands (β1,2 and 3) form a beta sheet. The β-sheets 1&2 with α-helix 1&2 as well as β-sheets 3 connect to the α-helix 4&5 by the α-helix region 3 and a variable linker region (Figure 8).
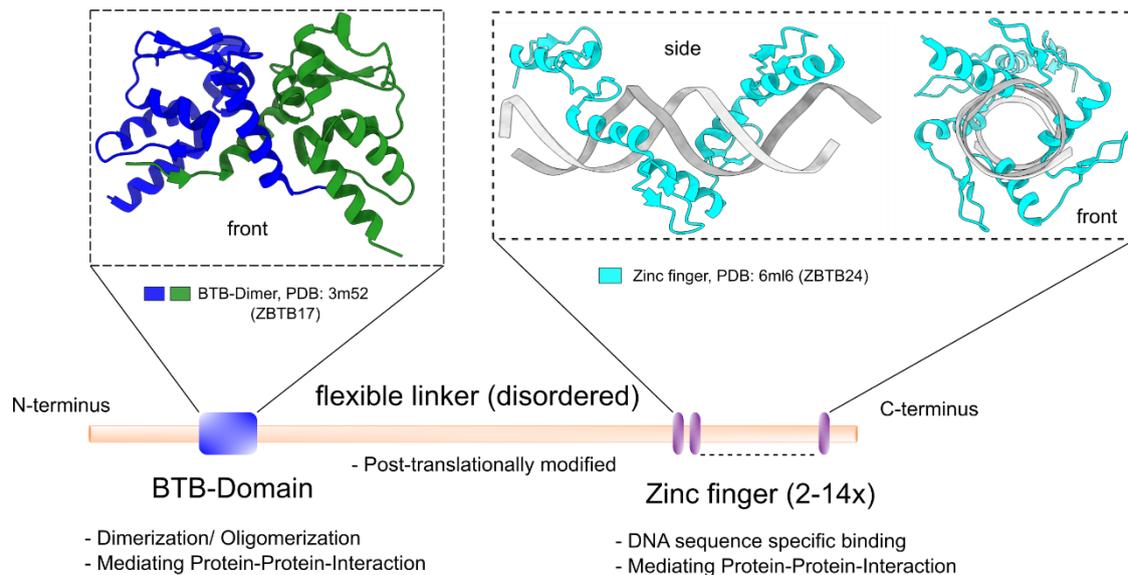
***Figure 8*: General architecture of a ZBTB protein.** At the N-term the BTB domain is responsible for dimerization/ oligomerization and mediating Protein-Protein-Interactions to influence cellular processes. As structural example (details see text) the crystal structure of the BTB domain from ZBTB17 (human, PDB 3m52) is shown[405]. At the C-terminus the zinc finger region comprising of C2H2 type can be found, mediating DNA specific contact but also Protein-Protein-Interactions. As example the zinc finger region of ZBTB24 bound to a methylated DNA sequence is shown (front and side few; *mus musculus*, PDB: 6ml6)[406]. Both domains are spaced by a flexible linker, known to be disordered and often post-translationally modified to alter the function of the ZBTB protein.

In Figure 8 the general architecture of a ZBTB protein is depicted. The N-terminal BTB domain (here PDB 3m52) of ZBTB17 shows the characteristic fold as described. The C-terminal C2H2-type fingers (here PDB 6ml6) of ZBTB24 shows how the DNA (major groove) is contacted by the zinc fingers. A special note here, the depicted structure shows the binding to a methylated DNA sequence. Notably, a complete structure of a full-length ZBTB protein has not been reported yet, since the structural determination due to the very flexible linker region is difficult. Moreover, researchers were sometimes more interested in one or the other domain and did not include in their observation the influence of each domain towards the other. Further studies have to be conducted to determine the influence of the BTB domain towards the Zinc finger region and vice versa. Other domains which are frequently associated with BTB containing proteins are Kelch, T1 and MATH domains. It has been shown that these domains function as recognition substrate for Cullin 3 based ubiquitin ligase complexes or voltage-gated potassium channels, thus expanding the functional context of the BTB domains within the cell[391,401,407]. Moreover, the BTB domain

recruits transcriptional coregulators, mediating rapid epigenetic changes in chromatin structure through histone methylation or acetylation[394].

Table 1: **A selection of ZBTB proteins and their associated phenotype/ cellular function**[394,408].
*The phenotype described here is when the protein is mutated leading to certain diseases.

ZBTB in human	Synonym	How many Zinc fingers present	Phenotype*/ cellular function
ZBTB1	ZNF909	8	Hematopoietic/ Lymphoid development
ZBTB2	ZNF437	4	? / Cell proliferation
ZBTB4	KAISO-L1, ZNF903	6	? / Cell cycle regulation, repression of p21
ZBTB7A	pokemon	4	Hematopoietic, embryogenesis, tumorgenesis/ Cell cycle, survival
ZBTB9	ZNF919	2	?
ZBTB14	ZnF478	5	? / Cell cycle regulation, dopamine transporter regulator
ZBTB17	Miz-1, ZNF151, ZNF60	13	Hematopoietic, embryogenesis / Gastrulation, cell cycle progression
ZBTB22	ZNF297A	3	?
ZBTB28	Bcl6b	6	Hematopoietic/ activation T-Cell, angiogenesis
ZBTB32	PLZP	3	Hematopoietic/ T-Cell proliferation, cytokine secretion
ZBTB40	ZNF923	12	? / Bone mineral density
ZBTB49	ZNF509	7	?

The dimerization of ZBTB proteins is an important feature to regulate the activity and DNA specific recognition to repress or activate transcriptional processes[394,409]. It has been shown that ZBTB proteins can facilitate DNA loop formation to mediate enhancer-promotor interaction, hence broadening their structural/ organizational role in chromatin architecture[394,410,411]. As a complementary result, recent findings described how ZBTB7b specifically targets DNA sequences to the inner nuclear membrane by interacting with HDAC3 (Histone deacetylase) and Lap2ß (lamina-associated protein), which results in transcriptional silencing or particular regions in the genome[394,412]. Notably, the BTB domain mediates the association with different corepressors and histone modifying enzymes (e.g. BCoR, NCoR-1/2, NuRD and HDAC), which remodel chromatin architecture and accessibility[394,412]. All these findings suggest an important role of ZBTB proteins within the chromatin organization and many members of the ZBTB family have been studied in the context of several cancer types. In case of ZBTB7A, it has been classified as therapeutic target and has been associated with cancer types in brain, lung, or prostate[413]. ZBTB proteins are associated in many different cellular processes and are till today poorly understood. The following study will shed light on a member of the ZBTB family, ZBTB9 which might be connected to the centromeric chromatin.

4.7 Chromatin Remodelers facilitate Chromatin access: The ISWI Chromatin Remodelling family

The DNA dependent biological processes rely on the accessibility of Chromatin and can be achieved by Chromatin remodelers. These complexes can catalyse reactions concerning nucleosome restructuring such as nucleosome sliding, conformational changes in nucleosomal DNA and histone exchange[35,414]. These changes regulate the transcriptional activity and accessibility of DNA for the fundamental processes in DNA recombination, DNA replication, transcription, histone variant deposition, and DNA repair[415]. The general architecture of a remodeler contains an ATP hydrolysing-DNA translocase unit, required to perform nucleosomal DNA translocation and therefore disrupting the contact between histone and DNA by utilizing the hydrolysis of ATP to ADP. There are four distinct families of ATP-dependent chromatin remodelers, which have distinct modes of remodeling[35,415]:

- Family 1: Switch/ sucrose-non-fermenting (SWI/SNF)
- Family 2: Imitation switch (ISWI)
- Family 3: Chromodomain-helicase-DNA binding (CHD)

- Family 4: Inositol requiring 80 (INO80)

Moreover, each remodeler type has their specific domains such as bromodomain in SWI/SNF, HAND-SANT-SLIDE (HSS) domains in ISWI, chromodomains in CHD remodelers or HAS domains in INO80 remodelers[415].

The following paragraph will highlight the ISWI chromatin remodler family since the project worked on characterizing these. For a comprehensive review on other chromatin remodeler families the reviews of Monica Tyagi et.al (2016) and Markert J, Luger K. (2021) are recommended[415,416].

The ISWI remodeler family is known for their regular spacing and assembly mechanisms, hence limiting chromatin accessibility and gene expression. In contrast, the NURF complex (nucleosome remodeling factor complex) associated with other subunits resulting in altered functioning, where it is active in transcription [35,417]. Chromatin remodelers are grouped in two categories based on their mode of action: 1) mediating histone posttranslational modifications on histones and 2) alter DNA-histone contact within the nucleosome through ATP hydrolysis. The latter action utilizes the energy to reconstruct the nucleosome from ATP hydrolysis (~7.3 kcal/mole) [418], hence disrupting the contact between histone and DNA to regulate the dynamic access to Chromatin[415]. In comparison, ISWI chromatin remodelers are smaller in organization, since they only compose of two to four subunits, each having the nucleosome-dependent ATPase ISWI. As described earlier, besides possessing a highly conserved ATPase domain, which is the engine of the chromatin remodeling complex, it possesses a HANDSANT-SLIDE domain with DNA activity [419].

The ISWI family is a well conserved ATPase family and possesses high conserved SWI2/SNF2 family ATPase domain, which belongs to a superfamily of DEAD/H-helicases. Some of the chromatin remodelling complexes with an ISWI ATPase were originally identified in *Drosophila* homologs and have been shown to be highly conserved in yeast and mammals. Complexes belonging to this family are NURF, ACF, CHRAC, NoRC, RSF and WICH. It has been shown, that CHRAC and ACF complexes assist in nucleosome sliding, whereas NURF is involved in epigenetic regulation (higher-order chromatin structure)[415]. A closer look in mammals reveal, the organization of ISWI complexes. One of the ATPase unit (SMARCA5 or SMARCA1) forms a stable complex with one up to three non-catalytic subunits interchangeably, therefore leading up to 16 complexes (Table 2).

Table 2: All ISWI complexes build with either ATPase subunit SMARCA5 or SMARCA1. The table summarizes the different complexes, which can be built upon binding either to SMARCA5 or SMARCA1[420].

ATPase subunit	Subunit	Complex
SMARCA5	BAZ1B	WICH-5
	BAZ1A	ACF-5
	BAZ1A, CHRAC1, DPOE3	CHRAC-5
	BAZ2A	NoRC-5
	RSF1	RSF-5
	BAZ2B	BRF-5
	BPTF	NURF-5
	CECR2	CERF-5
SMARCA1	BAZ1B	WICH-1
	BAZ1A	ACF-1
	BAZ1A, CHRAC1, DPOE3	CHRAC-1
	BAZ2A	NoRC-1
	RSF1	RSF-1
	BAZ2B	BRF-1
	BPTF	NURF-1
	CECR2	CERF-1

The organization in different subunits possess different functional domains, assisting in performing different tasks at the Nucleosome level [420,421]. The ISWI proteins contain different functional domains, facilitating its diverse function at the chromatin site. The SANT domain is a unique histone-interaction module, which links histone binding to enzyme catalysis and is an important domain for nucleosome sliding activity and regulation of nucleosome spacing. The WAC domain specific to the ACF complex, helps the complex to bind to DNA but also mediates the binding of ACF-related factors to the DNA. The Bromodomain is a conserved motif, enabling the recognition of acetylated lysine residues on histones ore interacting partners. Some complex members harbour a MBD (methyl-CpG-binding) domain responsible for recognizing and binding methylated CpGs. The binding to these elements triggers the methylation of H3K9 and induced transcriptional repression. The PHD-type zinc finger binds to specific epigenetic marks on histone tails, hence leading to the recruitment of transcription factors and nucleosome-associated complexes. In other

subunit members a C-terminal NegC domain can be found, which binds to the core2 domain and functions as an allosteric element in response to extranucleosomal DNA length[420].

As previously mentioned, centromeric transcript are involved in the recruitment and loading of CENP-A. Recent studies in the field of chromatin remodelers described the role of non-coding RNAs (ncRNAs) interacting with ISWI or other chromatin remodeling complexes to regulate chromatin modification and gene expression. The direct binding of ncRNAs to ISWI subcomplex members guides the complex to its anchoring site. Notably, these ncRNAs can also be incorporated into the complex formation and function as a scaffold in chromatin remodelling[420,422]. It remains elusive, if centromeric transcripts might also regulate/ influence ISWI complexes. Many of the known ISWI complex members have been implicated in cancer development and invasion and serve as prognostic markers. Therefore, the recruitment, function, and regulation of the ISWI chromatin remodeler family is an active area of research to facilitate the development of new therapeutic cues of different cancer types. The following study investigates the role of two ISWI remodelers (WICH-5 and CHRAC1-5 complex) in the context of centromere chromatin organization.

4.8 Cross-linking and Mass spectrometry (XL-MS) sheds light on protein connectivity

Due to the complexity and size of the kinetochore the interaction and changing conformation between each subcomplex becomes difficult to study. One way to overcome the limitations is by applying hybrid structural approaches. In general, are protein-protein interactions (PPIs) a fundamental process in biology and methodological advancements and developments paved the path to study the interactome of a given complex. There are many sophisticated methods such as co-immunoprecipitation, affinity purification combined with mass spectrometry (AP-MS), proximity-dependent biotin identification (BioID) or yeast two-hybrid screening (Y2H). All methods provide a good interaction resolution of a protein of interest but have their deficiencies[423]. Either the methods are time consuming or require tedious genetic alteration to study the subject. Moreover, information regarding structural and quantitative properties to complexes of interest are limited. One way to obtain structural and quantitative information is by chemical cross-linking and mass spectrometry (XL-MS). It is a versatile tool enabling the identification of protein connectivity and topology[424]. Mass spectrometry (MS) has become the method of choice for in-depth protein analysis near physiological conditions and characterization of macromolecular complexes in a cellular

environment[425]. MS enables to investigate the composition, regulation, and function of molecular complexes, resulting in discoveries of novel pathways[426]. The study of the cell proteome can be either achieved by Top-down or by Bottom-up proteomics. Top-down approaches analyse the protein in its entirety. The Bottom-up strategies follow the peptides generated by an enzymatic digest of the protein (e.g. trypsin) [426]. The peptides are ionized and transferred to the vacuum of a mass spectrometer, where fragmentation of peptides is initiated to generate a fragmentation spectrum (MS^2) to facilitate peptide identification. These spectra provide information on sequence identity and quantity of peptides, which can be referred to the actual protein[427]. The identification process of a peptide is built upon creating an *in-silico* library of theoretical peptides following the cleavage rule of used endopeptidase. The experimentally obtained spectra are matched against the theoretical database to identify the best masses' fit for a given protein identity[428]. An additional information can be drawn from the MS^1 scans, where relative abundance of a peptide corresponds to the spectral intensity observed[429]. Notably, the peak intensity is not direct proportional to the protein abundance, hence MS is not a quantitative method perse. Therefore, several quantification methods have been developed such as SILAC (stable isotope labelling by/with amino acids in cell culture), which is a metabolic labelling technique, whereas TMT (tandem mass tag) is an isobaric labelling on a peptide level[429,430]. These methods do not provide information on protein complex connectivity and topology. The rise of chemical cross-linking coupled to mass spectrometry paved the way to study multiple conformations of a given protein complex and its protein-protein interactions sites. In combination with native MS methods, the information gained provide a low-resolution insight into composition, stoichiometry, heterogeneity, and subunit assembly. Mostly XL-MS complements other structural methods such as X-ray crystallography, small-angle X-ray scattering (SAXS) or Electron (Cryo) Microscopy (Cryo-EM)[431,432].

Cross-linking of proteins can be achieved under near physiological conditions with homobifuntional reagents, such as Bis [SulfoSuccinimidyl] Suberate (BS^3)[433]. Amino acids in near spatial proximity are physically connected by a covalent bond with the reagent. This connection enables the localization of side chains involved in an interaction[434]. For XL-MS to provide information on proximity is restrained by the length of the reagent's spacer arm. For most cross-linkers this range from 7-30 Å[434].

In order to cross-link a protein complex, the cross-linker has to react and form covalent bonds with amino acid side chains. Many reagents target primary amine groups such as

those on lysine (K) side chain[433]. Nevertheless, other cross-linker form covalent bonds with carboxylic acids (aspartate, D; glutamate, Q) or thiols (cysteine, C)[435]. A cross-link reaction with lysine residues is preferred, since they are frequently exposed at physiological conditions with a positive charge on the protein surface and therefore easily accessible.[53] BS[3] consists of two reactive groups (Sulfo-N-hydroxysuccinimide (NHS) esters), which are separated by a spacer sequence. The primary amine on lysines side chains is nucleophilic and react with the NHS ester under a nucleophilic substitution reaction[433], to form a covalent amide bond[435]. The byproduct Sulfo-N-hydroxysuccinimide (NHS) leaves the reaction and a second nucleophilic attack can be performed on the opposing reactive group to form K-K bond[435].

The cross-link reaction could include incomplete or nonspecific reactions, leading to various cross-linked species[436]. There are four types of cross-linked peptides known: a) dead-end crosslinks (or mono-link) occur when one NHS reactive group formed a covalent bond with an amino acid residue, but the opposing group did not, due to hydrolysis of the reactive group[437]. b) loop-links occur when two residues within a peptide reacted with the cross-link reagent. c) inter-crosslinks occur, between residues of different proteins in spatial proximity[436]. d) intra-crosslinks occur when two residues within a peptide reacted with the cross-link reagent. The two latter types provide information on the conformational status of a protein and its interaction interface to other proteins, respectively [434,436].

5 Characterization of Novel Centromere associated Proteins and their Role in Centromere Transcription

5.1 Results

5.1.1 WICH-5 complex, CHRAC1-5 complex and ZBTB9 might be novel interactors of the Centromere

In an unpublished study (doctoral thesis, Dr. Götz Hagemann, 2020, LMU) novel centromeric interactors, which have been loosely associated with the centromere, were identified by ChIP-MS. The dataset ranges over hundreds of novel candidates in various cell cycle stages and need to be further validated to declare them as novel centromere interacting proteins. This study has chosen three proteins/ complexes for further investigations: ZBTB9 (Zinc finger and BTB domain- containing protein 9), the ISWI chromatin remodelers WICH-5 complex and CHRAC1-5 complex (Figure 9). The study was conducted as followed (Figure 9 A; a detail description can be found in the respective thesis): A stable HeLa T-Rex FlpIN of N-terminally tagged CENP-A or H3.3 was created. After confirmation of induced expression of CENP-A or H3.3 by addition of Doxycycline, the experiment for each timepoint was set. The cells were seeded out and expression of CENP-A or H3.3 was induced with Doxycycline post arrest state 48 h. Afterwards the cells were synchronized with a Thymidine Arrest to have the majority of cells in the beginning of G1-/S-Phase transition. By the addition of Lovastatin, the cells could be kept in G1-Phase. By releasing the cells in short time periods of 2 h, the different stages of S-Phase and lastly G2-Phase could be captured. After harvesting each timepoint (3 biological replicates of each), the cells were lysed, and the nuclei was enriched and pelleted. Afterwards the Chromatin was released with Micrococcal nuclease to obtain Oligo-Nucleosomes. The FLAG-tagged proteins were captured by anti-FLAG-beads and after subsequent washing, the FLAG-eluate was loaded onto Nickle-NTA beads and the resulting pulldown was digested on beads and subjected to LC-MS/MS Analysis.

The following plots were generated from the dataset presented in the mentioned thesis. This does not represent the complete dataset. Only the WICH complex, CHRAC1 complex and ZBTB9 were depicted here.

In Figure 9B and C a bar plot showing ZBTB9, CENP-B and the chromatin remodelers with their respective -log10 adj. p-values at each time point. In the following bar graphs the abundance of CENP-A and CENP-B is presented to depict the fluctuations at the centromere during the different cell cycle stages and to understand the abundance of the novel interacting proteins. Starting off with timepoint 0 (G1-/S-Phase transition), the abundance of CENP-B and ZBTB9 seem to be the same compared to CENP-A levels. When the cell enters the S-Phase the abundance of ZBTB9, CENP-B/A changes. In early S-Phase the abundance of CENP-B/A rises. When comparing the values to ZBTB9, then it seems that ZBTB9 levels rise more than CENP-A or CENP-Bs. During mid S-Phase CENP-Bs abundance is decreased, whereas ZBTB9 stays moderately the same with a slight decrease compared to early S-Phase. Afterwards in late S-Phase CENP-As abundance rises and obtains its peak compared to all timepoints. The abundance for CENP-B and ZBTB9 at this specific state are declining further but rises when the cell progresses towards G2-/M-Phase. After entering the S-Phase both change their occupancy. ZBTB9 seem to be slightly more enriched than CENP-B, whereas CENP-A and CENP-B share nearly same values. During mid S-Phase CENP-B ZBTB9 seem to be down regulated, followed by the same trend in late S-Phase. Upon entering G2-Phase, the levels of ZBTB9 and CENP-B recover, and CENP-A levels decrease slightly again. After the cell successfully finished one round of cell cycle and created its daughter cell, in G1-Phase the levels of CENP-A/B and ZBTB9 rise again. ZBTB9 has shown to be always associated with the centromere, where it has its highest occupation at early S-Phase and its lowest at the G1-/S-Phase transition.

The same analysis was performed for the Chromatin Remodelers of interest: WICH-5 and CHRAC1-5 complex (Figure 9C). Both complexes share the same ATPase (SMARCA5) and could be therefore interchangeable by just switching binding partners at certain cell cycle stages, whereas SMARCA5 remains associated with the centromere. At timepoint 0 only the subunit of CHRAC1 complex, DPOE3 has a high occupancy at the centromere. These levels could be explained that the cell primes for the replication process and DPOE3 is an important Polymerase subunit. Upon entering S-Phase the occupancy of the CHRAC1 complex becomes prominent. This does not change in the complete S-Phase, except that levels of BAZ1A do rise after late S-Phase and peak in G1-Phase. An explanation can be given by the fact that BAZ1A forms another complex with SMARCA5 (ACF-5) and SMARCA1 (ACF-1) without the subunits of CHRAC1 and DPOE3. In contrast to CHRAC1 complex, the WICH-5 complex subunit members peak in G1-Phase specifically. The abundance of BAZ1B has been low over the time of the cell cycle timepoints, it only peaks

in G1-Phase. As proposed in the unpublished study, WICH-5 complex might be a specific chromatin remodeler for the centromere specifically at G1-Phase. In case for CHRAC1 complex it can be a specific centromeric chromatin remodeler in S-Phase.

As described in this section, the novel centromere interactors do have a cell-cycle specific association with the centromere and were identified exclusively in the CENP-A Proteome screen. The next questions which shall be addressed in the next section are if the WICH and CHRAC1 complex bind to known centromeric proteins and if they have a link to the CENP-A containing Nucleosome.
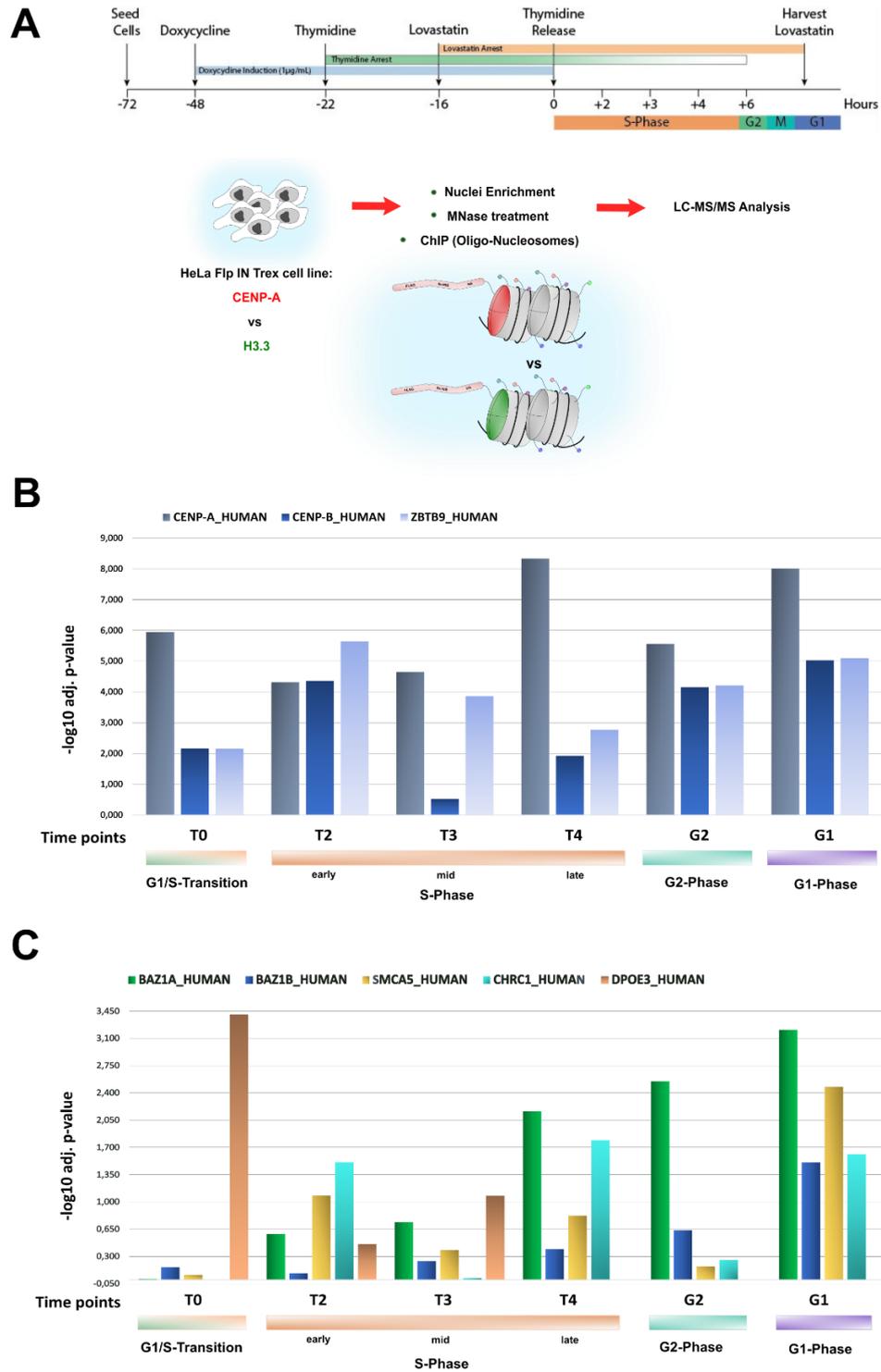
Figure 9: ChIP-MS of CENP-A proteome at diverse cell cycle time points. Original Data was kindly provided by Dr. Götz-Norman Hagemann, visualization and partial analysis was performed by Chandni Kumar. **A)** Experimental Design of the ChIP-MS Experiments conducted at different

timepoints. **B)** Barplot of -log10 adjusted p-values of ZBTB9, CENP-B and CENP-A at different cell cycle time points. **C)** Barplot of -log10 adjusted p-values of CHRAC1 and WICH complex components at different cell cycle stages.

5.1.2   The ISWI Remodelers WICH-5 complex and CHRAC1-5 complex can be linked to the Centromere

As described in the previous paragraph, the ISWI Chromatin Remodelers WICH-5 and CHRAC1-5 complex can be linked to the centromere and have a cell cycle dependent association with it. The first step in characterizing these remodelers, lies in the molecular weight determination and the stoichiometry of the complex. Each of the complexes were cloned into a Multi-Bac system to express them in insect cells. By subsequent purification, these complexes were further characterized. Therefore SEC-RALS experiments (collaboration with Gregor Witte) were conducted to determine the molecular weight of these complexes. In Figure 10 A, B for each complex a tabular description of the measured molecular weight is given. For CHRAC1-5 complex a molecular weight of 372.326 kDa has been measured. The complex consists of four complex components. If a stoichiometry of 1:1:2:2 (BAZ1A:SMARCA5:CHARC1:DPOE3) is assumed, then the resulting theoretical molecular weight would deviate about + 8 kDa. This result implies that this complex exists as a heterotetrametric complex and that CHARC1:DPOE3 exist in a tetrameric complex within the complex. Since CHRAC1 has been observed at S-Phase at CENP-A Nucleosomes, it was of interest whether this complex can bind CENP-A containing Nucleosomes. As shown at the right panel of Figure 10 A, the EMSA (Electro Mobility Shift Assay) shows a shift of CENP-A Nucleosome upon titrating in the CHRAC1 complex. The CENP-A Nucleosome wrapped around a Widom601 sequence was set at a concentration of 450 nM. With various concentrations of CHRAC1-5 complex (100-500 nM) the shift could be observed, by forming a higher molecular weight complex upon binding under non-denaturing conditions on a native gel. As a confirmation but also gaining more detailed insight into the binding of CHRAC1-5 or WICH-5 complex to the Nucleosome, cross-linking and mass spectrometry (XL-MS) enables a dynamic view on the conformational changes a protein complex can adopt. Therefore, XL-MS experiments were conducted where each chromatin remodeler was incubated with CENP-A Nucleosome [Widom601] and cross-linked using the homo-bifunctional crosslinker BS3. The network plots for each complex are depicted in Figure 10 A and B for each chromatin remodeler. The crosslinks were obtained by using the xQuest/xProphet software and crosslinks were filtered with 1% FDR and were

manually validated if the spectra apply to a manual validation system. Cross-links passing this validation step were taken and visualized in xVis.

The smaller subunits CHRAC1 and DPOE3 interact specifically with each other, and it has been shown that there are some existing cross-links showing a formation of dimers of each and supporting the findings seen in the molecular weight determination by SEC-RALS. Moreover, the map shows that DPOE3 is the main connector between the bigger subunits and the smaller ones. Only three cross-links were observed between CHRAC1 and BAZ1A (aa K10 to aa K1517, aa K102 to aa K197 or aa K102 to aa K28). The cross-links at the C-terminal domain of DPOE3 direct to two distinct domains at BAZ1A (DDT and Bromo domain). In case of CHRAC1, the cross-links at the C-term direct towards the N-term of DPOE3. The cross-links between SMARCA5 and BAZ1A are more versatile, implying the complex is highly dynamic upon binding to the Nucleosome. In a previous study, the binding site of SMARCA5 to BAZ1A was studied and the stretch of aa 667 to 933 is required on BAZ1A to sufficiently bind SMARCA5. The presented cross-link map does show other observed cross-links, where the situation might change upon binding to a Nucleosome. (in this case to CENP-A containing Nucleosomes). Nonetheless, the majority of cross-links from BAZ1A to SMARCA5 can be found within both Helicase domains and near the DEAH-box domain. A closer look at the cross-links obtained from the nucleosome reveal, that H2B is interacting with BAZ1A exclusively. The observed cross-links are found near the DDT domain as well to the zinc finger PHD type domain. At the site of H2B, the cross-links are directed towards the C-terminus specifically adjacent to the two glutamic acid side chains (E102, E110) known to be involved in the formation of the acidic patch in conjunction with H2A residues. The cross-link map does not provide information about H2As involvement or binding to one of these domains. Nonetheless, cross-links near the acidic patch of H2A to H2Bs acidic patch can be found. The map provides another interesting crosslink observed from CENP-A to the SMARCA5 subunit. When considering the other map of the WICH complex (containing the same ATPase unit) a possible binding mechanism of SMARCA5 to the CENP-A Nucleosome can be drawn. A previous study structurally determined by Cryo-EM the binding of SMARCA5's Helicase domain to the H3.2 Nucleosome[438]. There the Helicase was attached sideways to the Nucleosome. Only the position of the Helicase was determined in this structure but not how the SANT domains could be involved in it. Building up an *in silico* structure reconstitution of alphafold[330,331] predicted SMARCA5 and existing CENP-A Nucleosome structure with the knowledge gained from the cross-link maps, a binding mechanism of SMARCA5 can be proposed (Figure 11 A, B). The region

with the SANT domains helps the ATPase to bind and by a proposed flip mechanism the Helicase domain is placed sideways and helps to pump the DNA through to mid-position the Nucleosome by a sliding mechanism. By this means the observed cross-links between CENP-A and SMARCA5 can be explained.

After characterizing CHRAC1-5 complex, it was of interest how the WICH-5 complex differs in comparison to CHRAC1-5 complex. Hence as a first step of characterization, the molecular weight (MW) of WICH-5 complex was determined. WICH complex was measured with an MW of 330 kDa, where the difference to the theoretical MW is about 30 kDa less (296 kDa). Since the complex only consists of two subunits, the stoichiometry of this complex could be 1:1 (BAZ1B:SMARCA5), as judged by the molecular weight measured. Furthermore, the WICH complex was enriched mostly in G1-Phase. The question was asked if the complex binds to CENP-A containing Nucleosomes as seen for CHRAC1-5 complex. Therefore, an EMSA assay was performed to see if WICH-5 can bind CENP-A containing Nucleosomes. This was confirmed by titrating in WICH complex (100-500 nM) in different molar concentrations to 450 nM CENP-A Nucleosome^{Widom601}. Indeed, the WICH-5 complex can bind to CENP-A Nucleosomes, thus confirming the ChIP-MS observation. It was of interest whether the cross-link of WICH complex:CENP-A Nucleosome complex would differ compared to CHARC1-5 complex. The cross-link map In Figure 10 B shows distinct features different to the CHRAC1-5 complex. By stringent filtering of 1% FDR, main crosslinks between aa stretch 1000-1500 in BAZ1B and aa 740-930 in SMARCA5 can be found, proposing this bind to be the interaction site of both. The binding mode for these two are different as for BAZ1A to SMARCA5. As already seen in CHRAC1 complex, H2B is mainly interacting with BAZ1B, suggesting that H2B in conjunction with the other histones mediates the interaction to the bigger subunits, directing the ATPase to the Nucleosome. The cross-links from H2B to BAZ1B are in similar fashion as described for CHRAC1 complex, implying a straightforward binding mechanism for this type of remodeler. It has to be mentioned, that crosslinks to H4 were not identified, therefore making it difficult to have a complete understanding on how the Nucleosome is bound by the respective chromatin remodelers, since the mentioned publication with the Helicase domain of SMARCA5 bound to H3.2 Nucleosome, the binding of H4 plays an important role [438].

As previously mentioned, both chromatin remodelers were identified in respective cell cycle stages specifically enriched in the CENP-A Proteome. In a MS-affinity enrichment experiment, attempts were made to identify interaction partners of WICH-5 or CHRAC1-5

complex (see section 5.1.6), more precisely to find kinetochore proteins to link the remodelers directly to the centromere. Surprisingly CENP-B have been found in the WICH complex but not in CHRAC1. Moreover, in CHRAC1 no kinetochore proteins were found. If WICH can bind CENP-B and therefore confirm the MS-based experiment, an analytical size-exclusion experiment was conducted (Figure 10 B down panel), where the binding of CENP-B to WICH-5 was determined. Both form a stable complex by co-eluting and the binding was confirmed by an additional *in vitro* binding assay (Figure 13 A lower panel). This novel complex formation between WICH-5 complex and CENP-B could be an explanation for the presence of the WICH-5 complex in the CENP-A Proteome pulldowns. Maybe the presence of CENP-B tethers the WICH-5 complex to the centromere to remodel CENP-A in the replenishment phase of newly synthesized CENP-A and its subsequent incorporation in G1-Phase.

Taken all results together, the ISWI family chromatin remodelers WICH (BAZ1B, SMARCA5) and CHRAC1 (BAZ1A, SMARCA5, CHRAC1, DPOE3) complex have been identified in a ChIP-MS screen and biochemical data suggests, that indeed these complexes are involved in the maintenance of the centromere at different cell cycle stages, performing various tasks. Additionally, the cross-link data provides a hint, how the ATPase subunit SMARCA5 might interact with the Nucleosome through its SANT domain and through a proposed flip mechanism the Helicase domain can interact with the Nucleosome.
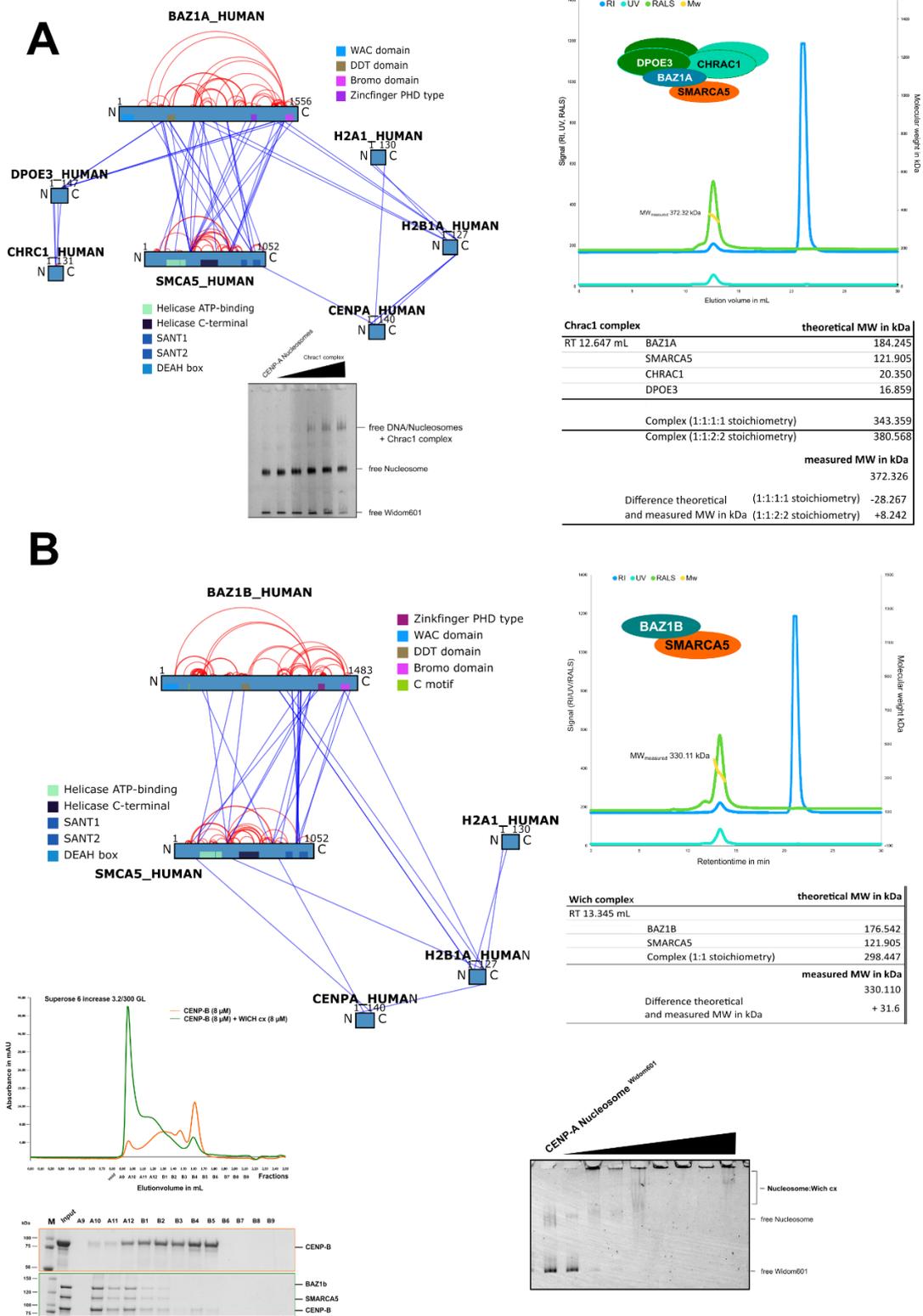
Figure 10: Cross-link Map of WICH and CHRAC1 complex bound to CENP-A Nucleosome and EMSA assays confirming the binding. **A)** Cross-link map of CHARC1 complex with CENP-A

Nucleosome^{Widom601} (BS^3 used as cross-linker, 1% FDR). Next to the map a SEC-RALS analysis of the Molecular weight of the CHRAC1 complex can be found. The complex exists as a four-component complex where CHRAC1 and DPOE3 exist as a heterotetrametric complex. Beneath the table an EMSA is shown for the binding of CHRAC1 complex with CENP-A Nucleosomes (Widom601). **B)** Cross-link map (BS^3 used as cross-linker, 1% FDR) of WICH complex with CENP-A Nucleosome^{Widom601}. The table at the right side shows the measured Molecular Weight of the complex, existing as a heterodimer with 1:1 stoichiometry. Beneath, an EMSA assay to confirm the binding of WICH to CENP-A Nucleosome (Widom601). Additionally, an analytical SEC run confirming the binding of CENP-B to WICH complex by co-elution.

A

K787^{SMARCA5}

K65^{CENP-A}

CENP-A Nucleosome; PDB: 6TEM
SMARCA5: Alpha-Fold: AF-O60264-F1

90°C

K1023^{SMARCA5}

K54^{CENP-A}

CENP-A Nucleosome; PDB: 6TEM
SMARCA5: Alpha-Fold: AF-O60264-F1

- Helicase domains of SMARCA5
- Alphafold prediction SMARCA5
- CENP-A (PDB 6TEM)
- H2B (PDB 6TEM)

B

K160^{SMARCA5}

K125^{CENP-A}

possible folding/flip
mechanism

of Helicase domain
to the
observed cross-link

CENP-A Nucleosome; PDB: 6TEM
SMARCA5: Alpha-Fold: AF-O60264-F1

H3.2 Nucleosome bound to SMARCA5 PDB: 6ne3
SMARCA5: Alpha-Fold: AF-O60264-F1

- Helicase domains of SMARCA5 (PDB 6ne3)
- Alphafold prediction SMARCA5
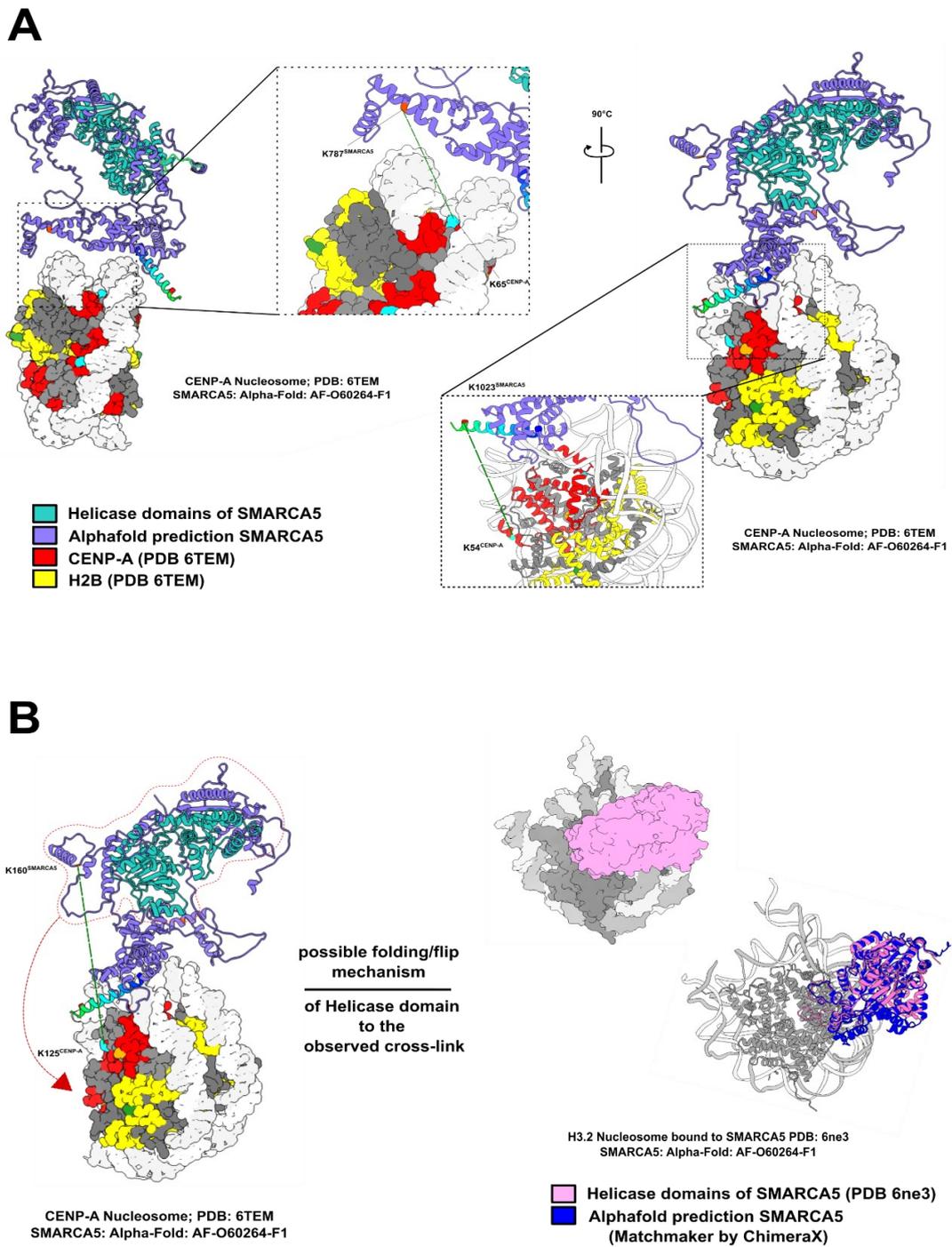 (Matchmaker by ChimeraX)

Figure 11: **Possible Structural Explanation of observed cross-links between CENP-A Nucleosome and SMARCA5. A)** Structural Depiction of *in silico* modelled alphafold predicted SMARCA5 to a known CENP-A Nucleosome Structure (PDB 6TEM) to explain the observed cross-links between CENP-A and SMARCA5 by considering the cross-links observed either in CHRAC1 or

WICH complex. **B)** Possible explanation how the observed crosslink between CENP-A and SMARCA5 (K125 to K160) can be in spatial proximity, even though the constructed structure shows a much larger distance between both residues. This can be explained by a possible flip mechanism of the helicase domain.

5.1.3 The Transcription Factor ZBTB9 is affiliated with the Centromere through the binding to CENP-B

As described in an aforementioned section, a transcription factor called Zinc finger and BTB containing protein 9 (ZBTB9) was consistently associated with CENP-A throughout the cell cycle stages. The protein itself is not well characterized and its function in the cell is unknown. The protein consists of two well-structured domains, joined by an unstructured linker region, which has a small acidic patch within (aa 212-225). The N-terminal domain (aa 48-112) is a BTB domain (**B**road-Complex, **T**ramtrack and **B**ric a brac) and at the C-terminus (aa 411-460) two Zinc finger of C2H2 can be found. The latter Zinc finger is an atypical C2H2 type, since it has additional cysteines and histidines to coordinate a zinc ion as seen in typical C2H2 type zinc fingers. The BTB domain is known as a dimerization/ oligomerization domain as well as a binding hub to mediate protein-protein interaction. The zinc finger in general do contain DNA-binding properties but also can commit to protein-protein interactions. Since ZBTB9 has been identified to be a potential specific centromere binding protein, efforts have been made to biochemically characterize this protein in this study. Of particular interest was, whether ZBTB9 binds to one of the kinetochore-associated proteins explaining the constant abundance with the centromere.

As shown in Figure 12 D, ZBTB9 was cloned as an MBP-fusion and various constructs were expressed in *E.coli* and recombinantly purified. To test specifically the inner CCAN, all subcomplexes were cloned into the Multi-Bac system and expressed in insect cells. Each complex was purified and tested for binding to ZBTB9 (Figure 12 A). Upon analysing the bound fraction, CENP-B was the only kinetochore member binding to ZBTB9. CENP-B is known to bind the CENP-A Nucleosome specifically with its N-terminal DNA helix-turn-helix binding motif to a specific DNA-sequence, so called CENP-B box. The other kinetochore members were not enriched in the binding assay, prompting CENP-B as the only binding partner of ZBTB9 and links the transcription factor directly to the centromere. After establishing the binding of ZBTB9 to CENP-B, it was of interest, where ZBTB9 binds CENP-

B and therefore XL-MS experiments as well as deletion-mutant *in vitro* binding assays were conducted.

Besides ZBTB9s binding to CENP-B also the complex formation between CENP-B and the C-terminal CENP-C fragment was of interest, since the latter one has not been studied in detail. Therefore, ZBTB9 with CENP-B and the C-terminal region of CENP-C (546-943) were mixed in equimolar ratios and crosslinked with DSSG for 15 min. This short cross-linking at a predefined cross-linker concentration provides an insight into the binding dynamics between these proteins without creating artefacts or aggregates. After validation (FDR 1%) the cross-links were visualized by xVis as a network plot (Figure 12 B). The network plot reveals that ZBTB9 (ZB) contacts CENP-B (CB) via its C-terminal Zinc finger region the N-terminal DNA-binding domain of CENP-B (ZB aa K390 to CB aa K47; ZB aa K382 to CB aa K103; ZB aa K390 to CB aa K278). Moreover, the map provides valuable information regarding ZBTB9s and CENP-Bs structural organization. The intra-links show that the BTB domain is structurally placed near the zinc finger region. The Alphafold model (AF-Q96C00-F1) of ZBTB9 shows similar structural features. However, when trying to fit the obtained cross-links into the model, the distance between each lysine residues did not match with the spacer length of the used cross-linker. This implies, that the cross-links have captured a confirmation, that cannot be depicted by Alphafold. Notably, the binding of CENP-B to CENP-C happens through the N-terminal DNA binding domain but also cross-links are observed near the acidic patch of CENP-B. A closer look at CENP-C reveals, that the binding of CENP-B happens within the region 550 to 700. This region does not include the cupin dimerization domain of CENP-C, thus CENP-B binds prior to this domain. The cross-links provide valuable insides on the very inner kinetochore axis between CENP-B and CENP-C.

After establishing the most probable binding site of CENP-B to ZBTB9, the actual binding region had to be biochemically determined. XL-MS can only provide data on spatial restraints of neighbouring lysine residues, that happened to be in the right angle or distance to cross-link to each other.

To determine the minimal binding region, deletion mutants lacking certain domains were created (Figure 12 C, D). A biochemical binding assay with MBP-immobilized ZBTB9 mutants confirmed the loss of binding when the zinc finger region is missing. The binding is lost upon complete deletion (aa 380-462; 91% binding loss) and an interesting phenotype reduction can be seen, when the zinc finger at aa 409-433 is depleted (50% binding loss),

whereas the second zinc finger only accounts for a binding reduction of 30 %. An interesting observation was made upon deletion of the BTB domain (42-112). Here the opposite can be seen, where CENP-B binds 15% better then seen in the wildtype scenario. This implies, that the BTB domain influences the binding capabilities of the zinc finger region to CENP-B, which is in line with the observed intra cross-links, since the BTB-domain seems to be in near spatial proximity to the zinc fingers. Indeed, some literature propose the possible influence of the BTB on the actual binding capacities of the zinc fingers.
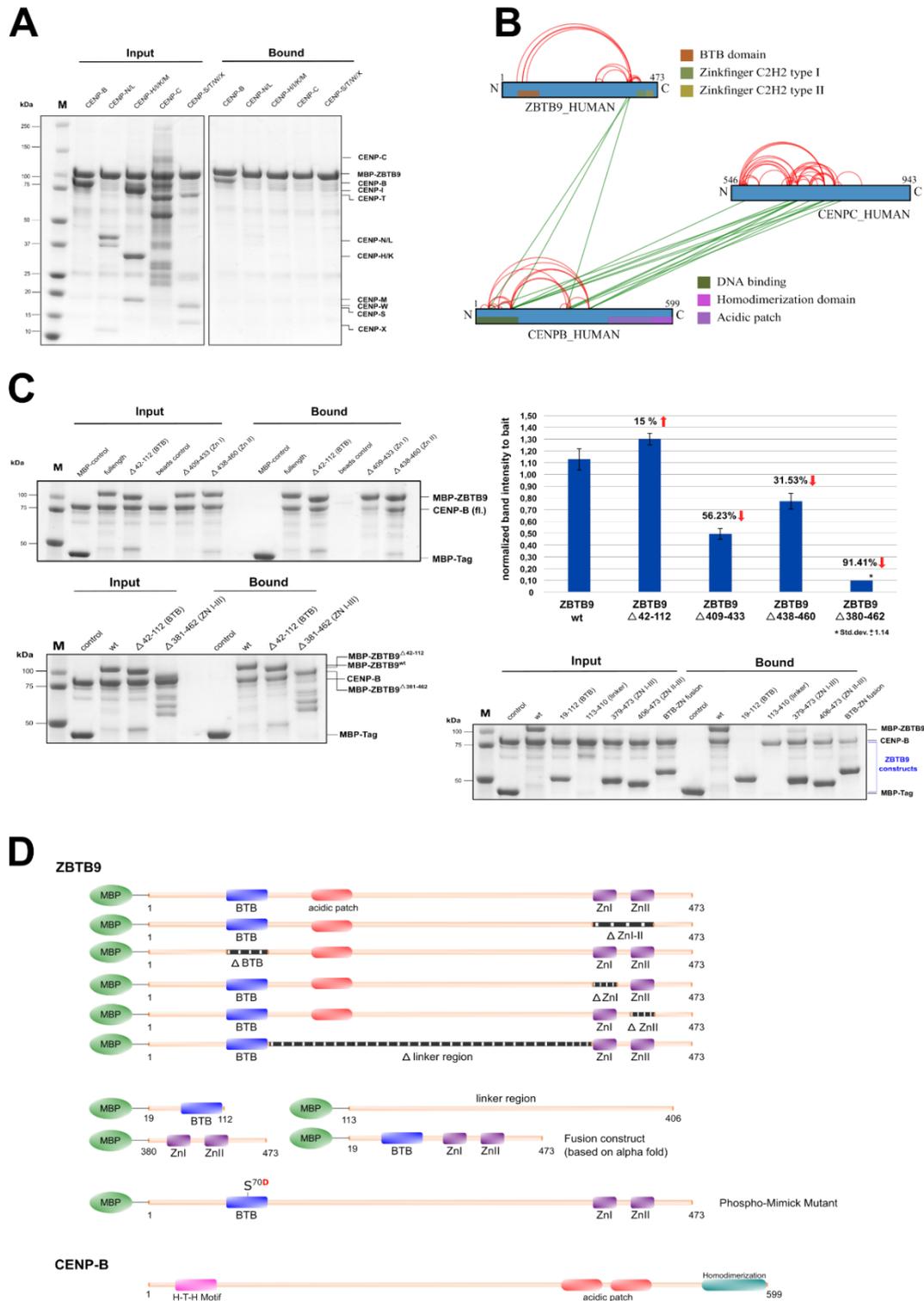
Figure 12: **ZBTB9 binds** *in vitro* **CENP-B.** *In vitro* Binding Assay to determine kinetochore binding partner. **A)** Different human inner kinetochore complexes were incubated with immobilized ZBTB9 on Maltose Resin by its N-terminal MBP-tag. Assays were performed in at least duplicates. **B)** XL-

MS confirmed binding of ZBTB9 to CENP-B and map provides information to the binding sites. **C)** Deletion mutant of ZBTB9 confirms C-terminal binding of ZBTB9 Zinc fingers to CENP-B DNA-Binding domain. Short constructs of each domain confirm the binding site in the zinc finger region. Each experiment was conducted at least in two biological replicates. Quantification of band intensities show differences within zinc finger mutants. Standard deviation (Std. Dev.) is shown as error bars. In case for the delta mutant 380-462 the calculated Std. Dev. was 1.14. For comparison reasons, the Std. Dev. was not shown. **D)** Overview of ZBTB9 constructs used in this study.

Nonetheless, detailed experimental data dissecting how the BTB influences the zinc finger on a structural level are still missing. In this study, the crosslinks provide such an insight and could explain this better binding phenotype. A closer look at cross-links from the zinc finger region (near the deletion mutant aa 409-433) to the BTB domain reveal their structural neighbouring and explains in addition why the deletion of the BTB leads to a higher affinity binding, since the BTB cannot regulate the binding to this region anymore.

The Alphafold model of ZBTB9 predicted a third zinc finger prior to the first annotated one with lower confidence (confidence level between 90-50). In conjunction with alphafill[439], the model is capable to correctly coordinate a zinc ion, therefore this could explain why the deletion of the region 409-433 only results in a 50% loss. The other 50% might be lying in this novel predicted zinc finger with the second zinc finger. As a confirmatory approach, the small domains were expressed as MBP-fusion constructs and the binding to CENP-B was tested. Indeed, CENP-B does only bind to the Zinc finger region, where the best binding was observed with the construct aa 380-473 (annotated zinc fingers with predicted one). A smaller construct comprising of the annotated two zinc finger regions (406-473) resulted in a 15 % binding reduction than compared to the previous construct containing all three zinc fingers. Interestingly, another construct where the BTB was fused to the zinc finger region by a small linker region based on the observed cross-links resulted in a loss of 52% binding. This suggests an important role for the unstructured linker region, ensuring the correct structural arrangement of both domains, so CENP-B can bind the zinc finger region correctly. This confirms that all three zinc fingers are required for proper binding of CENP-B to ZBTB9 with special emphasis on zinc finger region 409-433. The binding event of CENP-B to ZBTB9 was further tested in an analytical SEC experiment and it could be confirmed that both form a stable complex by co-eluting (Figure 13 A upper panel).

In the previous paragraph the binding of CENP-B to WICH complex have been investigated, it was of interest whether CENP-B is capable in binding both proteins at the same time. This has been the case as shown in the *in vitro* binding assay (Figure 13 A lower panel, left

assay as control). Here WICH complex was preincubated with CENP-B. After washing, ZBTB9 was titrated in. As a result, all three remain bound. In a similar experiment, ZBTB9 was immobilized on maltose beads and CENP-B was preincubated. Afterwards WICH has been titrated in, but this time WICH complex remained less bound (data not shown). The experiment leads to the conclusion, that CENP-B can bind both at the same time but when CENP-B is bound to ZBTB9 prior, WICH has limited access to CENP-B. For confirmation, this experiment has to be redone by immobilizing CENP-B again and preincubating with ZBTB9 and titrating in WICH complex. If the result is similar to the just mentioned binding assay, then there is a timely restricted binding of CENP-B to WICH or ZBTB9, which might correlate with the ChIP-MS Data.

When analysing the chromatogram of ZBTB9:CENP-B binding, a high molecular weight (MW) complex (1 mL elution volume on a Superose6 increase 3.2/300 GL column; corresponding to > 700 kDa) was observed. Since ZBTB9 should have a theoretical MW of 50 kDa, it was expected that a moderate higher MW of a single ZBTB9 molecule bound to a CENP-B dimer or maybe two ZBTB9 bound to one CENP-B homo dimer might elute. It was of interest, whether the high MW was induced by ZBTB9, since it is known that CENP-B forms as a homo dimer. Indeed, the MW determination of ZBTB9 (Figure 13 B, collaboration with Gregor Witte) confirmed, that ZBTB9 adopts a much higher MW, resulted in a measured MW of 774 kDa. By a theoretical mass of a ZBTB9 monomer (50 kDa; with Tag 56 kDa) would ZBTB9 assemble with itself 14 times or when taken two ZBTB9 molecules together due to the properties of the BTB domain to form dimers, it can assemble as a heptamer. The formation of such a ZBTB9 multimer in complex with the CENP-B homodimer explains the early elution from the column.
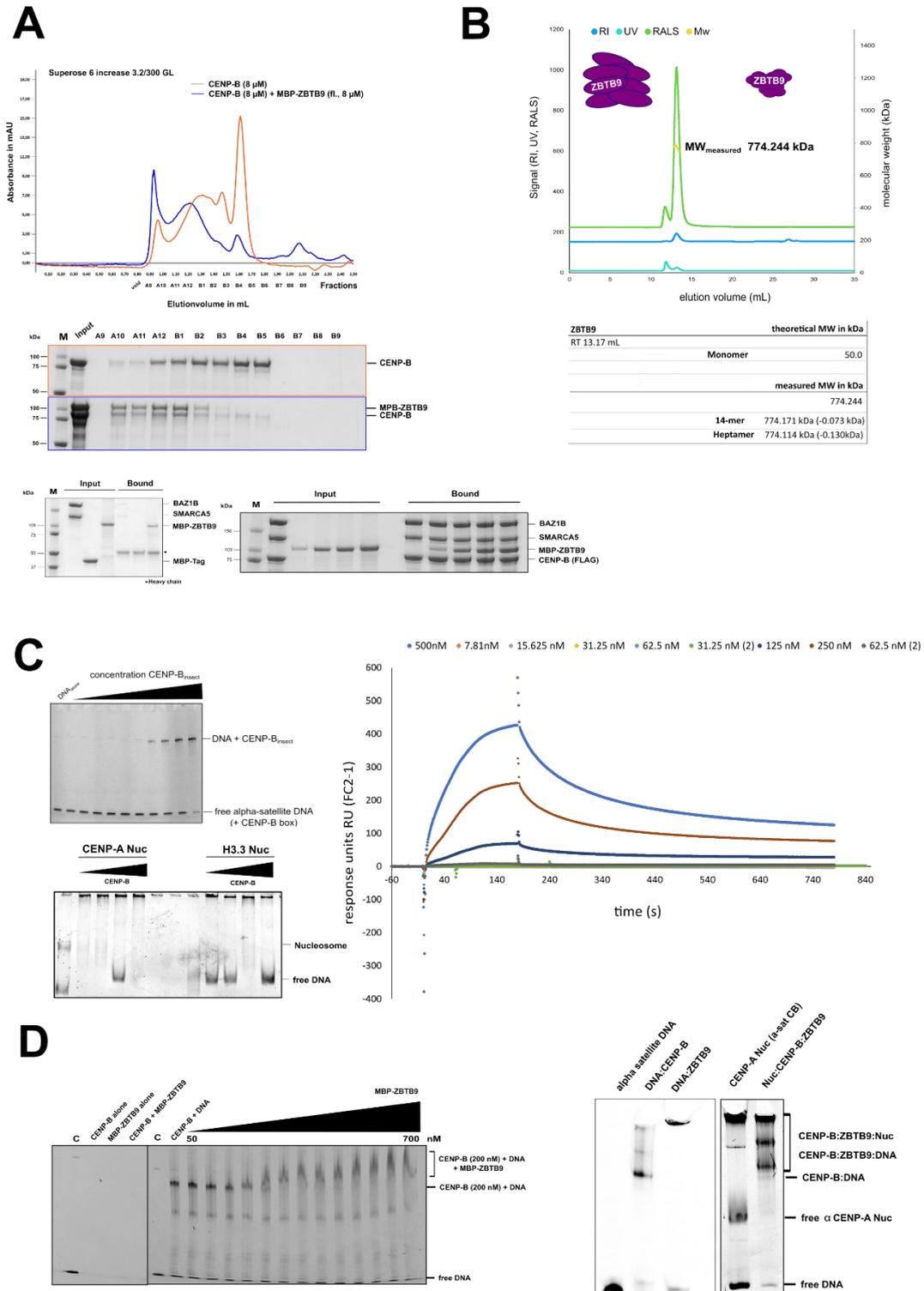
***Figure 13: In vitro* Characterization of ZBTB9:CENP-B Binding. A)** Analytical SEC with 8 µM CENP-B and 8 µM ZBTB9 on Superose6 increase 3.2/300 GL (Cytiva). Fractions containing complex were collected, TCA precipitated and loaded onto a SDS-PAGE and stained with Coomassie Blue.

For each fraction, the same amount of TCA precipitated sample was loaded. Beneath the SEC-Chromatogram an in vitro binding assay shows the binding of CENP-B to ZBTB9 and WICH complex. CENP-B was prebound to the WICH complex and ZBTB9 was titrated in. After washing, samples were loaded onto SDS-PAGE and stained with Coomassie Blue. **B)** SEC-RALS Chromatogram to measure the actual molecular weight of ZBTB9 (construct 8xHis-Twinstrep-precission-ZBTB9). *In vitro* Binding assay of WICH:CENP-B:ZBTB9 to determine if the binding to CENP-B is mutually exclusive. Another in vitro Binding assay showing the binding of CENP-B to different CENP-A peptides. **C)** EMSA assay to determine CENP-Bs binding to free alpha-satellite DNA containing a CENP-B box and the subsequent analysis with CENP-A Nucleosome and H3.3 Nucleosome with Widom601 sequence. **D)** EMSA to determine the binding of ZBTB9 to CENP-B:alpha-satellite DNA and subsequently the influence on CENP-A Nucleosome with alpha-satellite DNA containing a CENP-B-box.

After establishing that ZBTB9 exists as a multimer and can build a stable complex with CENP-B, it was of interest how the binding kinetics between both are. Therefore SPR (Surface Plasmon Resonance) Experiments on a Biacore system were conducted (collaboration with Dr. Gregor Witte, Figure 13 C right). Here a titration of MBP fused ZBTB9 wildtype (7.81 nM to 500 nM) to immobilized CENP-B (50 ng protein, via Twinstrep on a biotin chip) resulted in the measurement of two KDs (fitting curve for heterogenous ligand binding). The first KD is about 31.6 µM whereas the second ranges in 6.7 µM. Both KDs show the expected range for a protein-protein binding, nonetheless the kinetics of this binding cannot be fitted into a 1:1 binding curve. Even fitting to a heterogenous ligand or bivalent analyte is not complete satisfactory, since the fitting curve does not completely align with the measured values. This could be due to the multimerization of ZBTB9 and the existence of monomer/ dimer/ tetramer formation at different concentrations, making it difficult to establish a binding kinetic to CENP-B. Moreover, since CENP-B itself exists as a homodimer, theoretical two DNA-binding sites are available for binding. Now it could be that two zinc finger regions in two ZBTB9 molecules bind one DNA-binding sites of CENP-B followed by another dimer binding to the other freely available DNA-binding site. Another explanation could be the binding of the first zinc finger (first observed KD with higher affinity) and subsequently the next two zinc fingers follow (second measured KD with lower affinity; possible hinged mechanism) on one or two DNA-binding domains. Since the binding kinetics could not be unambiguously determined in this setting, one has to switch the system and immobilize ZBTB9 on a chip and have to titrate in the DNA-binding region of CENP-B. This setting might provide valuable information on how the binding occurs, when CENP-B exists as a monomeric component and how the multimerization of ZBTB9 acts on the

binding kinetics. It is also possible to try the BTB delta mutant to decrease the multimerization of ZBTB9, but it has been observed in the cross-links that the zinc finger region might also provide a second dimerization domain, since zinc fingers are known to form dimers[440]. Nevertheless, the SPR Experiment confirmed the binding of ZBTB9 to CENP-B and provided a glimpse in the multifaceted binding of this complex.

Next, CENP-B is known to bind the CENP-B box at the entry site of CENP-A containing Nucleosomes. By the formation of CENP-B:ZBTB9 complex specifically by establishing in this study the binding of the zinc finger region to the DNA-binding region, it is natural to investigate if the binding of ZBTB9 to CENP-Bs DNA binding region might have an influence on the binding affinity towards the alpha-satellite DNA with integrated CENP-B box. Therefore, several EMSA (Electro mobility shift assays) were performed (Figure 13 C left, D). In a first attempt the binding of CENP-B (expressed in insect cells) to the 171 bp long alpha-satellite containing CENP-B box (further referred as α-satCB) was investigated. The sequence was extracted by restriction enzyme cutting and subsequent purification (vector containing the sequence was kindly provided by Dr. Nikolina Sekulic). 5 nM of DNA was incubated with different CENP-B concentrations (5 nM-200 nM) in a buffer at physiological pH and 150 mM NaCl and incubated for 25 min at room temperature. The sample was loaded on a self-pre-cast native Tris-glycine gel and run at native conditions. The gel was stained with PAGE Red (Biotum) and visualized on a Typhoon imager. In this setting the binding is observed at concentration of 100 nM of CENP-B (1:20 ratio; DNA:CENP-B). To confirm the binding of CENP-B to Nucleosomes, CENP-B was incubated with fixed concentration of CENP-A containing or H3.3 containing Nucleosomes (200 nM; Widom601 with 167 bp) and CENP-B was titrated in (10-100 nM). CENP-B is shown to bind to the Nucleosome having a slight preference for CENP-A Nucleosomes. After establishing the binding of CENP-B to the α-satCB and Nucleosomes, the experiments were redone in presence of ZBTB9 (Figure 13 D). First the binding of CENP-B:DNA complex to ZBTB9 was investigated. CENP-B and α-satCB were incubated (200 nM CENP-B) prior. Afterwards ZBTB9 was titrated in (50-700 nM) and incubated at room temperature for 25 min. As seen on the gel the first shift of the complex was observed at equimolecular concentration between ZBTB9 (200 nM) and CENP-B:DNA complex. At higher molar concentrations of ZBTB9 multiple binding events are observed as seen by non-distinct staining of the complex.

Moreover, ZBTB9 seems to bind the α-satCB DNA at higher concentrations, whereas in lower ranges it only binds to the CENP-B:DNA complex. After establishing the binding of ZBTB9 to existing CENP-B:DNA complex, the binding of ZBTB9 to CENP-B:CENP-A Nucleosome was determined. A protocol for CENP-A containing Nucleosome with 171 bp wrapped α-satCB DNA was established. Nucleosomes obtained after heat shifting at 50°C and mid-positioned as judged by native PAGE were taken as substrate. The nucleosomes were kept in a low salt buffer (50 mM NaCl). 2 μM of CENP-A$^{α\text{-}satCB}$ Nucleosome were incubated with 500 nM CENP-B and 500 nM ZBTB9 for 25 min at room temperature. The controls show upon binding of CENP-B to the DNA, a complex can be observed as described previously. This complex compared to the binding to the Nucleosome runs slightly faster than the CENP-B:Nucleosome complex. Next the binding of ZBTB9 to the DNA alone was investigated. Since ZBTB9 forms this higher order multimers, the DNA shifts with ZBTB9 but does not enter the gel pockets due to its high molecular weight and partial aggregation. This phenomenon can be seen upon binding to the Nucleosome as well since intense staining were observed in the pockets. A note aside has to be given that this delicate CENP-A$^{α\text{-}satCB}$ Nucleosome is prone for quick degradation processes, resulting in the intense staining of decomposed nucleosomes as seen in the gel pocket. Unfortunately, this process could not be prevented since this specific Nucleosome assembly resembles the *in vivo* formation with its native sequence as compared to the artificially optimized Widom601 sequence used in common studies, which is much more stable. The last lane represents the binding events for the complex of Nucleosome:CENP-B:ZBTB9. The Nucleosome shifts completely, and two distinct complexes are formed. One can be the CENP-B:Nucleosome the other one is the formation of the trimeric complex in a 4:1:1 ratio.

Taken together all presented results, ZBTB9 can be seen as a novel centromere interacting protein through its specific binding to CENP-B based on *in vitro* biochemical and biophysical findings. These findings were performed by leaving out the cellular *in vivo* context. As a next step, *in vivo* microscopy of fixed human cells was conducted to determine the localization of ZBTB9 and its role in the *in vivo* situation. Moreover, knockdown experiments should provide an insight if ZBTB9 has an important or dispensable role in the cell (in collaboration with AG Leonhardt, Jeannette Koch/ Dr. Hartmann Harz, LMU).

As a first step, the co-localization of CENP-B with CENP-A was established. Therefore, U2OS cells were fixed and incubated with specific anti-CENP-A and anti-CENP-B antibodies. The nucleus was stained with DAPI. The ready microscopy slides were

visualized in a confocal microscope, able to capture different wavelengths (channels). The results are presented in Figure 14 A.
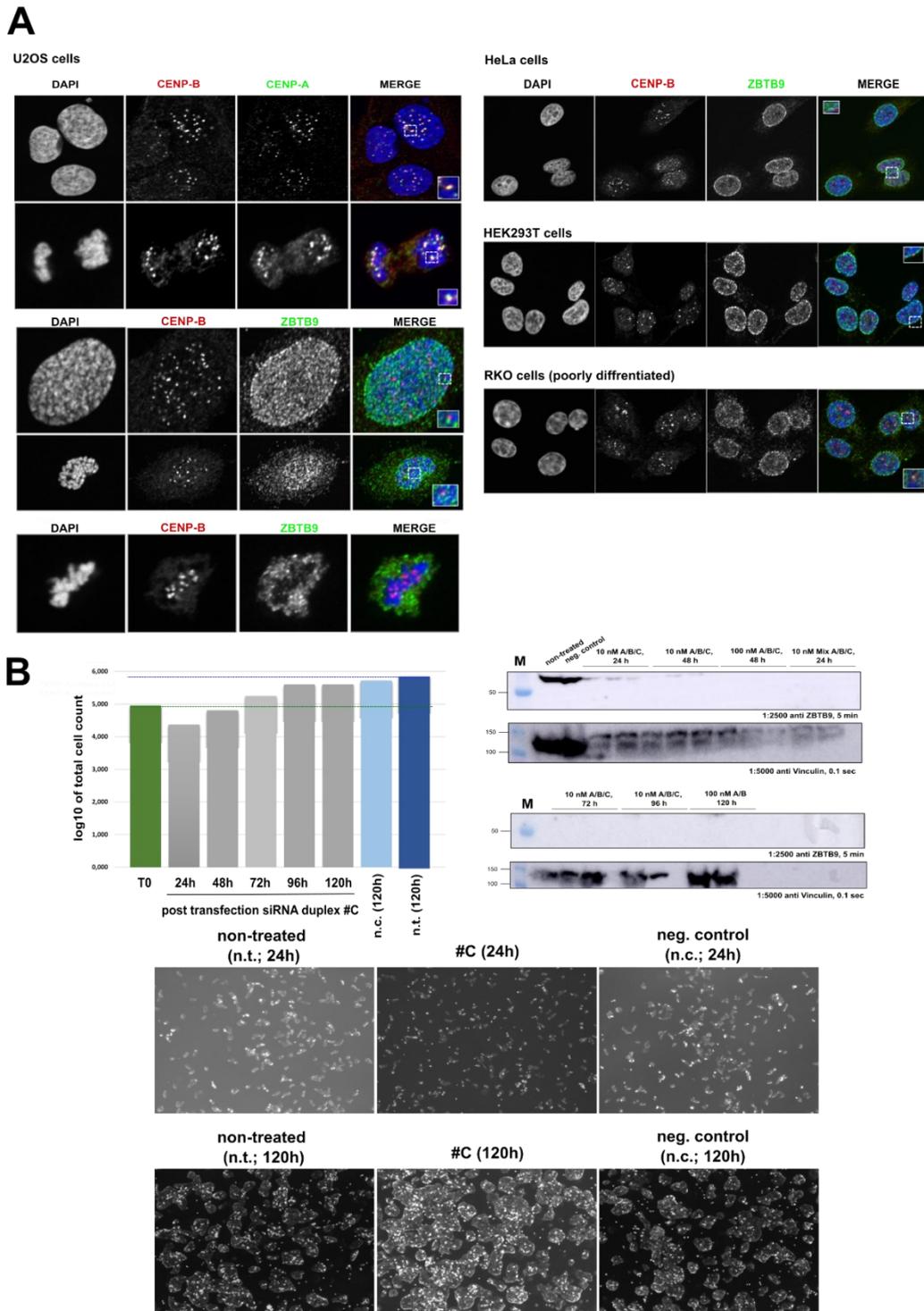

Figure 14: **Co-staining of CENP-A with CENP-B and ZBTB9 and CENP-B in human cells and knockdown experiments with ZBTB9.** The microscopy data for co-localization of CENP-B to

ZBTB9 or CENP-B to CENP-A as well as the knock down microscopy data were performed by Jeannette Koch (AG Leonhardt) **A)** Establishment of the co-staining between CENP-A and CENP-B with specific Antibodies in U2OS cells or ZBTB9 with CENP-B in U2OS, HeLa, HEK293T and RKO cells. Nuclei stained with DAPI. Co-staining were enhanced for some spots (red squares). **B)** Knockdown of ZBTB9 (siRNA duplex kit, Origen) in HeLa cells at specific timepoints and their morphological appearance after treatment. As a control non-treated HeLa cells as well as a transfection negative control were chosen. Cells were collected after treatment and a western blot was performed to confirm the loss of ZBTB9 during the treatment. Vinculin was chosen as loading control.

As presented in the microscopy overviews, the co-localization of CENP-B and CENP-A as reported in numerous publications could be confirmed. Two stages of the cell cycle in U2OS cells are shown (upper panel in interphase, lower panel in ana-/telophase, white rectangle). CENP-B and CENP-A are always co-localized and therefore associated with each other. Next a co-staining of ZBTB9 and CENP-B was prepared. Here after several optimization rounds, the staining of ZBTB9 could be visualized. The results are shown in Figure 14 A second panel. At different cell cycle stages ZBTB9 (here Prophase and Pro-metaphase) is only very seldomly co-localized to CENP-B. Interestingly, ZBTB9 shows a staining pattern towards the nuclear membrane. To verify this co-staining to the nuclear membrane, other cell lines (HeLa, HEK293T and RKO; Interphase) were tested (Figure 14 A right panel). The staining of ZBTB9 to the nuclear membrane could be confirmed. Additionally, some co-localization spots between ZBTB9 and CENP-B could be determined (see white rectangles and enhanced image at the bottom or top of each merged picture). The staining of ZBTB9 revealed an interesting staining pattern, where the nucleolus is left out. Therefore, suggesting that ZBTB9 is not located within the nucleolus. In comparison to the co-localization signal seen in CENP-A/CENP-B (upper panel), one cannot draw the conclusion that ZBTB9 might be an exclusive novel centromere protein. It seems ZBTB9 has a more complex role within the nucleus and one of it being at the centromere, which have been caught in the CENP-A ChIP-MS experiments and now have been described in this study with biochemical, XL-MS and biophysical data. An interesting observation was caught in the last panel (Figure 14 A lowest panel). It seems the chromosomes are aligned to the metaphase plate (mitotic stage: metaphase) and ZBTB9 is located around the DNA, whereas CENP-B (here marking the centromere) is located within the metaphase chromosomes, implying that ZBTB9 has an additional role in chromatin condensation process.

Since localization of ZBTB9 could be established it was of interest what phenotype the cells would show if ZBTB9 was knocked down. Here a time course experiment with three different siRNA duplexes (#A, #B, #C; ZBTB9 Kit, Origen) at different concentrations (10 nM, 100 nM) were conducted (Figure 14 B). A bar graph shows the log10 of total cell count to each time point (n=1). The green dotted line implicates the starting cell count at timepoint T0, whereas the blue dotted line shows the value for the non-treated cells after 120h. The western blot at the right panel shows an immediate decrease of ZBTB9 after 24 h incubation with all tested duplexes. A mixture of all duplexes at a concentration of 10 nM already resulted in a complete depletion of ZBTB9s western blot signal. After 72 h up to 120 h ZBTB9s western blot signal cannot be detected anymore. In addition, the total cell count was calculated to see if the treatment leads to an alteration in living cells. Indeed, in the first 24 h the total cell count for all tested duplexes decreased (log10 of total cell count was at 3.8) and can also be confirmed by a microscopic shot of the cells in the well (here only presented treatment with duplex #C as representative). Compared to not treated cells or treated with the negative control, most of the cells are detached or apoptotic. Upon continuing the treatment, the total cell count increases again and results at timepoint 120 h in a slight decreased total cell count (value 5.60) than compared to the grown non treated cells (value 5.86). Comparing these data with the microscopic data, the cells seem to recover and were growing. One difference between non-treated and the treatment with duplex #C resulted in a more cluster like formation of the cells. It seems that the knock down of ZBTB9 has an immediate effect on the cell, but it is not a dramatic one, since the majority of the cells are still alive and can cope with the absence stress of ZBTB9 by other yet unknown mechanisms or alternative pathway. The same analysis was performed for the other duplexes, and all showed the same phenotype. Unfortunately, an immune-fluorescence staining after ZBTB9 knock-down resulted in inconclusive staining patterns for ZBTB9 (data not shown), even though western blot data suggest near complete reduction of ZBTB9 (Figure 14 B). Therefore, this experiment has to be reperformed to describe the phenotype upon ZBTB9 knockdown and subsequently the faith of the cell.

Taken all presented evidence together provided for the *in vivo* role of ZBTB9 it shows that ZBTB9 is not exclusively co-localized to the centromere and shows staining pattern near the nuclear lamina (might be associated with heterochromatin) and as seen in mitotic phases, might be involved in chromatin condensation. In addition, no staining of ZBTB9 at nucleolus level could be determined. First knock-down experiments showed a subtle phenotype, where after 24 h treatment the cells seem to be detached and apoptotic.

Nonetheless, this phenotype cannot be seen in later time courses since the cells seem to recover and gain an increase in total cell count. Moreover, the cells seem to cluster more when ZBTB9 was depleted. These findings have to be further validated. As an additional characterization of ZBTB9, a nuclear fishing pulldown was performed to characterize potential binding partners of ZBTB9 within the nucleus.

5.1.4 ZBTB9 can be associated with diverse functions in the cell

ZBTB9 has not been yet described in depth regarding its interactors or its role in the cell. Efforts have been made to establish a FLAG-tagged stably expressing ZBTB9 wt or mutant HeLa T-Rex FlpIN cell line. It was possible to establish such a FLAG-cell line expressing ZBTB9 wt (identifier 4F03) and mutant (delta 380-462, identifier 4F04), but the extraction procedure used for CENP-A Pulldowns (doctoral thesis, Dr. Götz Norman Hagemann, 2020, LMU) could not be optimized for ZBTB9. In future attempts this protocol has to be adapted to ensure sufficient solubilisation of ZBTB9 for subsequent *in vivo* Pulldown analysis (Data not shown). Since the *in vivo* Pulldown of ZBTB9 was not straight forward, the idea was to obtain insights on binding partners of ZBTB9 by incubating both structurally well-defined domains (BTB and Zinc finger region) with human nuclear extract. Both domains were expressed in *E.coli* and subsequently purified. For this purpose, the structured part as judged by the alphafold model of ZBTB9 (Identifier AF-Q96C00-F1) were fused to an N-terminal MBP tag with a C-terminal 8xHis. For the BTB domain amino acid 19 to 112 were cloned and expressed. In case for Zinc finger region, the annotated zincfinger as well as the predicted third zinc finger, as seen in alphafold, was cloned (aa 380-473) (Figure 15).
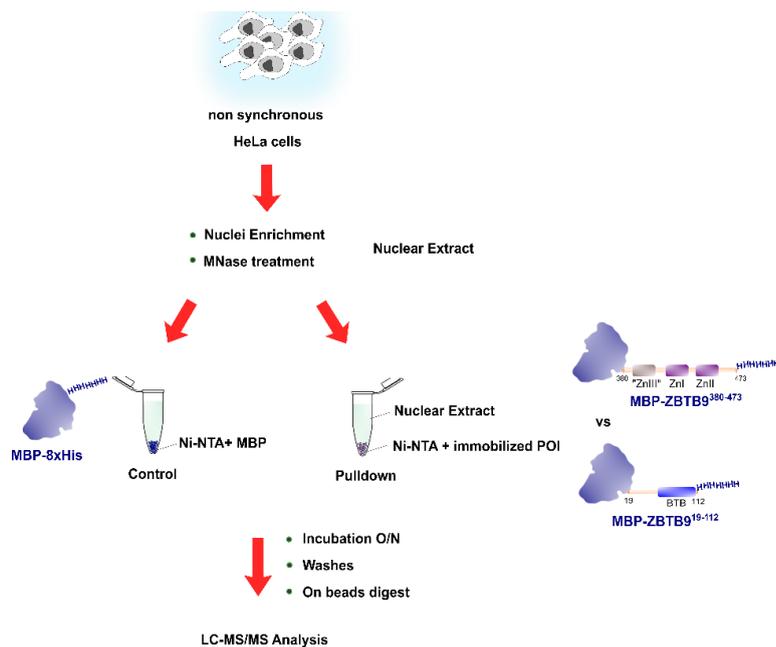
Figure 15: **Experimental Setup for the Nuclear Fishing Pulldown (NFPD).** The Graphic depicts how the Nuclear Fishing Experiments were conducted. A MNase digested Nuclear Extract was obtained from non-synchronous HeLa cells. The MNase was stopped with addition of EGTA. The Nickle-NTA beads were equilibrated with Buffer. Protein/control were incubated with beads for one hour. Afterwards non-bound protein was washed away and the beads with protein of interest were

incubated with Nuclear Extract over-night. After incubation, beads were washed, and protein complexes were digested and subsequently analysed by LC-MS/MS.

The linker region of ZBTB9 was not used as bait, since alphafold predicts it as disordered regions. By doing so, by incubating the domains with the nuclear extract, proteins or complexes naturally interacting with the BTB or the zinc finger domains shall remain bound and can be determined by MS. The experimental setup is depicted in Figure 15. Each Experiment was performed in three independent biological replicates and each biological replicate was measured in two independent technical replicates. The raw files were searched with MaxQuant. Resulting txt files for protein groups, evidence and peptides were taken as input parameter to perform statistical analysis in R (R- Statistical Analysis and DAVID Analysis was performed with help and by Dr. Victor Solis-Mezarino; STRING and partial DAVID Analysis was performed by Chandni Kumar). The Intensity Based Absolute Quantification (iBAQ) values were extracted and contaminants and proteins with low identification scores were filtered out. After subtracting the control background from the interaction background of each domain, the abundance of each protein in the set was calculated as fold change (log2) and the significance was calculated with a statistical t-test to obtain adjusted p-values (later on transformed to -log10 values). This statistical procedure enabled to identify with different FDR cut offs significant interacting partners of both domains and therefore paints a picture with whom ZBTB9 might interact and in which biological processes it is involved and proceeds its molecular functions. Proteins passing the threshold of 0.01-0.1 % FDR were taken as true interactors for each domain. To obtain a complete picture of ZBTB9, each significant interactors surpassing the threshold of 0.1% FDR were combined in the bioinformatic STRING and DAVID Analysis.

The total analysis of both pulldowns identified 1350 proteins in each pulldown. In the BTB-NFPD in total 160 proteins have been identified with an FDR ≤ 0.1 %, whereas in the Zinc finger-NFPD in total 203 proteins with an FDR ≤ 0.1 % could be identified. A short summary of identified proteins with different FDR cut offs is presented in the following table (Table 3). In this FDR cut offs are not included the identification of CENP-B and CENP-F. CENP-B was found in the Zinc finger-NFPD but with an FDR of 11% whereas CENP-F could be identified with an FDR of 23%. Moreover, CENP-B was enriched in the zinc finger pulldown, whereas in BTB it was not detected. Unfortunately, the control with the MBP tag showed in two of the three biological replicates the identification of CENP-B and CENP-F. Therefore, the *in vitro* assays in the previous paragraphs were partially confirmed by the NFPD but with low confidence.

Table 3: Significantly enriched proteins identified in the NFPD of the BTB or Zinc-finger domain of ZBTB9. At different FDR cut offs total significant enriched proteins are listed.

FDR cut off in %	Significantly identified proteins in the BTB-NFPD	Significantly identified proteins in the ZN-NFPD
≤ 0.01	74	113
≤ 0.10	160	203
≤ 1.00	269	361
≤ 5.00	385	612
	Total identified 1350	Total identified 1350

As a first analysis step it was of interest, whether the identified proteins (≤ 0.10 % FDR or p adjusted value 0.001) could be found in the initial ChIP-MS CENP-A pulldown. Indeed 22 proteins were in common with the original dataset (see Table 4).

Table 4: **Identified proteins within the Nuclear Fishing Experiments compared to the ChIP-MS CENP-A Pulldowns.** The most significant enriched proteins in the CENP-A Pulldown were compared to either the BTB or Zinc finger Nuclear Fishing pulldown. The p-adjusted value for each identified protein in the nuclear fishing pulldown (n=3) is stated as well. A short UniProt description for each protein is given as well.

Identified in ChIP-MS CENP-A Pulldown	Identified either in BTB or Zn NFPD	p adjusted value in NFPD	UniProt Description
SMARCA5	yes	≤ 0.001	Helicase, catalytic subunit of ISWI chromatin complexes
NPM3	yes	≤ 0.001	Involved in ribosome biogenesis, chromatin remodelling, protein chaperoning
DDX21	yes	≤ 0.01	RNA helicase, sensor of transcriptional status of RNA Polymerase I and II
NPM	yes	0.442	Involved in ribosome biogenesis, protein chaperoning, histone assembly
SMARCA1	yes	0.049	Component of SWI/SNF chromatin complex (npBAF), involved in transcriptional activation/ repression
ATRX	yes	0.048	Transcriptional regulation, facilitates DNA replication, catalytic component of chromatin remodeling complex ATRX:DAXX required for H3.

			Deposition at pericentric DNA repeats
UHRF1	yes	0.0050	Key epigenetic regulator bridging DNA methylation and chromatin modification
DEK	yes	0.983	Involved in chromatin organization
DDB1	yes	0.048	Involved in DNA repair and protein ubiquitination
JUNB	yes	≤ 0.001	Transcription factor
UHRF2	yes	0.0032	E3 ubiquitin ligase, plays role in DNA methylation, histone modification, cell cycle and DNA repair
CENP-B	yes (specific in Zinc finger)	0.118	Interacts with centromeric heterochromatin, binds specific 17 bp subset of alphoid satellite DNA (CENP-B box)
SSRP1	yes	0.041	Component of FACT complex, facilitates RNAPII transcription
SP16H	yes	0.0018	Component of FACT complex, facilitates RNAPII transcription
RSF1	yes	≤ 0.001	Regulatory subunit of RSF1 and RSF5 ISWI chromatin remodeler complex, facilitate DNA access during DNA replication, transcription and repair
RBBP4	yes	0.017	Core histone-binding subunit
RBBP7	yes	0.018	Core histone-binding subunit
CHD1	yes	0.237	ATP-dependent chromatin-remodelling factor as substrate recognition component of the transcription regulatory histone acetylation complex SAGA
CHD2	yes	0.268	DNA-binding helicase specifically bind to promoter of target genes
CHD8	yes	0.573	DNA helicase acts as chromatin remodelling factor and regulates transcription
CHD5	yes	0.951	Chromatin-remodelling protein binding DNA through histones and regulate gene expression

As provided in Table 4, 22 proteins significantly identified in the CENP-A proteome were in common with the ZBTB9 NFPD at different p-values observed. Some of these proteins are

involved in transcriptional regulation (e.g. JUNB or ATRX) or chromatin remodelling (members of the FACT complex, RSF-5 complex or CHD1,2,8,5). An interesting observation was encountered for NPM3 and NPM which are besides chromatin remodelling also involved in ribosome biogenesis. This implies, that ZBTB9 indeed has binding partners and complexes in common with the CENP-A proteome. Moreover, some of these were not associated with the centromere before (e.g. NPM3, DDX21, UHRF1 or 2 or RBBP4/ 7), expanding the complexity of the centromeric organization. Since some of the CENP-A associated proteins were found in the NFPD again, it was of interest, whether the BTB and Zinc finger domain share same protein partners, and which are unique to each domain. Therefore, the NFPDs were compared (p-adjusted value 0.0001 or 0.01% FDR) and in Table 5 a comprehensive summary of proteins in common in both NFPDs is given with their short respective UniProt description. In total 49 protein ID were in common with both NFPDs. From 74 proteins (≤ 0.01 % FDR) 25 were unique to BTB and in case for the zinc fingers, from 113 proteins (≤ 0.01 % FDR) 62 were unique for the zinc finger region.

Table 5: **Proteins identified common in both ZBTB9 NFPD identified (p adjusted value ≤ 0.0001) or for each domain uniquely with a short UniProt description.**

Identified proteins in both NFPD similarly	UniProt description	Identified proteins in both NFPD similarly	UniProt description
5NTD	Hydrolyses extracellular nucleotides into membrane permeable nucleosides	LUZP1	Involved in actin and cytoskeleton stabilization
ACL6A	Transcriptional activation/repression, component npBAF complex or INO80 complex	LYAR	Transcription regulation
ACL6B	Transcriptional activation/repression, component npBAF complex or INO80 complex	MCM5	Component of MCM2-7 complex, replicative helicase for DNA replication
CALL3	May function as specific light chain of myosin-10	MYH7B	Movement of organelles along Actin filament
CNTP5	Involved in cell adhesion, intercellular communication	MYO1A	Movement of organelles along Actin filament
CPSF6	Component of CFIm complex, activator of pre-	NKRF	Enhances ATPase activity of DHX15

Identified proteins in both NFPD similarly	UniProt description	Identified proteins in both NFPD similarly	UniProt description
	mRNA 3'-end cleavage and polyadenylation process		
CRX	Transcription factor	NPM3	Involved in ribosome biogenesis, chromatin remodelling, protein chaperoning
CTR9	Component of PAF1C complex, required for RNAP II transcription, involved in histone modification	P3C2A	Role in several intracellular trafficking events
DHX8	Involved in pre-mRNA splicing as component of spliceosome	PCNP	May be involved in cell cycle regulation
FA98A	Stimulates PRMT1-induced protein arginine methylation	PPHLN	Component of HUSH complex, epigenetic repression
GELS	Actin-modulating protein binding to plus ends of actin monomers or filaments, Calcium-regulated	PPIB	Assisting in protein folding (PPIase catalyzing cis-trans isomerization of proline imidic peptide bonds)
H2A2B	Histone core component	PRKDC	Serine/Threonine-protien kinase, molecular sensor for DNA damage
Identified proteins in both NFPD similarly	**UniProt description**	**Identified proteins in both NFPD similarly**	**UniProt description**
HDAC2	Histone deacetylase, transcriptional regulation	PRP4	Pre-mRNA splicing component of U4/U6-U5 tri-snRNP complex
HDGF	Transcriptional repressor	PRP6	Pre-mRNA splicing component of U4/U6-U5 tri-snRNP complex
HDGR3	Role in cell proliferation, enhances DNA synthesis	PSMD2	Component of 26S proteasome, ubiquitin degradation pathway
KHDR2	RNA-binding protein, plays a role in alternative splicing	RBM23	RNA-binding protein, acting as transcription coactivator and pre-mRNA splicing factor
LMNB2	Component of nuclear lamina	RHGBB	Promotes development and evolutionary expansion of the brain neocortex

Identified proteins	UniProt description	Identified proteins	UniProt description
PRP6	Pre-mRNA splicing component of U4/U6-U5 tri-snRNP complex	RHOJ	Plasma membrane associated small GTPase
PSMD2	Component of 26S proteasome, ubiquitin degradation pathway	RHOQ	Plasma membrane associated small GTPase
PTBP3	RNA-binding protein, mediating pre-mRNA alternative splicing regulation	SAFB2	Binding to scaffold/matrix attachment region DNA
RBM15	Key regulator of N6-methyladenosine methylation of RNAs, involved in X-chromosome inactivation	SMCA5	Helicase, catalytic subunit of ISWI chromatin complexes
SON	mRNA splicing cofactor efficient splicing of transcripts with weak splice sites (e.g. Aurora B, Akt1)	TOM20	Protein import into mitochondria
SREK1	Regulation of alternative splicing	TOP1M	Supercoiling and torsional tension release of DNA during mitochondrial DNA replication
SUZ12	Polycomb group (PcG), component of PRC2 complex, methylates H3K9 and H3K27 leading to transcriptional repression	TPM4	Binding to actin filaments, stabilizing cytoskeleton actin filaments
TBAL3	Constituent of microtubules	TPX2	Spindle assembly factor, required for mitotic spindle assembly
ZFR2	Not known		
Identified proteins unique to BTB NFPD	**UniProt description**	**Identified proteins unique to BTB NFPD**	**UniProt description**
ABC3A	DNA deaminase, may have a role in epigenetic regulation of gene expression through process of active DNA demethylation	IF4A1	ATP-dependent RNA helicase, involved in cap recognition and required for mRNA binding to ribosome
CBX3	Involved in transcriptional silencing in heterochromatin-like complexes, binds to	MOFA1	Not known

	H3K9me, involved in formation of functional kinetochore though interaction with Mis12 complex, mediates recruitment of methyltransferase (SUV39H1/2)		
CHMP3	Probable component of ESCRT-III complex, involved in late cytokinesis	MYH14	Plays a role in cytokinesis, cell shape
CLUS	Function as extracellular chaperone	ODF2L	Localizes to centrioles, acts as suppressor of ciliogenesis
CNOT1	Scaffoling component of CCR4-NOT complex, major cellular mRNA deadenylases	P210L	Component of the nuclear pore membrane
COA3	Core component of MITRAC (mitochondrial translation regulation assembly intermediate of cytochrome c oxidase) complex	P4HA1	Catalyses PTM of 4-hydroxproline in collagen and other proteins
DCAF7	In association with DIAPH1 controls GLI1 transcriptional activity	RL10A	Component of large ribosomal subunit (60S)
DJB11	Co-chaperone for HSPA5	RS15	Component of 40s ribosome
DNJA3	Modulates apoptotic signal transduction or effector structures within mitochondrial matrix	SMCA2	Component of SWI/SNF chromatin remodeling complexes, involved in transcriptional activation/ repression
GTD2A	Not known	TBB1	Constituent of microtubules
TIM13	Mitochondrial intermembrane chaperone		
UN13B	Plays a role in vesicle maturation during exocytosis		
YTHD1/2/3	Binds and recognizes N6-methyladenosine containing mRNA and regulates their stability		

BAF	Plays a role in mitotic nuclear assembly, chromatin organization, DNA damage response, gene expression		

Identified proteins unique to Zinc finger NFPD	UniProt description	Identified proteins unique to Zinc finger NFPD	UniProt description
ABC3C	DNA deaminase, may have a role in epigenetic regulation of gene expression through process of active DNA demethylation	AN32E	Histone chaperone, mediates removal of histone H2A.Z from nucleosomes
ABC3D	DNA deaminase, may have a role in epigenetic regulation of gene expression through process of active DNA demethylation	AP2A2	Component of adaptor protein complex 2, function in protein transport
ABC3F	DNA deaminase, may have a role in epigenetic regulation of gene expression through process of active DNA demethylation	AP2M1	Component of adaptor protein complex 2, function in protein transport
ADT4N	ADP:ATP antiporter mediating import of ADP into mitochondrial matrix for ATP synthesis	ARF	Acts as tumor suppressor, can induce G1- and G2-arrest
AN32A	Multifunctional protein, involved in cell cycle progression, transcription or apoptosis	CSK22	Catalytic subunit of constitutively active serine/threonine-protein kinase complex, during mitosis function as p53/TP53-dependent spindle assembly checkpoint maintaining CKD-cyclinB activity
AN32B	Not known, might be similar to AN32A	CSK23	Catalytic subunit of constitutively active serine/threonine-protein kinase complex, during mitosis function as

			p53/TP53-dependent spindle assembly checkpoint maintaining CKD-cyclinB activity
CDC42	Plasma membrane associated small GTPase, regulates bipolar attachment of spindle microtubules to kinetochores before chromosome congression in metaphase	CSK2B	Catalytic subunit of constitutively active serine/threonine-protein kinase complex, during mitosis function as p53/TP53-dependent spindle assembly checkpoint maintaining CKD-cyclinB activity
CDK1	Key role in eukaryotic cell cycle	CTBL1	Component of PRP19-CDC5L complex, integral part of spliceosome
CDK12	Kinase phosphorylates C-terminal domain of RNAPII, kex regulator of transcription elongation	CYTD	Cysteine protease inhibitor
CDK13	Similar as CDK12 and also required for RNA splicing by phosphorylating factors	DHX30	RNA-dependent helicase
CDK14	Kinase, acts in cell cycle	EIF2A	Functions in early steps of protein synthesis of small number of specific mRNAs
CDK15	Kinase	FBX3	Substrate recognition component of SCF (Skp1-Cul1-F-box protein)-type E3 ubiquitin ligase complex
CDK16	Kinase	FOXG1	Transcription repression factor
CDK17	Kinase	FOXL1	Transcription factor
CDK18	Kinase	GATA1	Transcription activator/ repressor
CDK4	Kinase, regulates G1-S-Transition	GATA2	Transcription activator/ repressor
CDK5	Kinase	GATA4	Transcription activator/ repressor

CDK6	Kinase, regulates G1-S-Transition	GATA5	Transcription activator/repressor
CIAO1	Key component of cytosolic iron-sulfur protein assembly complex, mediating incorporation of iron-sulfur clusters into extramitochondrial Fe/S proteins	GRSF1	Regulator of post-transcriptional mitochondrial gene expression
CMS1	Not known	ICA1L	Not known
CNBP	Protein binding to single-stranded DNA	ICA69	Not known
CSK21	Catalytic subunit of constitutively active serine/threonine-protein kinase complex, during mitosis function as p53/TP53-dependent spindle assembly checkpoint maintaining CKD-cyclinB activity	MSI1H	RNA binding protein regulates expression of target mRNAs at translational level
MSI2H	RNA binding protein regulates expression of target mRNAs at translational level	RL22L	60S ribosomal subunit
PB1	Involved in transcriptional activation and repression of selected genes by chromatin remodleling, required for stability of SWI/SNF pBAF remodeling complex	RSF1	Regulatory subunit of RSF-1 and RSF-5 ISWI chromatin remodeler complex, facilitate DNA access during DNA replication, transcription and repair
PELP1	Coactivator of estrogen receptor-mediated transcription and a corepressor of other nuclear hormone receptors and sequence-specific transcription factors	SETLP	Transcriptional activator
PPM1G	Protein Phosphatase	SMCA4	Involved in transcriptional activation and repression of selected genes by chromatin remodelling,

			component of SWI/SNF npBAF
RASA1	Inhibitory regulator of Ras-cyclic AMP pathway	SP9	Transcription factor
SRSF4	Plays role in alternative splicing		
STRAP	Involved in building blocks of spliceosome (snRNAPs)		
TEN4	Transmembrane protein		
TRI26	E3 ubiquitin-protein ligase		
ZCH10	Not known		

In the upper half of Table 5 all proteins are in common between the NFPD of the BTB and the zinc finger. Many proteins are involved in transcriptional regulation (CRX, LYAR, HDGF or HDAC2) and provides an insight on how ZBTB9 might be involved in gene transcription. An interesting observation is the enrichment of mRNA splicing factors or component of the spliceosome. This might indicate that ZBTB9 is involved in mRNA splicing events at the (peri)centromere. Moreover, ZBTB9 seems to be involved in chromatin organization by associating with different chromatin remodelling complexes (SMCA5, ACL6A/B (INO80 complex or npBAF complex) or NPM3) or DNA replication machinery (MCM5). In addition, ZBTB9 seems to be involved in maintaining heterochromatin marks by interacting with components of this process (SUZ12, RMB15 or PPHLN). Surprisingly, ZBTB9 also interacts with proteins involved in actin modulation, microtubule binding or spindle assembly factor required for mitotic spindle assembly (GELS, TBAL3 or TPX2). Hence, ZBTB9 might also be important during mitosis. This observation is partially supported by the *in vivo* microscopy data (Figure 14 A), where ZBTB9 is localized around the chromosomes arranged at the metaphase plate.

The second half of Table 5 describes the IDs associated specifically with each domain and following the UniProt description, the majority of the proteins are involved either in transcription activation/ repression (e.g., GATA1-4) or are involved in heterochromatin/ pericentromeric chromatin formation by interacting with deacetylases or readers of specific H3K9me and H3K27me mark, which is an important repressive mark (e.g., CBX3 or PB1) for heterochromatin and represses transcription. Moreover, ZBTB9 seems to be involved in ubiquitin-mediated degradation pathway by interacting with FBX3 or TRI26. As already mentioned in the previous paragraph, ZBTB9 is also involved in mRNA splicing activity and

might bind to ribosomal subunits and alter translational activity (RL22L, IF4A1, RL10A, RS15).

Large datasets generated by ChIP-sequencing or Affinity-IP-MS-based experiments makes it difficult to validate each protein individually. Therefore, it is important to have bioinformatical databases providing functional analysis of these identified proteins and a broader picture in which relation the bait is set to the identified proteins. Furthermore, the bioinformatic analysis might reveal insights in which pathways or biological processes it is mainly involved. In case for the NFPDs the dataset was analysed with the STRING and the DAVID bioinformatics database.

The STRING-Analysis provided a first insight in which biological processes ZBTB9 could be involved. To verify this finding in the STRING Analysis, a second bioinformatic analysis (**D**atabase of **A**nnotation, **V**isualization, and **I**ntegrated **D**iscovery-Analysis, DAVID) was performed [441,442].

As STRING and DAVID-Analysis input, the most significant protein interactors of each NFPD (FDR cut off 0.1% of each domain) were taken. The STRING Analysis (Figure 16) shows the clustering of six biological functions or molecular functions associated with the interactors of ZBTB9 (FDR < 0.1 % were taken as top functions). STRING is a database of known and predicted protein-protein interactions and provides organism-wide protein association networks. This includes the direct (physical) and indirect (functional) association. Data sources for STRING are automated text mining from scientific literature, computational interaction predictions form co-expression, high-throughput experiments, and genomic databases[443,444]. The complete STRING network was divided into six different clusters (MCL clustering option) and additionally the dataset was subjected to DAVID-Analysis to confirm the STRING annotated pathways but also was used to generate heatmaps of each biological process or molecular function (based on Gene Ontology) ZBTB9 might be involved in (Figure 17-21).

The STRING analysis shown in Figure 16 gave six clusters. The first cluster with red nodes shows interactors involved in RNA splicing, mRNA Processing or Transcription (e.g., SF3B2, HNRNPAB, CHERP, FUBP1). The second cluster (green node) points towards function in Translation or nonsense-mediated decay or SRP-dependent co-translational protein targeting (e.g., RPS3A, RPL27, EIF2S1). The yellow cluster (cluster 3) has interactors involved in Actin Filament based processes or Cytoskeleton organization (e.g.,

RAC1, LMNA, ACTG1). The next cluster (light cyan, cluster 4) in conjunction with cluster five (blue) and six show proteins involved in Transcription with RNA Polymerase II (RNAPII) with DNA-binding transcription activator activity (e.g., CDK9, HDAC1, SSRP1, XX1, PARP1) as well as regulation of gene expression (e.g., POU3F1/2/3/4, POU6F1, POU2F1/2/3). As a confirmatory approach, the same IDs were subjected to DAVID analysis. The most popular biological processes and molecular functions (with their respective protein IDs) in which ZBTB9 might be involved were taken as input and a heat map with a raw log2 scale was generated (Figure 17-21). The colour scale shows when a particular protein is enriched in either the control or the BTB/Zinc finger NFPD. The lighter the colour scale (light green to yellow), the less likely it is enriched in the given pulldown, likewise the other way around with a darker colour scale (here dark blue/ purple) the protein ID in its respective NFPD is highly enriched. By providing many IDs, a cluster can be visualized on the heatmap and which domain of ZBTB9 is more likely to be involved in the respective molecular function or biological process. The first impression of the DAVID analysis confirms the STRING analysis. The identified protein IDs for each domain draw a picture of what ZBTB9 might be involved in. Here again the top terms were associated with RNA mediated processes (RNA Polymerase II regulation, mRNA splicing events or spliceosome association), Cell cycle regulation or transcriptional coupled processes.
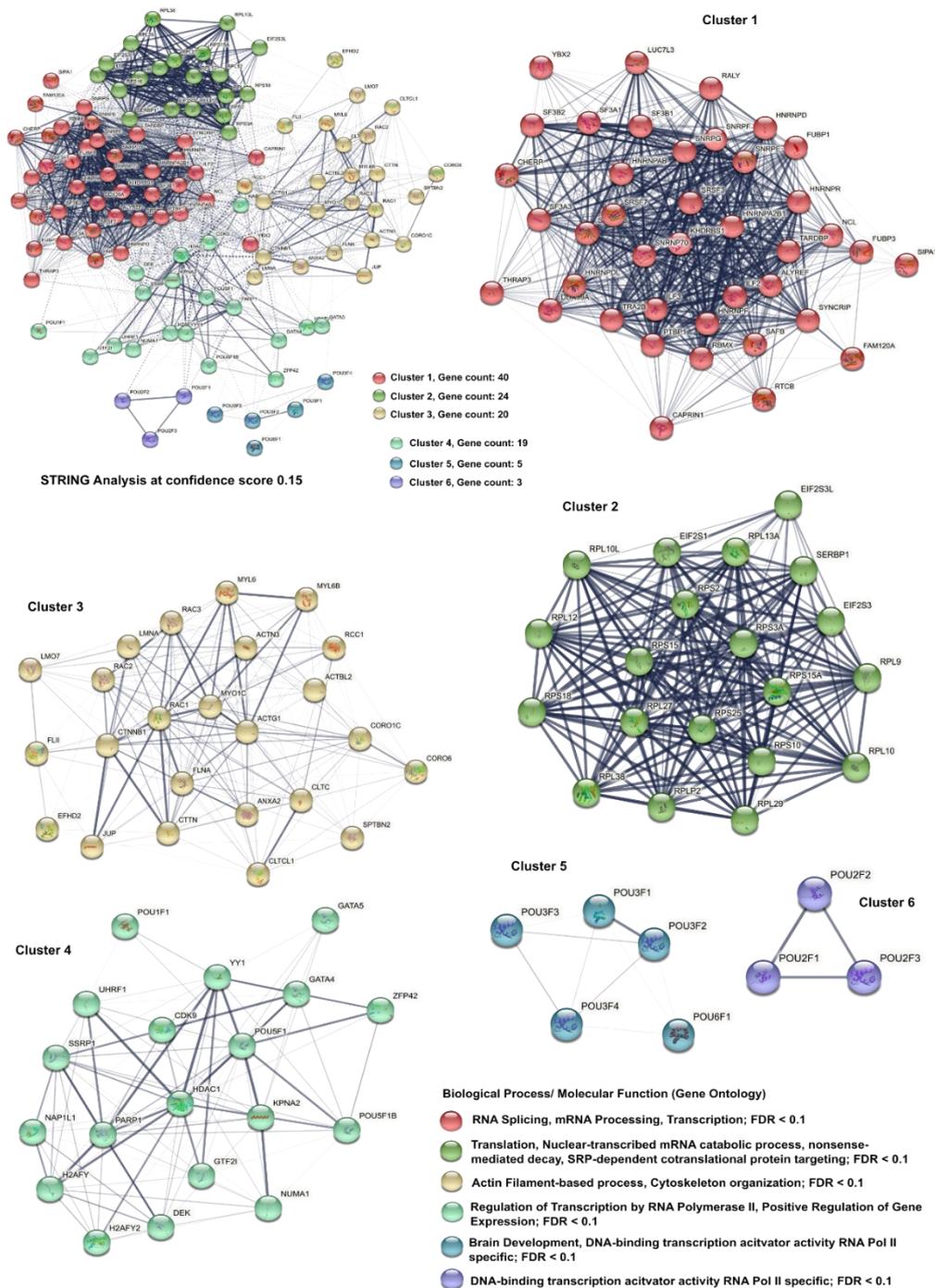
Figure 16: STRING Analysis of proteins identified in the Nuclear Fishing Experiments of the BTB and the Zinc finger Domain of ZBTB9. Proteins enriched for either BTB or Zinc finger domain with a cut off of 0.1% FDR were taken and put as input for the STRING analysis. The interaction score was set to 0.15. The clustering was performed with the settings of MCL clustering option. The most prominent biological processes or molecular functions with an FDR score of < 0.1 were taken.

Shown are p-values corrected for multiple testing within each category using the Benjamini-Hochberg procedure
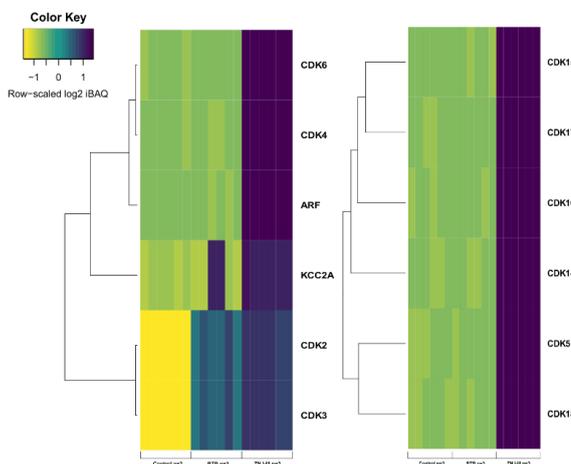
DAVID Analysis of protein interaction with ZBTB9 BTB and Zn Domain

DAVID Analysis: Biological Process

1) Regultion of Transcription from RNA Polymerase II Promotor
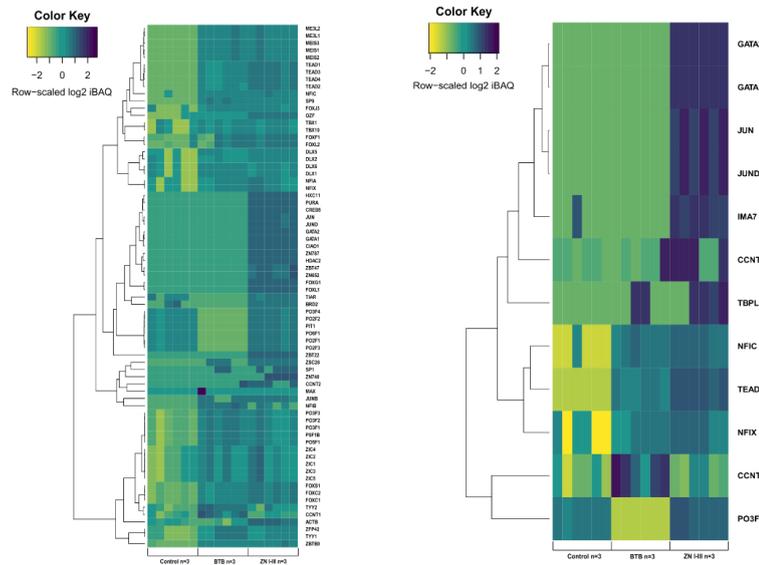
2) Regultion of Transcription involved in G1/S-Transition
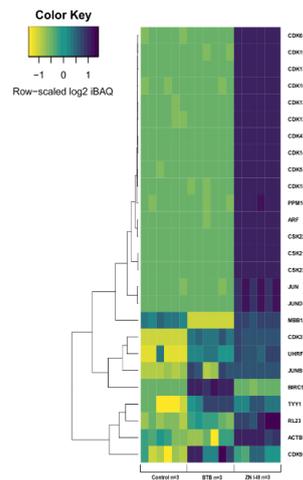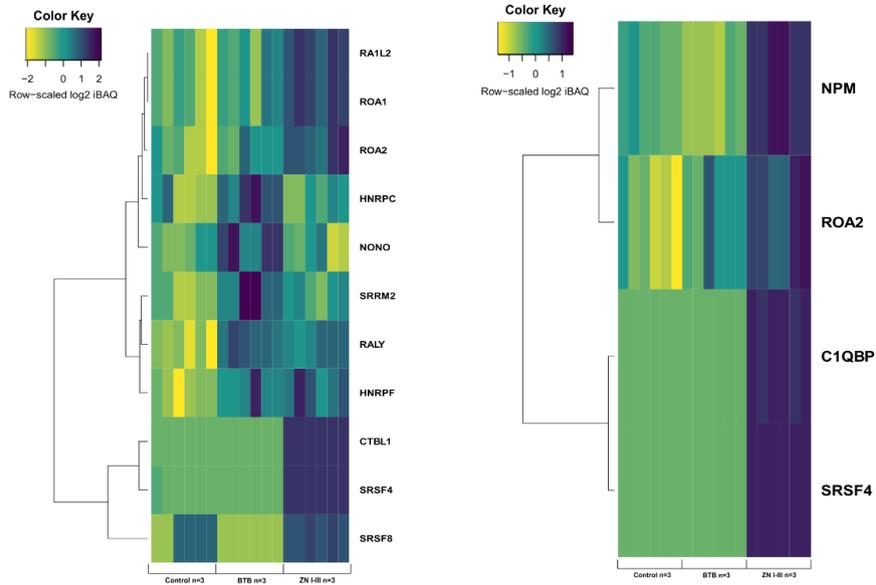
3) Regultion of Cell cycle

Figure 17: **Bioinformatical Analysis via DAVID of relevant biological processes involved in binding partners of ZBTB9.** Proteins enriched for either BTB or Zinc finger domain with an FDR cut off 0.1% were taken as input. Top biological processes and associated proteins were taken and

mapped with their corresponding log2 iBAQ value to obtain a heatmap corresponding to the DAVID-Analysis.

DAVID Analysis of protein interaction with ZBTB9 BTB and Zn Domain

DAVID Analysis: Biological Process

4) Regultion of mRNA splicing via splicosome

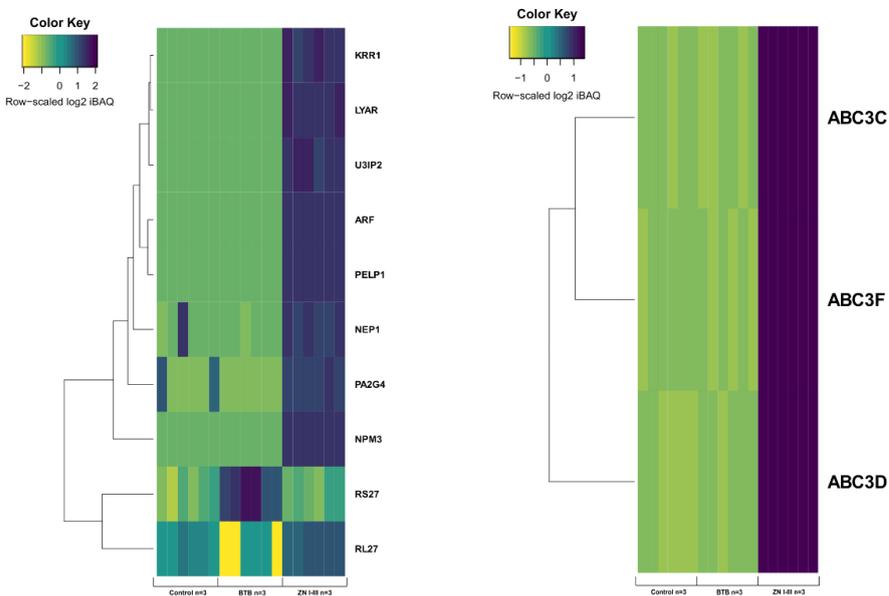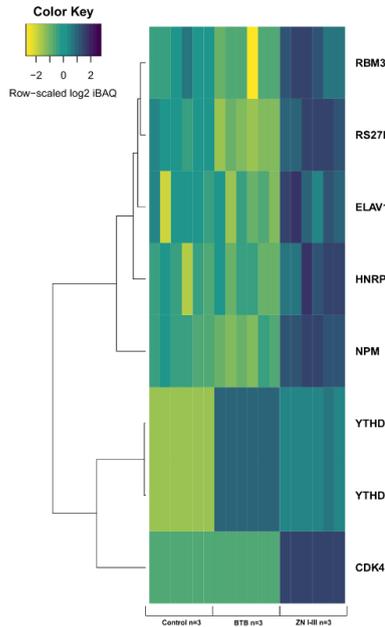**5) rRNA processing and cytidine to uridine editing**

Figure 18: **Bioinformatical Analysis via DAVID of relevant biological processes involved in binding partners of ZBTB9.** Proteins enriched for either BTB or Zinc finger domain with an FDR cut

off 0.1% were taken as input. Top biological processes and associated proteins were taken and mapped with their corresponding log2 iBAQ value to obtain a heatmap corresponding to the DAVID-Analysis.

DAVID Analysis of protein interaction with ZBTB9 BTB and Zn Domain

DAVID Analysis: Biological Process

6) Regulation of Translation

7) Microtubule cytoskeleton organization

8) Hippo Signaling Pathway

9-12) Import into nucleus
Nuclear migration/ Nuclear pore localization
Heterochromatin assembly

***Figure 19*: Bioinformatical Analysis via DAVID of relevant biological processes involved in binding partners of ZBTB9.** Proteins enriched for either BTB or Zinc finger domain with an FDR cut off 0.1% were taken as input. Top biological processes and associated proteins were taken and mapped with their corresponding log2 iBAQ value to obtain a heatmap corresponding to the DAVID-Analysis.
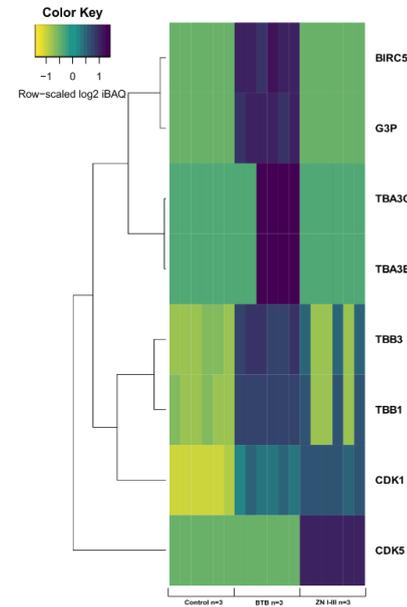
Figure 20: Bioinformatical Analysis via DAVID of relevant molecular functions involved in binding partners of ZBTB9. Proteins enriched for either BTB or Zinc finger domain with an FDR cut off 0.1% were taken as input. Top molecular functions and associated proteins were taken and mapped with their corresponding log2 iBAQ value to obtain a heatmap corresponding to the DAVID-Analysis.

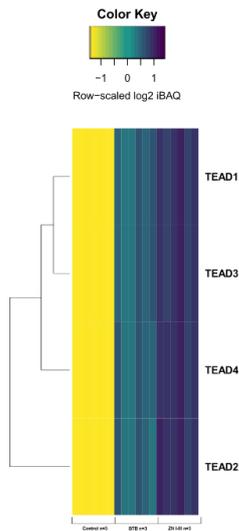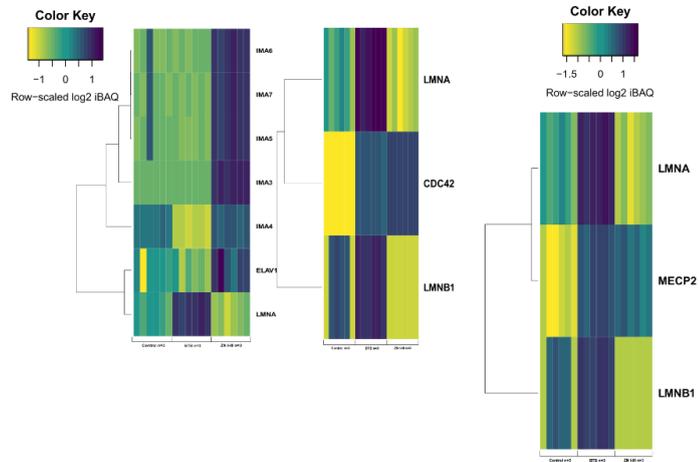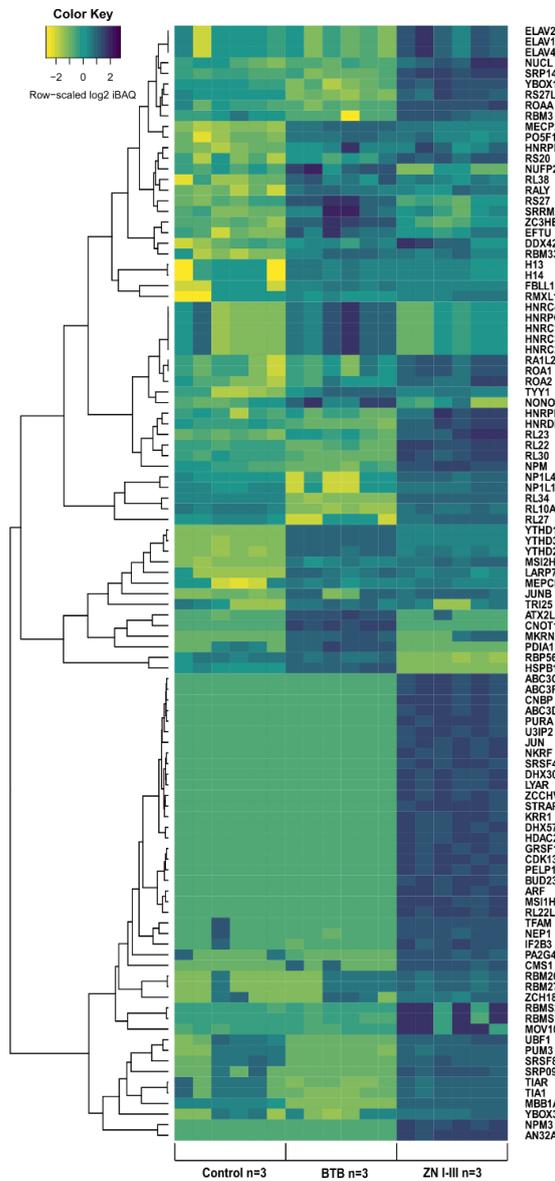**DAVID Analysis of protein interaction with ZBTB9 BTB and Zn Domain**

DAVID Analysis: Molecular Function

3) **RNA Polymerase II C-term domain kinase activity/ Activator activity, RNAP II transcription regulatory region sequence-specific binding**

4) **Transcription factor binding**
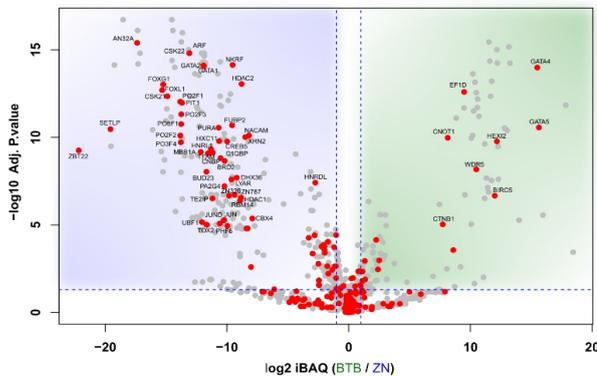
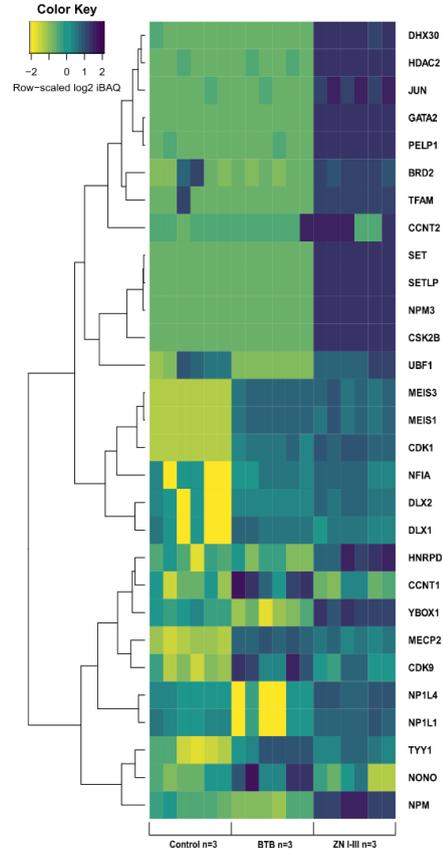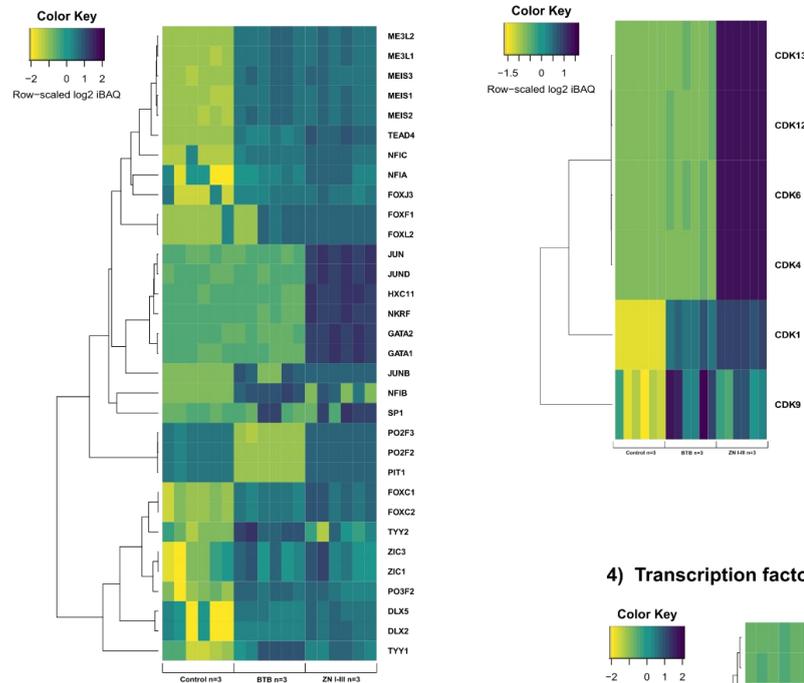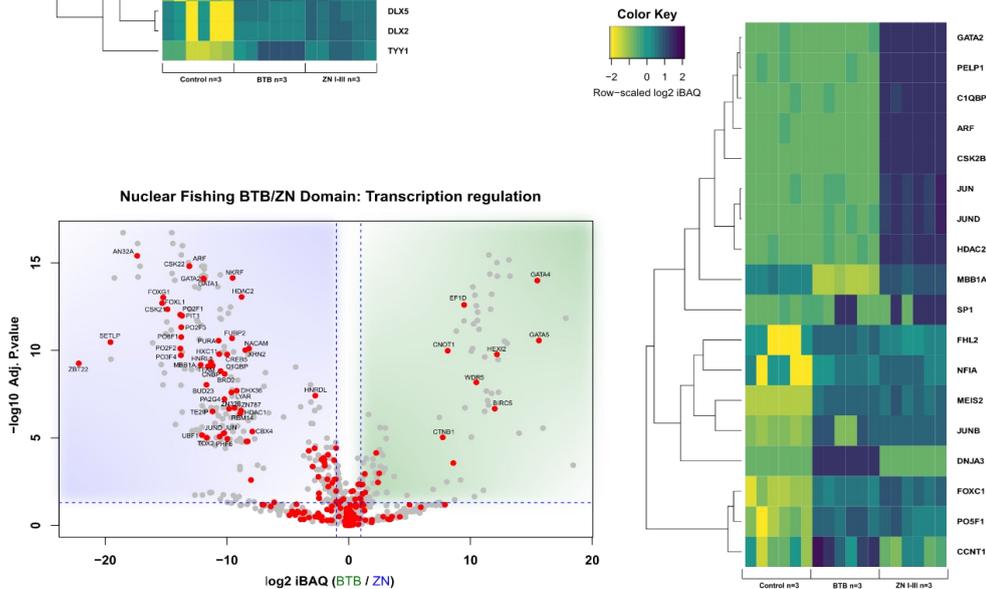*Figure 21:* **Bioinformatical Analysis via DAVID of relevant molecular functions involved in binding partners of ZBTB9**. Generation of heatmaps are explained in previous figures. The volcano plot for each domain was set at a cut off of 1 and all the interactors known in transcriptional regulation were plotted with their respective log2 iBAQ value and their -log10 adjusted p-value.

The top biological functions for ZBTB9 (when both domains were taken together) were *Regulation of Transcription from RNAPII promotor* (e.g., NONO, SRRM2, CTBL1, SRSF4, ZBTB22)*, Regulation of Transition from G1-/S-Phase and Regulation of Cell cycle* (e.g., CDKs, ARF, JUN, JUND, TYY1 or BIRC5). Moreover, other processes such as *RNA splicing* or *cytosine to uridine editing* (e.g., ABC3C/F/D, NEP, KRR1, RS27) were present too. Additionally, processes involved in *translational regulation, microtubule cytoskeleton organization, nuclear pore localization, nuclear import* (e.g., BIRC5, G3P, TBA3C, TBB3/1) were enriched as well. Surprisingly, in the biological processes analysis, the signalling *Hippo* pathway was present in in the pulldown for both domains. This implies, that ZBTB9 might be involved in the cascade pathway of the HIPPO pathway. It is known that the HIPPO pathway is responsible for tissue growth and homeostasis of the cell[445] but how this pathway is in interlinked with ZBTB9s function at the centromere or even transcriptional processes remains elusive.

On a more globular perspective, each domain of ZBTB9 is involved in the same biological processes and remain similar molecular functions. Nonetheless, there are some cases in which either the BTB domain or the zinc finger domain has a more prominent role in interacting with the specific enriched proteins. In case of transcriptional regulation/ RNA related processes (RNA Polymerase II, mRNA splicing, RNA binding), both domains seem to be similarly involved and show distinct clustering patterns in the heatmaps (darker colouring scale for higher enrichment), whereas these cannot be found in the control. Specifically, the zinc finger region is more occupied with transcription factor binding, implying that the zinc finger region is important not only for the binding to CENP-B. An interesting observation can be drawn from the BTB domain, which might be involved in cytoskeleton organization/ nuclear import and nuclear pore localization. This enrichment of specific proteins can be partially confirmed by the aforementioned *in vivo* staining patter of ZBTB9 in human cells. There ZBTB9 seem to be near the nuclear lamina. Moreover, this region is also known for heterochromatin occupation and as seen in the analysis of very inner protein binding partners, there are interactors identified which are known to be involved in heterochromatin formation.

The Nuclear Fishing Pulldown of the ZBTB9 BTB and Zinc finger domain revealed a diverse function for ZBTB9 in a cellular context. Ranging from transcriptional regulation to RNA splicing events up to cytoskeleton organization and nuclear import or nuclear localization signalling.

5.1.5 ZBTB9s function might be altered by phosphorylation as seen in the Nuclear Fishing and *in vitro* phosphorylation Experiments

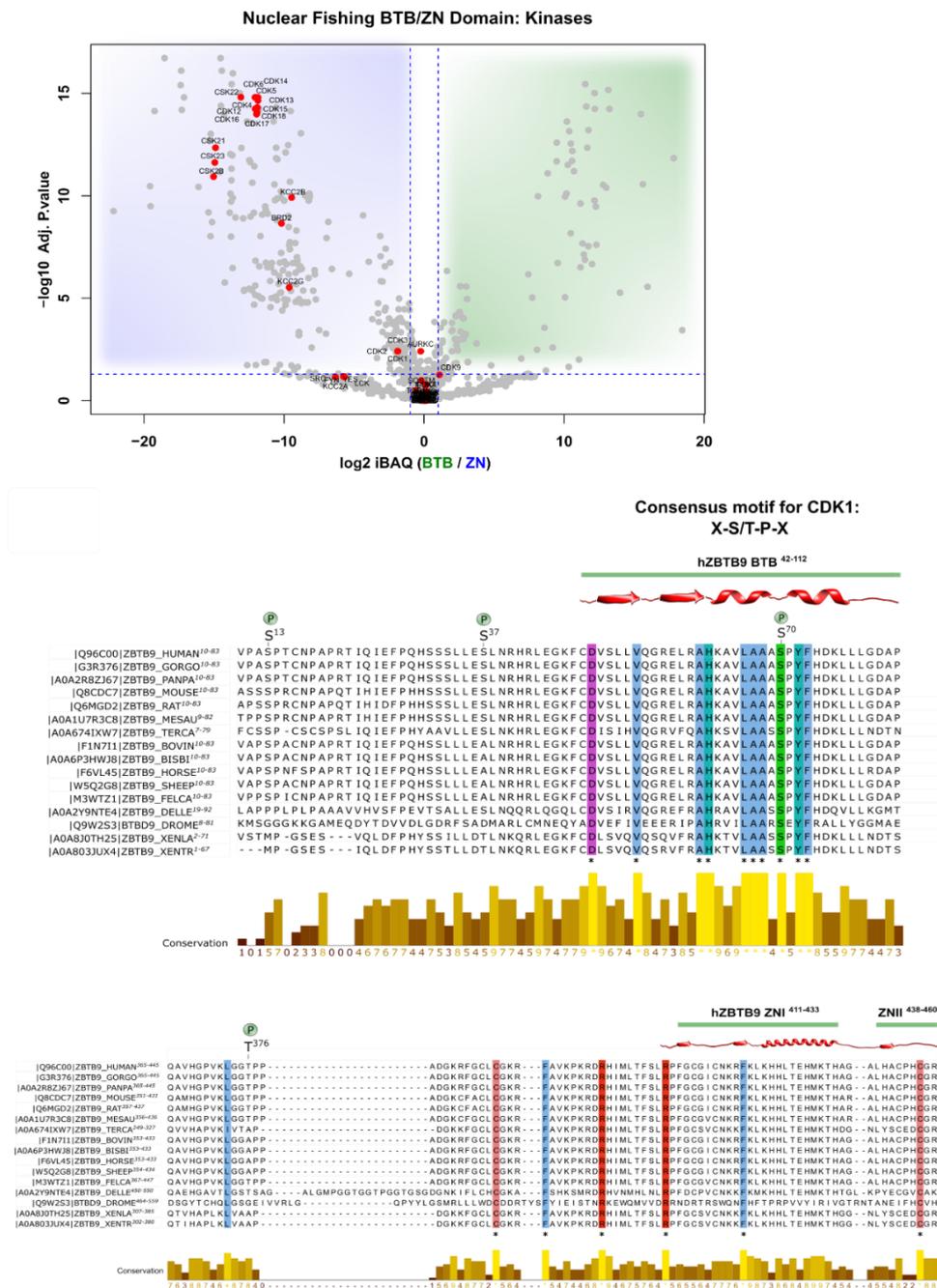
Figure 22: **Volcano Plot of identified Kinases in ZBTB9-NFPD and the Clustal Omega Analysis of ZBTB9.** The Vulcano Plot highlights the significant enriched kinases in each of the domains. Many kinases were detected beneath the cut off, whereas specifically in the zinc finger region many other kinases could be detected with a more stringent cut off. The sequence alignment was performed with

Clustal Omega and output parameter were loaded into the software Jalview to provide a more comprehensive alignment. Structural features from alphafold were added via the software Inkscape.

During the MaxQuant analysis, the search for phosphorylation sites on each given ZBTB9 Domain were investigated, since the volcano plot depicted in Figure 22 shows a high enrichment of diverse kinases for both domains, more specifically for the zinc finger domain (e.g., CSK21,22,23, CDK1,2,3,4,5,6,9,12,13,14,15,16 and CDK18). The phosphorylation sites were only identified by MS/MS runs in the BTB-NFPD. A quantitative statement cannot be given, since MaxQuant could not match the MS/MS spectra to a MS1 scan event. Therefore, the identified sites on ZBTB9s BTB Domain (site S70 and S37) could only be assumed. As seen in Figure 22, a sequence alignment on different ZBTB9 sequences across different eukaryotic species in Clustal Omega revealed one prominent serine site with CDK consensus sequence motif in the BTB domain (S70) which is conserved throughout all aligned species.

Other phosphorylation sites annotated by the database PhosphoSitePlus (Table 7) were not found in this analysis, since some of the sites were not present in the given constructs used for NFPD. Moreover, the identified sites in this NFPD have not been yet annotated, marking them as being novel phosphorylation sites. The serine at position 70 was also detected in an *in vitro* phosphorylation assay (Table 6), where different kinases phosphorylated ZBTB9 wildtype (from *E.coli* with an N-terminal MBP-tag) and phosphorylation sites were analysed by LC-MS/MS-Analysis (Ignasi Forné, ZfP, BMC, LMU). As the sequence alignment correctly displayed, the serine at position 70 can be phosphorylated by CDK1 (here yeast CDK1 complex was used). Additionally, at threonine at site 376 can be phosphorylated by CDK1 too. Since the timing of the *in vitro* phosphorylation was kept short (30 min), CDK1 only phosphorylated these two sites, promoting the idea, that besides the conserved site also another side is required to alter the structure of ZBTB9. A closer look at the sequence alignment reveals another CDK1 consensus sequence motif and therefore provides a platform for CDK1 phosphorylation but this site is not conserved. In contrast, human PLK1 does not share the same sites as CDK1 and mainly phosphorylates the unstructured linker region, except for two sites in the BTB domain (S51 and T22). The latter kinase was performed as a control, since PLK1 has not been identified in the NFPD. Although PLK1 was not originally identified in the pulldown the phosphorylation pattern of PLK1 was interesting, since it nearly only phosphorylated sites within the unstructured linker region adjacent to the acidic stretch (aa 211-226 VEEEEEEEEDDDDEDQ). Since it is known that the linker regions are often post-

translationally modified, this *in vitro* phosphorylation assay provides an insight on how and where this modification within ZBTB9 could resides. Nonetheless, it has to be confirmed if PLK1 is another kinase modulating the function of ZBTB9. Moreover, it is unclear which other kinases might phosphorylate the linker region of ZBTB9 and how this influences its function in a molecular level.

Since in the previous paragraph the relation of ZBTB9 and CENP-B was established, it was of interest whether phosphorylation of ZBTB9 could influence CENP-Bs binding capacity (Figure 23). Moreover, a phosphorylation mimic mutant (serine 70 to aspartic acid) was tested to see if one phosphorylation site on the BTB-domain might have an influence on the binding to the zinc finger region at the C-term of ZBTB9 to CENP-B. The result of such a binding assay is shown in Figure 23. The kinase assay reveals that phosphorylation of ZBTB9 by yCDK1 or PLK1 does slightly influence the binding to CENP-B. Quantification analysis shows a discrete decrease in binding.

The quantification reveals that the phosphorylation of ZBTB9 by yeast CDK1 decreases the binding of CENP-B to 19.30 %, whereas the phosphorylation by PLK1 only reduces the binding to 2.05%. The phosphorylation mimic mutant (S70D) only decreases the binding of CENP-B to this mutant about 4.70%. If both kinases are phosphorylating ZBTB9 the resulting phosphorylation pattern decreases the binding to 28%. The kinase assay provides an explanation why so many kinases were detected, since ZBTB9 seems to be strictly regulated by phosphorylation events. The kinases detected are not only responsible for G1/S-Phase transition or mitotic progression. Some of the CDKs (CDK9, CDK12, CDK13, CDK4/6) are also known to regulate the transcriptional process of the RNA Polymerase during the initiation, elongation, and termination process.

As stated in section 5.1.3, the fusion construct of ZBTB9 (missing the unstructured linker region) showed an interesting binding phenotype, where the binding of CENP-B is abrogated to about 50%. Since phosphorylation showed a slight influence into the binding of CENP-B, it was of interest whether the phosphorylation of this construct with CDK1 could lead to a more pronounced reduction in CENP-B binding. The phosphorylation site at S70 is still present (not T376) in this construct and subsequent phosphorylation with CDK1 lead to a decreased binding of up to 60 % (decrease in 8 %). As the phenotype of this construct has been shown before, an additional phosphorylation event changes the dynamic between ZBTB9 and CENP-B and therefore leads to a decreased binding event. As pointed out earlier, the linker region seems to be important for the overall structural integrity of ZBTB9,

since the fusion construct shows reduced binding, although the zinc finger region should be available for binding. Moreover, the construct shows how the BTB domain still influences the zinc finger region, since the phosphorylation could only occur on the one conserved CDK1 site within the BTB domain. Taken all results together, ZBTB9 seems to be regulated by phosphorylation events and might mediates through its zinc finger region other phosphorylation events which might be required in the transcriptional process and even initiates an unknown phosphorylation cascade leading to specific gene regulation events. This phenomenon has to be further investigated.

Table 6: *In vitro* **Phosphorylation sites on ZBTB9 identified by MS**. ZBTB9 was immobilized on maltose binding beads via its N-terminal MBP tag. The *in vitro* phosphorylation was performed with yeast CDK1 complex and with human PLK1 at 30°C for 45 min in a buffer containing ATP and $MgCl_2$. The reaction was stopped and subsequent addition of SDS-loading buffer. Protein was separated on SDS-PAGE and protein bands corresponding to ZBTB9 were cut and analysed by MS (Ignasi Forné, ZfP, BMC). Analysis of MS-Spectra were performed in Scaffold 5 program.

Sequence	Sequence identity	Phosphorylated by	Phosphorylation identfied?	ZBTB9 position
AVLAAASPYFHDK	100 %	CDK1	Yes	S70 (BTB)
LGGTPPADGKR	100 %	CDK1	Yes	T376 (before Zn)
FCDVSLLVQGR	100 %	PLK1	Yes	S51 (BTB)
TIQIEFPQHSSSLLESLNR	100 %	PLK1	Yes	T22 (Before BTB in N-term)
ELETSGGGISAR	100 %	PLK1	Yes	S144 (linker)
ELETSGGGISAR	100 %	PLK1	Yes	S149 (linker)
EEISGSGTQPGGAK	100 %	PLK1	Yes	S297 (linker)
EEISGSGTQPGGAK	100 %	PLK1	Yes	S299 (linker)
EEISGSGTQPGGAKEETK	96 %	PLK1	Yes	T310 (linker)
GGNSYHALLSTTSSTGGWCIR	100 %	PLK1	Yes	T162, T163 (linker)
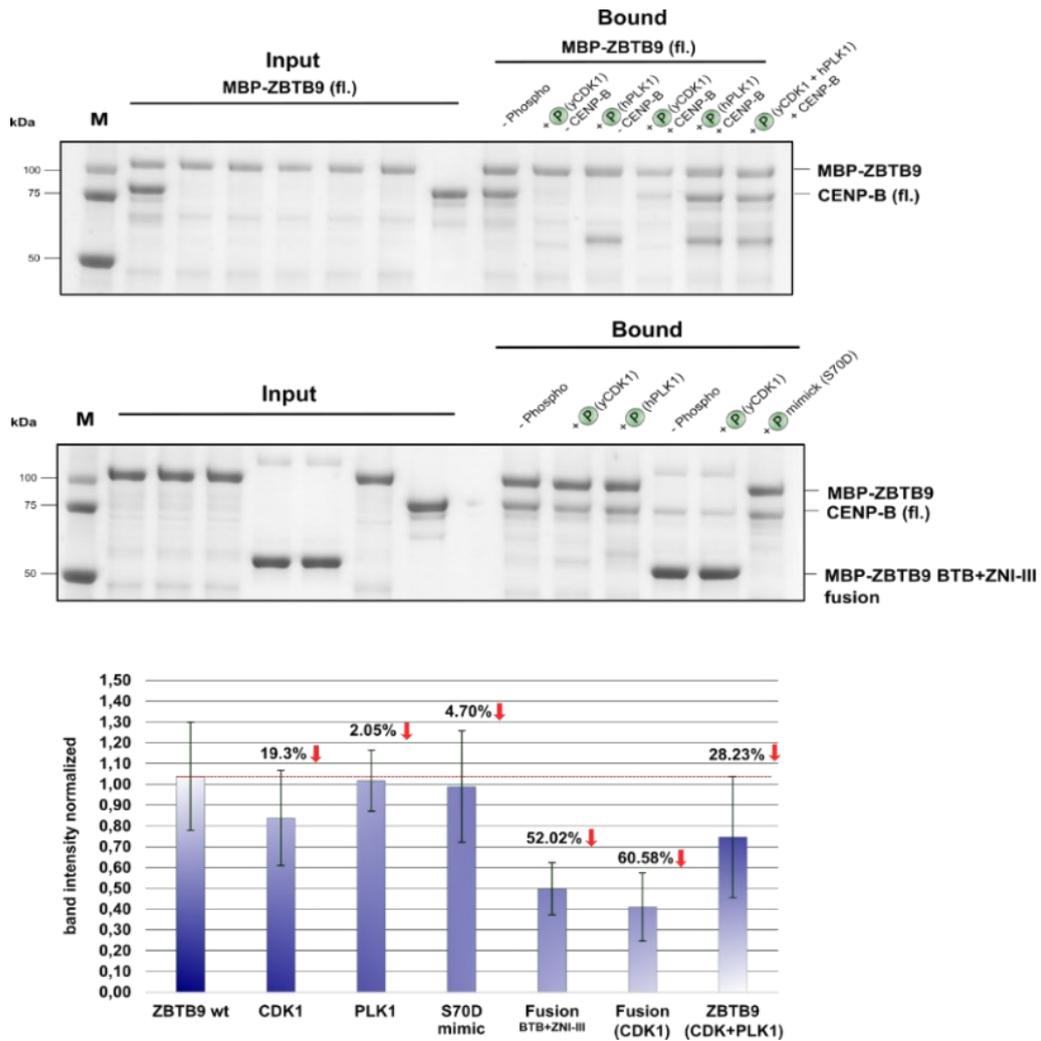
***Figure 23*: Kinase assay to determine the effect on ZBTB9 and CENP-B binding.** All assays were performed as biological triplicates. For a quantitative statement, the average mean of each experiment was taken as input for the quantification of the band intensities. Band intensities were measured by the software ImageJ. The standard deviation is presented as error bars. The red dotted line indicates the band intensity for the wildtype situation. The percentage at the top of each condition represents the decreased binding of CENP-B to ZBTB9 in percent.

Table 7: **Modified sites enlisted on the PhospoSitePlus V6.6.04 database for ZBTB9 based on HTP proteomic datasets.** The references are the number of publications mentioning this site identified with this specific modification. It is not known which enzyme is responsible for each modification. In this study two annotated phosphorylation sites could be linked to two kinases as well as two novel phosphorylation sites are detected.

Soure; https://www.phosphosite.org/proteinAction.action?id=2292678&showAllSites=true, accessed 30.09.2022[446].

Sequence	Modified site	Modification type	References by HTP-proteomic MS	Responsible Enzyme
PLPPVPASPTCNPAP	S 13	Phospho	6	?
TCNPAPRTIQIEFPQ	T 22	Phospho	1	?
GRELRAHKAVLAAAS	K 63	Ubiquitin	1	?
ASPYFHDKLLLGDAP	K 76	Ubiquitin	1	?
SARGGNSYHALLSTT	Y 156	Phospho	1	?
YHALLSTTSSTGGWC	T 163	Phospho	1	?, PLK1 [this study]
ALLSTTSSTGGWCIR	S 165	Phospho	1	?
TPQPQRVSGVFPRPH	S 241	Phospho	1	?
GPHPLPMTATPRKLP	T 256	Phospho	3	?
HPLPMTATPRKLPEG	T 258	Phospho	1	?
PPKIFXIKQEPFEPK	K 286	Ubiquitin	1	?
PPKIFXIKQEPFEPK	K 286	Sumo	5	?
KQEPFEPKEEISGSG	K 293	Ubiquitin	1	?
GTQPGGAKEETKVFS	K 307	Ubiquitin	1	?
GGGGPSWKPVDLHGN	K 344	Ubiquitin	1	?
QAVHGPVKLGGTPPA	K 372	Sumo	1	?
GPVKLGGTPPADGKR	S 376	Phospho	8	?, CDK1 [this study]
GTPPADGKRFGCLCG	K 382	Ubiquitin	3	?

It must be mentioned, since other kinases could not be tested, it cannot be ruled out that other kinases might phosphorylate other sites on ZBTB9 which might lead to complete loss of CENP-B binding. Some candidates are presented in the volcano plot such as CDK9, CSK21-23, CDK14 or the mitotic master kinases Aurora B.

5.1.6 The WICH-5 complex might have a role in CENP-A replenishment and Centromere transcription

In the previous paragraph the NFPD resulted in the identification of proteins involved in RNA-dependent processes or heterochromatin formation. The interactome of WICH-5 or CHRAC1-5 complex has not yet been studied. Therefore, the same NFPD was performed by using the complete complex for each pulldown. The CHRAC1-5 PD was performed only as one biological replicate, whereas the WICH-5 PD was performed in two biological replicates (each measured as one technical replicate). The following results display the interactors identified in the WICH-5 PD. The CHRAC1-5 PD will only be mentioned briefly. In the WICH-5 NFPD in total 1046 proteins were identified, after removing contaminants and low abundant proteins. After applying the t-test to obtain statistical p-adjusted value, proteins with an p-adjusted value of 0.01 (corresponding to 1% FDR) or 0.05 (5% FDR) were taken as input for further STRING and DAVID analysis (Figure 24, Table 8, Figure 25). When applying 1% FDR, 38 significantly enriched proteins were identified; in case for 5 % FDR in total 166 proteins could be significantly identified. The significant identified proteins were compared to the CENP-A ChIP-MS Pulldown as well as with the previous ZBTB9 NFPD. There were no proteins in common with the CENP-A ChIP-MS Pulldown. In case for ZBTB9 NFPD only four proteins (SMARCA2, CENP-X, GATA4 and DHX8) were in common.

Since WICH has been shown to bind CENP-B, it was of interest whether the binding can be confirmed in a NFPD but also what the role of WICH could be in a more cellular context. The binding to CENP-B could be confirmed, though with low confidence (19.5% FDR). Another kinetochore member CENP-X has been identified with an FDR or 10.9%. For STRING and DAVID Analysis the FDR was set to 5%. As seen in Figure 24, the STRING Analysis revealed that WICH complex might be involved in RNA splicing and Transcriptional Regulation (e.g., DHX15, DDX1, FOXG1, CCNL1 or ZIC3), similarly as ZBTB9 shown in the previous paragraph. Since the WICH complex belongs to the ISWI chromatin remodeler family, it is no surprise that Chromatin Binding and Nucleosome Assembly/Disassembly were one of the top molecular functions of this complex. Interestingly, the DAVID Analysis revealed the involvement of the WICH complex with nuclear import processes (IMA1, IMA5, IMA6) and an additional molecular function in RNA binding. An interesting finding was the enrichment of SWI/SNF chromatin remodeler INO80 or pBAF subunits (INO80, ARP8/5, ACL6A/B). Since the WICH-5 complex have been found to be associated with the SWI/SNF remodeler family members INO80 and pBAF complex, the interaction might facilitate CENP-

A replenishment. As for now the mechanism of CENP-A replenishment are still elusive, this might provide an insight by studying the interaction of these two chromatin remodeling families. Notably, subunit members of CHRAC1-5 and ACF-5/-1 were identified as well. Another interesting observation is the enrichment of RNA Polymerase II associated proteins in the WICH-5 NFPD (e.g., SPT5H, SPT4H, RPB1C, RPB3, RPAB1). The connection of WICH-5 with RNA Polymerase II was drawn in one study, where WICH bound to NM1 (MYO1C) and mediated the recruitment of histone acetylases and methylases to maintain acetylation and methylation marks for active transcription as seen in e

uchromatin[447]. Yet the NFPD presented here demonstrates that the WICH-5 complex can indeed associate with RNA Polymerase II proteins and might be involved in centromere transcription. In previous studies the WICH-5 complex was associated with RNA Polymerase I and transcription of ribosomal genes (rDNA) or with RNA Polymerase III in the transcriptional process of 5S rRNA genes[448,449].

In contrast to the ZBTB9-NFPD with a cut off from 5% FDR no kinase could be detected. With a higher cut off (≥ 50% FDR) kinases such as CDK9 (15.3% FDR), CDK1 (50% FDR), CD11A (10.3% FDR) could be detected. It has to be mentioned, since the WICH complex was purified from insect cells, phosphorylation sites might be altered by insect cell kinases, and therefore the kinases in the nuclear extract do not need to further phosphorylate any of the sites. Therefore, if it is of importance to see which kinases might be involved in WICH complex phosphorylation, the complex has to be expressed in *E.coli* and then the NFPD has to be replicated in this conditions.

A similar analysis with CHRAC1 complex has been performed but the NFPD was only performed in one biological replicate, since no kinetochore associated proteins were identified. In total 172 proteins were identified after subtracting the background. Since no CCAN protein has been detected, this complex might interact with CENP-A Nucleosomes directly and does not need any physical interaction with known kinetochore proteins to protrude its function. A quick analysis of this one biological replicate revealed annotations for Transcription, Regulation of Gene Expression or Chromatin assembly/ organization. Hence, the analysis of the one pulldown had similar annotations as the WICH complex but was not further investigated.

Nonetheless, the analysis of the WICH-NFPD showed that this ISWI chromatin remodeler is involved in RNAPII mediated transcription, RNA-binding, and RNA-splicing events as well in Transcriptional regulation. Additionally, it might have a role in nuclear import signalling.
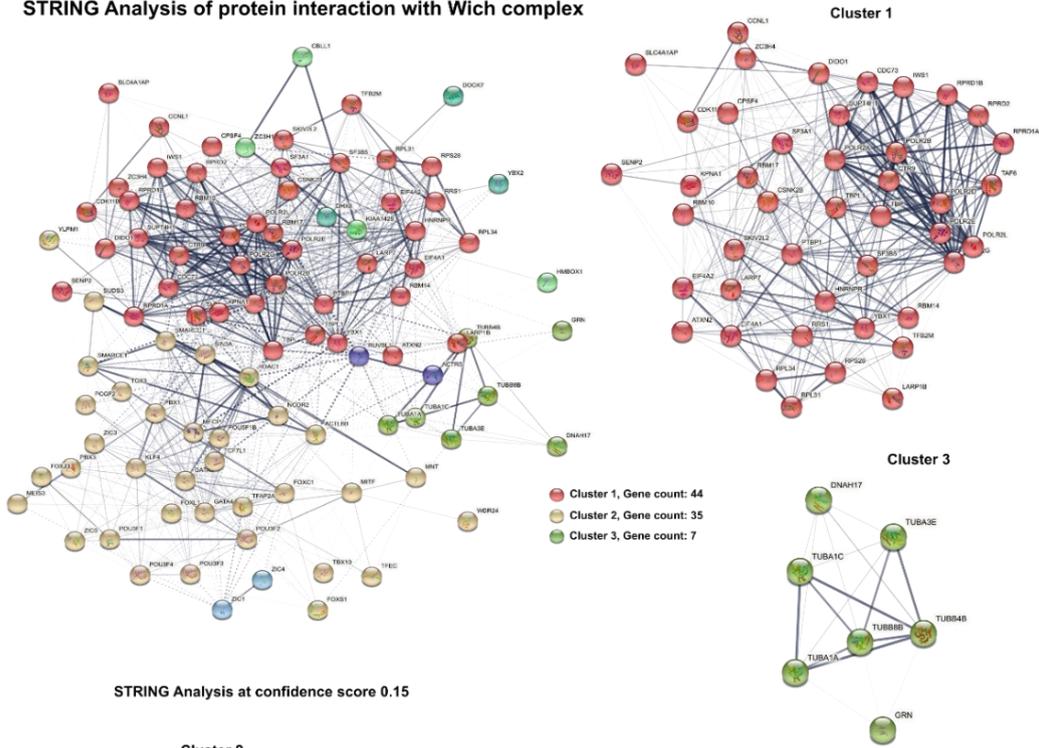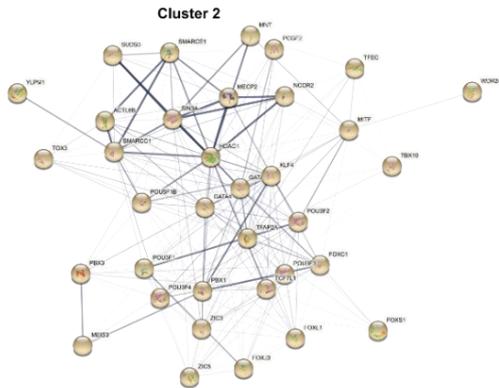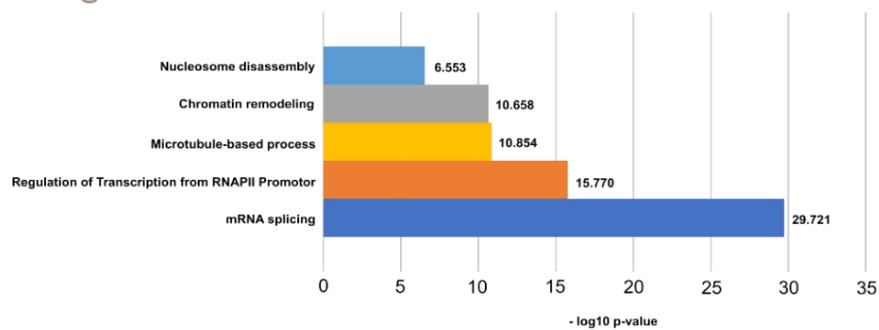
Figure 24: STRING and DAVID Analysis of WICH complex interaction partner. Interaction partners identified by an FDR of 5% were taken as Input for either STRING or DAVID Analysis. In STRING, the interaction score was set to 0.15. The clustering was performed with the settings of MCL clustering option. The most prominent biological processes or molecular functions with an FDR

score of < 0.1 were taken. The DAVID Analysis only depicts the five most named biological functions with protein interaction partners found with an FDR of 5%. The p-value provided by the DAVID analysis was -log10 transformed. Shown are p-values corrected for multiple testing within each category using the Benjamini-Hochberg procedure

Table 8: **Protein IDs identified within the NFPD of the WICH complex with different p-adjusted value cut offs and a short UniProt description.**

| Identified proteins in WICH NFPD	p adjusted value in NFPD	UniProt description
SPT5H	≤ 0.01	Component of DRB sensitive-inducing factor complex (DSIF), regulates mRNA processing and transcription elongation by RNAPII
NFIB	≤ 0.01	Transcriptional activator of GFAP, recognizes and binds to palindromic sequences 5'-TTGGCNNNNNNGCCAA-3'
DHX15	≤ 0.01	RNA helicase involved in mRNA processing
TBX10	≤ 0.01	Probable transcriptional regulator
DUS11	≤ 0.01	Binds to RNA, may participate in nuclear mRNA metabolism
ZIC3	≤ 0.01	Acts as transcriptional activator
ACL4A	≤ 0.01	Involved in transcriptional activation and repression by chromatin remodelling, component of BAF or INO80 complex
ACL6A	0.0097	Involved in transcriptional activation and repression by chromatin remodelling, component of BAF or INO80 complex
ACL6B	0.0064	Involved in transcriptional activation and repression by chromatin remodelling, component of BAF or INO80 complex
RU2B	≤ 0.01	Involved in pre-mRNA splicing as component of spliceosome
TBPL2	≤ 0.01	Transcription factor
PBX3	≤ 0.01	Transcription factor
RPB1C	≤ 0.01	DNA-dependent RNA polymerase catalyses transcription of DNA into RNA, component of RNAPII
ELOF1	≤ 0.01	Transcription elongation factor
SPT4H	≤ 0.01	Component of DISF complex, regulates mRNA processing and transcriptional elongation by RNAPII
SF3B3	≤ 0.01	Involved in pre-mRNA splicing as a component of splicing factor SF3B complex, constituent of spliceosome
SF3A2	≤ 0.01	Involved in pre-mRNA splicing as component of splicing factor SF3A
CDC73	≤ 0.01	Tumor suppressor, probably involved in transcriptional and post-transcriptional control pathways, involved in histone modifications
HAKAI	≤ 0.01	E3 ubiquitin-protein ligase, mediating ubiquitination

ZN444	≤ 0.01	Transcriptional regulator
PPIL4	≤ 0.01	Peptidyl-prolyl-cis-trans isomerase accelerate folding of proteins
DDX1	≤ 0.01	Acts as ATP-dependent RNA helicase, able to unwind RNA-RNA and RNA-DNA duplexes
SIN3A	≤ 0.01	Transcriptional repressor, cooperates with FOXK1 regulating cell cycle progression
ATX2	≤ 0.01	Involved in EGFR trafficking
FOXS1	≤ 0.01	Transcriptional repressor
FOXG1	≤ 0.01	Transcriptional repressor
NADAP	≤ 0.01	Not known
WDR61	≤ 0.01	Component of PAF1 complex, multiple functions during transcription by RNAPII, involved in histone modifications
ARP8	≤ 0.01	Important role in functional organization of mitotic chromosome, component of INO80 complex
ARP5	0.0446	Important role in functional organization of mitotic chromosome, component of INO80 complex
S30BP	≤ 0.01	Role in transcriptional repression by promoting histone deacetylase activity of histone H3
CCNL1	≤ 0.01	Involved in pre-mRNA splicing, associates with CDKs, may play a role in regulating RNAPII
RTRAF	≤ 0.01	RNA-binding protein involved in modulating RNA transcription by RNAPII
ACOT9	≤ 0.01	Acyl-CoA thioesterases,, regulating intracellular levels of acyl-CoA
HS71L	≤ 0.01	Molecular chaperone
SRP14	≤ 0.01	Component of signal recognition particle (SRP) complex
RL34	≤ 0.01	Component of 60S ribosomal unit
SET	≤ 0.01	Involved in nucleosome assembly, histone chaperoning, transcription and apoptosis
ANKH1	≤ 0.01	May play role in scaffolding proteins
ANR17	≤ 0.01	Could play pivotal role in cell cycle and DNA regulation
GATA4	≤ 0.01	Transcription activator/ repressor
DPOE3	0.573	Component of DNA Polymerase epsilon complex, forms CHRAC1 complex
BAZ1A	0.991	Regulatory subunit of ACF1 and ACF5 ISWI chromatin remodeler complex
CHRAC1	0.695	Forms complex with POLE3, forms CHRAC1 complex with ACF5 ISWI remodeler
SMARCA2	0.011	Component of SWI/SNF chromatin complex (npBAF), involved in transcriptional activation/ repression
SMARCA1	0.023	Component of SWI/SNF chromatin complex (npBAF), involved in transcriptional activation/ repression

SMRD1	0.019	Component of SWI/SNF chromatin complex (npBAF), involved in transcriptional activation/ repression
INO80	0.040	ATPase component of chromatin remodeling INO80 complex, involved in transcriptional regulation, DNA replication and DNA repair
KCTD1	0.037	BTB/POZ containing protein, Sequence specific DNA-binding protein
KCD21	0.032	BTB/POZ containing protein, Probable substrate specific adapter of BTB-Cul3-RBX1 E3 ubiquitin ligase complex, mediating ubiquitin of proteins (e.g. HDAC1)
HP1B3	0.012	Component of heterochromatin maintaining heterochromatin integrity during G1-/S-Phase progression
RPB3	0.029	DNA-dependent RNA Polymerase catalyses transcription, component of RNAPII
RPAB1	0.037	DNA-dependent RNA Polymerase catalyses transcription, component of RNAPII

DAVID Analysis of protein interactors with Wich complex

DAVID Analysis: Biological Process

1) Nuclear Import
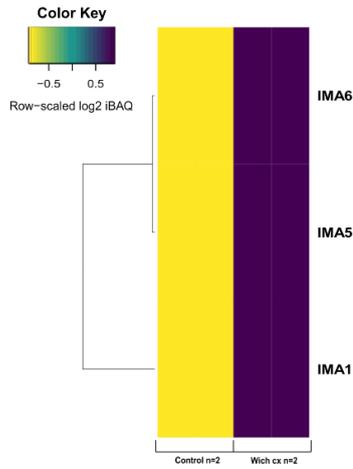
2) Nucleosome Assembly
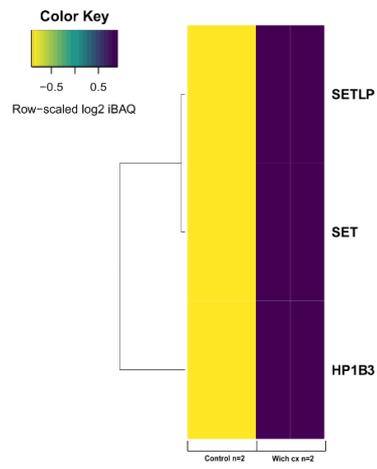
DAVID Analysis: Molecular Function

1) RNA Binding
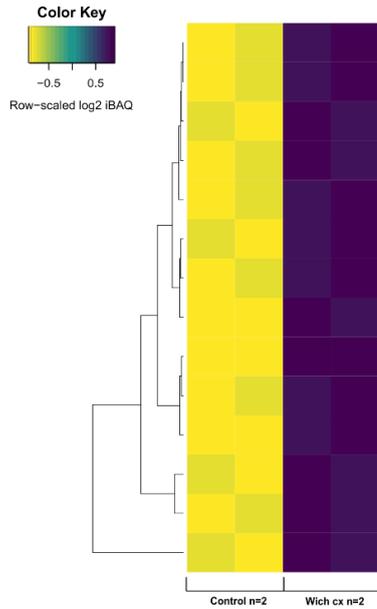
2) Chromatin Binding
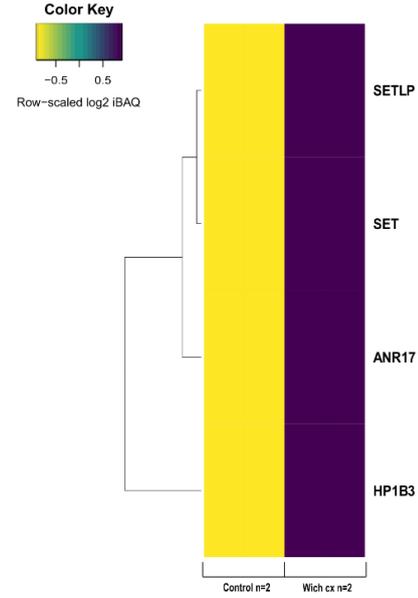
Figure 25: **Bioinformatical Analysis via DAVID of relevant biological processes and molecular functions associated with the binding partners of WICH complex.** Proteins enriched for WICH complex PD with an FDR cut off 5% were taken as input. Top biological processes and associated proteins were taken and mapped with their corresponding log2 iBAQ value to obtain a heatmap corresponding to the DAVID-Analysis.

5.2 Discussion

The constant maintenance of the centromere ensures faithful sister chromatid segregation during mitosis. During the cell cycle the centromere undergoes constant changes either through the incorporation of replenished CENP-A in G1 or through the assembly of the kinetochore to ensure correct bi-oriented spindle micro tubule attachment in mitosis. Latest studies suggested that transcription at the pericentromere and centromere is an integer part of this specific chromatin region. In early stages of centromere research, studies proposed a silent nature of the centromere region, where transcription was not determined. Moreover, since heterochromatin marks such as H3K4 trimethylation or H3K27 methylation were observed, it was clear that this region is only a nucleation hub for microtubule attachment through the kinetochore. Yet, many years later the picture is changing. The unique features of the highly repetitive centromere elements have shown it can be transcriptionally active, resulting in the formation of long-noncoding RNAs which are spliced into smaller RNA transcripts and fulfil various tasks at the centromere. Moreover, this process is vital for the stability and integrity of the centromere region. Depletion or knock down experiments of centromere transcripts resulted in instable centromere formation, defective CENP-A or CENP-C incorporation or segregation defects resulting in aneuploidy. To gain insights into a) how CENP-A assembles onto chromatin, b) how stably it is transmitted during DNA replication and c) which proteins are necessary to promote centromere identity, efforts have been made to study the CENP-A proteome during the different cell cycle stages. In a previous work of the lab, a preliminary data set identified novel centromere associated factors. The study presented in this thesis follows up and characterizes some of the interactors (here ZBTB9, WICH-5 and CHRAC1-5 complex) in respect to their contribution in centromere maintenance and function. The interest lies in finding answers asked by following questions. Are the ISWI chromatin remodeler WICH-5 and CHRAC1-5 centromere specific remodelers and if so, with whom do they interact directly? Moreover, what could be a proposed model of these chromatin remodelers at the centromere? Another interesting finding is the constant association of a transcription factor called ZBTB9. At the N-terminal region the protein harbours a BTB motif which is known for dimerization/ oligomerization but also mediating protein-protein interactions. At the C-terminal site there are zinc finger of the C2H2 type typically associated with direct DNA binding and protein-protein interactions. This protein has not been yet studied and it was not clear how it was associated with the centromere. Therefore, it was of interest to address with whom ZBTB9 interacts at the centromere and what function it could harbour at the centromere.

In the previous study of the lab, the ISWI remodelers WICH-5 and CHRAC1-5 complex were originally identified in the screening and were associated to certain cell cycle stages (G1- and S-Phase respectively). Another interesting interactor was the transcription factor ZBTB9, which was associated tightly to the centromere over the course of the entire cell cycle. The remodelers as well as ZBTB9 could be linked to the centromere through their binding to CENP-A Nucleosomes. Another interesting observation was made with ZBTB9, and WICH-5 complex, where both were able to bind CENP-B directly. In case of CHRAC1-5 only the direct binding to the CENP-A Nucleosome were a direct link to the centromere. Moreover, WICH-5 could bind the CENP-A Nucleosome directly without requiring the binding to CENP-B. In case of ZBTB9, CENP-B and ZBTB9 form a complex with the centromeric CENP-A Nucleosome and therefore links it as well to the centromere. Till today the binding of a chromatin remodeler to CENP-B has not been yet described and provides a novel insight into centromere maintenance through CENP-B. Furthermore, the binding of ZBTB9 to CENP-B was dissected and showed that the zinc finger domain of ZBTB9 is the primary binding site of CENP-B. Further analysis showed that the Zinc finger (380-462) region of ZBTB9 bind the N-terminal DNA-binding domain (1-129) of CENP-B. This novel binding to the DNA region of CENP-B has not yet been described and marks ZBTB9 as an integer part of the centromere landscape. Recently, yet another transcription factor ZFAT has been shown to bind to CENP-B at the centromere which alters centromeric transcription[450]. ZFAT harbours an AT-hook domain and 18-repeats of C2H2-type zinc fingers. A study could draw a link between centromeric ncRNA (noncoding RNA) transcription and ZFAT function at the centromere. ZFAT mediates the expression of ncRNA by recruiting KAT2B (histone acetyltransferase) which acetylates H4K8 and attracts the bromo-domain containing protein BRD4 to the centromere and thus facilitate centromere transcription[290]. Overexpression of ZFAT lead to increased expression of ncRNA at least at the centromere at chromosome 17 and X[290]. A follow up study by the same investigators found out that ZFAT binds to CENP-B. As seen for CENP-B binding to the CENP-B box, ZFAT also binds a specific 8-bp DNA sequence, the ZFAT box[290,450]. The box is well conserved and widely distributed at whole alpha satellite DNA region of every chromosome[290].

In case of ZBTB9 there could not be determined a specific DNA sequence that is bound specifically by ZBTB9. It is hypothesized that it might bind A-box sequences, since no protein has been described to bind such sequences. Further investigations have to address if ZBTB9 binds a specific DNA sequence and how the zinc fingers at the C-term recognize

these sequences. The ZBTB family has been shown to be specific DNA binders, specifically onto methylated sequences (ZBTB38 or ZBTB24)[406,451–453]. Notably, CENP-Bs' acidic domain is the main interaction site for the zinc finger region (C9-C12, aa 485-822) of ZFAT. The knockdown of CENP-B lead to a reduction of ZFAT levels and ncRNAs, connecting CENP-B to the ncRNA transcriptional process[450]. In contrast to ZFAT, ZBTB9 binds to the DNA-binding domain of CENP-B therefore marks it as a novel interaction and might alter the binding to the centromere/ CENP-A Nucleosome.

In the initial EMSA assay CENP-B and ZBTB9 could bind to the centromeric CENP-A Nucleosome and form a higher molecular weight complex. One thing that could not be tested is, if ZBTB9 might interfere with CENP-Bs binding, when the DNA is methylated. The centromere DNA is known to be methylated and studies showed a methylation sensitive binding of CENP-B to the CENP-B box. The CENP-B box contains two CpG sites and therefore could be methylated [200]. Association of DNA methylation to the centromere are also shown in the nuclear fishing pulldown with ZBTB9. Here the DNA methyltransferase DNMT1 have been found and might regulate with ZBTB9 and CENP-B the DNA methylation pattern at the (peri)centromere.

Another family member of ZBTB9, ZBTB24 have been seen to coordinate DNMT3B to control DNA methylation at heterochromatin marks[454]. ZBTB38 at the other end has been associated with methylated DNA[455]. Notably, CENP-C has been shown to bind to DNMT3B and therefore manage the DNA methylation landscape at the centromere.[202] Since the PWWP domain of DNMT3A and B is important for pericentric heterochromatin formation[192], it could be that the association of DNMT1 is more prominent for the actual centromeric part. It is known that DNMT1 associates at G1-, S- and G2/M-Phase with the DNA to maintain the DNA methylation pattern[456]. An interesting observation alongside with DNMT1 was the enrichment of the E3 ubiquitin ligase UHRF1 in the pulldowns, which was also present in the CENP-A ChIP-MS data as well. Recent publications point towards an important function of UHRF1: DNMT1 complex in regard to DNA methylation status[457]. Therefore, ZBTB9 might control DNA methylation status of the (peri)centromere by interacting with CENP-B and mediating UHRF1:DNMT1 localization to the centromere.

In addition, CENP-B is post-translationally modified by a trimethylation mark at glycine at position one (since the methionine is cleaved off)[304]. Future studies will address, whether the combination of trimethylation on CENP-B in the DNA binding region as well as the methylation of the CENP-B box have an impact on ZBTB9s binding to the DNA binding

region in CENP-B and if this might alter the binding of CENP-B to the CENP-B box. The direct binding of ZBTB9 to CENP-B explains why the interactor was associated with the centromere at each cell cycle stage since CENP-B is always associated with the centromere[198].

Other members of the CCAN were not involved in binding to ZBTB9. ZBTB9 and CENP-B have a common feature, both have an acidic patch within their protein sequence. In ZBTB9s case this lies in the unstructured linker region which connects the BTB and the zinc finger region. Recent studies showed how the acidic domain of CENP-B is required to recruit heterochromatin associated interactors to the centromere[309]. In addition, the formation of DNA loops by CENP-Bs C-terminal dimerization domain provides another functional role of CENP-B in the centromere organization[314]. By using a tetR-EYFP-CENP-B as bait, the study confirmed the pulldown of Suv39h1 but also additional factors such as ASH1L, NSD1, KMT2A, KDM4D, KDM68, KDM7A, PHF2 and HP1α, β and γ[309]. Since ZBTB9 binds to CENP-B it was of interest whether similar proteins could be found within the nuclear fishing pulldown experiments with the BTB or the zinc finger region of ZBTB9 but not including the unstructured linker region with the acidic patch. In the zinc finger pulldown, the binding of CENP-B could be confirmed with lower confidence. Interestingly, similar proteins binding to the acidic domain of CENP-B were found interacting with ZBTB9 either by the BTB or the zinc finger region (PHF3, PHF14, KDM2B, KDM2A, ASH2L). In addition, CBX 5,3,8,6 (CBX5 also known as HP1α) which are associated with heterochromatin formation were also found[127,130]. This implies that ZBTB9 is actively associated with heterochromatin formation at the (peri)centromere. Further evidence is given by the *in vivo* microscopy data, where ZBTB9 seems to be located near the nuclear lamina. The data also showed that ZBTB9 is not only exclusively located to the centromere and might occupy similar sites as shown for H3K9me2, which orchestrates spatial positioning of heterochromatin[458].

Further evidence can be found within the BTB-pulldown, which revealed the interaction of proteins associated with cytoskeleton organization, nuclear pore localization, nuclear lamina and microtubules. These proteins include LMNA (lamin-a type), LMNB1 (lamin-b type), Nup210L, HDAC1 and HDAC2 (histone deacetylases), CBX5 (HP1α), CBX3, DHX30, DHX8 and SUZ12 (polycomb group protein). Therefore, the BTB domain might anchor ZBTB9 near nuclear pore complexes to facilitate nuclear import but also associate the (peri)centromere near this region. The nuclear lamina network consists of A- and B-type lamins and their associated proteins at the inner nuclear membrane periphery[459,460]. In the

so-called lamina-associated domains the chromosomal regions interact with the nuclear membrane and are associated with weakly expressed genes. The Lamin-B-receptor links the heterochromatin mark of H3K9 di- and trimethylation as well as pericentromeric regions to the lamina associated domain by forming a complex with HP1α[461,462]. It has been shown that repressive marks known from heterochromatin (H3K9me2/3) are enriched beneath nuclear lamina through the spatial localization of histone deacetylases or methylases mediated by PPIs[463,464]. It has been shown that Nup98 binds to SETD1A and SETD1B (histone methyltransferases) and thus promote the deposition of H3K4 trimethylation mark (euchromatin), which stands for active transcription[465]. In addition, Nup98 has been linked in regulating splicing of specific mRNA transcripts by binding to DHX9[466]. Gene repression has been linked in progenitor cells by the recruitment of histone deacetylase HDAC3 through LAP2β and emerin, which are required to maintain the lamina associated domain[463,467].

Since the BTB domain of ZBTB9 has been associated with some of these processes at the nuclear lamina periphery, ZBTB9 might be involved in tethering the (peri)centromeric chromatin to these sites to ensure gene repression by mediating the interaction with HDAC1 or 2 as well as splicing events by binding to DHX30 and DHX8. How Nup210L plays a role in this complete process is not yet understood since it is poorly studied, therefore the relation between Nup210L and ZBTB9 have to be further investigated. In conjunction with heterochromatin association of the nuclear periphery, demethylases were associated either with CENP-B or as stated here with ZBTB9. In the nuclear fishing pulldown KDM2A and KDM2B were specifically enriched. Both demethylases belong to the Jumonji C domain-containing demethylase family. KDM2A and KDM2B target H3K36me2 but in case of KDM2B it also targets H3K4me3[468–471]. Some ZBTB proteins have been implicated as a transcriptional repressor by mediating the recruitment of HDAC or NcOR-1/2 specifically by the BTB domain[472]. Smad4 interacts with ZBTB7A, and increased interaction lead to enhanced recruitment of ZBTB7A to the Smad4-DNA complex which recruits HDAC1 to the complex and decreases the interaction between the downstream interactors p300[473]. In case of ZBTB32 recruits HDAC1 and HDAC2 to maintain the repressive chromatin state of the Il-4 locus to repress the expression of IL-4 in CD8+ Tc2 cells. Depletion of ZBTB32 resulted in hyperactivation and -proliferation of CD8+ T cells *in vitro*[474,475]. The presented examples show that indeed ZBTB9 can be involved in the heterochromatin repressive mark by mediating the localization of HDAC1 or 2 to the (peri)centromere, since aforementioned ZBTB family members have been shown to be involved in these processes.

This leads to the assumption that ZBTB9 strictly regulates the heterochromatin pattern at the (peri)centromere but also regulates the transcription at these sites by mediating between open and closed chromatin organization in conjunction with CENP-B but it is not clear how this process is performed. Future studies shall address this question how ZBTB9 mediates this process specifically at the (peri)centromere. Besides maintaining repressive sign at the (peri)centromere, ZBTB9 might also be able to mediate activation of transcription. The WICH-5 complex have been associated with Myo1c, which recruits the histone acetyl transferase PCAF and methyl transferase SET1/ASH2 to maintain H3K9ac and H3K4me3[447,476]. The link between WICH-5 and ZBTB9 can be drawn from the nuclear fishing experiment where specifically SMARCA5 have been identified. A proposed mechanism could be as followed: ZBTB9 recruits SMARCA5 by its zinc finger domain and mediating the assembly of the WICH-5 complex as well as the interaction with nuclear myosin 1 to recruit ASH2 to turn on transcription by modulating H3K4 methylation. Since CENP-B is capable binding both complexes at the same time as seen in the biochemical binding studies, it is of interest if there is a timely restricted interaction with each other or if they build a larger complex to recruit other interactors to the centromere and thus change the centromeric chromatin environment. Preliminary data (not shown) suggests, that ZBTB9 is not capable in binding the already assembled WICH-5 complex, which would underline the fact that BAZ1B was absent as an interactor of ZBTB9 nuclear fishing pulldown.

Besides the association with (peri)centromeric heterochromatin formation or maintenance, interactors found within the ZBTB9 nuclear fishing pulldown were linked to the RNA Polymerase II transcriptional processes as well as were component of the spliceosome or mRNA splicing process. The splicing event can be performed in two distinct ways: a) constitutive splicing where adjacent exons are joined or b) alternative splicing where the machinery self has to discriminate between multiple splicing sites[477]. The splicing of an mRNA transcript is mediated by a large macromolecular complex consisting of more than 100 individual components undergoing constant rearrangements to configure the catalytical core, the spliceosome[478,479].

As proven in several studies, the centromeric transcripts play a vital role in centromere integrity and perform different functions ranging from chromatin remodeling, transcriptional regulation, or overall chromatin organization[480,481]. So far only a hand full of proteins are associated with the process of centromeric transcription process, including centromere proteins CENP-A, CENP-B and CENP-C. Other components are HJURP, SGO1, Aurora B,

DHX38 and SUV39H1[158,284,482–488]. In CENP-Cs specific case it has been shown that the unstructured regions within CENP-C (between the N-terminal binding elements for Mis12 complex, CENP-N/L, CENP-H/I/K/M and the central CENP-A binding motif) were associated with RNA binding. In fact, the regions involved in DNA and Nucleosome binding (central region and CENP-C motif) were associated with binding to alpha satellite *RNA in vitro*[488,489]. Performing ChIP with anti CENP-C antibody, the pulldown contained alpha satellite RNA. Moreover, treatment with RNAse resulted in reduction of CENP-C localization during metaphase and eliminated centromeric localization of INCENP and Survivin (chromosomal passenger complex members)[158,490]. In case of ZBTB9, it is known that between the BTB and the zinc finger of C2H2 type an unstructured linker resides. It remains speculative but maybe ZBTB9 binds centromeric transcripts through the unstructured linker region as seen in CENP-C and might gain additional stabilizing effects in regard to CENP-B binding or to other centromere associated proteins. The binding to RNA might support the centromere organization and has an additional stabilizing effect on the building kinetochore. This thought is partially supported by an observation made during the purification process of the linker region alone (data not shown). During the purification process the absorbance of the protein was measured to determine the purity and quantity of the linker region. The absorbance value of protein to DNA ratio (A260/A280) yielded a high value which suggest high bound DNA or RNA content. After performing anion-exchange and size exclusion chromatography the ratio decreased resulting in low to no bound DNA or RNA contamination. In future the linker region has to be tested if it is capable to bind RNA and if so, if the binding results in a strong complex which protects the RNA from RNAses. If the binding can be confirmed *in vitro*, another experimental set up could be the deletion of the linker region *in vivo* to study its effect on RNA levels at the (peri)centromere.

In previous studies mass spectrometric analysis were not applied to detect potential factors associated with centromere transcription or splicing events. The nuclear fishing MS data of ZBTB9 presented in this study provides a unique insight into these processes and their possible involvement of centromere transcription and alternative splicing. Only DHX38 has been linked to be important for pre-mRNA splicing, yet the data around this topic is sparse[150,491–493]. One of the interactors in the ZBTB9 pulldowns was the RNA-binding protein SON, which acts as a mRNA splicing co-factor.

In an initial study, SON was found to be associated with the spliceosome, since component of the spliceosome were identified by mass spectrometric analysis, therefore links SON in

processing pre-mRNA[494]. Further analysis revealed upon depletion of SON, a decrease in cell proliferation was observed as well elevated levels of cells being in mitotic progression. The investigator found out that SON depletion activates the Spindle assembly checkpoint (SAC), specifically through MAD2. Moreover, SON deficiency resulted in altered chromosome congression due to activation of SAC which resulted in mis-aligned chromosomes and prolonged prometa and metaphase arrest[494]. In conjunction with SON, ZBTB9 is also affiliated with the SAC by interacting with Bub3, as seen in the nuclear fishing pulldowns. Bub3 has been shown to be essential for correct Kinetochore-Microtubule attachment and promotes stable end-on bipolar attachments[495]. In a previous study Bub3 has been shown to bind another zinc finger protein (ZNF207, BuGZ) which is associated with spindle microtubules and regulates chromosome alignment. Within BuGZ a conserved Bub3 interacting site (GLEBS, GLE-2-binding sequence) has been described, which directly binds and stabilizes Bub3. Moreover, the microtubule-binding domain enhances Bub3 loading onto kinetochores and this enhanced association is required for proper chromosome alignment and metaphase to anaphase progression[496]. Yet, with the nuclear fishing pulldown, ZBTB9 as another zinc finger containing protein have been associated with Bub3. The data is also in conjunction with ZBTB9's behaviour at the metaphase chromosomes, where it surrounds the DNA and mass spectrometric analysis (GO-terms chromatin condensation and microtubule/ cytoskeleton organization). Nonetheless, the association of ZBTB9 with Bub3 in the context of SAC has to be further evaluated.

Another study addressed how SON controls cell-cycle progression by coordinating the regulation of RNA splicing events[497]. A larger analysis of SON knockdown revealed its importance in cellular function by regulating gene expression. Pathways that were down or upregulation upon depletion of SON, were connected to DNA replication, recombination, or repair but also cell cycle, specifically G2/M checkpoint, mitosis and microtubule dynamics[497]. Moreover, RT-PCR experiments confirmed inefficient intron removal after SON was depleted. An interesting observation was the association of SON with transcripts of downregulated genes such as Aurora B and therefore indicates SON is directly involved in processing its target RNAs. In addition, the study pointed out the importance of SON in facilitating co-transcriptional assembly of SR proteins and other spliceosome components (U1 and U2AF) to elongating RNA Polymerase II complexes and therefore ensures efficient co-transcriptional RNA splicing[497]. Since SON has been shown to be specific interactor of ZBTB9 it is of interest how this connection is linked to the transcriptional process at the centromere. Moreover, ZBTB9 seems to be important as well in mitosis, since *in vivo*

microscopy and MS-data in this study suggests a role in chromosome condensation and ZBTB9 was also found to be linked to the CPC (chromosomal passenger complex) by interacting with Aurora B and Survivin (BIRC5). Specifically, Aurora B has been shown to be associated with centromeric transcription process, since pulldowns of Aurora B identified bound centromeric transcripts and facilitates its activity and localization at mitotic spindles[498]. Taken all together, ZBTB9 might be important for mitotic progression by interacting with CPC (Aurora B and Survivin), SON and the SAC (by Bub3) to mediate splicing of centromeric transcripts to stabilize the centromere in the mitotic phase. Moreover, ZBTB9 could be involved in correct bi-orientated microtubule attachment by interacting with CPC and SAC. Future studies have to determine whether the connection drawn here are valid *in vitro* and *in vivo* and if this might be one of the functional roles of ZBTB9 at the centromere.

Another less studied complex, the DBIRD complex, specifically the subunit ZN326 was associated with ZBTB9 in the nuclear fishing pulldowns. In one study DBIRD have been shown to directly interact with RNA Polymerase II (by size exclusion experiments and crude extracts) and regulates alternative splicing events of large exon substrates embedded in AT-rich DNA. Moreover, depletion experiments of DBIRD led to region-specific decrease in the transcript elongation process. The data suggests that DBIRD acts between mRNP (messenger ribonucleoprotein) particles and RNA Polymerase II by providing an interface for transcript elongation with the process of alternative splicing[499]. Furthermore, their data support the notion that DBIRD complex acts as an elongation factor facilitating transcript elongation in AT-rich regions, since the RNA Polymerase II has difficulties to transcribe at AT-rich sites and constitute very efficient elongation pause sites *in vitro*[500,501]. As shown in the nuclear fishing pulldown, ZN326 was specifically enriched and since the centromere consists of AT-rich alpha satellite sequences[502] a possible role of ZBTB9 could be, that it recruits the DBIRD complex by interacting with ZN326 to facilitate the transcriptional elongation process of centromeric transcripts. Nonetheless, this hypothesis has to be proven in future studies.

In the analysis of the ZBTB9 pulldown many kinases have been detected, specifically the master kinases CDKs driving the cell cycle. Since in the previous paragraph the notion has been established that ZBTB9 might be involved in transcription and splicing at the centromere, the regulating mechanism guiding the interaction have to be further elaborated. Specially the zinc finger pulldown showed a plethora of different CDKs (e.g., CDK1, CDK2,

CDK4/6, CDK5, CDK12/13) associated with ZBTB9. The discovery of different CDK kinases specifically at the zinc finger region promoted to study whether ZBTB9 itself is phosphorylation-dependent regulated and if it might mediate phosphorylation events required for transcription and RNA splicing. The *in vitro* phosphorylation assays with CDK1 and PLK1 confirmed the phosphorylation-dependent regulation of ZBTB9. Upon phosphorylation of the conserved S70 in the BTB domain and one not so conserved site near the zinc finger region, the binding to CENP-B could be reduced. Moreover, the phosphorylation of the linker region had no direct effect in CENP-B binding but might regulate its function in another cellular context, which has not yet been found. The question arises if ZBTB9 is regulated by so many different kinases at different cell cycle stages and/or if ZBTB9 mediates the phosphorylation of transcription or splicing associated interactors. Since ZBTB9 has been shown by sequence alignments with other species, that S70 is indeed a checkpoint of CDK1 phosphorylation and since it is deeply conserved it might serve as regulation knob for ZBTB9 interaction at the BTB domain. It can only be speculated that the phosphorylation at this specific site represses or increase the binding of an interaction partner. In case of CENP-B, the binding decreased but was not abrogated. CDK1 is known to facilitate the transition into mitosis[503]. Therefore, the phosphorylation of ZBTB9 by CDK1 might serve as a platform to progress into mitosis. Since in the previous paragraph ZBTB9 was associated with SAC and CPC this additional regulatory element would make sense and would imply an important role of ZBTB9 in the mitotic process. Similarly, CDK2 controls the G1/S and S/G2 transition[504]. Moreover, CDK2 has been shown to phosphorylate other transcription factors such as MYC or SMAD[505,506]. Therefore, a proposed mechanism could be that ZBTB9 is regulated by CDK2 during these transition cell cycle phases or might mediate CDK2 induced phosphorylation on other bound transcription factors to modulate their function. Next, CDK4/6 is required to enter S-Phase[504]. By phosphorylating retinoblastoma gene product (Rb) the Rb/E2F complex dissociates facilitating the transcription of E2F target genes including E-type cyclins[507]. In combination with CDK2 interaction, the activity of E2F target genes is increased and lead to the initiation of DNA synthesis[508,509]. In case of ZBTB9 the association with CDK4/6 might be to initiate the late S-Phase centromere DNA replication or mediate the phosphorylation of target substrates of CDK4/6 to guide the centromere DNA replication process. ZBTB9 was also associated with CDK9 which regulates RNA Polymerase II transcription elongation process[510]. This shows another example how ZBTB9 might be associated with centromeric transcription, by being regulated by CDK9 or mediating the phosphorylation to transcription

elongation associated proteins. In addition, CDK12 and 13 have been implicated in transition from transcriptional initiation to elongation by phosphorylating RNA Polymerase II at serine 2 but they are also regulating mRNA splicing, alternative splicing, 3'end processing as well as pre-replication complex assembly[504,511,512]. It remains to be investigated if ZBTB9 is phosphorylated at some point by CDK12 or 13 or if ZBTB9s presence at the centromere and binding to CENP-B influences the transcriptional landscape. Taken together, the high abundance of different CDKs in the nuclear fishing pulldown of ZBTB9 provided another glimpse onto its regulatory feature at the (peri)centromere.

In conclusion, ZBTB9 seems to be an integer part in the spliceosome environment and might have a role in modulating the mRNA processing steps at the alternative splicing complexes. Moreover, future studies have to clarify if components of the pre-mRNA splicing snRNP complex (e.g., RPM15, PRP4, SRSF4, PRP6, HNRPF, DHX39A or TADBP[478]) do interact with ZBTB9 and how ZBTB9 might be involved in the context of centromeric transcript processing and regulation.

As a last discussion part, the association of chromatin remodelling complexes within ZBTB9s nuclear fishing pulldown as well as with WICH-5 shall provide an insight on the possible role of ZBTB9 and WICH-5 in CENP-A replenishment and maintenance of (peri)centromere chromatin by interacting with different chromatin remodeler families. In the analysis of binding interactors in both pulldowns respectively, other chromatin remodeler families were associated with either ZBTB9 or WICH-5. Surprisingly, SMARCA5 was significantly enriched in the ZBTB9 pulldown, directly linking the ATPase domain of the WICH-5 complex to ZBTB9. Moreover, after establishing the binding of both complexes to CENP-B even at the same time, the complex might be involved directly/ indirectly in CENP-As replenishment at G1-Phase with or without the help of CENP-B. Since the previous dataset of the lab identified WICH-5 specifically in G1-Phase, this leads to the assumption that WICH-5 indeed could be involved in this process of CENP-A replenishment which have been restricted to G1-Phase. Till today a chromatin remodeler for CENP-A incorporation has not been determined. In the initial data presented in this thesis, specifically SMARCA2, SMARCA4, RSF1, SMARCA5, INO80, Arp5/4/8 were identified either in ZBTB9 or WICH-5 nuclear fishing pulldown respectively. Besides WICH-5 belonging to the ISWI family, the enriched interactors belong to two other novel chromatin remodeler families: SWI/SNF and INO80 family. SMARCA2/4 belong to the SWI/SNF family and build in their respective complexes the PBAF or BAF complex[422]. This remodeler family anchors to the histone

protein and translocate 1-2 bp of DNA along the surface of the nucleosome. Depending on which catalytical ATPase is present in the complex, the end product will result in nucleosome sliding or removal of H2A/H2B dimer from the histone core and creating an open chromatin environment[513,514]. The BAF complex is either recruited to actively or repressively marked chromatin and recruits further interactors[515]. In case of INO80 family the ATPase unit is formed by INO80 and includes several subunits such as actin-related ones (e.g., Arp4/5/8). The INO80 complex is known to remove canonical H2A and replace it with H2A.Z or mediates the removal of the aforementioned mark by the histone chaperone ANP32E as a primary step before DNA repair[516–518].

The question arises whether each complex exists in their respective boundaries or do different chromatin remodeler families share their subunits to create novel chromatin remodelers for specialized purposes? The remodeler community does not provide much information if subunits are shared between the families but recently it has been shown that CHD7 and CHD9 were interacting with PBAF/BAF as seen by reciprocal immunoprecipitation experiments. Moreover, the cooperation between CHD7 and PBAF at NC distal elements synergistically regulate NC gene expression and migration as seen *in vivo*.[519] Another recent paper described the expansion of the NuRD complex with ZBTB2 as a novel component[520]. The latter one implicates that ZBTB9 can expand the ISWI-family by its association with SMARCA5 and CENP-B, forming a novel complex.

As described earlier, subunits of the INO80 complex were detected. Recent studies showed a link of the INO80 complex to the maintenance of the centromere in various species. Previously, INO80 has been shown to be involved in removal of H3 containing nucleosomes and assisting direct CENP-A[Cnp1] chromatin assembly. Moreover, INO80 complex was shown to be associated with centromeric regions in a CENP-A[Cnp1] dependent manner in fission yeast. When INO80 complex was tether to non-centromeric regions resulted in spreading of CENP-A[Cnp1] into these regions[521]. In addition, in *S.pombe* the 35 kDa big Hap2 has been identified as an auxiliary subunit of the INO80 complex and is required for the conversion of H3 chromatin to CENP-A[Cnp1] chromatin on naïve centromere DNA. Moreover, the Hap2-INO80 promote a transcription-associated chromatin remodelling event, driving H3 nucleosome eviction and the assembly of CENP-A[Cnp1] [522]. This ties back to the initial findings of this study. Here ZBTB9 and WICH-5 have been shown to recruit the INO80 complex and might facilitate CENP-A incorporation maybe by a similar manner

(transcription replated) as described for *S.pombe*, since its centromere is a 'simplified' regional centromere but share similarities to human centromere[164].

In addition, another chromatin remodeler of the ISWI family have been identified in the ZBTB9 pulldowns: the RSF complex (RSF1 and SMARCA5). RSF1 has been transiently associated with CENP-A chromatin and it localizes at the centromere region in early mid-G1 phase but dissociates in transition to G1/S-Phase[280,523]. Depletion experiment of the RSF complex members have been shown to arrest cells in prometaphase and increase kinetochore misalignment in metaphase. Importantly, the RSF complex is required to deposit CENP-A into centromeric sites to establish the nucleation site of kinetochore attachment later in mitosis. *In vitro* assays showed a lower enzymatic activity of recombinant purified RSF complex, implying that additional factors are required for its function[280]. Till today additional factors have not been characterized. The current study might provide a unique insight by associating RSF with ZBTB9 and WICH-5 complex. In future, biochemical reconstitutions and *in vitro* nucleosome assays have to verify the finding and if the complexes are capable of binding each other or even associate at the same cell cycle phase *in vivo*. By performing depletion experiments the interplay between each complex member, this shall provide an insight on how the complexes are organized and if the depletion of one member results in abrogated CENP-A incorporation.

The multifaceted interaction network of ZBTB9 and WICH-5 expands further by the significant enrichment of the FACT complex in both pulldowns. FACT (Facilitates Chromatin Transcription, Spt16 and SSRP1) is a transcription elongation factor but also histone chaperone function in DNA replication by disassembling nucleosome in front of a processing RNA Polymerase II[524,525]. The complex is conserved among eukaryotes yet varies in its importance regarding certain cell lineages[526]. Interestingly, in WICH-5 also CENP-X have been found to interact with the complex. Similar behaviour has been seen in FACT binding to the CENP-T/W complex. The C-terminal region of Spt16 binds to histone fold in CENP-T/W and Spt16 have been shown to be important for CENP-T/W deposition at centromere sites[292]. An additional binding partner of FACT can be found within the CENP-H/I/K/M complex, were CENP-H is the main interactor for the FACT complex. Moreover, it has been shown that CENP-H/I/K/M complex is required for the octameric targeting of newly synthesized CENP-A[527]. In *Drosophila* it has been shown that CENP-A[CID] integrity is maintained and established by histone chaperone CAL1 interacting with FACT and also recruiting RNAPII directly during CENP-A deposition. The deletion of FACT causes

defective CENP-A^{CID} deposition and loss of transcription[159]. Furthermore, a recent study concluded that Spt6 is important for eviction of H3 but also recycles previous deposited CENP-A through a possible phosphorylation dependent interaction with Spt6. Therefore, Spt6 plays a pivotal role in *Drosophila* to maintain CENP-A centromeric mark in a long-term run[232]. The WICH-5 complex have been associated with the RNA Polymerase III by mediating the Max-dependent c-Myc binding[528], hence the data presented here includes the association with RNA Polymerase II transcriptional process by identifying factors such as RPB1C, RPB3, RPAB1, SPT4H, WDR61 or RTFAF. As stated earlier, the FACT complex can incorporate CENP-A in a transcriptional manner, the proposed mechanism of WICH-5 might be through the probable binding to CENP-X and the confirmed binding of CENP-B, that WICH-5 mediates CENP-As replenishment in G1-Phase in similar fashion as FACT has shown. As an additional mediator, ZBTB9 in conjunction with CENP-B might guide SMARCA5 to the centromeric site and facilitate the complex formation with BAZ1B to form the WICH-5. Accessory interactors such as RSF or INO80 promote a unique complex formation to ensure correct incorporation of CENP-A into the centromeric chromatin and ZBTB9 keeps the important euchromatin/ heterochromatin boundaries up to create the optimal environment for CENP-A.

These examples show that the possibility exists that ZBTB9 with WICH-5 and CENP-B as well with other remodeler family subunits would form a specialized complex or at least mediate the process to replenish CENP-A during G1-Phase. Future experiments have to verify if ZBTB9 co-exists in the presented chromatin remodelling complexes *in vivo* and if they biochemically form stable complexes *in vitro*. Moreover, it would be interesting to study if ZBTB9 has a more repressive or activating function within the remodeler complexes to facilitate CENP-A incorporation at the centromeric chromatin sites.

An interesting observation was made where some identified interactors of ZBTB9 were also involved in translation. This annotation was surprising since the centromere does not encode for proteins. Moreover, the centromeric transcripts made by RNA Polymerase II are spliced in smaller fragments and mostly involved in stabilizing the surrounding centromere associated proteins at the (peri)centromere and might also aid in heterochromatin formation. Recent studies suggested the existence of microproteins. It has been shown that the presence of small ORFs at the 5'-untranslated region (UTRs) of some mRNAs have an important function. Usually, these small ORFs in these UTRs are often referred as upstream open reading frames (uORFs). These elements regulate translation by engaging the

ribosome before it initiates the downstream ORF. It has been shown that the translation of this specific uORF is required for instance in the efficient translation of AdoMet carboxylase. Moreover, this provides the first evidence that such microproteins are functional relevant and expand our understanding of the proteome. The smORFs were further identified in noncoding regions, such as 3'UTR[529–531]. A recent study by Ruiz Cuevas *et.al* (2021) described a method to study non-canonical proteins in DLBCL (diffuse large B cell lymphoma) cells. They found 14,498 proteins of which 17% (2,503) were non-canonical. Moreover, of these 2,503 non-canonical proteins 28% were new isoforms and the majority were cryptic proteins (72%). It has been shown that these cryptic proteins are more disordered and unstable as classic proteins are. By combining Ribo-seq and RNA sequencing data, they created a non-redundant sample-specific protein database, which contains actively translated sequences. In combination with mass spectrometry, they were able to identify a subset of non-canonical proteins (9%). Deeper analysis showed that 12% of the protein sequences found, were derived from non-coding RNAs or processed transcripts[532]. Since ZBTB9 has been linked to the translational process as well it might be attempting to think, that maybe a transcription factor is mediating the translation of microproteins at the (peri)centromeric region, which might lead to stabilisation mechanisms or even signalling cascades facilitating the recruitment of the kinetochore to the centromere by these novel microproteins. Future studies have to dissect if microproteins do really exist in general and at the centromere and if they play a pivotal role in centromere organization.

As a concluding remark the proposed function of the novel centromere associated ZBTB9 as well as of the ISWI chromatin remodelers WICH-5 and CHRAC1-5 are summarized in Figure 26. In summary, the WICH-5 complex might be involved in the replenishment of CENP-A during G1 by interacting with ZBTB9 or maybe ZBTB9 might mediate the process. Moreover, both have been shown to bind to CENP-B and therefore provides a direct link to the centromere. ZBTB9 has been shown to harbour multifaceted functions at the (peri)centromeric region, ranging from heterochromatin 'gatekeeper' to mediator of transcriptional process and manager of splicing events. Moreover, ZBTB9 could additionally be important during mitosis by assisting correct micro-tubule attachment and therefore ensures faithful chromosome segregation by associating with CPC and SAC components. Future studies have to validate the data presented here in a cellular context and verify if ZBTB9 indeed is involved in these processes proposed.
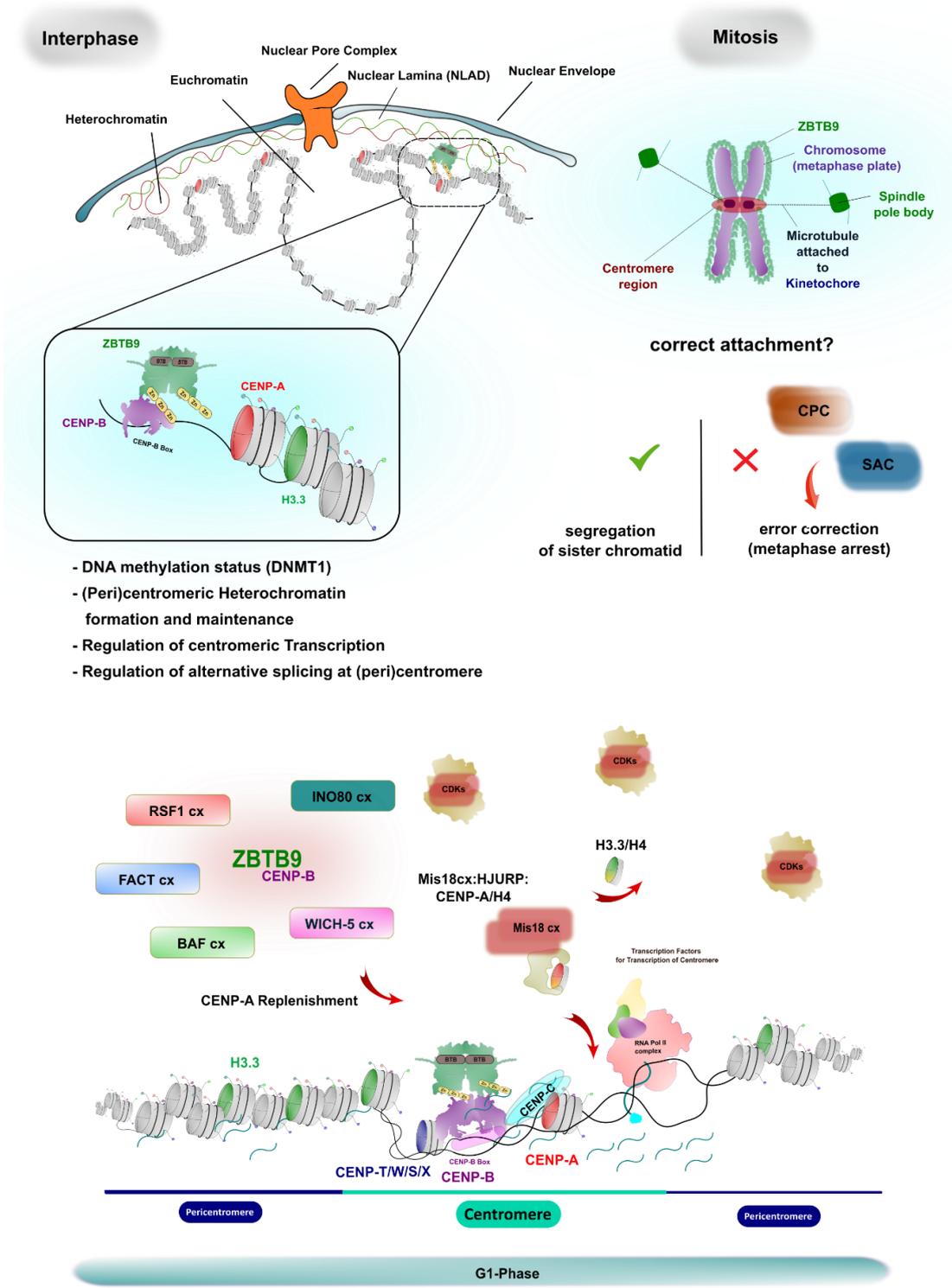
Figure 26: **Possible Function of ZBTB9 in the context of the centromere based on the experimental evidence obtained in this study.** ZBTB9 localizes probably with its BTB domain near the nuclear lamina and maintains by interacting with a plethora of different heterochromatin

associated proteins the centromeric chromatin integrity. It poses to have a function in managing DNA methylation patterns and regulating centromeric transcription and alternative splicing. It might be important during mitosis to condensate chromosomes but also interacting with CPC and SAC to sense incorrect microtubule attachments. The last scheme proposes how CENP-A could be replenished during G1-Phase. In the spotlight ZBTB9 and CENP-B as well as the different chromatin remodelers, which might contribute incorporation of CENP-A onto the centromere by evicting histone H3.3. The RNA Polymerase II transcribes small cenRNA, which helps to stabilize and maintain the centromere. CDKs and their associated cyclins at each cell cycle stage have a regulatory role.

6 Alteration of chromatin state by post-translational modifications: Development of an O-GlcNAc/ -GalNAc Peptide Enrichment Matrix coupled to LC-MS/MS

6.1 Results and Discussion

As in the previous paragraph discussed, the novel centromere interactors could be linked to the centromere and inherit important functions to maintain or establish the centromeric chromatin by facilitating the deposition of heterochromatin marks or opening the chromatin at this specific site for CENP-A incorporation by chromatin remodelers. Since phosphorylation has been shown to play a role in ZBTB9s function another modification has yet not been associated with the centromere: O-glycosylation. The following study was conducted to identify possible O-glycosylated nuclear proteins and investigating if centromere interactors might be O-glycosylated as well and if this has an impact on the centromere integrity. This project could not be finished during the time period, but preliminary results suggest potential role of O-glycosylation also at the centromere and provides a promising outlook for future studies.

6.1.1 Boronic acid-based enrichment matrix for specific O-glycopeptide enrichment

In the last decade many researchers developed multiple MS-based methods to characterize the Glycoproteome. A well characterized glycosylation form is the N-glycosylation on asparagine. Current methods are standardized, straightforward and glycans can be easily cleaved off by universal glyohydrolases such as PNGaseF. This method enables the identification of N-glycosylated peptides, the site occupancy, and its abundance level. Furthermore, latest publications took the standardized methods and implemented second enrichment matrices to detect phosphorylated peptides and glycosylated peptides in a single step[533,534].

Besides N-glycosylated proteins, the more complex and versatile O-glycosylation proposes till today analytical challenges and is therefore underrepresented in the literature. The overall low abundance of these modifications as well as their labile nature hinders its identification and quantification. Therefore, the development of an enrichment matrix to specifically enrich O-glycosylated peptides have emerged. Within the last few years,

researchers developed very sophisticated affinity-enrichment based matrices which rely on the specific chemical properties of these glycans attached to the affinity matrix. The resulting enriched peptides can be detected at high sensitivity due to the development of new mass spectrometric instruments and methods. The sugar residues can interact with different chemicals to form reversible interactions. One of these chemicals is boronic acid.

Boronic acid interacts with saccharides in a reversible manner and have been characterized since the 1950s[535]. The binding of carbohydrates to the boronic acid is influence by the pH, solvent, chemical group linked to the boron as well which type of diol can be bound. The interaction is mediated between the hydroxyl groups in the sugar moiety and the boronic acid. Through derivatization the specificity can be tailored to specifically enrich O-glycosylated proteins[535,536]. The project is based on a publication published in nature comminictions[537] , where the investigators developed a matrix capable of enriching N- and O-glycosylated peptides at alkaline conditions by screening first boronic acid derivatives. The investigators found out that Benzoboroxole has a high affinity to carbohydrates and therefore renders the interaction stronger. This ability helps to enrich low abundant glycopeptides as well. The peptides of a complex mixture are incubated with the self-synthesized boronic acid-based matrix. Through washing steps non-glycosylated peptides are removed and through acidic elution, the interaction between the boronic acid and carbohydrate can be disrupted. Leaving a reusable matrix. The acidic solution contains all glycosylated peptides with O-linked glycans attached. In case for N-glycopeptides, the peptides can be treated with PNGaseF to release the glycan. By adding a buffer containing heavy Oxygen (^{18}O), the attachment site on a given asparagine can be determined by a specific mass shift enabling an easy way to identify the site on a given peptide[537].
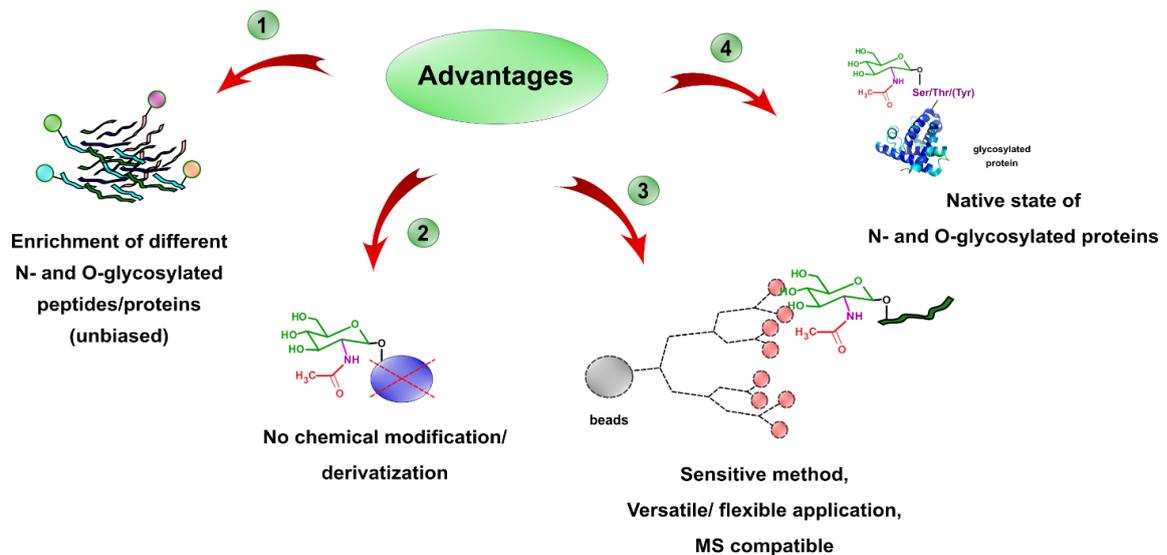
Figure 27: **Advantages of applying a boronic acid assisted matrix to enrich specifically O-glycosylated peptides.** The development of a boronic acid-based matrix to enrich O-glycosylated peptides leads to an unbiased enrichment. There is no need to chemically modify or derivatize the glycan attached. It is a sensitive method allowing the enrichment of low abundant peptides and can be applied in a universal manner with different suitable MS-based applications. Through this enrichment it is possible to preserve the native state of N- or O-glycosylated peptides and providing information on its structure and attachment side.

This enrichment matrix has following advantages (Figure 27): The enrichment of N- and O-glycosylated peptides can be performed simultaneously. By fractionating the cellular content (cytoplasm, nuclei, mitochondria, Golgi, vesicles, or peroxisomes) the identification of specific glycosylated peptides for each given component can be achieved. Secondly, there is no need to chemically modify or derivatize the attached glycan by harsh chemical reactions. This method has also been successfully applied, thus leading a bias to certain glycans, since not all glycans effectively react with a given chemical. The matrix can capture glycosylated proteins, therefor the native state of a N- or O-glycosylated site can be determined. This method due to its sensitivity and flexibility can be combined with different mass spectrometric methods.

This method described was the starting point for this project to address the open questions regarding the influence of O-glycosylation in chromatin biology, specifically investigating the influence of O-glycosylation modification on nuclear proteins and their effect on the centromere organization. By comparing different sources and methods of O-glycopeptide enrichment, this paper outperformed known enrichment techniques at that time (2019, e.g. lectin based enrichments or hydrophilic-liquid interaction chromatography). The idea to

perform a nuclear enrichment, which was not done in the original publication, led to the hypothesis that this might boost the identification of O-glycosylated peptides. The enrichment steps were carried out as outlined in Figure 28 A, by following the steps described in the publication. As "internal" standard the HILIC enriched peptides were taken as benchmark. If the Matrix can detect the prior HILIC enriched peptides (which have been separately enriched by the same method) then this indicates that the Matrix is capable of enriching glycosylated or phosphorylated peptides. As a drawback in the initial analysis, a MS run of the only HILIC enriched peptides (without further enrichment by the prototype Matrix) was not performed. Therefore, it is difficult to say if the peptides enriched were indeed glycosylated or phosphorylated. Moreover, when comparing the overall enrichment of non-prior enriched samples, the log2 iBAQ values seem to be less than for the actual BA-matrix pulldown. It was assumed that the HILIC enrichment would yield a higher abundance of these peptides. Unfortunately, that was not the case. It could be that the performance of the HILIC enrichment was performed incorrectly and have to be optimized in future. As depicted in Figure 27 A, the peptides generated from HeLa nuclear extracts were taken as input for the direct enrichment on the prototype BA-Matrix and were prior enriched by the HILIC method. Using alkaline conditions, the equilibrium of the interaction between boronic acid residues and the hydroxy-groups at the glycan or phosphor-group attached is shifted towards the binding of these. By washing the beads under these conditions, non-bound or non-specific peptides should be washed away, leaving only glycosylated, phosphorylated or a combination of both on a peptide bound to the matrix. By applying acidic conditions, the interaction of the boronic acid with the glycan/ phosphor-group is disrupted, leading to the elution of the desired modified peptide. After obtaining the enriched peptides, the most suitable MS-method have to be employed to ensure glycopeptide identification. As a starting point the method described in the aforementioned publication[537] was applied first, since they also used a Orbitrap Elite Instrument, which has been used in this study too. One alteration was done by performing the MS/MS acquisition in the ion trap, rather than in the orbitrap. Moreover, the fragmentation technique used was CID and HCD to investigate which might be superior in identifying glycopeptides. This already has been investigated by others in numerous studies. Often HCD has outperformed CID Fragmentation by providing more information regarding glycan attached and in general identification score for the actual glycopeptide[538,539]. The initial analysis is described in the next paragraph.

6.1.2 MS-Analysis reveals the potential enrichment of glycosylated and phosphorylated peptides as well as which protein groups might be targeted by O-glycosylation

The first analysis has shown that a prior enrichment with the HILIC method is not required. In fact, it reduced the overall identification of possible glycosylated peptides compared to the boronic acid method (Figure 29 B, C). In total 69 proteins could only be identified with the HILIC method, whereas the boronic acid method was capable of finding up to more than 800 proteins. When comparing the numbers to the study of *Xiao et.al*, the enrichment of glycoproteins by HILIC showed an enrichment of 500 proteins. Here the investigators were only comparing their enrichment with the HILIC but did not apply the same procedure as done in this study. Next, in total 234 O-Glycoproteins were identified [537]. With the current setting this study could enrich about 600 proteins, when applying certain filtering. Moreover, surprisingly the CID method performed in the identification of proteins better than the HCD method (Figure 29 C, bar graph) but when comparing abundance of each protein, the HCD method was more sensitive. In the initial study they did not compare if CID or HCD was a better fragmentation technique. Notably, the novel synthesized BA-Matrix in this study could detect centromere specific proteins (CENP-B, CENP-N, CENP-F and CENP-V) but in case of CID only CENP-V was detected. All the others were detected with HCD. In addition, other proteins, which have been identified in the Nuclear Fishing Pulldown of ZBTB9 or the ChIP-MS of CENP-A (see previous paragraph 5.1.4), were identified in this O-glycosylation enrichment. These include CHD1, BUB3, SMARCA5, BAF, NPM3, MCM3, MCM2 DEK, CBX5, SMARCA1, RSF1 or UHRF1. The listed proteins have been identified in previous O-glycosylation screens but till today the role of O-glycosylation on their function has not been studied. As already mentioned in the previous project, CENP-B and CENP-N are important interactors of the inner kinetochore. Yet, CENP-V has been less studied. CENP-V has been shown to be localized to the kinetochore during mitosis[540]. Moreover, RNAi experiments caused broadening of primary constriction at the centromere but did not affect heterochromatin levels. Its depletion causes delocalization of the chromosomal passenger complex components Sgo1. Moreover, CENP-V seems to be required for proper alignment of metaphase chromosome and correct cytokinesis and overexpression of CENP-V causes death of the cell[540]. But nothing is known on its post-translational modifications. Yet, in this current study suggests that glycosylation might play a part too. This expands the post translational modification of kinetochore proteins by bearing a possible carbohydrate attached. It has to be noted, that the identification is only made by the presence of the

peptides detected. Since the possible glycol-site could not be detected, it remains speculative but provides evidence for a possible alteration by glycosylation. If centromere proteins are truly altered by an attachment of O-GlcNAc and how this is coincides with phosphorylation remains elusive and shall be addressed in future studies. Since the HILIC enriched peptides identified served as "internal" standard, within the 22 identified proteins, the protein with the lowest log2 iBAQ value (n=2) was applied as filtering option for the larger boronic acid dataset. In both cases of each dataset (CID or HCD fragmentation applied) the protein SRRM2 was applied as benchmark for filtering. This protein had the lowest iBAQ and was one of the 22 common proteins identified in both sets with the HILIC enriched one. The benchmark for CID and HCD was set to a log 2 iBAQ value of 13.6 or 14.3 respectively. In case of the CID dataset after applying the threshold, in total 658 proteins were identified above the threshold. In case for the HCD data set in total 447 proteins were identified. When comparing both datasets, 100 proteins were in common with both datasets, yet the rest being unique to each fragmentation technique applied (Figure 28 B, C). By a manual assignment of the HILIC dataset, the proteins were compared to a recently published data source for glycosylation (GlyGen and O-GlcNAc database)[541,542]. It is possible to see which proteins are known to be glycosylated and if the site has been identified. In Table 15 and 16 an attempt was made to manually validate these proteins in common and if identified peptides might carry a glycosylation or could be phosphorylated. In case of Table 16, some of the proteins with their respective identified sequences are depicted to provide evidence for the possible enrichment of glycosylated peptides with the novel matrix synthesized. To compare the identified peptides with the phosphorylation status of the protein, the Phosphosite Plus[446] database was applied to identify if the peptide was solely or additionally phosphorylated. In total 22 proteins were in common, out of 16 proteins with their respective identified peptides (based also on their MS/MS count) were possibly glycosylated, only a small portion of the 22 proteins were actually only phosphorylated.

As in previous studies shown, O-glycosylation is a part of the histone code[543]. As proof of concept the histone H4 could be detected in this study with the possible glycosylation site (S48), which has been already identified in previous studies but its function in the nucleosome context remains elusive. Another histone H2B has been shown to be glycosylated too. In this study the matching peptide has been identified, providing evidence the matrix was able to enrich O-glycosylated peptides. In contrast to H4, H2B has been studied in its O-glycosylation context. It has been shown that S112 at the C-term of H2B is

targeted by OGT and promotes K119 monoubiquitination, which has been linked with active transcription[544]. Another notion to add here is, that H2B K119ub is also a mark at the centromere and has been shown to be associated there as well with active transcription[152], therefore O-glycosylation might play a role at the centromere too. Therefore, O-GlcNAcylation links to transcriptional activation. The investigators also interlinked the cellular availability of nutrients for the process of O-GlcNAc transfer by the Hexosamine-biosynthetic pathway. By depleting glucose in the media of human cultured cells or adding any drug to prevent the turnover of UDP-GlcNAc substrate it became evident that the glycosylation on S112 is directly linked with S119 ubiquitination process. Therefore, if glycosylation is impaired it is possible that transcription is downregulated[544].

This example of H2B S112 O-glycosylation provides an insight on how this specific modification modulates chromatin organization in order to facilitate transcription.

A

Enrichment by BA-PD | Prior Enrichment by HILIC

HILIC

1

2

HO–O–O–B
OH
B–O–O

Binding in alkaline conditions
(pH > 8.5)

3 | Wash (organic solvent)
pH 10

HO–O–O–B
HO–O–O

4 | Elution (acidic)
pH 2.0

5

LC-MS/MS
(CID vs HCD Fragmentation)

● Peptide with Phosphor-group (S/T)
● Peptide with O-GlcNAc/ O-GalNAc/ O-glyco-group (S/T)

B

Proteins identified in BA-PD
CID vs. HCD Fragmentation

CID   HCD

658   100   447

Proteins in common with
both datasets

C

Proteins enriched prior with HILIC
and enriched via BA-PD

found in O-GlcNAc-Database

CID   HCD

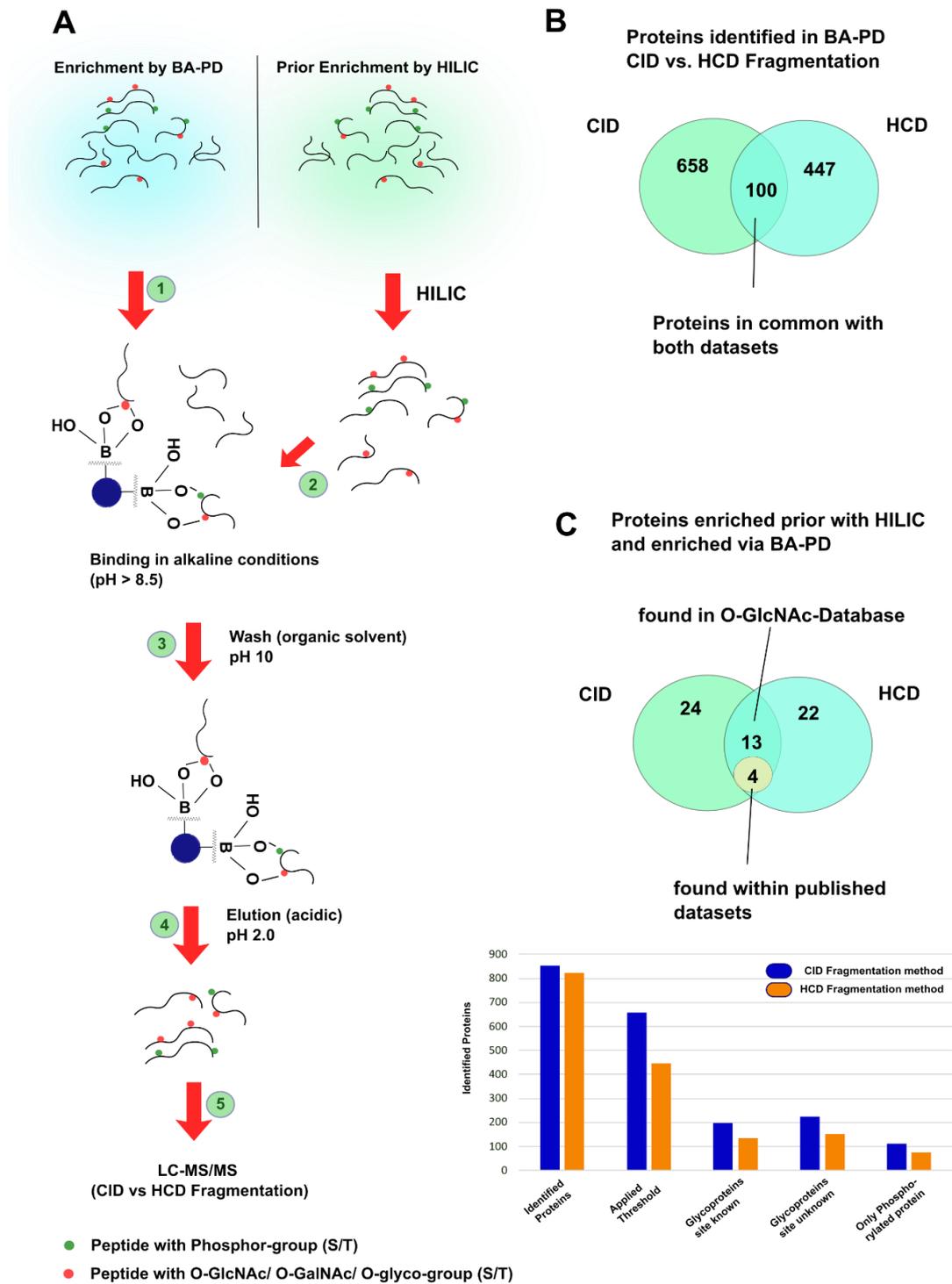24   13   22
      4

found within published
datasets

Figure 28: **Schematic of Boronic-Acid pulldown and integrated HILIC enrichment of simultaneous enrichment of glyco- and phospho-peptides and the performance of the synthesized prototype. A)** Schematic overview of the Boronic-Acid Pulldown (BA-PD) steps

developed in this study. Peptides obtained after digesting nuclear extract are either directly incubated with the BA-Enrichment Matrix under alkaline conditions, washed (organic solvent and alkaline conditions). Glyco- and phosphor-peptides are eluted under acidic conditions and subjected to LC-MS/MS analysis. **B)** Protein identified in each pulldown after filtering out contaminants, reverse identified proteins and only proteins identified once in one of the pulldowns (n=4). Each enrichment was measured once in CID mode and then in HCD mode and fragmentation spectra were generated in an ion trap. **C)** As "internal" standard served the enrichment of prior enriched HILIC peptides. Proteins that were enriched in both modes under HILIC conditions were taken as benchmark to obtain information on enrichment performance. After manual validation of proteins in the HILIC condition the actual numbers for glycoprotein and phosphoproteins were calculated (see bar graph)

Next, attempts to identify the glycol-peptide by the GlycReSoft[545] software developed by Joseph Zaia's group did not yield satisfactory results. The software could not find any glycopeptides within the dataset given. After contacting one of the developers (Klein, Joshua) to exclude any bugs within the software provided, it became evident that the dataset created did not recorded the specific oxonium ion needed to determine whether the peptide has been glycosylated. During the fragmentation, in the best scenario, the peptide has its labile carbohydrate still attached while being in gas phase and entering the mass spectrometer. By applying HCD, the fragmentation spectra shall provide information on the peptide backbone and therefore the sequence but also the detachment of the glycan attached. Many software's uses the detection of an oxonium ion to strengthen the identification of a glycosylated peptide. In case for O-GlcNAc, the oxonium ion is detected at m/z 204[546]. Since a bug in the software could be excluded, the method itself could be the problem why an oxonium ion could not be detected, which seemed to be the case. In the initial method the detection window for the ion trap was set to m/z 380. Therefore, the oxonium ion could not be detected in most of the cases. For a higher resolution on the oxonium ion the orbitrap had been the better choice. Therefore, this concludes that the current status of the method is, that it might have captured glycosylated peptides but cannot be confirmed by the current settings. In theory all single spectra for each protein have to be evaluated manually by judging the peptide backbone fragments combined with the precursor mass and the delta mass created by cleaving off the glycan by CID or HCD. Since a certain mass shift is expected, this should provide evidence that this might have been glycosylated. Unfortunately, the dataset was not remeasured with adjusted settings. Therefore, it remains to be optimised in future studies.
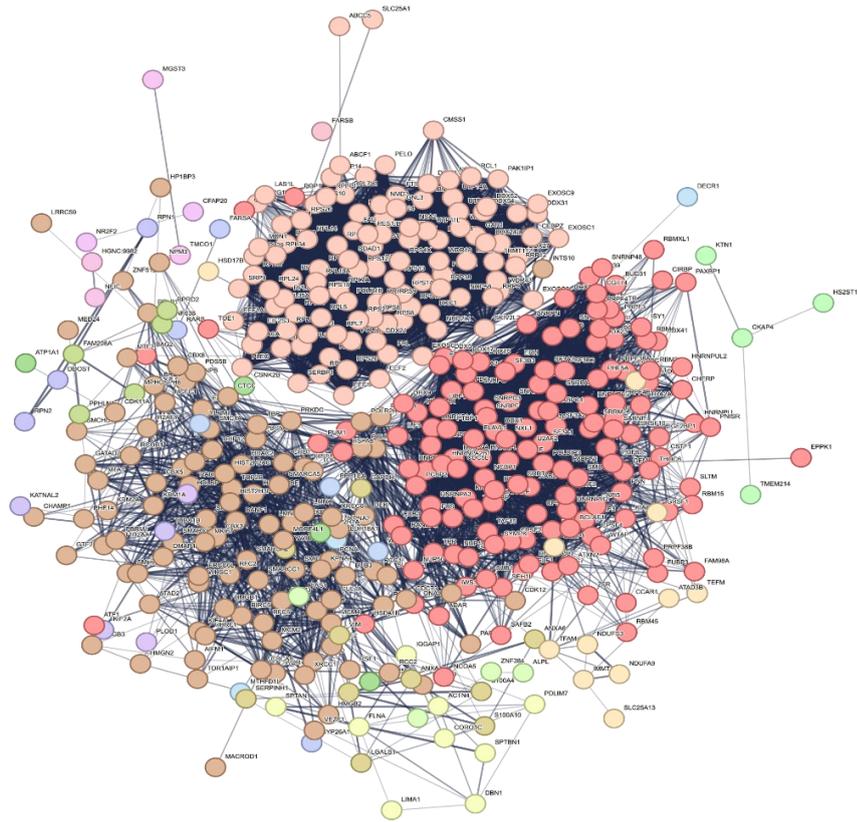
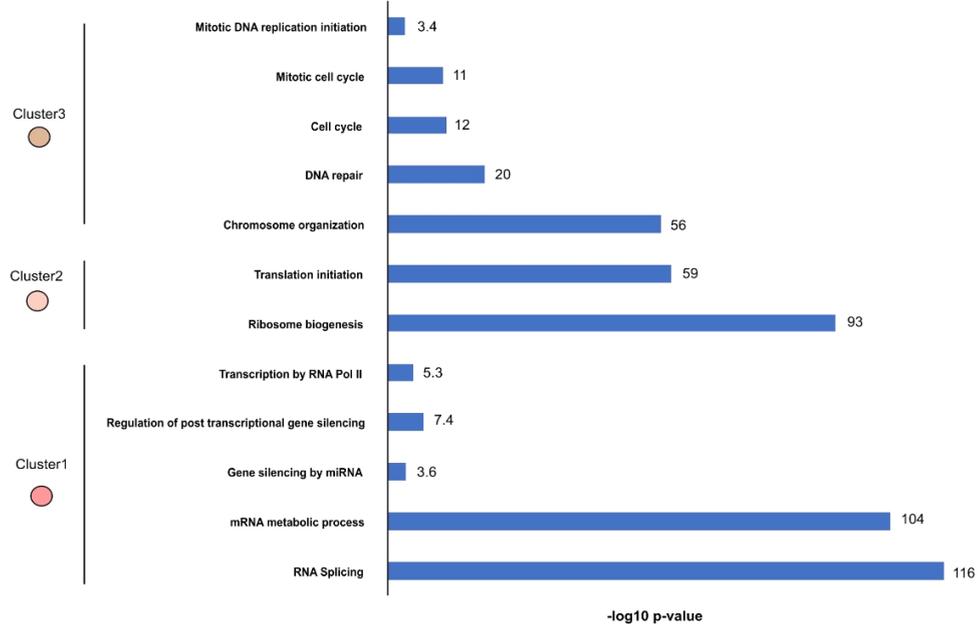**Figure 29**: **STRING and DAVID analysis of enriched proteins by the BA-Matrix prototype. A)** STRING network of identified proteins after applying a cut off above the lowest log2 iBAQ value

measured within the HILIC enrichment. **B)** Biological processes annotated with the clusters observed and their respective -log10 p-values. Shown are p-values corrected for multiple testing within each category using the Benjamini-Hochberg procedure.

Table 9: **Proof of concept by manually validating the HILIC enriched proteins identified in both data sets.** Each protein was looked up in the GlyGen and O-GlcNAc database as well as Phosphosite Plus site to determine whether the identified peptide could harbour the glyco-site known or a possible phosphorylation site or even both.

Protein shared between HILIC and BA	Identified glycosylated peptide?	Identified phosphorylated peptide?	MS/MS peptide count
DX39A	Potentially yes, site not identified yet	Several peptides could be phosphorylated	9
HMGB3	Potentially yes, site not identified yet	Several peptides could be phosphorylated	13
CPSF5	Potentially yes, site not identified yet (multiple sites possible)	Several peptides could be phosphorylated	8
ROA1	No, peptides missing	Enrichment of phosphorylated peptides possible	8
H2A.V	Potentially yes, site not identified yet (multiple sites possible)	Enrichment of phosphorylated peptides possible	5
HNRPL	Potentially yes, site not identified yet	Enrichment of phosphorylated peptides possible	5
H1.2	Potentially yes, site not identified yet	Enrichment of phosphorylated peptides possible (also acetylation)	8
ROA2	Yes, one of two annotated glycosites (S48)	Enrichment of several phospho-peptides	6
SFPQ	No, three sites not found in peptides	Enrichment of several phospho-peptides	6

MATR3	Yes, one peptide harbouring 6 glyco-sites (S61, S71, S74, S75, S77, S85)	Enrichment of several phospho-peptides	8
ROA3	Yes, three known glycosites (S84, T82, T173) in two peptides	Several phospo-peptides, some not annotated modifications possible	2
H4	Yes, only glycosite (S48)	no	24
TRA2B	Potentially yes, site not identified yet	Enrichment of several phospho-sites in one peptide (T201, T203, Y207, S212, S213, S215, S216)	13
EF1A1	Potentially yes, site not identified yet; O-GlcNAc Database (T261, T279) identified	No phosphorylation reported, might be the site of glycosylation occupation	16; 5; 6
H3.3	No, peptide with S11 missing	Enrichment of phospo-peptide with T58	6
HNRPU	No, peptide missing	Enrichment of phosphor-peptides	5
RL18	Yes, peptide with T39 identified	Enrichment of several phospo-peptides	4
DHX9	Yes, one peptide harbouring three known O-GlcNAc-sites (S1142, S1152, S1153)	Enrichment of several phosphor-peptides	4
SSRP1	No, known site peptide missing	Enrichment of several phosphor-peptides	5
NONO	No, known site peptide missing	Enrichment of several phosphor-peptides	7
HMGN3	Potentially yes, not known site, here no since peptide does not contain S/T	No phospho-site known, site known for acetylation/ ubiquitination	15

Protein			
SRRM2	Yes, S2259 identified in one peptide	Enrichment of possible phosphor-peptides	5

Table 10: Proteins identified in the HILIC enriched fraction and comparing the identified peptides with the known glycosylation sites and their mentioning in publications.

Protein	Sequence	Glyco-site known	Mentioned Publication
MATR3	LASLMNLGMSSSLNQQGAHSALSSASTSSH NLQSIFNIGSR	S61, S71, S74, S75, S77, S85	18
H4	RISGLIYEETR	S48	12
EF1A1	VETGVLKPGMVVTFAPVNVTTEVK	T261, T279	15
RL18	TNSTFNQVVLK	T39	10
DHX9	PSAAGINLMIGSTR	S1142, S1152, T1153	13
SRRM2	TPAIPTAVNLADSR	S2259	23

Although the unambiguously identification of the glycosylated peptides is not possible and therefore the total identification of glycosylated proteins, attempts were made to determine a rough estimation of the enrichment of glycosylated peptides by the matrix. After manually validating the identified peptides for the 22 common HILIC enriched proteins within the dataset, the numbers are calculated back to the HCD and CID dataset and are shown in the bar graph in Figure 29 C. It was estimated that about 30 % of the dataset were containing probably "true" glycosylated proteins where the site has been known (CID Dataset: 197 proteins; HCD Dataset: 143 proteins). Another 34 % are probably glycosylated as well but have not been identified yet by their site (CID Dataset: 223 proteins; HCD Dataset: 152 proteins). Since the boronic acid enrichment matrix has been shown to enrich phosphorylated peptides as well[547], it was assumed where peptides could not be assigned to a glycosylated sites or were not known to be glycosylated, were assigned to be phosphorylated (17%, CID Dataset: 112 proteins; HCD Dataset: 81 proteins). The remaining 19% are non-specific interactions or enrichment background. This has to be further confirmed, since it could be that these peptides might have multiple attachments which could not be assigned in this study.

Since the manual validation provided some hints if the dataset might contain possible glycosylated peptides and their corresponding proteins, the proteins obtained after filtering were bioinformatically analysed by STRING and DAVID Analysis (Figure 29 A, B). Here both analyses yielded similar annotations in biological processes and molecular function. Therefore, the annotation of the STRING database was taken and depicted in a bar graph beneath the STRING network. After applying the MCL clustering, three main clusters were obtained, of which were taken and analysed for their relevance in biological processes. As seen in cluster 1, proteins enriched were having a common theme in RNA splicing, mRNA metabolism or gene silencing. Till today there are no reports covering the influence of O-glycosylation in RNA splicing events or mRNA metabolism. The second cluster hinted the importance of O-glycosylation in ribosome biogenesis and translation initiation. Indeed, the latter biological process has been linked directly to O-glycosylation and its effects on the translational machinery by modifying core ribosomal proteins[548]. In a study published in 2010, the investigators studied the influence of O-GlcNAcylation on core ribosomal proteins, which have been identified prior in global glycoproteomic studies before. O-GlcNAcylation has been implicated to support protein stability and cellular protection during stress response. How this modification alters the translational machinery remains elusive. In the study several subunits of the core ribosomal units (40s and 60s) were shown to be glycosylated and their sites were identified. Moreover, the enzymes OGT and OGA are associated with different subpopulations of ribosomes and their interaction might propose a regulatory role in ribosome biogenesis and translational processes[31,548]. By preparing isolated polysomes from HepG2 cells the investigators could identify subpolysomal fractions as well as heavy and light polysomes to be O-GlcNAcylated[31]. In one specific instance, the RPS6 has been shown to be phosphorylated and glycosylated at the same time and mutations of the phosphorylation sites lead to abrogated O-GlcNAcylation[31], bridging both PTMs and underlines the importance of their cross talk. In another experiment they found out that overexpression of OGT and OGA increases their association with polysomes and therefore O-GlcNAcylation might play a role in ribosomal subunit homeostasis[31]. Another recent study investigated the O-Glycoproteome under stress response and also found transcription factors to be glycosylated. By using a metabolic labelling approach and subsequent click based chemistry the investigators could identify 990 azide-labelled O-GlcNAc proteins with high confidence[549]. Their analysis also resulted in gene ontology terms like RNA splicing, cell cycle or chromatin remodelling[549]. Indeed, the identified proteins within this study also contained ribosomal proteins and STRING/DAVID analysis resulted

in similar gene ontology terms, therefore this suggests that the BA-matrix prototype might have the potential to enrich glycosylated peptides, even though the MS/MS data could not assign glycosylated peptides due to the loss of the oxonium ion reporter. In future studies, this has to be further improved and validated.

The last cluster (nr.3) revealed that the chromatin organization indeed can be altered by glycosylation, since biological terms such as chromosome organization, DNA repair and even cell cycle were enriched. In case of DNA repair in a recent study OGA has been directly linked to the DNA damage repair machinery. It has been shown that upon DNA damage, OGT relocates to DNA lesion and catalyses O-GlcNAcylation at DNA lesions and the cross talk with other PTMs might add another layer of DNA damage repair regulation[550]. OGA has been shown in the study to be deglycosylating target proteins during DNA damage response. Moreover, its HAT domain plays a part in localizing OGA to the DNA lesion but also interacting with NONO and Ku70/80, two important components of the non-homologous end joining (NHEJ) DNA double strand break repair (DSB). Additionally, the deglycosylation of NONO is required for its chromatin association and NHJE repair mechanisms. Prolonged association of NONO due to its impaired deglycosylation lead to DSB repair defects. Therefore, the deglycosylation of NONO and Ku70/80 is important to avoid NHEJ repair defects at DNA lesions[550]. The glyco-enrichment matrix in this study could identify some of the proteins associated with DNA repair and therefore provides further evidence that his prototype could capture glycosylated peptides, also by the fact that NONO had been identified as well.

Taken all involved biological processes together, these were in common with already published studies, providing evidence that the matrix indeed is capable of capturing glycosylated peptides [537,551,552]. Notably, the distinctions of glycosylated to phosphorylated peptides makes it difficult to assign certain biological processes to one of each modification. The rational here is, since the matrix might only enrich 17% of the phosphorylated peptides (in contrast to maybe 30 % of possible glyco-enrichment) the majority of proteins in the analysis are either glycosylated or maybe having an additional phosphorylation site. The depiction of truly phosphorylated peptides is underrepresented. Nonetheless it has to be considered, since the cross talk between phosphorylation and glycosylation plays a role in a cellular context[21], the tandem purification might provide unique insights into the O-Glycoproteome and Phosphoproteome of the nucleus and the chromatin environment.

In conclusion, the presented data here provides the first steps in a method development to enrich specifically O-glycosylated peptides and simultaneously phosphorylated peptides in one experiment. The MS-method has to be further improved to ensure proper detection of the oxonium ion required for correct assignment of the glycosylated peptide by proving the fragmentation of the attached glycan. The binding conditions should be optimized to develop a protocol which favours at one hand the enrichment of glycosylated peptides or the other hand the phosphorylated peptides. At this stage of the development the matrix resulted in a prototype which might be capable of identifying glycosylated and phosphorylated peptides in one run. Future steps to complete the development of this enrichment matrix includes the complete synthesis of the matrix and its control experiments to ensure correct synthesis steps. Moreover, the limit of detection and limit of quantification has to be determined to find the optimal standard concentration of different peptides generated from lysates to efficiently enrich the desired modification at low abundance, since tissue or rare patient materials are limited in their sample size and amount. Therefore, it would be good to know how sensitive the method can be. Another cornerstone of the method should be the integration of specific enzyme cleavage of attached glycans at specific sites. By using specialized human glycohydrolases, the removal of the glycan and subsequent attachment of a reporter label at this site could improve the side identification. In this study attempts were made to purify some of these enzymes recombinantly. The expression and purification of OGA and the less studied HEXD was promising. HEXD (Hexosaminidase D, ß-6-SO$_3$-N-acetylgluosaminidase) and NAGA (Alpha-N-acetylgalactosaminidase). HEXD is capable of cleaving off O-GlcNAc and O-GalNAc from proteins. It has a preference for O-GalNAc cleavage[553]. This protein is reported to be in the nucleus as well. By using this enzyme for cleavage, peptides having either O-GlcNAc or O-GalNAc could be detected. In combination with OGA treated peptides it can be determined, which proteins might be targeted by one or the other to determine their specificity. Unfortunately, it could not be tested whether the enzymes were active and if they would cleave a glycosylated substrate. By integrating this enzymatic cleavage with the enrichment matrix, provides versatile information on glycosylated proteins in one experimental condition. Future attempts have to address the aforementioned improvements to ensure highly specific enrichment of O-glycosylated peptides/ protein and also include novel strategies which have been developed within the last two years[549,551,554]. The first steps were initiated.

6.2 Material and Methods

6.2.1 Cloning of human kinetochore, chromatin remodelers, ZBTB9 constructs and CENP-A/H3.3 Histones

The human kinetochore constructs were cloned from a cDNA library kindly provided by Prof. Andrea Musacchio. Genes encoding each kinetochore protein was PCR amplified with specific gene primers as well with specific overhangs for Gibson assembly to the vector backbone of a pLIB vector. Each construct was verified by sequencing. For complex expression, the individual pLIBs served as template to assemble one pBIG vector with all complex members in different cassettes via Gibson assembly. Each construct was verified by sequencing. Following proteins were tagged N- or C-terminally with a 8xHis-tag followed by a small linker and a second Twinstrep-Tag (TW) and a precission protease cleavage site (see Table 9). If needed, the proteins were incubated with the 3C precission protease to cleave off the tag. The histones were pre-assembled in a modified pET-28 vector for a polycistronic expression in *E.coli*. Since this strategy did not work, the histones were re-cloned for single expression in the same vector backbone. Each histone was tagged with an N-Terminal 8xHis Tag followed by a 3C precission cleavage site. The starter methionine of each histone was not cloned to obtain tagged version of the histones. In case of CENP-A, another strategy was followed. Here CENP-A (untagged, with starter methionine) was cloned into a pRSF-Duet vector backbone (cassette I) and its binding partner H4 (untagged, with starter methionine) into cassette II. By subsequent co-expression, the soluble tetramer complex could be obtained. The ZBTB9 wt as well as its constructs were PCR amplified from a cDNA clone (Origen) with overhangs for Gibson assembly and cloned into a modified pET-28 vector, cloning an N-terminal MBP-Tag followed by a precission protease site and a C-terminal 8xHis tag. PCR primers were designed with Tm-Calculator of NEB or ThermoFischer. PCR primers were ordered at Sigma-Aldrich as HPLC-purified version. The amplicon was obtained with standard PCR protocols (NEB) with High Fidelity Phusion Polymerase (NEB). Purified PCR products (Kit, Invitrogen) were mixed in a molar ratio of Vector:Insert (1:5) and the standard protocol for Gibson assembly at 50°C for 60 min was performed. The reaction was transformed into DH5 alpha cells and clones showing correct restriction digest pattern were sequenced and correct insertion of the GOI was confirmed. These constructs were further used for biochemical studies.

Table 11: **Proteins cloned with or without indicated Tag and expressed in different expression systems.**

Protein	N-term Tag	C-term Tag	Vector	Expression system
ZBTB9 (full-length)	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9 Δ 42-112	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9 Δ 381-462	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9 Δ 409-433	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9 Δ 438-460	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9$^{19\text{-}112}$	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9$^{113\text{-}410}$	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9$^{379\text{-}473}$	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9$^{406\text{-}479}$	MBP-precission	8xHis	pET-28	*E.coli*
ZBTB9$^{19\text{-}112+380\text{-}473}$ Fusion	MBP-precission	8xHis	pET-28	*E.coli*
CENP-A	-	-	pRSF-Duet	E.coli
H4	-	-		*E.coli*
H2A	6xHis-precission	-	pET-Duet/ pET-28	*E.coli*
H2B	-	-	pET-Duet/ pET-28	*E.coli*
H3.3	8xHis-precission	-	pET-28	*E.coli*
Baz1a	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
Baz1b	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
Snf2h (SMARCA5)	-	-	pLIB/pBIG	Insect cell

Chrac1	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
DPOE3	-	-	pLIB/pBIG	Insect cell
CENP-C (full-length)	GST-precission/ 8x-His-TW-precission	8xHis	pLIB/pBIG	Insect cell
CENP-C$^{2-118}$	GST-precission	8xHis	pLIB/pBIG	Insect cell
CENP-C$^{+PEST119-719}$	GST-precission	8xHis	pLIB/pBIG	Insect cell
CENP-C$^{-PEST\ 419-719}$	GST-precission	8xHis	pLIB/pBIG	Insect cell
CENP-C$^{546-943}$	GST-precission	8xHis	pLIB/pBIG	Insect cell
CENP-B	8xHis-TW-precission or 1x FLAG	8xHis (for FLAG construct)	pLIB/pBIG	Insect cell
CENP-B$^{2-129}$	MBP-precission	8xHis	pET-28	*E.coli*
CENP-N	-	-	pLIB/pBIG	Insect cell
CENP-L	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
CENP-I	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
CENP-H	-	-	pLIB/pBIG	Insect cell
CENP-K	-	-	pLIB/pBIG	Insect cell
CENP-M	-	-	pLIB/pBIG	Insect cell
CENP-O	-	-	pLIB/pBIG	Insect cell
CENP-P	-	-	pLIB/pBIG	Insect cell
CENP-Q	-	-	pLIB/pBIG	Insect cell
CENP-U	8xHis-TW-precission	-	pLIB/pBIG	Insect cell
CENP-R (iso5)	8xHis-TW-precission	-	pLIB/pBIG	Insect cell

6.2.2 Chemical crosslinking and MS-Analysis

The complex containing CENP-A Nucleosome$^{Widome601,168\ bp}$ with either purified WICH or CHRAC1 complex was assembled in equimolar ratios in solution. The complex was incubated for 20 min at room temperature before crosslinking it. It was cross-linked using

an equimolar mixture of isotopically light (hydrogen) and heavy (deuterium) labelled bis[sulfosuccinimidyl]suberate (BS3, H12/D12) (Creative Molecules) at a final concentration of 0.25–0.5 mM at 30°C for 15 min in a buffer containing 30 mM HEPES, pH 8.2, 150 mM NaCl, 5% glycerol and 1 mM DTT. The reaction was quenched by adding ammonium bicarbonate to a final concentration of 100 mM for 10 min at 30 °C subsequent protein digest and mass spectrometry (see below).

The cross-link reaction for ZBTB9, CENP-B and CENP-C[546-943] was performed as followed. The proteins were mixed in an equimolar ratio and incubated for 20 min at room temperature. Afterwards the crosslink reaction was set up with an equimolar mixture of isotopically light and heavy labelled BS2 (shorter linker then BS3, H12/D12; Creative Molecules) at a final concentration of 0.5 mM at 30°C for 15 min in a buffer containing 30 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol and 1 mM DTT.

Cross-linked samples were denatured by adding two sample volumes of 8 M urea, reduced with 5 mM TCEP (ThermoFisher) and alkylated by the addition of 10 mM iodoacetamide (Sigma-Aldrich) for 40 min at RT in the dark. Proteins were digested with Lys-C (1:50 (w/w), FUJIFILM Wako Pure Chemical Corporation) at 35°C for 2 hr, diluted with 50 mM ammonium bicarbonate, and digested with trypsin (1:50 w/w, Promega) overnight. Peptides were acidified with trifluoroacetic acid (TFA) at a final concentration of 1% and purified by reversed phase chromatography using C18 cartridges (Sep-Pak, Waters). Cross-linked peptides were enriched on a Superdex Peptide PC 3.2/30 column using water/acetonitrile/TFA (75/25/0.1, v/v/v) as mobile phase at a flow rate of 50 µl/min and were analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) using an Orbitrap Elite instrument (ThermoFisher).

The mass spectrometer was operated in data-dependent mode, selecting up to 10 precursors from a MS1 scan (resolution 60,000) in the range of m/z 350–1800 for collision-induced dissociation excluding singly and doubly charged precursor ions and precursors of unknown charge states. Dynamic exclusion was activated with a repeat count of 1, exclusion duration of 30 s, list size of 300, and a mass window of ±50 ppm. Fragment ions were detected at low resolution in the linear ion trap.

Fragment ion spectra were searched, and cross-links were identified by the dedicated software *xQuest* (Walzthoeni et al., 2012). The results were filtered according to the following parameters: Δscore ≤ 0.85, MS1 tolerance window of −3 to 3 ppm and score ≥ 25.

The quality of all cross-link spectra passing the filter was manually validated (with 1% FDR) and cross-links were visualized as network plots using the webserver *xVis* (Grimm et al., 2015).

6.2.3 Expression and Purification of ZBTB9 wt and constructs from *E.coli*

pET-28 vectors containing ZBTB9 wt or other constructs mentioned in this study were transformed into 50 µL Rosetta (DE3) cells with the heat shock method for 43 seconds at 42°C. Cells were put on ice for 2 min before recovering them in 100 µL SOC media for one hour at 37°C. Cells were streaked out on chloramphenicol/ kanamycin plates and grown over night. The next day, two-three clones of the plate were taken as overnight culture, grown in LB-media and corresponding resistance. The following day a preculture of 600 mL with antibiotic resistance were inoculated with the overnight culture and grown additionally for 3 hours at 150 rpm. 12 L of YT-media (2xLB) were prepared in 6x 5L Erlenmeyer Flasks by filling in each 2 L of media with culture (100 mL in each). The cells were grown at 37°C till the OD_{600} reaches 0.2. Then the cells were cooled down till 24°C till they reach OD_{600} 0.4. Afterwards the temperature is again decreased to 16°C till they reach OD_{600} 0.6. Then the cells were induced with 100 µM IPTG over night at 16°C (18 h). After induction, the cells were harvested, flash frozen in liquid nitrogen and stored at -80°C or directly subjected for purification. The cells were resuspended in a buffer containing 50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 1 mM DTT, 20 µM $ZnCl_2$, 0.05% Tween-20, 10 mM Imidazole, protease inhibitors and subsequently lysed by sonication in a cooled metal beaker. The sonication process was performed three times for 5 min intervals of duty cycle 40% (output control 5) with two minutes cooling session in between. The lysed cells were centrifuged for 25 min at 17,500 rpm at 4°C. The clear lysate was loaded onto pre-equilibrated Ni-NTA beads (Qiagen) for batch purification and incubated 1 h while rotating at 4°C. The beads were washed afterwards with a buffer containing 50 mM HEPES (pH 7.5), 1 M NaCl, 5% glycerol, 1 mM DTT, 20 µM $ZnCl_2$, 20 mM Imidazole. Another washing step with 500 mM NaCl and 30 mM Imidazole was performed. The protein was eluted with a buffer containing 50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 1 mM DTT, 200 mM Imidazole. The elution's containing the protein of interest were diluted to a salt concentration to 200 mM NaCl with a buffer containing 30 mM HEPES (pH 7.8), 1 mM DTT. The eluates were loaded onto a HiTrap Heparin FF (Cytiva) and washed with 30 mM HEPES (pH 7.8), 250 mM NaCl. The protein was eluted with a gradient B till 100 % (2 M NaCl). Fractions containing the protein were collected and concentrated for the last purification step, size exclusion

chromatography. Here a superose6 increase 10/300 GL (Cytiva) was used with a buffer containing 30 mM HEPES (pH 7.5), 250 mM NaCl, 5% glycerol, 1 mM DTT. The protein was eluted with a linear gradient and fractions containing soluble protein were collected and concentrated to the desired concentration used in the biochemical assays. The protein was aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

6.2.4 Expression and Purification of human kinetochore complexes or chromatin remodelers from insect cells

A P2 (generation 2) Virus was generated with SF-21 cell line for each human kinetochore subcomplex or for chromatin remodelers. 2-6 L of HighFive cells were transfected with P2 Virus (1:100 ratio) of each complex and incubated for 48-72 h at 27°C, 95 rpm. The cell viability of each construct and cell diameter was checked to ensure protein expression and no contamination. Cells were harvested and flash frozen in liquid nitrogen or directly subjected for purification. The human kinetochore protein subcomplexes or single proteins were lysed in a buffer containing 50 mM HEPES (pH 8.0), 300 mM NaCl, 5% glycerol, 1 mM DTT, 1 mM $MgCl_2$, 0.05% Tween-20, 10 mM Imidazole, protease inhibitor. In case for CENP-C and CENP-B the salt concentration was set to 500 mM. The cells were resuspended in the buffer and dounced in a glass pestle with 30 strokes. The lysate was centrifuged at 17,500 rpm for 25 min at 4°C. The supernatant was incubated with pre-equilibrated Ni-NTA beads (Qiagen) and incubated for 1 h at 4°C rotating. After incubation, the beads were washed in a buffer containing 50 mM HEPES (pH 8.0), 300 mM NaCl, 5% glycerol, 1 mM DTT, 35 mM Imidazole. In case of CENP-C and CENP-B an additional washing step containing 1 M NaCl was performed to remove bound DNA. After washing the protein complexes were eluted in 50 mM HEPES (pH 8.0), 300 mM NaCl, 5% glycerol, 1 mM DTT, 200 mM Imidazole. The subsequent purification steps after eluting the complexes from Ni-NTA as described in Yatskevich *et al.* (Science, 2022) were followed for all the subcomplexes, by using a HiTrap Heparin FF or a HiTrapQ FF column (Cytiva). In case of CENP-B and CENP-C, no additional step was performed. The proteins were directly subjected to size exclusion chromatography. As a last step all complexes (except CENP-O/P/Q/U/R) were subjected to size exclusion to obtain stoichiometric and soluble complex (mostly Superdex200 increase 10/300 GL, Cytiva) in a Buffer containing 30 mM HEPES (pH 8.0), 250 mM NaCl, 5% glycerol, 1 mM DTT. The fractions containing the complex were collected and concentrated to the desired concentration (1 mg/mL), aliquoted, flash frozen in liquid nitrogen and stored at -80°C for further use in biochemical assays.

The chromatin remodelers (CHRAC1 and WICH complex) were transfected and expressed as described for the human kinetochore subcomplexes. Cells were harvested and directly subjected for purification by resuspending in a buffer containing 50 mM HEPES (pH 8.2), 420 mM NaCl, 5% glycerol, 1 mM DTT, 20 µM $MnCl_2$, 20 µM $ZnCl_2$, 0.05% Tween-20, 10 mM Imidazole. The cells were dounced as described and the lysate was centrifuged at 17,500 rpm for 25 min at 4°C. The ready supernatant was incubated with pre-equilibrated Ni-NTA beads and incubated for 1 h at 4°C while rotating. The beads were washed with a buffer containing 50 mM HEPES (pH 8.2), 420 mM NaCl, 5% glycerol, 1 mM DTT, 35 mM Imidazole. The remodelers were eluted with a buffer containing 50 mM HEPES (pH 8.2), 420 mM NaCl, 5% glycerol, 1 mM DTT, 200 mM Imidazole. An optional anion-exchange step (MonoQ 5/50, Cytiva) was included where the protein solution was diluted to a salt concentration of 250 mM with a buffer containing 30 mM HEPES (pH 8.2), 1 mM DTT. The complex was eluted with a linear gradient till 100% B (2 M NaCl). Fractions containing the complex were subjected for polishing step with size exclusion chromatography in a buffer containing 30 mM HEPES (pH 8.2), 420 mM NaCl, 5% glycerol, 1 mM DTT. Fractions containing the complex were collected and concentrated (1 mg/mL). Small aliquots were flash frozen in liquid nitrogen and stored at -80°C.

6.2.5 Expression and native purification of CENP-A/H4 Tetramer and H2A/B Dimer or single Histone purification under denaturing conditions from *E.coli*

Since the purification of CENP-A via denaturing condition did not yield pure amounts of histone, the purification of CENP-A was changed. CENP-A/H4 were co-expressed in Rosetta2 (DE3) cells. The plasmid was transformed into the Rosetta2 (DE3) strain and the heat shock method was applied. The transformed cells were plated out on LB plates containing chloramphenicol/ kanamycin resistance. The cells were incubated over night at 37°C. Two clones were picked to inoculate the overnight culture with corresponding antibiotic resistance. The following day a preculture of 300 mL with antibiotic resistance was inoculated with the overnight culture and grown for 3 h at 37°C. Afterwards the culture was equally distributed onto 5 L Erlenmeyer Flasks containing 2 L of LB media with corresponding antibiotics. The cells were grown till OD_{600} 0.2 at 37°C, then the temperature was decreased to 24°C till OD_{600} 0.4 is reached. Upon reaching OD_{600} O.4, the temperature was decreased again to 20°C till the cells were at OD_{600} 0.6. The protein expression was induced with 300 µM IPTG over night at 20°C for 18 h. The cells were harvested and

subjected directly for purification. The cells were resuspended in 80 mM Phosphate-Buffer (Na_2HPO_4/NaH_2PO_4, pH 6.9), 500 mM NaCl, 2 mM DTT, protease inhibitors. The French press was used to lyse the cells through one cycle at 18 kpsi. The lysate was centrifuged at 26,000 x g for 15 min at 4°C. The supernatant was sonicated at a duty cycle of 80% (output control 1) for 20 sec (three times). The sonicated supernatant was again centrifuged for 26,000 x g for 15 min at 4°C. The supernatant was loaded onto a HiTrap SP FF (Cytiva) prior equilibrated with Lysis buffer and then washed with Wash buffer (80 mM Na_2HPO_4/NaH_2PO_4, pH 6.9, 500 mM NaCl, 2 mM DTT) until the UV absorbance base line was reached. The complex was eluted with a gradient of Buffer A (80 mM Na_2HPO_4/NaH_2PO_4, pH 6.9) and B (80 mM Na_2HPO_4/NaH_2PO_4, pH 6.9, 2 M NaCl), from 25% B to 50% B for 20 min, then 50 – 100% B for 10 min, collecting 500 μL fractions. Pure CENP-A/H4 tetramer eluted at higher salt concentrations (1 M NaCl). Fraction samples were then concentrated using a 10 kDa concentrator up to 1 mg/ml. SEC was performed using a Superdex200 increase 10-300 GL (Cytiva) columns in SEC buffer (25 mM Tris-HCl, pH 7.5, 2 M NaCl, 5% Glycerol, 2 mM DTT). Peak fraction containing soluble CENP-A/H4 Tetramer were collected and concentrated to 1 mg/ml. 20% glycerol was added for long term storage, frozen in liquid nitrogen and stored at -80°C.

The expression of histone H2A/H2B was performed similarly. The purification was different to the just mentioned protocol. Cells were lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 2 M NaCl, 5% glycerol, 1 mM DTT, 10 mM Imidazole, protease inhibitor. The cells were sonicated at 40% duty cycle (output control 5) for three times with 5 min interval (2 min cooling phase in between). The lysate was centrifuged for 25 min at 15,500 rpm at 4°C. The supernatant was loaded onto a pre-equilibrated HisTrap FF (Cytiva). After washing (30 mM Imidazole) the complex was eluted with a buffer containing 50 mM Tris-HCl (pH 7.5), 2 M NaCl, 5% glycerol, 1 mM DTT, 200 mM Imidazole and fractions containing protein complex were subjected to size exclusion chromatography in a buffer containing 25 mM Tris-HCl (pH 7.5), 2 M NaCl, 5% glycerol, 1 mM DTT on a Superdex200 increase 10-300 GL (Cytiva).

The expression and purification of single histones and subsequent reconstitution to an octamer, were performed as described by Luger and co-workers at denaturing conditions. Briefly, Rosetta (DE3) cells were transformed with vector containing each histone singly (tagged with N-terminal 6-8x His-precission). Clones were picked to inoculate overnight cultures. The next day 12 L of LB media were inoculated with a preculture grown out of the

overnight culture. The cells grew till OD_{600} 0.6 at 37°C. Protein expression was induced with 400 µM IPTG for 18 h at 37°C so histones were mainly found in inclusion bodies later. The cells were after expression harvested and stored at -80°C or directly subjected for purification. The cells were lysed in a buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM DTT, protease inhibitors. The cells were lysed by sonication at 40% duty cycle (output control 5) for three times in a 5 min interval (2 min cooling down). The lysate was cleared with a centrifugation step at 15,500 rpm, 4°C for 25 min. The supernatant was discarded, and the pellet was resuspended in buffer B (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1 mM DTT, 1% Triton-X100, protease inhibitors) and then centrifuged at the same speed as describe previously. The pellet was treated in total three times with this buffer. Next the pellet was treated with the lysis-buffer two additional times and centrifuged. After the last centrifugation step, the supernatant was discarded, and the pellet was resuspended in 1 mL of DMSO and set for 30 min at room temperature. Afterwards the pellet was resuspended in an unfolding buffer containing 50 mM Tris-HCl (pH 8.0), 7 M guanidine hydrochloride, 500 mM NaCl, 1 mM DTT. The extraction process was performed for one hour. After histone extraction, the lysate was centrifuged at 15,500 rpm, 4°C for 25 min. The supernatant was dialysed against a buffer containing 50 mM Tris-HCl (pH 8.0), 8 M Urea, 500 mM NaCl, 1 mM DTT, 10 mM Imidazole and 10 mM L-Lysine. After dialysis the supernatant containing unfolded histone was incubated with pre-equilibrated Ni-NTA beads (Qiagen) for one hour at room temperature. The beads were washed with the same buffer prior mentioned. The histone was eluted with a buffer containing 50 mM Tris-HCl (pH 8.0), 8 M Urea, 500 mM NaCl, 1 mM DTT, 200 mM Imidazole. Eluates containing pure histone fractions are dialysed overnight in a buffer containing 15 mM Tris-HCl (pH 8.0). If aggregates occur, centrifuge the eluate at high speed in a tabletop centrifuge. The eluate contains soluble histone and is further subjected for anion-exchange purification. Since histones are highly positively charged, they will pass through the HiTrapQ column, but DNA remains bound. Flowthrough was collected and concentrated to 5 mg/mL. The protein solution was lyophilized by a vacuum freeze dryer. After obtaining lyophilized histones, the histones were stored at -80°C until further use.

For reconstituting an octamer, the lyophilized histones were resuspended in unfolding buffer (25 mM Tris-HCl (pH 7.5), 7 M guanidine hydrochloride, 1 mM DTT) and mixed in equimolar ratios with slight excess (2% more) of H2A and H2B to ensure proper octamer assembly. After obtaining an equimolar ratio, the mixture was diluted to 1 mg/mL in unfolding buffer. The histone mixture was dialysed overnight in reconstitution buffer (25 mM Tris-HCl (pH

7.5), 2 M NaCl, 1 mM DTT). Aggregates were centrifuged down and soluble octamer is subjected to size exclusion chromatography in buffer containing 25 mM Tris-HCl (pH 7.5), 2 M NaCl, 1 mM DTT. Fractions containing soluble octamer were concentrated to 1 mg/mL and 20% glycerol was added as preservative. The octamer solution was aliquoted and stored at -80°C.

6.2.6 *In vitro* Binding assays

Proteins were measured at the nanodrop for their absorbance at 280 nm. The estimated extinction coefficient was taken from the EXPASY ProtParam Tool for the given protein sequence. The molarity of the protein solution was calculated. For each reaction, a 50 µL reaction was set up. The proteins used were mixed at 2 µM equimolar. The remaining solution was filled up with Binding buffer (30 mM HEPES (pH 7.5), 150 mM NaCl, 3% glycerol, 0.5 mM DTT) to a total reaction of 50 µL. 1/3rd of the reaction was taken out as input. The rest of the solution was transferred to the pre-equilibrated beads used. For MBP-Pulldowns (MBP-fused proteins) 15 µL bed volume were used. In case for FLAG-IP (FLAG-tagged bait), 15 µL bed volume have been used. The beads were prior equilibrated in Binding buffer. The beads were centrifuged at 1500 rpm for 3 min at 4°C. The supernatant was taken off and the beads were quickly dried, and protein solution was directly added to the beads. The interaction took place in a thermomixer for one hour at 1200 rpm at 4°C. The beads were washed three times with Wash Buffer HS (30 mM HEPES (pH 7.5), 300 mM NaCl, 3% glycerol, 0.5 mM DTT, 0.05 % Tween-20) and one time with Binding buffer. The beads were centrifuged in between, and the supernatant was sucked off. After the last washing step, the beads were dried and 15 µL of Elution buffer was added (for MBP: 30 mM HEPES (pH 7.5), 150 mM NaCl, 3% glycerol, 0.5 mM DTT, 20 mM Maltose; for FLAG: 30 mM HEPES (pH 7.5), 150 mM NaCl, 3% glycerol, 0.5 mM DTT, 1 µg/mL FLAG-Peptide) and incubated for 10 min at 30°C, then 5 µL 4x SDS-loading Dye was added. Equal amounts of input and binding assay were loaded onto a TGX precast SDS-PAGE gel (BioRad) and run for 40 min at 230 V. The gel was stained in brilliant Coomassie Blue and destained in 10% acetic acid till background diminished. Each assay was at least performed in duplicates or triplicates. For Phosphorylation-Assays, the bait was prior bound to MBP-beads with the same method described. 1/3rd of the solution was taken out as input. The remaining protein solution was incubated. The beads with bound bait were washed with 30 mM Phosphate-Buffer (pH 7.5), 150 mM NaCl, 1.5 mM MgCl$_2$ and 2 mM ATP. In a separate tube the kinase (here yeast CDK1 or human PLK1) and the reaction buffer 30 mM HEPES (pH 7.5), 150

mM NaCl, 1.5 mM MgCl$_2$ and 2 mM ATP were added (0.5-1 µM kinase) were combined and added to the beads. The reaction was performed for 35 min at 30°C at 1200 rpm. The reaction was stopped by washing the beads three times with Wash buffer HS and once with Binding buffer. Then the prey protein was added and incubated for one hour with the phosphorylated bait protein. The beads were processed as mentioned above and samples were loaded onto a TGX SDS-PAGE Gel for analysis. For MS- phosphorylation site analysis, the corresponding bands of phosphorylated bait protein was cut out and in gel digested (Ignasi Forné, ZfP, BMC, LMU Munich) and subjected to MS-Analysis. Data was evaluated with Scaffold 5 Program.

6.2.7 *In vitro* reconstitution of Nucleosomes

The *in vitro* reconstitution of nucleosomes was performed by the salt dialysis method (slight modified protocol from Luger and Co-workers; AG Hopfner (LMU) modified protocol). In a pre-experiment (total 10 µL) the optimal DNA:octamer or DNA:tetramer:dimer ratio was estimated for future larger reconstitutions. The DNA used was either a 167 bp long Widom601 sequence:

5'ACTTACATGCACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATC
CCCTTGGCGGTTAAAACGCGGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGA
GCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTCCAGGGCGGCCAGT3'
or a 171 bp alpha satellite DNA containing the CENP-B box (a kind gift by Dr. Nikolina Sekulic, Norway; <u>CENP-B box</u> is highlighted:

5'ATATGGGA<u>TTTCGTTGGAAACGGGA</u>TCAACTTCCCATAACTGAACGGAAGCAAACT
CAGAACATTCTTTGTGATGTTTGTATTCAACTCACAGAGTTGAACCTTCCTTTGATAGT
TCAGGTTTGCAACACCCTTGTAGTAGAATCTGCAAGTGTATATTTTGACCACTTTG3'.
In case for DNA:octamer different molar ratios were tested to find the optimal nucleosome assembly. In case of DNA:tetramer:dimer reconstituted CENP-A/H4 tetramer (10 µM) were mixed with H2A/H2B dimer (20 µM). The reaction was assembled as followed (4°C). First the buffer was pipetted, then the DNA and on top the octamer or tetramer-dimer. The reactions were dialysed (Slide-a-lyzer mini dialysis tubes) over night from high salt concentrations (N2000 buffer: 25 mM Tris-HCl, pH 7.5, 2 M NaCl, 0.25 mM DTT) to low salt concentration, gradually at a flowrate of 1.5 mL/min overnight (16 h) (N50 buffer: 25 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.25 mM DTT). After 16 h the nucleosomes were placed in a new beaker and incubated with fresh N50 buffer for 3 h. In case for alpha satellite wrapped

nucleosomes, a heat shifting method was applied prior loading a native gel. The nucleosomes were incubated for 2-3 h at 50°C. Samples were taken and run on a 5% Tris-glycine native PAGE at 150 V for 1 h at 4°C.

The gel was stained using SYBR Gold or Gel Red PAGE and visualized in a Typhoon Imager (GE-Cytiva). After verification of correct assembly of the nucleosome, samples were purified using anion-exchange (MonoQ 5/50, Cytiva) and subsequent dialysis in a low salt buffer to obtain pure assembled nucleosomes. 20% glycerol was added and flash frozen with liquid nitrogen and stored at -80°C. Nucleosomes not directly frozen were stored in fridge for only two weeks.

6.2.8 Electromobility Shift assays (EMSA)

For DNA binding reactions (10 µL, low binding tubes) with CENP-B or ZBTB9, 100 nM of DNA (alpha satellite with CENP-B box) was used as fixed amount and variable molar concentrations of CENP-B (5-200 nM) or ZBTB9 (50-700 nM) was titrated in with a buffer containing 30 mM HEPES (pH 8.0), 150 mM NaCl, 3% glycerol, 0.5 mM DTT. The reaction was incubated for 25 min at room temperature. 5 µL of a 4x native Loading dye (containing glycerol, sucrose, phenol blue, tris) were added. The reaction was loaded onto a self-casted 5% Tris-glycine native PAGE. The gel was pre-run prior applying the samples at 100 V for 45 min. The samples were loaded and run for 1 h at 120 V. The gel was stained prior with SYBR Gold or Gel Red PAGE and visualized on a Typhoon Imager (GE-Cytiva).

In case for Nucleosome binding assays with chromatin remodelers, 450 nM Nucleosome (Widom601) were used as fixed amount. Various molar concentrations were titrated in for CHRAC1 or WICH complex (100-500 nM) in a buffer containing 30 mM HEPES (pH 8.0), 150 mM NaCl, 3% glycerol, 0.5 mM DTT. The reaction was incubated for 25 min at room temperature. EMSA assays containing alpha satellite containing CENP-A Nucleosomes (2 µM) were used in fixed amount and 0.5 µM of CENP-B and ZBTB9 were added with a buffer containing 30 mM HEPES (pH 8.0), 150 mM NaCl, 3% glycerol, 0.5 mM DTT. The reaction was incubated for 25 min at room temperature. After adding 5 µL of 4x native sample buffer, the reaction was loaded onto a self-casted 5% Tris-glycine native PAGE gel and run for 1 h at 120 V at 4°C. The gel was stained prior with SYBR Gold or Gel Red PAGE and visualized on a Typhoon Imager (GE-Cytiva).

6.2.9 Molecular weight determination of WICH/ CHRAC1 complex and ZBTB9 by Size exclusion chromatography coupled to right-angle light scattering (SEC-RALS) and Binding kinetics by Surface Plasmon Resonance (SPR) of CENP-B:ZBTB9 binding

Prior to the actual measurements, a BSA control (two independent runs) was performed. Proteins were concentrated to 3-5 mg/mL (100 µL) and were injected into a ÄKTAmicro (Cytiva). Under isocratic conditions (in case of WICH and CHRAC1 complex) separation was performed on a Superose6 increase 3.2/300 GL or (in case for ZBTB9) on a Superdex200 increase 3.2/300 GL. The buffer for WICH/ CHRAC1 complex was 30 mM HEPES (pH 8.2), 420 mM KCl, 2% glycerol (at 18°C). In case for ZBTB9 the buffer composition contained HEPES (pH 8.0), 150 mM KCl, 2% glycerol (at 18°C). The ÄKTAmicro was coupled to a Viscotek RALS with an RI to determine the right-angle light scattering data.

In case for the SPR experiments (Biacore X100, Cytiva), CENP-B (with an N-terminal Twinstrep tag) was immobilized by a self-coupled Biotin chip in a HBS buffer (HEPES (pH 7.4), 150 mM NaCl, 0.05% Tween-20) in a concentration of 50 ng. Different molar concentrations of ZBTB9 (500, 250, 125, 62.5, 21.25, 7.81 nM) were injected into the flow cell 2. Flow cell 1 was set as control. Binding events in control were subtracted from the actual binding events in flow cell 2. Efforts have been made to fit the binding kinetics to different binding models but since the binding of ZBTB9 and CENP-B seem to have multiple modes, the KD could not be satisfactory determined. The data of both experimental setups was conducted by Dr. Gregor Witte (collaboration, AG Hopfner).

6.2.10 *In vivo* Microscopy and siRNA assays

Human cell lines (U2OS, HeLa, HEK293T, RKO) were grown on cover slips in a six well plate to a total confluency of 70% in DMEM (Dulbecco's modified Eagle's medium) media supplemented with 10% FBS (fetal bovine serum) and 2 mM L-glutamine at 37°C and 5% CO_2. The cells were fixed with 2% Formaldehyde in DPBS for 10 min at room temperature. Afterwards the cells were permeabilized and blocked with a buffer containing 2% BSA or BlockAID (Thermo Fisher) with 0.02% Tween-20 and incubated for 1h at room temperature. The fixed cells were washed three times á 5 min with PBS containing 0.02% Tween-20. The first antibody (anti-CENP-A/ anti-CENP-B/ anti-ZBTB9) were added in a dilution of 1:100 or 1:200 (ZBTB9) in a buffer containing 2% BSA with 0.02% Tween-20 in PBS and incubated

with primary antibody for 1h at room temperature. Afterwards cells were washed again with wash buffer three times á 5 min. Subsequently, cells were treated with secondary antibody in a dilution 1:200 (anti-rabbit or anti mouse with coupled fluorophores) for 1h at RT in a buffer containing 2% BSA with 0.02% Tween-20 in PBS. Before imaging the cells were washed with wash buffer three times á 5 min at room temperature. Cells were imaged on a PDV system (General Electric; Bioimaging Facility LMU).

For knockdown assays of ZBTB9 via siRNA duplexes (ZBTB9 human siRNA Oligo Duplex, Locus ID221504, Origen), the cells were seeded out and grown to a confluency of 70%. For an initial screen, the cells were treated with three different duplexes (#A, #B, #C) at a concentration of 1 nM, 10 nM and 100 nM for 24-72h by lipofectamine transfection. The cells were harvested, and samples were analysed on Western Blot to obtain optimum level of decreased protein levels of ZBTB9 in HeLa cells. After establishing the right concentration and incubation time, cell viability assays were performed by measuring the total count of treated cells with non-treated control cells at a given time point. The data presented here was originally obtained by Jeannette Koch (AG Leonhardt, Human Biology & Bioimaging, LMU). This protocol is only a short summary of steps performed. A detailed protocol can be requested.

6.2.11 Western Blot analysis

Samples for Western Blot were loaded onto a self-casted 15% SDS-PAGE gel and run at 180 V for 45 min. The gel was incubated in transfer buffer (Tris-glycine buffer with 1% methanol) before setting up the Western Blot cassette. A PVDF membrane was pre-activated with 100% methanol for 1-2 min. The cassette was assembled, and the transfer reaction was carried out for 90 min at 220 mA in transfer buffer. After successful transfer, the blot was washed 3 times with 1xTBST-Buffer (tris-glycine buffer containing 1% Tween-20) for 5 min. The membrane was blocked for 1 hour at room temperature with 5% skim-milk solution in 1xTBST. After blocking, the primary antibody was incubated overnight in 5% skim-milk solution (anti-ZBTB9, 1:2500 dilution). The next day, the blot was washed 3 times with 1x TBST buffer for 15 min. The secondary antibody (anti-rabbit, 1:15,000) was added to 5% skim-milk solution and incubated for 1h at room temperature. As a final step, the blot was washed 3 time with 1x TBST for 5 min. For detection the Amersham ECL (Cytiva) was used, and the blot was developed on a ImageQuant (Cytiva) system. The exposure time was set manually, and each blot was treated with the same settings.

6.2.12 Nuclear Fishing Experiments and Mass Spectrometric Analysis

Human cell line HeLa were maintained in DMEM (Dulbecco's modified Eagle's medium without pyruvate, Gibco) supplemented with 10% FBS (Fetal Bovine Serum, Gibco) and 2 mM L-glutamine at 37°C and 5% CO_2. Asynchronous cells were grown on square plate dishes (245 mm, 500 cm^2) and grown for two-three days to obtain enough cell material (final $1x10^8$ cells/mL). Cells were trypsinized and centrifuged for 10 min at 1500 rpm (4°C). The cells were washed with DPBS and centrifuged again (1500 rpm, 10 min, 4°C). Cell pellet containing $1x10^8$ cells/mL were flash frozen in liquid nitrogen and stored at -80°C.

The nuclear extract was performed as followed (initial protocol developed by Sansoni *et.al*, 2014 and modified by Dr. Götz Hagemann). All steps were caried out on ice. Cell amount corresponding to $1-4x10^9$ were thawn and resuspended/ lysed in 5 mL in a buffer containing DBPS +0.3% Triton X-100 and complete protease inhibitor for exactly 10 min while rotating. The cells were centrifuged for 10 min at 810 rcf at 4°C. Supernatant was discarded (cytosolic fraction) and pelleted nuclei were washed in 6 mL DPBS with protease inhibitor. The nuclei were centrifuged for 5 min at 1300 rcf. The Nuclei were resuspended in 500 µL of EX-100 buffer (10 mM HEPES (pH 7.6), 100 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA, 10%(v/v) glycerol, 10 mM ß-glycerol phosphate, 1 mM DTT, protease inhibitor) and transferred into 1.5 mL low binding Eppendorf tubes. For MNase digest, $CaCl_2$ to a final concentration of 2 mM were added and 150 U of micrococcal nuclease was added and incubated for 12 h at 4°C at 1400 rpm. The reaction was stop by adding EGTA to a final concentration of 10 mM and 0.5% Tween-20. The digested extract was centrifuged for 30 min at 20,000 rcf and the supernatant (S1) contains the chromatin fraction which was used for Nuclear Fishing Pulldowns.

For Nuclear Fishing Pulldowns (at 4°C), the recombinant domains of ZBTB9 (MBP-fused BTB [19-112] and Zinc finger [380-473]) or full-length version of recombinant WICH/ CHRAC1 complex were immobilized on Ni-NTA beads (Qiagen). 10 µL of bed volume were used, corresponding of a binding capacity of total 500 µg (~50 µg/µL). The beads were pre-equilibrated with binding buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 2% glycerol, 0.5 mM DTT). The protein solution was diluted to a concentration of total 500 µg in 500 µL to cover the complete beads. As a control for each pulldown either only MBP-tag or empty beads were taken. The constructs/ complexes were incubated for 1h while rotating. After incubation, non-bound construct/ complex was removed by applying three washing steps with binding buffer. After drying the beads, the nuclear extract (1 mL) was added and

incubated over night at 4°C. The beads were centrifuged at 1500 rpm for 5 min. Afterwards the beads are washed three times with 20 mM HEPES (pH 7.5), 150 mM KCl, 2% glycerol, 0.5 mM DTT and once with a buffer containing 10 mM HEPES (pH 7.5), 100 mM KCl, 5% glycerol. The beads are dried and resuspended in 10 µL of resuspension buffer (10 mM HEPES (pH 7.5), 100 mM KCl, 5% glycerol). For protein amount estimation, 2 µL of beads were taken out and visualized on a precast TGX SDS-PAGE gel (BioRad), stained by the silverstain method. After roughly estimating the protein amount on the beads, the proteins were subjected to on-bead tryptic digestion. To do so, the beads were incubated for 20 minutes at 25°C in resuspension buffer. Subsequently, the sample was denatured by the addition of two sample volumes of 8M urea (Sigma). The sample was reduced by adding 5mM tris(2-carboxyethyl)phosphine (TCEP, Thermo) for 15 minutes at 35°C shaking and alkylated by addition of 10mM iodoacetamide for 30 minutes at room temperature in the dark. Proteins were digested with lysyl endopeptidase (Wako) at an enzyme: substrate ratio of 1:50 (w/w) at 35°C for 2 hours. The sample was diluted with 50mM ammonium bicarbonate to a urea concentration of 1M, 1/50 (w/w) trypsin (Promega) was added and then incubated at 35°C, shaking overnight. The next morning concentrated trifluoroacetic acid (TFA) was added to the reaction to stop the tryptic digest to a final concentration of 1%, and acetonitrile was added to a final concentration of 3%. The pH was checked with an indicator strip (Merck) to be around pH 2 and proceeded to peptide clean-up and lyophilization in a speed vac at 42°C. The samples were reconstituted in a 1% ACN (acetonitrile), 0.2% FA (formic acid).

LC-MS/MS analysis was carried out on an Ultimate 3000 RSLC coupled with a Q Exactive HF-X (both Thermo Scientific). Aliquots of 1 µg peptides were injected onto a trap column (PepMap 100 C18, 100 µm × 2 cm, 5 µM particles, Thermo Scientific) at a flow rate of 5 µL/min (mobile phase: 0.1 % formic acid and 1 % acetonitrile in water). Peptide separation was performed with an EASY-Spray column (PepMap RSLC C18, 75 µm × 50 cm, 2 µm particles, Thermo Scientific) and a flow rate of 250 nL/min. A two-step gradient was used: first from 3 % B (0.1 % formic acid in acetonitrile) to 25 % B in 160 min and second a 10 min ramp to 40 % B 5 (A: 0.1 % formic acid in water). Peptides were analysed in the data-dependent acquisition mode with up to 15 MS/MS scans per cycle.

MS-Files were searched in MaxQuant (v1.6.5.0). Protein Groups txt files, peptides txt files and evidence txt files were used for statistical analysis in R (performed by Dr. Victor Solis-Mezarino or partially by Chandni Kumar). All contaminants and reverse identified proteins

were excluded from the analysis. By subtracting the background of the samples with the control, proteins found to be highly enriched in each domain or both in similar terms were further subjected to statistical t-test or wilcoxen-tests. By obtaining the t-test p-value (or adjusted p-value) proteins could be separated by different FDR scores. The FDR values were used to perform further bioinformatical analysis (STRING or DAVID) and producing the heatmaps corresponding to DAVID-Analysis GO-Terms.

6.2.13 Synthesis of Enrichment matrix with functionalized benzoboroxol

The synthesis of the dendrimer coupled to the PEGA-resin (amino coupling technique) and subsequent addition of benoboroxol were performed by Stefan Marchner (PhD student at Prof. Anja Hoffmann-Röder, Department Chemistry and Pharmacy, LMU). The initial synthesis as described in Xiao *et.*al (2018) could not be recreated, that is why a new strategy was developed. Briefly, PEGA resin with N-amino coupling were coupled to dendrimer conjugates. The dendrimer conjugates were decorated with benzoboroxol molecules. Upon incubation in organic solvents, the resin can expand itself and increase its surface, facilitating glycol-peptide enrichment. The complete and detailed synthesis can be found in Stefan Marchner's PhD Thesis.

6.2.14 Expression and purification of O-glycohydrolases enzymes (OGA, HEXD, NAGA)

The enzymes were cloned from cDNAs with standard PCR methods and Gibson assembly with an N-terminal 8xHis-tag followed by a 3C precission tag cleavage site into a modified pET-28 vector or a pLIB vector. Inserted genes were verified by sequencing. The vector for expression OGA or HEXD were transformed into Rosetta (DE3) cells via heat shock method. Cells were expressed at 37°C till OD_{600} reached 0.6. The protein expression was induced with 400 µM IPTG at 16°C for 18 h. The cells were harvested after induction and subjected to purification. The cells were resuspended in lysis buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 10 mM Imidazole, 1.5 mM $MgCl_2$, 0.01% Tween, 1 mM DTT, protease inhibitor) and sonicated with 40% duty cycle (output control 5) for three times á 5 min with 2 min cooling sessions. The lysate was centrifuged at 19,500 rpm for 25 min at 4°C. Pre-equilibrated Ni-NTA were incubated with supernatant for 1h at 4°C. The beads are washed with wash buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 35 mM Imidazole, 1 mM DTT) and protein was eluted with elution buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 200 mM Imidazole, 1 mM DTT). Fractions containing enzyme

were subjected to size exclusion chromatography onto Superdex200 increase 10/300 GL (Cytiva) in a buffer containing 30 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol. Fractions containing soluble enzyme were collected, concentrated to 1 mg/mL, aliquoted and flash frozen in liquid nitrogen and stored at -80°C.

In case of NAGA, a P2 virus was used to transfect HighFive cell lines with a C-terminal 8xHis-precission-tagged NAGA construct. The signalling peptide at the N-term was kept in the construct. A ratio of 1:100 of P2 virus was used to transfect the cell line and grown for 72 h at 27°C, 95 rpm. Cells were checked for viability and diameter to ensure protein expression. The cells were harvested and lysed in a buffer containing 50 mM Phosphate buffer (pH 7.0; NaH_2PO_4/Na_2HPO_4), 250 mM NaCl, 5% glycerol, 0.01% Tween, 10 mM Imidazole, 1.5 mM $MgCl_2$, protease inhibitor. The cells were lysed with a glass pestle by douncing for 30 times. The lysate was centrifuged at 19,500 rpm for 25 min at 4°C. Pre-equilibrated Ni-NTA were incubated with supernatant for 1h at 4°C. The beads are washed with wash buffer (50 mM Phosphate buffer (pH 7.0), 500 mM NaCl, 5% glycerol, 35 mM Imidazole, 1 mM DTT) and protein was eluted with elution buffer (50 mM Phosphate buffer (pH 7.0), 500 mM NaCl, 5% glycerol, 200 mM Imidazole, 1 mM DTT). Fractions containing enzyme were subjected to size exclusion chromatography onto Superdex200 increase 10/300 GL (Cytiva) in a buffer containing 30 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol. Fractions containing soluble enzyme were collected, concentrated to 1 mg/mL, aliquoted and flash frozen in liquid nitrogen and stored at -80°C.

6.2.15 HILIC Enrichment of nuclear O-glycosylated peptides with iSPE-HILIC columns

iSPE columns (Hilicon AB, Sweden) were conditioned with one column volume 100% ACN by gravity flow. Then the column was primed first with 0.1% (v/v) TFA in pure water and then with 1% (v/v) TFA in 80% ACN in water (2 column volumes). Peptides (500 µg of protein extract) generated after digest were loaded two times onto the pre-equilibrated iSPE-HILIC column. The column was washed two times with 1% (v/v) TFA in 80% ACN in water. The column was dried by vacuum. Glyco-peptides were eluted twice with 0.8% (v/v) TFA in water. Before elution, the column was incubated with elution buffer for 1-2 min and bypassed by gravity flow. The second elution was performed in the same way. Peptides were lyophilized.

6.2.16 Enrichment of O-GlcNAc, O-GalNAc or other O-glycosylated peptides out of nuclear extract

The nuclear extract was prepared as described in 5.6.11. The proteins were digested in 500 µg aliquots with the same procedure as described. The peptides were purified by the peptide clean up procedure and lyophilized in a speed vac at 42°C.

The synthesized Glyco-Enrichment beads with dendrimer conjugated benzoboroxol (prototype synthesis scale) were stored in Ethanol. 100 µL beads slurry (~ 50 µL bed volume) were transferred into a 1.5 mL low binding Eppendorf tube and washed three times with 100 mM ammonium acetate buffer (pH 11). The beads were centrifuged in between each wash step at 1500 rpm for 10 min at room temperature. To increase bead size, 1 mL of DMSO was added to the beads and rotated for 4h on a wheel at room temperature. Afterwards the beads were washed twice with a buffer containing DMSO and 0.5% TEA. In between the beads are centrifuged at 1500 rpm, for 10 min at room temperature. Lyophilized peptides or HILIC enriched peptides (see 6.5.2) were dissolved in DMSO containing 0.5% TEA for 1h at 30°C. After dissolving, the peptides were incubated with the enrichment matrix for 90 min at room temperature. Shortly after incubation the beads are centrifuged at 1500 rpm for 10 min at room temperature. Non-glycosylated peptides are washed away with 70% DMSO containing 30% 100 mM ammonium acetate (pH 11) for four times, in between the beads are centrifuged as described earlier. Supernatant was discarded after each step. In the last washing step, the beads are carefully dried and incubated three times with elution buffer I (acidic conditions) ACN:H_2O:TFA (50:49:1) at 37°C for 30 min at 1000 rpm. The beads are additionally treated with 5% FA solution at 56°C for 5 min. All eluates are combined and lyophilized. Enzymatic cleavage was not performed.

6.2.17 LC-MS/MS-Analysis and Identification

Benzoboroxol enriched peptides were resuspended in solvent containing 1.0% ACN and 0.2% FA and 10 µL were loaded onto a 15 cm x 0.075 mm Acclaim PepMap C18 column (2 µm particle size, 100 Å pore size) and separated at a flow rate of 300 nL/min using an EASY-nLC 1200 coupled to an LTQ.Orbitrap Elite mass spectrometer (Thermo Fisher). Following gradient was used: 0-5 min 3% B and 5-65 min with 3-65% B (ACN:H_2O:FA, 98:2:0.1). The mass spectrometer was operated in data-dependent mode (Top20), selecting up to 10 precursors from a MS1 scan (resolution 60,000) in the range of m/z 350–1800 for collision-induced dissociation (CID) or higher-energy collision induced dissociation

(HCD) excluding singly and doubly charged precursor ions and precursors of unknown charge states. Dynamic exclusion was activated with a repeat count of 1, exclusion duration of 30 s, list size of 300, and a mass window of ±50 ppm. Fragment ions were detected at low resolution in the linear ion trap.

MS-Files were searched in MaxQuant (v1.6.5.0) against the human proteome (reviewed, Uniprot, downloaded 2019). Protein Groups txt files, peptides txt files and evidence txt files were used for statistical analysis in R (performed by Chandni Kumar with help of Dr. Victor Solis-Mezarino). All contaminants and reverse identified proteins were excluded from the analysis. The HILIC enriched and identified proteins served as benchmark for further analysis. The proteins confidently identified were searched against the O-GlcNAc Database (GlyGen or O-GlcNAc database).

Table 12: Enzyme list used for purification or MS purpose.

Enzyme	Description
DNase I, RNase free	DNase I, Lot: 00633533, Thermo Scientific
Benzonase	EMPROVE® bio, Benzonase®, Lot: K50109095817, Merck
Trypsin	Sequencing Grade modified Trypsin, Lot. 0000048875, Promega
LysC	Lys-C (1:50 wt/wt), Wako pure chemical industries

Table 13: Antibodies used for Western blot or IF analysis.

Antibodies	Description
Anti-ZBTB9 rabbit polyconal	Anti-ZBTB9, polyclonal, rabbit, Sigma-Aldrich, AV39929
Anti-ZBTB9 rabbit polyconal	Anti-ZBTB9, Atlas Antibody, #HPA05871, Lot R82134
Anti-CENP-B mouse polyclonal	Novus Biologicals, Cat #H00001059-B01P
Anti-CENP-A rabbit polyclonal	Novus Biologicals, Cat #NBP2-92632
Anti-Rabbit secondary antibody	Anti-Rabbit-monoclonal, goat, HRP conjugated, abcam

6.2.18 Chemicals and reagents

Table 14: Chemicals and reagents used in this study.

Chemical	Description
Acetic Acid	Acetic acid 100 % molecular biology grade, Lot. 3V001842, Applichem
Acrylamide	Acrylamide-Solution (40 %)-mix 37.5:1 for molecular biology, Lot. 6H011032, Applichem Panreac
Agarose	Agarose Bioenzyme LE Genetic Pure, Lot No. 000488616, Bioenzyme Scientific GmbH

Ampicillin (1:1000)	Ampicillin Natriumsalz ≥ 97 % für die Molekularbiologie und Biochemie, Charge. 355232377, Carl Roth GmbH + Co. KG
APS	Ammonium peroxidsulfate ≥ 98 %, p.a., ACS, Charge: 084201798, Carl Roth GmbH + Co. KG
Chloramphenicol	Chloramphenicol >98,5%, Ph. Eur., für die Biochemie, Carl Roth GmbH + Co. KG
DMSO	Dimethyl sulfoxide molecular biology grade, Lot. 3N009790, Applichem Panreac
DNA gel stain	Roti Safe DNA gel stain, Charge: 404220375, Carl Roth GmbH + Co. KG
DNA ladder marker	DNA Ladder Marker,2-Log DNA ladder 0.1-10.0 kb, #N300L, Lot. 0591204, New England Biolabs
DNA loading dye	DNA loading dye, Gel loading Dye purple (6x), #B7024S, Lot. 0051309, New England Biolabs
dNTPs	dNTPs Set 4 x 25 µmol in 4 x 250 µL, Lot. #SNBG4568N, Sigma-Aldrich
DTT	1,4-Dithiothreit ≥ 99 %, p.a., Charge. 355230614, Carl Roth GmbH + Co. KG
Flag-Peptide	Flag-Peptide, Zhejiang Outores Biochtechnologies Co. Ltd., P1607473, Lot: 0P092916SF-01, MW: 4429.46 Da, Purity: 90.45 %
EDTA	EDTA microbiology grade, Lot. 2A009309, Applichem Panreac
Gentamycin (1:1000)	Gentamycin sulphate, ≥ 590 I.U./mg for biochemistry, Charge. 456232377, Carl Roth GmbH + Co. KG
Glycine	Glycine p.A, Lot. 3V007614, Applichem Panreac

HCl	Hydrochloric acid (HCl), 37 %, p.c., Lot. 50011871, Applichem Pancreac
HEPES	HEPES, >99,5%, Bioscience-Grade, Carl Roth GmbH + Co. KG
Imidazole	Imidazole for buffer solutions, Lot. 4H013658, Applichem Panreac
Iodoacetamide	Iodoacetamide, Lot # 041M53051V, Sigma
Kanamycine (1:1000)	Kanamycinsuflat, 25 g, > 750 I.U./mg, Carl Roth GmbH
LB-media	LB-Medium für die Molekularbiologie, LB Broth (Luria/Miller), Charge. 057254379, Carl Roth GmbH + Co. KG
LDS (4x)	NuPAGE® LDS sample buffer (4x), Lot. 1772815, Novex (by life technologies)
L-Glutamine (insect cultue)	L-Glutamine, CELLPURE®, for cell culture and biochemistry, Lot # 155228536, Carl Roth GmbH + Co. KG
Methanol	Methanol Analytical reagent grade, Lot. 1668488, Fisher Scientific UK
Milk powder	Skim Milk Powder for microbiology, Lot # SZBF3210V, Sigma-Aldrich
Na_3PO_4	Sodium phosphate, 96 %, Lot # SZBF1570V, ALDRICH® Chemistry
NaCl	Sodium chloride puriss. p.a., ACS reagent, reag. ISO, ≥ 99.8 %, Lot # 125225535, Carl Roth GmbH + Co. KG
NaN_3	Sodium azide pure, Natriumazid reinst. Lot. 3W005899, Applichem Panreac
NaOH	Sodium hydroxide solution, 50 %, Charge. 461178869, Carl Roth GmbH + Co. KG

PBS	Dulbecco's Phosphate Buffered saline 10 x, Lot. RNBD7140, Sigma-Aldrich®
DPBS	DPBS (1x), Dulbecco's Phosphate Buffered Saline, Lot. 1813438, Gibco®, Invitrogen (cell culture)
DMEM	DMEM, Gibco®, Invitrogen (cell culture)
Penicillin-Streptomycin (insect culture)	P4333 Penicillin-Streptomycin, Lot # 055M4784V, Sigma
Pluronic (insect culture)	Pluronic® F-127, Lot # RNVD5140, Sigma
Propanol	AnalaR NORMAPUR, 2-Propanol, ACS, Reag. Ph. Eur., Batch. 15K240507, VWR Prolabo® Chemicals
Protein Marker	Page Ruler™ Prestained Protein Ladder, Lot # 00590517, Thermo Scientific
Roche Tablet	Complete Tablets Protease Inhibitor cocktail tablets, EDTA free easy pack, Lot. 187624200, Roche
SDS-solution	SDS-solution 20 % for molecular biology, Lot. 6V015700, Applichem Panreac
SOC-media	SOC Medium, Lot. # 1607043A, Takara clonetech
TEMED	TEMED 99 %, p.a. für die Elektrophorese, Charge: 074208389, Carl Roth GmbH + Co. KG
Tris	TRIS Pufferan® ≥ 99,3 % Buffer Grade, Lot/Charge. 125225535, Carl Roth GmbH und Co. KG
Tween	Tween®20 molecular biology grade, Batch. 5P015148, VWR chemicals
ZBTB9 siRNA Kit	ZBTB9 Human siRNA Oligo Duplex (Locus ID 221504), Cat#: SR316624, Origen

6.2.19 Laboratory Equipment and material

Table 15: Laboratory instruments and materials used in the study.

Instruments and Utensils	Description
Affinity beads	Protino® Ni-NTA agarose, Lot: 1703/001, Machery-Nagel GmbH & Co. KG ANTI-FLAG® M2 affinity gel, Lot: #SLBV3817, Sigma
Centrifugal filter unit	Centrifugal Filter Units, Amicon® Ultra-4, Regenerated Cellulose 10.000 NMWL, Ultra cel® 10 k, Lot R5H14346, Merck Milipore Ltd.
Centrifuge, big	Avanti ® Centrifuge, J-26 XP, Beckman & Coulter™, ID Nr. 404259 **Rotors** JLA 8.1000, 8,000 rpm JA-25.50, 25.000 rpm (07E2778)
Centrifuge, small	Eppendorf centrifuge 5427R (25 °C, 4 °C)
Chromatographic column	Ecno-Column Chromatography columns, Batch # 7372522, BioRad
Mass Spec vials	X100 11 MM PP Vial Criup/Snap 250 µL, Lot. 00251883, Thermo Scientific Snap Cap pink, Lot. 00252488, Thermo Scientific
PCR machine	Flexlid, Eppendorf
PCR tubes	PCR soft tubes, 0.2 mL farblos, Flachdeckel, DNA-, DNAse- und RNAsefrei, Biozyme Biochech Trading GmbH
Powerpac	PowerPac™ Basic Supply, for nucleic acid and protein gel electrophoresis, BioRad Power supply and Electrophoresis chamber, Mini-PROTEAN® Tetra Vertical Electrophoresis Cell, PowerPac™ Basic, #1658025FC, BioRad

Roller	Rollenmischer, RM 5-30 V, Ingenieurbüro CAT, M. Zipper GmbH
Sep-Pak column	SEP-Pak® Vac 1cc (50 mg), +C18 cartridges, Lot. 010337074B, Waters
Thermocycler	Thermo mixer F 1.5, Eppendorf
Tube, 1.5 mL	Microtube 1.5 mL EASYCAP, Lot. 6081511, SARSTEDT
Tube, 14 mL plastic	FALCON® 14 mL Polypropylene Round-Bottom Tube, Lot. 19915037, VWR
Tube, 15 mL plastic	Cellstar® Tubes, 15 mL graduated, conical bottom, blue screw, sterile, Lot. E17043N8, Grainer Bio-one
Tube, 50 mL plastic	Cellstar® Tubes, 50 mL graduated, conical bottom, blue screw, sterile, Lot. E16083GC, Grainer Bio-one
Western blot membrane	Westernblot membrane, Hybond™ 0.45 PVDF, PVDE Blotting membrane, Nucleic acid and protein application, Lot. A21197991, GE healthcare

Table 16: Ready to use Kits used in this study.

Kits	Description
Q5 Mutagenisis Kit	Q5® Site-Directed Mutagenesis Kit, Lot. M7706S, NEB
KLD reaction Kit	KLD Enzyme Mix, Lot. # M0544S, NEB
PCR clean-up Kit	NucleoSpin®, Gel and PCR clean-up, Lot. 1310/001, Macherey-Nagel
Plasmid extraction Kit	NucleoSpin® Plasmid, Lot. 1520/006, Macherey-Nagel

7 References

1. Suski, J. M., Braun, M., Strmiska, V. & Sicinski, P. Targeting cell-cycle machinery in cancer. *Cancer Cell* vol. 39 759–778 Preprint at https://doi.org/10.1016/j.ccell.2021.03.010 (2021).

2. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* **9**, 153–166 (2009).

3. Sherr, C. J. & Roberts, J. M. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev* **18**, 2699–2711 (2004).

4. Martínez-Alonso, D. & Malumbres, M. Mammalian cell cycle cyclins. *Seminars in Cell and Developmental Biology* vol. 107 28–35 Preprint at https://doi.org/10.1016/j.semcdb.2020.03.009 (2020).

5. Quandt, E., Ribeiro, M. P. C. & Clotet, J. Atypical cyclins: the extended family portrait. *Cellular and Molecular Life Sciences* **77**, 231–242 (2020).

6. Tan, E. P., Duncan, F. E. & Slawson, C. The sweet side of the cell cycle. *Biochem Soc Trans* **45**, 313–322 (2017).

7. Li, N. *et al.* Cyclin C is a haploinsufficient tumour suppressor. *Nat Cell Biol* **16**, 1080–1091 (2014).

8. Li, H. *et al.* Molecular Cloning and Chromosomal Localization of the Human Cyclin C (CCNC) and Cyclin E (CCNE) Genes: Deletion of theCCNCGene in Human Tumors. *Genomics* **32**, 253–259 (1996).

9. Hwang, H. C. & Clurman, B. E. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24**, 2776–2786 (2005).

10. Moiseeva, T. N. & Bakkenist, C. J. Regulation of the initiation of DNA replication in human cells. *DNA Repair (Amst)* **72**, 99–106 (2018).

11. Teixeira, L. K. & Reed, S. I. Cyclin E Deregulation and Genomic Instability. in 527–547 (2017). doi:10.1007/978-981-10-6955-0_22.

12. Sava, G. P., Fan, H., Coombes, R. C., Buluwela, L. & Ali, S. CDK7 inhibitors as anticancer drugs. *Cancer and Metastasis Reviews* **39**, 805–823 (2020).

13. Dang, F., Nie, L. & Wei, W. Ubiquitin signaling in cell cycle control and tumorigenesis. *Cell Death and Differentiation* vol. 28 427–438 Preprint at https://doi.org/10.1038/s41418-020-00648-0 (2021).

14. Rape, M., Reddy, S. K. & Kirschner, M. W. The Processivity of Multiubiquitination by the APC Determines the Order of Substrate Degradation. *Cell* **124**, 89–103 (2006).

15. Otto, T. & Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer* **17**, 93–115 (2017).

16. Singer, J. D., Gurian-West, M., Clurman, B. & Roberts, J. M. Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. *Genes Dev* **13**, 2375–2387 (1999).

17. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).

18.   Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: Signal transduction and O-GlcNAc. *Science* Preprint at https://doi.org/10.1126/science.1058714 (2001).

19.   Hart, G. W. Dynamic O-Linked Glycosylation of Nuclear and Cytoskeletal Proteins. *Annu Rev Biochem* **66**, 315–335 (2002).

20.   Biwi, J., Biot, C., Guerardel, Y., Vercoutter-Edouart, A.-S. & Lefebvre, T. The Many Ways by Which O-GlcNAcylation May Orchestrate the Diversity of Complex Glycosylations. *Molecules* **23**, 2858 (2018).

21.   Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu Rev Biochem* **80**, 825–858 (2011).

22.   van Tol, W., Wessels, H. & Lefeber, D. J. O-glycosylation disorders pave the road for understanding the complex human O-glycosylation machinery. *Curr Opin Struct Biol* **56**, 107–118 (2019).

23.   de Queiroz, R. M., Carvalho, Ã. & Dias, W. B. O-GlcNAcylation: The Sweet Side of the Cancer. *Front Oncol* **4**, 1–10 (2014).

24.   Indellicato, R. & Trinchera, M. Epigenetic regulation of glycosylation in cancer and other diseases. *International Journal of Molecular Sciences* vol. 22 1–16 Preprint at https://doi.org/10.3390/ijms22062980 (2021).

25.   Wang, Z. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* **3**, 1–22 (2010).

26.   Tan, E. P., Caro, S., Potnis, A., Lanza, C. & Slawson, C. O-linked N-acetylglucosamine cycling regulates mitotic spindle organization. *Journal of Biological Chemistry* **288**, 27085–27099 (2013).

27.   Fisi, V. *et al.* O-linked N-acetylglucosamine transiently elevates in hela cells during mitosis. *Molecules* **23**, (2018).

28.   Lanza, C. *et al.* Reduced O-GlcNAcase expression promotes mitotic errors and spindle defects. *Cell Cycle* **15**, 1363–1375 (2016).

29.   Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* **32**, 645–655 (2013).

30.   Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561–564 (2013).

31.   Zeidan, Q., Wang, Z., de Maio, A. & Hart, G. W. O-GlcNAc Cycling Enzymes Associate with the Translational Machinery and Modify Core Ribosomal Proteins. *Mol Biol Cell* **21**, 1922–1936 (2010).

32.   Piovesan, A. *et al.* On the length, weight and GC content of the human genome. *BMC Res Notes* **12**, 106 (2019).

33.   Webster, M., Witkin, K. L. & Cohen-Fix, O. Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. *J Cell Sci* **122**, 1477–1486 (2009).

34.   Walther Flemming. Zellsubstanz, Kern und Zelltheilung. *F.C.W. Vogel* (1882).

35. Morrison, O. & Thakur, J. Molecular complexes at euchromatin, heterochromatin and centromeric chromatin. *Int J Mol Sci* **22**, (2021).

36. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science (1979)* **184**, 868–871 (1974).

37. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**, 809–814 (2003).

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

39. Martire, S. & Banaszynski, L. A. The roles of histone variants in fine-tuning chromatin organization and function. *Nature Reviews Molecular Cell Biology* vol. 21 522–541 Preprint at https://doi.org/10.1038/s41580-020-0262-8 (2020).

40. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *EMBO Rep* **16**, 1467–1481 (2015).

41. Zhu, D., Zhang, Y. & Wang, S. Histone citrullination: a new target for tumors. *Molecular Cancer* vol. 20 Preprint at https://doi.org/10.1186/s12943-021-01373-z (2021).

42. Xu, H., Wu, M., Ma, X., Huang, W. & Xu, Y. Function and Mechanism of Novel Histone Posttranslational Modifications in Health and Disease. *BioMed Research International* vol. 2021 Preprint at https://doi.org/10.1155/2021/6635225 (2021).

43. Jorgensen, S., Schotta, G. & Sorensen, C. S. Histone H4 Lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res* **41**, 2797–2806 (2013).

44. Castellano-Pozo, M. *et al.* R Loops Are Linked to Histone H3 S10 Phosphorylation and Chromatin Condensation. *Mol Cell* **52**, 583–590 (2013).

45. Palmer, D., O'Day, K., Wener, M., Andrews, B. & Margolis, R. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J Cell Biol* **104**, 805–815 (1987).

46. Earnshaw, W., Bordwell, B., Marino, C. & Rothfield, N. Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. *Journal of Clinical Investigation* **77**, 426–430 (1986).

47. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin Compaction by a Polycomb Group Protein Complex. *Science (1979)* **306**, 1574–1577 (2004).

48. Sanulli, S. *et al.* HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, 390–394 (2019).

49. Filion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224 (2010).

50. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).

51. Sims, R. J., Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends in Genetics* **19**, 629–639 (2003).

52. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Mol Cell* **11**, 709–719 (2003).

53. Murton, B. L., Chin, W. L., Ponting, C. P. & Itzhaki, L. S. Characterising the Binding Specificities of the Subunits Associated with the KMT2/Set1 Histone Lysine Methyltransferase. *J Mol Biol* **398**, 481–488 (2010).

54. Venkatesh, S. & Workman, J. L. Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol* **2**, 685–700 (2013).

55. Strahl, B. D. *et al.* Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Mol Cell Biol* **22**, 1298–1306 (2002).

56. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931–21936 (2010).

57. Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452–455 (2012).

58. Gao, Y. *et al.* Acetylation of histone H3K27 signals the transcriptional elongation for estrogen receptor alpha. *Commun Biol* **3**, 165 (2020).

59. Bian, C. *et al.* Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J* **30**, 2829–2842 (2011).

60. Adam, M., Robert, F., Larochelle, M. & Gaudreau, L. H2A.Z Is Required for Global Chromatin Integrity and for Recruitment of RNA Polymerase II under Specific Conditions. *Mol Cell Biol* **21**, 6270–6279 (2001).

61. Mizuguchi, G. *et al.* ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science (1979)* **303**, 343–348 (2004).

62. Hong, J. *et al.* The Catalytic Subunit of the SWR1 Remodeler Is a Histone Chaperone for the H2A.Z-H2B Dimer. *Mol Cell* **53**, 498–505 (2014).

63. Willhoft, O. *et al.* Structure and dynamics of the yeast SWR1-nucleosome complex. *Science (1979)* **362**, (2018).

64. Shia, W.-J., Li, B. & Workman, J. L. SAS-mediated acetylation of histone H4 Lys 16 is required for H2A.Z incorporation at subtelomeric regions in *Saccharomyces cerevisiae*. *Genes Dev* **20**, 2507–2512 (2006).

65. Draker, R. *et al.* A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation. *PLoS Genet* **8**, e1003047 (2012).

66. Winkler, C. *et al.* Hodgkin's lymphoma RNA-transfected dendritic cells induce cancer/testis antigen-specific immune responses. *Cancer Immunology, Immunotherapy* **61**, 1769–1779 (2012).

67. Chadwick, B. P. & Willard, H. F. A Novel Chromatin Protein, Distantly Related to Histone H2a, Is Largely Excluded from the Inactive X Chromosome. *Journal of Cell Biology* **152**, 375–384 (2001).

68. Tolstorukov, M. Y. *et al.* Histone Variant H2A.Bbd Is Associated with Active Transcription and mRNA Processing in Human Cells. *Mol Cell* **47**, 596–607 (2012).

69. Sansoni, V. *et al.* The histone variant H2A.Bbd is enriched at sites of DNA synthesis. *Nucleic Acids Res* **42**, 6405–6420 (2014).

70.   Ahmad, K. & Henikoff, S. The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Mol Cell* **9**, 1191–1200 (2002).

71.   Bano, D., Piazzesi, A., Salomoni, P. & Nicotera, P. The histone variant H3.3 claims its place in the crowded scene of epigenetics. *Aging* **9**, 602–614 (2017).

72.   Ray-Gallet, D. *et al.* HIRA Is Critical for a Nucleosome Assembly Pathway Independent of DNA Synthesis. *Mol Cell* **9**, 1091–1100 (2002).

73.   Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences* **107**, 14075–14080 (2010).

74.   Voon, H. P. J. *et al.* ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Rep* **11**, 405–418 (2015).

75.   Campos, E. I. *et al.* The program for processing newly synthesized histones H3.1 and H4. *Nat Struct Mol Biol* **17**, 1343–1351 (2010).

76.   Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell* **116**, 51–61 (2004).

77.   Goldberg, A. D. *et al.* Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140**, 678–691 (2010).

78.   Drané, P., Ouararhni, K., Depaux, A., Shuaib, M. & Hamiche, A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* **24**, 1253–1265 (2010).

79.   Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421–434 (2011).

80.   Mosammaparast, N. A role for nucleosome assembly protein 1 in the nuclear transport of histones H2A and H2B. *EMBO J* **21**, 6527–6538 (2002).

81.   Valieva, M. E. *et al.* Large-scale ATP-independent nucleosome unfolding by a histone chaperone. *Nat Struct Mol Biol* **23**, 1111–1116 (2016).

82.   Kemble, D. J., McCullough, L. L., Whitby, F. G., Formosa, T. & Hill, C. P. FACT Disrupts Nucleosome Structure by Binding H2A-H2B with Conserved Peptide Motifs. *Mol Cell* **60**, 294–306 (2015).

83.   Hondele, M. *et al.* Structural basis of histone H2A–H2B recognition by the essential chaperone FACT. *Nature* **499**, 111–114 (2013).

84.   Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: How is it established, and why does it matter? *Dev Biol* **339**, 258–266 (2010).

85.   Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J. & Groth, A. Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol* **18**, 141–158 (2017).

86.   Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* **16**, 178–189 (2015).

87.   Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**, 267–273 (2013).

88. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**, 1499–1504 (1980).

89. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

90. Jjingo, D., Conley, A. B., Yi, S. v., Lunyak, V. v. & Jordan, I. K. On the presence and role of human gene-body DNA methylation. *Oncotarget* **3**, 462–474 (2012).

91. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).

92. Smale, S. T. & Kadonaga, J. T. The RNA Polymerase II Core Promoter. *Annu Rev Biochem* **72**, 449–479 (2003).

93. Hampsey, M. Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiology and Molecular Biology Reviews* **62**, 465–503 (1998).

94. Tserel, L. *et al.* Genome-wide promoter analysis of histone modifications in human monocyte-derived antigen presenting cells. *BMC Genomics* **11**, 642 (2010).

95. Jin, C. *et al.* H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* **41**, 941–945 (2009).

96. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).

97. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465–476 (2008).

98. Hughes, A. L., Kelley, J. R. & Klose, R. J. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194567 (2020).

99. Birnbaum, R. Y. *et al.* Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res* **22**, 1059–1068 (2012).

100. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**, 717–728 (1983).

101. Core, L. J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**, 1311–1320 (2014).

102. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & de Laat, W. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol Cell* **10**, 1453–1465 (2002).

103. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* **76**, 473-484.e7 (2019).

104. Heitz, E. Das Heterochromatin der Moose. *Bornträger, Stuttgart* (1928).

105. Berger, F. Emil Heitz, a true epigenetics pioneer. *Nat Rev Mol Cell Biol* **20**, 572–572 (2019).

106. Kosak, S. T. *et al.* Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science (1979)* **296**, 158–162 (2002).

107. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (1979)* **326**, 289–293 (2009).

108. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

109. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).

110. Costanzi, C., Stein, P., Worrad, D. M., Schultz, R. M. & Pehrson, J. R. Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. *Development* **127**, 2283–2289 (2000).

111. Gamble, M. J., Frizzell, K. M., Yang, C., Krishnakumar, R. & Kraus, W. L. The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes Dev* **24**, 21–32 (2010).

112. Plys, A. J. *et al.* Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev* **33**, 799–813 (2019).

113. Britten, R. J. & Kohne, D. E. Repeated Sequences in DNA. *Science (1979)* **161**, 529–540 (1968).

114. McClintock, B. CHROMOSOME ORGANIZATION AND GENIC EXPRESSION. *Cold Spring Harb Symp Quant Biol* **16**, 13–47 (1951).

115. Peters, A. H. F. M. *et al.* Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* **107**, 323–337 (2001).

116. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).

117. Bernard, P. *et al.* Requirement of Heterochromatin for Cohesion at Centromeres. *Science (1979)* **294**, 2539–2542 (2001).

118. Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. v. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences* **97**, 11319–11324 (2000).

119. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**, 2733–2742 (2002).

120. Volpe, T. A. *et al.* Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science (1979)* **297**, 1833–1837 (2002).

121. Volpe, T. *et al.* RNA interference is required for normal centromere function in fission yeast. *Chromosome Research* **11**, 137–146 (2003).

122. Martienssen, R. A., Zaratiegui, M. & Goto, D. B. RNA interference and heterochromatin in the fission yeast Schizosaccharomyces pombe. *Trends in Genetics* **21**, 450–456 (2005).

123. Canzio, D. *et al.* Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Mol Cell* **41**, 67–81 (2011).

124. Canzio, D. *et al.* A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* **496**, 377–381 (2013).

125. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

126. Azzaz, A. M. *et al.* Human Heterochromatin Protein 1α Promotes Nucleosome Associations That Drive Chromatin Condensation. *Journal of Biological Chemistry* **289**, 6850–6861 (2014).

127. Muchardt, C. *et al.* Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1α. *EMBO Rep* **3**, 975–981 (2002).

128. Huo, X. *et al.* The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. *Mol Cell* **77**, 368-383.e7 (2020).

129. Thakur, J. & Henikoff, S. Architectural RNA in chromatin organization. *Biochem Soc Trans* **48**, 1967–1978 (2020).

130. Machida, S. *et al.* Structural Basis of Heterochromatin Formation by Human HP1. *Mol Cell* **69**, 385-397.e8 (2018).

131. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

132. Feil, R. & Berger, F. Convergent evolution of genomic imprinting in plants and mammals. *Trends in Genetics* **23**, 192–199 (2007).

133. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. *Science (1979)* **347**, 1017–1021 (2015).

134. Kung, J. T. *et al.* Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. *Mol Cell* **57**, 361–375 (2015).

135. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev* **28**, 723–734 (2014).

136. Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell* **76**, 395-411.e13 (2019).

137. Saldaña-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* **76**, 412-422.e5 (2019).

138. Henikoff, S., Ahmad, K. & Malik, H. S. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science (1979)* **293**, 1098–1102 (2001).

139. Furuyama, T. & Henikoff, S. Centromeric Nucleosomes Induce Positive DNA Supercoils. *Cell* **138**, 104–113 (2009).

140. Nurk, S. *et al.* The complete sequence of a human genome. *Science (1979)* **376**, 44–53 (2022).

141. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science (1979)* **376**, (2022).

142. Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V. & Yurov, Y. Alpha-satellite DNA of primates: old and new families. *Chromosoma* **110**, 253–266 (2001).

143. Thakur, J. & Henikoff, S. Unexpected conformational variations of the human centromeric chromatin complex. *Genes Dev* **32**, 20–25 (2018).

144. Henikoff, J. G., Thakur, J., Kasinathan, S. & Henikoff, S. A unique chromatin complex occupies young α-satellite arrays of human centromeres. *Sci Adv* **1**, (2015).

145. Arunkumar, G. & Melters, D. P. Centromeric transcription: A conserved swiss-army knife. *Genes* vol. 11 1–23 Preprint at https://doi.org/10.3390/genes11080911 (2020).

146. Senaratne, A. P. *et al.* Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. *Current Biology* **31**, 173-181.e7 (2021).

147. Gassmann, R. *et al.* An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. *Nature* **484**, 534–537 (2012).

148. Bresson, S. & Tollervey, D. Surveillance-ready transcription: nuclear RNA decay as a default fate. *Open Biol* **8**, 170270 (2018).

149. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763–776 (2009).

150. Nishimura, K. *et al.* DEAH box RNA helicase DHX38 associates with satellite I noncoding RNA involved in chromosome segregation. *Genes to Cells* **24**, 585–590 (2019).

151. Leclerc, S. & Kitagawa, K. The Role of Human Centromeric RNA in Chromosome Stability. *Frontiers in Molecular Biosciences* vol. 8 Preprint at https://doi.org/10.3389/fmolb.2021.642732 (2021).

152. Smurova, K. & de Wulf, P. Centromere and Pericentromere Transcription: Roles and Regulation … in Sickness and in Health. *Frontiers in Genetics* vol. 9 Preprint at https://doi.org/10.3389/fgene.2018.00674 (2018).

153. Zhang, X. *et al.* Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. *Int J Mol Sci* **20**, 5573 (2019).

154. Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F. & Crespi, M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res* **46**, 2169–2184 (2018).

155. Yao, R.-W., Wang, Y. & Chen, L.-L. Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21**, 542–551 (2019).

156. Montefalcone, G., Tempesta, S., Rocchi, M. & Archidiacono, N. Centromere Repositioning. *Genome Res* **9**, 1184–1188 (1999).

157. Melters, D. P. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**, R10 (2013).

158. McNulty, S. M., Sullivan, L. L. & Sullivan, B. A. 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 (2017).

159. Chen, C. C. *et al.* Establishment of Centromeric Chromatin by the CENP-A Assembly Factor CAL1 Requires FACT-Mediated Transcription. *Dev Cell* **34**, 73–84 (2015).

160. Bobkov, G. O. M., Gilbert, N. & Heun, P. Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. *Journal of Cell Biology* **217**, 1957–1972 (2018).

161. Marczynski, G. T. & Jaehning, J. A. A transcription map of a yeast centromere plasmid: unexpected transcripts and altered gene expression. *Nucleic Acids Res* **13**, 8487–8506 (1985).

162. Ohkuni, K. & Kitagawa, K. Endogenous Transcription at the Centromere Facilitates Centromere Activity in Budding Yeast. *Current Biology* **21**, 1695–1703 (2011).

163.   Ling, Y. H. & Yuen, K. W. Y. Point centromere activity requires an optimal level of centromeric noncoding RNA. *Proceedings of the National Academy of Sciences* **116**, 6270–6279 (2019).

164.   Allshire, R. C. & Ekwall, K. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harb Perspect Biol* **7**, a018770 (2015).

165.   Kixmoeller, K., Allu, P. K. & Black, B. E. The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. *Open Biology* vol. 10 Preprint at https://doi.org/10.1098/rsob.200051 (2020).

166.   Musacchio, A. & Desai, A. A molecular view of kinetochore assembly and function. *Biology (Basel)* **6**, (2017).

167.   Mahlke, M. A. & Nechemia-arbely, Y. Guarding the genome: Cenp-a-chromatin in health and cancer. *Genes* vol. 11 1–24 Preprint at https://doi.org/10.3390/genes11070810 (2020).

168.   Nakaseko, Y., Kinoshita, N. & Yanagida, M. A novel sequence common to the centromere regions of Schizosaccharomyces pombe chromosomes. *Nucleic Acids Res* **15**, 4705–4715 (1987).

169.   Clarke, L. & Carbon, J. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287**, 504–509 (1980).

170.   Bloom, K. & Costanzo, V. Centromere Structure and Function. in 515–539 (2017). doi:10.1007/978-3-319-58592-5_21.

171.   McFarlane, R. J., Mian, S. & Dalgaard, J. Z. The many facets of the Tim-Tipin protein families' roles in chromosome biology. *Cell Cycle* **9**, 700–705 (2010).

172.   Garavís, M. *et al.* The structure of an endogenous Drosophila centromere reveals the prevalence of tandemly repeated sequences able to form i-motifs. *Sci Rep* **5**, 13307 (2015).

173.   Schalch, T. & Steiner, F. A. Structure of centromere chromatin: from nucleosome to chromosomal architecture. *Chromosoma* **126**, 443–455 (2017).

174.   Heckmann, S. *et al.* Holocentric Chromosomes of Luzula elegans Are Characterized by a Longitudinal Centromere Groove, Chromosome Bending, and a Terminal Nucleolus Organizer Region. *Cytogenet Genome Res* **134**, 220–228 (2011).

175.   Steiner, F. A. & Henikoff, S. Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *Elife* **3**, (2014).

176.   Talbert, P. B. & Henikoff, S. What makes a centromere? *Exp Cell Res* **389**, (2020).

177.   Joseph, A., Mitchell, A. R. & Miller, O. J. The organization of the mouse satellite DNA at centromeres. *Exp Cell Res* **183**, 494–500 (1989).

178.   Garavís, M., Escaja, N., Gabelica, V., Villasante, A. & González, C. Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by at Hoogsteen Base Pairs. *Chemistry - A European Journal* **21**, 9816–9824 (2015).

179.   Muro, Y. *et al. Centromere Protein B Assembles Human Centromeric a-satellite DNA at the 17-bp Sequence, CENPB Box. The Journal of Cell Biology* vol. 116 (1992).

180.   Fukagawa, T. & Earnshaw, W. C. The centromere: Chromatin foundation for the kinetochore machinery. *Developmental Cell* vol. 30 496–508 Preprint at https://doi.org/10.1016/j.devcel.2014.08.016 (2014).

181.   Hara, M. & Fukagawa, T. Kinetochore assembly and disassembly during mitotic entry and exit. *Current Opinion in Cell Biology* vol. 52 73–81 Preprint at https://doi.org/10.1016/j.ceb.2018.02.005 (2018).

182.   Li, E. & Zhang, Y. DNA Methylation in Mammals. *Cold Spring Harb Perspect Biol* **6**, a019133–a019133 (2014).

183.   Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

184.   Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13**, 335–340 (1997).

185.   Scelfo, A. & Fachinetti, D. Keeping the centromere under control: A promising role for DNA methylation. *Cells* vol. 8 Preprint at https://doi.org/10.3390/cells8080912 (2019).

186.   Robertson, K. D. & Wolffe, A. P. DNA methylation in health and disease. *Nat Rev Genet* **1**, 11–19 (2000).

187.   Bestor, T. H. The DNA methyltransferases of mammals. *Hum Mol Genet* **9**, 2395–2402 (2000).

188.   Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**, 219–220 (1998).

189.   Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257 (1999).

190.   Arand, J. *et al.* In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet* **8**, e1002750 (2012).

191.   Easwaran, H. P., Schermelleh, L., Leonhardt, H. & Cardoso, M. C. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep* **5**, 1181–1186 (2004).

192.   Chen, T., Tsujimoto, N. & Li, E. The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. *Mol Cell Biol* **24**, 9048–9058 (2004).

193.   Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484–492 (2012).

194.   Walton, E. L., Francastel, C. & Velasco, G. Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics* **6**, 1373–1377 (2011).

195.   Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**, 5594–5605 (2003).

196.   Wong, N. C. *et al.* Permissive Transcriptional Activity at the Centromere through Pockets of DNA Hypomethylation. *PLoS Genet* **2**, e17 (2006).

197.   Cedar, H. & Bergman, Y. Programming of DNA Methylation Patterns. *Annu Rev Biochem* **81**, 97–117 (2012).

198.   Fachinetti, D. *et al.* DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Dev Cell* **33**, 314–327 (2015).

199. Gamba, R. & Fachinetti, D. From evolution to function: Two sides of the same CENP-B coin? *Experimental Cell Research* vol. 390 Preprint at https://doi.org/10.1016/j.yexcr.2020.111959 (2020).

200. Tanaka, Y., Kurumizaka, H. & Yokoyama, S. CpG methylation of the CENP-B box reduces human CENP-B binding. *FEBS Journal* **272**, 282–289 (2005).

201. Okada, T. *et al.* CENP-B Controls Centromere Formation Depending on the Chromatin Context. *Cell* **131**, 1287–1300 (2007).

202. Gopalakrishnan, S., Sullivan, B. A., Trazzi, S., della Valle, G. & Robertson, K. D. 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 (2009).

203. Kim, I. S. *et al.* Roles of Mis18α in Epigenetic Regulation of Centromeric Chromatin and CENP-A Loading. *Mol Cell* **46**, 260–273 (2012).

204. Fujita, Y. *et al.* Priming of Centromere for CENP-A Recruitment by Human hMis18α, hMis18β, and M18BP1. *Dev Cell* **12**, 17–30 (2007).

205. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology* **13**, 1192–1200 (2003).

206. Saksouk, N. *et al.* Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Mol Cell* **56**, 580–594 (2014).

207. Guetg, C. *et al.* The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J* **29**, 2135–2146 (2010).

208. Nan, X. *et al.* Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proceedings of the National Academy of Sciences* **104**, 2709–2714 (2007).

209. Nan, X., Tate, P., Li, E. & Bird, A. DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol* **16**, 414–421 (1996).

210. Guttenbach, M. & Schmid, M. Exclusion of Specific Human Chromosomes into Micronuclei by 5-Azacytidine Treatment of Lymphocyte Cultures. *Exp Cell Res* **211**, 127–132 (1994).

211. Suzuki, T., Fujii, M. & Ayusawa, D. Demethylation of classical satellite 2 and 3 DNA with chromosomal instability in senescent human fibroblasts. *Exp Gerontol* **37**, 1005–1014 (2002).

212. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic Nuclear Envelope Collapse in Cancer Cell Micronuclei. *Cell* **154**, 47–60 (2013).

213. Ly, P. *et al.* Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat Cell Biol* **19**, 68–75 (2017).

214. Kokalj-Vokac, N. *et al.* Specific induction of uncoiling and recombination by azacytidine in classical satellite-containing constitutive heterochromatin. *Cytogenet Genome Res* **63**, 11–15 (1993).

215. Hirano, T., Kobayashi, R. & Hirano, M. Condensins, Chromosome Condensation Protein Complexes Containing XCAP-C, XCAP-E and a Xenopus Homolog of the Drosophila Barren Protein. *Cell* **89**, 511–521 (1997).

216. D., F., J., B.-S. & B., D. Complex relationships between 5-aza-dC induced DNA demethylation and chromosome compaction at mitosis. *Chromosoma* **111**, 37–44 (2002).

217. Collings, C. K., Waddell, P. J. & Anderson, J. N. Effects of DNA methylation on nucleosome stability. *Nucleic Acids Res* **41**, 2918–2931 (2013).

218. Choy, J. S. *et al.* DNA Methylation Increases Nucleosome Compaction and Rigidity. *J Am Chem Soc* **132**, 1782–1783 (2010).

219. Keshet, I., Lieman-Hurwitz, J. & Cedar, H. DNA methylation affects the formation of active chromatin. *Cell* **44**, 535–543 (1986).

220. Schalch, T. & Steiner, F. A. Structure of centromere chromatin: from nucleosome to chromosomal architecture. *Chromosoma* **126**, 443–455 (2017).

221. Bergmann, J. H. *et al.* Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J* **30**, 328–340 (2011).

222. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* **11**, 1076–1083 (2004).

223. Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *EMBO Journal* vol. 28 2511–2531 Preprint at https://doi.org/10.1038/emboj.2009.173 (2009).

224. Hinshaw, S. M. & Harrison, S. C. Kinetochore Function from the Bottom Up. *Trends in Cell Biology* vol. 28 22–33 Preprint at https://doi.org/10.1016/j.tcb.2017.09.002 (2018).

225. Tanaka, T. U., Stark, M. J. R. & Tanaka, K. Kinetochore capture and bi-orientation on the mitotic spindle. *Nature Reviews Molecular Cell Biology* vol. 6 929–942 Preprint at https://doi.org/10.1038/nrm1764 (2005).

226. Musacchio, A. The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Current Biology* **25**, R1002–R1018 (2015).

227. Maresca, T. J. & Salmon, E. D. Welcome to a new kind of tension: translating kinetochore mechanics into a wait-anaphase signal. *J Cell Sci* **123**, 825–835 (2010).

228. Uchida, K. S. K. *et al.* Kinetochore stretching inactivates the spindle assembly checkpoint. *Journal of Cell Biology* **184**, 383–390 (2009).

229. Hara, M. & Fukagawa, T. Dynamics of kinetochore structure and its regulations during mitotic progression. *Cellular and Molecular Life Sciences* vol. 77 2981–2995 Preprint at https://doi.org/10.1007/s00018-020-03472-4 (2020).

230. Joglekar, A. A Cell Biological Perspective on Past, Present and Future Investigations of the Spindle Assembly Checkpoint. *Biology (Basel)* **5**, 44 (2016).

231. Mitra, S. *et al.* Genetic screening identifies a SUMO protease dynamically maintaining centromeric chromatin. *Nat Commun* **11**, 501 (2020).

232. Bobkov, G. O. M. *et al.* Spt6 is a maintenance factor for centromeric CENP-A. *Nat Commun* **11**, (2020).

233. Black, B. E. & Cleveland, D. W. Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. *Cell* **144**, 471–479 (2011).

234. Black, B. E., Brock, M. A., Bédard, S., Woods, V. L. & Cleveland, D. W. An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. *Proceedings of the National Academy of Sciences* **104**, 5008–5013 (2007).

235. Falk, S. J. *et al.* CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. *Nat Struct Mol Biol* **23**, 204–208 (2016).

236. Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* **476**, 232–235 (2011).

237. Dalal, Y., Wang, H., Lindsay, S. & Henikoff, S. Tetrameric Structure of Centromeric Nucleosomes in Interphase Drosophila Cells. *PLoS Biol* **5**, e218 (2007).

238. Dimitriadis, E. K., Weber, C., Gill, R. K., Diekmann, S. & Dalal, Y. Tetrameric organization of vertebrate centromeric nucleosomes. *Proceedings of the National Academy of Sciences* **107**, 20317–20322 (2010).

239. Furuyama, T. & Henikoff, S. Centromeric Nucleosomes Induce Positive DNA Supercoils. *Cell* **138**, 104–113 (2009).

240. Arimura, Y. *et al.* Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci Rep* **4**, 7115 (2015).

241. Lacoste, N. *et al.* Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. *Mol Cell* **53**, 631–644 (2014).

242. Bui, M. *et al.* Cell-Cycle-Dependent Structural Transitions in the Human CENP-A Nucleosome In Vivo. *Cell* **150**, 317–326 (2012).

243. Nechemia-Arbely, Y. *et al.* Human centromeric CENP-A chromatin is a homotypic, octameric nucleosome at all cell cycle points. *Journal of Cell Biology* **216**, 607–621 (2017).

244. Padeganeh, A. *et al.* Octameric CENP-A Nucleosomes Are Present at Human Centromeres throughout the Cell Cycle. *Current Biology* **23**, 764–769 (2013).

245. Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of (CENP-A–H4)2 reveals physical features that mark centromeres. *Nature* **467**, 347–351 (2010).

246. Boopathi, R. *et al.* Phase-plate cryo-EM structure of the Widom 601 CENP-A nucleosome core particle reveals differential flexibility of the DNA ends. *Nucleic Acids Res* **48**, 5735–5748 (2020).

247. Roulland, Y. *et al.* The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol Cell* **63**, 674–685 (2016).

248. Ali-Ahmad, A. & Sekulić, N. CENP-A nucleosome—a chromatin-embedded pedestal for the centromere: lessons learned from structural biology. *Essays Biochem* **64**, 205–221 (2020).

249. Gonzalez, M., He, H., Sun, S., Li, C. & Li, F. Cell cycle-dependent deposition of CENP-A requires the Dos1/2–Cdc20 complex. *Proceedings of the National Academy of Sciences* **110**, 606–611 (2013).

250. Jansen, L. E. T., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *Journal of Cell Biology* **176**, 795–805 (2007).

251. Choi, E. S. *et al.* Factors That Promote H3 Chromatin Integrity during Transcription Prevent Promiscuous Deposition of CENP-ACnp1 in Fission Yeast. *PLoS Genet* **8**, e1002985 (2012).

252. Choi, E. S. *et al.* Identification of Noncoding Transcripts from within CENP-A Chromatin at Fission Yeast Centromeres. *Journal of Biological Chemistry* **286**, 23600–23607 (2011).

253. Grenfell, A. W., Heald, R. & Strzelecka, M. Mitotic noncoding RNA processing promotes kinetochore and spindle assembly in Xenopus. *Journal of Cell Biology* **214**, 133–141 (2016).

254. Wong, L. H. *et al.* Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* **17**, 1146–1160 (2007).

255. Chan, F. L. *et al.* Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proceedings of the National Academy of Sciences* **109**, 1979–1984 (2012).

256. Bobkov, G. O. M., Gilbert, N. & Heun, P. Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. *Journal of Cell Biology* **217**, 1957–1972 (2018).

257. Qué, D. & Dalal, Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife* (2014) doi:10.7554/eLife.03254.001.

258. Bobkov, G. O. M., Gilbert, N. & Heun, P. Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. 1–16 (2018).

259. Shukla, M. *et al.* Centromere DNA Destabilizes H3 Nucleosomes to Promote CENP-A Deposition during the Cell Cycle. *Current Biology* **28**, 3924-3936.e4 (2018).

260. Dunleavy, E. M., Almouzni, G. & Karpen, G. H. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. *Nucleus* **2**, 146–157 (2011).

261. Foltz, D. R. *et al.* Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137**, 472–484 (2009).

262. Dunleavy, E. M. *et al.* HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* **137**, 485–497 (2009).

263. Tachiwana, H. *et al.* HJURP involvement in de novo CenH3CENP-A and CENP-C recruitment. *Cell Rep* **11**, 22–32 (2015).

264. Chen, C.-C. *et al.* CAL1 is the Drosophila CENP-A assembly factor. *Journal of Cell Biology* **204**, 313–329 (2014).

265. Williams, J. S., Hayashi, T., Yanagida, M. & Russell, P. Fission Yeast Scm3 Mediates Stable Assembly of Cnp1/CENP-A into Centromeric Chromatin. *Mol Cell* **33**, 287–298 (2009).

266. Stoler, S. *et al.* Scm3, an essential Saccharomyces cerevisiae centromere protein required for G2/M progression and Cse4 localization. *Proceedings of the National Academy of Sciences* **104**, 10571–10576 (2007).

267. Rošić, S., Köhler, F. & Erhardt, S. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *Journal of Cell Biology* **207**, 335–349 (2014).

268. Nechemia-Arbely, Y. *et al.* DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. *Nat Cell Biol* **21**, 743–754 (2019).

269. ten Hagen, K. G., Gilbert, D. M., Willard, H. F. & Cohen, S. N. Replication timing of DNA sequences associated with human centromeres and telomeres. *Mol Cell Biol* **10**, 6348–6355 (1990).

270. Jansen, L. E. T., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *Journal of Cell Biology* **176**, 795–805 (2007).

271. Nechemia-Arbely, Y., Fachinetti, D. & Cleveland, D. W. Replicating centromeric chromatin: Spatial and temporal control of CENP-A assembly. *Exp Cell Res* **318**, 1353–1360 (2012).

272. Hayashi, T. *et al.* Mis16 and Mis18 Are Required for CENP-A Loading and Histone Deacetylation at Centromeres. *Cell* **118**, 715–729 (2004).

273. Stellfox, M. E., Nardi, I. K., Knippler, C. M. & Foltz, D. R. Differential Binding Partners of the Mis18α/β YIPPEE Domains Regulate Mis18 Complex Recruitment to Centromeres. *Cell Rep* **15**, 2127–2135 (2016).

274. Moree, B., Meyer, C. B., Fuller, C. J. & Straight, A. F. CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *Journal of Cell Biology* **194**, 855–871 (2011).

275. Silva, M. C. C. *et al.* Cdk Activity Couples Epigenetic Centromere Inheritance to Cell Cycle Progression. *Dev Cell* **22**, 52–63 (2012).

276. McKinley, K. L. & Cheeseman, I. M. Polo-like Kinase 1 Licenses CENP-A Deposition at Centromeres. *Cell* **158**, 397–411 (2014).

277. Spiller, F. *et al.* Molecular basis for Cdk1-regulated timing of Mis18 complex assembly and CENP-A deposition. *EMBO Rep* **18**, 894–905 (2017).

278. Jansen, L. E. T., Black, B. E., Foltz, D. R. & Cleveland, D. W. Propagation of centromeric chromatin requires exit from mitosis. *Journal of Cell Biology* **176**, 795–805 (2007).

279. Stellfox, M. E., Nardi, I. K., Knippler, C. M. & Foltz, D. R. Differential Binding Partners of the Mis18α/β YIPPEE Domains Regulate Mis18 Complex Recruitment to Centromeres. *Cell Rep* **15**, 2127–2135 (2016).

280. Perpelescu, M., Nozaki, N., Obuse, C., Yang, H. & Yoda, K. Active establishment of centromeric cenp-a chromatin by rsf complex. *Journal of Cell Biology* **185**, 397–407 (2009).

281. Huang, H. *et al.* A unique binding mode enables MCM2 to chaperone histones H3–H4 at replication forks. *Nat Struct Mol Biol* **22**, 618–626 (2015).

282. Dunleavy, E. M., Almouzni, G. & Karpen, G. H. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. *Nucleus* **2**, 146–157 (2011).

283. Shukla, M. *et al.* Centromere DNA Destabilizes H3 Nucleosomes to Promote CENP-A Deposition during the Cell Cycle. *Current Biology* **28**, 3924-3936.e4 (2018).

284. Johnson, W. L. *et al.* RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *Elife* **6**, (2017).

285. Burton, A. *et al.* Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol* **22**, 767–778 (2020).

286. Olszak, A. M. *et al.* Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol* **13**, 799–808 (2011).

287. Chan, D. Y. L., Moralli, D., Khoja, S. & Monaco, Z. L. Noncoding Centromeric RNA Expression Impairs Chromosome Stability in Human and Murine Stem Cells. *Dis Markers* **2017**, 1–8 (2017).

288. Zhu, Q. *et al.* Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Mol Cell* **70**, 842-853.e7 (2018).

289. Zhu, Q. *et al.* BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**, 179–184 (2011).

290. Ishikura, S., Nakabayashi, K., Nagai, M., Tsunoda, T. & Shirasawa, S. ZFAT binds to centromeres to control noncoding RNA transcription through the KAT2B–H4K8ac–BRD4 axis. *Nucleic Acids Res* **48**, 10848–10866 (2020).

291. Chen, E. S. *et al.* Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* **451**, 734–737 (2008).

292. Prendergast, L. *et al.* The CENP-T/-W complex is a binding partner of the histone chaperone FACT. *Genes Dev* **30**, 1313–1326 (2016).

293. Herrera-Moyano, E., Mergui, X., García-Rubio, M. L., Barroso, S. & Aguilera, A. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription–replication conflicts. *Genes Dev* **28**, 735–748 (2014).

294. Richter, W. F., Nayak, S., Iwasa, J. & Taatjes, D. J. The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nature Reviews Molecular Cell Biology* Preprint at https://doi.org/10.1038/s41580-022-00498-3 (2022).

295. Carlsten, J. O. *et al.* Mediator Promotes CENP-A Incorporation at Fission Yeast Centromeres. *Mol Cell Biol* **32**, 4035–4043 (2012).

296. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *Journal of Cell Biology* **109**, 1963–1973 (1989).

297. Henikoff, J. G., Thakur, J., Kasinathan, S. & Henikoff, S. A unique chromatin complex occupies young α-satellite arrays of human centromeres. *Sci Adv* **1**, (2015).

298. Earnshaw, W. C., Ratrie, H. & Stetten, G. Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma* **98**, 1–12 (1989).

299. Hemmerich, P. *et al.* Dynamics of inner kinetochore assembly and maintenance in living cells. *Journal of Cell Biology* **180**, 1101–1114 (2008).

300. Tachiwana, H. *et al.* Nap1 regulates proper CENP-B binding to nucleosomes. *Nucleic Acids Res* **41**, 2869–2880 (2013).

301. Mohibi, S. *et al.* Alteration/Deficiency in Activation 3 (ADA3) Protein, a Cell Cycle Regulator, Associates with the Centromere through CENP-B and Regulates Chromosome Segregation. *Journal of Biological Chemistry* **290**, 28299–28310 (2015).

302. Tan, T., Chen, Z., Lei, Y., Zhu, Y. & Liang, Q. A Regulatory Effect of INMAP on Centromere Proteins: Antisense INMAP Induces CENP-B Variation and Centromeric Halo. *PLoS One* **9**, e91937 (2014).

303. Sugimoto, K. & Himeno, M. Casein Kinase II Site of Human Centromere Protein B (CENP-B) is Phosphorylated in Vitro. *Biosci Biotechnol Biochem* **56**, 1174–1175 (1992).

304. Dai, X. *et al.* Identification of novel α-N-methylation of CENP-B that regulates its binding to the centromeric DNA. *J Proteome Res* **12**, 4167–4175 (2013).

305. Keiten-Schmitz, J., Schunck, K. & Müller, S. SUMO Chains Rule on Chromatin Occupancy. *Front Cell Dev Biol* **7**, (2020).

306. Maalouf JE *et al.* CENP-B dynamics at centromeres is regulated by a SUMOylation/ubiquitination and proteasomal-dependent degradation mechanism involving the SUMO-targeted ubiquitin E3 ligase RNF4. *bioRxiv* (2018).

307. Suzuki, N. *et al.* CENP-B Interacts with CENP-C Domains Containing Mif2 Regions Responsible for Centromere Localization. *Journal of Biological Chemistry* **279**, 5934–5946 (2004).

308. Erliandri, I. *et al.* Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucleic Acids Res* **42**, 11502–11516 (2014).

309. Otake, K. *et al.* CENP-B creates alternative epigenetic chromatin states permissive for CENP-A or heterochromatin assembly. *J Cell Sci* **133**, (2020).

310. Miyazaki, H. *et al.* Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. *PLoS Genet* **9**, e1003897 (2013).

311. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* **128**, 735–745 (2007).

312. Shono, N. *et al.* CENP-C and CENP-I are key connecting factors for kinetochore and CENP-A assembly. *J Cell Sci* (2015) doi:10.1242/jcs.180786.

313. Müller, S. & Almouzni, G. Chromatin dynamics during the cell cycle at centromeres. *Nat Rev Genet* **18**, 192–208 (2017).

314. Chardon, F. *et al.* CENP-B-mediated DNA loops regulate activity and stability of human centromeres. *Mol Cell* **82**, 1751-1767.e8 (2022).

315. Kato, H. *et al.* A Conserved Mechanism for Centromeric Nucleosome Recognition by Centromere Protein CENP-C. *Science (1979)* **340**, 1110–1113 (2013).

316. Guo, L. Y. *et al.* Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. *Nat Commun* **8**, 15775 (2017).

317. Falk, S. J. *et al.* CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. *Nat Struct Mol Biol* **23**, 204–208 (2016).

318. Falk, S. J. *et al.* CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science (1979)* **348**, 699–703 (2015).

319. Walstein, K. *et al.* *Assembly principles and stoichiometry of a complete human kinetochore module*. *Sci. Adv* vol. 7 https://www.science.org (2021).

320. Ariyoshi, M. *et al.* Cryo-EM structure of the CENP-A nucleosome in complex with phosphorylated CENP-C. *EMBO J* **40**, (2021).

321. Zhou, K. *et al.* CENP-N promotes the compaction of centromeric chromatin. *Nat Struct Mol Biol* **29**, 403–413 (2022).

322. Yatskevich, S. *et al.* *Structure of the human inner kinetochore bound to a centromeric CENP-A nucleosome*. https://www.science.org (2022) doi:DOI: 10.1126/science.abn3810.

323.   Pesenti, M. E. *et al.* Structure of the human inner kinetochore CCAN complex and its significance for human centromere organization. *Mol Cell* (2022) doi:10.1016/j.molcel.2022.04.027.

324.   Roulland, Y. *et al.* The Flexible Ends of CENP-A Nucleosome Are Required for Mitotic Fidelity. *Mol Cell* **63**, 674–685 (2016).

325.   Yan, K. *et al.* Structure of the inner kinetochore CCAN complex assembled onto a centromeric nucleosome. *Nature* **574**, 278–282 (2019).

326.   Tachiwana, H. *et al.* Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* vol. 476 232–235 Preprint at https://doi.org/10.1038/nature10258 (2011).

327.   Musacchio, A. & Desai, A. A molecular view of kinetochore assembly and function. *Biology* vol. 6 Preprint at https://doi.org/10.3390/biology6010005 (2017).

328.   Weir, J. R. *et al.* Insights from biochemical reconstitution into the architecture of human kinetochores. *Nature* **537**, 249–253 (2016).

329.   Basilico, F. *et al.* The pseudo GTPase CENP-M drives human kinetochore assembly. *Elife* **3**, (2014).

330.   Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

331.   Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).

332.   Pentakota, S. *et al.* Decoding the centromeric nucleosome through CENP-N. *Elife* **6**, 1–25 (2017).

333.   Klare, K. *et al.* CENP-C is a blueprint for constitutive centromere–associated network assembly within human kinetochores. *Journal of Cell Biology* **210**, 11–22 (2015).

334.   Guse, A., Carroll, C. W., Moree, B., Fuller, C. J. & Straight, A. F. In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature* **477**, 354–358 (2011).

335.   Nishino, T. *et al.* CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. *Cell* **148**, 487–501 (2012).

336.   Prendergast, L. *et al.* Premitotic Assembly of Human CENPs -T and -W Switches Centromeric Chromatin to a Mitotic State. *PLoS Biol* **9**, e1001082 (2011).

337.   Logsdon, G. A. *et al.* Both tails and the centromere targeting domain of CENP-A are required for centromere establishment. *Journal of Cell Biology* **208**, 521–531 (2015).

338.   Amano, M. *et al.* The CENP-S complex is essential for the stable assembly of outer kinetochore structure. *Journal of Cell Biology* **186**, 173–182 (2009).

339.   Hori, T. *et al.* CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. *Cell* **135**, 1039–1052 (2008).

340.   Takeuchi, K. *et al.* The centromeric nucleosome-like CENP–T–W–S–X complex induces positive supercoils into DNA. *Nucleic Acids Res* **42**, 1644–1655 (2014).

341.   Huis in, P. J. *et al.* Molecular basis of outer kinetochore assembly on CENP-T. doi:10.7554/eLife.21007.001.

342.   Takenoshita, Y., Hara, M. & Fukagawa, T. Recruitment of two Ndc80 complexes via the CENP-T pathway is sufficient for kinetochore functions. *Nat Commun* **13**, (2022).

343.   Petrovic, A. *et al.* Structure of the MIS12 Complex and Molecular Basis of Its Interaction with CENP-C at Human Kinetochores. *Cell* **167**, 1028-1040.e15 (2016).

344.   Hori, T., Okada, M., Maenaka, K. & Fukagawa, T. CENP-O Class Proteins Form a Stable Complex and Are Required for Proper Kinetochore Function. *Mol Biol Cell* **19**, 843–854 (2008).

345.   McKinley, K. L. *et al.* 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 (2015).

346.   Pesenti, M. E. *et al.* Reconstitution of a 26-Subunit Human Kinetochore Reveals Cooperative Microtubule Binding by CENP-OPQUR and NDC80. *Mol Cell* **71**, 923-939.e10 (2018).

347.   Hinshaw, S. M. & Harrison, S. C. The structure of the Ctf19c/CCAN from budding yeast. *Elife* **8**, (2019).

348.   Alushin, G. M. *et al.* The Ndc80 kinetochore complex forms oligomeric arrays along microtubules. *Nature* **467**, 805–810 (2010).

349.   Ciferri, C. *et al.* Implications for Kinetochore-Microtubule Attachment from the Structure of an Engineered Ndc80 Complex. *Cell* **133**, 427–439 (2008).

350.   Ciferri, C. *et al.* Architecture of the Human Ndc80-Hec1 Complex, a Critical Constituent of the Outer Kinetochore. *Journal of Biological Chemistry* **280**, 29088–29095 (2005).

351.   Valverde, R., Ingram, J. & Harrison, S. C. Conserved Tetramer Junction in the Kinetochore Ndc80 Complex. *Cell Rep* **17**, 1915–1922 (2016).

352.   Pekgöz Altunkaya, G. *et al.* CCAN Assembly Configures Composite Binding Interfaces to Promote Cross-Linking of Ndc80 Complexes at the Kinetochore. *Current Biology* **26**, 2370–2378 (2016).

353.   Nishino, T. *et al.* CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J* **32**, 424–436 (2013).

354.   Petrovic, A. *et al.* The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *Journal of Cell Biology* **190**, 835–852 (2010).

355.   Kline, S. L., Cheeseman, I. M., Hori, T., Fukagawa, T. & Desai, A. The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. *Journal of Cell Biology* **173**, 9–17 (2006).

356.   Nishino, T. *et al.* CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J* **32**, 424–436 (2013).

357.   Gascoigne, K. E. *et al.* Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell* **145**, 410–422 (2011).

358.   Screpanti, E. *et al.* Direct Binding of Cenp-C to the Mis12 Complex Joins the Inner and Outer Kinetochore. *Current Biology* **21**, 391–398 (2011).

359.   Petrovic, A. *et al.* Modular Assembly of RWD Domains on the Mis12 Complex Underlies Outer Kinetochore Organization. *Mol Cell* **53**, 591–605 (2014).

360. Kasuboski, J. M. *et al.* Zwint-1 is a novel Aurora B substrate required for the assembly of a dynein-binding platform on kinetochores. *Mol Biol Cell* **22**, 3318–3330 (2011).

361. Woo Seo, D. *et al.* Zwint-1 is required for spindle assembly checkpoint function and kinetochore-microtubule attachment during oocyte meiosis. *Sci Rep* **5**, 15431 (2015).

362. Varma, D. *et al.* Spindle assembly checkpoint proteins are positioned close to core microtubule attachment sites at kinetochores. *Journal of Cell Biology* **202**, 735–746 (2013).

363. Kiyomitsu, T., Murakami, H. & Yanagida, M. Protein Interaction Domain Mapping of Human Kinetochore Protein Blinkin Reveals a Consensus Motif for Binding of Spindle Assembly Checkpoint Proteins Bub1 and BubR1. *Mol Cell Biol* **31**, 998–1011 (2011).

364. Godek, K. M., Kabeche, L. & Compton, D. A. Regulation of kinetochore–microtubule attachments through homeostatic control during mitosis. *Nat Rev Mol Cell Biol* **16**, 57–64 (2015).

365. Khodjakov, A. & Pines, J. Centromere tension: a divisive issue. *Nat Cell Biol* **12**, 919–923 (2010).

366. Barbosa, J., Sunkel, C. E. & Conde, C. The Role of Mitotic Kinases and the RZZ Complex in Kinetochore-Microtubule Attachments: Doing the Right Link. *Frontiers in Cell and Developmental Biology* vol. 10 Preprint at https://doi.org/10.3389/fcell.2022.787294 (2022).

367. Jeyaprakash, A. A. *et al.* Structural and Functional Organization of the Ska Complex, a Key Component of the Kinetochore-Microtubule Interface. *Mol Cell* **46**, 274–286 (2012).

368. Schmidt, J. C. *et al.* The Kinetochore-Bound Ska1 Complex Tracks Depolymerizing Microtubules and Binds to Curved Protofilaments. *Dev Cell* **23**, 968–980 (2012).

369. Gaitanos, T. N. *et al.* Stable kinetochore–microtubule interactions depend on the Ska complex and its new component Ska3/C13Orf3. *EMBO J* **28**, 1442–1452 (2009).

370. Monda, J. K. *et al.* Microtubule Tip Tracking by the Spindle and Kinetochore Protein Ska1 Requires Diverse Tubulin-Interacting Surfaces. *Current Biology* **27**, 3666-3675.e6 (2017).

371. Abad, M. A. *et al.* Structural basis for microtubule recognition by the human kinetochore Ska complex. *Nat Commun* **5**, 2964 (2014).

372. Abad, M. A. *et al.* Ska3 Ensures Timely Mitotic Progression by Interacting Directly With Microtubules and Ska1 Microtubule Binding Domain. *Sci Rep* **6**, 34042 (2016).

373. Magidson, V. *et al.* Adaptive changes in the kinetochore architecture facilitate proper spindle assembly. *Nat Cell Biol* **17**, 1134–1144 (2015).

374. Rieder, C. L. The Formation, Structure, and Composition of the Mammalian Kinetochore and Kinetochore Fiber. in 1–58 (1982). doi:10.1016/S0074-7696(08)61672-1.

375. Liang, C. *et al.* Centromere-localized Aurora B kinase is required for the fidelity of chromosome segregation. *Journal of Cell Biology* **219**, (2020).

376. Hadders, M. A. *et al.* Untangling the contribution of Haspin and Bub1 to Aurora B function during mitosis. *Journal of Cell Biology* **219**, (2020).

377. Broad, A. J. & DeLuca, J. G. The right place at the right time: Aurora B kinase localization to centromeres and kinetochores. *Essays Biochem* **64**, 299–311 (2020).

378. Broad, A. J., DeLuca, K. F. & DeLuca, J. G. Aurora B kinase is recruited to multiple discrete kinetochore and centromere regions in human cells. *Journal of Cell Biology* **219**, (2020).

379. Lampson, M. & Grishchuk, E. Mechanisms to Avoid and Correct Erroneous Kinetochore-Microtubule Attachments. *Biology (Basel)* **6**, 1 (2017).

380. Ye, A. A. *et al.* Aurora A Kinase Contributes to a Pole-Based Error Correction Pathway. *Current Biology* **25**, 1842–1851 (2015).

381. Chmátal, L., Yang, K., Schultz, R. M. & Lampson, M. A. Spatial Regulation of Kinetochore Microtubule Attachments by Destabilization at Spindle Poles in Meiosis I. *Current Biology* **25**, 1835–1841 (2015).

382. Nijenhuis, W., Vallardi, G., Teixeira, A., Kops, G. J. P. L. & Saurin, A. T. Negative feedback at kinetochores underlies a responsive spindle checkpoint signal. *Nat Cell Biol* **16**, 1257–1264 (2014).

383. Sarangapani, K. K., Koch, L. B., Nelson, C. R., Asbury, C. L. & Biggins, S. Kinetochore-bound Mps1 regulates kinetochore–microtubule attachments via Ndc80 phosphorylation. *Journal of Cell Biology* **220**, (2021).

384. Moura, M. & Conde, C. Phosphatases in Mitosis: Roles and Regulation. *Biomolecules* **9**, 55 (2019).

385. Saurin, A. T. Kinase and Phosphatase Cross-Talk at the Kinetochore. *Front Cell Dev Biol* **6**, (2018).

386. Liu, D. *et al.* Regulated targeting of protein phosphatase 1 to the outer kinetochore by KNL1 opposes Aurora B kinase. *Journal of Cell Biology* **188**, 809–820 (2010).

387. Kim, T. Recent Progress on the Localization of PLK1 to the Kinetochore and Its Role in Mitosis. *International Journal of Molecular Sciences* vol. 23 Preprint at https://doi.org/10.3390/ijms23095252 (2022).

388. Papavassiliou, K. A. & Papavassiliou, A. G. Transcription Factor Drug Targets. *J Cell Biochem* 2693–2696 (2016) doi:10.1002/jcb.25605.

389. Bonchuk, A., Denisov, S., Georgiev, P. & Maksimenko, O. Drosophila BTB/POZ Domains of "ttk Group" Can Form Multimers and Selectively Interact with Each Other. *J Mol Biol* **412**, 423–436 (2011).

390. Meng, Y. *et al.* Z-DNA is remodelled by ZBTB43 in prospermatogonia to safeguard the germline genome and epigenome. *Nat Cell Biol* **24**, 1141–1153 (2022).

391. Chaharbakhshi, E. & Jemc, J. C. Broad-complex, tramtrack, and bric-à-brac (BTB) proteins: Critical regulators of development. *Genesis* vol. 54 505–518 Preprint at https://doi.org/10.1002/dvg.22964 (2016).

392. Cheng, Z. Y., He, T. T., Gao, X. M., Zhao, Y. & Wang, J. ZBTB Transcription Factors: Key Regulators of the Development, Differentiation and Effector Function of T Cells. *Frontiers in Immunology* vol. 12 Preprint at https://doi.org/10.3389/fimmu.2021.713294 (2021).

393. Zhu, C., Chen, G., Zhao, Y., Gao, X. M. & Wang, J. Regulation of the development and function of B cells by ZBTB transcription factors. *Frontiers in Immunology* vol. 9 Preprint at https://doi.org/10.3389/fimmu.2018.00580 (2018).

394. Chevrier, S. & Corcoran, L. M. BTB-ZF transcription factors, a growing family of regulators of early and late B-cell development. *Immunology and Cell Biology* vol. 92 481–488 Preprint at https://doi.org/10.1038/icb.2014.20 (2014).

395. Brayer, K. J. & Segal, D. J. Keep your fingers off my DNA: Protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochemistry and Biophysics* vol. 50 111–131 Preprint at https://doi.org/10.1007/s12013-008-9008-5 (2008).

396. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA Recognition by Cys $_2$ His $_2$ Zinc Finger Proteins. *Annu Rev Biophys Biomol Struct* **29**, 183–212 (2000).

397. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**, 1609–1614 (1985).

398. Matthews, J. M. & Sunde, M. Zinc Fingers--Folds for Many Occasions. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **54**, 351–355 (2002).

399. Rakhra, G. & Rakhra, G. Zinc finger proteins: insights into the transcriptional and post transcriptional regulation of immune response. *Mol Biol Rep* **48**, 5735–5743 (2021).

400. Martin, P. J., Delmotte, M.-H., Formstecher, P. & Lefebvre, P. PLZF is a negative regulator of retinoic acid receptor transcriptional activity. *Nucl Recept* **1**, 6 (2003).

401. Hudson, A. M. & Cooley, L. Drosophila Kelch functions with Cullin-3 to organize the ring canal actin cytoskeleton. *Journal of Cell Biology* **188**, 29–37 (2010).

402. Pintard, L. *et al.* The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature* **425**, 311–316 (2003).

403. Kelly, K. F. & Daniel, J. M. POZ for effect – POZ-ZF transcription factors in cancer and development. *Trends Cell Biol* **16**, 578–587 (2006).

404. Maeda, T. Regulation of hematopoietic development by ZBTB transcription factors. *Int J Hematol* **104**, 310–323 (2016).

405. Stogios, P. J., Cuesta-Seijo, J. A., Chen, L., Pomroy, N. C. & Privé, G. G. Insights into Strand Exchange in BTB Domain Dimers from the Crystal Structures of FAZF and Miz1. *J Mol Biol* **400**, 983–997 (2010).

406. Ren, R. *et al.* Structural basis of specific DNA binding by the transcription factor ZBTB24. *Nucleic Acids Res* **47**, 8388–8398 (2019).

407. Aravind, L. & Koonin, E. v. Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain 1 1Edited by F. Cohen. *J Mol Biol* **285**, 1353–1361 (1999).

408. Siggs, O. M. & Beutler, B. The BTB-ZF transcription factors. *Cell Cycle* vol. 11 3358–3369 Preprint at https://doi.org/10.4161/cc.21277 (2012).

409. Marianayagam, N. J., Sunde, M. & Matthews, J. M. The power of two: protein dimerization in biology. *Trends Biochem Sci* **29**, 618–625 (2004).

410. Barna, M. *et al.* Plzf Mediates Transcriptional Repression of HoxD Gene Expression through Chromatin Remodeling. *Dev Cell* **3**, 499–510 (2002).

411. Katsani, K. R. Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *EMBO J* **18**, 698–708 (1999).

412. Zullo, J. M. *et al.* DNA Sequence-Dependent Compartmentalization and Silencing of Chromatin at the Nuclear Lamina. *Cell* **149**, 1474–1487 (2012).

413. Singh, A. K. *et al.* Role of ZBTB7A zinc finger in tumorigenesis and metastasis. *Molecular Biology Reports* vol. 48 4703–4719 Preprint at https://doi.org/10.1007/s11033-021-06405-x (2021).

414. Rippe, K. *et al.* DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences* **104**, 15635–15640 (2007).

415. Tyagi, M., Imam, N., Verma, K. & Patel, A. K. Chromatin remodelers: We are the drivers!! *Nucleus* vol. 7 388–404 Preprint at https://doi.org/10.1080/19491034.2016.1211217 (2016).

416. Markert, J. & Luger, K. Nucleosomes Meet Their Remodeler Match. *Trends in Biochemical Sciences* vol. 46 41–50 Preprint at https://doi.org/10.1016/j.tibs.2020.08.010 (2021).

417. Xiao, H. *et al.* Dual Functions of Largest NURF Subunit NURF301 in Nucleosome Sliding and Transcription Factor Interactions. *Mol Cell* **8**, 531–543 (2001).

418. Khorasanizadeh, S. The Nucleosome. *Cell* **116**, 259–272 (2004).

419. Petty, E. & Pillus, L. Balancing chromatin remodeling and histone modifications in transcription. *Trends in Genetics* **29**, 621–629 (2013).

420. Li, Y. *et al.* The emerging role of ISWI chromatin remodeling complexes in cancer. *Journal of Experimental and Clinical Cancer Research* vol. 40 Preprint at https://doi.org/10.1186/s13046-021-02151-x (2021).

421. Goodwin, L. R. & Picketts, D. J. The role of ISWI chromatin remodeling complexes in brain development and neurodevelopmental disorders. *Molecular and Cellular Neuroscience* **87**, 55–64 (2018).

422. Neve, B., Jonckheere, N., Vincent, A. & van Seuningen, I. Long non-coding RNAs: the tentacles of chromatin remodeler complexes. *Cellular and Molecular Life Sciences* vol. 78 1139–1161 Preprint at https://doi.org/10.1007/s00018-020-03646-0 (2021).

423. Smits, A. H. & Vermeulen, M. Characterizing Protein???Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends in Biotechnology* vol. 34 825–834 Preprint at https://doi.org/10.1016/j.tibtech.2016.02.014 (2016).

424. O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* **25**, 1000–1008 (2018).

425. Cravatt, B. F., Simon, G. M. & Yates, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* Preprint at https://doi.org/10.1038/nature06525 (2007).

426. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

427. Lössl, P., Waterbeemd, M. & Heck, A. J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J* **35**, 2634–2657 (2016).

428. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).

429. Ong, S.-E. & Mann, M. Mass spectrometry–based proteomics turns quantitative. *Nat Chem Biol* **1**, 252–262 (2005).

430. Thompson, A. *et al.* Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal Chem* **75**, 1895–1904 (2003).

431. Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr Opin Struct Biol* **35**, 100–108 (2015).

432. Sinz, A. Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein–Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angewandte Chemie International Edition* **57**, 6390–6396 (2018).

433. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Current Opinion in Structural Biology* vol. 23 252–260 Preprint at https://doi.org/10.1016/j.sbi.2013.02.008 (2013).

434. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* **41**, 20–32 (2016).

435. Rappsilber, J. The beginning of a beautiful friendship : Cross-linking / mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **173**, 530–540 (2011).

436. Tran, B. Q., Goodlett, D. R. & Goo, Y. A. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1864**, 123–129 (2016).

437. Leitner, A. *et al.* Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Molecular & Cellular Proteomics* **9**, 1634–1649 (2010).

438. Armache, J. P. *et al.* Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. *Elife* **8**, (2019).

439. Hekkelman, M. L., de Vries, I., Joosten, R. P. & Perrakis, A. AlphaFill: enriching the AlphaFold models with ligands and co-factors. doi:10.1101/2021.11.26.470110.

440. McCarty, A. S., Kleiger, G., Eisenberg, D. & Smale, S. T. Selective Dimerization of a C2H2 Zinc Finger Subfamily. *Mol Cell* **11**, 459–470 (2003).

441. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

442. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).

443. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605–D612 (2021).

444. Snel, B. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**, 3442–3444 (2000).

445. Hillmer, R. E. & Link, B. A. The Roles of Hippo Signaling Transducers Yap and Taz in Chromatin Remodeling. *Cells* **8**, 502 (2019).

446. Hornbeck, P. v. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512–D520 (2015).

447. Almuzzaini, B., Sarshad, A. A., Farrants, A.-K. Ö. & Percipalle, P. Nuclear myosin 1 contributes to a chromatin landscape compatible with RNA polymerase II transcription activation. *BMC Biol* **13**, 35 (2015).

448. Vintermist, A. *et al.* The Chromatin Remodelling Complex B-WICH Changes the Chromatin Structure and Recruits Histone Acetyl-Transferases to Active rRNA Genes. *PLoS One* **6**, e19184 (2011).

449. Sadeghifar, F., Böhm, S., Vintermist, A. & Östlund Farrants, A.-K. The B-WICH chromatin-remodelling complex regulates RNA polymerase III transcription by promoting Max-dependent c-Myc binding. *Nucleic Acids Res* **43**, 4477–4490 (2015).

450. Ishikura, S. *et al.* CENP-B promotes the centromeric localization of ZFAT to control transcription of noncoding RNA. *Journal of Biological Chemistry* **279**, (2021).

451. Aktar, S., Sasaki, H. & Unoki, M. Identification of ZBTB24 protein domains and motifs for heterochromatin localization and transcriptional activation. *Genes to Cells* **24**, 746–755 (2019).

452. de Dieuleveult, M. & Miotto, B. DNA Methylation and Chromatin: Role(s) of Methyl-CpG-Binding Protein ZBTB38. *Epigenetics Insights* vol. 11 Preprint at https://doi.org/10.1177/2516865718811117 (2018).

453. Marchal, C., Defossez, P.-A. & Miotto, B. Context-dependent CpG methylation directs cell-specific binding of transcription factor ZBTB38. *Epigenetics* 1–22 (2022) doi:10.1080/15592294.2022.2111135.

454. Thompson, J. J. *et al.* ZBTB24 is a transcriptional regulator that coordinates with DNMT3B to control DNA methylation. *Nucleic Acids Res* **46**, 10034–10051 (2018).

455. Pozner, A. *et al.* The C-Terminal Zinc Fingers of ZBTB38 are Novel Selective Readers of DNA Methylation. *J Mol Biol* **430**, 258–271 (2018).

456. Hervouet, E., Nadaradjane, A., Gueguen, M., Vallette, F. M. & Cartron, P.-F. Kinetics of DNA methylation inheritance by the Dnmt1-including complexes during the cell cycle. *Cell Div* **7**, 5 (2012).

457. Ren, Y. Regulatory mechanism and biological function of UHRF1–DNMT1-mediated DNA methylation. *Funct Integr Genomics* **22**, 1113–1126 (2022).

458. Poleshko, A. *et al.* H3K9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. doi:10.7554/eLife.49278.001.

459. Gruenbaum, Y. & Foisner, R. Lamins: Nuclear Intermediate Filament Proteins with Fundamental Functions in Nuclear Mechanics and Genome Regulation. *Annu Rev Biochem* **84**, 131–164 (2015).

460. Buchwalter, A., Kaneshiro, J. M. & Hetzer, M. W. Coaching from the sidelines: the nuclear periphery in genome regulation. *Nature Reviews Genetics* vol. 20 39–50 Preprint at https://doi.org/10.1038/s41576-018-0063-5 (2019).

461. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* **41**, 246–250 (2009).

462. Polioudaki, H. *et al.* Histones H3/H4 form a tight complex with the inner nuclear membrane protein LBR and heterochromatin protein 1. *EMBO Rep* **2**, 920–925 (2001).

463. Poleshko, A. *et al.* Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction. *Cell* **171**, 573-587.e14 (2017).

464. Towbin, B. D. *et al.* Step-Wise Methylation of Histone H3K9 Positions Heterochromatin at the Nuclear Periphery. *Cell* **150**, 934–947 (2012).

465. Franks, T. M. *et al.* Nup98 recruits the Wdr82–Set1A/COMPASS complex to promoters to regulate H3K4 trimethylation in hematopoietic progenitor cells. *Genes Dev* **31**, 2222–2234 (2017).

466. Capitanio, J. S., Montpetit, B. & Wozniak, R. W. Human Nup98 regulates the localization and activity of DExH/D-box helicase DHX9. *Elife* **6**, (2017).

467. Somech, R. *et al.* The nuclear-envelope protein and transcriptional repressor LAP2β interacts with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *J Cell Sci* **118**, 4017–4025 (2005).

468. Janzer, A. *et al.* The H3K4me3 Histone Demethylase Fbxl10 Is a Regulator of Chemokine Expression, Cellular Morphology, and the Metabolome of Fibroblasts. *Journal of Biological Chemistry* **287**, 30984–30992 (2012).

469. He, J., Nguyen, A. T. & Zhang, Y. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **117**, 3869–3880 (2011).

470. Iuchi, S. & Paulo, J. A. Lysine-specific demethylase 2A enhances binding of various nuclear factors to CpG-rich genomic DNAs by action of its CXXC-PHD domain. *Sci Rep* **9**, 5496 (2019).

471. Zhou, J. C., Blackledge, N. P., Farcas, A. M. & Klose, R. J. Recognition of CpG Island Chromatin by KDM2A Requires Direct and Specific Interaction with Linker DNA. *Mol Cell Biol* **32**, 479–489 (2012).

472. Melnick, A. *et al.* Critical Residues within the BTB Domain of PLZF and Bcl-6 Modulate Interaction with Corepressors. *Mol Cell Biol* **22**, 1804–1818 (2002).

473. Yang, Y. *et al.* Pokemon (FBI-1) interacts with Smad4 to repress TGF-β-induced transcriptional responses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**, 270–281 (2015).

474. Piazza, F., Costoya, J. A., Merghoub, T., Hobbs, R. M. & Pandolfi, P. P. Disruption of PLZP in Mice Leads to Increased T-Lymphocyte Proliferation, Cytokine Production, and Altered Hematopoietic Stem Cell Homeostasis. *Mol Cell Biol* **24**, 10456–10469 (2004).

475. Omori, M. *et al.* CD8 T Cell-Specific Downregulation of Histone Hyperacetylation and Gene Activation of the IL-4 Gene Locus by ROG, Repressor of GATA. *Immunity* **19**, 281–294 (2003).

476. Sadeghifar, F., Böhm, S., Vintermist, A. & Östlund Farrants, A.-K. The B-WICH chromatin-remodelling complex regulates RNA polymerase III transcription by promoting Max-dependent c-Myc binding. *Nucleic Acids Res* **43**, 4477–4490 (2015).

477. Dvinge, H. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Letters* vol. 592 2987–3006 Preprint at https://doi.org/10.1002/1873-3468.13119 (2018).

478. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**, 108–121 (2014).

479. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718 (2009).

480. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell* **43**, 904–914 (2011).

481. Yao, R.-W., Wang, Y. & Chen, L.-L. Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21**, 542–551 (2019).

482. Qué, D. & Dalal, Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. doi:10.7554/eLife.03254.001.

483. Bergmann, J. H. *et al.* Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J* **30**, 328–340 (2011).

484. Nishimura, K. *et al.* DEAH box RNA helicase DHX38 associates with satellite I noncoding RNA involved in chromosome segregation. *Genes to Cells* **24**, 585–590 (2019).

485. Ideue, T., Cho, Y., Nishimura, K. & Tani, T. Involvement of satellite I noncoding RNA in regulation of chromosome segregation. *Genes to Cells* **19**, 528–538 (2014).

486. Liu, H. *et al.* Mitotic Transcription Installs Sgo1 at Centromeres to Coordinate Chromosome Segregation. *Mol Cell* **59**, 426–436 (2015).

487. Chan, F. L. *et al.* Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proceedings of the National Academy of Sciences* **109**, 1979–1984 (2012).

488. Wong, L. H. *et al.* Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* **17**, 1146–1160 (2007).

489. Talbert, P. B. & Henikoff, S. Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. *Trends in Genetics* vol. 34 587–599 Preprint at https://doi.org/10.1016/j.tig.2018.05.001 (2018).

490. Ideue, T., Cho, Y., Nishimura, K. & Tani, T. Involvement of satellite I noncoding RNA in regulation of chromosome segregation. *Genes to Cells* **19**, 528–538 (2014).

491. Thillainadesan, G. *et al.* Conserved protein Pir2ARS2 mediates gene repression through cryptic introns in lncRNAs. *Nat Commun* **11**, 2412 (2020).

492. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem* **89**, 359–388 (2020).

493. Vijayakumari, D. *et al.* Early splicing functions of fission yeast Prp16 and its unexpected requirement for gene Silencing is governed by intronic features. *RNA Biol* **16**, 754–769 (2019).

494. Huen, M. S. Y. *et al.* SON is a spliceosome-associated factor required for mitotic progression. *Cell Cycle* **9**, 2679–2685 (2010).

495. Logarinho, E., Resende, T., Torres, C. & Bousbaa, H. The Human Spindle Assembly Checkpoint Protein Bub3 Is Required for the Establishment of Efficient Kinetochore–Microtubule Attachments. *Mol Biol Cell* **19**, 1798–1813 (2008).

496.   Jiang, H. *et al.* A Microtubule-Associated Zinc Finger Protein, BuGZ, Regulates Mitotic Chromosome Alignment by Ensuring Bub3 Stability and Kinetochore Targeting. *Dev Cell* **28**, 268–281 (2014).

497.   Ahn, E.-Y. *et al.* SON Controls Cell-Cycle Progression by Coordinated Regulation of RNA Splicing. *Mol Cell* **42**, 185–198 (2011).

498.   Blower, M. D. Centromeric Transcription Regulates Aurora-B Localization and Activation. *Cell Rep* **15**, 1624–1633 (2016).

499.   Close, P. *et al.* DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* **484**, 386–389 (2012).

500.   Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Mol Cell* **38**, 202–210 (2010).

501.   Saeki, H. & Svejstrup, J. Q. Stability, Flexibility, and Dynamic Interactions of Colliding RNA Polymerase II Elongation Complexes. *Mol Cell* **35**, 191–205 (2009).

502.   Barra, V. & Fachinetti, D. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications* vol. 9 Preprint at https://doi.org/10.1038/s41467-018-06545-y (2018).

503.   Malumbres, M. Cyclin-dependent kinases. *Genome Biol* **15**, 122 (2014).

504.   Łukasik, P., Załuski, M. & Gutowska, I. Cyclin-dependent kinases (Cdk) and their role in diseases development–review. *International Journal of Molecular Sciences* vol. 22 1–33 Preprint at https://doi.org/10.3390/ijms22062935 (2021).

505.   Matsuura, I. *et al.* Cyclin-dependent kinases regulate the antiproliferative function of Smads. *Nature* **430**, 226–231 (2004).

506.   Hydbring, P. *et al.* Phosphorylation by Cdk2 is required for Myc to repress Ras-induced senescence in cotransformation. *Proceedings of the National Academy of Sciences* **107**, 58–63 (2010).

507.   Narasimha, A. M. *et al.* Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *Elife* **3**, (2014).

508.   Weintraub, S. J. *et al.* Mechanism of active transcriptional repression by the retinoblastoma protein. *Nature* **375**, 812–816 (1995).

509.   Harbour, J. W., Luo, R. X., Santi, A. D., Postigo, A. A. & Dean, D. C. Cdk Phosphorylation Triggers Sequential Intramolecular Interactions that Progressively Block Rb Functions as Cells Move through G1. *Cell* **98**, 859–869 (1999).

510.   Egloff, S. CDK9 keeps RNA polymerase II on track. *Cellular and Molecular Life Sciences* **78**, 5543–5567 (2021).

511.   Wood, D. J. & Endicott, J. A. Structural insights into the functional diversity of the CDK–cyclin family. *Open Biology* vol. 8 Preprint at https://doi.org/10.1098/rsob.180112 (2018).

512.   Greifenberg, A. K. *et al.* Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep* **14**, 320–331 (2016).

513.   Zhang, X. *et al.* Transcriptional Repression by the BRG1-SWI/SNF Complex Affects the Pluripotency of Human Embryonic Stem Cells. *Stem Cell Reports* **3**, 460–474 (2014).

514. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**, 407–422 (2017).

515. Giles, K. A. *et al.* Integrated epigenomic analysis stratifies chromatin remodellers into distinct functional groups. *Epigenetics Chromatin* **12**, 12 (2019).

516. Brahma, S. *et al.* INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nat Commun* **8**, 15616 (2017).

517. Alatwi, H. E. & Downs, J. A. Removal of H2A.Z by <scp>INO</scp> 80 promotes homologous recombination. *EMBO Rep* **16**, 986–994 (2015).

518. Zlatanova, J. & Thakar, A. H2A.Z: View from the Top. *Structure* **16**, 166–179 (2008).

519. Bajpai, R. *et al.* CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958–962 (2010).

520. Russo, R. *et al.* ZBTB2 protein is a new partner of the Nucleosome Remodeling and Deacetylase (NuRD) complex. *Int J Biol Macromol* **168**, 67–76 (2021).

521. Choi, E. S., Cheon, Y., Kang, K. & Lee, D. The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3. *Nat Commun* **8**, (2017).

522. Singh, P. P. *et al.* Hap2-Ino80-facilitated transcription promotes de novo establishment of CENP-A chromatin. *Genes Dev* **34**, 226–238 (2020).

523. Cai, G., Yang, Q. & Sun, W. RSF1 in cancer: interactions and functions. *Cancer Cell International* vol. 21 Preprint at https://doi.org/10.1186/s12935-021-02012-9 (2021).

524. Kulaeva, O. I., Hsieh, F.-K., Chang, H.-W., Luse, D. S. & Studitsky, V. M. Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**, 76–83 (2013).

525. Foltman, M. *et al.* Eukaryotic Replisome Components Cooperate to Process Histones During Chromosome Replication. *Cell Rep* **3**, 892–904 (2013).

526. Formosa, T. & Winston, F. The role of FACT in managing chromatin: Disruption, assembly, or repair? *Nucleic Acids Research* vol. 48 11929–11941 Preprint at https://doi.org/10.1093/nar/gkaa912 (2021).

527. Okada, M. *et al.* The CENP-H–I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat Cell Biol* **8**, 446–457 (2006).

528. Sadeghifar, F., Böhm, S., Vintermist, A. & Östlund Farrants, A.-K. The B-WICH chromatin-remodelling complex regulates RNA polymerase III transcription by promoting Max-dependent c-Myc binding. *Nucleic Acids Res* **43**, 4477–4490 (2015).

529. Hill, J. R. & Morris, D. R. Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the cis-acting upstream open reading frame. *J Biol Chem* **268**, 726–31 (1993).

530. Gray, N. K. & Wickens, M. CONTROL OF TRANSLATION INITIATION IN ANIMALS. *Annu Rev Cell Dev Biol* **14**, 399–458 (1998).

531. Rathore, A., Martinez, T. F., Chu, Q. & Saghatelian, A. Small, but mighty? Searching for human microproteins and their potential for understanding health and disease. *Expert Rev Proteomics* **15**, 963–965 (2018).

532. Ruiz Cuevas, M. V. *et al.* Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep* **34**, (2021).

533. Yang, L. *et al.* Integrated Pipeline of Isotopic Labeling and Selective Enriching for Quantitative Analysis of N-Glycome by Mass Spectrometry. *Anal Chem* **91**, 1486–1493 (2019).

534. Du, Z. *et al.* A New Strategy for High-Efficient Tandem Enrichment and Simultaneous Profiling of N-Glycopeptides and Phosphopeptides in Lung Cancer Tissue. *Front Mol Biosci* **9**, (2022).

535. Arnaud, J., Audfray, A. & Imberty, A. Binding sugars: From natural lectins to synthetic receptors and engineered neolectins. *Chem Soc Rev* **42**, 4798–4813 (2013).

536. Xiao, H., Suttapitugsakul, S., Sun, F. & Wu, R. Mass Spectrometry-Based Chemical and Enzymatic Methods for Global Analysis of Protein Glycosylation. *Acc Chem Res* **51**, 1796–1806 (2018).

537. Xiao, H., Chen, W., Smeekens, J. M. & Wu, R. An enrichment method based on synergistic and reversible covalent interactions for large-scale analysis of glycoproteins. *Nat Commun* **9**, (2018).

538. Jedrychowski, M. P. *et al.* Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Molecular & Cellular Proteomics* (2011) doi:10.1074/mcp.m111.009910.

539. Zhang, Y. *et al.* Systems analysis of singly and multiply O-glycosylated peptides in the human serum glycoproteome via EThcD and HCD mass spectrometry. *J Proteomics* **170**, 14–27 (2018).

540. Tadeu, A. M. B. *et al.* CENP-V is required for centromere organization, chromosome alignment and cytokinesis. *EMBO J* **27**, 2510–2522 (2008).

541. Wulff-Fuentes, E. *et al.* The human O-GlcNAcome database and meta-analysis. *Sci Data* **8**, (2021).

542. York, W. S. *et al.* GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology* **30**, 72–73 (2020).

543. Sakabe, K., Wang, Z. & Hart, G. W. β-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences* **107**, 19915–19920 (2010).

544. Fujiki, R. *et al.* GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* **480**, 557–560 (2011).

545. Klein, A. J. & Zaia, Joseph (Bioinformatics Program, D. of B. B. U. Protocol for analysis of glycoproteomics LC-MS data using GlycReSoft. **6762**, 1–21 (2020).

546. Madsen, J. A., Farutin, V., Lin, Y. Y., Smith, S. & Capila, I. Data-independent oxonium ion profiling of multi-glycosylated biotherapeutics. *MAbs* 1–11 (2018) doi:10.1080/19420862.2018.1494106.

547. Luo, B. *et al.* Boronic Acid-Functionalized Magnetic Metal–Organic Frameworks via a Dual-Ligand Strategy for Highly Efficient Enrichment of Phosphopeptides and Glycopeptides. *ACS Sustain Chem Eng* **7**, 6043–6052 (2019).

548. Zeidan, Q., Wang, Z., de Maio, A. & Hart, G. W. O-GlcNAc Cycling Enzymes Associate with the Translational Machinery and Modify Core Ribosomal Proteins. *Mol Biol Cell* **21**, (1922).

549. Liu, Y. *et al.* Proteomic profiling and genome-wide mapping of O-GlcNAc chromatin-associated proteins reveal an O-GlcNAc-regulated genotoxic stress response. *Nat Commun* **11**, (2020).

550. Cui, Y. *et al.* OGA is associated with deglycosylation of NONO and the KU complex during DNA damage repair. *Cell Death Dis* **12**, (2021).

551. Du, Z. *et al.* A New Strategy for High-Efficient Tandem Enrichment and Simultaneous Profiling of N-Glycopeptides and Phosphopeptides in Lung Cancer Tissue. *Front Mol Biosci* **9**, (2022).

552. Cho, K. C., Chen, L., Hu, Y., Schnaubelt, M. & Zhang, H. Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides. *ACS Chem Biol* **14**, 58–66 (2019).

553. Alteen, M. G. *et al.* Mechanism of Human Nucleocytoplasmic Hexosaminidase D. *Biochemistry* **55**, 2735–2747 (2016).

554. Xu, S., Zheng, J., Xiao, H. & Wu, R. Simultaneously Identifying and Distinguishing Glycoproteins with O-GlcNAc and O-GalNAc (the Tn Antigen) in Human Cancer Cells. *Anal Chem* **94**, 3343–3351 (2022).

8 List of Figures

9 List of Tables

10 Abbreviation

Abbreviation	Full description
APC	Anaphase-promoting complex
CCAN	Centromere Associated Network
CDK	Cyclin dependent kinase
CENP	Centromere protein
cenRNA	Centromeric RNA
ChiP-MS	Chromatin immuno precipitation Mass Spectrometry
CID	Collision-induced dissociation (MS-fragmentation mode)
CMG complex	Cdc45-MCM-GINS complex
CPC	Chromosome Passenger Complex
DNMT	DNA methyltransferases
HCD	Higher-energy collisional dissociation (MS-fragmentation mode)
HDAC	Histone Deacetylases
ISWI	Imitation SWItch remodeler family
KIP/CIP	CDK interacting protein/Kinase inhibitory protein
KMN	Knl1/Mis12/Ndc80 complex (outer kinetochore)
KT-MT	Kinetochore-Microtubule
lncRNA	Long noncoding RNA
MCM complex	Minichromosome maintenance complex
MS	Mass Spectrometry
NFPD	Nuclear Fishing Pull Down
OGA	Protein O-GlcNAcase

OGT	O-linked N-acetylglucosaminyltransferase
PcG	Polycomb group proteins
ppGALNT	Polypeptide N-acetylgalactosaminyltransferase
PTM	Post-translational modification
RB1	Retinoblastoma protein
SAC	Spindle Assembly Checkpoint complex
TET enzyme	ten eleven translocation enzyme
TFIIH	Transcription factor IIH
TSS	Transcription Start Site
UDP-GalNAc	Uridine diphosphate *N*-acetylgalactosamine
UDP-GlcNAc	Uridine diphosphate *N*-acetylglucosamine
XL-MS	Cross-linking Mass Spectrometry

11 Acknowledgements

The last three to four years of my PhD were very exciting but also exhausting and frustrating at the same time. In this paragraph I would like to thank the people I met and became friends over the time who helped me to succeed in obtaining my PhD degree.

First, I would like to thank **Dr. Franz Herzog** giving me the opportunity to work in his laboratory and could develop myself as a scientist. It was a great experience, and I am grateful for his supervision and the time he spent to develop the projects and guide me through the process of gaining confidence in my results for the projects I worked on.

Moreover, a special thank goes to **Prof. Roland Beckmann** and **Prof. Karl-Peter Hopfner** especially in my last phase of my PhD at the Gene Center. The incredible support as well as their supervision helped me to get through this phase and finally obtaining my degree. I thank the board members of my PhD defence committee who supported me and helped me to reach my goal. Moreover, I would like to thank the **GRK1721**-members as well as the coordinator **Dr. Beate Hafner** for their support and the scientific input I got during my time there.

Another major thank you goes to **PD Dr. Gregor Witte**. It was great working with you, and I really appreciated your advice in any life situation. It really helped to find my focus and the fruitful discussions around my subject meant a lot to me. I would like to thank **Franziska Kunert** and **Stephan Woike** as well. We went through a lot, and I will really miss the coffee breaks with you on the roof top or the spontaneous conversation about our topics and exchanging ideas or just venting when something or someone just got on our nervous. I wish you both the best.

Learning how to maintain insect cells got easy with **Brigitte Keßler** and **Manuela Moldt**. Thanks for your great time at the insect cell room and the help I got how to maintain the insect stocks. Without your help we all would be completely lost.

For a short time of period, I could supervise two very talented master students: **Stefan Georgiev** and **Diego Ninez**. You both really helped me through difficult

times and made things happen, which I would have not dreamt of. I really enjoyed supervising you both and I hope you could learn something from me and could gain some knowledge. I wish you a bright future.

The next passage is dedicated to **Romy Böttcher, Joanna Musial, Alicia Musial and Andrea Gilmozzi**. It has been a wild ride for me, from being an innocent PhD student to become a lab manager/ postdoc, sometimes even PI in one person. You all helped me to focus on myself and guided me how to do certain things to reach my goals. I will not forget how much you helped me out in just little things such as making media or how to make competent cells. Through your help I got a better scientist, since I learned everything from the very start. But what I even more appreciate is your emotional support during this time. It helped me to survive and get stronger. I will not forget how Joanna always says: "*Alles was Dich nicht umbringt, macht Dich nur noch stärker*!". Thanks for your endless support!

Another major thank you goes to **Timo Denk, Matthias Thoms, Timur Mackens-Kiani, Shuangshuang Niu** and **Leona Chitoiu**. Whenever I had a problem or just needed advice, I could always come and ask, or we found a solution together. Your scientific input regarding topics crossing over your subjects were really helpful and guided me through the journey of my first project. I could not believe that it got so interesting in the end. Thanks for being there and helped me out. I wish you all the best.

My biochemical data of my first project would not have such an impact without the help of my collaborators **Jeannette Koch** und **Dr. Hartmann Harz**. Both helped me to obtain cell biological data to understand why my protein is acting as it is. But I really enjoyed the uncomplicated scientific help and fruitful discussions around this topic.

Another thanks for an collaborator goes to **Stefan Marchner** (PhD at Prof. Anja Hoffmann-Röder group, Chemistry Department, LMU). Even though our Glycosylation Project was not that successful I think our ideas and synthesis had potential. I will really miss coming over to the chemistry department and just talk to

you about synthesis steps or enzyme cascades for our projects. I wish you all the best for your future and it was a great pleasure to work with you.

I think I would not be at that stage where I am now and that I can manage myself without the help of my former colleagues **Dr. Sylvia Sing, Dr. Götz-Norman Hagemann** and **Dr. Victor-Solis Mezarino**. From Sylvie I learned how to clone but also how to organize myself in the very first moments and ask myself critical questions regarding my scientific results and what controls had to be implemented. I could always come to you and ask you anything, you always had an answer and even when you were not sure, you knew how to obtain it. It's been heart-breaking for me that we could not spend more time together, especially on Götze's and my project. I think with your ideas and support we could have made even more great experiments to find out, what was really going on. I cannot thank you for being still there and that I can always contact you and having an ear for my problems. I wish you all the best in future and would love to stay in touch with you. Unfortunately, Götz left to soon too. I learned from you not only the very challenging maintenance of the MS machine and the preparation of a human nuclear extract but also how to maintain human cells and even creating my own cell lines. Moreover, your constant input and critical view on our project helped me to focus more and become a better scientist. Your data about the CENP-A time course made this thesis possible and the fact that you are still present, helping me out even having a child to take care of, means a lot to me and I am grateful to know you and have you at my side to hopefully publish this great data. I thank you for your input and support and I hope we can manage to get this data out into the world. I wish your family all the best and hope to stay in touch with you. The challenges I had to face for my MS-data was huge, but I could always rely on Victor. Imaging having your own business and even a child did not stop him from helping me out with his impressive knowledge and great expertise and how to visualize my data but also giving meaning to it. It feels to me insane what kind of capacity you have; it always impresses me. Without your endless support and encouragement, I would be completely lost and the things I have discussed could not be implemented, since I would have never discovered them. You made a better data scientist out of me, since I am eager to learn how to

do it, so I can perform it on my own. You were a great inspiration and the fact that we become friends over time honours me. I wish you and your family all the best and I would love to stay in touch with you.

I have already mentioned the difficulties of my last PhD phase in previous paragraphs but there is one person I should mention here, who really deserves much attention. I thank Mr. **Prof. Axel Imhof** and his group members for integrating me in your group and taking me as a "scientific refugee". I will not forget how all of us were at the Schneeferner Haus and had a great time talking about science but also coming together as a group and becoming friends. I cannot describe how much this means to me. Moreover, the weekly lab meetings helped me to become a better scientist by learning what you all are working on, and I am always mind blown by the expertise the PhDs, PostDocs and scientific coordinators are bringing along. Mr Prof. Imhof gave me shelter and very importantly gave me back the excitement I felt in the beginning when I started my research career. I felt lost along the way of my PhD but with his supervision, I got the spark/ passion back I lost during the time. I am so excited to start in his lab as a PostDoc and I am looking forward to discovering the limitations of our imagination. I thank you from the bottom of my heart.

Finally, I thank my family, my parents **Pawan Kumar** and **Gabriele Kumar** as well my little sister **Priya Vanessa Kumar**. You knew from the very beginning that I would like to do something in research or even become a medical doctor. After having some realization that becoming a medical doctor might not be the best option, you all always supported me whenever you could in my dream of becoming a scientist. I cannot forget the sleepless nights we had and the hard work you had to put into our former shop to get me through this phase. I hope I could make you proud by obtaining this degree and being now settled in my life. I do not know how I can every repay this never-ending depth but by giving you a stage here, I would like to say: Thank you for being my family and supporting me so hard. I know I was never the easiest person, and I am sorry for giving you sleepless nights. You helped me to find my passion, and this helped me to fight through all the difficulties and circumstances together. You made me to the person, who I am today. Without your endless support I would have given up on a very early stage but especially the last

third of my PhD showed me how fortunate I am to have you and call you, my family. Moreover, we went through a lot in a short time, and I would like to see how the dark clouds vanish from the sky and we all can finally enjoy the warmth of the sun. It might be cheesy, but I really love you and thank you form the bottom of my heart for your endless support and love.