Aus dem Adolf-Butenandt-Institut für Molekularbiologie
der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Peter Becker

PWWP2A: a novel H2A.Z nucleosome interactor involved in cell cycle regulation

Dissertation zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der Ludwig-Maximilians-Universität München

vorgelegt von

Sebastian Pünzeler
aus Düren

2015

**Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München**

Eingereicht am: 24.03.2015

Betreuerin: PD Dr. Sandra B. Hake

Zweitgutachter: Prof. Dr. Axel Imhof

Dekan: Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR

Tag der mündlichen Prüfung: 10.07.2015

Eidesstattliche Versicherung

Ich, Sebastian Pünzeler, erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

"PWWP2A: a novel H2A.Z nucleosome interactor involved in cell cycle regulation"

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den …………………… …………………………...

(Sebastian Pünzeler)

TABLE OF CONTENTS

SUMMARY

In the nucleus of eukaryotic cells, DNA is assembled with basic histone proteins into chromatin, with the nucleosome as its fundamental entity. The nucleosome consists of DNA wrapped around a histone-octamer, which itself is comprised of two copies of each core histone H2A, H2B, H3 and H4. In order to allow DNA-related processes like replication, transcription or repair to occur, access to the packaged DNA has to be granted. This is achieved by different interconnected epigenetic regulatory mechanisms, where the exchange of core histones through a replacement histone variant is one possibility. This replacement can influence structural dynamics of the respective chromatin domain, but also enrich it with new sites of posttranslational chemical modifications or enable the interaction with specific effector proteins. To date, several human histone H2A variants are known and among them, H2A.Z is most thoroughly investigated. This variant is essential in many organisms and implicated in important cellular processes like transcriptional regulation, DNA repair and cell cycle control. In vertebrates, two genes, *H2AFZ* and *H2AFV*, give rise to two H2A.Z isoforms, H2A.Z.1 and H2A.Z.2, differing in only three amino acids. In my PhD thesis I report the identification of the human H2A.Z isoform-containing nucleosome interactome by employing a label-free quantitative mass spectrometry based approach for the first time on mononucleosomes containing GFP-tagged H2A.Z variants. H2A.Z-containing nucleosomes are associated with the SRCAP complex, a chaperone/remodeling complex that exchanges nucleosomal H2A-H2B for free H2A.Z-H2B dimers, and with members of several other chromatin-related complexes. Besides these, also new chromatin interactors were identified that were either not yet assigned to any complex or not yet fully characterized. Further investigations focused on the PWWP domain-containing candidate protein PWWP2A as a novel H2A.Z nucleosome binder. *In vitro* binding and *in vivo* FRAP assays not only recapitulated PWWP2A's strong H2A.Z chromatin interaction but also revealed that a distinct internal amino acid stretch of PWWP2A conveys H2A.Z specificity and general nucleosome binding. Additionally, its conserved PWWP domain is necessary for a strong chromatin interaction. Native MNase ChIP-sequencing revealed PWWP2A's colocalization with H2A.Z at gene promoters and correlation with microarray transcriptome data suggested a potential involvement of PWWP2A in transcriptional regulation. Interaction proteomics with GFP-tagged PWWP2A not only validated PWWP2A's binding to H2A.Z nucleosomes but also discovered its association with a minimal NuRD complex that might mediate chromatin deacetylation without remodeling. Functional consequences of PWWP2A's interaction with

SUMMARY

H2A.Z at active genes and its partly contradictory connections with other proteins and complexes as revealed by its interactome are currently investigated. However, depletion of PWWP2A in cells by RNAi resulted in impaired cellular proliferation with a decrease of cells in S-phase and an accumulation of cells in metaphase of mitosis. How PWWP2A influences the cell cycle remains elusive to date but the underlying molecular mechanism might require PWWP2A's action in transcriptional regulation and its interaction with certain chromatin domains.

ZUSAMMENFASSUNG

Im Zellkern einer eukaryotischen Zelle ist die DNS mit Hilfe von basischen Histonproteinen zu Chromatin verpackt, mit dem Nukleosom als grundlegender Einheit. Das Nukleosom besteht aus DNS, die um ein Histon-Oktamer gewunden ist, welches jeweils zwei Kopien der Kernhistone H2A, H2B, H3 und H4 umfasst. Um DNS-basierte Prozesse wie die Replikation, Transkription oder DNS-Reparatur zu ermöglichen, muss die DNS unterschiedlichen Proteinen zugänglich sein. Dieses wird durch verschiedene epigenetische Regulationsmechanismen ermöglicht, wobei der Austausch von Kernhistonen durch eine Histonvariante eine von mehreren Möglichkeiten ist. Dieser Austausch kann die Struktur der jeweiligen Chromatindomäne beeinflussen, sie aber auch mit zusätzlichen posttranslationalen chemischen Modifizierungen ausstatten oder die Interaktion mit spezifischen Effektorproteinen ermöglichen. Bis heute sind mehrere humane Histon H2A Varianten bekannt und von diesen ist H2A.Z am gründlichsten untersucht. Die Variante ist essentiell in vielen Organismen und in wichtigen zellulären Prozessen, wie die transkriptionelle Regulation, die DNS-Reparatur und die Zellzykluskontrolle involviert. In Vertebraten kodieren zwei Gene, *H2AFZ* und *H2AFV*, für zwei H2A.Z Isoformen, H2A.Z.1 und H2A.Z.2, die sich in nur drei Aminosäuren unterscheiden. In meiner Dissertation habe ich das Interaktom von Nukleosomen, die GFP-fusionierte H2A.Z Isoformen enthalten, durch quantitative Massenspektrometrie identifiziert. Diese Methode wurde zum ersten Mal mit Nukleosomen durchgeführt und von mir etabliert. Ich konnte zeigen, dass H2A.Z Nukleosomen mit dem SRCAP Komplex, einem Chaperon- und Remodellierungskomplex, der nukleosomale H2A-H2B Dimere gegen freie H2A.Z-H2B Dimere austauscht, und mit Mitgliedern von mehreren anderen Chromatin-bindenden Komplexen assoziiert sind. Daneben wurden auch neue Chromatin Bindeproteine identifiziert die entweder noch keinem Komplex zugeteilt oder noch nicht vollständig charakterisiert wurden. Weiterführende Untersuchungen fokussierten sich auf das bislang uncharakterisierte Protein PWWP2A als neues H2A.Z Nukleosomen Bindeprotein, welches über eine PWWP Domäne verfügt. *In vitro* Bindungsstudien und *in vivo* FRAP-Analysen bestätigten nicht nur die starke Bindung von PWWP2A an H2A.Z Nukleosomen, sondern enthüllten auch, dass ein bestimmter interner Aminosäureabschnitt die Spezifität von PWWP2A für H2A.Z sowie eine generelle Nukleosomeninteraktion vermittelt. Zusätzlich ist die PWWP Domäne für eine starke Nukleosomenbindung *in vivo* nötig. Native Immunpräzipitationen von MNase-verdauten Chromatinfragmenten und anschließende Sequenzierung der präzipitierten DNS, sowie die

3

ZUSAMMENFASSUNG

Korrelierung dieser Daten mit Transkriptomanalysen, weisen auf eine Beteiligung von PWWP2A an der Regulation der Transkription hin. Ein Interaktionsproteom von GFP-fusioniertem PWWP2A validierte die Interaktion mit H2A.Z Nukleosomen und führte darüberhinaus zu der Entdeckung eines PWWP2A-assoziierten, minimalen NuRD Komplexes, der Chromatin möglicherweise deacetyliert ohne es zu remodellieren. Die funktionalen Konsequenzen, die sich aus der Interaktion von PWWP2A mit H2A.Z an aktiven Genen und seinen - teilweise gegensätzlichen - Verknüpfungen mit weiteren Proteinen und Komplexen ergeben, werden zu Zeit untersucht. Die Abreicherung von PWWP2A in Zellen mittels RNS-Interferenz führte jedoch zu einer beeinträchtigten Zellvervielfältigung, einhergehend mit einer reduzierte Zahl von replizierenden Zellen und einem Arrest von Zellen in der Metaphase der Mitose. Wie PWWP2A den Zellzyklus beeinflusst ist zur Zeit noch unklar. Der zugrundeliegende molekulare Mechanismus könnte eine Aktivität PWWP2As in der transkriptionellen Regulation erfordern sowie seine Interaktion mit verschiedenen Chromatindomänen.

1 Introduction

1.1 Chromatin structure

DNA does not exist as a naked molecule in the nucleus of a eukaryotic cell but is assembled with basic histone proteins and non-histone proteins into chromatin. In 1882, W. Flemming suggested the term 'chromatin' for "that substance in the cell nucleus which is readily stained" (Flemming, 1882). 'Chroma' is the Greek expression for 'color', referring to the ability of chromatin to be stainable with dyes for light microscopy. Since this early description of chromatin many significant scientific milestones, ranging from the discovery of the histones (Kossel, 1911), the identification of DNA as the carrier of genetic information (Avery et al., 1944), its double-helical structure (Watson and Crick, 1953) to the electron-microscopic and, later, x-ray visualization of chromatin and its repeating substructure (Luger et al., 1997; Olins and Olins, 1974; Woodcock et al., 1976) had an immense impact on our understanding of chromatin structure today. The basic unit of chromatin is the nucleosome (Kornberg, 1974; Oudet et al., 1975) consisting of ~145-147 bp of superhelical, left-handed DNA organized in 1.65 turns by the histone octamer composed of two copies of each histone H2A, H2B, H3 and H4 (Luger et al., 1997; Richmond and Davey, 2003) (**Figure 1.1.1**). These 'core' histones can be structurally and functionally divided into two distinct regions: the histone fold and the histone tail.
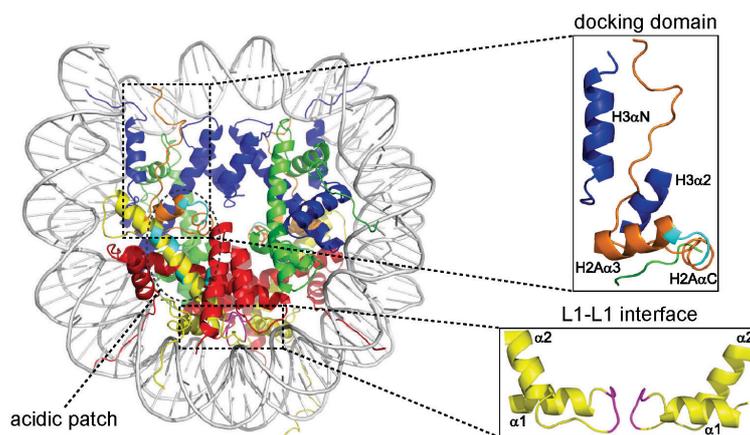
Figure 1.1.1 The nucleosome crystal structure. The crystal structure of the nucleosome as it was solved in 1997 by Luger et al. (Luger et al., 1997). The DNA is depicted in grey, histone H2A in yellow, H2B in red, H3 in blue and H4 in green. Important structural features are shown as blow-ups on the right and are highlighted as follows: the acidic patch in cyan, the L1-L1 interface in magenta and the docking domain in orange. Figure taken from (Bonisch and Hake, 2012).

Whereas the histone tail is largely unstructured, the histone fold domain (HFD) exhibits distinct secondary structures and facilitates interactions between the histones but also with the DNA (Andrews and Luger, 2011). The HFD forms three helices, α1 - α3, connected by two loops, L1 and L2. In solution, these regions arrange as antiparallel heterodimers, where H3 only pairs with H4 and H2A only pairs with H2B. Two H3-H4 pairs form a (H3-H4)$_2$ tetramer through a 4-helix bundle (4HB). In the presence of DNA (or high salt) two H2A-H2B dimers interact with the tetramer through a 4HB between H2A and H4 to form the nucleosome. The C-terminus of H2A, the 'docking domain', which contacts the N-terminal tail of H3 and the C-terminus of H4, mediates additional interaction. Moreover, the interaction between the H2A L1 loops of both H2A-H2B dimers (L1-L1 interface) supports nucleosomes stabilization. The DNA interacts base-unspecific with the histones, its minor grooves always facing the octamer (Luger et al., 1997). Nucleosome assembly is supported by different histone chaperones that bind histones, guide them to the DNA and ATP-independently assemble nucleosomes without becoming part of them (Elsasser and D'arcy, 2012). Adjacent nucleosomes, associated by the linker DNA, form the primary chromatin structure, the beads-on-a-string nucleosome array. These nucleosomal arrays assemble into three-dimensional higher-order chromatin structures that ultimately build the entire metaphase chromosome. The linker histone H1 plays an important role during the compaction process (Thoma et al., 1979) but its biological role is still not completely understood (Fan et al., 2005). It interacts tightly with nucleosomes possessing linker DNA and leads to the organization of 20 additional base pairs of DNA to form a chromatosome (Caterino and Hayes, 2011; Thoma et al., 1979). *In vitro* data suggest that nucleosomal arrays form a secondary structure under physiological salt conditions that is termed the 30nm fiber. Its exact structure has not been solved yet but it is proposed to either form a solenoid helical structure or a zigzag helical structure (Tremethick, 2007). The later does not include the linker histone H1, thus questioning its biological relevance as most metazoan nucleosomes are bound by H1 (Luger et al., 2012). However, *in vivo* additional architectural chromatin proteins exist (like Heterochomatin protein 1 (HP1), Methyl-CpG-binding protein 2 (MeCP2) or the Polycomb complex) that modulate chromatin structure thereby influencing DNA accessibility (McBryant et al., 2006).

1.2 Modes of 'epigenetic' regulation

The compaction of DNA into chromatin to fit the genetic information into a eukaryotic nucleus is not only an issue of storage management but is in fact key to modulate the accessibility of DNA. More open and in general transcriptionally active chromatin is known as euchromatin, whereas densely and tightly packed chromatin is termed constitutive (always transcriptionally silent) or facultative (transcriptionally silent but context dependent) heterochromatin (Trojer and Reinberg, 2007). However, chromatin is the template for not only gene transcription, but of all DNA-related processes. Therefore, also DNA repair and DNA replication depend on proper chromatin regulation. To alter DNA packaging dynamics and chromatin conformation different epigenetic mechanisms have evolved. Conrad Waddington suggested the term 'epigenetics' in 1942 as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1942). Today, epigenetics denominates heritable changes in genome function that occur independent of the DNA sequence (Probst et al., 2009). The inheritance aspect is fundamental to this definition and includes the self-maintenance of the epigenetic information independent of the stimulus, that lead to the phenotype in the progenitor cell or parental organism (Campos et al., 2014). However, the underlying molecular mechanisms that are responsible for epigenetic inheritance, may it be during cell divisions or even transgenerational, remain elusive and are highly debated (Heard and Martienssen, 2014). In principle, six 'epigenetic' mechanisms have been proposed involving different but interconnected key players (Bonisch et al., 2008): DNA methylation, ATP-dependent chromatin remodeling, chromatin localization within the nuclear architecture, non-coding RNAs, posttranslational histone tail modifications (PTMs) and the incorporation of histone variants. In the following sections the four chromatin regulatory mechanisms most important for this thesis will be briefly introduced.

1.2.1 DNA methylation

One of the best understood examples of epigenetic regulation is mediated by the methylation of the fifth position of cytosine (5mC) in the genome of animals and plants (Feng et al., 2010). In mammals, DNA methylation is primarily occurring in the CpG dinucleotide (cytosine and guanine separated by phosphate) context where 60 – 80 % of CpG sites are methylated (Smith and Meissner, 2013). Cytosine methylation is mainly associated with gene silencing and influences not only pluripotency, X-chromosome inactivation and genomic

imprinting (Bird, 2002) but also genome stability (Howard et al., 2008). The inhibitory effect on transcription is presumably mediated by interference with transcription factor binding sites or recruitment of methyl-CpG binding proteins that attract repressive complexes (Fuks, 2005). Genomic methylation patterns are established and maintained by a group of DNA methyltransferases (Denis et al., 2011). DNMT3A and DNMT3B are responsible for the *de novo* CpG methylation during embryonic development (Okano et al., 1999), whereas DNMT1 methylates the unmodified and freshly synthesized progeny strand during DNA replication, thereby maintaining the parental methylation pattern (Bestor and Ingram, 1983). The existence of a reverse mechanism that mediates demethylation of CpG sites was only discovered in the last years (Delatte et al., 2014). Key players are the Ten Eleven Translocation (TET) enzymes (TET1, TET2 and TET3) that oxidize 5-methylcytosine (5mC) iteratively to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5-caC) (Ito et al., 2011; Tahiliani et al., 2009). This process initiates DNA demethylation either by passive dilution of 5hmc or through active base excision repair and cytosine replacement (Guo et al., 2011; He et al., 2011).

1.2.2 ATP-dependent chromatin remodeling

Crucial to endow chromatin with dynamic properties are molecular 'machines' that use the energy of ATP hydrolysis to remodel nucleosomes (Ho and Crabtree, 2010) thereby allowing binding of regulatory molecules. Consequences of chromatin modulations are either the exposure of DNA sites by nucleosomal repositioning (sliding), eviction or unwrapping, or the variation of nucleosome composition by histone ejection and histone variant replacement (Clapier and Cairns, 2009). According to their respective ATPase subunit, the multi-subunit remodeling complexes can be divided into four major subfamilies: SWI/SNF, ISWI, CHD and INO80 (Eberharter and Becker, 2004). It appears, that the high number of different complexes is due to their involvement in processes beyond remodeling (Morrison and Shen, 2009) or a functional specialization through association with different interaction partners (Ho and Crabtree, 2010). The process of histone variant incorporation for example often requires the combined action of chromatin remodelers and histone chaperones and is discussed in the histone variants section below.

1.2.3 Posttranslational histone modifications

Not only the DNA is covalently modified (see DNA methylation section) but also the histone proteins that form the histone octamer in every nucleosome. These PTMs are large in number:

acetylation, phosphorylation, methylation, ubiquitination, sumoylation, ADP ribosylation, glycosylation, propionylation, butyrylation, formylation, crotonylation, deimination and proline isomerization are known and reported to affect over 100 different residues on histone proteins (Kouzarides, 2007; Rothbart and Strahl, 2014; Tan et al., 2011). Especially the N-terminal unstructured histone tail can be decorated but also the globular histone fold domain is subjected to post-translational modification (Jack et al., 2013; Jack and Hake, 2014). Histone modification is in general a very dynamic process that highly affects genomic conformation and is tightly regulated by specific enzymes (Bannister and Kouzarides, 2011; Henikoff and Shilatifard, 2011). One group of these "modifying" enzymes ('writers'), like histone acetyltransferases (HATs), methyltransferases (HMTs) or kinases, is responsible for establishing modifications on certain residues like acetylation (lysine), methylation (lysine or arginine) and phosphorylation (serine, threonine or tyrosine). These PTMs can in turn be removed by a second group of modifiers ('erasers') like histone deacetylases (HDACs), histone demethylases or phosphatases. PTMs can influence the chromatin landscape essentially in three ways (Kouzarides, 2007). First, some affect chromatin structure by changing the charge of histones. Histone acetylation, for example, neutralizes the positive charge of the lysine residue and potentially loosens histone-DNA interaction, thereby mediating chromatin unfolding or even preventing formation of higher order chromatin structure (Tremethick, 2007). Second, PTMs can serve as binding platform to recruit non-histone proteins ('readers') carrying enzymatic activities that further affect chromatin structure. Third, these interactions can be prevented by other specific histone marks. Molecules interacting with PTMs possess specific domains that facilitate binding to distinct modifications (de la Cruz et al., 2005; Kouzarides, 2007; Maurer-Stroh et al., 2003; Mellor, 2006). Methyl-groups are bound by PHD-fingers, chromo-, tudor-, malignant brain tumor (MBT)- and PWWP-domains as well as WD40 repeats. Acetyl-groups are bound by bromodomains, phosphorylated residues by a domain in 14-3-3 proteins and unmodified histone residues by SANT-domains and PHD-fingers. The above-mentioned PTM-mediated mechanisms cooperate with other chromatin regulators and PTM-binding molecules to shape global chromatin environments and define active/accessible and silent/inaccessible chromatin (Kouzarides, 2007). Briefly, H3 lysine 4, 36 or 79 (tri)methylation and general H3 and H4 acetylation are associated with transcriptionally active chromatin (Ciabrelli and Cavalli, 2014). H3K27 trimethylation together with the Polycomb-repressive complexes 1 and 2 (PRC1 and PRC2) as well as H4K20 trimethylation and H3K9 di- and trimethylation in concert with HP1α are associated with transcriptionally inactive chromatin (Ciabrelli and

Cavalli, 2014). Besides transcription, other chromatin related processes like DNA replication or repair are associated with PTMs (Kouzarides, 2007) because they depend on the recruitment of effector molecules that help unraveling chromatin and manipulating the DNA. To orchestrate actions on chromatin, the "histone code" as an underlying principle was put forward several years ago (Stahl und Allis 2000). It proposed the existence of certain PTM combinations that facilitate important and specific downstream functions. Whether this code actually exists is highly debated and experimentally not verified yet. Notably, more and more proteins are identified that possess more then one PTM binding domain enabling them to 'read' a putative PTM code (Taverna et al., 2007). Ultimately, it is not clear, whether histone modifications really define certain chromatin states or if those chromatin features themselves attract the establishment of certain modifications (Henikoff and Shilatifard, 2011). However, recent genome-wide deep sequencing studies associating several chromatin regulators (e.g. Pol II and CTCF), a plethora of histone modifications and histone variants (H2A.Z) with a certain local chromatin environment (Ernst et al., 2011; Filion et al., 2010; Sexton et al., 2012; Shen et al., 2012) allowed to further subdivide chromatin into several active or repressive subtypes of chromatin (Ciabrelli and Cavalli, 2014).

1.2.4 Histone variants

In addition to histone PTMs, the exchange of the canonical histones through histone variants (**Figure 1.2.1**) alters nucleosome composition and influences the global chromatin landscape. Conventional (canonical) histones are transcribed replication-dependent from genomic clusters containing multiple copies of intron-less genes encoding the core histones H2A, H2B, H3 and H4 and the linker histone H1. In vertebrates, all 14 H4 genes encode the same protein, 15 H2A and 17 H2B genes encode for several and 13 H3 genes for 2 canonical isoforms (H3.1 and H3.2) with small amino acid differences (Marzluff et al., 2002; Maze et al., 2014). Both, the high number of canonical histone genes and their timed expression limited to S-phase, fit the needs of freshly synthesized and deposited histones behind the replication fork. Instead of a poly(A)-tail, canonical histone transcripts posses a 26 bp sequence forming a unique stem-loop structure that is responsible for S-phase expression, as well as translation and degradation of histone mRNA (Marzluff, 2005). In contrast, genes coding for histone variants are mostly single copy genes. They can contain introns and give rise to polyadenylated mRNAs that are expressed replication-independently throughout the cell cycle (Marzluff et al., 2008). Histone variant incorporation has been associated with diverse biological processes (Talbert and Henikoff, 2010) including pluripotency and the putative

contribution to cancer initiation and progression when deregulated (Skene and Henikoff, 2013; Vardabasso et al., 2014). Key to influence these processes is the sequence (and therefore structural and PTM) variation that discriminate histone variants from their canonical counterparts. Such changes can be rather subtle (e.g. H3 variants) or more significant (H2A variants) (Bernstein and Hake, 2006).
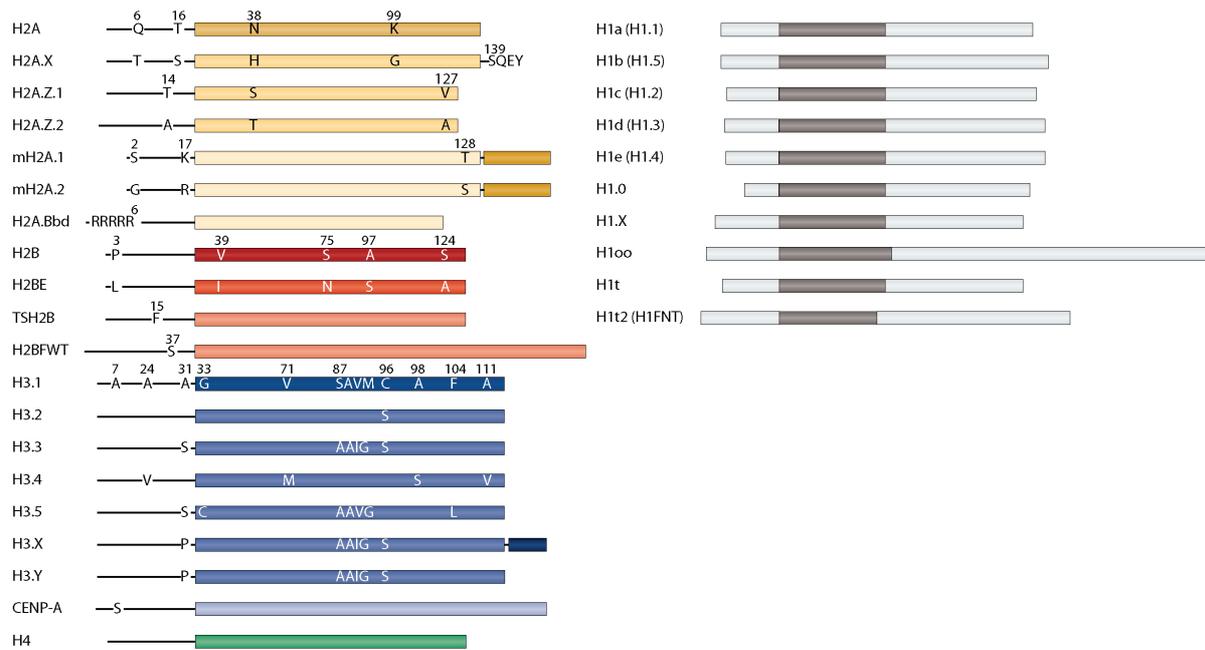
Figure 1.2.1 Human core and linker histone variants. Variants of the core histones H2A (yellow), H2B (red) and H3 (blue) as well as linker histone H1 variants (grey) are shown. So far, no human H4 variant (green) is known. Important residues reflecting key differences within a histone family are highlighted (except for the high divergent H1 variants). Globular macro domains of mH2A.1 and mH2A.2 (macroH2A.1 and macroH2A.2) are depicted in brown. Unstructured N- and C-terminal domains in H1 variant are shown in light grey. Figure taken from (Maze et al., 2014) and modified slightly.

In addition, tissue-specific (e.g. testis) expression, localization within the chromatin landscape (e.g. centromeres, telomeres or promoters) guided by variant-specific histone chaperone/ remodeling complexes shape the functional outcome of histone variants. Variant incorporation can modify nucleosome structure and charge and/or offer the establishment of distinct PTMs that recruit specific interacting factors. Furthermore, the variant itself in its specific nucleosomal context might attract different effector molecules than its canonical sibling. In higher eukaryotes, replacement variants for all canonical histone variants are known except for H4 (Malik and Henikoff, 2003). In the following section I introduce some of these histone variants, referring to their human counterparts unless otherwise stated.

1.2.4.1 H1 variants

H1 variants compose the largest group of histone variants (seven somatic and three germ-line specific variants) but appear to be highly redundant (Bustin et al., 2005). Besides their important role in chromatin compaction, H1 variants have been identified as transcriptional regulators localizing to active and repressed loci (Izzo et al., 2013) and are associated with haematological malignancies when mutated (Maze et al., 2014).

1.2.4.2 H2B variants

Currently, three H2B variants are known and their functions remain largely elusive (Bernstein and Hake, 2006), which might be a consequence of their peculiar expression in (primate) testis and brain. The human testis-specific H2B variant hTSH2B is involved in spermatogenesis where it destabilizes the nucleosome and supports the general chromatin reorganization occurring during chromatin-to-nucleoprotamine transition in the maturing sperm (Maze et al., 2014). The primate-specific H2BFWT shares only 45% identity with canonical H2B and is also expressed in testis and potentially involved in telomere formation and transmission (Bernstein and Hake, 2006). A fascinating third variant - H2BE - was just recently discovered that is exclusively expressed in olfactory chemosensory neurons. Here, it is only expressed and incorporated into chromatin when sensory experience is reduced and thereupon it mediates neuronal cell death (Santoro and Dulac, 2012). This is the first discovery of histone variant regulation in post-mitotic cells controlled by an activity-dependent mechanism.

1.2.4.3 H3 variants

To date, eight human histone H3 variants are known: the canonical H3.1 and H3.2 and the replacements variants H3.3, H3.4, H3.5, H3.X, H3.Y and CENP-A (Talbert and Henikoff, 2010). H3.4 and H3.5 are testis-specific variants and their function in sperm development is poorly understood (Maze et al., 2014). The remaining H3 variants are described in the following paragraphs.

H3.3 is encoded by two genes (*H3F3A* and *H3F3B*) in vertebrates. It differs from its canonical counterparts H3.1 and H3.2 in five and four amino acids, respectively (Ederveen et al., 2011), which are proposed to be responsible for specific H3.3 functionality. Indeed, the patch comprising of A87, I89 and G90 is involved in coupling H3.3 with distinct chaperone/remodeling machines that mediate its incorporation at specific target sites. Whereas H3.1 and H3.2 incorporation is dependent on the histone chaperone CAF1 (Latreille

et al., 2014; Verreault et al., 1996), H3.3 is deposited on the one hand by the chaperone complex HIRA at promoters and bodies of active genes (Goldberg et al., 2010; Tagami et al., 2004) and on the other hand by the chaperone Death domain-associated protein, DAXX, (together with chromatin remodeler α-Thalassemia/mental retardation X linked, ATRX) at telomeres of mouse embryonic stem (mES) cells and in pericentric heterochromatin (Drane et al., 2010; Goldberg et al., 2010). Another unique amino acid is S31 in the N-terminal tail of H3.3 that, in its phosphorylated state, is found close to centromeres in metaphase HeLa cells (Hake et al., 2005) suggesting an involvement in mitosis. Interesting is H3.3's partnership with H2A.Z in the same nucleosome (see below), rendering the nucleosome unstable (Jin et al., 2009), potentially supporting active transcription. Its role as transcriptional regulator is of special interest when deregulated, for example overexpressed in cancer (Vardabasso et al., 2014). Remarkably, recent studies uncovered mutations in *H3F3A* in glioblastoma, which is the first evidence for mutated histone-variant genes in cancer (Schwartzentruber et al., 2012).

CENP-A, histone H3-like centromeric protein A, is an essential H3 variant that localizes to eukaryotic centromeres, where it not only forms and epigenetically maintains the centromere but also serves as docking station for kinetochore assembly during mitosis (Black and Cleveland, 2011). It is 60% identical to the HFD of canonical H3 but the global structure of nucleosomes containing CENP-A is still quite similar to the structure of conventional nucleosomes (Tachiwana et al., 2011). Several competing models have been proposed regarding the composition of centromeric nucleosomes (Black and Cleveland, 2011) ranging from conventional octamers with two copies of CENP-A (homotypic nucleosomes) over a tetrasome lacking H2A-H2B dimers to a hemisome with only one copy of each histone. However, also dynamic transitions between octamers in S-phase and hemisomes in other cell cycle stages have been reported and recent studies point more and more towards an octameric structure (Fukagawa and Earnshaw, 2014). CENP-A is deposited replication-independently in late mitosis/early G_1 and guided in mammals by the Mis18 complex and the chaperone HJURP (Fukagawa and Earnshaw, 2014). Given their importance in proper chromosome segregation during mitosis it is not surprising that CENP-A and HJURP are overexpressed in several cancers suggesting a contribution to aneuploidy and chromosomal instability in theses cells (Vardabasso et al., 2014).

H3.Y, a primate-specific histone H3 variant was discovered in our lab together with H3.X (Wiedemann et al., 2010). H3.Y is expressed in a neuronal subpopulation in the human hippocampus, localizes to euchromatic regions in the nucleus and its depletion in tissue culture cells leads to proliferation inhibition.

Excitingly, its usually low expression can be increased by nutritional stress in osteosarcoma derived cancer cells.

1.2.4.4 H2A variants

Among the core histones, most variation is found within the H2A family. Whereas most of the isoforms belong to the group of canonical histones and possess only minor amino acid changes, several functionally specialized variants have evolved that differ especially in their C-termini (Ausio and Abbott, 2002): the universal variants H2A.Z and H2A.X, as well as H2A.Bbd and macroH2A.

H2A.X was discovered in the 1980's together with H2A.Z (West and Bonner, 1980), is present in all eukaryotes and incorporated throughout the genome in a replication-independent manner with the deposition machinery still unclear (Vardabasso et al., 2014). Upon DNA damage, its unique serine 139 gets phosphorylated and foci of phosphorylated H2A.X (γH2A.X) form around sites of DNA double strand breaks. It is still unclear whether this rapid phosphorylation has an influence on the stability of the damaged chromatin, but it appears essential for the retention of the DNA damage repair (DDR) machinery, in particular the associated chromatin remodelers (Pinto and Flaus, 2010). With its role in DDR in mind it is not surprising that H2A.X mutations or deletions are found in tumors showing chromosomal instability (Vardabasso et al., 2014) and that mice lacking H2A.X are impaired in class switch recombination (Petersen et al., 2001).

macroH2A (mH2A), a vertebrate-specific H2A variant, has three isoforms and is encoded in two independent genes in mammals: *H2AFY* encoding mH2A.1 and *H2AFY2* encoding mH2A.2 (Chadwick and Willard, 2001a). The first gene can be alternatively spliced, giving rise to mH2A.1.1 and mH2A.1.2 (Rasmussen et al., 1999). The mH2As are outstanding histone variants because they possess an exceptional structure: the N-terminal histone fold domain, which shares 65% of sequence identity with H2A, is connected by a linker sequence with a large (30 kDa) and highly conserved non-histone macro domain (Pehrson and Fried, 1992). The macro domain appears to protrude from the nucleosomal core serving either as binding platform for distinct protein interactions or it might inhibit putative interactions in a steric manner (Buschbeck and Di Croce, 2010). One fascinating interaction is mediated by the macro domain of mH2A.1 only: it binds nicotinamide adenine dinucleotide (NAD) metabolites (like poly(ADP)ribose) thereby, for example, recruiting poly-ADP-ribose polymerase 1 (PARP1) and potentially other ADP ribosylated proteins like sirtuins that could in turn influence chromatin structure (Posavec et al., 2013). On the other hand, transcription

factor binding has been reported to be inhibited by mH2A, and also SWI/SNF-mediated remodeling is less effective, both arguing for a role of mH2A in transcriptional repression (Angelov et al., 2003). Concordantly, mH2A.1 was identified to coat the inactive X chromosome where it contributes to the transcriptional silencing of the X chromosome in female mammalian cells (Costanzi and Pehrson, 1998). Moreover, mH2A was shown to act as a tumor suppressor in malignant melanoma, where it inhibits tumor progression by suppressing the oncogene CDK8 (Kapoor et al., 2010). The mode of mH2A incorporation into chromatin remains elusive but recently ATRX, a SWI/SNF helicase, was reported as negative regulator of mH2A deposition (Ratnakumar et al., 2012).

H2A.Bbd is a mammalian histone H2A variant that is only 50% identical to canonical H2A, expressed almost exclusively in testis and excluded from the X-chromosome (Bbd = Barr-body deficient) (Chadwick and Willard, 2001b; Eirin-Lopez et al., 2008). Strikingly, it has a shorter primary sequence lacking not only an acidic patch (needed for internucleosomal contacts) but also a part of the docking domain (H3-H4)$_2$-tetramer contacts). Lack of this structural features leads to a reduction of nucleosome stability and a generally more open chromatin structure (Bonisch and Hake, 2012). This is consistent with the finding that H2A.Bbd localizes to transcription start sites (TSS) and gene bodies of active genes and promotes transcription (Tolstorukov et al., 2012). Interestingly, H2A.Bbd is overexpressed in Hodgkin's lymphoma and was suggested to drive proliferation in these cancer cells by elevating S-phase progression (Sansoni et al., 2014).

Structure and function of H2A.Z is fundamental to this thesis and will be discussed in detail in the next section.

1.3 The histone variant H2A.Z

H2A.Z has been subject of extensive studies since its first description in 1980 (West and Bonner, 1980). It is highly conserved in eukaryotes (~90% between different species) and only ~60 % identical to canonical H2A arguing for its functional specialization within the H2A family (Zlatanova and Thakar, 2008) (**Figure 1.3.1**). In fact, H2A.Z is essential in many organisms like *D. melanogaster* (fruit fly), *M. musculus* (mouse), *T. thermophile* (tetrahymena) and *X. laevis* (frog) (Faast et al., 2001; Iouzalen et al., 1996; Liu et al., 1996; van Daal and Elgin, 1992) but not in *S. cerevisiae* (budding yeast) and *S. pombe* (fission yeast) that show impaired growth upon deletion (Carr et al., 1994; Jackson and Gorovsky, 2000).

```
                10        20        30        40        50        60        70        80
H2A      --SGRGKQGGKARAKAKTRSSRAGLQFPVGRVHRLLRKG-NYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTRI
H2A.Z.1  AGGKAGKDSGKAKTKAVSRSQRAGLQFPVGRIHRHLKSRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLKVKRI
          . **:.***:;** :**.**********:** *:.  .    **** * ** **:;*******;*******;:* * .**

                90       100       110       120       130
H2A      IPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK
H2A.Z.1  TPRHLQLAIRGDEELDSLI-KATIAGGGVIPHIHKSLIGKKGQQKTV----
          ********* ****:.*: *.*** ***:*.*:  *: ** :.:
```

Figure 1.3.1 H2A and H2A.Z protein primary structure. Alignment of human histone H2A with human histone H2A.Z.1 protein sequences. Important structural features are colored according to Figure 1.1.1: the acidic patch in cyan, the L1-L1 interface in magenta and the docking domain in orange. Consensus symbols: fully conserved residue (*), conserved strongly similar residue (.) and conserved weakly similar residue (:). Alignments were carried out using the web-browser based Clustal Omega tool. Figure adapted from (Bonisch and Hake, 2012).

In vertebrates, two non-allelic genes encode for H2A.Z proteins: *H2AFZ* (H2A.Z.1) and *H2AFV* (H2A.Z.2), which both contain introns and give rise to mRNAs with poly(A)-tail (Dryhurst et al., 2009). Interestingly, while the nucleotide sequence of both genes is unique and allows independent targeting by RNA interference (RNAi), on the protein level H2A.Z.1 and H2A.Z.2 differ in only 3 amino acids (Coon et al., 2005; Eirin-Lopez et al., 2009) (**Figure 1.3.2**). This nearly 100% identity has so far impeded the characterization of both variants separately, at least by employing antibody-based assays. The primary RNA transcript of *H2AFV* is subject to alternative splicing leading to the expression of two proteins: H2A.Z.2.1 (the formerly mentioned H2A.Z.2 protein) and the primate-specific H2A.Z.2.2 variant (Bonisch et al., 2012).

```
            ●  ●   ●
                10        20        30        40        50        60        70        80
H2A.Z.1    AGGKAGKDSGKAKTKAVSRSQRAGLQFPVGRIHRHLKSRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLKVKR
H2A.Z.2.1  AGGKAGKDSGKAKAKAVSRSQRAGLQFPVGRIHRHLKTRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLKVKR
H2A.Z.2.2  AGGKAGKDSGKAKAKAVSRSQRAGLQFPVGRIHRHLKTRTTSHGRVGATAAVYSAAILEYLTAEVLELAGNASKDLKVKR
           *************:*********************:*******************************************

                                                           ■■    ■
                90       100       110       120
H2A.Z.1    ITPRHLQLAIRGDEELDSLIKATIAGGGVIPHIHKSLIGKKGQQKTV
H2A.Z.2.1  ITPRHLQLAIRGDEELDSLIKATIAGGGVIPHIHKSLIGKKGQQKTA
H2A.Z.2.2  ITPRHLQLAIRGDEELDSLIKATIAGGEKRRCS--------------
           ***************************
```

Figure 1.3.2 H2A.Z isoform primary structures. Alignment of human histone H2A variant H2A.Z.1 with H2A.Z.2.1 and H2A.Z.2.2 protein sequences. Important structural features are colored according to Figure 1.1.1: the acidic patch in cyan and the docking domain in orange. Amino acid differences between H2A.Z.1 and H2A.Z.2.1 (T14A, S38T and V127A) are highlighted in red. Black circles highlight acetylated residues and black squares ubiquitinated residues. Consensus symbols: fully conserved residue (*) and conserved weakly similar residue (:). Alignments were carried out using the web-browser based Clustal Omega tool. Figure adapted from (Bonisch and Hake, 2012).

In 2000, the crystal structure of a nucleosome containing H2A.Z was solved and revealed high similarity to the canonical nucleosome despite the existing sequence difference between

H2A.Z and H2A (Suto et al., 2000). However, the stability of the H2A.Z-containing nucleosome has been addressed in many studies reporting both stabilizing but also destabilizing effects (Bonisch and Hake, 2012; Zlatanova and Thakar, 2008). Two structural features appear to be different between H2A and H2A.Z potentially leading to altered nucleosome stability: the L1 region and the docking domain. Indeed, Suto et al. postulated that a combination of one H2A and one H2A.Z chain within the same nucleosome (heterotypic nucleosome) would lead to a steric clash due to an incompatibility of their L1 loops and therefore could not exist *in vivo*. Although homotypic H2A.Z nucleosomes seem to be more stable according to the crystal structure and were reported to occupy the transcription start site (TSS) of active genes in *D. melanogaster* (Weber et al., 2010), the existence of heterotypic H2A.Z-H2A nucleosomes was shown *in vitro* and *in vivo* in different model organisms (Chakravarthy et al., 2004; Luk et al., 2010; Viens et al., 2006; Weber et al., 2010). Another evidence for a putative destabilized H2A.Z nucleosome is based on the altered C-terminal docking domain that loses hydrogen bonds with H3 thereby causing subtle destabilization (Suto et al., 2000). Interestingly, the docking domain of the alternatively spliced H2A.Z variant H2A.Z.2.2, which differs from the H2A.Z.1 / H2A.Z.2.1 docking domain in 6 amino acids at the uttermost C-terminus, is more flexible then the H2A.Z.2.1 C-terminus, thereby reducing interactions with the $(H3-H4)_2$-tetramer and rendering the nucleosome highly instable (Bonisch et al., 2012).

The destabilizing role of H2A.Z is also recapitulated when it teams up with the histone H3 variant H3.3 in the same nucleosome (Jin and Felsenfeld, 2007). Nucleosomes prepared from native chromatin containing both variants were shown to be less stable than nucleosomes containing only H3.3 under physiological salt conditions (Henikoff, 2009). Moreover, genome-wide assays in human HeLa cells employing low salt conditions mapped H3.3-H2A.Z nucleosomes to nucleosome depleted regions (NDRs) of active promoters and to insulator regions (CTCF binding sites) that are depleted of H2A.Z under physiological conditions (Jin et al., 2009). This suggests that double-variant nucleosomes at least transiently occupy regulatory sites that were regarded as nucleosomes free. What exactly renders these nucleosomes unstable remains unclear (Henikoff, 2009) and is also challenged by the results of *in vitro* studies that could not detect drastic stability changes (Thakar et al., 2009). Apparently contradictory to its potential destabilizing character on the level of mononucleosomes, H2A.Z incorporation has been implicated in higher order chromatin structure formation (Chen et al., 2013). Compared to H2A, H2A.Z possesses an extended acidic patch that has a higher affinity to the H4 tail then H2A (Fan et al., 2004). This mediates

interaction between neighboring nucleosomes in one chromatin fiber and leads to its compaction (intrafiber folding) and favored formation of a stable secondary structure (Bonisch and Hake, 2012). In addition, recruitment of HP1α to H2A.Z-chromatin secondary structures enhances intrafiber folding and could also be of relevance for H2A.Z presence in pericentric heterochromatin and centric chromatin (Greaves et al., 2007).

H2A.Z is subject of posttranslational modification with different functional consequences (Sevilla and Binda, 2014; Thambirajah et al., 2009). C-terminally ubiquitinated H2A.Z colocalizes with the inactive X-chromosome in female mammals (Sarcinella et al., 2007), sumoylation of H2A.Z was reported to be involved in DNA repair in *S. cerevisiae* (Kalocsay et al., 2009) and monomethylation of H2A.ZK7 by SETD6 was suggested to be involved in transcriptional repression in mouse ES cells (Binda et al., 2013). Several lysine residues in the N-terminal tail of H2A.Z can be acetylated (Billon and Cote, 2013). Acetylated H2A.Z (H2A.Zac) is generally associated with transcriptionally active chromatin in different organisms (Bruce et al., 2005; Millar et al., 2006; Ren and Gorovsky, 2001) where it localizes to the TSS of active genes (Bellucci et al., 2013; Valdes-Mora et al., 2012). Together with its potential ability to destabilize nucleosomes (Thambirajah et al., 2006; Tolstorukov et al., 2009) dynamic acetylation is proposed to act as a switch-like mechanism associating H2A.Zac with activating (destabilizing) and unmodified H2A.Z with repressive (stabilizing) function (Thambirajah et al., 2009).

H2A.Z genome-wide dynamics are dependent on its deposition or eviction by multifactor ATP-dependent chromatin remodeling complexes (Billon and Cote, 2013). In *S. cerevisae*, a nucleosomal H2A-H2B dimer is exchanged for a free H2A.Z-H2B dimer by the SWR-1 complex (Mizuguchi et al., 2004). Targeting to sites of H2A.Z incorporation depends on a certain DNA sequence (Raisner et al., 2005), nucleosome acetylation (Altaf et al., 2010; Watanabe et al., 2013) or the presence of the NDR (Ranjan et al., 2013; Yen et al., 2013). In higher eukaryotes, H2A.Z incorporation is achieved by two related complexes, p400/NuA4/TIP60 and SRCAP (Billon and Cote, 2013), that are presumably working context dependent. Removal of H2A.Z is facilitated by the INO80 complex in *S. cerevisiae* (Papamichos-Chronakis et al., 2011), a complex including the histone chaperone NAP1 and members of the SWI/SNF and INO80 family in mouse embryonic stem cells (Li et al., 2012) and the histone chaperone ANP32E in concert with p400/NuA4/TIP60 at enhancers and insulators in human cells (Obri et al., 2014).

H2A.Z incorporation has been implicated in a plethora of biological processes like transcription regulation, DNA repair, genome stability, cell cycle progression,

heterochromatin formation, telomere integrity, chromosome segregation, mitosis and recently also in brain memory formation (Bonisch and Hake, 2012; Marques et al., 2010; Thambirajah et al., 2009; Vardabasso et al., 2014; Zlatanova and Thakar, 2008; Zovkic et al., 2014). However, the role of H2A.Z in transcription regulation has been in focus since Allis et al. initially suggested it in 1986 (Allis et al., 1986). In the last decade the invention of genome-wide assays (ChIP-chip and ChIP-seq) allowed to further dissect H2A.Z's role in transcription. Employing these techniques it was shown, that H2A.Z is enriched around the TSS at gene promoters in yeast (Albert et al., 2007; Raisner et al., 2005), human (Barski et al., 2007; Schones et al., 2008), mouse (Conerly et al., 2010a), fly (Schauer et al., 2013) and plants (Zilberman et al., 2008), as well as on other regulatory regions, such as enhancers and insulators (Zlatanova and Thakar, 2008). In yeast, human and mouse both nucleosomes flanking the NDR (the -1 and +1 nucleosomes) at the TSS contain H2A.Z (**Figure 1.3.3**). High-resolution subnucleosomal mapping achieved with ChIP-exo followed by sequencing (Rhee and Pugh, 2011) just recently revealed preferential localization to the NDR-distal half of the +1 nucleosome in *S. cerevisiae*, towards the direction of transcription (Rhee et al., 2014). Interestingly, in *S. cerevisiae* the presence of H2A.Z at promoters was reported to be not (Raisner et al., 2005) or inversely (Guillemette et al., 2005) correlated with transcription. In contrast, genome-wide studies in human T cells and fly embryos reported a positive correlation with transcription as H2A.Z localized to the TSS of active genes (Barski et al., 2007; Mavrich et al., 2008). A recent report suggests that H2A.Z is a common feature decorating the TSS of active and inactive genes across the human genome, where H2A.Z in active genes is acetylated (Valdes-Mora et al., 2012). As mentioned before, also the cooperative action with other histone variants within the same nucleosome such as H3.3 (Chen et al., 2013; Jin et al., 2009) or H2A.Bbd (Soboleva et al., 2012) argue for a participation in gene activation. Moreover, whereas the +1 nucleosome is thought to act as a barrier for RNA polymerase II (RNAPII) (Nock et al., 2012), this barrier function is modulated by the presence of H2A.Z that renders the +1 nucleosome more labile, thereby allowing RNAPII passage during transcription (Weber et al., 2014). Taken together, the impact on and final result of gene expression is not dictated by the mere deposition/eviction of H2A.Z alone, but rather needs the concerted action of H2A.Z, PTMs, other histone variants, transcription factors and the H2A.Z deposition machinery.
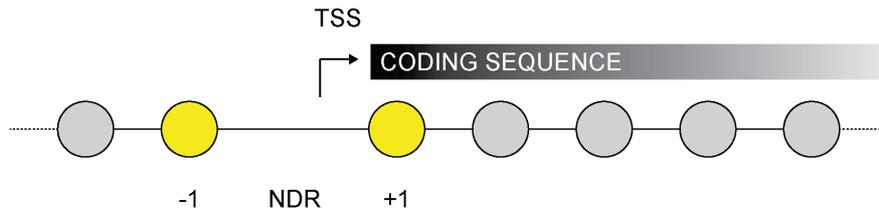
Figure 1.3.3 Nucleosome organization at gene promoters. H2A.Z localizes to the -1 and +1 nucleosomes that flank the nucleosome-depleted region (NDR) at the transcription start site (TSS) of genes. H2A.Z-containing nucleosomes are depicted in yellow, other nucleosomes in grey.

H2A.Z is also reported to regulate DNA replication in *S. cerevisiae*, where it is thought to influence the expression of genes implicated in G_1/S transition (Dhillon et al., 2006). It localizes to origins of replication (Oris) in *S. cerevisiae* (Dhillon et al., 2006) and *A. thaliana* (Costas et al., 2011) and recently also human origins were mapped to TSS of genes that are generally occupied by H2A.Z (Dellino et al., 2013).

Moreover, knockout studies in *S. cerevisiae* suggested, that loss of H2A.Z increases DNA damage, the sensitivity to DNA damage-inducing agents during replication (Billon and Cote, 2013) and genomic instability (Morillo-Huesca et al., 2010). DNA double strand break (DSB) repair by homologous recombination involves sumoylated H2A.Z at the DNA brakeage site (Kalocsay et al., 2009) in this model organism. As an early reaction in DNA damage response in human cells, p400/NuA4/TIP60 deposits H2A.Z not only to support an open chromatin conformation upon DSB occurrence but also the recruitment of members of the DSB repair machinery (Xu et al., 2012). Notably, H2A.Z.2 was recently reported to be important for the reorganization of damaged chromatin and RAD51 foci formation at sites of DSBs in DT40 cells (Nishibuchi et al., 2014).

A role for H2A.Z in cell cycle progression includes beside its association with S-phase regulation also an involvement in mitosis and especially in chromatin condensation and segregation. The condensation of chromatin and the formation of chromosomes highly depends on the presence of the structural maintenance of chromosomes (SMC) complexes, which include cohesin, condensin and Smc5/6 in eukaryotes (Jeppsson et al., 2014). Cohesin topologically embraces and thereby links sister chromatids. Starting already dynamically in telophase of the previous mitosis, the whole chromosome gets stably decorated with cohesin during replication. During prophase, cohesin is removed from chromosome arms and just remains at centromeres, where it is cleaved during sister chromatid separation in anaphase. Chromatin association of condensin peaks in prophase of mitosis and mediates further chromosome compaction but is also important for disentanglement of sister chromatids during

separation in anaphase. Smc5/6 generally mimics the chromosomal association of cohesion but its actual role in chromosome condensation is less defined (Jeppsson et al., 2014). H2A.Z has been shown to interact with condensin in *S. pombe* (Kim et al., 2009). Together with H2A, other factors and the chromosomal passenger kinase Aurora B that phosphorylates condensin, H2A.Z assists in loading condensin onto chromatin during mitosis (Tada et al., 2011). This study also revealed an interaction of human H2A.Z with condensin. H2A.Z has also been implicated in cohesion dynamics in yeast. In *S. cerevisiae*, lack of H2A.Z leads to defects in sister chromatid cohesion, but the exact involvement of H2A.Z is unclear, however independent of cohesin (Sharma et al., 2013). A function for H2A.Z in cohesion has also been reported in *S. pombe* (Tapia-Alveal et al., 2014), where it is required to keep cohesin at chromosome arms. Chromosome segregation defects in cells lacking H2A.Z have been reported early in fission yeast (Carr et al., 1994) but also in human cells (Rangasamy et al., 2004). Here, in addition to potential interconnections with the SMC machinery, also H2A.Z's presence in pericentric (Boyarchuk et al., 2014; Rangasamy et al., 2003) and centric (Greaves et al., 2007) heterochromatin could by crucial as already mentioned before.

H2A.Z is overexpressed in many cancers like bladder, breast, colorectal and lung cancers but also in malignant melanoma and also the H2A.Z deposition machinery is implicated in cancer (Vardabasso et al., 2014). A direct involvement of H2A.Z in cancer promotion was reported in hormone-dependent cancers like breast (Gevry et al., 2009; Svotelis et al., 2010) or prostate cancer (Dryhurst et al., 2012; Valdes-Mora et al., 2012), often including the oncogene Myc.

The discovery that two highly similar H2A.Z isoforms exist in vertebrates raised the question whether H2A.Z.1 and H2A.Z.2 are redundant or possess distinct functionality. In fact, both variants can by acetylated at the same N-teminal lysine residues, are nearly equally expressed in different tissues, show euchromatic nuclear localization and similar Fluorescence Recovery After Photobleaching (FRAP) mobilities (Bonisch et al., 2012; Dryhurst et al., 2009). However, the analysis of the *H2AFZ* and *H2AFV* promoter sequence revealed significant differences in promoter structure arguing for distinct expression regulation (Dryhurst et al., 2009). Furthermore, the specific knockdown of H2A.Z.2 but not H2A.Z.1 in chicken DT40 cells resulted in BCL6 down regulation and increased apoptosis, suggesting a functional specialization of both variants that is supported by H2A.Z.1 knockdown studies in mice (Faast et al., 2001). How could a potential functional specialization be achieved? First, each H2A.Z variant could possess its own histone chaperone/remodeling machinery that determines genomic localization of the individual variant, as it is the case for the H3 variants.

As already mentioned, two distinct complexes are known in mammals: SRCAP and p400/NuA4/TIP60. However, data from our lab and others show, that reduction of SRCAP activity influences the deposition of both variants (Matsuda et al., 2010) and SRCAP and p400/NuA4/TIP60 complex members are enriched on all chromatin-free H2A.Z variants as identified by SILAC quantitative MS (unpublished data and (Bonisch et al., 2012)). Second, each variant could independently occupy distinct genomic positions, a situation which is in fact true for the H3 variants. Several studies have determined genome-wide binding pattern of H2A.Z in different organisms (Vardabasso et al., 2014). However, the employed ChIP-techniques relied on antibodies that were not able to discriminate both variants. Third, nucleosomes containing H2A.Z.1 or H2A.Z.2 could interact with a different set of effector proteins or could be decorated with variable PTMs, both potentially mediating distinct downstream functionality. Recently, BRD2 was identified to preferentially interact with H2A.Z.1-containing nucleosomes and mediate activation of downstream gene expression (Draker et al., 2012). Draker et al. showed, that BRD2 is recruited to androgen receptor (AR) regulated genes in dependence of H2A.Z.1 and hyperacetylated H4 to promote AR-regulated gene expression. This was, irrespective of the interaction with the H2A.Z deposition machinery, the first discovery of a protein 'reading' H2A.Z chromatin in a variant-specific manner thereby influencing transcription.

1.4 Objectives

In this thesis I aimed to establish a method to quantitatively identify the interactome of human nucleosomes containing H2A.Z variants. Therefore, I applied label-free interaction proteomics previously employed for modified histone tail peptides (Eberl et al., 2013) for the first time on a nucleosomal template. Furthermore, I aimed to investigate, whether nucleosomes containing the different H2A.Z isoforms, H2A.Z.1 or H2A.Z.2, attract distinct interaction partners. The outcome of one such interaction was further elaborated in collaboration with Chiara Vardabasso (PhD), a PostDoc in Prof. Emily Bernstein's group at the Mount Sinai Hospital in New York, focusing on the role of the H2A.Z nucleosome binder BRD2 in H2A.Z.2-mediated malignant melanoma progression. Moreover, I aimed to unravel the functional role of a second H2A.Z nucleosome interactor, PWWP2A, in HeLa Kyoto cells. Therefore, I analyzed how PWWP2A binds specifically to H2A.Z nucleosomes, discovered its genomic localization also in correlation with H2A.Z isoform localization and aimed to uncover its cellular function through knockdown studies.

2   Materials and Methods

2.1   Materials

2.1.1   Technical devices

Description	Supplier
-20°C Freezer	Bosch, Liebherr
-80°C Freezer	Thermo Scientific
4°C Fridge	Siemens, Liebherr
37°C Incubator (bacteria)	Memmert
37°C Incubator (mammalian cells)	Heraeus
Agarose gel chamber	Repair shop of Adolf-Butenandt-Institute
Autoclave (Varioklav)	H+P
2100 Bioanalyzer	Agilent
CASY Cell Counter	Innovatis
Centrifuges	Eppendorf 5417R
	Eppendorf 5515D
	Eppendorf 5810R
	Heraeus Biofuge pico
	Heraeus Cryofuge 6000i
	Heraeus Megafuge 2.0
	Hettich Rotina 46
	Sorvall RC6 Plus (SS-34 rotor)
Developer machine Curix 60	AGFA
Gel documentation system	Peqlab
Hood	CleanAir
Incubation shaker (Multitron)	Infors
LightCycler 480 II	Roche
Magnetical particle concentrator	Dynal
Microscopes	Leica DMIL LED
	Leica SP5 II confocal scanning
	PerkinElmer UltraVIEW VoX spinning disc
	Zeiss Axiovert 200M epifluorescence
Microwave	Severin

MilliQ-water	Millipore
pH 720 pH-meter	inoLab
PhosphoImager FLA3000	Fuji
Pipetboy	Integra Biosciences
Pipettes	Gilson
Power supply	BioRad
Proteingel chamber (Novex Mini Cell)	Invitrogen
Repetman Multistep pipet	Gilson
Rotating wheel	Neolab, VWR
Q Exactive	Thermo Scientific
Qubit fluorometer	Invitrogen
RTCA DP Analyzer xCELLigence	Roche
Scales	Sartorius
Shaker	Roth
Sonifier	Branson MD-250
	Diagenode Bioruptor
Spectrophotometer	Peqlab Nanodrop ND1000
	Pharmacia Biotech Ultrospec 2000
Thermomixer	Eppendorf 5436
	Eppendorf compact
Trans Blot SD Semi-dry transfer cell	BioRad
UV Stratalinker 1800	Stratagene
Vortex Genie 2	Bachofer
Water bath	Memmert

2.1.2   Chemicals and consumables

Unless stated otherwise, all common chemicals are purchased in analytical grade from Merck.

Description	Supplier
1.5 ml and 2 ml reaction tubes	Greiner, Sarstedt
1.5 ml low binding tubes	Sarstedt
15 ml and 50ml tubes	Sarstedt
Acetic acid	Sigma
Agarose	Bio & Sell
Ampicillin	Roth

AMPure XP beads	Beckman Coulter
LB Agar	Serva
Cling film	Saran
BSA 98%	Sigma
2-Chloroacetamide	Sigma
Chlorophorm	VWR
Cellculture plates	Sarstedt
Combitips plus	Eppendorf
Complete Protease Inhibitor Cocktail Tablets	Roche
Coomassie Brilliant Blue	Sigma
Cover slips (round, 12 mm/15 mm)	Hecht-Assistent
Cryovials	Roth
DAPI	Invitrogen
Developer	Agfa
DMSO	Sigma
DNA oligonucleotides	Sigma
DNA Chips	Agilent
dNTP mix	NEB
DTT	Roth
DMEM	Sigma
ECL Western Blotting Detection Reagents	Amersham
EDTA	Sigma
EGTA	Sigma
Ethanol, absolute	VR
Ethidium bromide	Sigma
FCS dialyzed	Sigma
Filter paper Whatman 3MM	Whatman
Filter tips	Biozym, Gilson
Fixer	AGFA
x-tremeGENE HP Transfection Reagent	Roche
G418-sulfate	Sigma
GFP-trap_M	Chromotek
Glass pipettes 5 ml and 10 ml	Hirschmann

Glassware	Schott
Glutathione sepharose beads	GE
Glycerol	VWR
IPTG	Roth
Isoamyl alcohol	Merck
Laboratory film	Parafilm
LightCycler R 480 Multiwell Plate 384, white	Sarstedt
LightCycler R 480 Sealing Foil	Sarstedt
Magna ChIP Protein G magnetic beads	Millipore
MaXtract High Density column	Qiagen
beta-Mercaptoethanol	Sigma
Methanol	Sigma
Microscope slides SuperFrost	Roth
Multiply -µStripPro with 8 x 0.2ml tubes	Sarstedt
1 ml NORM-JECT syringe	Henke Sass Wolf
NP-40	Sigma
Oligofectamine Transfection Reagent	Invitrogen
Opti-MEM Reduced-Serum Medium	Invitrogen
Pasteur pipettes	Brand
Penicillin / streptomycin	Sigma
Petridishes	Greiner
Pipette tips	Biozym, Greiner, Sarstedt
PMSF	Sigma
Ponceau S solution	Sigma
Propidium iodide	Sigma
Protein gel cassettes (disposable)	Invitrogen
Protein gels precast	Serva
Protran Nitrocellulose Transfer Membrane	Whatman
Qubit assay tubes	Invitrogen
Rotiphorese Acrylamide/bisacrylamide mix	Roth
RTCA E-Plate 16	Roche
SDS	Serva
siRNAs	Eurofins MWG Operon

Sterican needles	Braun
TEMED	Roth
TFA	Sigma
TSA	Sigma
Tris	Invitrogen
Triton X-100	Sigma
Trypsin/EDTA (cellculture)	Sigma
Trypsin (mass spec)	Promega
Tween 20	Sigma
Vectashield Mounting Medium	Vector Laboratories
Water, PCR-grade	Roche
X-ray films	Fujifilm

2.1.3   Kits, enzymes and markes

Description	Supplier
100bp DNA marker	NEB
1kb DNA marker	NEB
Annexin V FITC Detection Kit	eBioscience
DNA 1000 Kit	Agilent
Gel extraction Kit	Qiagen
Fast SYBR Green Master Mix	Applied Biosystems
Maxi- and Midiprep Kit	Qiagen, Promega
Micrococcal nuclease	Sigma
MicroPlex Library Preparation Kit	Diagenode
PCR-purification Kit	Qiagen, Sigma
peqGOLD Protein Marker IV, V	Peqlab
Phusion R DNA Polymerase	Finnzymes
Pfu Turbo DNA Polymerase	Stratagene/Agilent
ProtoScript First Strand cDNA Synthesis Kit	NEB
Qubit dsDNA HS Assay Kit	Invitrogen
Restriction endonucleases	NEB
RNase-Free DNase Set	Qiagen
RNeasy Kit	Qiagen
RNeasy MinElute Cleanup Kit	Qiagen

Taq DNA Polymerase   NEB

2.1.4   Antibodies

2.1.4.1   Primary antibodies

Name (product#)	Supplier	Application	Dilution
goat αGFP	A. Ladurner, LMU Munich	ChIP	2 µl of 'red' purification
mouse αGFP (11814460001)	Roche	WB	1:10000
rabbit αZNHIT1 (HPA01904)	Sigma	WB	1:1000
rabbit αH3K9me3 (39162)	Active motif	WB	1:1000
rabbit αH3K4me3 (C15410003)	Diagenode	WB	1:1000
rabbit αH3K36me3 (61102)	Active motif	WB	1:1000
rabbit αPWWP2A (NBP2-13833)	Novus Biologicals	WB IF	1:1000 1:100
rabbit αBRD2 (A302-583A)	Bethyl Laboratories	WB	1:1000
rabbit αH3K27ac (39134)	Active motif	WB	1:1000
rabbit αH3 (ab1791)	Abcam	WB	1:10000
rabbit αH2A (ab13923)	Abcam	WB	1:1000
rabbit αH2A.Z (ab4174)	Abcam	WB	1:3000
rat αGST (Klon 6c9 R-2A)	E. Kremmer, Munich	WB	1:50
rabbit αH3S10phos (39636)	Active motif	FACS IF	1:200 1:1000

2.1.4.2 Secondary antibodies

Name	Supplier	Application	Dilution
α rabbit HRP	GE	WB	1:10000
α mouse HRP	GE	WB	1:10000
α rat HRP	GE	WB	1:10000
α rabbit Alexa 488	Jackson	IF	1:1000

2.1.5 Plasmids

Name	Source	Description	Marker
pT7Blue-3	Novagen	Subcloning	Amp, Kan
pIRESneo-eGFP	C. Bönisch	Expression of N-terminally GFP-tagged fusion proteins in mammalian cells	Amp, Neo
pIRESneo-eGFP -PWWP2A_fl -PWWP2A_CT -PWWP2A_NT -PWWP2A_I -PWWP2A_I_S_PWWP	this thesis	Expression of N-terminally GFP-tagged PWWP2A constructs in mammalian cells	Amp, Neo
pGEX6P1-GST	Amersham	Expression of N-terminally GST-tagged fusion proteins in *E. coli* cells	Amp
pGEX6P1-GST -PWWP2A_fl -PWWP2A_NT -PWWP2A_CT -PWWP2A_P1_P2_I -PWWP2A_P1 -PWWP2A_P2_I_S_PWWP -PWWP2A_P2_I_S -PWWP2A_P2_I -PWWP2A_I_S_PWWP -PWWP2A_I_S	this thesis	Expression of N-terminally GST-tagged PWWPA constructs in *E. coli* cells	Amp

-PWWP2A_I

-PWWP2A_IN

-PWWP2A_IC

-PWWP2A_conI

pEGFP-C1-H2A.Z.1	(Bonisch et al.,	Expression of N-terminally GFP-tagged H2A.Z in mammalian cells	Kan, Neo
pEGFP-C1-H2A.Z.2	2012)		

2.1.6 Oligonucleotides

Oligonucleotide sequences are always depicted from 5' to 3' end.

2.1.6.1 Oligonucleotides for cloning

Name	Sequence	Description
PWWP2A Fwd#2	GGAGTTGGAGGAGGGAGAAG	Cloning of PWWP2A into pT7Blue-3
PWWP2A Rev#2	TTCCAATGGTCTTGCCTACC	

2.1.6.2 Oligonucleotides for qPCR

Table 2.1.1 qPCR

Name	Sequence	Efficiency	Description
HPRT1 Fwd	AAGGGTGTTTATTCCTCATGGA	2.04	Reference gene
HPRT1 Rev	AATCCAGCAGGTCAGCAAAG		
HMBS Fwd	AGTGTGGTGGCAACATTGAA	1.93	Reference gene
HMBS Rev	GCATGTTCAAGCTCCTTGGT		
PWWP2A_all Fwd1	ACGGTGTCGCAACTGATCC	1.93	PWWP2A
PWWP2A_all Rev1	CCATGGGGCCCAAACCTTTT		transcripts a,b,c
PWWP2A_2 Fwd1	GAAGTGCGGGCTTTGTTGAC	1.94	PWWP2A
PWWP2A_2 Rev1	CTCCAATCTGGCCACGCTAT		transcript b
H2A.Z-FWD	GGCAGGAAATGCATCAAAAG	1.87	H2A.Z.1
H2A.Z-REV	TGGATGTGTGGAATGACACC		transcript
H2AFV-1&2_FOR	GAGCTGGCAGGTAATGCTTC	1.83	H2A.Z.2.1
H2AFV-Var1_spez_REV	TTTGTGGATGTGAGGGATCA		transcript

Table 2.1.2 ChIP-qPCR. gb, gene body; **+**, +1 nucleosome; **-**, -1 nucleosome.

Name	Sequence	Efficiency	Position
RPL11_gb2 Fwd	ACAGCTTTGGGTGATGCAGT	2.04	24021666 - 24021685
RPL11_gb2 Rev	TTGTTGGACCAAAACACGGC		24021755 - 24021736
YIPF_gb2 Fwd	TCTAGCTGCCCCATGCTCTA	1.98	54339580 - 54339599
YIPF_gb2 Rev	GCATTGGGCCACATAGAGGT		54339704 - 54339685
PARS2+ Fwd	GGGATGCAAGTGGGAAAAC	1.90	55230014 - 55230032
PARS2+ Rev	ATTGCGGTAGGTGAACGTG		55230139 - 55230121
PARS2- Fwd	AGACGCCTTTATTACAGTGCCC	1.94	55230622 - 55230643
PARS2- Rev	TCTACGTGGTAGCAGCTCAAAA		55230712 - 55230691
RNF11+ Fwd	CCGCAGCCCGAGGAATATG	2.00	24018399 - 24018417
RNF11+ Rev	AAAACGATGGCAGGGAAAGTG		24018508 - 24018488
RNF11- Fwd	TTGGCGTTCTCATCACACAAG	1.91	24017997 - 24018017
RNF11- Rev	AAAGCTCCCGAATCAGCACG		24018096 - 24018077
UHMK1+ Fwd	GCTGGCGGAGATGTGACC	1.95	162467707 - 162467724
UHMK1+ Rev	CGCCATCGGTGTGGGTTAAG		162467796 - 162467777
UHMK1- Fwd	GGCCGGGTTTTATTTTTCGGGT	1.95	162467346 - 162467367
UHMK1- Rev	TACCATTCCAGGCCCAGTGTTA		162467466 - 162467445

NUF2+ Fwd	CACTGTAGGTGAGCGCGAGA	1.96	163291849 - 163291868
NUF2+ Rev	CGCTGAGCACGACGAAAACA		163291968 - 163291949
NUF2- Fwd	GCATCTAACAAAACCCGGCAC	1.98	163291342 - 163291362
NUF2- Rev	GTCCGAGTTGAAGAGCAAAACC		163291484 - 163291463
SWT1+ Fwd	CTCCTTTGGCTTGGGGCTC	1.94	185126397 - 185126415
SWT1+ Rev	GCCAGTATACTTGGGGCGG		185126518 - 185126500
SWT1- Fwd	GGTCAGAGAACTGACGTGAATG	1.97	185125900 - 185125921
SWT1- Rev	TACTAACAGGATTTGGTCACTGGT		185126020 - 185125997

2.1.6.3 Oligonucleotides for RNAi

Name (#number)	Sequence (sense)	Description
Luciferase	CUUACGCUGAGUACUUCGATT	Control siRNA
H2A.Z.1#2	ACUUAAAGGUAAAGCGUAUTT	siRNA targeting H2A.Z.1 mRNA
H2A.Z.2.1#3	GGAAAAGCAUAGACAAUUATT	siRNA targeting H2A.Z.2.1 mRNA
PWWP2A#1	GGACAGAAGUCAAGUGUGAUTT	siRNAs targeting
PWWP2A#2	GCUAUUAAACUACGACCCAUTT	PWWP2A mRNA
PWWP2A#3	GAAGACAGGACUUGAGAAAUTT	

2.1.7 Bacterial strains and cell lines

2.1.7.1 E. coli strains

Strain	Genotype	Supplier
DH5	F⁻ Φ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)$_{U169}$ *rec*A1 *end*A1 *hsd*R17 (r$_K$⁻, m$_K$⁺) *pho*A *sup*E44 λ⁻ thi-1 *gyr*A96 *rel*A1	Genentech
BL21-CodonPlus (DE3)-RIL	B F⁻ *omp*T hsdS(r$_B$⁻, m$_B$⁻) *dcm*+ Tet^r gal λ /DE3) *end*A The [*arg*U *ile*Y *leu*W Cam^r]	Stratagene

2.1.7.2 Human cell lines

Table 2.1.3 Human cell lines.

Cell line	Origin	Source
HeLa Kyoto (HK)	Cervical cancer	H. Leonhardt, LMU Munich
SK-mel147 (SK)	Metastatic melanoma	E. Bernstein, Mount Sinai, NY

Table 2.1.4 Stable human cell lines. HeLa Kyoto cell lines based on pEGFP-C1 plasmids were raised from G418-selected and FACS sorted single cell clones; HeLa Kyoto cell lines based on pIRESneo-eGFP plasmids were raised from G418-selected cell populations; SK-mel147 cell lines were lentiviraly transduced. #, clone or population number.

Cell line	Plasmid	Source
HK GFP-only#4	pEGFP-C1	(Bonisch et al., 2012)
HK GFP-H2A#4	pEGFP-C1-H2A	
HK GFP-H2A.Z.1#4	pEGFP-C1-H2A.Z.1	
HK GFP-H2A.Z.2.1#5	pEGFP-C1-H2A.Z.2.1	
HK GFP-PWWP2A_fl#5.2	pIRESneo-eGFP-PWWP2A_fl	this thesis
HK GFP-PWWP2A_NT#2	pIRESneo-eGFP-PWWP2A_NT	
HK GFP-PWWP2A_I#2	pIRESneo-eGFP-PWWP2A_I	
SK GFP-only	pEGFP-C1	(Vardabasso et al., 2015)
SK GFP-H2A	pEGFP-C1-H2A	(under revision)
SK GFP-H2A.Z.1	pEGFP-C1-H2A.Z.1	
SK GFP-H2A.Z.2.1	pEGFP-C1-H2A.Z.2.1	

2.1.8 Software

Application	Software
Image Processing	Adobe Photoshop CS5
	Adobe Illustrator CS5
	Fiji / ImageJ 2.0.0

	Huygens Essentials (4.4)
In silico modeling & visualization	iTASSER (web browser-based)
	Chimera (1.8.0)
Phylogeny analyses	Protein BLAST (web browser-based)
	CLC Main Workbench 7
Primer design	Primer 3 (web browser-based)
	Primer-BLAST (web browser-based)
qMS	Perseus (1.3.10.0)
	RStudio / R (3.0.2)
qPCR	LC480 software
Sequencing (ChIP)	BioViz Integrated Genome Browser
Sequence alignment	Clustal Omega (web browser-based)

2.1.9 Buffers and solutions

Ampicillin stock solution	100 mg/ml	Ampicillin (1000x)
Coomassie destaining solution	10%	Acetic acid (v/v)
	30%	Methanol (v/v)
Coomassie staining solution	10%	Acetic acid (v/v)
	50%	Methanol (v/v)
	0.1%	Coomassie Brilliant Blue (w/v)
Ethidium bromide stock solution	10 mg/ml	Ethidium bromide (20000x)
5 x Laemmli loading buffer	314 mM	Tris
(adjust pH to 6.8 with HCl)	50%	Glycerol (v/v)
	5%	SDS (v/v)
	5%	beta-Mercaptoethanol (v/v)
	0.01%	Bromphenol blue (w/v)
Laemmli running buffer	25 mM	Tris
	192 mM	Glycine
	0.1%	SDS (w/v)
LB agar plates	1.5%	LB Agar
LB medium	1.0%	Tryptone (w/v)
	0.5%	Yeast extract (w/v)
	1.0%	NaCl (w/v)

PBS	140 mM	NaCl
	2.7 mM	KCl
	10 mM	Na_2HPO_4
	1.8 mM	KH_2PO_4

SDS PAGE		
Separating gel	15%	Acrylamide
	0.4%	Bis-acrylamide
	363 mM	Tris (pH 8.8)
	0.1%	SDS (w/v)
	0.1%	TEMED (v/v)
	0.1%	Ammonium persulfate (w/v)
Stacking gel	4%	Acrylamide
	0.1%	Bis-acrylamide
	125 mM	Tris (pH 6.6)
	0.1%	SDS (w/v)
	0.1%	TEMED (v/v)
	0.1%	Ammonium persulfate (w/v)

TBE	45 mM	Tris
	45 mM	Boric acid
	1 mM	EDTA

Transfer buffer	48 mM	Tris
	39 mM	Glycine
	0.0375%	SDS (w/v)
	20%	Methanol (v/v)

2.2 Cell biological methods

2.2.1 Cultivation and manipulation of mammalian cells

2.2.1.1 Cultivation of mammalian cells

Adherent HeLa Kyoto and SK-mel147 wild type and stably transfected cell lines were kept in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 1% penicillin/streptomycin at 37°C and 5% CO_2 in a humidified incubator. Usually, cells were passaged in 10 cm tissue culture dishes (covered with 10 ml growth medium) and expanded to several 15 cm dishes (covered with 20 ml growth medium) if needed for experiments. Cells were splitted in a ratio of 1:10 every second to third day depending on doubling time and confluence. Therefore, old medium was removed, cells washed with sterile PBS and trypsin/EDTA was added (1:10 dilution in PBS). After 5 min of incubation at 37°C detaching cells were resuspended in fresh medium, cell suspension partly discarded (according to splitting ratio) and replaced by fresh medium. For experiments, cells were trypsinized, resuspended in DMEM and cell viability and cell number determined with the CASY cell counter (Innovatis). The respective amount of cells in suspension was then aliquoted, pelleted by centrifugation at 1200 rpm (tabletop centrifuge) and washed once with PBS. Regularly, cells were frozen down for storage. Thus, trypsinized cells were resuspended in FCS + 10% DMSO, transferred into cryo vials and stored at -80°C. For long time storage, cells were kept in liquid nitrogen. Cells in culture were replaced every 1 – 2 months with freshly thawed cells. Cryo vials were thawed quickly in a water bath (37°C) and the cells washed once with full growth medium to remove DMSO. Afterwards, cells were plated in DMEM on an appropriate tissue culture plate.

2.2.1.2 Establishment of stably transfected human cell lines

4×10^5 HeLa Kyoto cells were seeded into 6-well plated 24h prior to transfection to reach a confluence of 80 – 90%. 1 µg of plasmid DNA was diluted in 100 µl of Opti-MEM (final volume). 3 µl X-tremeGENE transfection reagent was added and incubated for 30 min at room temperature (RT). Dropwise, transfection suspension was added to the cells and dispensed by shaking the plate. Every plasmid was transfected in triplicates. Control cells were not transfected. In the case of plasmids coding for eGFP (GFP), transfection efficiency was monitored 24 h – 48 h after transfection by epifluorescence microscopy or fluorescence activated cell sorting (FACS). Selection for positive cells was mediated by selection medium,

containing 400 – 600 µg/ml G418-sulfate. Selection was carried out until non-transfected control cells were dead and controlled by FACS. Cell populations of 80 – 100% positive thus stable cells were aspired and employed for further experiments.

2.2.1.3   mRNA knockdown by RNA interference (RNAi)

2×10^5 HeLa Kyoto cells were seeded into 6-well plates 24 h prior to transfection to reach a confluence of 30 – 40%. Small interfering RNA (siRNA) stock solutions (100 µmol in 1x Universal Buffer) were diluted 1:5 in sterile water. 4 µl Oligofectamine transfection reagent was diluted in 11 µl Opti-MEM and incubated at RT for 5 min. Meanwhile, 10 µl of siRNA dilution was mixed with 175 µl Opti-MEM. 15 µl transfection reagent dilution was added and incubated for 20 min at RT. Cells were washed twice with sterile PBS, growth medium replaced by 800 µl DMEM not containing FCS or antibiotics and the transfection mix added drop wise. After 4 h at 37°C in the incubator 500 µl DMEM supplemented with 30% FCS was added. Cells were harvested and analyzed 2 to 3 days after transfection.

2.2.2   Proliferation analysis of adherent human cells

The proliferation of HeLa Kyoto cells upon RNAi treatment was monitored using the xCELLigence RTCA DP Analyzer equipped with 16-well E-plates. A baseline was set by measuring 100 µl medium per well and then 5000 cells were added in additional 100 µl medium 2 days after siRNA transfection. Cells were allowed to settle for 30 min at RT and then the measurement was started. Every 15 min, the incubator-housed xCELLigence analyzer recorded the impedance in every well over 96 h, providing quantitative information about parameters like cell number, viability and morphology. Recorded values of replicate experiments were exported and depicted as graphs including standard errors with Excel.

2.2.3   Immunofluorescence (IF) staining

Human cells were grown on 12 mm confocal-grade round coverslips in 24-wells, washed once in PBS and then fixed for 10 min in 2% formaldehyde in PBS at RT. Next, the fixation solution was replaced with PBS-T (PBS + 0.02% Tween20) by aspirating half of the fixation solution and filling up with PBS-T before complete exchange and two additional washes with PBS-T. Then, cells were permeabilized in PBS-T + 0.5% Triton X-100 for 10 min, washed twice with PBS-T and blocked in PBS containing 2% BSA for 1 h. Cells were incubated stepwise in primary and secondary antibody dilutions (in blocking solution) at the desired concentrations (see 2.1.4) for 1 h each. After washing 3x with PBS-T, the DNA was

counterstained with 4',6-Diamidin-2-phenylindol (DAPI) for 5 min at a concentration of 200 ng/ml in PBS, the coverslips were washed once in PBS and once in distilled water and then mounted using Vectashield mounting medium (Vector laboratories).

2.2.4   Immunofluorescence microscopy

IF stained cells on coverslips were imaged with the Zeiss epifluorescence microscope LSM200 or the confocal microscope Leica SP5 II. Confocal images were taken utilizing the Argon laser (488 nm) for GFP or Alexa 488 and the UV-diode (405 nm) for DAPI with the appropriate photomultiplier tubes (PMTs) in sequential mode. Optimal PMT performance was reached with a smart gain of 600 – 800 V. Further settings: see table below.

Laserlines:	Argon (488 nm, 15%)
	UV-diode (405 nm, 100%)
Acousto-optical tunable filters (AOTF):	405 nm: 15%, 488 nm: 30% - 40%
Photomultiplier tubes (PMTs):	PMT1: DAPI (415 – 470 nm)
	PMT2: GFP/Alexa 488 (498 – 570 nm)
Frame averaging:	4 frames per channel
Objective:	PL APO CS 63x 1.3 Gly 21°C UV
Glycerol:	Leica Type G n_e = 1.450
Pinhole:	default
Scanning speed:	400 Hz
Image depth:	16bit
Image size:	1024 x 1024 pixel (single cells)
	2048 x 2048 pixel (overviews)
Zoom:	6x (single cells)
	2x (overviews)
Pixel size:	40 nm x 40 nm (single cells)
	60 nm x 60 nm (overviews)

Deconvolution of confocal images was carried out using the Huygens Essential (4.4) software package, applying a recorded point spread function derived from fluorescence 270 nm beads. Confocal raw-images (.lif) with a pixel size of 40 nm x 40 nm were opened with Huygens and processed utilizing the 'Deconvolution Wizard'. The deconvolution settings are summarized

below, images cropped to 512 x 512 pixels and background intensities manually set for every image. Processed images were saved as 16 bit .tiff files.

Iterations:	40
Signal/noise:	15
Quality threshold:	0.000005
Iteration mode:	optimized
Bricklayout:	auto

2.2.5 Fluorescence Recovery After Photobleaching (FRAP) and live cell imaging

FRAP analyses were performed by our collaborator Dr. Katrin Schneider (LMU BioCenter, Munich) essentially as described in (Schneider et al., 2013). Live cell imaging and FRAP experiments were typically performed on an UltraVIEW VoX spinning disc microscope with integrated FRAP PhotoKinesis accessory (PerkinElmer) assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 100×/1.4 NA Plan-Apochromat oil immersion objective. The microscope was equipped with a heated environmental chamber set to 37°C. Fluorophores were excited with 488 nm solid-state diode laser line. Confocal image series were typically recorded with 14-bit image depth, a frame size of 512×512 pixels, a pixel size of 68.5 nm and with time intervals of 154 ms. For photobleaching experiments, the bleach regions, typically with a length of 8–10 μm, were chosen to cover the anterior half of the oval-shaped nucleus. Photobleaching was performed using two iterations with the acousto-optical tunable filter (AOTF) of the 488 nm laser line set to 100% transmission. Typically, 20 prebleach and 330 postbleach frames were recorded for each series. Data correction, normalization and quantitative evaluations were performed by automated processing with ImageJ (http://rsb.info.nih.gov/ij/) using a set of self-developed macros followed by calculations in Excel.

2.2.6 Fluorescence Activated Cell Sorting (FACS) analyses

FACS analyses were carried out using the FACSCanto machine (BDI Bioscience), equipped with the FACS Diva software package. Further analyses were performed with the FlowJo (8.8.7) software suit.

2.2.6.1 Transfection efficiency / stable cell lines

In order to monitor GFP-expression of cells transiently or stably expressing a GFP-tagged transgene (histone variants or PWWP2A) cells were subjected to FACS analyses after harvesting. Cell suspensions in medium were measured and gated according to forward (FSC) and sideward (SSC) scatter into the viable, non-debris cell population. The GFP fluorescence of this population was then visualized utilizing a histogram.

2.2.6.2 Cell cycle analysis by propidium iodide (PI) staining

To determine the cell cycle phase of a HeLa Kyoto cell population upon RNAi treatment, cells were stained with propidium iodide (PI), a non-permeable fluorescent dye intercalating into the DNA. Briefly, 1×10^6 cells were harvested, washed once with PBS and then resuspended in 300 µl PBS. Cells were then permeabilized by adding 700 µl 100% methanol while vortexing the suspension. Next, samples were incubated for 30 min at -20°C, washed once with 1 ml PBS and the cell pellet incubated in a 200 µl staining solution containing 100 µg/ml RNase A and 50 µg/ml PI at 37°C for 30 min. Without further washing the cells were then analyzed by FACS. PI positive cells were plotted in a linear histogram, allowing a discrimination of cells in G_1, S or G_2/M phase of the cell cycle according to their DNA content. Further analyses were carried out using the FlowJo 'cell cycle' tool and utilizing the Dean-Jett-Fox model with default settings that determines the percentage of cells in every respective cell cycle phase (Fox, 1980).

2.2.6.3 Mitose (H3Ser10phos)

Discrimination between cells in G_2 and M phase was achieved by specifically staining mitotic cells with an antibody recognizing the phosphorylated serine 10 of histone H3 that only occurs during mitosis. Essentially, cells were prepared as described for PI staining. After permeabilization and washing, cells were resuspended in 200 µl PBS containing the H3S10phos antibody diluted 1:200 and incubated at RT for 1 h on a rotating wheel. Cells were then washed once in PBS and subjected to incubation with an Alexa 488-coupled secondary antibody in PBS (1:200 in 200 µl PBS) at RT for 1 h in the dark. After washing the cells once with PBS they were PI stained as described above and analyzed by FACS. A gate was set to isolate G_2/M-phase cells in the PI histogram and this population separated into G_2- and M-phase cells in a scatter plot according to their PI and Alexa 488 intensities.

2.2.6.4 Apoptosis (Annexin V)

Apoptotic cells were detected utilizing the Annexin V staining kit (eBioscience) including DNA staining with PI according to manufacturer's instructions.

2.2.7 Trichostatin A treatment

A trichostatin A (TSA, MW = 302.37 g/mol) stock solution of 1 mg/ml was reconstituted with 1 ml DMSO from 1 mg lyophilized TSA, aliquoted and stored at -20°C. This 3307.206 µM stock solution was diluted with DMEM to 100 µM and then added at a final concentration of 200 nM to cells supplied with fresh growth medium. Control cells were treated with DMSO only and cells harvested 2 h after treatment for subsequent assays.

2.3 Molecular biological methods

2.3.1 Agarose gel electrophoresis

DNA fragments originating from cloning or MNase digestions were separated using agarose gels prepared with TBE buffer and the respective amount of agarose (1% agarose for cloning fragments and 2% agarose for MNase fragments). DNA was visualized by the intercalating and fluorescent agent ethidium bromide, added to the gels at a final concentration of 0.5 µg/ml.

2.3.2 RNA extraction of human cells

Total RNA from human cell lines was isolated utilizing the RNeasy Kit (Qiagen) according to manufacturer's instructions. The quality of the RNA (amount and purity) was assessed using a Nanodrop ND1000 spectrophotometer. Isolated RNA was stored at -80°C.

2.3.3 cDNA synthesis

cDNA was transcribed from 1 µg RNA with the ProtoScript First Strand cDNA Synthesis Kit (NEB) following the manufacturer's instructions. Oligo-dT primers were used and samples without reverse transcriptase were utilized as control for genomic DNA contamination. A test PCR amplifying GAPDH from the prepared cDNA followed by agarose gel electrophoresis served as quality control.

2.3.4   Quantification of mRNA levels with quantitative PCR (qPCR)

mRNA levels of genes of interest were detected by qPCR employing a Roche LightCylcer 480 II (LC 480) and expression assessed relative to two housekeeping genes (HPRT1 and HMBS). Primer pairs were designed using the web-based Primer3 tool or the 'Pick Primer' (Primer-BLAST) tool on the NCBI Nucleotides homepage. All Primers are listed in table 2.1.6.2. Using non-template controls, melting curves and cDNA dilutions, primer pair specificity was tested. Primer efficiencies were assessed by measuring serial cDNA dilutions and calculating standard curves with the 'Advanced Absolute Quantification' tool that is part of the LC 480 software package. Samples were analyzed in technical triplicates with a total reaction volume of 15 µl. Per reaction, 0.15 µl cDNA (prepared as described above) were diluted with qPCR-grade water to 5 µl and distributed to different wells of a 384-well plate with a multistep pipet and 0.1 ml combitips. Furthermore, 7.5 µl Fast SYBR Green Master Mix (Applied Biosystems) were mixed with primers (to a final concentration of 0.44 µM per primer) and qPCR-grade water to a volume of 10 µl and subsequently added to wells already containing the cDNA dilution with a multistep pipet and 0.2 ml combitips. After sealing the plate with an adhesive foil, it was centrifuged 2 min at 2000 rpm (Heraeus Megafuge 2.0) and then loaded into the LighCycler. The following PCR program was applied:

	Pre-Incubation	95°C	20 sec
45 x	Amplification	95°C	3 sec
		60°C	20 sec
	Melting curve	95°C	5°C/min

Resulting crossing point (CP) values were analyzed employing the 'Advanced Relative Quantification' tool that is part of the LC 480 software package. It considers pre-assessed primer efficiencies and computes the mRNA levels of the respective gene of interest relative to two reference genes (HPRT1 and HMBS). Background levels can be quantified implementing no RT controls mentioned in the cDNA synthesis section.

2.3.5   Validation of MNase ChIP-sequencing results with quantitative PCR (MNase ChIP-qPCR)

MNase ChIP-sequencing results were validated by quantitative PCR. Primer pairs amplifying MNase fragments located at the +1 and -1 nucleosomes around the TSS of selected genes (PARS2, RNF11, UHMK1, NUF2 and SWT1/TRMT1L) and in the gene body of two

controls genes (RPL11 and YIPF) on human chromosome 1 were designed employing the Integrated Genome Browser software (BioViz) and the NCBI Gene tools. For details on primer sequence and location please review section 2.1.6.2. Primer efficiencies were determined using dilutions of MNase digested GFP-H2A input DNA and calculated with the 'Advanced Absolute Quantification' tool that is part of the LC 480 software. Samples were analyzed in technical triplicates with a total reaction volume of 15 µl. Per reaction, 1 µl ChIP DNA (diluted 1:10 in qPCR-grade water) was diluted with qPCR-grade water to 5 µl and distributed to different wells of a 384-well plate with a multistep pipet and 0.1 ml combitips. Furthermore, 7.5 µl Fast SYBR Green Master Mix (Applied Biosystems) were mixed with primers (to a final concentration of 0.44 µM per primer) and qPCR-grade water to a volume of 10 µl and subsequently added to wells already containing the MNase ChIP DNA dilution with a multistep pipet and 0.2 ml combitips. In addition, one input triplicate per primer pair with an amount of 1 ng was included as standard for the subsequent absolute quantification. For PCR program please see section 2.3.4. Enrichments at selected TSS were normalized twice. First, logarithmized input values were subtracted from logarithmized IP values (for TSS and control loci) and subsequent, values for one control gene body locus (YIPF) were subtracted from TSS loci values. In this way, the final enrichments around the TSS were depicted relative to a negative control region and the input. Fold enrichments were visualized in Excel bar plots.

2.3.6 Cloning of PWWP2A and PWWP2A truncations

A DNA fragment of 2332 bp containing the PWW2A coding sequence and parts of the 5' and 3' UTR was amplified with gene specific primers and Phusion DNA Polymerase from SK-mel147 cDNA. PCR products were subcloned into the EcoRV-cut and dephosphorylated shuttle vector pT7blue3 allowing for blue/white screening after transformation of chemical competent *E. coli* (DH5α) with the ligation reaction. DNA of several white (positive) clones was isolated from over night (ON) 5ml LB-Amp cultures with a Miniprep Kit (Qiagen). Subsequent sequencing (Eurofins MWG Operon) of the isolated plasmid DNA revealed one positive clone. Initially, this full-length PWWP2A fragment was used in Gateway cloning by Andrea Schmid (head of the Gateway cloning facility in our institute) to create a N-terminally GFP-tagged PWWP2A fusion protein construct (pIRESneo-eGFP-PWWP2A_fl) for *in vivo* studies and a GST-tagged PWWP2A fusion construct (pGEX6P1-PWWP2A_fl) for recombinant bacterial expression. In addition, Andrea created several PWWP2A truncation

constructs for eukaryotic and bacterial expression (see section 2.1.5) based on our requirements.

2.4 Biochemical methods

2.4.1 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated using polyacrylamide gels fitting the percentage of the separating gels and the number of gel pockets to particular needs. Self-poured gels in disposable cassettes were used as well as precast gels from Serva. peqGOLD Protein Marker IV or V were used as size markers. Protein samples were denatured in Laemmli loading buffer at 95°C for 5 min prior to loading. Gels were run at 90 V until the marker started to separate, then continued at 180 V. Subsequently, gels were stained with Coomassie or used for immunoblotting

2.4.2 Coomassie staining of polyacrylamide gels

To detect separated proteins, polyacrylamide gels were incubated in Coomassie staining solution for 1 h at RT and then destained in destaining solution until protein bands become visible. Destained gels were washed with water and scanned with the following settings: transparency mode, 300 dpi resolution, 16bit HDR grey scale, tiff-format.

2.4.3 Immunoblotting

Polyacrylamid gels were blotted on nitrocellulose membranes (Amersham) employing a semi-dry blotting device (Bio-Rad). Before assembling a blotting sandwich, 3mm Whatman paper, membranes and gels were soaked for 2 min in transfer buffer. Stacking of two Whatman papers, membrane, gel and two additional Whatman papers assembled a blot. Proteins were transferred onto the membrane applying 200 – 300 mA for 1 h. Membranes were blocked 1 h at RT in PBS-T (PBS + 0.1 % Tween20) containing 5% milk powder and subsequently incubated with primary antibody diluted in blocking solution ON at 4°C (see section 2.1.4 for antibody dilutions). Membranes were washed 3 times for 10 min with PBS-T and incubated with the secondary antibody diluted in blocking solution for 1 h at RT. Next, blots were washed 3 times for 10 min with PBS-T and incubated with ECL detection reagent for 2 min at RT. X-ray films (Fujifilm) were exposed in a dark room and developed utilizing a developer machine (AGFA). Films were subsequently scanned as described for Coomassie gels.

2.4.4 Mononucleosome preparation

Human cell lines were grown in 15 cm tissue culture dishes and harvested as described in section 2.2.1.1. Cells were counted with the CASY counter (Innovatis), 4×10^7 cells aliquoted into a 15 ml falcon tube and washed once with PBS. All following steps were performed on ice or at 4°C and protease inhibitors (Roche) were added to all buffers. Nuclei were prepared by resuspending the pelleted cells in 5 ml PBS containing 0.3% Triton X-100 and incubating them for exact 10 min at 4°C while rotating thereby lysing the plasma membrane but not the nuclear membrane. Nuclei were pelleted for 5 min at 2000 rpm (Eppendorf 5810 R), washed in 5 ml PBS and centrifuged again. The supernatant was discarded and remaining buffer removed by pipet. Pelleted nuclei were resuspended in 500 µl ice-cold EX100 buffer per 4×10^7 cells and transferred to 1.5 ml low-binding reaction tubes. Calcium, important for subsequent MNase digestion, was adjusted to a final concentration of 2 mM with $CaCl_2$ (4 µl of a 250 mM stock solution). Lyophilized micrococcal nuclease (MNase, Sigma) was reconstituted with sterile water to gain an enzymatic activity of 1 U/µl, aliquoted, frozen in liquid nitrogen and stored at -80°C. 1.5 U MNase was added to the nuclei in suspension and incubated at 26°C for 20 min. The reaction was stopped by adding EGTA to a final concentration of 10 mM (10 µl of a 500 mM stock solution) and centrifuged at full speed (Eppendorf 5417 R) at 4°C for 25 min. The supernatant (S1) was transferred into a fresh tube, containing almost exclusively mononucleosomes and was used as input for subsequent immunoprecipitations (see section 2.4.6). 25 µl S1 were boiled in Laemmli loading buffer for 5 min at 95°C as input fraction for SDS-PAGE analysis. A second fraction (S2) could be extracted by incubating the pellet over night at RT in 500 µl resuspension buffer and centrifuging the suspension the next day on full speed (supernatant = S2). The resulting pellet (P) was analyzed by resuspending the pellet in 200 µl 1 x Laemmli buffer and incubating it at for 10 min at 95°C. Then, benzonase (500 U) was added, incubated for at least 2 h at 37°C and inactivated by cooking the suspension again at 95°C for 10 min.

EX100 buffer

	10 mM	Hepes pH 7.6
	100 mM	NaCl
	1.5 mM	$MgCl_2$
	0.5 mM	EGTA
	10%	Glycerol (v/v)
	10 mM	β-Glycerol phosphate
add prior to use:	1 mM	DTT
	1 x	Protease inhibitors

<u>**Resuspension buffer**</u>

1 x	PBS
150 mM	NaCl
2 mM	EGTA
0.1%	Triton X-100 (v/v)
add prior to use:	1 x	Protease inhibitors

2.4.5 Purification of MNase digested DNA to determine the digestion degree

To determine the degree of MNase digestion, proteins were removed from the S1/S2 fraction by phenol/chlorophorm/isoamylalcohol extraction and DNA isolated by ethanol precipitation. To do so, 25 µl S1/S2 were filled up to 200 µl with 5 mM Tris-HCl (pH 8.0), 200 µl phenol was added and the tube inverted several times. Then, 200 µl chlorophorm/isoamlyalcohol (ratio 24:1) were added, inverted and transferred to pre-spun (11.000 rpm, tabletop centrifuge) maXtract tubes (Qiagen). Centrifugation at 11000 rpm resulted in separation of aqueous and organic phase and enabled to transfer the aqueous phase (200 µl) containing the digested DNA into a fresh tube. Glycogen was added to a final concentration of 200 µg/ml (2 µl of a 20 mg/ml stock solution), as well as sodium acetat to a final concentration of 0.3 M (20 µl of a 3 M stock solution) and 500 µl 100 % ethanol. DNA was precipitated for 20 min at -20°C, spun down for 30 min at 13300 rpm (tabletop centrifuge) and 4°C and whitish pellet washed with 500 µl of 70 % ethanol. The almost transparent pellet dried for 5 min at RT, resuspended in 30 µl 5 mM Tris-HCl (pH 8.0) and DNA concentration assessed using a Nanodrop ND1000 spectrophotometer (Peqlab). Digested DNA was visualized on a 2% agarose gel (1 µg DNA, 45 min at 100 V) or a 1000 DNA BioAnalyzer chip (Agilent).

2.4.6 Mononucleosome chromatin immunoprecipitation (MNase ChIP)

Preparation of mononucleosomes from human cell lines was performed as described in 2.4.4. All steps were done on ice or at 4°C and protease inhibitors (Roche) were added to all buffers. Immunoprecipitations were carried out with mononucleosomes from 4×10^7 cells in 1.5 ml low-binding tubes. 25 µl slurry GFP-trap magnetic beads (Chromotek) per IP were equilibrated in EX100 buffer, added to the mononucleosomes and incubated for 2.5 h at 4°C while rotating. Beads were quickly spun down, separated in a magnetic rack and the supernatant kept as non-bound fraction. Beads were washed 2 x 5 min with 1 ml wash buffer 1 and 2 x 5 min with 1 ml wash buffer 2 while rotating. Clumpy beads were resuspended carefully by pipetting. Beads were separated in a magnetic rack and boiled in 25 µl 1 x Laemmli loading buffer (5 min at 95°C) after discarding the supernatant.

Immunoprecipitations were analyzed by SDS-PAGE followed by Coomassie staining or immunoblotting together with input and non-bound fractions.

Wash buffer 1

	10 mM	Tris pH 7.5
	150 mM	NaCl
	0.1%	NP-40 (v/v)
add prior to use:	1 mM	DTT
	1 x	Protease inhibitors

Wash buffer 2

	10 mM	Tris pH 7.5
	150 mM	NaCl
add prior to use:	1 mM	DTT
	1 x	Protease inhibitors

2.4.7 Mononucleosome chromatin immunoprecipitations followed by quantitative label-free interaction proteomics (MNase ChIP-qMS)

2.4.7.1 MNase ChIP

Preparation of mononucleosomes from human cell lines was performed as described in 2.4.4. IPs were essentially carried out as described in section 2.4.6 with the following variations. IPs were conducted with mononucleosomes from 8×10^7 cells and in technical triplicates ($3 \times 8 \times 10^7$ cells). After MNase digestion, all supernatants containing mononucleosomes ($6 \times \sim 500$ µl) were combined, in total 50 µl S1 kept as input fraction for future SDS-PAGE analyses and determination of the digestion degree and the rest aliquoted into three replicate reactions with equal volume (~ 1000 µl) of mononucleosomes (in low-binding tubes). IP and wash steps were performed as described before. 50 µl of the last wash step (resuspended beads in wash buffer 2; equals 5% of respective IP) of every replicate were combined (equals 15% of a single IP), beads magnetically separated and boiled in 10 µl 1 x Laemmli loading buffer as IP fraction for SDS-PAGE analysis.

2.4.7.2 On-beads tryptic digest

To prepare the precipitated proteins for on-beads tryptic digestion, the remaining suspension was magnetically separated, the washing solution discarded and the beads incubated with buffer E1 for 20 min at 25 °C and mild shaking (Eppendorf thermomixer compact, 500 rpm). After magnetic separation, the supernatant was transferred to a fresh low-binding tube and 2-chloroacetamide (CAA, Sigma) added to a final concentration of 5 mM. Beads were

resuspended in Buffer E2, shook for 5 min and then trypsin (Promega) was added to a final concentration of 5 µg/ml. The reaction was incubated for 1 h at 25°C while shaking (500 rpm), magnetically separated and the supernatant added to the first supernatant. The mixture was incubated ON at RT. The tryptic digest was stopped the next morning by adding 1 µl concentrated trifluoracetic acid (TFA). Remaining beads were resuspended in 25 µl 1 x Laemmli loading buffer (5 min at 95°C) to check complete protein elution from the beads in SDS-PAGE and Coomassie staining later.

Basic buffer

50 mM	Tris pH 7.5
2M	Urea

E1 buffer

1 mM	EDTA in basic buffer

E2 buffer

5 mM	CAA in basic buffer

2.4.7.3 StageTips

Digested immunoprecipitations were kept at 4°C until StageTip loading. StageTips were employed to clean and concentrate tryptic peptides (Rappsilber et al., 2007) and were kindly provided by our collaborator Eva Keilhauer (MPI of Biochemistry). A working and a backup set of StageTips for every replicate IP were labeled and activated with 100 µl methanol. Methanol was removed by centrifugation (Eppendorf 5415 D, 4000 rpm, 1 min) and tips washed two times with 100 µl 0.5% acetic acid. The digestion reactions were divided into two parts: 50 µl were loaded on the working tips and 50 µl on the backup tips. Next, tips were washed once with 50 µl 0.5% acetic acid and dry tips stored at 4°C until measuring.

2.4.7.4 LC-MS/MS Analysis

Peptides were eluted from the C18 StageTips according to the standard protocol (Rappsilber et al., 2007). Samples were analyzed by reversed-phase liquid chromatography on an EASY-nLC 1000 system (Thermo Fisher Scientific) directly coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific) using a nanoelectrospray source (Proxeon Biosystems, now Thermo Fisher Scientific) by Eva Keilhauer. High performance liquid chromatography (HPLC) columns of 20 cm length and an inner diameter of 75 µm were in-

house packed with ReproSil-Pur 120 C18-AQ 1.9 µm particles (Dr. Maisch GmbH). Peptide mixtures were separated using linear gradients of 140 min (total run time + washout) and a two buffer system: buffer A++ (0.1% formic acid) and buffer B++ (80% acetonitrile in 0.1% formic acid). The flowrate was set to 250 nl/min and the column was heated to 50°C using a column oven (Sonation GmbH). Peptides eluting from the column were directly sprayed into the mass spectrometer; with the spray voltage set to 2.4 kV and the capillary temperature set to 250°C. The mass spectrometer was operated in a data-dependent mode acquiring survey scans at a resolution of 70.000 with an AGC target of 3E06 ions and a maximum ion injection time of 20 msec. Subsequently, the top 10 most abundant peaks were selected for fragmentation with an isolation window of 2m/z and fragmented by higher energy collisional dissociation (HCD) with a normalized collision energy of 25. Fragmentation spectra were acquired at a resolution of 17.500 with a target value of 1E05 ions and 120 msec as maximum ion injection time. To minimize re-sequencing of peptides, dynamic exclusion was enabled within a time window of 20 sec.

2.4.7.5   Raw MS Data analysis

MS raw files were processed by Eva Keilhauer using MaxQuant (Cox and Mann, 2008) version 1.3.9.20, leading to the identification of > 1000 proteins in each technical replicate. MS/MS spectra were searched against a human sequence database obtained from Uniprot on 2/25/2012 using the Andromeda search engine (Cox et al., 2011). Cysteine carbamidomethylation was set as a fixed modification; N-terminal acetylation and methionine oxidation were set as variable modifications. Trypsin was chosen as specific enzyme, with 2 maximum missed cleavages allowed. Peptide and protein identifications were filtered at a 1% FDR. Label-free quantification was enabled, with a minimum ratio count of 1. The match between runs option was enabled, while the requantify option was disabled. All other parameters were left at standard settings.

2.4.8   Mononucleosome chromatin immunoprecipitations followed by next generation sequencing (MNase ChIP-seq)

MNase IPs of GFP-PWWP2A or nucleosomes containing GFP-H2A, GFP-H2A.Z.1 or GFP-H2A.Z.2 were carried out in biological duplicates. Cells expressing GFP-only served as negative control. Sequencing was done by Dr. Gregor Gilfillan (Norwegian Sequencing Center, NGS, Oslo) and Dr. Stefan Krebs (Laboratory of Functional Genome Analysis,

LAFUGA, Munich). Bioinformatic analyses were performed by our collaborator Dr. Tobias Straub (head of our in-house bioinformatics core facility).

2.4.8.1 MNase ChIPs

MNase ChIPs employing GFP-trap magnetic beads (GFP-H2A, GFP-H2A.Z.1, GFP-H2A.Z.2 and GFP-only control MNase IPs) were carried out essentially as described in section 2.4.6 with the following variations. After immunoprecipitations, beads were washed 2 x 5 min with 1 ml of each wash buffer 1, wash buffer 2 and LiCl buffer, and 1 x 5 min with 1 ml TE + 0.2% Triton X-100 and 1 x 5 min with 1 ml TE. For GFP-PWWP2A IPs, magnetic G-protein beads (Millipore) coupled to a goat anti-GFP antibody (kind gift of Prof. Andreas Ladurner, LMU Munich) turned out to outperform the GFP-trap magnetic beads with respect to specificity: the amount of DNA pulled down unspecific in the GFP-only control IP was reduced. Mononucleosomes prepared from the stable HeLa Kyoto GFP-PWWP2A cell line were incubated over night with 2 µl of GFP antibody at 4°C while rotating. Next, 20 µl of magnetic G-protein beads equilibrated in EX100 buffer were incubated with the reaction for additional 4 h at 4°C while rotating. Beads were washed as described for GFP-trap IPs, excluding the LiCl buffer wash step.

Wash buffer 1

	10 mM	Tris pH 7.5
	1 mM	EDTA
	0.1%	SDS (w/v)
	0.1%	Sodiumdeoxycholate (w/v)
	1%	Triton X-100 (v/v)
add prior to use:	1 x	Protease inhibitors

Wash buffer 2

	10 mM	Tris pH 7.5
	1 mM	EDTA
	0.1%	SDS (w/v)
	0.1%	Sodiumdeoxycholate (w/v)
	1%	Triton X-100 (v/v)
	150 mM	NaCl
add prior to use:	1 x	Protease inhibitors

LiCl buffer

	10 mM	Tris pH 7.5
	1 mM	EDTA
	0.5%	NP-40 (v/v)
	0.5%	Sodiumdeoxycholate (w/v)
	250 mM	LiCl
add prior to use:	1 x	Protease inhibitors

TE buffer

	10 mM	Tris pH 7.5
	1 mM	EDTA
add prior to use:	1 x	Protease inhibitors

2.4.8.2 *Purification of immunoprecipitated DNA*

Purification of precipitated DNA and respective input fraction DNA was conducted essentially as described in (Cuddapah et al., 2009). Briefly, washed beads were resuspended in 100 µl TE, 3 µl 10% SDS and 5 µl of 20 mg/ml proteinase K were added and incubated for 1 h at 65 °C. Suspensions were vortexed briefly, magnetically separated and the supernatant transferred to a fresh tube. Beads were washed once with 100 µl TE containing 0.5 M NaCl, magnetically separated and the supernatant mixed with the first supernatant. Input fractions were processed in parallel. Phenol/chlorophorm/isoamylalcohol extraction and DNA precipitation with ethanol were carried out as described in section 2.4.5 with the following variations. Samples were not filled-up with Tris-HCl buffer, as they already possess the right volume (200 µl). After washing the DNA precipitate, the IP DNA pellet was resuspended in 12 µl and the input DNA pellet in 32 µl 10 mM Tris-HCl (pH 7.5). DNA concentrations were determined with the Qubit dsDNA hs Kit (Invitrogen) and DNA size monitored on a 1000 DNA BioAnalyzer chip (Agilent).

2.4.8.3 *Library preparation*

Illumina Sequencing libraries were established with the MicroPlex Library Preparation Kit (Diagenode) following the manufacturer's instructions with following variations. The number of step 5 amplification cycles was scaled according to the amount of input material (see Table 2.4.1). 10 µl of library material were send for sequencing. When purifying libraries after amplification, samples were incubated 15 min with AMpure beads instead of 5 min, ethanol washed beads were dried for 3 min at RT instead at 37°C and DNA was eluted in 22 µl 10 mM Tris-HCL pH 7.5 instead of TE.

Table 2.4.1 HeLa Kyoto MNase ChIP library preparation

Name	Type	Replicate	Input conc. [ng]	Index	Step 5 cycles	Library conc. [ng]
H2A IP rep1	ChIP	1	50	1	3	246
H2A IP rep2		2		2		214
H2A.Z.1 IP rep1		1		3		220
H2A.Z.1 IP rep2		2		4		238
H2A.Z.2.1 IP rep1		1		5		200
H2A.Z.2.1 IP rep2		2		6		214
H2A Input rep1	Input	1	50	11	3	181
H2A Input rep2		2		12		292
PWWP2A#5.2 IP 1st	ChIP	1	20	3	6	330
PWWP2A#5.2 IP 2nd		2		1		302
PWWP2A#5.2 Input 1st	Input	1	20	4	6	212
PWWP2A#5.2 Input 2nd		2		2		236

2.4.8.4   Illumina Sequencing

Our collaborator Dr. Gregor Gilfillan (NGS) performed next generation sequencing of MNase ChIPs of GFP-tagged histone H2A and histone H2A.Z variant nucleosomes. Next-generation sequencing of GFP-PWWP2A MNase ChIPs was performed by Dr. Stefan Krebs at the Laboratory of Functional Genome Analysis (LAFUGA) in Munich (Gene Center). Libraries were sequenced on an Illumina HiSeq 2000 using V3 clustering and sequencing reagents (50 bp read length) according to manufacturer's instructions. Image analysis and base calling were performed using Illumina's RTA software version 1.13.48. Reads were filtered to remove those with low base call quality using Illumina's default chastity criteria.

2.4.9 Recombinant bacterial expression and affinity purification of GST-tagged PWWP2A and PWWP2A truncations

Nina Wommelsdorf, a master student whom I supervised, carried out all assays with recombinant PWWP2A. BL21 chemical competent *E. coli* cells were transformed with pGEX6P1-PWWP2A constructs (see section 2.1.5). For this purpose, 50 µl of bacterial culture were thawed and incubated on ice with 50 ng of plasmid DNA. Following a heat shock of 45 sec at 42°C, bacterial cells were cooled down on ice for 3 min, 100 µl LB-medium without antibiotics was added and the cell suspension incubated for 1 h at 37°C.

Cells were spread on LB-agar plates containing ampicillin (stock: 100 mg/ml, 1:1000 dilution →100 µg/ml) and single clone colonies grown ON at 37°C. A 5 ml LB-medium pre-culture containing ampicillin (100 µg/ml) was inoculated with a single colony on the next day and incubated ON at 37°C shaking at 160 rpm (Infor shaker). 2 ml of this pre-culture were then used to inoculate 200 ml LB-medium (containing 100 µg/ml ampicillin) the next morning and grown at 37°C and 160 rpm until an OD_{600} of 0.4 – 0.5 was reached. Next, the culture was cooled down to 18°C. Reaching an OD_{600} of 0.6 – 0.7 a sample of 1 ml was taken (uninduced culture) for SDS-PAGE, the bacteria pelleted for 5 min full speed (tabletop centrifuge) and boiled at 95°C in 100 µl 1 x Laemmli loading buffer. Adding 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) induced recombinant protein expression of the remaining culture. Moreover, additional ampicillin (50 µg/ml) was added. After 16 – 18 h at 18°C another 1 ml sample was taken for SDS-PAGE (induced culture), OD_{600} determined again and the induced sample diluted relative to the OD_{600} of the uninduced sample to assure constant loading conditions in SDS-PAGE. The main culture was centrifuged at 4000 rpm (Heraeus Cryofuge 6000i) for 15 min, the bacterial pellet shock frozen in liquid nitrogen and stored at -80°C. Protein purification was carried out on ice or at 4°C, utilizing ice-cold buffers containing protease inhibitors (Roche) in every step. Resuspension in 10 ml lysis buffer and sonication in iced water for 2 x 2 min (30% output, 1 sec on, 1 sec off) with a sonifier (Branson) lysed cells. The lysate was cleared by centrifugation at 15000 rpm (Sorvall RC6 Plus) at 4°C. For SDS-PAGE, 100 µl of cleared lysate were kept as input fraction and boiled in 20 µl 5 x Laemmli loading buffer at 95°C. Purification of GST-tagged recombinant proteins was carried out employing glutathione sepharose beads (GE). 200 µl of resuspended slurry beads were washed twice with 1 ml PBS for 5 min on a rotating wheel. Centrifugation of beads is limited to 500 g. Bacterial lysate was added to washed beads and incubated in a 15 ml falcon at 4°C on a rotating wheel for 2 h. Beads were pelleted by centrifugation, 100 µl supernatant kept as non-bound fraction for SDS-PAGE and beads washed three times for 10 min at 4°C while rotating with PBS first, then with PBS + 0.4% NaCl and concluding with PBS again. 100 µl of the first wash step were kept for SDS-PAGE. Beads were transferred into a fresh tube and stored in PBS at 4°C. Purification was assessed by SDS-PAGE on 10% polyacrylamide gels and subsequent Coomassie staining or immunoblotting.

Lysis buffer

1 x	PBS
500 mM	NaCl
0.1%	NP-40 (v/v)
add prior to use: 1 x	Protease inhibitors

Table 2.4.2 Recombinant GST-PWWP2A full length and truncations. MW, expected molecular weight.

PWWP2A_	Amino acid stretch	MW incl. GST-tag (kDa)
full length / fl	1 – 755	110
NT	1 – 654	99
P1	1 – 240	53
P1_P2_I	1 – 574	90
P2_I	147 – 574	76
P2_I_S	147 – 654	85
P2_I_S_PWWP	147 – 755	96
I	292 – 574	60
IN	292 – 422	43
conI	301 – 516	53
IC	423 – 574	45
I_S	292 – 654	69
I_S_PWWP	292 – 755	80
CT	655 – 755	40

2.4.10 *In vitro* binding assays with recombinant PWWP2A constructs

Recombinantly expressed GST-PWWP2A fusion proteins were kept on glutathione sepharose beads after purification from bacterial lysates and used for *in vitro* binding assays with mononucleosomes prepared from human cell lines or with recombinant mononucleosomes. All steps were carried out on ice or at 4°C and protease inhibitors (Roche) were added to all buffers. Interaction of PWW2A and mononucleosomes was analyzed by SDS-PAGE and Coomassie staining or immunoblotting.

2.4.10.1 Binding assays with cellular mononucleosomes

Mononucleosomes were prepared from HeLa Kyoto and SK-mel147 cell lines as described in 2.4.4. Mononucleosomes from 4×10^7 cells were incubated with 25 µl PWWP2A_fl or a truncated version coupled to glutathione sepharose beads ON at 4°C on a rotating wheel.

Next, beads were pelleted by mild centrifugation (2300 rpm, table top centrifuge) and the supernatant kept as non-bound fraction for subsequent SDS-PAGE analysis. A reduced EX100 buffer was used for 4 x 10 min washes on a rotating wheel at 4°C. The first two washes included additional 150 mM NaCl and 0.1% NP-40 in the buffer, the last two washes additional 100 mM NaCl only and no NP-40. Proteins were eluted from the beads by resuspending them in 25 µl 1 x Laemmli loading buffer and boiling them at 95°C for 5 min. Samples were analyzed by SDS-PAGE and Coomassie staining or immunoblotting.

EX100 buffer, reduced

	10 mM	Hepes pH 7.6
	100 mM	NaCl
	1.5 mM	$MgCl_2$
add prior to use:	1 mM	DTT
	1 x	Protease inhibitors

2.4.10.2 Binding assays with recombinant mononucleosomes

Instead of using mononucleosomes prepared from human cells, recombinantly expressed and *in vitro* reconstituted nucleosomes (assembled by Clemens Bönisch, (Bonisch et al., 2012)) were used. Briefly, 10 µl of the nucleosome preparation were incubated ON in 500 µl EX100 with 25 µl protein on glutathione sepharose beads at 4°C on a rotating wheel. Washes, elution and analysis were conducted as described above.

2.4.10.3 Electrophoretic Mobility Shift Assay (EMSA)

GST-PWWP2A recombinant proteins were eluted from gluthatione-sepharose beads by incubating the beads three times for 15 min in 200 µl 1 x PBS containing 30 mM glutathione and collecting the cleared supernatant (= eluate). Three eluates were pooled and stored at -80°C. Protein concentrations were determined by Bradford assay (Bio-Rad Protein Assay) employing a BSA standard curve. Protein-DNA-interactions were examined by using an Electrophoretic Mobility Shift Assay. Recombinant GST-PWWP2A proteins were cleared by centrifugation (10 min, full speed, 4°C) and then diluted in protein buffer to a final concentration of 1 µM in a 12 µl reaction volume. In order to test different protein concentrations, a 1:1 dilution series was performed by pipetting 6 µl of previous protein dilution to 6 µl of protein buffer. Double-stranded DNA containing the 601 sequence (CCCGGTGCCGAGGCCGCTCAATTGG), labeled at the 5 'end with Cy5 was a kind gift from the Peter Becker group. The DNA (concentration: 10 µM) was diluted to 50 nM in EMSA buffer and 6 µl DNA was added to different GST-PWWP2A dilutions. Binding

reactions were performed at RT for 15 min and analyzed on a 4.5% native TBE gel (100 V, 65 min, 4°C, running buffer: 0.3 x TBE) and afterwards imaged with the PhosphoImager FLA3000 (Fuji).

EMSA buffer

1 x	PBS
0.1%	Tween20
5%	Glycerol

Protein buffer

1 x	PBS
0.1%	Tween20

Native gel

30%	1 x TBE
15%	Acrylamide
8.4%	Glycerol
1%	Ammonium persulfate
0.1%	TEMED

2.5 Bioinformatics

2.5.1 MNase ChIP-qMS analyses

Analyses of label-free interaction proteomics data were performed using the freely available Perseus software (version 1.3.10.0). Identified proteins were filtered as follows: contaminants, hits to the reverse database and proteins only identified with modified peptides were eliminated. Additionally, at least 2 unique or razor peptides were required per protein. Label-free intensities were logarithmized, then samples were grouped into triplicates and identifications were filtered to require three valid values in at least one group. To enable statistical analysis, missing values were imputed with values representing a normal distribution around the detection limit of the mass spectrometer. To that end, mean and standard deviation of the real distribution of intensities were determined, then a new distribution with a downshift of 1.8 standard deviations and a width of 0.3 standard deviations was created. Interacting proteins were identified by performing two-sample *t*-tests with permutation-based false discovery rate (FDR) statistics essentially as previously described (Eberl et al., 2013; Hubner et al., 2010; Keilhauer et al., 2015; Tusher et al., 2001; van Nuland

et al., 2013a). 250 permutations were performed and different FDR values required (see table below). The parameter S_0 was empirically optimized to separate outliers from the background.

	GFP-H2A.Z.1		GFP-H2A.Z.2		GFP-PWWP2A	
	FDR	S_0	FDR	S_0	FDR	S_0
SK-mel147 Exp. 1	0.2	0.8	0.15	1.0	-	-
SK-mel147 Exp. 2	0.08	1.0	0.1	1.0	-	-
HeLa Kyoto Exp. 5	0.05	1.0	0.05	1.0	-	-
HeLa Kyoto Exp. 1	-	-	-	-	0.0005	1.0

2.5.1.1 GFP-histone variant MNas ChIPs

As a first step, background binders were removed by employing a two sample *t*-test (without FDR statistics) and keeping only proteins that were enriched (*t*-test difference > 0) in the GFP-histone IPs (GFP-H2A, -H2A.Z.1, -H2A.Z.2) compared to the control IP (GFP-only). After GFP-background filtering, all remaining proteins were subjected to a second two sample *t*-tests (including FDR statistics), now comparing GFP-H2A.Z.1 respectively -H2A.Z.2 with GFP-H2A to identify interacting proteins specifically enriched on H2A.Z.1- or H2A.Z.2-containing mononucleosomes versus H2A-containing mononucleosomes. The respective *p*-values (technical replicates) and *t*-test differences were plotted against each other in volcano plots using R (3.0.2). Proteins within the cutoff line (based on the FDR and S_0) were selected as promising candidates for further investigations.

2.5.1.2 GFP-PWWP2A MNase ChIPs

GFP-PWWP2A IPs were compared to GFP-only IPs employing a two sample t-test (including FDR statistics), *p*-values (of technical triplicates) and *t*-test differences plotted against each other in volcano plots using R and interacting proteins visualized by including a cutoff line representing both the pre-determined FDR of 0.00005 and the *t*-test difference.

2.5.1.3 Heatmaps

In addition to the volcano plots, interacting proteins were also visualized in heatmaps. Therefore, the ratio (fold change) of the logarithmized and averaged triplicate LFQ intensities of GFP-histones (H2A, H2A.Z.1 and H2A.Z.2) and GFP-only was hierarchically clustered and plotted as heatmaps with Perseus. Heatmaps of proteins interacting with GFP-PWWP2A just visualize the respective logarithmized and averaged LFQ intensities.

2.5.2 MNase ChIP-seq analyses

2.5.2.1 Read mapping, generation of coverage vectors and peak calling

Raw sequencing reads were mapped to the human genome (hg19) using bowtie (version 0.12.9) omitting reads with more than one match. Coverage vectors were created after read extension to 150 bp corresponding to the expected fragment size after MNase digestion.

Peak calling was performed on the pooled replicates of each target against the pooled input chromatin libraries using Homer (v4.7) applying parameters -style histone, -fragLength 150 and -inputFragLength 150. Peak annotations, genomic feature enrichment statistics and gene ontology enrichment analysis were obtained using the annotatePeaks script of the Homer package.

2.5.2.2 Heatmaps and cumulative plots

Heatmaps were generated by sampling coverage values in 5 bp bins in a 2 kb window around features of interest and subsequent smoothing by a 9-bin moving average. Heatmaps were plotted after sorting the windows based on either the average PWWP2A signal in the central 1 kb window or the expression level of the corresponding gene measured by microarray.

For TSS-centered heatmaps at genes with PWWP2A peaks, genes were selected based on whether a Homer-called peak has been matched to its promoter. For heatmaps around intergenic PWWP2A peaks selection was based on defining peaks that do not map closer than 2 kb to genes. Cumulative plots were obtained by averaging the signals across all genes.

H3K4me3 ChIP-sequencing data were downloaded from the Gene Expression Omnibus data base: GEO accession GSM733682.

2.5.3 PWWP2A conservation

The full PWWP2A amino acid sequence (NCBI reference sequence NP_001124336.1 (NM_001130864.1) or Uniprot identifier Q96N64-1) was searched employing the web browser-based NCBI Basic Local Alignment Search Tool (BLAST). BLAST search was executed utilizing default settings and 500 aligned sequences were displayed. Hits down to a query covery of 70% (100 hits) were selected to calculate a distance tree with BLAST by pairwise alignment applying default settings. The tree was converted into a circular phylogram with the CLC Main Workbench (7.0) software.

2.5.4   PWWP2A protein structure prediction

A potential structure of the PWWP domain of PWWP2A was modeled employing the web browser-based tool iTASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and visualized with the freely available software Chimera (1.8.0). Template was the already published structure of the PWWP2B PWWP domain (Qin and Min, 2014), Protein Data Bank (PDB) code 4LD6. The predicted PWWP2A structure was compared to the published structure of the DNA-binding PWWP domain of PSIP1 (Eidahl et al., 2013), PDB code 4FU6). The Electrostatic Surface Potential (ESP) of PSIP1 and PWWP2A was calculated utilizing the Coulombic Surface Coloring algorithm, which is part of the Chimera (1.8.0) software package, using default settings.

3   Results

3.1   The H2A.Z nucleosome interactome

In order to shed light on the rather enigmatic functions of H2A.Z in chromatin biology and to elucidate whether the variants H2A.Z.1 and H2A.Z.2 are functionally distinct, a label-free quantitative mass spectrometry (qMS) approach was established (**Figure 3.1.1**) to identify potential H2A.Z nucleosome interacting proteins.

Nuclei isolation

↓   *in replicates: GFP, GFP-H2A, GFP-H2A.Z.1 & GFP-H2A.Z.2*

MNase digest

↓   *mononucleosomes (S1)*

ChIP

↓   *GFP-trap beads*

Tryptic digest

↓   *on-beads*

MS: Q Exactive

↓

Analyses	Software	Number of proteins in qMS run		
		SK-mel147 Exp. 1	SK-mel147 Exp. 2	HeLa Kyoto Exp. 5
Identification & LFQ*	*MaxQuant*	1668	1479	1518
Filtering**	*Perseus*	1305	1078	1216
t-test vs. GFP	*Perseus*	893	571	881
t-test vs. H2A	*Perseus*	**Z.1**: 334 / **Z.2**: 450	**Z.1**: 322 / **Z.2**: 332	**Z.1**: 495 / **Z.2**: 383
Volcano plots with FDR threshold line	*R*	**Z.1**: 55 / **Z.2**: 69	**Z.1**: 59 / **Z.2**: 52	**Z.1**: 55 / **Z.2**: 77

Figure 3.1.1 Label-free chromatin interaction proteomics workflow. Nuclei were isolated under physiological conditions from cell lines expressing GFP-only or GFP-H2A, GFP-H2A.Z.1 or GFP-H2AZ.2 and subjected to MNase digestion. Mononucleosomes were salt-extracted, immunoprecipitated in triplicates with GFP-trap beads and precipitated material digested on-beads with trypsin. Tryptic peptides were cleaned and concentrated with StageTips, peptides eluted and analyzed by LC-MS/MS employing a Q Exactive mass spectrometer. (*) MS raw files were processed using MaxQuant version (1.3.9.2) and label-free quantification (LFQ) enabled. (**) Filtering and analyses of LFQ proteomics data were performed using Perseus (1.3.10.0). For details see Materials and Methods section 2.5.1. A first two-sample *t*-test allowed to remove background binders by keeping only proteins that were enriched in the GFP-histone IPs (GFP-H2A, -H2A.Z.1, -H2A.Z.2) compared to the control IP (GFP). A second two-sample *t*-test implementing a permutation-based false discovery rate (FDR) compared GFP-H2A.Z.1 or GFP-H2A.Z.2 with GFP-H2A and allowed for the identification of proteins specifically enriched on nucleosomes containing GFP-H2A.Z.1 or GFP-H2A.Z.2.

This approach had been successfully applied to discover histone tail peptide interactors (Eberl et al., 2013) but this is the first time that this technique is utilized to determine interaction partners of mononucleosomes derived from cells.

To prepare nuclei and chromatin under physiological conditions and maintain nuclear integrity and chromatin compaction, a salt extraction method was applied to MNase-digested chromatin that was already published almost 40 years ago (Sanders, 1978) and rediscovered and further developed in the last years (Henikoff et al., 2009; Sansoni et al., 2014). The extraction method (for details see 2.4.4) initially tested with chromatin from different stable HeLa Kyoto cell lines yielded three fractions that differed in chromatin composition as well as in protein content (**Figure 3.1.2**)
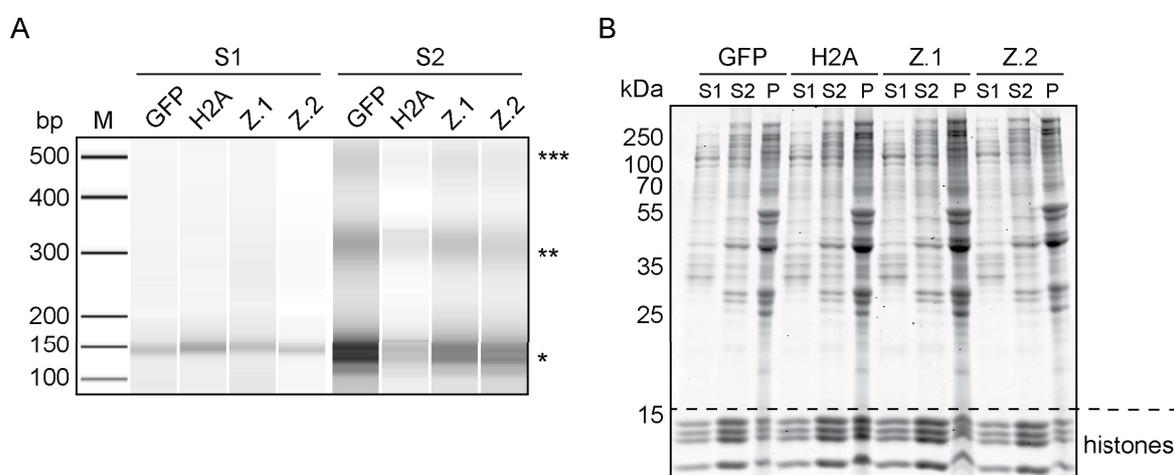
Figure 3.1.2 Evaluation of MNase digest conditions. MNase digested chromatin was prepared form nuclei isolated from HeLa Kyoto cell lines stably expressing GFP-only (GFP), GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) or GFP-H2A.Z.2 (Z.2). S1 fractions are obtained after 20 min incubation with 1.5 U/µl MNase at 26°C. S2 fractions are obtained by ON incubation with a resuspension buffer at RT. The remaining pellet (P) fraction was cooked in Laemmli buffer and subsequently digested with benzonase (500 U) for 2 hours. **(A)** DNA obtained from chlorophorm/isoamylacohol extraction and ethanol precipitation of the S1 and S2 fractions was loaded on a 1000 bp DNA BioAnalyzer chip. S1 fractions contained almost pure mononucleosomes (*), the S2 fractions also contained di- (**) and trinucleosomes (***). **(B)** The protein content of all fractions was analyzed by SDS-PAGE and Coomassie staining. All fractions contained the core histones (below dashed line) but differed in their respective protein content.

The protocol was optimized to obtain mainly mononucleosomes contained in a first fraction (S1), representing open and therefore highly MNase-accessible chromatin (**Figure 3.1.2A**). A second fraction (S2) could be obtained after ON incubation, as well as a pelleted fraction (P), both containing higher order chromatin that was less to little MNase-accessible (**Figure 3.1.2A** only shows S1 and S2 DNA). All three fractions contained histones but differed in the overall protein content (**Figure 3.1.2B**). In order to enable identification of proteins

interacting with mononucleosomes and sequencing of equal MNase ChIP material, the S1 fraction was chosen for subsequent assays (**Figure 3.1.1**).

First quantitative mass spectrometry results were obtained with mononucleosomes from the metastatic melanoma-derived cell line SK-mel147. Experimental work of our collaborator Chiara Vardabasso (Emily Bernstein group at the Mount Sinai in New York) uncovered that H2A.Z isoforms are overexpressed in patients with metastatic melanoma and are correlated with a poor survival prognosis. RNAi-mediated knockdown of H2A.Z.2 but not H2A.Z.1 led to impaired proliferation of different metastatic melanoma cell lines (including SK-mel147) and transcriptome analyses revealed, that H2A.Z.2 regulates transcription of cell cycle regulatory genes in these cells. In order to unravel the mechanism behind H2A.Z.2 functionality in melanoma biology, I initially established H2A.Z isoform interactome data from SK-mel147 chromatin (**Figure 3.1.1**). Moreover, cervical cancer-derived HeLa Kyoto cell lines were utilized to confirm the quantitative mass spectrometry results in a different cell line and served as cellular systems for further biochemical assays.

Aiming for a H2A.Z nucleosome interactome, nuclei were isolated from SK-mel147 cells stably expressing GFP-only or GFP-tagged H2A, H2A.Z.1 or H2A.Z.2 (Vardabasso et al., 2015) (under revision) and digested with micrococcal nuclease (MNase) to obtain chromatin in fraction S1 fragmented to almost 100% into mononucleosomes (Sansoni et al., 2014), (**Figure 3.1.3A**). Subsequently, mononucleosomes containing the GFP-tagged histone variants were immunoprecipitated with GFP-trap beads in triplicates and subjected to on-beads trypsin digestion. Analyzing 10% of the IP material before tryptic digests by SDS-PAGE and Coomassie staining revealed the successful enrichment of the GFP-tagged histone variant, the presence of all core histones as well as several interacting factors. In addition, beads cooked in 1x Laemmli buffer after tryptic on-beads digest were loaded to monitor the successful digestion (IP*) (**Figure 3.1.3B**). In collaboration with Eva Keilhauer, a PhD student in Matthias Mann's group at the MPI of Biochemistry in Munich, tryptic peptides were analyzed on a Q Exactive mass spectrometer. Identification of immunoprecipitated proteins and label-free quantification was carried out using MaxQuant software (1.3.9.2) (Eberl et al., 2013; Luber et al., 2010) and lead to the identification of >1000 proteins (**Figure 3.1.1** and **Figure 3.1.3C**).
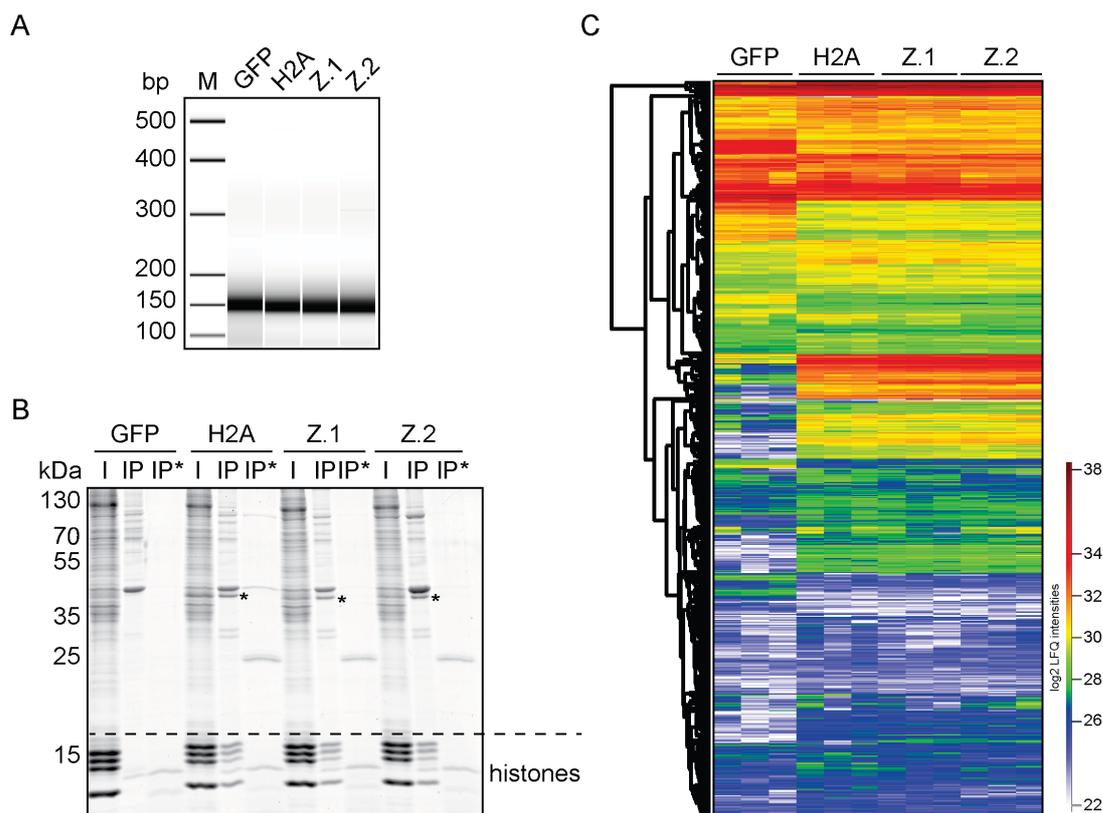
Figure 3.1.3 First SK-mel147 MNase ChIP experiment followed by qMS.
(A) MNase digested chromatin was prepared form nuclei isolated from SK-mel147 cell lines stably expressing GFP, GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) or GFP-H2A.Z.2 (Z.2). MNase digestion yielded almost pure mononucleosomes (~150 bp) as controlled by a 1000 bp DNA BioAnalyzer chip. **(B)** Mononucleosomes containing GFP-tagged histone variants and GFP-only control (I = Input S1 fractions) were subjected to immunoprecipitations utilizing magnetic GFP-trap beads. 1% Input, 10% of the IP material and 50% of the beads after tryptic elution (IP*) were analyzed by SDS-PAGE and Coomassie staining, revealing the successful pull down of histones (around 15 kDa), the GFP-tagged variant (* ~40 kDa) and interacting proteins as well as a complete tryptic digest (band ~13 kDa: GFP-trap) **(C)** Tryptic IP peptides were analyzed by LC-MS/MS employing a Q Exactive mass spectrometer, MS raw data processed by MaxQuant and logarithmized label-free quantification (LFQ) intensities of 1305 proteins enriched by replicate IPs visualized using hierarchical clustering in a heatmap by Perseus.

The freely available software Perseus (1.3.10.0) was utilized for initial filtering and further analysis (see Materials and Methods section 2.5.1 and **Figure 3.1.1**). On the finally processed dataset a hierarchical row clustering was performed that allowed to distinguish the GFP-histone IPs from GFP-control IPs (**Figure 3.1.3C**) and revealed allover good replicate quality. Next, background binders were removed by keeping only proteins that were enriched in the GFP-histone IPs compared to the control IPs. Two sample t-tests comparing GFP-H2A.Z.1 (or GFP-H2A.Z.2, respectively) with GFP-H2A provided (technical) p-values and t-test differences that were plotted in volcano plots using R (3.0.2) to visualize association of co-

precipitated proteins with either GFP-H2A.Z.1- or GFP-H2A.Z.2-containing nucleosomes versus GFP-H2A-containing nucleosomes (**Figure 3.1.4**).

Figure 3.1.4 Identification of proteins interacting with GFP-H2A.Z.1 and GFP-H2A.Z.2 nucleosomes in SK-mel147 cell lines. Proteins interacting with GFP-H2A.Z.1 or GFP-H2A.Z.2 nucleosomes (right half) were separated from proteins enriched on GFP-H2A nucleosomes (left half) by plotting *p*-values and *t*-test differences obtained by two-sample *t*-test in volcano plots with R, highlighting interesting candidates with a threshold line based on FDR-statistics (H2A.Z.1: FDR = 0.2, S_0 = 0.8; H2A.Z.2: FDR = 0.15, S_0 = 1). Colored and labeled are only GFP-H2A.Z.1 / Z.2 nucleosome interactors that were reproduced in a second SK-mel147 screen (see Appendix **Figure A.1**); green: H2A.Z, blue: SRCAP complex, red = candidates, yellow = candidates, but below threshold. For a complete interactor list see Appendix **Table A.1**.

The thresholds for reliable interactions were set based on FDR statistics and the ratio between GFP-H2A.Z and GFP-H2A (*t*-test difference) (see table in section 2.5.1). Accordingly, 55 proteins were identified to be enriched on H2A.Z.1 nucleosomes and 69 proteins to be enriched on H2A.Z.2 nucleosomes in this first experiment.

The comparison with a second independent MNase ChIP-qMS experiment (Appendix **Figure A.1**) enabled us to identify 43 proteins including H2A.Z that were reproducibly enriched on GFP-H2A.Z nucleosomes over GFP-H2A nucleosomes (**Table 3.1.1**, **Figure 3.1.5A** and **B**).

Table 3.1.1 SK-mel147 H2A.Z nucleosome interactors, complex affiliations and functions

Complex	Complex members	Function
SRCAP	ACTR6, DMAP1, RUVBL1, RUVBL2, SRCAP, VPS72, YEATS4, ZNHIT1	H2A.Z deposition
p400/NuA4/TIP60	DMAP1, EP400, RUVBL1, RUVBL2, TRRAP, VPS72, YEATS4	Histone acetylation and H2A.Z deposition
MLL	MLL/KMT2A, WBP7/MLL2, HCFC2, MEN1, RBBP5	H3K4 (tri)methylation
NuRD	HDAC2, MTA1, RBBP4, RBBP7	Chromatin remodeling and histone deacetylation
Cullin E3 Ligase	CUL4A, CUL4B, BRWD3, NEDD8,	Protein ubiquitination
No name given yet	HMG20A, PHF14, RAI1, TCF20	Repelled by H3K4me3 (Eberl et al., Mol. Cell, 2013)
Not assigned to complexes yet	BAHD1	Transcriptional repressor (BAH domain)
	BCORL1	Transcriptional corepressor (3 ANK repeats)
	BRD2	Transcriptional regulator (2 bromodomains)
	CDYL	Transcriptional corepressor (chromdomain)
	DIDO1	Transcription factor (PHD finger)
	KDM2A	Histone demethylase (PHD finger)
	MAGEA10	Unknown (MAGE-domain)
	MIER1	Transcriptional repressor (ELM2, SANT)
	MTA1-3	Unknown, MTA1 isoform (BAH, ELM2, SANT)
	MYPOP	Transcriptional repressor (Myb-like domain)
	PHF14	Unknown (2 PHD finger)
	PHF14-2	Unknown, PHF14 isoform 2 (2 PHD finger)
	PHF20L1	DNMT1 stabilization (PHD finger, 2 tudordomains)
	PWWP2A	Unknown (PWWP domain)
	ZFX/Y	Unknown (13 zinc finger)
	ZNF512B	Unknown (7 zinc finger)

Underlined proteins found also in HeLa Kyoto qMS screen. Chromatin-related domains are depicted in brackets.

Within this group we consistently found eight out of ten described members of the SWR1-related ATP-dependent chromatin remodeling complex SRCAP (blue dots in volcano plots in **Figure 3.1.4** and Appendix **Figure A.1D**; see also **Table 3.1.1** and **Figure 3.1.5A**) that exchanges nucleosomal H2A-H2B dimers for free H2A-H2A.Z dimers in mammalian cells (Ruhl et al., 2006), which confirmed the accuracy and reliability of the assay. Actin could not be identified and Actin-like protein 6A (ACTL6A) did not fulfill the selection criteria, but was identified and highlighted in volcano plots and heat maps for the sake of completeness (**Figure 3.1.4** and Appendix Figure **A.1D**). MNase IPs followed by immunoblotting validated

this result: ZNHIT1, a member of the SRCAP complex, bound to GFP-H2A.Z-containing but not GFP-H2A-containing nucleosomes (**Figure 3.1.5C**).
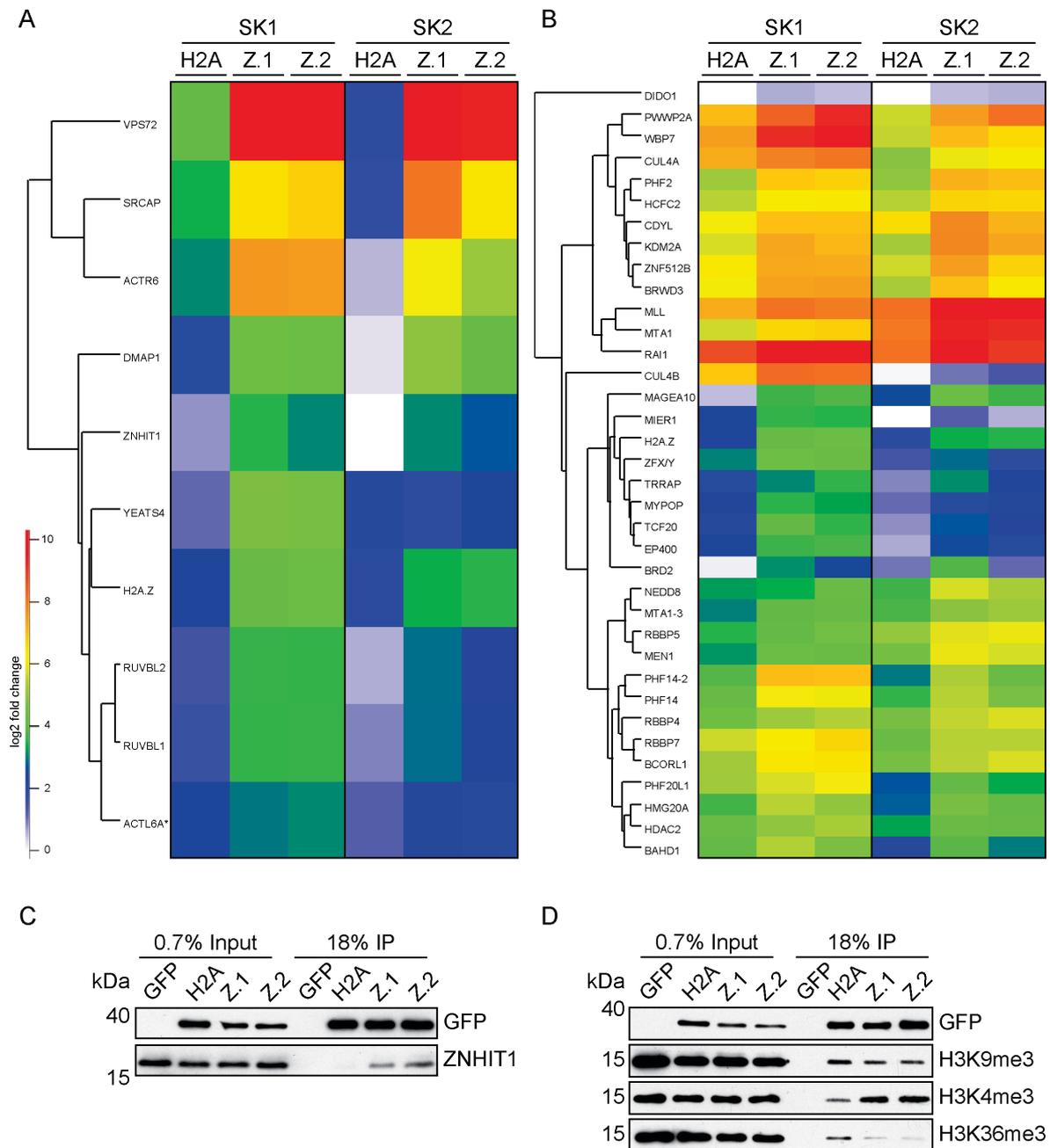
Figure 3.1.5 Interactor-enrichment on GFP-H2A.Z.1 and GFP-H2A.Z.2 versus GFP-H2A nucleosomes revealed in both SK-mel147 qMS screens and nucleosome PTM-analysis. The ratio (fold change) of the logarithmized and averaged triplicate LFQ intensities of GFP-histones (H2A, H2A.Z.1 and H2A.Z.2) and GFP-only was hierarchically clustered and plotted as heatmap with Perseus. The log2 fold change is presented for the members of the SRCAP complex (**A**) and all other GFP-H2A.Z nucleosome interactors (**B**). (**C**) SRCAP complex member ZNHIT1 interaction with GFP-H2A.Z nucleosomes was validated by MNase IPs with chromatin derived from SK-mel147 cell lines stably expressing GFP, GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) or GFP-H2A.Z.2 (Z.2) followed by immunoblotting. (**D**) PTMs associated with immunoprecipitated nucleosomes containing GFP-H2A, GFP-H2A.Z.1 or GFP-H2A.Z.2 are shown by immunoblotting.

Several components of the SRCAP complex are also present in a second mammalian H2A.Z depositing complex, the p400/NuA4/TIP60 complex (Billon and Cote, 2013). Surprisingly, only two out of ten specific members of this larger complex were precipitated (**Table 3.1.1**): Transformation/transcription domain-associated protein (TRRAP) and E1A-binding protein p400 (EP400); but not BRD8, EPC1/2, MEAF6, ING3, TIP60, MRG15, MRGBP, and MRGX that we and other have found to be specific components (Billon and Cote, 2013; Bonisch et al., 2012).

Furthermore, several members of the nucleosome remodeling and deacetylase (NuRD) complex were identified: the Histone deacetylase 2 (HDAC2), the Histone-binding proteins 4 and 7 (RBBP4 and RBBP7) and the Metastasis-associated protein 1 (MTA1) that supports recruitment of the NuRD complex (Allen et al., 2013); but not HDAC1, MBD2/3, CHD3/4/5 and GATAD2A/B. Moreover, members of the Mixed lineage leukemia 1 and 2 (MLL/KMT2A and WBP7/MLL2) complexes (van Nuland et al., 2013a) were found to interact with GFP-H2A.Z nucleosomes: Retinoblastoma-binding protein 5 (RBBP5), Host cell factor 2 (HCFC2) and Menin 1 (MEN1). Also, the Cullin E3 ligases 4A and 4B (CUL4A and CUL4B) were present on GFP-HA.Z nucleosomes, as well as the ubiquitin-like protein NEDD that was reported to covalently attach to Cullin ligases (Lydeard et al., 2013). Interestingly, a previously identified but yet unnamed complex that was reported to be repelled by H3K4me3 (Eberl et al., 2013) was identified to enrich on H2A.Z nucleosomes. It consists of the High mobility group protein 20A (HMG20A), PHD finger protein 14 (PHF14), Retinoic acid-induced protein 1 (RAI1) and Transcription factor 20 (TCF20). In addition, several proteins were discovered that are chromatin-associated but either not yet assigned to any complex or not yet fully characterized (**Table 3.1.1** and **Figure 3.1.5B**). Among them are many factors associated with transcriptional regulation like Bromo adjacent homology domain-containing protein 1 (BAHD1), Mesoderm induction early response protein 1 (MIER1), Bromodomain-containing protein (BRD2) or Death-inducer obliterator 1 (DIDO1). However, also proteins with unknown function like PWWP domain-containing protein 2A (PWWP2A) or Melanoma-associated antigen 10 (MAGEA10) are present.

Importantly, the interaction of BRD2 with H2A.Z nucleosomes in melanoma cells is part of the collaborative work with the group of Emily Bernstein already mentioned before. Here we show that BRD2 acts in concert with H2A.Z.2, acetylated H4 and transcription factor E2F1 to regulate melanoma cell proliferation, thereby influencing cancer progression. Currently, this second-author contribution is in the experimental revision process for publication in Molecular Cell.

To further characterize H2A.Z-specific chromatin regions, MNase IPs were followed by immunoblots probing for different posttranslational histone modifications. The results revealed an enrichment of H3K4me3 on GFP-H2A.Z-containing nucleosomes, whereas H3K9me3 and H3K36me3 were reduced compared to GFP-H2A nucleosomes (**Figure 3.1.5D**). This was already shown before (Kim et al., 2013) and is in line with the identification of members of the MLL complexes that establish H3K4me3, as well as with KDM2A associating with H2A.Z nucleosomes, which demethylates H3K36 (Tsukada et al., 2006) and PHF2 that removes H3K9 methyl groups (Wen et al., 2010)

Combining the data of both biological replicates it was possible to identify the enrichment of certain proteins on nucleosomes with one specific H2A.Z-variant (**Figure 3.1.6**). Thus, 14 proteins appeared to bind preferentially to GFP-H2A.Z.1-containing nucleosomes (most prominent BRD2), whereas 6 proteins seemed to prefer GFP-H2A.Z.2-containing nucleosomes (most prominent PWWP2A). Notably, and in accordance with results from the group of Cheung (Draker et al., 2012), BRD2 is enriched on nucleosomes that contain GFP-H2A.Z.1 in comparison to those containing GFP-H2A.Z.2.
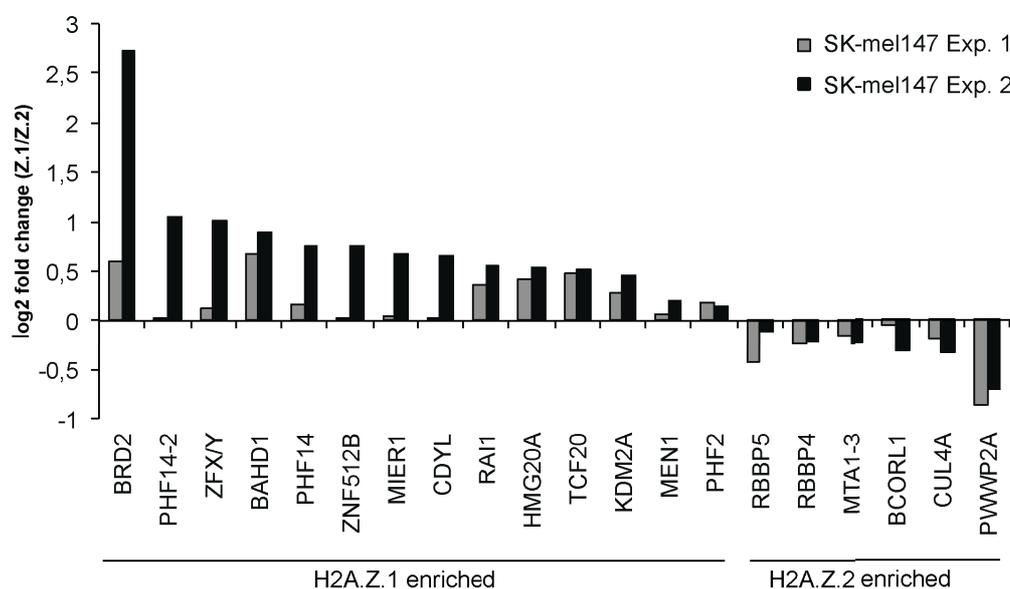
Figure 3.1.6 Preferential interaction of proteins with either GFP-H2A.Z.1 or GFP-H2A.Z.2 nucleosomes. The ratio (fold change) of the logarithmized and averaged triplicate LFQ intensities of GFP-H2A.Z.1 and GFP-H2A.Z.2 was plotted in bar plots with Excel demonstrating a potential H2A.Z isoform preference of some binding proteins.

In summary, I have established a quantitative label-free mass spectrometry assay to identify nucleosome interactors. I could show that in SK-mel147 cells several known complexes associate with H2A.Z nucleosomes, but also uncharacterized proteins were identified. Interestingly, the overlap between proteins interacting with H2A.Z.1 and H2A.Z.2 nucleosomes is substantially high (**Table 3.1.1**), still allowing for some preferential interactions (**Figure 3.1.6**) and also unique binding factors (**Table A.1**).

Next, I wondered whether the identified H2A.Z nucleosome interacting proteins are cell type specific or rather general H2A.Z nucleosome binders. Therefore, the same screening workflow (**Figure 3.1.1**) was applied to HeLa Kyoto cells stably expressing GFP-only or GFP-H2A, -H2A.Z.1 or H2A.Z.2 (Bonisch et al., 2012) (**Figure 3.1.7**).
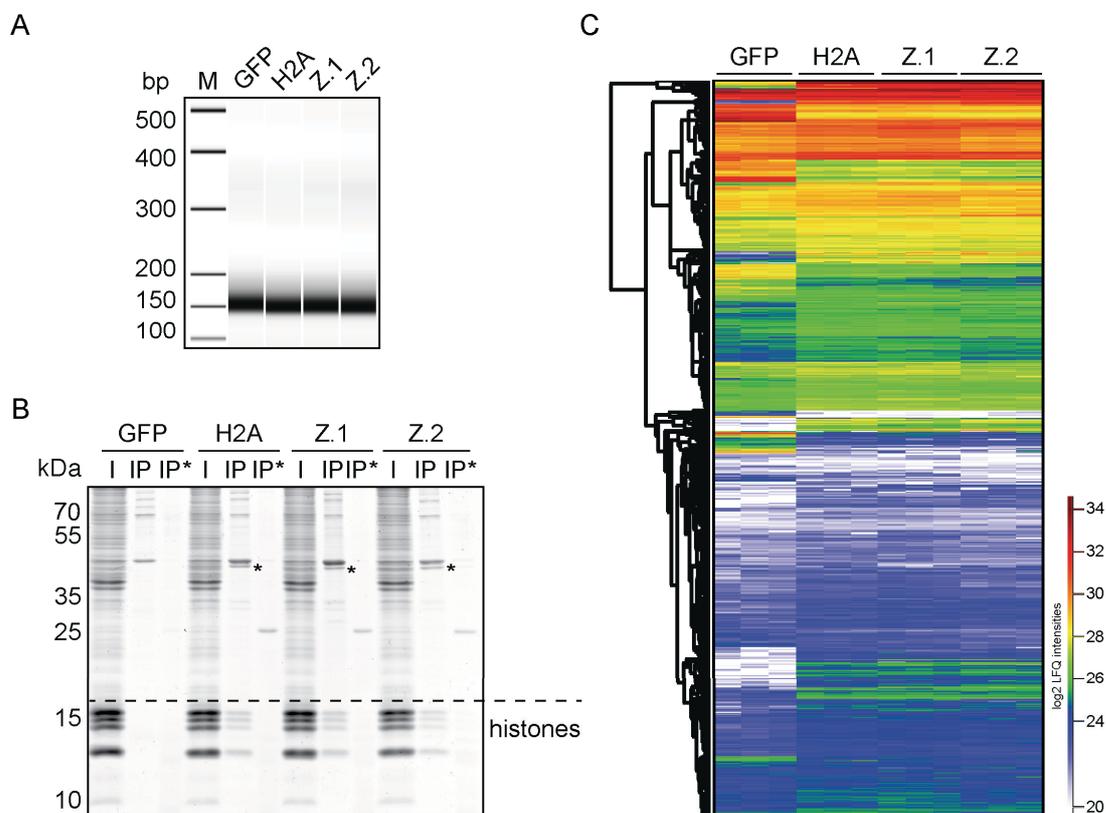
Figure 3.1.7 MNase ChIP experiment followed by qMS with HeLa Kyoto cell lines. **(A)** MNase digested chromatin was prepared form nuclei isolated from HeLa Kyoto cell lines stably expressing GFP, GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) or GFP-H2A.Z.2 (Z.2). MNase digestion yielded almost pure mononucleosomes (~150bp) as controlled by a 1000 bp DNA BioAnalyzer chip. **(B)** Mononucleosmes containing GFP-tagged histone variants and GFP-only control (I = Input S1 fractions) were subjected to immunoprecipitations utilizing magnetic GFP-trap beads. 1% Input, 10% of the IP material and 50% of the beads after tryptic elution (IP*) were analyzed by SDS-PAGE and Coomassie staining, revealing the successful pull down of histones (around 15 kDa), the GFP-tagged variant (* ~40 kDa) and interacting proteins as well as a complete tryptic digest. **(C)** Tryptic IP peptides were analyzed by LC-MS/MS employing a Q Exactive mass spectrometer, MS raw data processed by MaxQuant and logarithmized label-free quantification (LFQ) intensities of 1216 proteins enriched by replicate IPs visualized using hierarchical clustering in a heatmap by Perseus.

Isolated nuclei from these cells were MNase digested, fragmented chromatin prepared and the MNase digestion degree of S1 fractions monitored by agarose gel electrophoresis (**Figure 3.1.7A**). Similar to the previous experiments with SK-mel147 cells, mononucleosomes were obtained in roughly 100% purity with only minor contaminations in di-, and trinucleosomes. This material was then used for IPs with GFP-trap beads, success of IPs and subsequent tryptic on-beads digest were controlled by SDS-PAGE and Coomassie staining (**Figure 3.1.7B**) and samples ultimately analyzed by mass spectrometry as described for the melanoma cell lines (**Figure 3.1.7C** and **Figure 3.1.8**).

Unfortunately, the Hela Kyoto experiments turned out to be more difficult than expected. The identification of SRCAP complex members, serving as a quality criterion based on the previous SK-mel147 experiments, as well as of other chromatin-related proteins was hampered by the enrichment of various background binders. In some experiments, only the H2A.Z.1 results looked promising, in others only the H2A.Z.2 results, but the reasons for this variability remained despite intense trouble-shooting unresolved. In the end, the last out of five HeLa Kyoto qMS experiments (**Figure 3.1.8**) passed the quality control judged by the presence of SRCAP complex members (blue dots in volcano plots in **Figure 3.1.8**; see also **Table 3.1.2**). Interestingly, a general overlap of interactors found in both, the pioneering SK-mel147 experiments and this HeLa Kyoto experiment was observed (red dots in volcano plots in **Figure 3.1.8** and underlined proteins in **Table 3.1.1**). However, not all candidates derived from the melanoma experiments were also found to be H2A.Z nucleosome binders in HeLa Kyoto and amongst those, that are also present in the HeLa Kyoto experiment, not all were identified on both GFP-H2A.Z variant nucleosomes as summarized in **Table 3.1.2**. Candidates are also presented in heatmaps in comparison with GFP-H2A and with respect to their log2 fold enrichment over the GFP control (**Figure 3.1.9**). Moreover, all identified candidates associating with H2A.Z.1 or H2A.Z.2 nucleosomes in HeLa Kyoto cells are listed in the Appendix section (**Table A.2**).
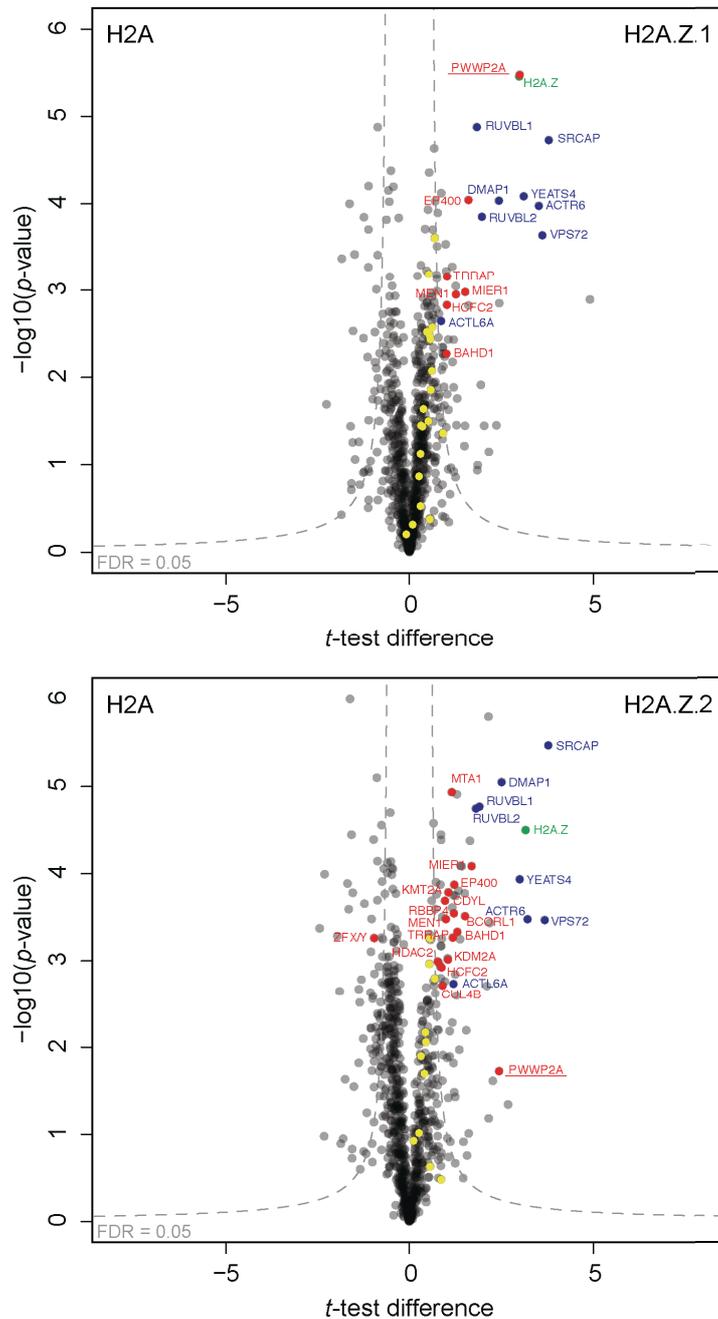
Figure 3.1.8 Identification of proteins interacting with GFP-H2A.Z.1 and GFP-H2A.Z.2 nucleosomes derived from HeLa Kyoto cell lines. Proteins interacting with GFP-H2A.Z.1 or GFP-H2A.Z.2 nucleosomes (right half) were separated from proteins enriched on GFP-H2A nucleosomes (left half) by plotting p-values and t-test differences obtained by two-sample t-test in volcano plots with R, highlighting interesting candidates with a threshold line based on FDR-statistics (FDR = 0.05, S_0 = 1). Colored and labeled are only GFP-H2A.Z.1 / Z.2 nucleosome interactors that were among the 36 reproduced candidates identified by both SK-mel147 qMS screens; green: H2A.Z, blue: SRCAP complex, red = candidates, yellow = candidates, but below threshold. For a complete list please see Appendix **Table A.2**.

Table 3.1.2 Hela Kyoto H2A.Z nucleosome interactors and complex affiliation

Complex	Complex members on Z.1	Complex members on Z.2
SRCAP	ACTR6, ACTL6A, DMAP1, RUVBL1, RUVBL2, SRCAP, VPS72, YEATS4	ACTR6, ACTL6A, DMAP1, RUVBL1, RUVBL2, SRCAP, VPS72, YEATS4
p400/NuA4/TIP60	DMAP1, EP400, RUVBL1, RUVBL2, TRRAP, VPS72, YEATS4	DMAP1, EP400, RUVBL1, RUVBL2, TRRAP, VPS72, YEATS4
MLL	MEN1, HCFC2	MLL/KMT2A, MEN1, HCFC2
NuRD		HDAC2, MTA1, RBBP4
Cullin E3 Ligase		CUL4B
Not assigned to complexes yet	BAHD1, MIER1, PWWP2A	BAHD1, BCORL1, CDYL, KDM2A, MIER1, PWWP2A

Table only contains interactors also found in both SK-mel147 screens. For a detailed list of interactors see Appendix **Table A.2**.
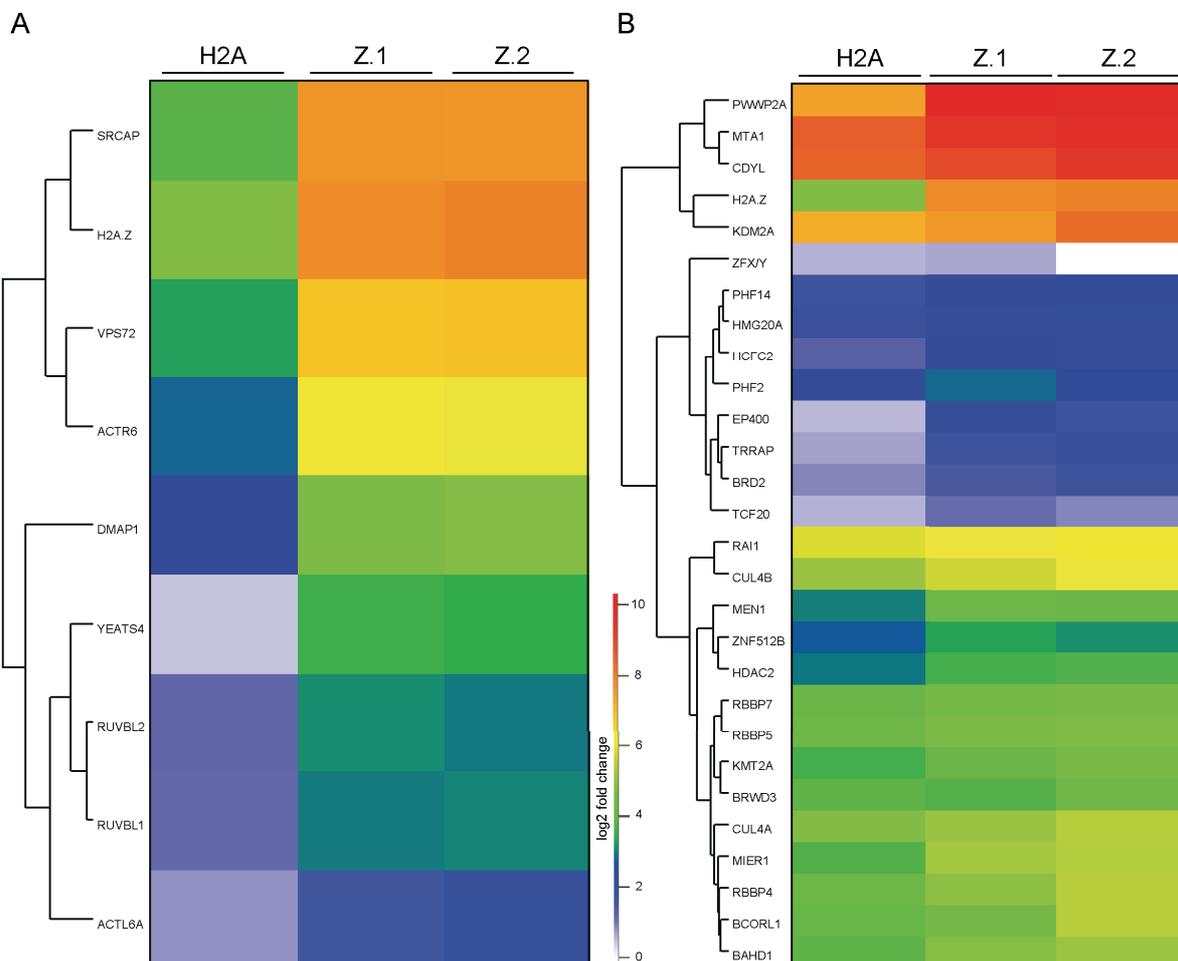
Figure 3.1.9 Interactor-enrichment on GFP-H2A.Z.1 and GFP-H2A.Z.2 versus GFP-H2A nucleosomes revealed in HeLa Kyoto qMS screen. The ratio (fold change) of the logarithmized and averaged triplicate LFQ intensities of GFP-histones (H2A, H2A.Z.1 and H2A.Z.2) and GFP was hierarchically clustered and plotted as heatmap with Perseus. The log2 fold change is presented for the members of the SRCAP complex **(A)** and all GFP-H2A.Z nucleosome interactors reproducibly identified in both SK-mel147 qMS screens **(B)**.

ZNHIT1, a member of the SRCAP complex that was present in the melanoma screens, could not be identified in the presented HeLa Kyoto GFP-H2A.Z interactome. However, it was featured in previous HeLa Kyoto experiments (data not shown) and could also be detected by MNase IPs followed by immunoblotting (**Figure 3.1.10**). ACTL6, a member of the SRCAP complex that did not pass the threshold set for the SK-mel147 screens, was now present in the list of HeLa Kyoto GFP-H2A.Z interactors. In accordance with the melanoma results, two specific p400/NuA4/TIP60 members (EP400 and TRRAP) were identified. Also MLL1 complex members were again found to be enriched on GFP-H2A.Z nucleosomes. Members of the NuRD complex were also identified but this time only with H2A.Z.2-containing nucleosomes. Three proteins not assigned to any complex that were present on SK-mel147 GFP-H2A.Z nucleosomes were also identified in the HeLa Kyoto experiment: BAHD1, MIER1 and PWWP2A. Similar to the SK-mel147 experiments, PTMs associated with the precipitated GFP-H2A.Z nucleosomes were detected by immunoblotting (**Figure 3.1.10**). The rather active state of the SK-mel147 GFP-H2A.Z chromatin could be recapitulated also in the HeLa Kyoto cells: GFP-H2A.Z nucleosomes are decorated with H3K4me3 and have reduced levels of H3K9me3 compared to GFP-H2A nucleosomes.
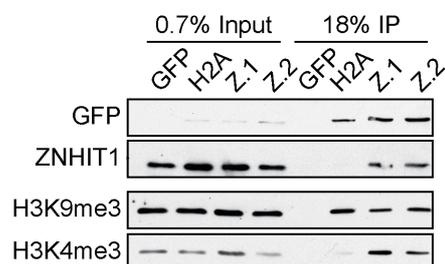
Figure 3.1.10 Presence of ZNHIT1 and two histone PTMs on GFP-H2A.Z and GFP-H2A nucleosomes in HeLa Kyoto cell lines. SRCAP complex member ZNHIT1 was shown to interact with GFP-H2A.Z nucleosomes by MNase IPs with chromatin of HeLa Kyoto cell lines stably expressing GFP-only (GFP), GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) and GFP-H2A.Z.2 (Z.2) followed by immunoblotting although not identified in the HeLa Kyoto qMS screen. MNase IPs and immunoblotting demonstrated H3K9me3 and H3K4me3 decoration of immunoprecipitated HeLa Kyoto nucleosomes containing GFP-H2A, GFP-H2A.Z.1 or GFP-H2A.Z.2.

In conclusion, the HeLa Kyoto qMS results recapitulate the presence of several major protein complexes and interesting uncharacterized proteins interacting with H2A.Z nucleosomes as observed in the SK-mel147 qMS screens.

3.2 PWWP2A interacts with H2A.Z nucleosomes

Having, for the first time, been able to identify a large cohort of proteins to interact with GFP-H2A.Z nucleosomes in a cell type-independent manner, it was important to next validate these results with further biochemical assays. Here, the PWWP2A protein was chosen, because (1) it was found to associate with H2A.Z in both cell lines, (2) it harbors a PWWP domain shown in other proteins to facilitate chromatin-binding through different modes (see below) and (3) has not been studied by other groups yet.

PWWP2A consists of 755 aa, has a molecular mass of roughly 82 kDa and is a protein of so far unknown function. The NCBI and UniProt databases predict the existence of one canonical and two additional PWWP2A isoforms but are not consistent in their nomenclature. NCBI refers to canonical PWWP2A that was identified in the H2A.Z qMS screen as isoform b (NCBI reference sequence NP_001124336.1 / NM_001130864.1). However, UniProt depicts it as isoform 1 (UniProt identifier Q96N64-1). Phylogenetic analyses determine PWWP2A's conservation among vertebrates (Appendix **Figure A.2**), arguing for an important function within these species. It consists of two proline-rich stretches (P1: aa 61 – 146 and P2: aa 240 – 291) and one serine-rich stretch (S: aa 575 – 632) and belongs to the PWWP domain family of proteins, as it possesses a PWWP domain (PWWP: aa 655 – 715) close to its C-terminus (**Figure 3.2.1**). The domain was named after the central core motif that was found in the first PWWP-domain containing protein discovered, the Wolf-Hirschhorn syndrome candidate protein 1 (WHSC1): (P) proline - (W) tryptophane - (W) tryptophane - (P) proline (Stec et al., 1998) and is mostly involved in chromatin binding of more then 20 human PWWP domain containing proteins (Qin and Min, 2014).
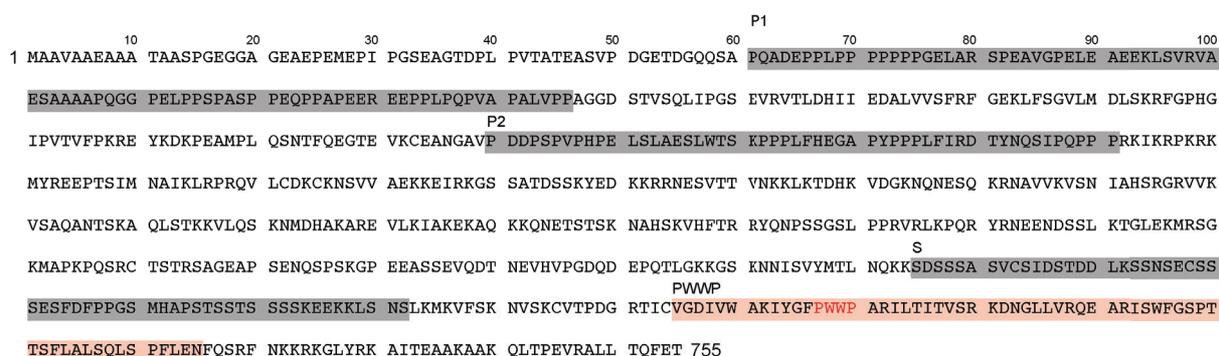
Figure 3.2.1 Primary structure of PWWP2A. PWWP2A consists of 755 amino acids and possesses two proline-rich stretches (P1: aa 61 – 146 and P2: aa 240 – 291, grey) and one serine-rich stretch (S: aa 575 – 632) highlighted with grey boxes. PWWP2A belongs to the PWWP domain-containing protein family because it harbors a PWWP domain (PWWP: aa 655 – 715) close to its C-terminus. The core PWWP domain is boxed in light red, the (P) proline - (W) tryptophane - (W) tryptophane - (P) proline motif is highlighted in red. NCBI reference sequence NP_001124336.1 (NM_001130864.1) or Uniprot identifier Q96N64-1.

Together with the Tudor domain, the MBT domain and the chromodomain, the PWWP domain belongs to the "Royal Family" (Maurer-Stroh et al., 2003) whose common feature is to bind histones with methylated lysine or arginine residues (Adams-Cioaba and Min, 2009). Indeed, also PWWP domain containing proteins were shown to possess histone lysine methylation affinity, especially towards H3K36me3 (Dhayalan et al., 2010; Eidahl et al., 2013; Li et al., 2013; Vezzoli et al., 2010), but also towards H4K20me3 and H3K79me3 (Wu et al., 2011). The PWWP domain of certain proteins was also shown to mediate contacts with DNA, like for example the PWWP domains of PC4 and SFRS1 interacting protein 1 (PSIP1) or Hepatoma-derived growth factor (HDGF) (Qin and Min, 2014)

First, PWWP2A binding to H2A.Z nucleosomes needed to be validated in an assay supplementary to qMS. Therefore, MNase ChIPs (HeLa Kyoto and SK-mel147 mononucleosomes) followed by immunoblotting with a commercially available PWWP2A antibody were performed. Consistently with the qMS results, PWWP2A showed a preferred association with GFP-H2A.Z-containing nucleosomes in comparison to GFP-H2A-containing ones (**Figure 3.2.2A** and **Figure 3.2.2B**). Wondering whether hyperacetylation of nucleosomes would enhance PWWP2A-binding to nucleosomes as it was shown for BRD2 (Draker et al., 2012), cells were treated with the histone deacetylase inhibitor trichostatin A (TSA) prior to chromatin isolation and MNase digestion. Enhanced H3K27ac levels demonstrated successful HDAC inhibition (**Figure 3.2.2C**).
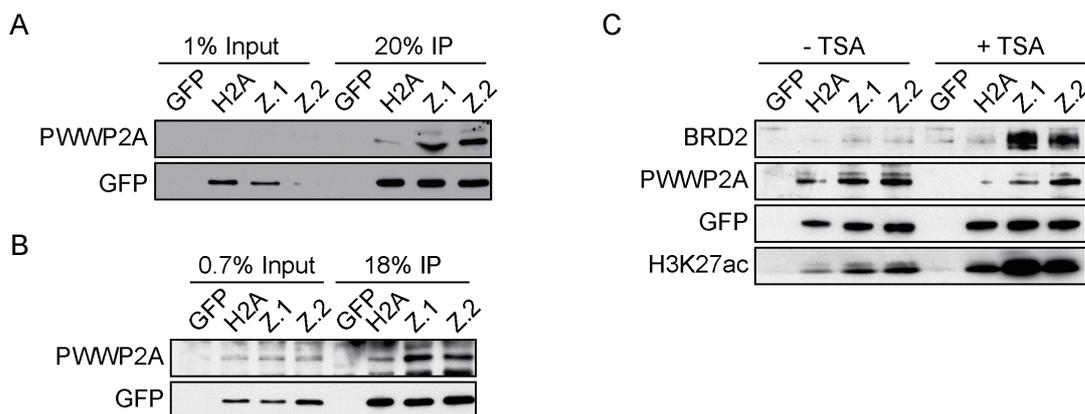
Figure 3.2.2 PWWP2A preferentially interacts with H2A.Z nucleosomes in SK-mel147 and HeLa Kyoto cell lines independent of histone acetylation. MNase IPs with chromatin from **(A)** SK-mel147 and **(B)** HeLa Kyoto cell lines stably expressing GFP, GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) and GFP-H2A.Z.2 (Z.2) were followed by immunoblotting. Detection of the GFP-tagged histone variant and PWWP2A validated PWWP2A's preferential presence on GFP-H2A.Z nucleosomes as observed by qMS. **(C)** MNase IPs with chromatin from SK-mel147 cell lines treated with the HDAC inhibitor TSA followed by immunoblotting demonstrated PWWP2A independence on hyperacetylated GFP-H2A.Z nucleosomes. Checking H3K27 acetylation-levels controlled for hyperacetylation induced by TSA. Probing for BRD2 served as a control for enhanced protein binding as this protein is attracted by hyperacetylated H2A.Z nucleosomes.

Indeed, immunoprecipitated GFP-tagged variant-containing nucleosomes from TSA treated SK-mel147 cells showed an increase in BRD2 binding compared to control cells (**Figure 3.2.2C**). However, no PWWP2A enrichment could be detected on hyperacetylated nucleosomes. Interestingly, PWWP2A binding to GFP-H2A.Z nucleosomes was rather reduced by enhanced histone acetylation levels.

In order to gain insights into the cellular function of PWWP2A, its coding sequence was amplified from SK-mel147 cDNA, subcloned into a pT7blue-3 shuttle vector and then transferred into the pIRESneo-eGFP vector by Gateway cloning. This vector allowed for stable expression of a N-terminal GFP-PWWP2A fusion protein in HeLa Kyoto cells (**Figure 3.2.3A**). FACS analysis of a G418-selected uniform cell population (HeLa Kyoto GFP-PWWP2A#5.2) showed a rather mild GFP-PWWP2A overexpression. Confocal microscopy revealed the solely nuclear localization of GFP-PWWP2A (interphase chromatin, **Figure 3.2.3B**). Interestingly, this localization was maintained also during mitosis (metaphase chromatin, **Figure 3.2.3B**), which argues for a comparably strong association of PWWP2A with chromatin.
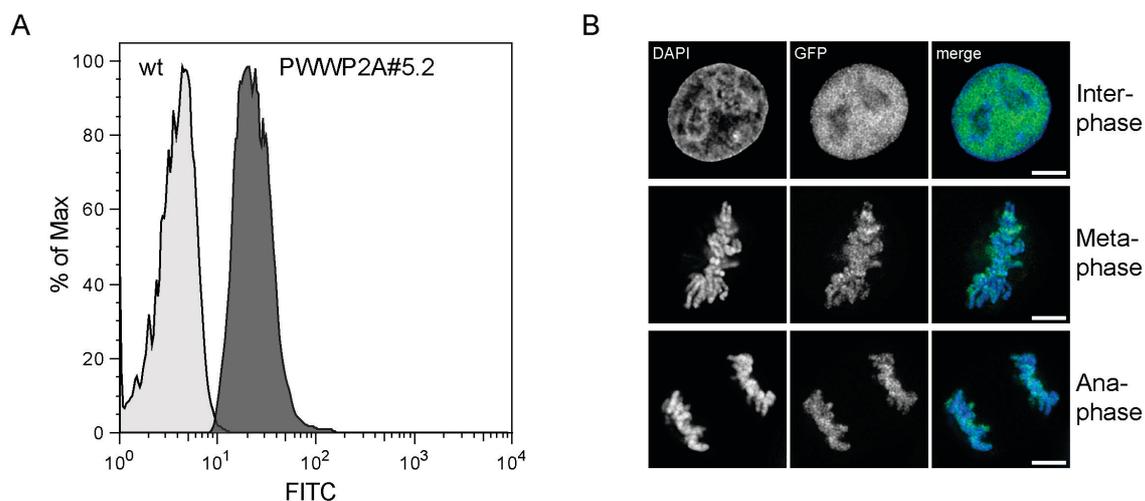
Figure 3.2.3 GFP-PWWP2A is stably expressed in HeLa Kyoto cells and localizes to interphase and mitotic chromatin. HeLa Kyoto cells were transfected with the pIRESneo-eGFP-PWWP2A construct and selected with G418. **(A)** Comparative FACS analysis of wild type (wt) HeLa Kyoto cells and the selected GFP-PWWP2A cell line revealed a homogenous cell population mildly overexpressing GFP-PWWP2A (PWWP2A#5.2). GFP-fluorescence was detected in the FITC channel and plotted in a histogram. **(B)** Fixation on coverslips and DNA counterstaining with DAPI of stable GFP-PWWP2A HeLa Kyoto cells followed by SP5 confocal microscopy showed stable association of GFP-PWWP2A with interphase but also meta- and anaphase chromatin. Confocal images were deconvolved utilizing Huygens Essentials (4.4) and a recorded point spread function. Scale bars = 10 µm.

To validate the interaction of PWWP2A with H2A.Z-containing nucleosomes MNase ChIPs followed by immunoblotting were carried out with mononucleosomes isolated from the stable

HeLa Kyoto GFP-PWWP2A cell line, including a GFP and a GFP-H2A.Z.1 cell line as controls. Indeed, nucleosomes co-precipitated with PWWP2A and predominantly contained endogenous H2A.Z compared to H2A as revealed by immunoblotting (**Figure 3.2.4**). Interestingly, H2A.Z nucleosomes interacting with PWWP2A were not decorated with H3K4me3, which was not due to an absence of general H3 (**Figure 3.2.4**). This is interesting, because the initial reciprocal experiment suggested, that H2A.Z nucleosomes could possess this PTM and bind PWWP2A (**Figure 3.1.8**, **Figure 3.1.10** and **Figure 3.2.4**). However, the result confirms the observation of Eberl et al., who found PWWP2A to be repelled by histone H3 tail peptides containing trimethylated lysine 4 (Eberl et al., 2013). Importantly, no chemical crosslinkers were necessary in this ChIP experiment, supporting the finding of PWWP2A interacting rather stable with chromatin as previously also shown by confocal microscopy and observed in qMS experiments.
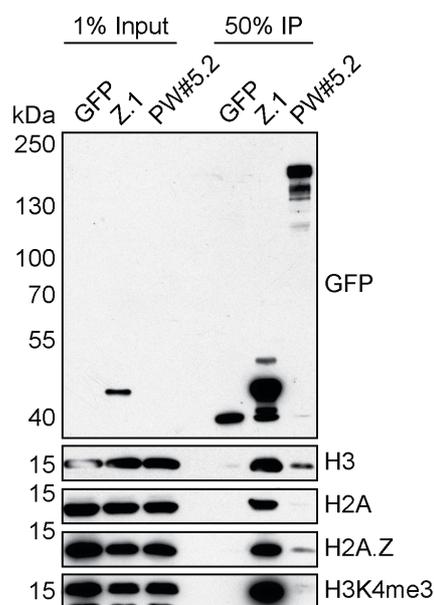
Figure 3.2.4 GFP-PWWP2A preferentially interacts with H2A.Z nucleosomes lacking H3K4 trimethylation. MNase ChIPs with chromatin from the HeLa Kyoto cell lines stably expressing GFP-PWWP2A (PW#5.2) GFP-H2A.Z.1 (Z.1) and GFP were followed by immunoblotting detecting histones H2A, H2A.Z and H3 as well as the PTM H3K4me3. PWWP2A's preferential interaction with H2A.Z nucleosomes could be validated, but these nucleosomes were depleted in H3K4me3. Control MNase ChIPs utilizing chromatin from the HeLa Kyoto GFP-H2A.Z.1 cell line recapitulated the enrichment of H3K4 trimethylation.

Based on these results a native MNase ChIP-sequencing approach was utilized to investigate the genome-wide *in vivo* localization of GFP-PWWP2A with respect to sites of H2A.Z-variant containing chromatin. Briefly, mononucleosomes from stable HeLa Kyoto cell lines expressing GFP-tagged H2A, H2A.Z.1, H2A.Z.2 and PWWP2A were generated as described for the qMS experiments and subjected to GFP immunoprecipitations. Sequencing libraries were prepared from immunoprecipitated DNA and input DNA of two independent biological replicates and sequenced by Illumina sequencing in Oslo (Norwegian Sequencing Center) or Munich (LAFUGA). Histone variant MNase ChIP-sequencing was also carried out using SK-mel147 cell lines, but was implemented into the collaborative study with the Bernstein group

mentioned earlier (data not shown). Notably, our collaborators validated the GFP-ChIP results and determined the localization of endogenous H2A.Z utilizing an isoform-unspecific H2A.Z antibody. Thus, the epitope tag-based results reflect well the physiological state. For statistical and bioinformatic analyses we collaborated with Dr. Tobias Straub, head of the bioinformatics core facility of our institute. Generally, genome wide GFP-PWWP2A peaks were found to overlap in great parts with GFP-H2A.Z (**Figure 3.2.5A**): 66% of all GFP-PWWP2A peaks are shared by GFP-H2A.Z.1 and 44% are shared by GFP-H2A.Z.2. A feature enrichment analysis revealed that GFP-PWWP2A peaks were mainly associated with genic regions (TSS, TTS, 5'UTR and exons) (**Figure 3.2.5B**). Except for the TTS, GFP-H2A.Z.1 and GFP-H2A.Z.2 were also found predominantly in genic regions. At these sites GFP-PWWP2A seemed to be slightly enriched compared to the GFP-H2A.Z isoforms. The general enrichment of all three proteins at genic sites was expected, as the ChIP input material predominantly contained MNase accessible, less compacted chromatin. In this line, peaks in intergenic regions that might posses a more compact chromatin structure appeared underrepresented (**Figure 3.2.5B**).
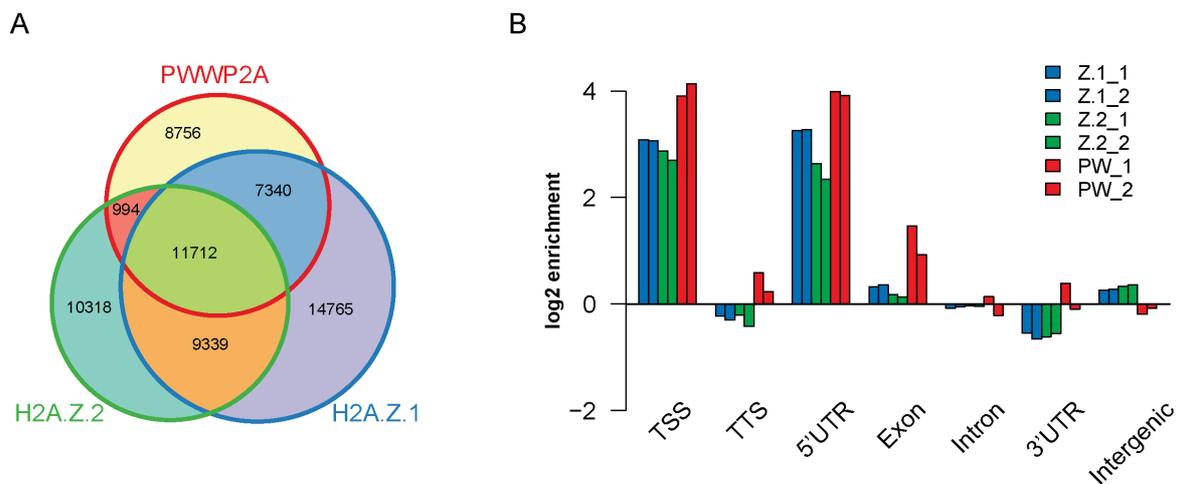
Figure 3.2.5 Peak overlap between and feature enrichments of GFP-PWWP2A and GFP-H2A.Z.1 / GFP-H2A.Z.2 nucleosome genome-wide locations. (A) Peaks identified from native MNase ChIP-sequencing employing mononucleosomes from stable GFP-H2A.Z.1, GFP-H2A.Z.2 and GFP-PWWP2A cell lines showed a large but not complete spatial overlap. Cumulative data of two biological replicates. **(B)** Feature enrichments suggest an accumulation of GFP-H2A.Z.1 (Z.1), GFP-H2A.Z.2 (Z.2) and GFP-PWWP2A (PW) in genic regions, especially at the promoter of genes. Peak calling and feature enrichment statistics were carried out using the Homer (4.7) software package. Two biological replicates are shown (_1 and _2).

In order to learn more about the local distribution of the immunoprecipitated proteins, a closer look was taken on genomic features like intergenic regions, gene bodies or the TSS of genes. Three representative snapshots are shown in **Figure 3.2.6**, depicting the localization of GFP-

PWWP2A, -H2A.Z.1 and -H2A.Z.2 as uncovered by replicate ChIP-sequencing. The close-ups demonstrate that GFP-PWWP2A is found at specific sites within the presented genomic features and does not span larger chromatin domains. Moreover, it closely overlapped with GFP-H2A.Z.1 and GFP-H2A.Z.2 peaks in these regions, suggesting a correlation with H2A.Z-containing nucleosomes.
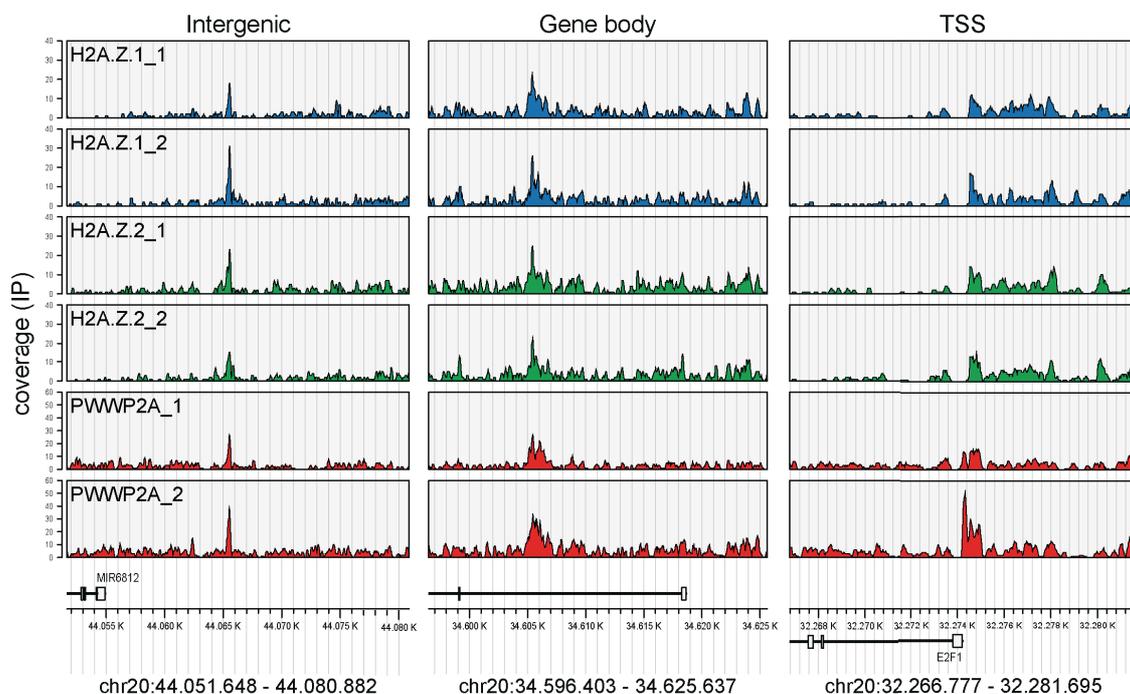
Figure 3.2.6 Peak overlap of GFP-PWWP2A with GFP-H2A.Z.1 and GFP-H2A.Z.2 at specific genomic sites. Snapshots of exemplary intergenic, gene body and TSS regions showed the close overlap of GFP-PWWP2A with GFP-H2A.Z.1 and -H2A.Z.2. Two biological replicates are shown (_1 and _2). Depicted below the snapshots are the respective genomic coordinates on human chromosome 20.

Next, promoters of genes (**Figure 3.2.7** top) or intergenic regions (**Figure 3.2.7** bottom) that were occupied by GFP-PWWP2A were sorted according to GFP-PWWP2A peak intensity (strongest peaks at the top). The respective GFP-PWWP2A sequencing read densities of both biological replicates were then plotted in heatmaps and contrasted with the heatmaps of GFP-H2A.Z.1 and GFP-H2A.Z.2 sorted accordingly. In the heatmaps, promoters were depicted centered to the respective TSS of the gene with a 1 kb up- and downstream window. This kind of representation illustrated the NDR at the center of the TSS that is surrounded by GFP-H2A.Z-containing nucleosomes (**Figure 3.2.7** top). Interestingly, GFP-PWWP2A also surrounded the NDR and seemed to predominantly occupy the regions downstream of the NDR (see also later in the text). This tendency is also found generally for GFP-H2A.Z, maybe slightly more pronounced for GFP-H2A.Z.2. Intergenic regions were defined as peaks that did not map closer than 2 kb to genes and were aligned at their center in the heatmaps (**Figure**

3.2.7 bottom). GFP-PWWP2A occupied also these regions and showed a colocalization with GFP-H2A.Z.1 and GFP-H2A.Z.2 However, the GFP-PWWP2A distribution appeared less broad compared to GFP-H2A.Z.
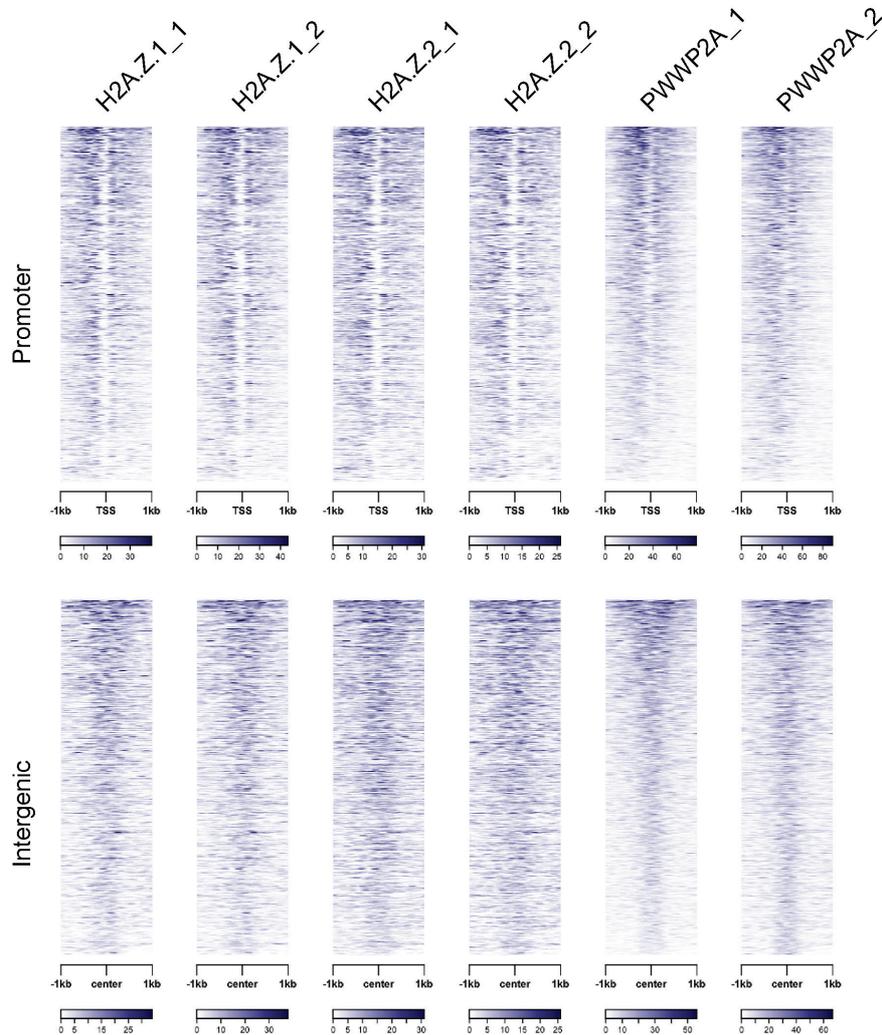
Figure 3.2.7 Promoter and intergenic regions occupied by GFP-PWWP2A overlap with GFP-H2A.Z.1 and GFP-H2A.Z.2. Read densities at promoters of genes (top) or intergenic regions (bottom) that are GFP-PWWP2A, GFP-H2A.Z.1 or GFP-H2A.Z.2 targets were depicted in heatmaps and sorted according to the PWWP2A peak intensity (strongest at the top). Promoters were TSS centered with a 1 kb up- and downstream window; Intergenic domains were defined as peaks that did not map closer than 2 kb to genes and then centered. Two biological replicates are shown (_1 and _2). Below the heatmaps: distance to center of the respective feature, heatmap color scale depicting read density.

The localization of GFP-PWWP2A to promoters of genes suggested a potential involvement in transcriptional regulation as it was postulated for H2A.Z. However, it is highly debated, whether and how localization of H2A.Z to the TSS of genes directly fosters gene transcription. Many studies have, on the first sight, conflicting results, depending on the gene (locus), cell type or differentiation state, and correlate H2A.Z with gene activation or

repression (Marques et al., 2010). The histone mark H3K4me3 is also found at the TSS of genes and is reported to associate especially with active genes (Bannister and Kouzarides, 2011). In order to investigate a potential role for PWWP2A in transcriptional regulation, all genes were sorted with respect to their expression levels (from top to bottom) in HeLa Kyoto cells as obtained by microarray transcriptome analysis (data not shown). Again, GFP-PWWP2A sequencing read densities were plotted in heatmaps and contrasted with the heatmaps of GFP-H2A.Z.1 / -H2A.Z.2 that were sorted accordingly (**Figure 3.2.8**).
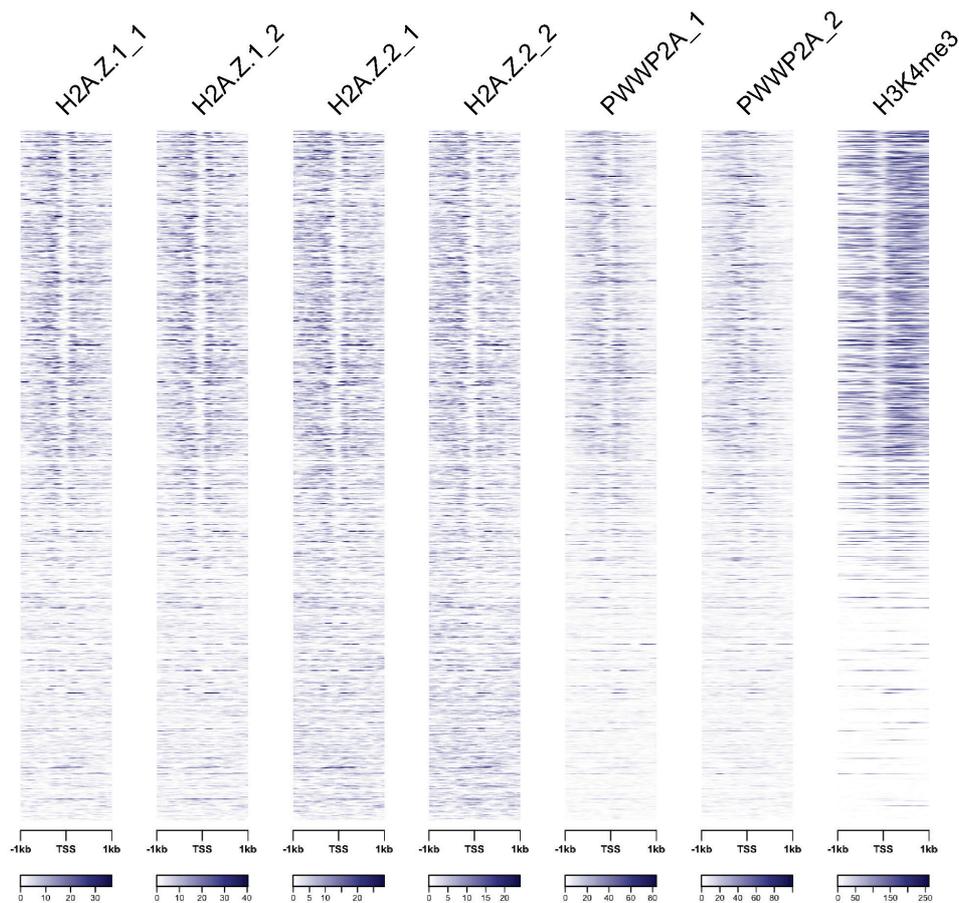
Figure 3.2.8 GFP-PWWP2A is associated with the TSS of active genes. Read densities at promoters of all sequenced genes were depicted in heatmaps and sorted according to gene expression levels obtained by HeLa Kyoto microarray analysis (from top to bottom: high expressed to low expressed). Included were also H3K4me3 ChIP-sequencing data from HeLa cells (Gene Expression Omnibus data base: GEO accession GSM733682) further highlighting actively transcribed genes. Promoters were TSS centered with a 1 kb up- and downstream window. Two biological replicates are shown (_1 and _2). Below the heatmaps: distance to center of the respective feature, heatmap color scale depicting read density

In addition, available H3K4me3 ChIP-sequencing data from HeLa cells (Gene Expression Omnibus data base: GEO accession GSM733682) were plotted in promoter heatmaps according to HeLa Kyoto gene expression (far right heatmap in **Figure 3.2.8**). As expected,

H3K4me3 is indeed found at promoters of active genes (**Figure 3.2.8** upper part of the heatmaps) but not at promoters of 'silent' genes (**Figure 3.2.8** lower part of the heatmaps), confirming the accurate listing of genes according to their gene expression levels. Strikingly, GFP-PWWP2A strongly associated with promoters of actively transcribed genes (**Figure 3.2.8**). Also, both GFP-H2A.Z variants localized to promoters of actively transcribed genes. Focusing only on the promoters of actively transcribed genes, the occupancy of GFP-PWWP2A and the GFP-H2A.Z isoforms at these promoters was depicted in a cumulative plot (**Figure 3.2.9**).
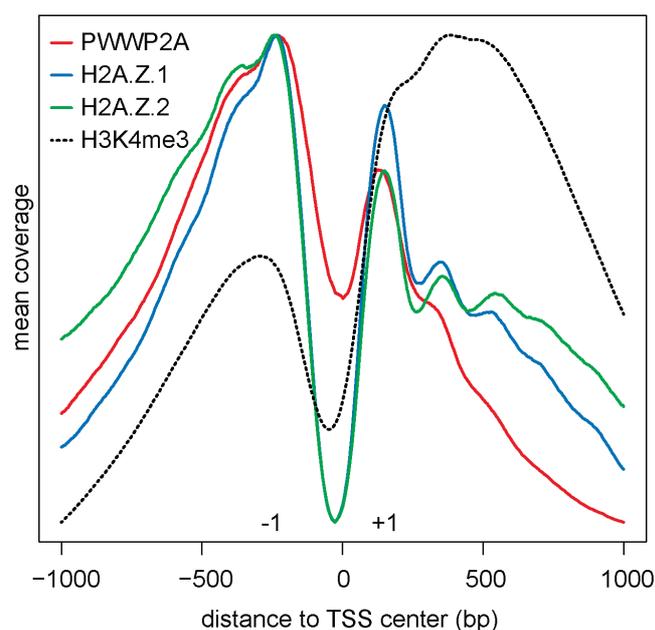
Figure 3.2.9 Nucleosomes flanking the NDR are occupied by GFP-PWWP2A, but also by GFP-H2A.Z.1 and GFP-H2A.Z.2. A cumulative plot of all actively transcribed genes is shown. GFP-PWWP2A, GFP-H2A.Z.1, GFP-H2A.Z.2 and H3K4me3 ChIP-sequencing signals were averaged across all active genes (mean coverage) and displayed centered to the TSS. Nucleosomes flanking the NDR are depicted as -1 and +1 nucleosomes. Combined data from two biological replicates are shown. H3K4me3 ChIP-sequencing data taken from the Gene Expression Omnibus data base (GEO accession GSM733682).

The available data from H3K4me3 ChIP-sequencing were again included into the analysis. This form of visualization made it immediately apparent that GFP-PWWP2A was particularly prominent on -1 and +1 nucleosomes flanking the NDR as it was already observed in the promoter heatmaps (**Figure 3.2.7**). Here, it tightly colocalizes with GFP-H2A.Z.1 and GFP-H2A.Z.2. H2A.Z's localization to the nucleosomes flanking the NDR was already published (Barski et al., 2007) (see also **Figure 1.3.3**) but is now for the first time separately shown for both H2A.Z isoforms. Interestingly, GFP-PWWP2A predominantly mapped to the -1 nucleosome (red line in **Figure 3.2.9**) and this is also true for the GFP-H2A.Z isoforms, albeit more pronounced for GFP-H2A.Z.2 (green line in **Figure 3.2.9**). Also, the NDR obtained from GFP-PWWP2A is not as prominent as the NDR from GFP-H2A.Z. However, this might be due to differences in the MNase digestion degree.

Notably, the tendency of GFP-PWWP2A (and GFP-H2A.Z) to occupy -1 nucleosomes is counteracted by the clear enrichment of H3K4me3 on +1 nucleosomes (dashed black line in

Figure 3.2.9). This might suggests a functional discrimination of the NDR-flanking nucleosomes and could also explain, why nucleosomes interacting with PWWP2A tend to accumulate less H3K4me3 as seen previously by immunoblotting of immunoprecipitated GFP-PWWP2A (**Figure 3.2.4**). The observed -1 and +1 nucleosome occupancy of GFP-PWWP2A, GFP-H2A.Z.1 and GFP-H2A.Z.2 were successfully validated by independent MNase ChIP experiments and qPCR. Here, several target genes were chosen that showed an enrichment of all three proteins around the TSS (PARS2, RNF11, UHMK1, NUF2 and SWT1/TRMT1L) (Appendix **Figure A.3A**). Results of the qPCR validation are shown in Appendix **Figure A.3B.**

In conclusion, the presented MNase ChIP-sequencing analyses argue for a correlated localization of GFP-PWWP2A and GFP-H2A.Z.1 / -H2A.Z.2 at the TSS of genes, thereby underlining the observed interaction of PWWP2A with H2A.Z nucleosomes. Moreover, the strong association with transcriptionally active genes suggests a role for PWWP2A in transcriptional regulation.

3.3 An internal region of PWWP2A but not its PWWP domain is sufficient to mediate nucleosome binding *in vitro*

PWWP2A being strongly associated with chromatin raised the question of how binding to nucleosomes in general and H2A.Z-containing nucleosomes in particular is achieved. To answer this question, an experimental strategy combining recombinant PWWP2A protein and mononucleosomes prepared from cells was employed. Together with Nina Wommelsdorf, a master student in our laboratory whom I supervised, full-length PWWP2A N-terminally fused with glutathione S-transferase (GST-PWWP2A_fl) was recombinantly expressed in BL21 *E. coli* cells (**Figure 3.3.1A**). After purification with gluthatione sepharose beads, GST-tagged PWWP2A_fl still immobilized to the beads or beads only were incubated with mononucleosomes isolated from HeLa Kyoto cells. Immunoblotting of the precipitated material revealed that GST-PWWP2A_fl is able to pull down nucleosomes containing H2A.Z (**Figure 3.3.1B**), thereby confirming the prior results and demonstrating the feasibility of this *in vitro/ex vivo* assay. Since the PWWP domain of other proteins interacts with chromatin, it is a strong candidate region responsible for nucleosome interaction. To test this hypothesis, two truncations of the full-length GST-PWWP2A_fl construct were established, then expressed in *E. coli* and purified: GST-PWWP2A_NT that lacks the PWWP domain and the

very C-terminus, and GST-PWWP2A_CT that just consists of the PWWP domain plus the C-terminus (**Figure 3.3.1A**).
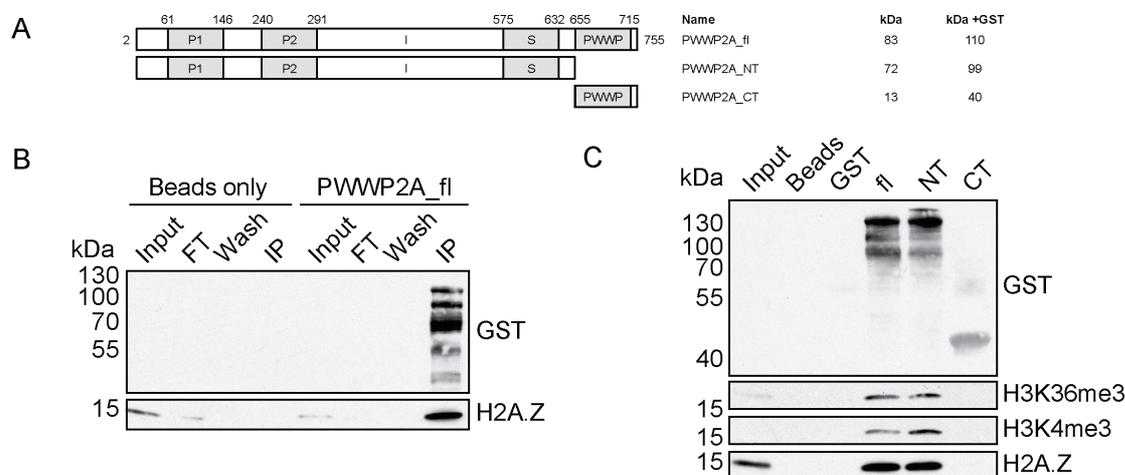
Figure 3.3.1 GST-PWWP2A interacts with HeLa Kyoto (H2A.Z) nucleosomes independent of the PWWP domain and is enriched on H3K4me3-depleted nucleosomes (A) Schematic of PWWP2A wild type and domain deletion proteins. The expected molecular weight of every construct with and without GST-tag is indicated in kDa. **(B)** GST-PWWP2A_fl immobilized to sepharose beads was incubated with mononucleosomes derived from HeLa Kyoto cells, precipitated material was separated by SDS-PAGE and GST and H2A.Z detected by immunoblotting. **(C)** GST-tagged PWWP2A_fl (fl), PWWP2A_NT (NT) and PWWP2A_CT (CT) were incubated with HeLa Kyoto mononucleosomes followed by immunoblotting detecting GST, H3K36me3, H3K4me3 and H2A.Z. Beads-only (Beads) and GST-only (GST) served as negative controls.

GST-IPs with HeLa Kyoto mononucleosomes were performed as described above and analyzed by immunoblotting. Surprisingly, GST-PWWP2A_NT was able to pull down (H2A.Z-containing) nucleosomes although lacking the PWWP domain (**Figure 3.3.1C**). In fact, the interaction with GST-PWWP2A_NT was comparable to the interaction with GST-PWWP2A_fl. Interestingly, the PWWP domain alone (GST-PWWP2A_CT) did not precipitate nucleosomes, arguing for its inability to bind nucleosomes *in vitro* or at least suggesting very weak or transient interaction (**Figure 3.3.1C**). This suggested, that PWWP2A is competent of interacting with nucleosomes independent of its PWWP domain, at least *in vitro*. Notably, nucleosomes precipitated with the full-length protein were not enriched for H3K36me3 compared to the mutant lacking the PWWP domain (**Figure 3.3.1C**) Furthermore, preliminary data suggest that also H4K20me1 and H4K20me3 are not present on these nucleosomes (data not shown). However, H3K4me3 appeared to be enriched on nucleosomes pulled down with GST-PWWP2A_NT, suggesting that the PWWP domain rejects binding of nucleosomes modified on H3 lysine 4 (**Figure 3.3.1C**). This is in

accordance with the result from *in vivo* MNase pull downs described earlier (**Figure 3.2.4**) and the already mentioned study published by Eberl et al., which shows PWWP2A to be repelled by histone H3 tail peptides trimethylated on K4 and attracted by unmodified H3 tail peptides (Eberl et al., 2013).

Intrigued by these findings, we wanted to narrow down the actual region in PWWP2A that facilitates binding to mononucleosomes. Thus, eight additional truncation constructs were generated (**Figure 3.3.2A**), followed by recombinant expression in BL21 and purification with glutathione sepharose.
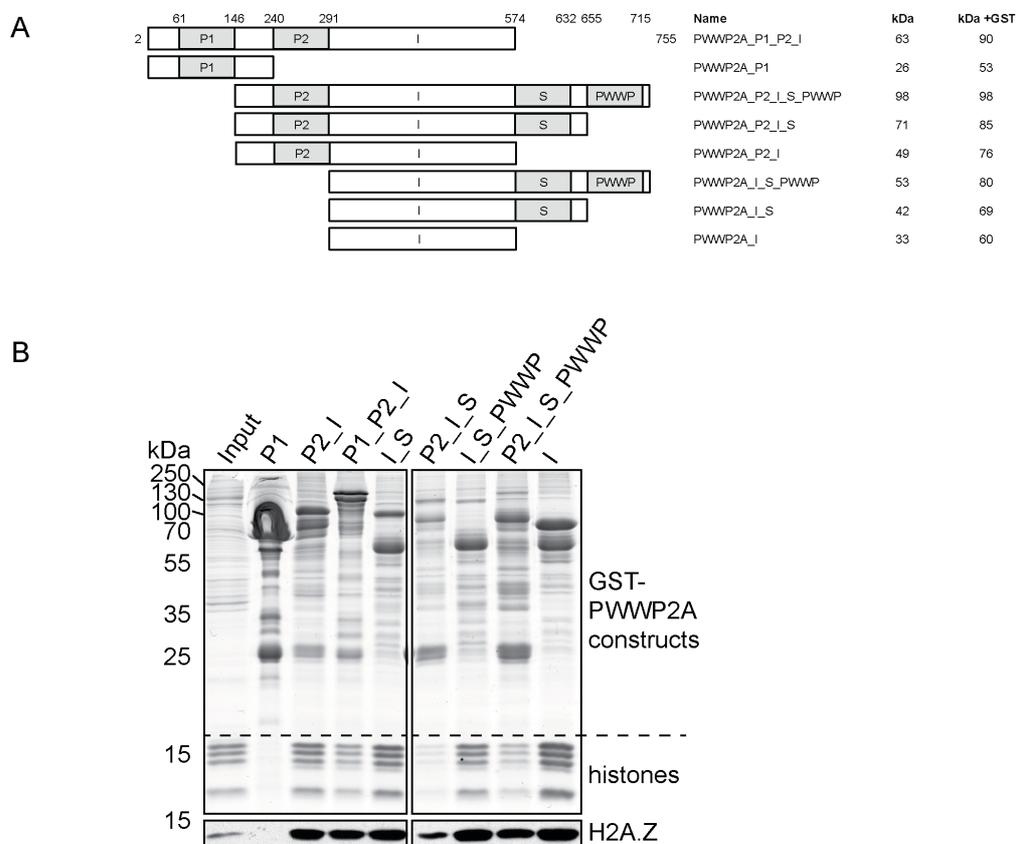
Figure 3.3.2 The internal domain of PWWP2A is needed and sufficient for (H2A.Z) nucleosome binding. (A) Schematic depiction of eight PWWP2A truncations. The expected molecular weight of every construct with and without GST-tag is indicated in kDa. **(B)** GST-tagged PWWP2A truncations immobilized to sepharose beads were incubated with mononucleosomes from HeLa Kyoto cells, precipitations separated by SDS-PAGE and gels stained with Coomassie (upper part) or utilized for immunoblotting and H2A.Z detection (lower part).

Again, these GST-fusions were incubated with HeLa Kyoto mononucleosomes and immunoprecipitated proteins analyzed by immunoblotting and Coomassie staining (**Figure 3.3.2B**). Surprisingly, all constructs that successfully pulled down all four core histones

contained a central region within PWWP2A that is flanked N-terminally by the second proline-rich stretch and C-terminally by the serine-rich stretch (aa 291 – 573). We decided to call this region the 'internal' domain (I). This stretch seems to be needed and to be sufficient to bind nucleosomes as one tested PWWP2A truncation that did not harbor the internal domain (P1) was not capable of precipitating nucleosomes (**Figure 3.3.2B**).

Cross species alignments of the PWWP2A protein sequence revealed a high conservation not only of the PWWP domain but also of the internal domain, underlining the suggested importance of these domains (Appendix **Figure A.4**). To map nucleosome binding more precisely, three additional constructs based on the internal region were established (**Figure 3.3.3A**): GST-PWWP2A_conI represents the conserved internal stretch, GST-PWWP2A_IN the N-terminal half and GST-PWWP2A_IC the C-terminal half of the human PWWP2A internal domain. Employing the same experimental strategy as for the other PWWP2A truncations revealed that GST-PWWP2A_conI but also the N-terminal internal construct (GST-PWWP2A _IN) were able to pull down nucleosomes (**Figure 3.3.3B**). Interestingly, although the IN-truncation precipitated nucleosomes, these nucleosomes did not contain H2A.Z, offering the possibility that this part of the internal domain is a general nucleosomes binder loosing the affinity for H2A.Z. The C-terminal half (GST-PWWP2A_CT), however, is only weakly interacting with nucleosomes. Strikingly, H2A.Z is enriched comparably to the I- and conI-constructs as detected by immunoblotting, arguing for a specific pull down of nucleosomes containing H2A.Z.
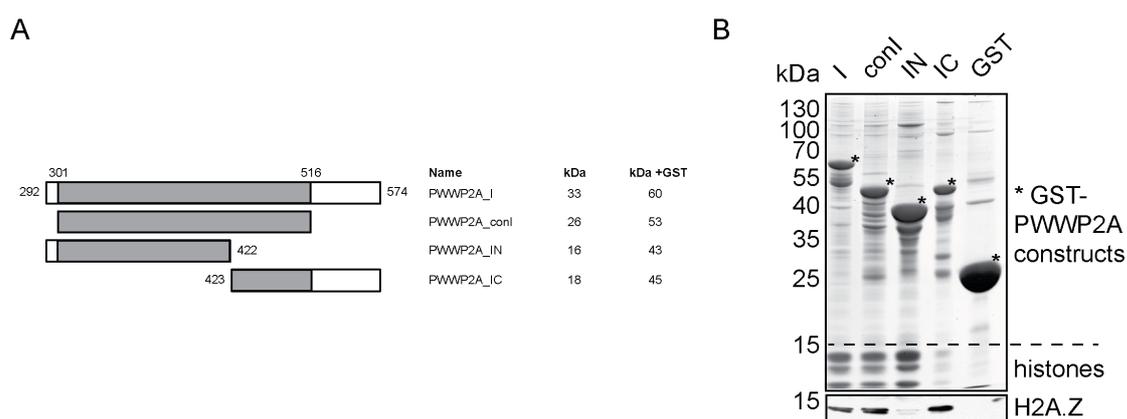
Figure 3.3.3 Nucleosome binding of PWWP2A is mediated by the IN part of the internal domain, whereas the IC part specifically interacts with H2A.Z. **(A)** Schematic depiction of internal deletion constructs. The expected molecular weight of every construct with and without GST-tag is indicated in kDa. **(B)** GST-tagged PWWP2A internal domain truncations immobilized to sepharose beads were incubated with mononucleosomes from HeLa Kyoto cells, precipitations separated by SDS-PAGE and gels stained with Coomassie (upper part) or utilized for immunoblotting and H2A.Z detection (box below). GST-only (GST) served as negative control.

3.4 The internal domain mediates H2A.Z nucleosome specificity

To find out whether there is an actual preference for H2A.Z-containing nucleosomes, binding assays with recombinant GST-PWWP2A full-length and truncations were carried out using mononucleosomes that contained GFP-tagged H2A, H2A.Z.1 or H2A.Z.2 from stable HeLa Kyoto cell lines. In fact, GST-PWWP2A_fl and GST-PWWP2A_NT preferentially bound to nucleosomes containing either GFP-H2A.Z.1 or GFP-H2A.Z.2 compared to nucleosomes containing only GFP-H2A (**Figure 3.4.1A**).
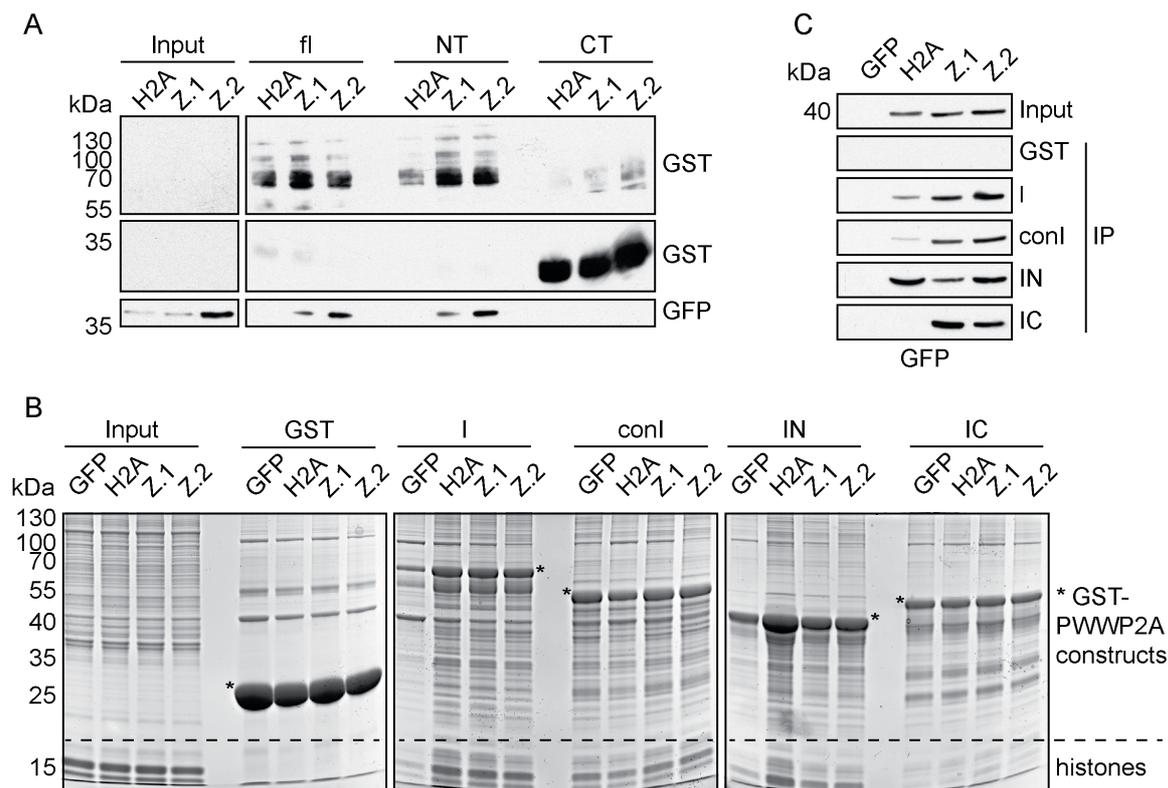
Figure 3.4.1 The internal domain of PWWP2A contains independent regions distinguishing between nucleosome-binding and H2A.Z-specificity. GST-PWWP2A full-length (fl), GST-PWWP2A_NT (NT), GST-PWWP2A_CT (CT) as well as different internal domain truncations (see **Figure 3.3.3**) immobilized to gluthatione sepharose beads were incubated with mononucleosomes from HeLa Kyoto cells, expressing GFP-H2A (H2A), GFP-H2A.Z.1 (Z.1) or GFP-H2A.Z.2 (Z.2). Precipitated proteins were separated by SDS-PAGE and gels either stained with Coomassie **(B)** or utilized for immunoblotting **(A** and **C)**. Cells expressing GFP as well as precipitations with GST served as negative controls.

Again, the PWWP domain was not necessary for this *in vitro* interaction and subsequently not involved in achieving H2A.Z specificity. As expected from the previous results, GST-PWWP2A_CT did not pull down nucleosomes. To further investigate the role of the internal domain in achieving H2A.Z specificity, the same assay was repeated employing the different truncations of the internal domain (see **Figure 3.3.3A**). Analyzing the pull downs by Coomassie staining after separation with SDS-PAGE recapitulated the results obtained with

wild type HeLa Kyoto mononucleosomes (**Figure 3.4.1B**). The internal domain (GST-PWWP2A_I) but also the conserved internal region (GST-PWWP2A-conI) and the internal N-terminal half (GST-PWWP2A_IN) were able to precipitate nucleosomes derived from all cell lines, whereas the C-terminal half (GST-PWWP2A_IC) showed a rather low nucleosomes affinity (**Figure 3.4.1B**). When detecting the GFP-tagged histone variants by immunoblotting, GST-PWWP2A_I and _conI show the preferential binding to nucleosomes containing GFP-H2A.Z.1 and GFP-H2A.Z.2 when comparing with GFP-H2A nucleosomes (**Figure 3.4.1C**), which was also previously observed with the full-length PWWP2A and the PWWP-deletion (**Figure 3.4.1A**). However, GST-PWWP2A_IN was not able to distinguish between nucleosomes containing either GFP-H2A.Z or GFP-H2A (**Figure 3.4.1C**), arguing for a H2A variant unspecific nucleosome affinity mediated by this part of the domain. Strikingly, the C-terminal half (GST-PWWP2A_IC) pulled-down less nucleosomes (possesses a lower nucleosome affinity) (**Figure 3.4.1B**) and showed a specific interaction with GFP-H2A.Z (**Figure 3.4.1C**). Fascinatingly, this implies that the internal region divides labor with regards to nucleosome binding and H2A.Z-specificity.

3.5 The PWWP domain is important for chromatin binding strength *in vivo*

To investigate the chromatin binding behavior of different PWWP2A truncations also in living cells, I teamed up with Dr. Katrin Schneider from the laboratory of Prof. Heinrich Leonhardt at the LMU BioCenter to perform Fluorescence Recovery After Photobleaching (FRAP) assays (**Figure 3.5.1A**). HeLa Kyoto cells were transiently transfected with different GFP-tagged PWWP2A full-length and truncation constructs (see also **Figure 3.3.2A)** but also with controls like GFP and GFP-tagged H2A.Z.1 and H2A.Z.2. As expected, all expressed fusion proteins and GFP localized to the nucleus of the transfected cells (**Figure 3.5.1B**). Briefly, half of the GFP-positive nucleus of a cell was bleached with a confocal microscope laser and subsequently the recovery time of the GFP signal recorded. Assuming, that chromatin-associated proteins are highly immobile, a slow recovery time would be expected. This is true for the GFP-tagged histone variants H2A.Z.1 and H2A.Z.2, which recovered very slowly. In contrast, highly mobile proteins recover fast and are unlikely to be tightly associated with chromatin. This is exemplified by the behavior of GFP that showed a fast kinetic. As reasoned from the previous results, the full-length PWWP2A protein possessed a remarkably slow FRAP kinetic, suggesting that it is tightly associated with chromatin.
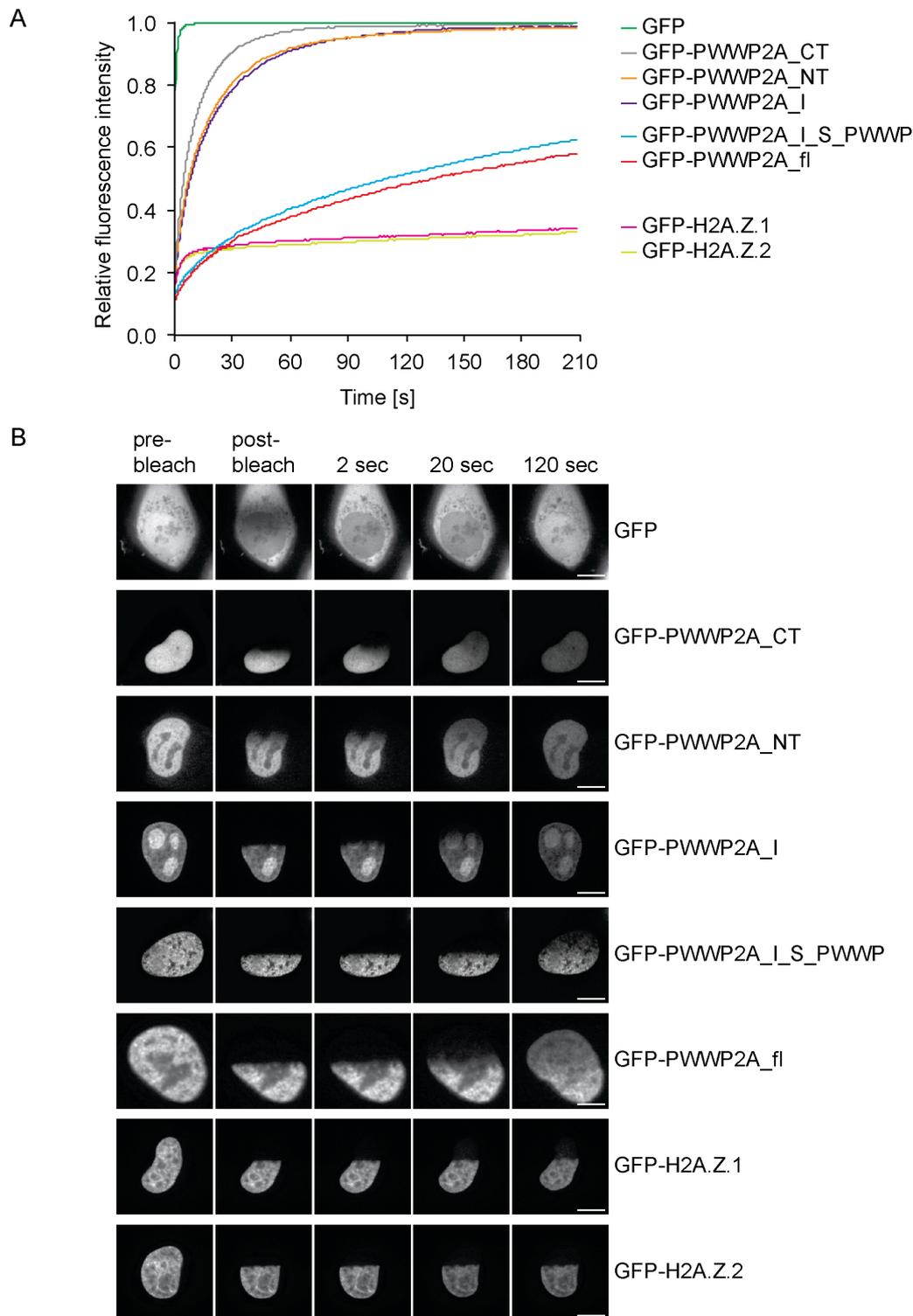
Figure 3.5.1 The PWWP domain and the internal domain combined are necessary for chromatin association of PWWP2A. FRAP assay of HeLa Kyoto cells transiently transfected with GFP-tagged PWWP2A full-length and truncations, GFP-H2A.Z.1 / H2A.Z.2 or GFP alone. GFP fluorescence was bleached in half of the nucleus and the recovery of the GFP fluorescence measured over 210 sec by spinning disc microscopy. **(A)** Quantification of fluorescence recovery rates after photobleaching. Mean-curves of 10 – 15 cells (n) are shown. Error bars are omitted for clarity. GFP-H2A.Z.1 (n = 10), GFP-H2A.Z.2 (n = 11), GFP-PWWP2A full-length (n = 10), GFP-PWWP2A_NT (n = 10), GFP-PWWP2A_I (n = 10), GFP-PWWP2A_CT (n = 14), GFP-PWWP2A_I_S_PWWP (n = 15) and GFP-only (n = 11). **(B)** Confocal microscopy GFP images of one representative cell per construct before and after photobleaching. Scale bars = 10 μm.

Surprisingly and contrary to their binding behavior *in vitro*, GFP-PWWP2A_I and GFP-PWWP2A_NT showed a rather fast recovery, explained by a higher mobility and arguing for only weak or potentially transient binding to chromatin. Notably, both truncations lack the PWWP-domain. For GFP-PWWP2A_CT, which consists of the PWWP-domain and the very C-terminus, an even faster recovery was observed, recapitulating its inability to pull down nucleosomes in the *in vitro* binding assays and suggesting that this domain is not sufficient for chromatin interaction. Interestingly, a truncation that harbors both the internal domain and the PWWP-domain (bridged by the serine-rich stretch, I_S_PWWP) mimicked the FRAP kinetics of the full-length protein, thereby clearly suggesting, that the combined action of the internal region and the PWWP-domain is crucial for a proper chromatin binding. Importantly, this teamwork is independent of the serine-rich stretch, as a truncation containing only this stretch and the PWWP domain (S_PWWP) possessed very fast discovery (preliminary data not shown). Notably, FRAP kinetics of stable HeLa Kyoto cell lines (where available) recapitulated the findings established by the presented FRAP assays with transiently transfected cells (**Figure 3.5.2**).

These results demonstrate the importance of both the internal and the PWWP domain for chromatin interaction *in vivo* and raise the question of how the PWWP domain contributes to this strong chromatin association (see discussion).
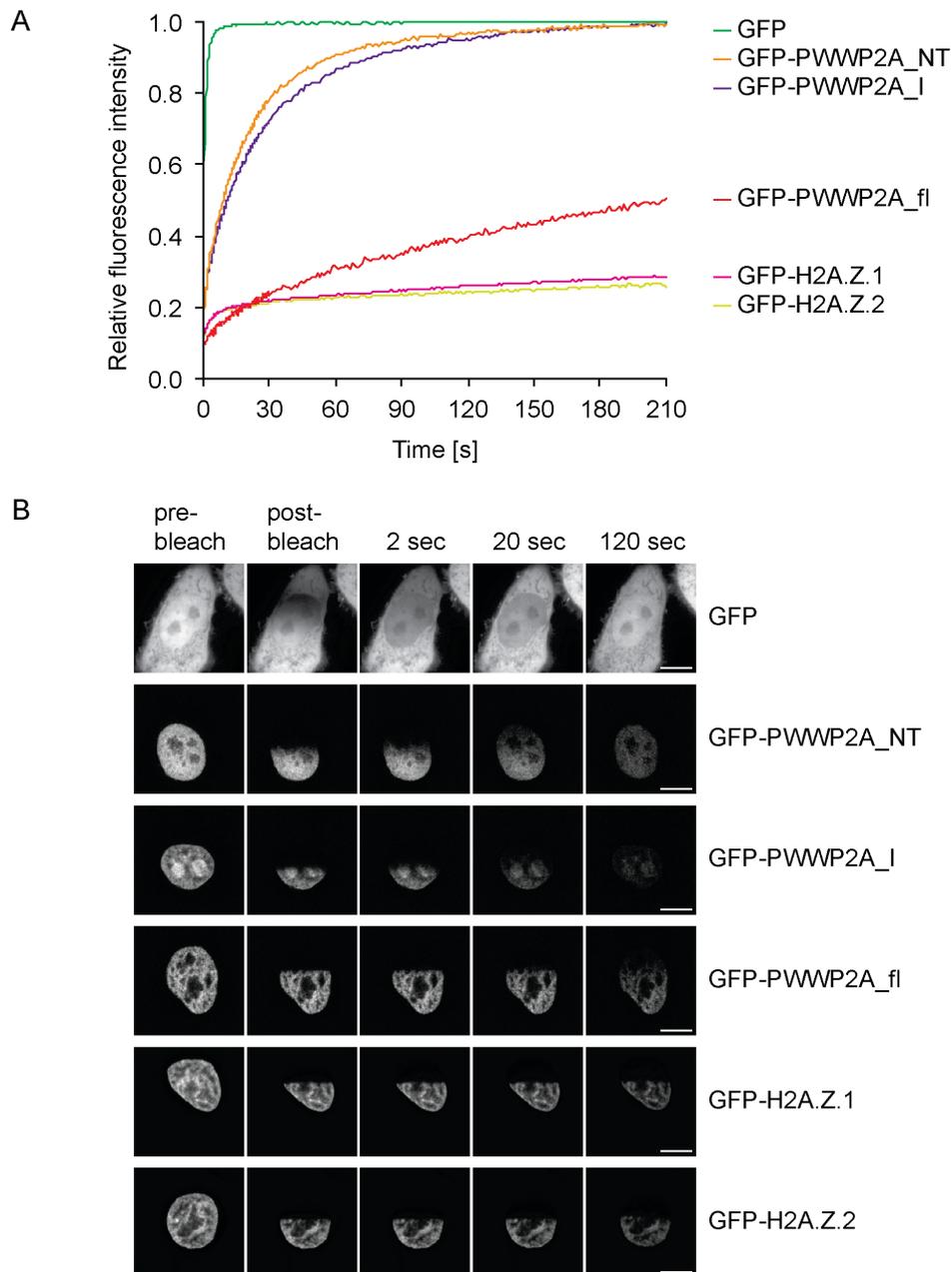
Figure 3.5.2 Validation of transient FRAP results in stable HeLa Kyoto cell lines. The FRAP experiment with stable cell lines was done as previously described in **Figure 3.5.1**. **(A)** Quantification of fluorescence recovery rates after photobleaching. Mean-curves of 10 or 11 cells (n) are shown. Error bars are omitted for clarity. GFP-H2A.Z.1 (n = 11), GFP-H2A.Z.2 (n = 11), GFP-PWWP2A full-length (n = 10), GFP-PWWP2A_NT (n = 10), GFP-PWWP2A_I (n = 11) and GFP-only control (n = 10) **(B)** Images of one representative cell per cell line before and after photobleaching. Scale bars = 10 µm.

3.6   The PWWP2A interactome

PWWP2A interacts with chromatin preferentially via its connection with H2A.Z-containing nucleosomes. In order to investigate if additional but so far unknown factors are potentially involved in supporting PWWP2A's chromatin binding and to get a first insight into PWWP2A function, the identification of putative PWWP2A interaction partners was very important. This experiment was carried out essentially as described earlier for the H2A.Z interactome (**Figure 3.1.1**). Briefly, mononucleosomes from HeLa Kyoto cells stably expressing GFP-PWWP2A or GFP were generated (**Figure 3.6.1A**) and immunoprecipitations against the GFP-tag were conducted with GFP-trap beads.
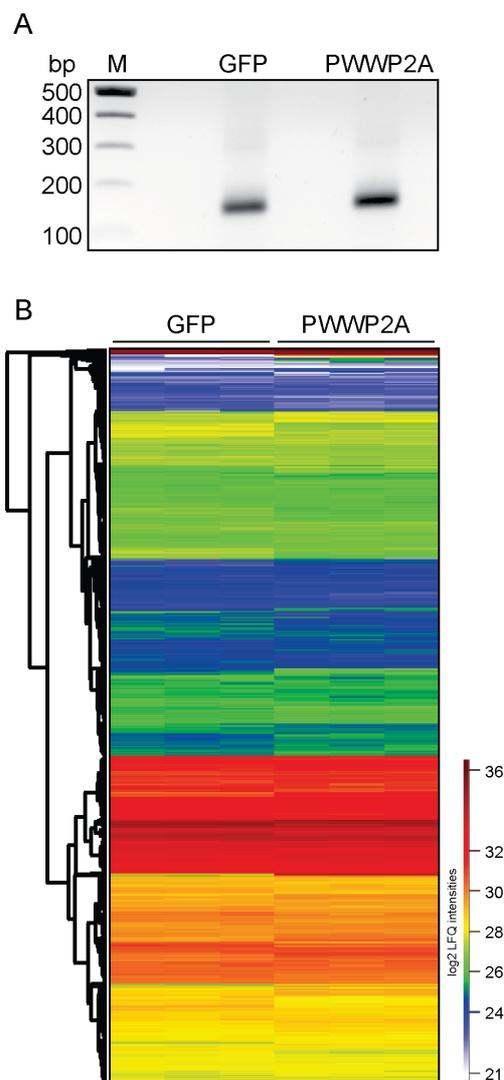
Figure 3.6.1 MNase ChIP experiment followed by qMS with a HeLa Kyoto cell line stably expressing GFP-PWWP2A. (A) MNase digested chromatin was prepared from nuclei isolated from HeLa Kyoto cell lines stably expressing GFP or GFP-PWWP2A. MNase digestion yielded almost pure mononucleosomes (~150bp) as controlled by agarose gel electrophoresis and ethidium bromide staining. **(B)** Tryptic IP peptides were analyzed by LC-MS/MS employing a Q Exactive mass spectrometer, MS raw data processed by MaxQuant and logarithmized label-free quantification (LFQ) intensities of 1387 proteins enriched by replicate IPs visualized using hierarchical clustering in a heatmap by Perseus.

After on-beads tryptic digestion technical triplicates were run on the Q Exactive mass spectrometer, peptides were identified and label-free quantification executed with MaxQuant (1.3.9.2). The screening workflow led to the identification of 1804 proteins (**Figure 3.6.1B**), of which 648 proteins were enriched on PWWP2A over GFP background binders. FDR-based statistical analysis executed with Perseus (1.3.10.0) and plotted in volcano plots with R (3.0.2) determined 46 proteins to be valid and promising interaction partners of PWWP2A (**Figure 3.6.2**). Complementing the H2A.Z interactome results, this reciprocal experiment identified H2A.Z as part of the PWWP2A interactome.
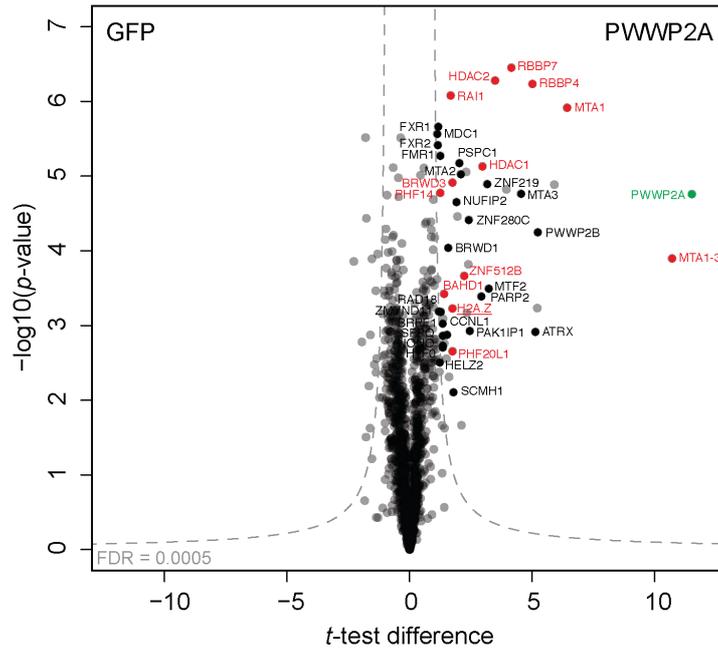
Figure 3.6.2 Identification of proteins interacting with GFP-PWWP2A in HeLa Kyoto cells. Proteins interacting with GFP-PWWP2A (right half) were separated from proteins enriched on GFP control (left half) by plotting *p*-values and *t*-test differences obtained by two-sample *t*-test in volcano plots with R, highlighting interesting candidates with a threshold line based on FDR-statistics (FDR = 0.0005, S_0 = 1). Colored and labeled are only interactors associated with chromatin; green: PWWP2A, red = H2A.Z nucleosome interactors, black = other chromatin interactors. For a complete list please see **Table 3.6.1**.

Furthermore, almost all members of the NuRD complex (Kloet et al., 2014) that were also identified in H2A.Z-mononucleosome qMS analyses, are prominently featured (**Table 3.6.1**): HDAC1, HDAC2, RBBP4, RBBP7 and MTA1. Also MTA2 and MTA3 were found, whereas Methyl-CpG-binding domain proteins 2 or 3 (MBD2/3), the Chromodomain-helicase-DNA-binding proteins 3, 4 and 5 (CHD3/4/5) as well as the histone binding proteins transcriptional repressors p66-alpha and -beta (GATAD2A/2B) were not identified. PWWP2A has previously been shown to be part of the HDAC1-interactome (Joshi et al., 2013), a result that is now recapitulated and consolidated by the presented findings. Members of the NuRD complex were also identified in the H2A.Z-interactome screen, establishing an exciting connection between H2A.Z, PWWP2A and histone deacetylation. Other proteins that were also identified to interact with H2A.Z nucleosomes are RAI1 and PHF14, members of a putative H3K4me3 repeller complex (Eberl et al., 2013), as well as BRWD3, BAHD1, PHF20L1 and ZNF512B (**Table 3.6.1**). All interacting candidates related to chromatin are plotted in two heatmaps for better visualization of their specific enrichment over the GFP

control, separating NuRD complex members from other chromatin interactors (**Figure 3.6.3A** and **B**).

Table 3.6.1 PWWP2A interactome, potential complex association and function

Group	Proteins	Complex/Function
H2A.Z binder	PWWP2A	This study
	H2A.Z	Histone H2A replacement variant
	MTA1, RBBP4, RBBP7, HDAC2	NuRD complex, chromatin remodeling and histone deacetylation
	RAI1, PHF14	No name given yet, repelled by H3K4me3
	BRWD3	Cullin E3 Ligase complex
	BAHD1	Transcriptional repressor
	PHF20L1	DNMT1 stabilization
	ZNF512B	Unknown
Other chromatin binder	MTA2, MTA3, HDAC1	NuRD complex, chromatin remodeling and histone deacetylation
	FXR1, FXR2, FMR1	Fragile X syndrome related proteins, RNA-binder
	PSPC1, NONO, SFPQ	Speckle formation, transcription regulation, splicing
	ATRX	H3.3-dependent chromatin remodeling, telomere structure
	BRPF1	MOZ/MORF complex, H3 acetylation
	BRWD1	Regulation of cell morphology, contains bromodomain
	CCNL1	Cyclin-L1, transcriptional regulator, splicing
	HELZ2	Helicase, transcriptional coactivator
	H1	Linker histone
	MDC1	DNA repair
	MTF2	PRC2 complex,
	NUFIP2	Proliferation induction
	PAK1IP1	Proliferation regulation
	PARP2	Base excision repair
	PWWP2B	Unknown, contains PWWP-domain
	RAD18	E3 Ligase, DNA repair
	SCMH1	PRC1-like complex
	ZMYND11	Transcription elongation, tumor suppression, pre-mRNA processing
	ZNF219	Transcription factor
	ZNF280C	Transcription factor
Other	ALPP	Alkaline phosphatase
	ATXN2L	Unknown, putative RNA-binder
	BMP2K	Kinase
	HERC5	E3 Ligase
	IMPDH2	Dehydrogenase
	KPNA4	Importin subunit, nuclear protein import
	LSM12	Unknown, putative RNA-binder
	MYO1D	Motor molecule
	OASL	RNA-binder, Antiviral activity
	REPS1	EGF receptor regulator
	TNS1	Focal adhesion, actin-binder
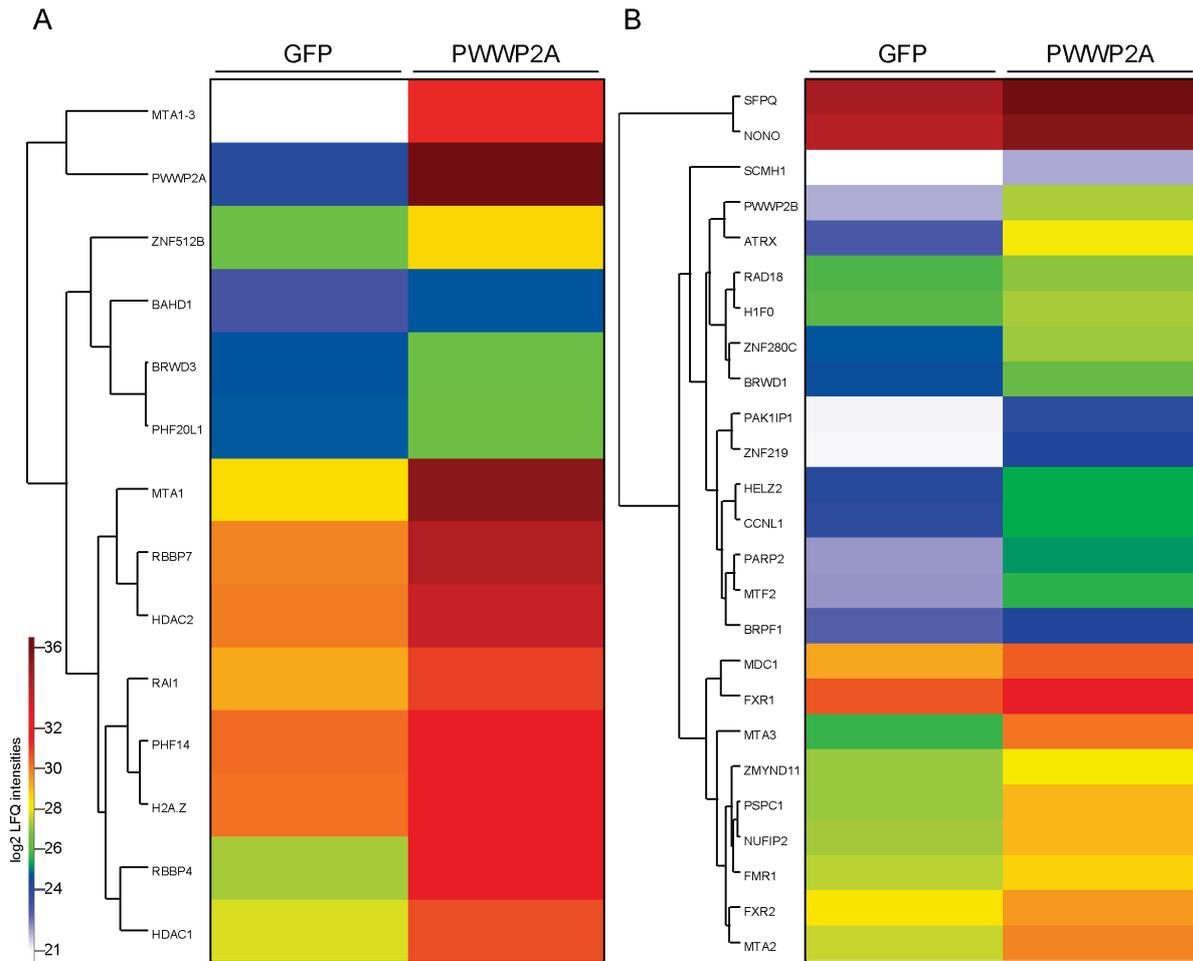
Figure 3.6.3 Interactors of GFP-PWWP2A and their LFQ intensities in GFP-PWWP2A versus GFP control pull downs revealed by qMS. Logarithmized and averaged LFQ intensities of GFP-PWWP2A interactors in GFP-only and GFP-PWWP2A ChIPs were hierarchically clustered and plotted as heatmap with Perseus. The log2 LFQ intensities are presented for GFP-PWWP2A binders that were **(A)** also identified to bind H2A.Z nucleosomes or **(B)** associated with chromatin in general.

3.7 PWWP2A is involved in cell cycle progression

Given its strong association with H2A.Z-containing nucleosomes, its *in vivo* co-occupancy with H2A.Z promoter peaks and its ability to recruit components of the NuRD complex, the question arose, what functional outcome these different PWWP2A connections may offer. Therefore, functional analyses of PWWP2A-depleted cells were conducted (**Figure 3.7.1**).
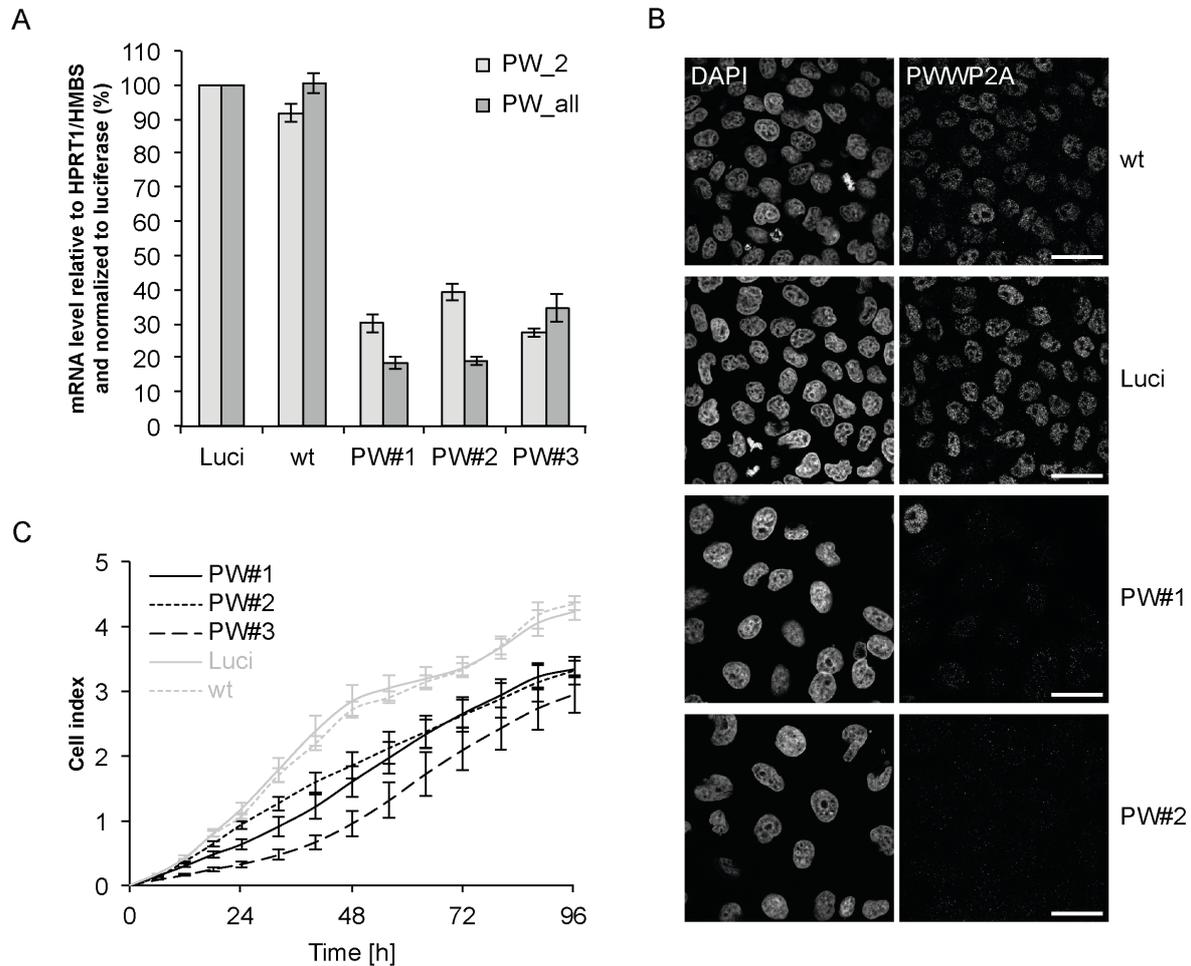
Figure 3.7.1 RNAi mediated knockdown of PWWP2A leads to impaired cell proliferation. (A) qPCR was employed to determine PWWP2A knockdown efficiency. HeLa Kyoto cells were independently transfected with three different siRNAs (PW#1, PW#2, PW#3). Two days upon transfection, RNA was isolated and cDNA prepared for qPCR analysis. RNA from wild type (wt) HeLa Kyoto cells and luciferase-control siRNA transfections (Luci) served as controls. Primer pair 'PW_2' detected only canonical PWWP2A; 'PW_all' detected all three isoforms. PWWP2A mRNA levels are depicted relative to reference gene mRNA levels (HPRT1 and HMBS) and normalized to luciferase control. Error bars represent the standard error (n = 4). **(B)** IF analysis of PWWP2A knockdown efficiency. PWWP2A knockdown was carried out in HeLa Kyoto cells using siRNAs PW#1 and PW#2. Cells were seeded on coverslip two days after transfection. Cells were fixed on day three, DNA counterstained with DAPI and endogenous PWWP2A probed with a commercial PWWP2A antibody (secondary antibody: Alexa 488 coupled). Imaging was done utilizing the Leica SP5 II confocal scanning microscope. Scale bars = 20 µm **(C)** Two days upon PWWP2A knockdown with PWWP2A siRNAs #1 or #2 (PW#1 and PW#2), HeLa Kyoto cells as well as control cells (HeLa Kyoto wt and HeLa Kyoto luciferase-knockdown) were seeded into an E-plate (time = 0h) and cell proliferation monitored for 96 hours in the incubator employing the xCELLigence RTCA DP system. Error bars represent the standard error (n = 3)

Thus, three different siRNAs that targeted all three potential PWWP2A isoforms were designed to knock down PWWP2A transcription and to analyze the resulting cellular consequences. In HeLa Kyoto cells a satisfying and reproducible 60% – 80% percent reduction of PWWP2A transcripts after three days (depending on siRNA and primer pair) could be detected by qPCR with all three siRNAs (**Figure 3.7.1A**). In order to verify the knockdown efficiency also on the protein level, IF stainings were utilized, probing wild type and knockdown cells with the commercial PWWP2A antibody. The IF results depicted in **Figure 3.7.1B** not only confirmed the actual knockdown of the PWWP2A protein in HeLa Kyoto cells but also demonstrated the antibody specificity. Mere observation of the cells with a tissue culture microscope upon PWWP2A knockdown revealed a decrease in cell number compared to non-treated or control-transfected (luciferase siRNA) cells. Therefore, growth curves of transfected and control cells were established over several days with the xCELLigence system (Roche), that monitors cell proliferation in real time in the tissue culture incubator. HeLa Kyoto cells were transfected with either one of three PWWP2A siRNAs or luciferase siRNA or not transfected at all and seeded into xCELLigence E-plates in duplicates of 5.000 cells each at day two after transfection. Proper knockdown of PWWP2A transcripts was controlled by qPCR (data not shown). With the help of the automatically recorded growth curves, the microscopically observed growth defect could be recapitulated for all three PWWP2A knockdowns reproducibly in three independent experiments (**Figure 3.7.1C**).

In order to identify the cause of the proliferation defect, cell cycle analysis via DNA content staining with PI and subsequent FACS analysis was applied. This experiment revealed that loss of PWWP2A led to a reduction of cells in S-phase accompanied by an increase of cells in G_2/M-phase compared to control cells (**Figure 3.7.2A** and quantification of 5 independent experiments **Figure 3.7.2B**).

To determine whether cells accumulate in G_2-phase or mitosis, cells were stained for H3 serine 10 phosphorylation (H3S10phos), a marker of mitotic chromatin (Kouzarides, 2007), and analyzed by FACS upon PWWP2A knockdown (**Figure 3.7.3A**). Strikingly, an almost 2-fold increase of mitotic cells was observed in knockdown cells compared to control cells (**Figure 3.7.3B**) suggesting a functional role for PWWP2A in progression of mitosis.
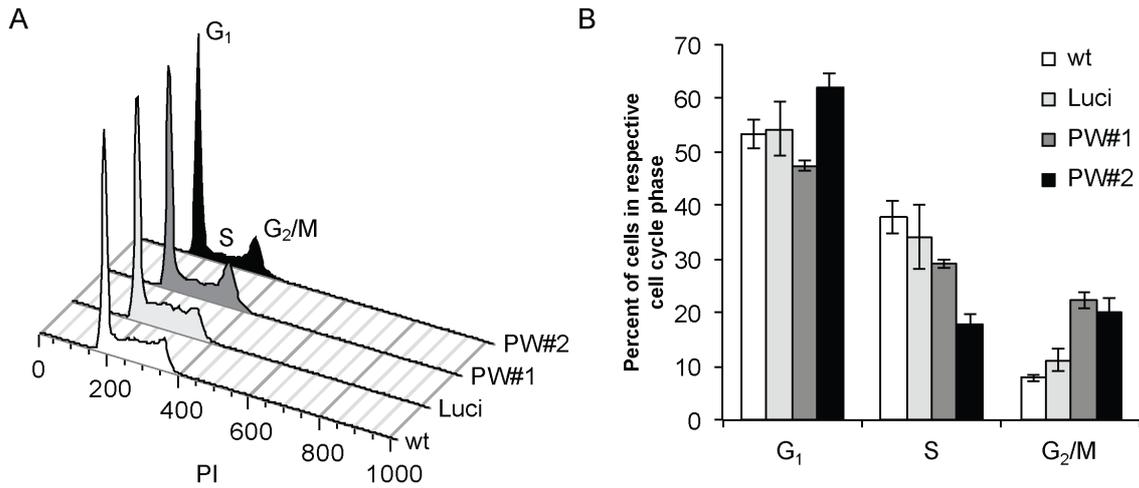
Figure 3.7.2 Number of HeLa Kyoto cells in S-phase is reduced but increased in G_2/M phase upon PWWP2A knockdown. (A) FACS histogram plot showing the distribution of PWWP2A knockdown and control cells upon PI staining into distinct cell cycle phases. Three days upon PWWP2A knockdown with PWWP2A siRNAs #1 or #2 (PW#1 and PW#2), HeLa Kyoto cells and control cells (HeLa Kyoto, wt and HeLa Kyoto luciferase-knockdown, Luci) were stained with PI and analyzed by FACS determining the number of cells in G_1-, S- and G_2/M-phase according to DNA content. **(B)** Quantification of five independent FACS experiments (see **(A)**). Error bars represent the standard error (n = 5).
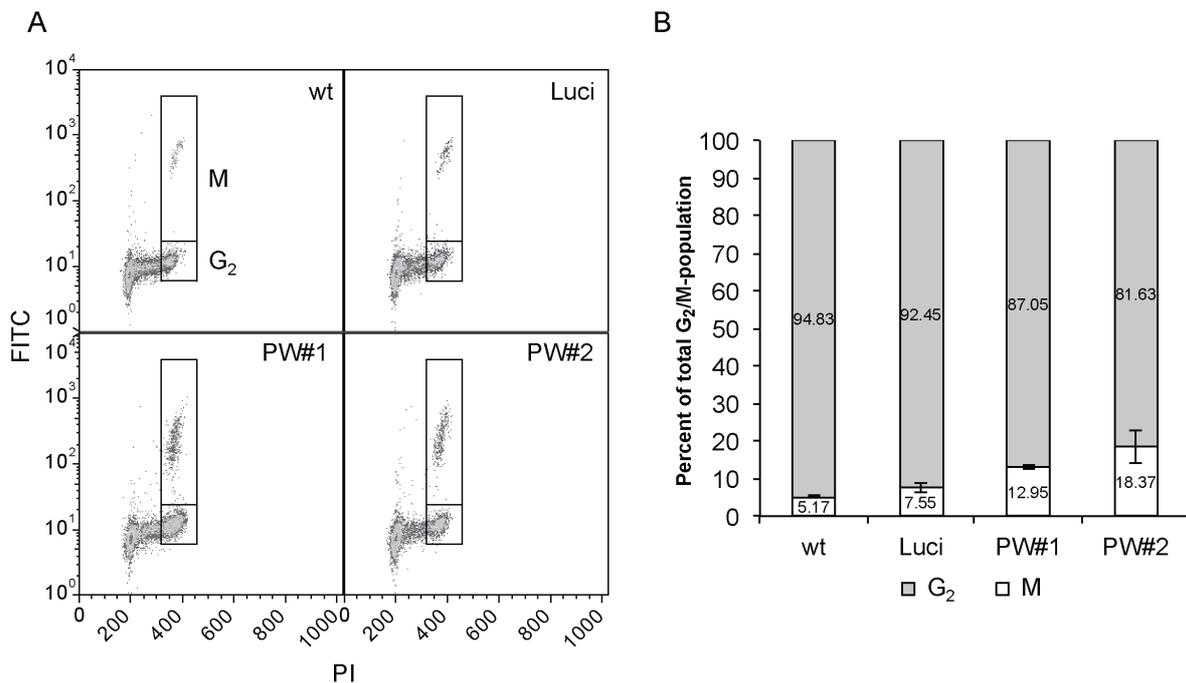
Figure 3.7.3 HeLa Kyoto cells accumulate in mitosis upon PWWP2A loss. (A) Three days upon PWWP2A knockdown with PWWP2A siRNAs #1 or #2 (PW#1 and PW#2), HeLa Kyoto cells and controls (HeLa Kyoto, wt and HeLa Kyoto luciferase-knockdown, Luci) were co-stained with PI (DNA content) and H3S10phos antibody (secondary antibody FITC-labeled) and subjected to FACS analyses. H3S10phos staining allowed to distinguish mitotic (M) cells from G2 cells. **(B)** Quantification of three independent FACS experiments (see **(A)**). Shown is the percentage of cells in G_2- or M-phase with respect to the total G_2/M-population. G_2-phase cells are depicted in grey, mitotic cells (M) in white. Error bars represent the standard error (n = 3).

99

To further dissect, which phase of mitosis was affected by the loss of PWWP2A, knockdown and control cells were seeded on coverslips, stained for H3S10phos and DNA counterstained with DAPI. Thereupon, overview images were taken with an epifluorescence microscope, on average 100 H3S10phos-positive mitotic cells were counted per treatment (knockdown and controls) and visually discriminated according to their mitotic phases (prophase, prometa-/metaphase, anaphase) (Hake et al., 2005) (**Figure 3.7.4A**). The analysis of four independent replicates revealed that PWWP2A knockdown appeared to accumulate cells in prometa-/metaphase concomitant with a decrease of cells in pro- and anaphase (**Figure 3.7.4B**), thereby delaying cell cycle progression and potentially explaining the observed phenotype. Notably, mitotic chromatin in knockdown cells appeared less condensed and chromosomes showed misorientation during anaphases (**Figure 3.7.4A** and **Figure 4.3.2**). However, a thoroughly quantification and documentation of these phenotypes will be object of future experiments.
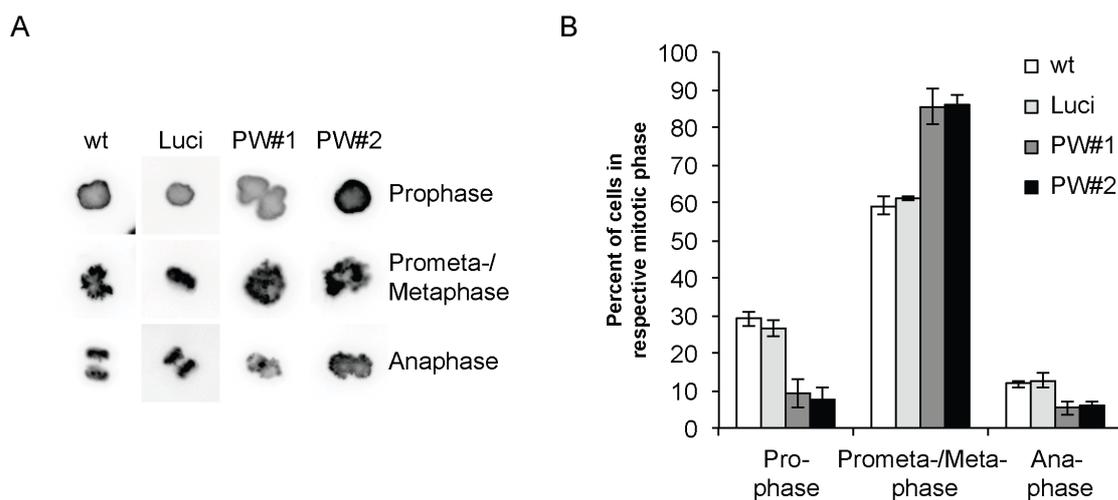
Figure 3.7.4 HeLa Kyoto cells are reduced in pro- and anaphase but accumulate in prometa/metaphase upon PWWP2A knockdown. (A) Two days upon PWWP2A knockdown with PWWP2A siRNA #1 or #2 (PW#1 and PW#2), HeLa Kyoto cells and controls were seeded onto coverslips, stained for H3S10phos (mitotic cells) and DNA counterstained with DAPI. Cells were counted and visually discriminated according to their mitotic phases into prophase, prometa-/metaphase and anaphase. Examples of H3S10phos-positive mitotic stages are shown. **(B)** Quantification of four independent experiments as described in **(A)** Error bars represent the standard error (n = 4).

In summary, the loss of PWWP2A led to defects in cellular proliferation. However, the underlying molecular mechanism, may it require PWWP2A's action in regulation of S-phase or mitosis, may it involve transcriptional regulation or chromatin structure, remains unclear to this end and will be the focus of future experimental work.

4 Discussion and future perspectives

4.1 The H2A.Z nucleosome interactome

The histone H2A replacement variant H2A.Z is associated with various cellular processes. To investigate consequences of H2A.Z chromatin incorporation a label-free quantitative mass spectrometry approach was employed analyzing the interactome of H2A.Z-containing nucleosomes. To my knowledge, this is the first time label-independent interaction proteomics was applied to histone variant chromatin. Pioneering this approach with mononucleosomes derived from MNase digested chromatin of SK-mel147 cell lines, 42 established but also new H2A.Z interactions could be identified (**Table 3.1.1**).

The SRCAP complex that exchanges nucleosomal H2A-H2B dimers for free H2A.Z-H2B dimers (Ruhl et al., 2006) was identified almost entirely confirming the feasibility of the presented approach. A second mammalian SWR1-related complex is the p400/NuA4/TIP60 (Gevry et al., 2007) complex that shares several members with the SRCAP complex (Billon and Cote, 2013) and functions apart from its acetylation capability also as H2A.Z depositor. Of its unique members only TRRAP and EP400 could be identified, which are big in size (440 kDa and 340 kDa, respectively), therefore produce more peptides and are more likely to be detected by qMS. It is possible that this rather big complex needs more than one nucleosome as platform for interaction, possesses a lower binding affinity or interacts rather transient. Moreover, the identified proteins could constitute a subcomplex (a p400 or TRRAP or p400/TRRAP complex) as proposed by Billon and Côté (Billon and Cote, 2013). Finally, the p400/NuA4/TIP60 complex seems to act at promoters of specific genes (Gevry et al., 2007; Gevry et al., 2009) and might simply not be as abundant as the SRCAP complex.

H2A.Z's partly contradictory role in gene expression regulation and its involvement in DNA repair and chromatin structure are also reflected by the diversity of interacting factors identified by qMS. However, is very likely that the presented interactors are not necessarily directly binding H2A.Z but rather depend on the whole H2A.Z-nucleosomal context, which includes reading associated PTMs with specific chromatin-related domains and recruitment through complex members (**Table 3.1.1** and **Table 3.1.2**).

In accordance with H2A.Z's localization at the TSS of active genes and its role in transcription (Marques et al., 2010), nucleosomes containing H2A.Z interact with members of the COMPASS-like family of histone methyltransferase complexes, the MLL1 and MLL2 complex, which methylate H3 on lysine 4 (van Nuland et al., 2013a), creating a hallmark of

transcriptionally active chromatin (Bannister and Kouzarides, 2011). In fact, H2A.Z nucleosomes not only attract the methyltransferases but are as a possible consequence also enriched with trimethylated H3K4 as shown by immunoblotting (**Figure 3.1.5D** and **Figure 3.1.10**) and suggested by ChIP-sequencing (**Figure 3.2.9**).

Four members of the NuRD complex were identified to interact with H2A.Z mononucleosomes (**Table 3.1.1** and **Table 3.1.2**): HDAC2, RBBP4, RBBP7 and MTA1 (Allen et al., 2013). The complex is known to regulate gene transcription through binding to promoters and enhancers and tissue- and context-specifically modify chromatin (Basta and Rauchman, 2014) namely by sliding nucleosomes and deacetylating histones (Xue et al., 1998). The latter would presumably cause chromatin condensation and could support H2A.Z mediated chromatin secondary structure formation. However, NuRD activity is not necessarily statically involved in silencing transcription but rather seen as a fine tuner maintaining equilibrium between acetylated and deacetylated histone tails (Hu and Wade, 2012). Interestingly, besides the histone binding components GATAD2A/B and the helicase components CHD3/4/5 of the NuRD complex also the methyl-CpG-binding proteins MBD2/3 were not found to be enriched on H2A.Z nucleosomes. DNA methylation and H2A.Z deposition are reported to be mutually exclusive in plants (Zilberman et al., 2008) and mammals (Conerly et al., 2010b). This could implicate the existence of a minimal NuRD complex that is recruited independently of DNA methylation to sites of H2A.Z-dependent transcriptional regulation where deacetylation rather than nucleosome remodeling is necessary (discussed also below).

The Cullin E3 ligases CUL4A and CUL4B were also found to be associated with H2A.Z nucleosomes, potentially linking histone ubiquitination and H2A.Z biology. Ubiquitination of H2A and H2B is involved in gene expression regulation (Bannister and Kouzarides, 2011). Also here, the influence on transcription is conflictive and context dependent: H2AK119ub is associated with Polycomb-dependent transcriptional repression, whereas H2BK120ub is required for trimethylation of H3K4 (histone crosstalk), thereby supporting transcriptional activation (Kouzarides, 2007). Cullin 4A and 4B have been reported to be involved in DNA repair, where H3 and H4 get ubiquitinated (Dai and Wang, 2006), and in nucleosome assembly, where Cullin 4A ubiquitinates H3 and regulates H3 deposition (Han et al., 2013). Interestingly, all Cullin ligases are modified by attachment of the ubiquitin-like protein NEDD8, which stimulates assembly and activity of Cullin containing ubiquitin ligases (Lydeard et al., 2013). Convincingly, NEDD8 was also identified to be part of the H2A.Z interactome (**Table 3.1.1**). In summary, this might hint towards a potential ubiquitination of

H2A.Z-containing nucleosomes or H2A.Z itself by neddylated Cullin E3 ligases with yet unknown consequences. Indeed, H2A.Z is mono-ubiquitinated in the inactive X-chromosome in female mammalian cells by a RING1b E3 ligase, which is part of the Polycomb complex (Sarcinella et al., 2007), underlining the involvement of ubiquitinated histones in transcriptional repression.

All members of a previously identified but yet unnamed complex were enriched on H2A.Z-containing nucleosomes (**Table 3.1.1**): HMG20A, PHF14, RAI1 and TCF20. Eberl et al. found RAI1, TCF20 and PHF14 in histone tail peptide pulldowns to be enriched on unmodified H3 tails but depleted on trimethylated H3K4 tails. Together with HMG20A they form a chromatin associated complex with unknown function (Eberl et al., 2013). The observed H3K4me3 repelling behavior is counterintuitive, because the complex is attracted by nucleosomes containing H2A.Z that are also H3K4 trimethylated in HeLa Kyoto and SK-mel147 cell lines according to immunoblot analysis (**Figure 3.1.5D** and **Figure 3.1.10**) and correlation with HeLa H3K4me3 ChIP sequencing data **Figure 3.2.9**). But, this does not generally mean that every H2A.Z nucleosome is also decorated with H3K4me3, a notion that is underlined by the differential distribution of H3K4me3 and H2A.Z on both nucleosomes flanking the NDR (**Figure 3.2.9**). Hence, it is very likely that different populations of H2A.Z containing nucleosomes show distinct modification pattern and attract specific factors. It will be crucial to unravel H3K4me3 localization in the respective cell lines by ChIP sequencing and compare these data with the genomic localization of H2A.Z and the respective complex members. Furthermore, extending the interaction proteomics approach by using assembled nucleosomes carrying H3K4me3 and/or H2A.Z could help to determine whether one chromatin mark potentially outcompetes the other.

Several proteins were identified that are not or not yet associated with any complex (**Table 3.1.1** and **Table 3.1.2**). Interestingly, among these factors are also transcriptional regulators like BAHD1 or MIER1, both supposedly involved in transcriptional repression (Bierne et al., 2009; Ding et al., 2003) underlining again H2A.Z's role in gene expression regulation. Some of the identified proteins even seem to preferentially interact with either H2A.Z.1 or H2A.Z.2, the strongest candidates being BRD2 (H2A.Z.1) and PWWP2A (H2A.Z.2) in melanoma cells (**Figure 3.1.6**). This isoform-specialization generally hints towards the suggested functional diversification of H2A.Z.1 and H2A.Z.2 but might also critically depend on the respective cell model and organism as well as on the PTMs that accompany nucleosomes containing the respective variant. Strikingly, preliminary data on posttranslational modifications of nucleosomes containing either H2A.Z.1 or H2A.Z.2 that I obtained in collaboration with

Bhanu Natarajan (PhD), a PostDoc in Prof. Benjamin Garcia's group at the University of Pennsylvania, point towards a differential modification pattern. On the one hand, we identified a new PTM present on H2A.Z.1 only, H2A.Z.1K37ac, which could specifically support H2A.Z.1 specific protein interactions and variant functionality (data not shown). Indeed, acetylation of H2A.Z has been connected to active transcription and an open chromatin state (Ishibashi et al., 2009). K37 is just one amino acid adjacent to H2A.Z.1 T38 and H2A.Z.2 A38, which represent one out of three amino acids that are altered between the two variants (**Figure 1.3.2**). Potentially, this differential amino acid might affect binding of the respective histone acetyltransferase and is therefore crucial for the establishment of this specific acetylation mark. On the other hand, further preliminary mass spectrometry results suggest, that also other histones within a H2A.Z nucleosome can be differentially modified depending on the isoform, but this has yet to be thoroughly validated and replicated before interpreting these data. However, hyperacetylation of histone H4 in H2A.Z.1 containing nucleosomes was shown to support interaction of BRD2 with this histone variant (Draker et al., 2012). BRD2 was also detected in the quantitative H2A.Z interactome screen that is part of this thesis and its function in H2A.Z.2 mediated control of cell cycle regulator expression and metastatic melanoma progression is extensively described in our collaborative study mentioned earlier (Vardabasso et al., 2015)(under revision). This example of a factor that facilitates downstream functionality upon binding H2A.Z nucleosomes illustrates the importance of uncovering new candidates potentially involved in transmitting H2A.Z function. One of these possible factors is PWWP domain-containing protein 2A (PWWP2A), which was identified to bind H2A.Z in both SK-mel147 qMS screens (**Table 3.1.1**) and could be validated also in the presented HeLa Kyoto screen (**Table 3.1.2**) as well as in independent pull downs followed by immunoblotting (**Figure 3.2.2A** and **B**).

Notably, applying the same native MNase ChIP workflow but analyzing the precipitated DNA by next generation sequencing instead of the enriched proteins by qMS allowed for the first time to establish an H2A.Z isoform-specific genomic map (see section 3.2). This assay revealed that GFP-H2A.Z.1 and GFP-H2A.Z.2 genomic peaks overlap in large parts, especially at the TSS of genes (**Figure 3.2.7**). Still there are loci that seemed not to be occupied by both variants leaving room for isoform specific influence on distinct genes potentially mediated by different interaction partners. However, both H2A.Z isoforms seemed to especially enrich at TSS of genes that are actively transcribed (**Figure 3.2.9**). Due to antibody limitations previous H2A.Z mapping studies could not distinguish between H2A.Z.1

and H2A.Z.2, but in general, H2A.Z localization at TSS of genes has been associated in many model systems like human or fly with active transcription (Barski et al., 2007; Mavrich et al., 2008) but also with transcriptional repression in the yeast *S. cerevisiae* (Guillemette et al., 2005). This data set associates both isoforms with active transcription for the first time. Transcriptome changes upon specific variant knockdown could elucidate further if both variants indeed mediate specific changes in gene expression. Unpublished microarray data from our lab in fact show a variant specific downregulation of distinct proteins upon H2A.Z.1 or H2A.Z.2 knockdown in HeLa cells, albeit with no obvious effect on the cell phenotype. In contrast, our collaborative work with Emily Bernsteins group in metastatic melanoma revealed a deregulation of cell cycle genes upon H2A.Z.2 but not H2A.Z.1 knockdown that led to impaired proliferation of melanoma cells (Vardabasso et al., 2015) (under revision).

4.2 PWWP2A interaction with the H2A.Z nucleosome

In vitro nucleosome binding assays uncovered the role of the internal (I) domain in mediating not only nucleosome interaction but also promoting H2A.Z nucleosome specificity. This newly discovered, large and rather central amino acid stretch in PWWP2A is flanked N-terminally by a proline-rich (P2) and C-terminally by a serine-rich (S) region (**Figure 3.3.2A**). Intriguingly, this region possesses together with the PWWP domain and parts of the P2 stretch a high cross species conservation (Appendix **Figure A.4**), suggesting its importance.


```
         IN
 292 ┊RKIKRPKRKM  YREEPTSIMN  AIKLRPRQVL  CDKCKNSVVA  EKKEIRKGSS  ATDSSKYEDK


    KRRNESVTTV  NKKLKTDHKV  DGKNQNESQK  RNAVVKVSNI  AHSRGRVVKV  SAQANTSKAQ

                  IC                                    461
    LSTKKVLQSK  N┊MDHAKAREV  LKIAKEKAQK  KQNETSTSKN  AHSKVHFTR R  YQNPSSGSLP

       475
    PRV R LKPQRY  RNEENDSSLK  TGLEKMRSGK  MAPKPQSRCT  STRSAGEAPS  ENQSPSKGPE


    EASSEVQDTN  EVHVPGDQDE  PQTLGKKGSK  NNISVYMTLN  QKK┊574
```
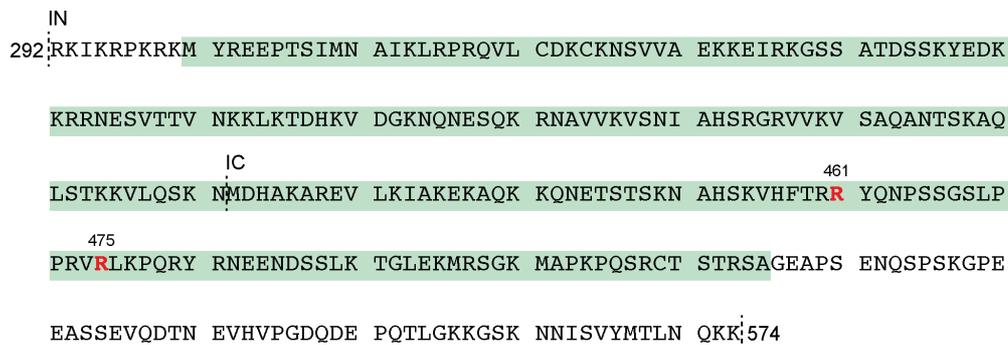
Figure 4.2.1 The internal domain of PWWP2A. The internal domain (aa 291 – 573) is functionally separable: the N-terminal half (IN) strongly binds nucleosomes without H2A variant specificity, whereas the C-terminal half (IC) preferentially binds H2A.Z-containing nucleosomes. Two residues within this IC part are frequently mutated in different cancers (R461Q and R475C) and are highlighted in red. The cross species-conserved conI region is highlighted in green.

Fascinating is its further separation into two subdomains, IN and IC (**Figure 4.2.1**), with distinct functionality: whereas IN facilitates variant-unspecific nucleosome binding, IC is able to specifically attract H2A.Z-containing nucleosomes. Moreover, a basic local alignment

search by BLAST revealed that no other protein (except other vertebrate PWWP2A proteins) possesses a domain comparable to IC or the primary I region. These results make it highly attractive to speculate that the IC domain endows PWWP2A with unique abilities to interact with H2A.Z nucleosomes. Notably, the IC domain harbors two residues frequently mutated in cancer, R461Q and R475C, as identified by cBioPortal ((Cerami et al., 2012; Gao et al., 2013) and E. Bernstein personal communication) (**Figure 4.2.1**). Potentially, these amino acids are involved in determining H2A.Z nucleosome specificity, which could be investigated by *in vitro* and *in vivo* binding assays employing full-length PWWP2A or only the IC domain with these residues mutated. In order to investigate the actual mode of interaction, we started a collaboration with Prof. Joel Mackay from the University of Sydney, Australia, to solve the structure of the internal domain. Preliminary structure predictions suggest, that it is largely unstructured but possesses an in part helical secondary structure (data not shown). Potentially, interaction with a binding partner, for example the H2A.Z nucleosome, could render it structured. Therefore, an attempt to co-crystalize it with nucleosomes appears reasonable. Also, MNase ChIP assays with formaldehyde cross-linked chromatin utilizing GFP-tagged internal and internal subdomains followed by immunoblotting or even sequencing could further elucidate PWWP2A's *in vivo* preference for H2A.Z nucleosomes. However, to date it is not clear, how IN binds to nucleosomes and whether IC directly interacts with H2A.Z within the nucleosome or if potentially other factors like PTMs, a certain DNA-sequence or a so far unknown third factor are involved here. Preliminary experiments, utilizing recombinant GST-PWWP2A full-length and truncations and incubating them with nucleosomes that were reconstituted from recombinant octamers with a 601-positioning sequence *in vitro* (Bonisch et al., 2012), showed no preference for recombinant nucleosomes containing H2A.Z.2 or recombinant nucleosomes containing H2A (**Figure 4.2.2**). However, these experiments have to be repeated more carefully including an exact titration of the employed recombinant components to avoid biases. Also, the utilized washing conditions, that implement physiological but rather low salt concentrations (150 mM), might not to be optimal. Actually, the preferential binding of bacterially expressed Swc2, a component of the *S. cerevisaie* SWR1 complex, to H2A.Z-H2B dimers was only reported to be specific under higher salt conditions (300 mM) (Wu et al., 2005). A crystal structure if available in the future might also help to elucidate how and if PWWP2A can distinguish between conventional and H2A.Z nucleosomes. Nevertheless, if the *in vitro* results hold true they suggest that albeit PWWP2A's nucleosome affinity is most reasonably mediated by the IN domain, additional

factors that are not present in the recombinant nucleosome preparations but in *ex vivo* mononucleosomes from cells are needed to gain H2A.Z nucleosome specificity.
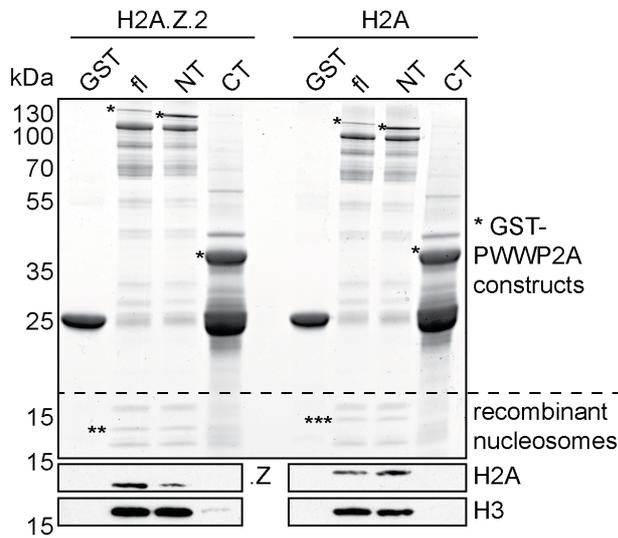
Figure 4.2.2 Recombinant PWWP2A does not posses variant specificity using recombinant nucleosomes. GST-tagged PWWP2A_fl, _NT and _CT (*) were incubated with nucleosomes reconstituted from recombinant octamers containing either H2A.Z.2 (**) or H2A (***) with a 601-positioning sequence *in vitro*. GST-precipitations were separated by SDS-PAGE and gels Coomassie-stained (upper part) or utilized for immunoblotting detecting H2A.Z (.Z), H2A or H3 (boxes below).

In contrast to the *ex vivo* nucleosomes/*in vitro* GST-PWWP2A assays, FRAP experiments with HeLa Kyoto cells stably and transiently transfected with GFP-PWWP2A full-length and truncations showed that only the combined action of both the I but also the PWWP domain is necessary to achieve full chromatin binding (**Figure 3.5.1**). For the PWWP domain alone, no or at least very weak nucleosome binding as suggested by the *in vitro* binding assays could be recapitulated by the FRAP analyses. The internal domain possessed slower FRAP kinetics compared to the PWWP domain alone, but appeared to be faster than full-length PWWP2A as demonstrated *in vitro*. This strongly suggests a synergistic-binding mechanism and assigns a pivotal role to the PWWP domain that implicates strengthening of an otherwise weak interaction. But how is this support achieved? Several mechanisms that are not mutually exclusive can be envisioned and include (1) the interaction with methylated histone lysines, (2) an association with DNA or (3) the binding to a yet unknown bridging factor.

The PWWP domains of several proteins possess the ability to bind methylated histone lysines. A combined alignment of the over 20 human PWWP domains and available structural information for several PWWP domains showed, that although the overall primary sequence is not highly conserved, structural features among theses domains are preserved (Qin and Min, 2014). In general, theses structural features are: a barrel-shaped β-strand core often containing a short η-helix between the β4 and β5, a linker of variable length and structure between β1 and β2 and finally a unique α-helical bundle at the C-terminus consisting of 1 - 6 α-helices (Wu et al., 2011) (**Figure 4.2.3**). Import for methyllysine binding is an aromatic

cage, formed by three varying aromatic amino acids within the β-barrel core, which is also present in the PWWP domain of PWWP2A consisting of phenylalanine (F666), tryptophane (W669) and again tryptophane (W695).

Figure 4.2.3 Modeling of the PWWP2A PWWP domain structure. A potential structure of the PWWP domain of PWWP2A was modeled employing the web browser-based tool iTASSER and visualized with the freely available software Chimera (1.8.0). Template was the already published structure of the PWWP2B PWWP domain (Qin and Min, 2014), Protein Data Bank (PDB) code 4LD6, that covered 95% of all amino acids (see also alignment below). Shown is the core PWWP domain (the β-barrel) as well as the additional C-terminal α-helical features in cartoon view. The β-barrel is colored in red and the β-strands numbered from β1 to β5. The three residues F666, W669 and W695 forming the aromatic cage are highlighted in green and depicted in stick mode. The η-helix between β4 and β5 is colored in blue, as well as the three α-helices (α1 - α3). Structural elements not belonging to a classical PWWP domain are colored in grey. NT = N-terminus, CT = C-terminus. The amino acid alignment of PWWP2A and PWWP2B below the modeled structure was carried out using the web browser-based tool Clustal Omega. Structural features of the PWWP domain of PWWP2A are highlighted in the sequence by color according to the model and schematically represented below the alignment. Consensus symbols: fully conserved residue (*), conserved strongly similar residue (.) and conserved weakly similar residue (:).

In fact, PWWP2A is part of a subgroup within the PWWP domain family that is DNMT3B-related and contains besides DNMT3B also DNTM3A, PWWP2B and Methyl-CpG-binding domain protein 5 (MBD5), the latter not possessing an aromatic cage (**Figure 4.2.4**).

```
                    •    •                              •
PWWP2A 655 VGDIVWAKIYGFPWWPARILTITVSRKDNGLLVRQEARISWFGSPTTSFLALSQLSPFLEN-------------- 715
PWWP2B 490 VGDIVWGKIHGFPWWPARVLDISLGQKEDGEPSWREAKVSWFGSPTTSFLSISKLSPFSEF-------------- 550
DNMT3A 292 IGELVWGKLRGFSWWPGRIVSWWMTGRSRAA--EGTRWVMWFGDGKFSVVCVEKLMPLSSF-------------- 350
DNMT3B 225 IGDLVWGKIKGFSWWPAMVVSWKATSKRQAM--SGMRWVQWFGDGKFSEVSADKLVALGLF-------------- 283
MBD5   1385 VGDLVWGQIKGLTSWPGKLVREDDVHNSCQ-----------QSPEEGKVEPEKLKTLTEGLEAYSRVRKRNRKS 1447
           :*::**.:: *: **. ::      .          .   . : .:*  :
```

Figure 4.2.4 Alignment of the core PWWP domain sequences of DNMT3-related PWWP domain-containing proteins. The PWWP domain amino acid sequences of all five members of the human DNMT3-related PWWP domain-containing protein family (PWWP2A [Q96N64], PWWP2B [Q6NUJ5], DNMT2A [Q9Y6K1], DNMT3B [Q9UBC3] and MBD5 [Q9P267]) as defined by UniProt [ID] were aligned employing Clustal Omega. Residues forming the aromatic cage are colored in green and highlighted with a green circle. The central core motif (PWWP) is highlighted with a grey box. Consensus symbols: fully conserved residue (*), conserved strongly similar residue (.) and conserved weakly similar residue (:).

DNMT3A and DNMT3B are *de novo* methyltransferases targeting DNA methylation to chromatin (Chen et al., 2004) and have been shown to bind trimethylated H3K36 (Baubec et al., 2015; Dhayalan et al., 2010). This of course raised the possibility that PWWP2A also contacts H3K36me3 with its PWWP domain. However, as shown by the already mentioned *in vitro* binding assays, nucleosomes precipitated with the full-length protein were not enriched with H3K36me3 compared to the truncation without the PWWP domain (**Figure 3.3.1B**). Taking into account, that H2A.Z-containing nucleosomes are depleted in H3K36me3 (**Figure 3.1.5D**), that the H3K36 demethylase KDM2A is associated with H2A.Z nucleosomes, (**Table 3.1.1**) and that this mark is generally found in gene bodies (Wagner and Carpenter, 2012), it is not surprising that PWWP2A chromatin interaction might not rely on this modification.

H3K4me3, which is present on H2A.Z nucleosomes (**Figure 3.1.5D**, **Figure 3.1.10** and (Kim et al., 2013)), represents a potential candidate PTM for PWWP domain binding. Albeit not shown before for PWWP domain proteins, it was reported just recently that WDR5 binds to H3K4 di- and trimethylated nucleosomes that contained H2A.Z dependent on its WD repeat domain (Kim et al., 2013). Fascinating but contrariwise, H3K4me3 was depleted on nucleosomes pulled down with the full-length GST-PWWP2A compared to GST-PWWP2A_NT (**Figure 3.3.1B**). Although this finding could be reproduced only in three out of six experiments and needs to be further validated, it suggests that H3K4me3 might repel PWWP2A via the PWWP domain. This idea is also supported by MNase IPs using chromatin from stable GFP-PWWP2A HeLa Kyoto cells, demonstrating in immunoblots the absence of H3K4me3 on nucleosomes precipitated with GFP-PWWP2A (**Figure 3.2.4**). Moreover, the study of Eberl et al. showed PWWP2A to be repelled by H3K4me3 histone tail peptides but attracted by the unmodified H3 tail (Eberl et al., 2013). As mentioned earlier for the identified

repeller complex (RAI1, TCF20, HMG20A and PHF14), these results appear not intuitive, because PWWP2A is attracted by nucleosomes containing H2A.Z that are coincidentally H3K4 trimethylated in HeLa Kyoto and SK-mel147 cell lines according to immunoblot analyses (**Figure 3.1.5D** and **Figure 3.1.10**). In line with H2A.Z's role in manifold biological functions, however, it is very likely that different populations of H2A.Z-containing nucleosomes exist throughout the genome that differ in their localization (TSS, gene body, enhancers, insulators, intergenic regions etc.), possess specific modifications (like H3K4me3 and others) and are occupied by different factors (like PWWP2A, BRD2 or WDR5). In an attempt to further characterize this situation, colocalization of PWWP2A and H2A.Z.1 / H2A.Z.2 nucleosomes was investigated by native MNase ChIP sequencing. This assay clearly associated PWWP2A with H2A.Z nucleosomes at the TSS of active genes (**Figure 3.2.8**) and also correlated it with the presence of H3K4me3. However, mapping of the latter mark was achieved in HeLa cells and not by our group (GEO accession GSM733682). Interestingly though, it appears, as if the distribution of the H3K4me3 mark is shifted towards the +1 nucleosome (**Figure 3.2.9**), whereas H2A.Z is rather equally distributed around the TSS, potentially allowing PWWP2A to interact preferentially with the -1 nucleosome. In fact, the TSS pile up visualizes the tendency of PWWP2A to enrich on the -1 nucleosome (**Figure 3.2.9**). Determining H3K4me3's (and other's) localization by ChIP sequencing in the cell lines used in the presented thesis and correlating these data with H2A.Z occupancy and genomic localization of PWWPA (and other interactors) could help to further investigate this issue. Even a subnucleosomal resolution could be obtained employing the recently established 'upgrade' of ChIP sequencing, ChIP-exo sequencing (Rhee et al., 2014), which would allow for more precise and even base-pair exact mapping. Furthermore, the applied native MNase ChIP-sequencing approach favors MNase accessible and generally more open chromatin. Therefore, a chromatin shearing-based ChIP-sequencing assay could be applied in future studies, presumably allowing a more detailed look at potential PWWP2A associations with so far underrepresented compact chromatin structures. Finally, one has to keep in mind that the aforementioned H3K4me3 repelling data ground on assays utilizing peptide tails only (Eberl et al., 2013). They could be repeated employing assembled nucleosomes carrying H3K4me3 (or other interesting PTMs) and could also include H2A.Z to determine if one chromatin mark can outcompete the other with respect to binding and how combinations of H2A.Z and different PTMs could change the interactome composition.

Albeit H3K36me3 appears to be favored by many PWWP domain-containing proteins, it might also be possible that the PWWP domain of PWWP2A binds, if at all, other

methyllysine modifications. H4K20me1, for example, was reported to attract the fission yeast PWWP domain protein Pdp1 (Wang et al., 2009). Also, H4K20me3 potentially mediates binding as shown for HDGF2 (Wu et al., 2011). However, preliminary results suggest, that nucleosomes precipitated with GST-PWWP2A_fl are not enriched in these marks compared to the PWWP domain-deletion (GST-PWWP2A_NT) (data not shown). In addition, HDGF2 is able to interact with H3K79me3 peptides, another potential PTM (Wu et al., 2011). Interestingly, albeit the tendency of this mark is to be enriched at gene bodies it is also found at promoters of genes (Barski et al., 2007), therefore potentially also colocalizing with H2A.Z in the employed cell lines, but future work must show, if this PTM plays a role in PWWP2A recruitment. An attempt to incubate recombinant GST-PWWP2A_fl on peptide arrays to elucidate a potential affinity for certain PTMs did not yield any reasonable results so far but could be addressed again in the future.

Second, PWWP2A could potentially bind free (linker) DNA with its PWWP domain, thereby stabilizing its chromatin interaction. The PWWP domains of several proteins like DNMT3B, HDGF, PSIP1 or DNA mismatch repair protein Msh6 (MSH6) were actually shown to enable DNA binding (Laguri et al., 2008; Qiu et al., 2002; Shun et al., 2008; Yang and Everett, 2007). To answer the question whether the PWWP domain of PWWP2A is able to bind DNA, preliminary Electrophoretic Mobility Shift Assays (EMSA) were conducted. At first, recombinant GST-PWWP2A full-length protein was incubated with increasing amounts of DNA. As a positive control, the yeast heat shock protein Hsp15 that is known to possess a strong ability to bind naked DNA (Korber et al., 1999) was included. Next, the potential formation of a protein-DNA complex was analyzed by native gel electrophoresis. Interestingly, when using high amounts of the GST-protein (1000 – 3000 nM) a faint shift could be observed (**Figure 4.2.5A**) Keeping these conditions, also recombinant GST-PWWP2A_NT and _CT were tested for their ability to shift DNA mobility (**Figure 4.2.5B**) No such shift could be detected with the PWWP2A truncation that lacks the PWWP domain (GST-PWWP2A_NT). However, GST-PWWP2A_CT, which contains only the PWWP domain and the very PWWP2A C-terminus, as also GST-PWWP2A full-length showed a rather faint DNA shift. These first results suggest, that the PWWP domain potentially owns a weak binding affinity towards DNA and therefore might be needed to establish the observed rigid interaction of PWWP2A with chromatin. Maybe this interaction is too weak to pull down nucleosomes, which could explain the results obtained in the presented *in vitro* binding experiments with mononucleosomes derived from cells.
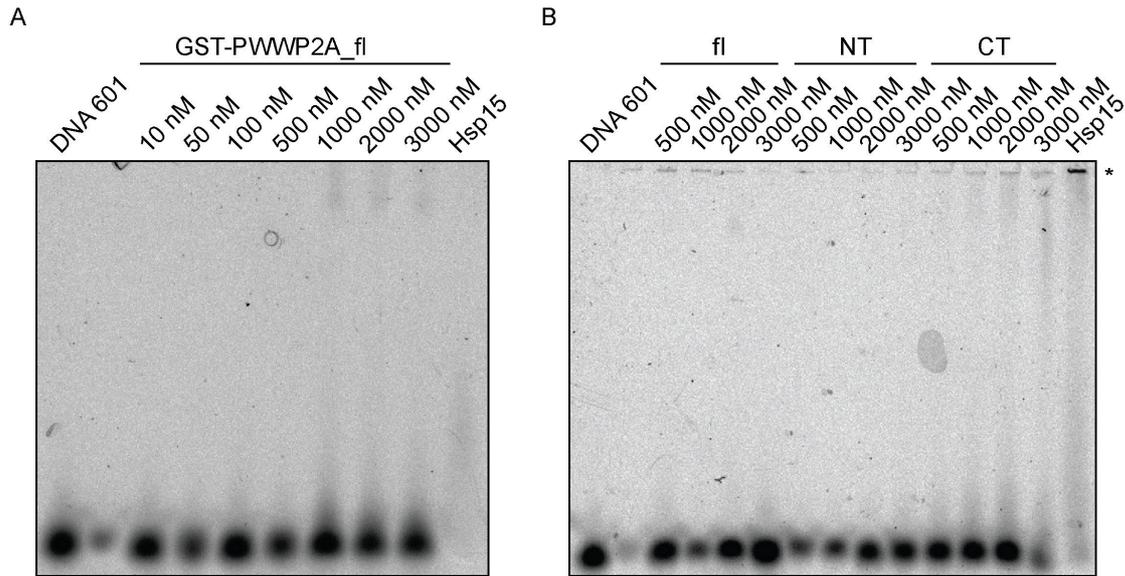
Figure 4.2.5 The PWWP domain of PWWP2A weakly binds DNA. (A) EMSA assay with recombinant GST-PWWP2A full-length protein incubated in increasing amounts with DNA (601-sequence 25mer). As a positive control, the yeast heat shock protein Hsp15 (Korber et al., 1999) was included. Native gel electrophoresis and subsequent imaging with the PhosphoImager analyzed the potential formation of a protein-DNA complex. **(B)** Keeping the conditions, also recombinant GST-PWWP2A_NT and _CT were tested for their ability to shift DNA mobility. (*) Immobile sample in the gel pockets. Preliminary data.

Furthermore, it might be possible that a naked DNA stretch longer then 25bp is necessary to mediate stronger binding. This could include the linker DNA that connects two nucleosomes but also the NDR at the TSS of genes, the latter possibly explaining the observed enrichment of PWWP2A at the NDR (**Figure 3.2.9**). However, comparison of electrostatic potential surfaces of different PWWP domain-containing proteins (PSIP1, HDGF and MSH6) revealed a positively charged DNA-binding surface, which potentially enables the domain to sequence-independently bind the negatively charged phosphate backbone of the DNA through electrostatic interactions (Qin and Min, 2014). In fact, the PWWP domain of PWWP2A has a theoretical isoelectrical point of 9.7 (according to the web browser-based ExPASy tool ProtParam), demonstrating its positive charge. Furthermore, predicting a 3D structure for the PWWP domain with the web browser-based software iTASSER (Roy et al., 2010) and calculating its electrostatic potential surface with the free protein structure software Chimera (1.8.0) revealed that also PWWP2A possesses a positively charged surface, potentially involved in DNA-binding (**Figure 4.2.6**).

Interestingly, dual-binding mechanisms involving both DNA and methyllysine binding are shown *in vitro* for the chromodomains of Male-specific lethal 3 (MSL3) or RBBP1 (Gong et al., 2012; Kim et al., 2010), or the Tudor domain of PHF1 (Musselman et al., 2013).

Moreover, the *in vitro* affinity of the PWWP domain of PSIP1 for H3K36me3-methylated nucleosomes was four orders of magnitude higher than for a H3K36me3 peptide and two orders of magnitude higher than for DNA only (van Nuland et al., 2013b), raising the importance of the presence of an entire and modified nucleosome.
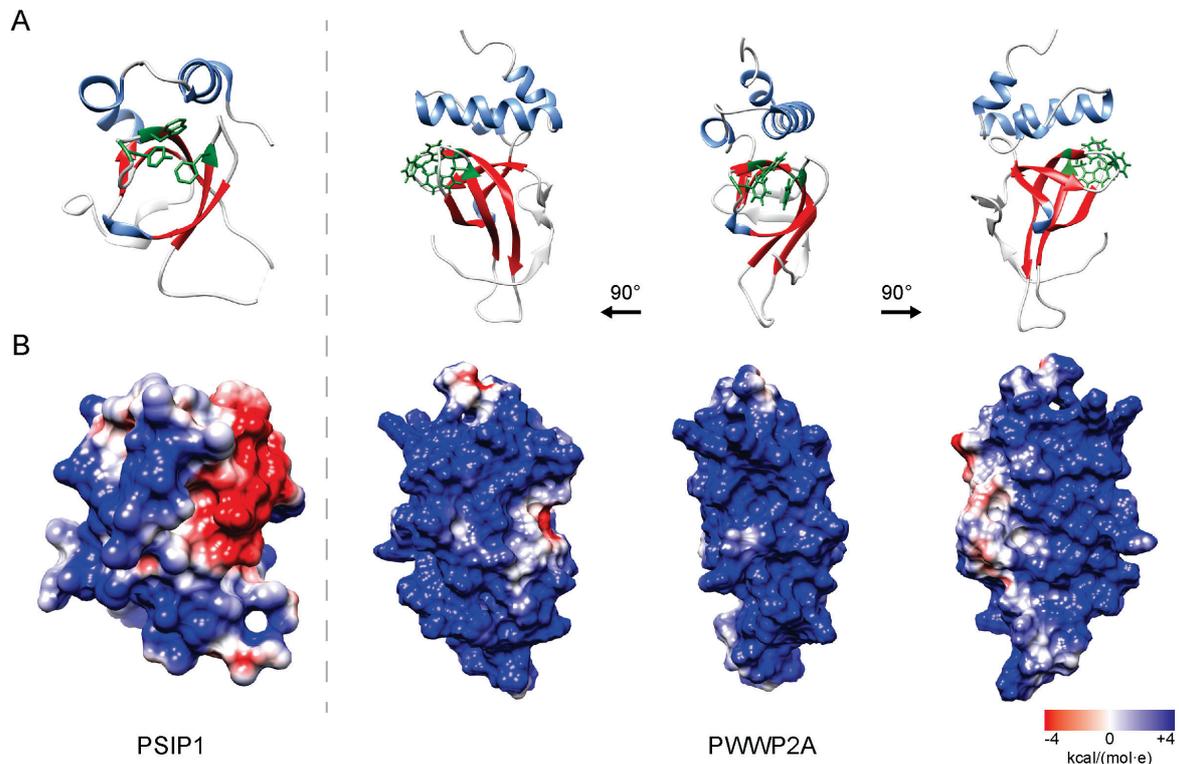
Figure 4.2.6 Electrostatic Surface Potential (ESP) modeling of the PWWP2A PWWP domain reveals positive charged surface. A potential structure of the PWWP domain of PWWP2A was modeled and visualized as described in **Figure 4.2.3** and compared to the published structure of the DNA-binding PWWP domain of PSIP1 ((Eidahl et al., 2013), PDB code 4FU6). **(A)** Cartoon view of the PWWP domain of PSIP1 and PWWP2A with specific structural features highlighted according to **Figure 4.2.3**. **(B)** The ESP of PSIP1 and PWWP2A was calculated utilizing the Coulombic Surface Coloring algorithm, which is part of the Chimera (1.8.0) software package, using default settings. The displayed ESP of the PSIP1 PWWP domain recapitulates the experimentally established and published electrostatic potential mapped to the surface of the domain (Eidahl et al., 2013) that revealed a positive charge (blue). A large part of the surface of the PWWP2A PWWP domain is also positively charged (blue) according to the simulation of the ESP, potentially offering an interface for DNA-interaction. For better visualization of the respective part of the domain surface, it was rotated by 90° clock- and counterclockwise. ESP color values are in units of kcal/(mol*e) at 298 K.

Third, a binding partner that is shared by PWWP2A and H2A.Z could potentially indirectly link PWWP2A with H2A.Z nucleosomes. This must not necessarily depend on the PWWP domain and could involve also the IN or even the IC domain (the latter with respect to H2A.Z nucleosome specificity). However, when the PWWP domain was described first in WHSC1 it was predicted to be a protein-protein interaction domain (Stec et al., 2000). In fact,

the PWWP domain of DNMT3A was shown to directly bind Sal-like protein 3 (SALL3), an inhibitor of the DNA methyltransferase (Shikauchi et al., 2009). Interestingly, my PWWP2A interactome data set suggests that PWWP2A is part of a larger protein network (**Figure 3.6.2**). Interestingly, there are several proteins identified that are also part of the H2A.Z interactome (**Table 3.6.1**). They are potential candidates promoting PWWP2A's interaction with H2A.Z nucleosome and potentially contribute to PWWP2A's functions (see below). However, future experiments have to demonstrate, whether theses proteins really influence the PWWP2A-H2A.Z connection, whether they are direct interactors and which domain are needed for binding. For example, knockdown of a candidate and subsequent GFP-PWWP2A FRAP or ChIP analyses could show if loss of the candidate leads to decreased PWWP2A chromatin/H2A.Z nucleosome association. Also *in vitro* binding assays with recombinant candidate proteins could shed more light on the involvement of these factors.

In summary, PWW2PA interaction with nucleosomes is established by its internal domain, where the IC part is necessary to achieve H2A.Z nucleosome specificity. The PWWP domain mediates the strength of this interaction by a so far unknown mechanism.

4.3 PWWP2A functions in cell cycle progression

Despite its physical association with H2A.Z and a potential role in H2A.Z-mediated transcriptional regulation, the function of PWWP2A is unclear. However, the specific depletion of PWWP2A transcript by RNAi and the subsequent decrease of PWWP2A protein reproducibly provoked a growth phenotype in HeLa Kyoto cells that is affiliated to the misregulation of the cell cycle. In fact, two effects presumably cause the inhibition of proliferation in a population of cells depleted of PWWP2A: first, a defect in replication, as the number of cells in S-Phase was reduced (**Figure 3.7.2**) and second, a defect in cell division, as cells accumulated in mitosis, particular in metaphase (**Figure 3.7.3** and **Figure 3.7.4**). How can these phenotypes be explained by PWWP2A loss? As PWWP2A's function remains unclear to date, considerations with respect to its role in human cells should include H2A.Z function but also the function of PWWP2A interaction partners identified in this thesis.

Depletion of H2A.Z is lethal in many organisms, emphasizing its fundamental role (Zlatanova and Thakar, 2008). However, in fission and budding yeast, H2A.Z is not essential and knockout leads to impaired proliferation (Carr et al., 1994; Jackson and Gorovsky, 2000). Interestingly, also RNAi mediated H2A.Z knockdown in several human cancers inhibited cell

proliferation, mainly by downregulation of cell-cycle regulatory genes that rely on H2A.Z presence (Kim et al., 2013; Svotelis et al., 2010; Vardabasso et al., 2015). Given that PWWP2A's genomic localization at the TSS of genes correlates with active transcription (**Figure 3.2.8**), it could influence the expression of cell-cycle regulatory genes in concert with or independent of H2A.Z. However, experimental data generated by a stable H2A.Z and/or PWWP2A knockdown by shRNA-mediated RNAi or the deletion of the genes by CRISPR/Cas technology are still missing. Of course, PWWP2A depletion should also be tested in other cell systems, for example in SK-mel147 cells whose proliferation is deregulated upon a specific H2A.Z.2 knockdown (Vardabasso et al., 2015)(under revision). This is of special interest, because in these cells PWWP2A tends to favor H2A.Z.2 nucleosomes over H2A.Z.1 nucleosomes (**Figure 3.1.6**). Notably, preliminary Gene Ontology (GO) analyses of GFP-PWWP2A target genes derived from GFP-PWWP2A MNase ChIP-sequencing (**Figure 3.2.8**) that implement KEGG database pathways (Kanehisa et al., 2012) revealed cell cycle regulatory genes (like CDC7, WEE1, ANAPC7, MYC or CDC14A) as first hit and suggest that PWWP2A might be involved in the transcriptional regulation of theses genes. However, RNA sequencing to unravel the transcriptome upon PWWP2A RNAi could help to elucidate PWWP2A's potential involvement in transcriptional regulation of cell cycle progression in future studies

It is possible, that PWWP2A-mediated transcriptional activity also affects DNA replication. At least H2A.Z has been implicated to regulate DNA replication by influencing the expression of genes important for G_1/S transition in yeast (Dhillon et al., 2006). However, the number of cells in G_1 phase seemed not to be affected upon PWWP2A knockdown (**Figure 3.7.2**). Therefore, another possibility is that the decreased number of cells in S-phase could be explained by a loss of cells through apoptosis upon replicative stress. Nevertheless, first Annexin V FACS assays to detect apoptotic cells did not convincingly show an increase of death cells upon PWWP2A knockdown (**Figure 4.3.1**). On the other hand, PWWP2A could be involved in the recruitment of the replication machinery in S-phase. H2A.Z has been shown to genetically interact with origin recognition complex (ORC) proteins and localizes to origins in *S. cerevisiae* (Dhillon et al., 2006) and *A. thaliana* (Costas et al., 2011). Interestingly, a recent study also mapped human origins to the TSS of genes (Dellino et al., 2013), where also H2A.Z and PWWP2A localize. In addition, the absence of H2A.Z has been associated with increased DNA damage, an increased sensitivity to DNA damage agents during replication (Billon and Cote, 2013) and increased genomic instability (Morillo-Huesca et al., 2010) in *S. cerevisiae*. In this organism, DNA double strand break (DSB) repair by

homologous recombination depends on sumoylated H2A.Z that is incorporated at sites of DNA damage (Kalocsay et al., 2009).
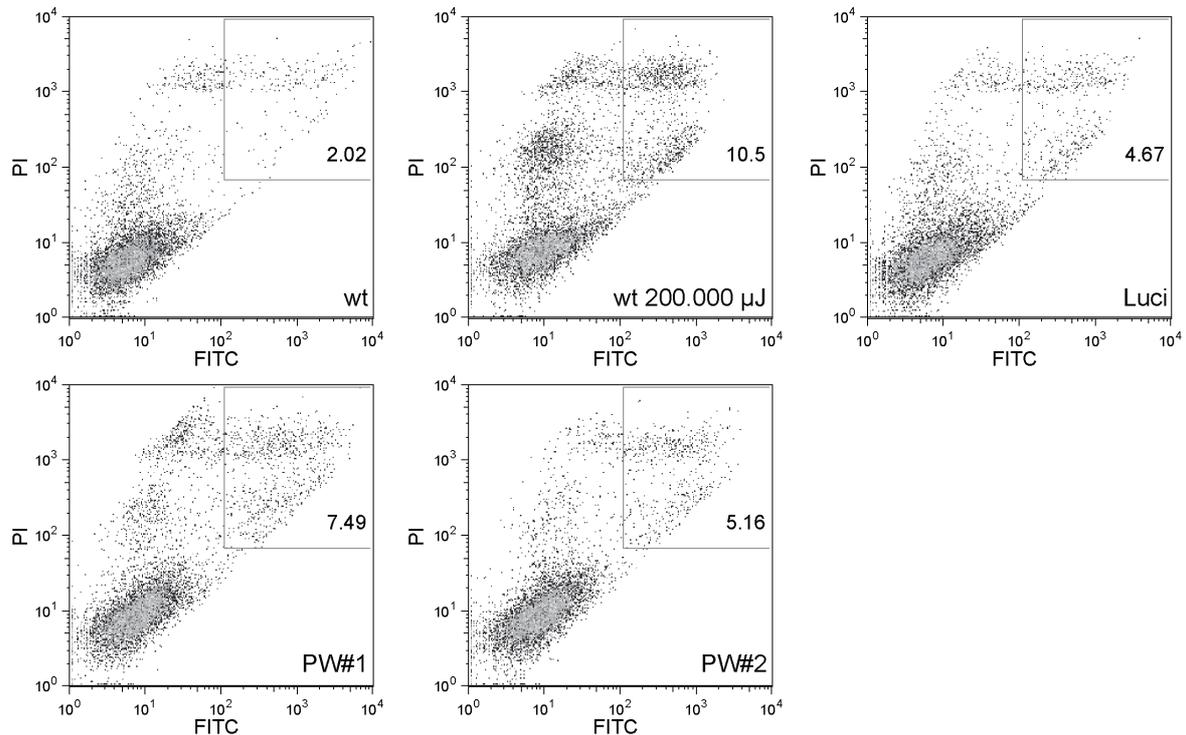
Figure 4.3.1 No induction of apoptosis upon PWWP2A knockdown in HeLa Kyoto cells. Three days upon PWWP2A knockdown with PWWP2A siRNAs PW#1 or PW#2, transfected HeLa Kyoto cells as well as controls (HeLa Kyoto wt cells, HeLa Kyoto wt cells subjected to UV-irradiation [200.000 µJ, UV Stratalinker 1800, Stratagene] and HeLa Kyoto luciferase knockdown cells), were harvested and subjected to Annexin V treatment and PI staining without methanol treatment. Annexin V labeled with a FITC-fluorochrome interacts with phosphatidylserines on the surface of apoptotic cells. PI can only enter cells and stain DNA in the absence of a permeabilizing agent like methanol when cells are apoptotic. Thus, double positive cells (PI and FITC) are apoptotic cells and gated in the upper right corner of the FACS plots. Percentage of the total cell count is depicted. The increase of apoptotic cells upon PWWP2A knockdown compared to wt cells was minimally higher if not comparable to the effect of the unspecific control siRNA (Luci). Treatment with UV-light demonstrated the functionality of the assay. Preliminary data.

However, in human cells H2A.Z is incorporated at DSBs by its deposition machinery (p400/NuA4/TIP60) and thought to locally promote an open chromatin structure and the loading of essential and early DSB repair components like Breast cancer 1 (BRCA1) and Ku70/Ku80 (Xu et al., 2012). It is tempting to speculate that also PWWP2A is involved in DNA damage response (DDR) and loss of its functionality could lead to p53-mediated apoptosis upon replicative stress. Preliminary IF stainings to detect phosphorylation of H2AX, one of the first steps in the DDR, or p53-binding protein 1 (53BP1), another important player in DDR, did not show an increased number of γH2AX or 53BP1 foci in HeLa Kyoto cells depleted of PWWP2A (data not shown). However, Mediator of DNA-damage

checkpoint protein 1 (MDC1), that actually binds γH2AX and is another key component of the DDR signaling cascade recruiting several protein kinases (Lukas et al., 2011b; Reinhardt and Yaffe, 2013), was detected in the presented qMS screen to interact with PWWP2A (**Table 3.6.1**). Furthermore E3 ubiquitin-protein ligase RAD18 and PARP2 were pulled down with PWWP2A, both involved in DNA repair processes (Beck et al., 2014; Yamada et al., 2014), keeping a door open for PWWP2A's potential involvement in DDR. In future experiments, DNA damage could be forced by agents like aphidicolin, which inhibits DNA Polymerase during replication (Lukas et al., 2011a), to study a potential role of PWWP2A in DDR upon replicative stress.

Interestingly, also some members of the NuRD complex interact with PWWP2A (**Table 3.6.1**). This histone deacetylase and nucleosome remodeling complex was reported to be involved in DDR (Allen et al., 2013). Depletion of NuRD subunits leads to hypersensitivity to DNA damage, NuRD tends to accumulate at sites of DSBs dependent on PARP1 and is required to accumulate DDR proteins like BRCA1 (Polo et al., 2010; Smeenk et al., 2010). However, the helicase activity of CHD4 appears to be essential for NuRD's ability to maintain genomic stability (Basta and Rauchman, 2014) but no helicase has been identified in the presented PWWP2A interactome. Moreover, the subunits, which facilitate DNA binding and also interact with each other, GATA2A/2B and MBD2/3, are not present, suggesting the existence of a minimal NuRD (miniNuRD) complex connected with PWWP2A biology. Whether such a miniNuRD lacking DNA-binding and methyl-CpG-binding capability and helicase functionality plays a role in DDR remains unanswered. Nevertheless, the existence of a NuRD complex without the chromatin remodeling components is discussed in the NuRD field (personal communication). A miniNuRD complex, consisting of HDAC1 or 2, the RBBPs 4 and 7, as well as MTA1/2/3 would at least with respect to its composition still be able to assemble, bind histone H4 in nucleosomes and remove acetyl groups from lysine residues on histone tails or other proteins. The latter is of special interest, because it has been shown previously that PWWP2A interacts with HDAC1 (Joshi et al., 2013). The MTA proteins might also support DNA binding, as their SANT, ELM2 and GATA domains are predicted to interact with DNA (Allen et al., 2013). Whereas histone acetylation is known to relax chromatin structure, thereby promoting DNA related processes like transcription and DNA repair, deacetylation is subsequently associated with the compaction of chromatin. In this line, the loss of PWWP2A could impede miniNuRD recruitment and therefore lead to decompacted chromatin. Notably, HeLa cells depleted of PWWP2A show slightly enlarged nuclei, as observed in IF microscopy by DAPI staining (**Figure 3.7.1B**) and FACS analysis

(data not shown), suggesting less compacted chromatin. The decreased number of cells in S-phase upon PWWP2A knockdown could be explained by less compaction. It is well established, that euchromatin is replicated in early S-phase whereas heterochromatin is replicated rather late (Gilbert, 2002). Potentially, opening up chromatin by deregulated histone acetylation could allow for earlier heterochromatin replication and therefore speed up S-phase. Backing up the current FACS and IF data with live cell imaging upon PWWP2A knockdown to analyze the progression through the cell cycle and also including synchronized cell populations in FACS could help to clarify a potential role of PWWP2A during DNA replication. Whether loss of PWWP2A really leads to enhanced chromatin decompaction will also be focus of future work. For example, high-resolution microscopy methods like 3D-SIM could show if PWWP2A can localize to heterochromatic regions and if it colocalize with members of the NuRD complex. H2A.Z knockdown, subsequent IF stainings and confocal microscopy could illuminate whether PWWP2A localization depends on H2A.Z localization in these regions. IF staining of PWWP2A knockdown cells could show if and which histone acetylation pattern is affected and whether heterochromatin components like HP1 or linker histone H1, which is also part of the PWWP2A interactome, are still present. Also FRAP could be employed to measure whether such heterochromatin proteins become more mobile upon PWWP2A knockdown. Furthermore, chromatin could be digested with MNase to find out whether MNase accessibility changes upon PWWP2A knockdown. 'Euchromatinization' could potentially also lead to an upregulation of transcripts, which could be detected by RNA sequencing upon PWWP2A knockdown (see above).

Compaction of chromatin also plays an essential role during mitosis, where chromatin begins to condense during prophase and is present in its most compacted structure as a chromosome during metaphase. Upon PWWP2A knockdown, cells tend to accumulate in metaphase and metaphase chromosomes seemed to look less condensed (**Figure 3.7.4**). Moreover, mislocalization of chromosomes during anaphase were observed (**Figure 4.3.2**) as well as reduction of cells in anaphase (**Figure 3.7.4**). This phenotype is potentially due to an impeded condensation of the chromosomes during onset of mitosis as well as problems with chromatin segregation later on, implicating PWWP2A to function during these processes. A loss of condensin is generally associated with condensation defects during metaphase and persistent entanglement during anaphase (Jeppsson et al., 2014). Interestingly, PWWP2A remains associated with chromosomes during mitosis as detected by IF and confocal microscopy in meta- and anaphase cells (**Figure 3.2.3**). Perhaps, PWWP2A supports recruitment of condensin in vertebrates and reduction of PWWP2A protein on chromatin leads to altered

condensin dynamics. In turn, this could prolong metaphase, as entanglement of chromosomes might be affected during the onset of anaphase.
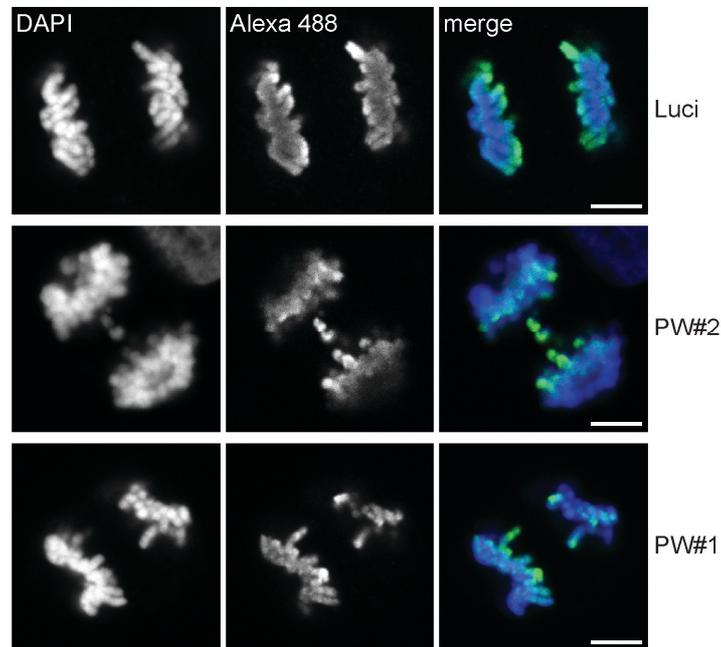
Figure 4.3.2 PWWP2A knockdown HeLa Kyoto cells show impaired anaphases. Two days upon PWWP2A knockdown (siRNAs PW#1 or PW#2), HeLa Kyoto cells were seeded on coverslips, stained to detect H3S10phos (secondary antibody: Alexa 488 coupled) and DNA counterstained with DAPI. Confocal microscopy revealed an increase of anaphase cells that showed mislocalized chromosomes compared to luciferase control (Luci). Scale bars = 10 µm.

Generally, non-functional cohesion abrogates chromosome alignment in metaphase and bipolar attachment to the spindle apparatus (Jeppsson et al., 2014). If PWWP2A plays a role in the cohesion process (dependent of H2A.Z and/or cohesin or not), its absence could lead to a premature sister chromatid separation as reported for H2A.Z mutants in *S. cerevisiae* (Sharma et al., 2013). However, this phenotype was not observed in human PWWP2A knockdown cells. Future experiments will show, whether PWWP2A is involved in structural maintenance of chromosomes (SMC) biology. For example, co-stainings with antibodies targeting SMC components could on the one hand visualize a potential colocalization with PWWP2A and on the other hand demonstrate what happens with this interaction in the case of PWWP2A knockdown. MNase ChIPs of chromatin from mitotic cells and subsequent SMC component immunoblotting could also reveal an interaction with PWWP2A. Beyond its potential implication in SMC biology, PWWP2A could also by other means influence chromatin condensation. Recently, it was reported in *S. cerevisiae*, that recruitment of the histone deacetylase Hst2p to sites of H3S10phos (a hallmark of mitotic chromatin) leads to

removal of the acetyl group from H4K16. Subsequently, this led to interaction of the free H4 tail with neighboring nucleosomes and condensation of chromatin fibers independent of the SMC complexes (Wilkins et al., 2014). Theoretically, PWWP2A could recruit the miniNuRD complex to chromatin, which in turn deacetylates H4. This hypothesis will be tested in future experiments, including IF confocal microscopy, immunoblotting or mass spectrometry to examine the acetylation status of control and PWWP2A knockdown cells and genome-wide mapping of miniNuRD members.

Apart from the discussed role of the SMC complexes in chromosome segregation, the concerted action of a plethora of centromere and kinetochore proteins, modifying enzymes like kinases, the centric and pericentric (hetero)chromatin, histone PTMs but also the mitotic spindle itself is utterly important for proper mitosis. This complex network leaves many potential roles for PWWP2A involvement. Preliminary IF microscopy revealed, however, that upon depletion of PWWP2A, localization of CENP-A, tubulin, Aurora B and phosphorylated H3S10 appear to be comparable to control cells (data not shown). Future confocal microscopy and potentially 3D-SIM will be employed to have a closer look at theses candidates. It will also be of interest to investigate if there is a connection between PWWP2A and pericentric heterochromatin (PCH). This heterochromatic region flanking the centromeres is seen as boundary, separating euchromatin from kinetochore chromatin, it possesses distinct chromatin marks (DNA methylation, H4K20me3, H3K9me2/3, HP1) and is able to recruit cohesin (Boyarchuk et al., 2011; Fukagawa and Earnshaw, 2014). Interestingly, ATRX, a protein found to interact with PWWP2A (**Table 3.6.1**), is highly enriched at PCH in mouse and human cells (McDowell et al., 1999), where it mediates H3.3 incorporation at least in mouse ES cells (Szenker et al., 2011). Additionally, ATRX is implicated in chromosome segregation (Ritchie et al., 2008) and depletion of ATRX was shown to lead to reduced sister chromatid cohesion and a delay in metaphase. Moreover, the NuRD complex is involved in PCH assembly and maintenance (Sims and Wade, 2011). Notably, also the NuRD complex is, as PWWP2A, repelled by H3K4me3 (Eberl et al., 2013), which nicely fits to a chromatin domain that is heterochromatic. Strikingly, H2A.Z is part of PCH, too (Boyarchuk et al., 2014; Rangasamy et al., 2003), but also of centromeric chromatin (Greaves et al., 2007), where its ability to generate compact secondary structures was suggested to be important for the special centromeric 3D structure including also HP1 (Verni and Cenci, 2015). It is tempting to speculate, that PWWP2A is also part of PCH, thereby possibly influencing centromere function.

Perhaps, PWWP2A even has a far more direct impact on mitosis as yet discussed. Repeatedly, and independent from each other, our collaboration partners working with GFP-PWWP2A cell lines in FRAP assays reported a localization of PWWP2A to the spindle apparatus (**Figure 4.3.3**) during mitosis. If this holds true and is not evoked by ectopically expressing the GFP-tagged protein, it suggests that PWWP2A interacts with the spindle and might be important for its stability during chromosome segregation.
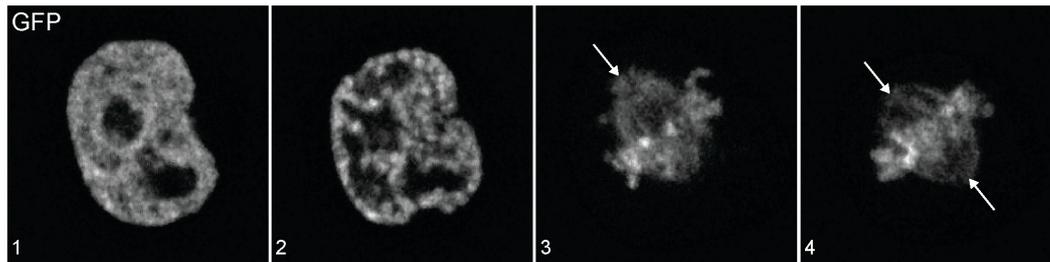
Figure 4.3.3 GFP-PWWP2A attaches to the spindle during mitosis. HeLa Kyoto cells stably expressing GFP-PWWP2A were imaged with a spinning disc microscope applying FRAP assay / live cell imaging conditions. Images were taken every 30 minutes. Depicted is one example cell transitioning from interphase (1) over prophase (2) to metaphase (3 and 4). White arrows depict GFP-PWWP2A signal at the mitotic spindle. Preliminary data.

In summary, PWWP2A influences cell cycle progression by a to date unknown mechanism. However, it is very likely that the afore discussed recruitment of specific complexes carrying distinct enzymatic functionalities, for example a deacetylating miniNuRD complex, or single proteins involved in certain cellular processes like MDC1 or ATRX, to sites where PWWP2A interconnects with a certain chromatin landscape (may this be independent or dependent of H2A.Z and the presence of particular PTMs like H3K4me3) enables PWWP2A to influence key biological processes like transcription, DNA repair and mitosis.

ABBREVIATIONS

4HB	Four helix bundle
aa	Amino acid
AOTF	Acousto Optic Tunable Filter
AR	Androgen receptor
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
CAA	2-chloroacetamide
ChIP	Chromatin immunoprecipitation
CP	Crossing point
CpG	Cytosine phosphate guanine
DAPI	4',6-diamidino-2-phenylindole
DDR	DNA damage repair
DMEM	Dulbecco's modified Eagle medium
DMSO	Dimethylsulfoxide
DSB	Double strand break
EMSA	Electrophoretic mobility shift assay
ES	Embryonic stem
ESP	Electrostatic surface potential
FACS	Fluorescence Activated Cell Sorting
FCS	Fetal calf serum
FDR	False discovery rate
FRAP	Fluorescence Recovery After Photobleaching
FSC	Forward scatter
G1	Gap 1
G2	Gap 2
GEO	Gene expression omnibus
GFP	Green fluorescent protein
GO	Gene ontology
HCD	High energy collisional
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HFD	Histone-fold domain
HK	HeLa Kyoto
HRP	Horseradish peroxidase
IF	Immunofluorescence
IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilo base
kDa	Kilo Dalton
L1	Loop 1
LC-MS/MS	Liquid chromatography-mass spectrometry
LFQ	Label-free quantification
Luci	Luciferase

M	Mitosis
MBT	Malignant brain tumor
MNase	Micrococcal nuclease
ncRNAs	non-coding RNAs
NDR	Nucleosome-depleted region
ON	Over night
P	Pellet
P/S	Penicillin/streptomycin
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCH	Pericentric heterochromatin
PDB	Protein data bank
PMT	Photomultiplier tube
PTM	Posttranslational modification
PWWP	(P) proline - (W) tryptophane - (W) tryptophane - (P) proline
qMS	Quantitative mass spectrometry
qPCR	Quantitative PCR
RNAi	RNA interference
RT	Room temperature
S	Synthesis
S1	Supernatant 1
S2	Supernatant 2
SANT	Swi3, Ada2, N-Cor and TFIIIB
SDS	Sodium dodecyl sulfate
siRNA	Short interfering RNA
SK	SK-mel147
SMC	Structural maintenance of chromosomes
SSC	Sideward scatter
TFA	Trifluoroacetic acid
TSA	Trichostatin A
TSS	Transcription start site
TTS	Transcription termination site
UTR	Untranslated region
WD40	(W) tryptophan - (D) aspartic acid repeat
wt	Wild type
α	Anti

BIBLIOGRAPHY

Adams-Cioaba, M.A., and Min, J. (2009). Structure and function of histone methylation binding proteins. Biochemistry and cell biology = Biochimie et biologie cellulaire *87*, 93-105.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Allen, H.F., Wade, P.A., and Kutateladze, T.G. (2013). The NuRD architecture. Cellular and molecular life sciences : CMLS *70*, 3513-3524.

Allis, C.D., Richman, R., Gorovsky, M.A., Ziegler, Y.S., Touchstone, B., Bradley, W.A., and Cook, R.G. (1986). hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. The Journal of biological chemistry *261*, 1941-1948.

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R.T., Gaudreau, L.*, et al.* (2010). NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. The Journal of biological chemistry *285*, 15966-15977.

Andrews, A.J., and Luger, K. (2011). Nucleosome structure(s) and stability: variations on a theme. Annual review of biophysics *40*, 99-117.

Angelov, D., Molla, A., Perche, P.Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P., and Dimitrov, S. (2003). The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. Molecular cell *11*, 1033-1041.

Ausio, J., and Abbott, D.W. (2002). The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function. Biochemistry *41*, 5945-5949.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. J Exp Med *79*, 137-158.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research *21*, 381-395.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Basta, J., and Rauchman, M. (2014). The nucleosome remodeling and deacetylase complex in development and disease. Translational research : the journal of laboratory and clinical medicine.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature.

Beck, C., Robert, I., Reina-San-Martin, B., Schreiber, V., and Dantzer, F. (2014). Poly(ADP-ribose) polymerases in double-strand break repair: focus on PARP1, PARP2 and PARP3. Experimental cell research *329*, 18-25.

Bellucci, L., Dalvai, M., Kocanova, S., Moutahir, F., and Bystricky, K. (2013). Activation of p21 by HDAC inhibitors requires acetylation of H2A.Z. PloS one *8*, e54102.

Bernstein, E., and Hake, S.B. (2006). The nucleosome: a little variation goes a long way. Biochemistry and cell biology = Biochimie et biologie cellulaire *84*, 505-517.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proceedings of the National Academy of Sciences of the United States of America *80*, 5559-5563.

Bierne, H., Tham, T.N., Batsche, E., Dumay, A., Leguillou, M., Kerneis-Golsteyn, S., Regnault, B., Seeler, J.S., Muchardt, C., Feunteun, J.*, et al.* (2009). Human BAHD1 promotes heterochromatic gene silencing. Proceedings of the National Academy of Sciences of the United States of America *106*, 13826-13831.

Billon, P., and Cote, J. (2013). Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. Biochimica et biophysica acta *1819*, 290-302.

Binda, O., Sevilla, A., LeRoy, G., Lemischka, I.R., Garcia, B.A., and Richard, S. (2013). SETD6 monomethylates H2AZ on lysine 7 and is required for the maintenance of embryonic stem cell self-renewal. Epigenetics : official journal of the DNA Methylation Society *8*, 177-183.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Black, B.E., and Cleveland, D.W. (2011). Epigenetic centromere propagation and the nature of CENP-a nucleosomes. Cell *144*, 471-479.

Bonisch, C., and Hake, S.B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research *40*, 10719-10741.

Bonisch, C., Nieratschker, S.M., Orfanos, N.K., and Hake, S.B. (2008). Chromatin proteomics and epigenetic regulatory circuits. Expert review of proteomics *5*, 105-119.

Bonisch, C., Schneider, K., Punzeler, S., Wiedemann, S.M., Bielmeier, C., Bocola, M., Eberl, H.C., Kuegel, W., Neumann, J., Kremmer, E.*, et al.* (2012). H2A.Z.2.2 is an alternatively

spliced histone H2A.Z variant that causes severe nucleosome destabilization. Nucleic acids research *40*, 5951-5964.

Boyarchuk, E., de Oca, R.M., and Almouzni, G. (2011). Cell cycle dynamics of histone variants at the centromere, a model for chromosomal landmarks. Current opinion in cell biology *23*, 266-276.

Boyarchuk, E., Filipescu, D., Vassias, I., Cantaloube, S., and Almouzni, G. (2014). The histone variant composition of centromeres is controlled by the pericentric heterochromatin state during the cell cycle. Journal of cell science *127*, 3347-3359.

Bruce, K., Myers, F.A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D.J., Thorne, A.W., and Crane-Robinson, C. (2005). The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. Nucleic acids research *33*, 5633-5639.

Buschbeck, M., and Di Croce, L. (2010). Approaching the molecular and physiological function of macroH2A variants. Epigenetics : official journal of the DNA Methylation Society *5*, 118-123.

Bustin, M., Catez, F., and Lim, J.H. (2005). The dynamics of histone H1 function in chromatin. Molecular cell *17*, 617-620.

Campos, E.I., Stafford, J.M., and Reinberg, D. (2014). Epigenetic inheritance: histone bookmarks across generations. Trends in cell biology *24*, 664-674.

Carr, A.M., Dorrington, S.M., Hindley, J., Phear, G.A., Aves, S.J., and Nurse, P. (1994). Analysis of a histone H2A variant from fission yeast: evidence for a role in chromosome stability. Molecular & general genetics : MGG *245*, 628-635.

Caterino, T.L., and Hayes, J.J. (2011). Structure of the H1 C-terminal domain and function in chromatin condensation. Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire *89*, 35-44.

Cerami, E., Gao, J.J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E.*, et al.* (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov *2*, 401-404.

Chadwick, B.P., and Willard, H.F. (2001a). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Human molecular genetics *10*, 1101-1113.

Chadwick, B.P., and Willard, H.F. (2001b). A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. The Journal of cell biology *152*, 375-384.

Chakravarthy, S., Bao, Y., Roberts, V.A., Tremethick, D., and Luger, K. (2004). Structural characterization of histone H2A variants. Cold Spring Harbor symposia on quantitative biology *69*, 227-234.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X.*, et al.* (2013). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes & development *27*, 2109-2124.

Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology *24*, 9048-9058.

Ciabrelli, F., and Cavalli, G. (2014). Chromatin-Driven Behavior of Topologically Associating Domains. Journal of molecular biology.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annual review of biochemistry *78*, 273-304.

Conerly, M.L., Teves, S.S., Diolaiti, D., Ulrich, M., Eisenman, R.N., and Henikoff, S. (2010a). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome research *20*, 1383-1390.

Conerly, M.L., Teves, S.S., Diolaiti, D., Ulrich, M., Eisenman, R.N., and Henikoff, S. (2010b). Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome research *20*, 1383-1390.

Coon, J.J., Ueberheide, B., Syka, J.E., Dryhurst, D.D., Ausio, J., Shabanowitz, J., and Hunt, D.F. (2005). Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *102*, 9463-9468.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Costas, C., de la Paz Sanchez, M., Stroud, H., Yu, Y., Oliveros, J.C., Feng, S., Benguria, A., Lopez-Vidriero, I., Zhang, X., Solano, R.*, et al.* (2011). Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. Nature structural & molecular biology *18*, 395-400.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology *26*, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res *10*, 1794-1805.

Cuddapah, S., Barski, A., Cui, K., Schones, D.E., Wang, Z., Wei, G., and Zhao, K. (2009). Native chromatin preparation and Illumina/Solexa library construction. Cold Spring Harbor protocols *2009*, pdb prot5237.

Dai, Q., and Wang, H. (2006). "Cullin 4 makes its mark on chromatin". Cell division *1*, 14.

de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A. (2005). Do protein motifs read the histone code? Bioessays *27*, 164-175.

Delatte, B., Deplus, R., and Fuks, F. (2014). Playing TETris with DNA modifications. The EMBO journal *33*, 1198-1211.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome research *23*, 1-11.

Denis, H., Ndlovu, M.N., and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO reports *12*, 647-656.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. The Journal of biological chemistry *285*, 26114-26120.

Dhillon, N., Oki, M., Szyjka, S.J., Aparicio, O.M., and Kamakaka, R.T. (2006). H2A.Z functions to regulate progression through the cell cycle. Molecular and cellular biology *26*, 489-501.

Ding, Z.H., Gillespie, L.L., and Paterno, G.D. (2003). Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. Molecular and cellular biology *23*, 250-258.

Draker, R., Ng, M.K., Sarcinella, E., Ignatchenko, V., Kislinger, T., and Cheung, P. (2012). A combination of H2A.Z and H4 acetylation recruits Brd2 to chromatin during transcriptional activation. PLoS genetics *8*, e1003047.

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

Dryhurst, D., Ishibashi, T., Rose, K.L., Eirin-Lopez, J.M., McDonald, D., Silva-Moreno, B., Veldhoen, N., Helbing, C.C., Hendzel, M.J., Shabanowitz, J.*, et al.* (2009). Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates. BMC biology *7*, 86.

Dryhurst, D., McMullen, B., Fazli, L., Rennie, P.S., and Ausio, J. (2012). Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated

transcription and is upregulated in a model of prostate cancer progression. Cancer letters *315*, 38-47.

Eberharter, A., and Becker, P.B. (2004). ATP-dependent nucleosome remodelling: factors and functions. Journal of cell science *117*, 3707-3711.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Molecular cell *49*, 368-378.

Ederveen, T.H., Mandemaker, I.K., and Logie, C. (2011). The human histone H3 complement anno 2011. Biochimica et biophysica acta *1809*, 577-586.

Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S.*, et al.* (2013). Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic acids research *41*, 3924-3936.

Eirin-Lopez, J.M., Gonzalez-Romero, R., Dryhurst, D., Ishibashi, T., and Ausio, J. (2009). The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. BMC evolutionary biology *9*, 31.

Eirin-Lopez, J.M., Ishibashi, T., and Ausio, J. (2008). H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. Faseb J *22*, 316-326.

Elsasser, S.J., and D'arcy, S. (2012). Towards a mechanism for histone chaperones. Bba-Gene Regul Mech *1819*, 211-221.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X.L., Wang, L., Issner, R., Coyne, M.*, et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-U52.

Faast, R., Thonglairoam, V., Schulz, T.C., Beall, J., Wells, J.R., Taylor, H., Matthaei, K., Rathjen, P.D., Tremethick, D.J., and Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. Current biology : CB *11*, 1183-1187.

Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. Molecular cell *16*, 655-661.

Fan, Y.H., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. Cell *123*, 1199-1212.

Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J.*, et al.* (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell *143*, 212-224.

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

Fox, M.H. (1980). A Model for the Computer analysis of Synchronous DNA Distributions Obtained by Flow Cytometry. Cytometry *1*.

Fukagawa, T., and Earnshaw, W.C. (2014). The centromere: chromatin foundation for the kinetochore machinery. Developmental cell *30*, 496-508.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev *15*, 490-495.

Gao, J.J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y.C., Jacobsen, A., Sinha, R., Larsson, E.*, et al.* (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal *6*.

Gevry, N., Chan, H.M., Laflamme, L., Livingston, D.M., and Gaudreau, L. (2007). p21 transcription is regulated by differential localization of histone H2A.Z. Genes & development *21*, 1869-1881.

Gevry, N., Hardy, S., Jacques, P.E., Laflamme, L., Svotelis, A., Robert, F., and Gaudreau, L. (2009). Histone H2A.Z is essential for estrogen receptor signaling. Genes & development *23*, 1522-1533.

Gilbert, D.M. (2002). Replication timing and transcriptional control: beyond cause and effect. Current opinion in cell biology *14*, 377-383.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Gong, W., Zhou, T., Mo, J., Perrett, S., Wang, J., and Feng, Y. (2012). Structural insight into recognition of methylated histone tails by retinoblastoma-binding protein 1. The Journal of biological chemistry *287*, 8531-8540.

Greaves, I.K., Rangasamy, D., Ridgway, P., and Tremethick, D.J. (2007). H2A.Z contributes to the unique 3D structure of the centromere. Proceedings of the National Academy of Sciences of the United States of America *104*, 525-530.

Guillemette, B., Bataille, A.R., Gevry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005). Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS biology *3*, e384.

Guo, J.U., Su, Y.J., Zhong, C., Ming, G.L., and Song, H.J. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell *145*, 423-434.

Hake, S.B., Garcia, B.A., Kauer, M., Baker, S.P., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2005). Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. Proceedings of the National Academy of Sciences of the United States of America *102*, 6344-6349.

Han, J., Zhang, H., Zhang, H., Wang, Z., Zhou, H., and Zhang, Z. (2013). A Cul4 E3 ubiquitin ligase regulates histone hand-off during nucleosome assembly. Cell *155*, 817-829.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q.Y., Ding, J.P., Jia, Y.Y., Chen, Z.C., Li, L.*, et al.* (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303-1307.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95-109.

Henikoff, S. (2009). Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. Nature genetics *41*, 865-866.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome research *19*, 460-469.

Henikoff, S., and Shilatifard, A. (2011). Histone modification: cause or cog? Trends Genet *27*, 389-396.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474-484.

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404-408.

Hu, G., and Wade, P.A. (2012). NuRD and pluripotency: a complex balancing act. Cell stem cell *10*, 497-503.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of cell biology *189*, 739-754.

Iouzalen, N., Moreau, J., and Mechali, M. (1996). H2A.ZI, a new variant histone expressed during Xenopus early development exhibits several distinct features from the core histone H2A. Nucleic acids research *24*, 3947-3952.

Ishibashi, T., Dryhurst, D., Rose, K.L., Shabanowitz, J., Hunt, D.F., and Ausio, J. (2009). Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome. Biochemistry *48*, 5007-5017.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300-1303.

Izzo, A., Kamieniarz-Gdula, K., Ramirez, F., Noureen, N., Kind, J., Manke, T., van Steensel, B., and Schneider, R. (2013). The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. Cell reports *3*, 2142-2154.

Jack, A.P., Bussemer, S., Hahn, M., Punzeler, S., Snyder, M., Wells, M., Csankovszki, G., Solovei, I., Schotta, G., and Hake, S.B. (2013). H3K56me3 is a novel, conserved heterochromatic mark that largely but not completely overlaps with H3K9me3 in both regulation and localization. PloS one *8*, e51765.

Jack, A.P., and Hake, S.B. (2014). Getting down to the core of histone modifications. Chromosoma *123*, 355-371.

Jackson, J.D., and Gorovsky, M.A. (2000). Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. Nucleic acids research *28*, 3811-3816.

Jeppsson, K., Kanno, T., Shirahige, K., and Sjogren, C. (2014). The maintenance of chromosome structure: positioning and functioning of SMC complexes. Nature reviews Molecular cell biology *15*, 601-614.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature genetics *41*, 941-945.

Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cristea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. Molecular systems biology *9*, 672.

Kalocsay, M., Hiller, N.J., and Jentsch, S. (2009). Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. Molecular cell *33*, 335-343.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research *40*, D109-114.

Kapoor, A., Goldberg, M.S., Cumberland, L.K., Ratnakumar, K., Segura, M.F., Emanuel, P.O., Menendez, S., Vardabasso, C., Leroy, G., Vidal, C.I.*, et al.* (2010). The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. Nature *468*, 1105-1109.

Keilhauer, E.C., Hein, M.Y., and Mann, M. (2015). Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). Molecular & cellular proteomics : MCP *14*, 120-135.

Kim, D., Blus, B.J., Chandra, V., Huang, P., Rastinejad, F., and Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. Nature structural & molecular biology *17*, 1027-1029.

Kim, H.S., Vanoosthuyse, V., Fillingham, J., Roguev, A., Watt, S., Kislinger, T., Treyer, A., Carpenter, L.R., Bennett, C.S., Emili, A.*, et al.* (2009). An acetylated form of histone H2A.Z regulates chromosome architecture in Schizosaccharomyces pombe. Nature structural & molecular biology *16*, 1286-1293.

Kim, K., Punj, V., Choi, J., Heo, K., Kim, J.M., Laird, P.W., and An, W. (2013). Gene dysregulation by histone variant H2A.Z in bladder cancer. Epigenetics & chromatin *6*, 34.

Kloet, S.L., Baymaz, H.I., Makowski, M., Groenewold, V., Jansen, P.W., Berendsen, M., Niazi, H., Kops, G.J., and Vermeulen, M. (2014). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. The FEBS journal.

Korber, P., Zander, T., Herschlag, D., and Bardwell, J.C.A. (1999). A new heat shock protein that binds nucleic acids. Journal of Biological Chemistry *274*, 249-256.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kossel, A. (1911). Ueber die chemische Beschaffenheit des Zellkerns. Munchen Med Wochenschrift.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Laguri, C., Duband-Goulet, I., Friedrich, N., Axt, M., Belin, P., Callebaut, I., Gilquin, B., Zinn-Justin, S., and Couprie, J. (2008). Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. Biochemistry *47*, 6199-6207.

Latreille, D., Bluy, L., Benkirane, M., and Kiernan, R.E. (2014). Identification of histone 3 variant 2 interacting factors. Nucleic acids research *42*, 3542-3550.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G.M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell *153*, 590-600.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell *151*, 1608-1616.

Liu, X., Li, B., and GorovskyMa (1996). Essential and nonessential histone H2A variants in Tetrahymena thermophila. Molecular and cellular biology *16*, 4305-4311.

Luber, C.A., Cox, J., Lauterbach, H., Fancke, B., Selbach, M., Tschopp, J., Akira, S., Wiegand, M., Hochrein, H., O'Keeffe, M.*, et al.* (2010). Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity *32*, 279-289.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nature reviews Molecular cell biology *13*, 436-447.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luk, E., Ranjan, A., Fitzgerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome. Cell *143*, 725-736.

Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grofte, M., Chan, K.L., Hickson, I.D., Bartek, J.*, et al.* (2011a). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. Nat Cell Biol *13*, 243-253.

Lukas, J., Lukas, C., and Bartek, J. (2011b). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. Nat Cell Biol *13*, 1161-1169.

Lydeard, J.R., Schulman, B.A., and Harper, J.W. (2013). Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO reports *14*, 1050-1061.

Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nature structural biology *10*, 882-891.

Marques, M., Laflamme, L., Gervais, A.L., and Gaudreau, L. (2010). Reconciling the positive and negative roles of histone H2A.Z in gene transcription. Epigenetics : official journal of the DNA Methylation Society *5*, 267-272.

Marzluff, W.F. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. Current opinion in cell biology *17*, 274-280.

Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. Genomics *80*, 487-498.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nature reviews Genetics *9*, 843-854.

Matsuda, R., Hori, T., Kitamura, H., Takeuchi, K., Fukagawa, T., and Harata, M. (2010). Identification and characterization of the two isoforms of the vertebrate H2A.Z histone variant. Nucleic acids research *38*, 4263-4273.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends in biochemical sciences *28*, 69-74.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C.*, et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

Maze, I., Noh, K.M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nature reviews Genetics *15*, 259-271.

McBryant, S.J., Adams, V.H., and Hansen, J.C. (2006). Chromatin architectural proteins. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *14*, 39-51.

McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J.*, et al.* (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proceedings of the National Academy of Sciences of the United States of America *96*, 13983-13988.

Mellor, J. (2006). It takes a PHD to read the histone code. Cell *126*, 22-24.

Millar, C.B., Xu, F., Zhang, K., and Grunstein, M. (2006). Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. Genes & development *20*, 711-722.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343-348.

Morillo-Huesca, M., Clemente-Ruiz, M., Andujar, E., and Prado, F. (2010). The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. PloS one *5*, e12143.

Morrison, A.J., and Shen, X.T. (2009). Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. Nat Rev Mol Cell Bio *10*, 373-384.

Musselman, C.A., Gibson, M.D., Hartwick, E.W., North, J.A., Gatchalian, J., Poirier, M.G., and Kutateladze, T.G. (2013). Binding of PHF1 Tudor to H3K36me3 enhances nucleosome accessibility. Nature communications *4*, 2969.

Nishibuchi, I., Suzuki, H., Kinomura, A., Sun, J., Liu, N.A., Horikoshi, Y., Shima, H., Kusakabe, M., Harata, M., Fukagawa, T.*, et al.* (2014). Reorganization of damaged chromatin by the exchange of histone variant H2A.Z-2. International journal of radiation oncology, biology, physics *89*, 736-744.

Nock, A., Ascano, J.M., Barrero, M.J., and Malik, S. (2012). Mediator-regulated transcription through the +1 nucleosome. Molecular cell *48*, 837-848.

Obri, A., Ouararhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P.T., Mak, T.W.*, et al.* (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. Nature *505*, 648-653.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Oudet, P., Grossbellard, M., and Chambon, P. (1975). Electron-Microscopic and Biochemical Evidence That Chromatin Structure Is a Repeating Unit. Cell *4*, 281-300.

Papamichos-Chronakis, M., Watanabe, S., Rando, O.J., and Peterson, C.L. (2011). Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell *144*, 200-213.

Pehrson, J.R., and Fried, V.A. (1992). MacroH2A, a core histone containing a large nonhistone region. Science *257*, 1398-1400.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R.*, et al.* (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660-665.

Pinto, D.M., and Flaus, A. (2010). Structure and function of histone H2AX. Sub-cellular biochemistry *50*, 55-78.

Polo, S.E., Kaidi, A., Baskcomb, L., Galanty, Y., and Jackson, S.P. (2010). Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. Embo Journal *29*, 3130-3139.

Posavec, M., Timinszky, G., and Buschbeck, M. (2013). Macro domains as metabolite sensors on chromatin. Cellular and molecular life sciences : CMLS *70*, 1509-1524.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nature reviews Molecular cell biology *10*, 192-206.

Qin, S., and Min, J. (2014). Structure and function of the nucleosome-binding PWWP domain. Trends in biochemical sciences.

Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nature structural biology *9*, 217-224.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell *123*, 233-248.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D.J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. The EMBO journal *22*, 1599-1607.

Rangasamy, D., Greaves, I., and Tremethick, D.J. (2004). RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. Nature structural & molecular biology *11*, 650-655.

Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. Cell *154*, 1232-1245.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols *2*, 1896-1906.

Rasmussen, T.P., Huang, T., Mastrangelo, M.A., Loring, J., Panning, B., and Jaenisch, R. (1999). Messenger RNAs encoding mouse histone macroH2A1 isoforms are expressed at similar levels in male and female cells and result from alternative splicing. Nucleic acids research *27*, 3685-3689.

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D.Y.*, et al.* (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. Genes & development *26*, 433-438.

Reinhardt, H.C., and Yaffe, M.B. (2013). Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response. Nature reviews Molecular cell biology *14*, 563-580.

Ren, Q., and Gorovsky, M.A. (2001). Histone H2A.Z acetylation modulates an essential charge patch. Molecular cell *7*, 1329-1335.

Rhee, H.S., Bataille, A.R., Zhang, L., and Pugh, B.F. (2014). Subnucleosomal Structures and Nucleosome Asymmetry across a Genome. Cell *159*, 1377-1388.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Ritchie, K., Seah, C., Moulin, J., Isaac, C., Dick, F., and Berube, N.G. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. The Journal of cell biology *180*, 315-324.

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. Bba-Gene Regul Mech *1839*, 627-643.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols *5*, 725-738.

Ruhl, D.D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M.P., Conaway, R.C., Conaway, J.W., and Chrivia, J.C. (2006). Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. Biochemistry *45*, 5671-5677.

Sanders, M.M. (1978). Fractionation of Nucleosomes by Salt Elution from Micrococcal Nuclease-Digested Nuclei. Journal of Cell Biology *79*, 97-109.

Sansoni, V., Casas-Delucchi, C.S., Rajan, M., Schmidt, A., Bonisch, C., Thomae, A.W., Staege, M.S., Hake, S.B., Cardoso, M.C., and Imhof, A. (2014). The histone variant H2A.Bbd is enriched at sites of DNA synthesis. Nucleic acids research *42*, 6405-6420.

Santoro, S.W., and Dulac, C. (2012). The activity-dependent histone variant H2BE modulates the life span of olfactory neurons. eLife *1*, e00070.

Sarcinella, E., Zuzarte, P.C., Lau, P.N., Draker, R., and Cheung, P. (2007). Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Molecular and cellular biology *27*, 6457-6468.

Schauer, T., Schwalie, P.C., Handley, A., Margulies, C.E., Flicek, P., and Ladurner, A.G. (2013). CAST-ChIP maps cell-type-specific chromatin states in the Drosophila central nervous system. Cell reports *5*, 271-282.

Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W.H., Wolf, P., Alvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V., *et al.* (2013). Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. Nucleic acids research *41*, 4860-4876.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887-898.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Sevilla, A., and Binda, O. (2014). Post-translational modifications of the histone variant H2AZ. Stem cell research *12*, 289-295.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458-472.

Sharma, U., Stefanova, D., and Holmes, S.G. (2013). Histone variant H2A.Z functions in sister chromatid cohesion in Saccharomyces cerevisiae. Molecular and cellular biology *33*, 3473-3481.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shikauchi, Y., Saiura, A., Kubo, T., Niwa, Y., Yamamoto, J., Murase, Y., and Yoshikawa, H. (2009). SALL3 interacts with DNMT3A and shows the ability to inhibit CpG island methylation in hepatocellular carcinoma. Molecular and cellular biology *29*, 1944-1958.

Shun, M.C., Botbol, Y., Li, X., Di Nunzio, F., Daigle, J.E., Yan, N., Lieberman, J., Lavigne, M., and Engelman, A. (2008). Identification and characterization of PWWP domain residues critical for LEDGF/p75 chromatin binding and human immunodeficiency virus type 1 infectivity. Journal of virology *82*, 11555-11567.

Sims, J.K., and Wade, P.A. (2011). Mi-2/NuRD complex function is required for normal S phase progression and assembly of pericentric heterochromatin. Molecular biology of the cell *22*, 3094-3102.

Skene, P.J., and Henikoff, S. (2013). Histone variants in pluripotency and disease. Development *140*, 2513-2524.

Smeenk, G., Wiegant, W.W., Vrolijk, H., Solari, A.P., Pastink, A., and van Attikum, H. (2010). The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. The Journal of cell biology *190*, 741-749.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews Genetics *14*, 204-220.

Soboleva, T.A., Nekrasov, M., Pahwa, A., Williams, R., Huttley, G.A., and Tremethick, D.J. (2012). A unique H2A histone variant occupies the transcriptional start site of active genes. Nature structural & molecular biology *19*, 25-30.

Stec, I., Nagl, S.B., van Ommen, G.J., and den Dunnen, J.T. (2000). The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS letters *473*, 1-5.

Stec, I., Wright, T.J., van Ommen, G.J., de Boer, P.A., van Haeringen, A., Moorman, A.F., Altherr, M.R., and den Dunnen, J.T. (1998). WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphy gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. Human molecular genetics *7*, 1071-1082.

Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nature structural biology *7*, 1121-1124.

Svotelis, A., Gevry, N., Grondin, G., and Gaudreau, L. (2010). H2A.Z overexpression promotes cellular proliferation of breast cancer cells. Cell Cycle *9*, 364-370.

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

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.Y., *et al.* (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232-235.

Tada, K., Susumu, H., Sakuno, T., and Watanabe, Y. (2011). Condensin association with histone H2A shapes mitotic chromosomes. Nature *474*, 477-483.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nature reviews Molecular cell biology *11*, 264-275.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., *et al.* (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016-1028.

Tapia-Alveal, C., Lin, S.J., Yeoh, A., Jabado, O.J., and O'Connell, M.J. (2014). H2A.Z-Dependent Regulation of Cohesin Dynamics on Chromosome Arms. Molecular and cellular biology *34*, 2092-2104.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature structural & molecular biology *14*, 1025-1040.

Thakar, A., Gupta, P., Ishibashi, T., Finn, R., Silva-Moreno, B., Uchiyama, S., Fukui, K., Tomschik, M., Ausio, J., and Zlatanova, J. (2009). H2A.Z and H3.3 histone variants affect nucleosome structure: biochemical and biophysical studies. Biochemistry *48*, 10852-10857.

Thambirajah, A.A., Dryhurst, D., Ishibashi, T., Li, A., Maffey, A.H., and Ausio, J. (2006). H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation. The Journal of biological chemistry *281*, 20036-20044.

Thambirajah, A.A., Li, A., Ishibashi, T., and Ausio, J. (2009). New developments in post-translational modifications and functions of histone H2A variants. Biochemistry and cell biology = Biochimie et biologie cellulaire *87*, 7-17.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of Histone-H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin. Journal of Cell Biology *83*, 403-427.

Tolstorukov, M.Y., Goldman, J.A., Gilbert, C., Ogryzko, V., Kingston, R.E., and Park, P.J. (2012). Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. Molecular cell *47*, 596-607.

Tolstorukov, M.Y., Kharchenko, P.V., Goldman, J.A., Kingston, R.E., and Park, P.J. (2009). Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes. Genome research *19*, 967-977.

Tremethick, D.J. (2007). Higher-order structures of chromatin: The elusive 30 nm fiber. Cell *128*, 651-654.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: Is there a distinctive molecular signature? Molecular cell *28*, 1-13.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America *98*, 5116-5121.

Valdes-Mora, F., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., Nair, S.S., Patterson, K.I., Tremethick, D.J., Stirzaker, C., and Clark, S.J. (2012). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. Genome research *22*, 307-321.

van Daal, A., and Elgin, S.C. (1992). A histone variant, H2AvD, is essential in Drosophila melanogaster. Molecular biology of the cell *3*, 593-602.

van Nuland, R., Smits, A.H., Pallaki, P., Jansen, P.W., Vermeulen, M., and Timmers, H.T. (2013a). Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. Molecular and cellular biology *33*, 2067-2077.

van Nuland, R., van Schaik, F.M., Simonis, M., van Heesch, S., Cuppen, E., Boelens, R., Timmers, H.M., and van Ingen, H. (2013b). Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics & chromatin *6*, 12.

Vardabasso, C., Gaspar-Maia, A., Hasson, D., Pünzeler, S., Valle-Garcia, D., Straub, T., Keilhauer, E.C., Strub, T., Panda, T., Chung, C.Y.*, et al.* (2015). Histone variant H2A.Z.2 mediates proliferation and drug sensitivity of malignant melanoma. Mol Cell under revision.

Vardabasso, C., Hasson, D., Ratnakumar, K., Chung, C.Y., Duarte, L.F., and Bernstein, E. (2014). Histone variants: emerging players in cancer biology. Cellular and molecular life sciences : CMLS *71*, 379-404.

Verni, F., and Cenci, G. (2015). The Drosophila histone variant H2A.V works in concert with HP1 to promote kinetochore-driven microtubule formation. Cell Cycle, 0.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell *87*, 95-104.

Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Gottgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. Nature structural & molecular biology *17*, 617-619.

Viens, A., Mechold, U., Brouillard, F., Gilbert, C., Leclerc, P., and Ogryzko, V. (2006). Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms. Molecular and cellular biology *26*, 5325-5335.

Waddington, C.H. (1942). The Epigenotype. Endeavour, 18-20.

Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat Rev Mol Cell Bio *13*, 115-126.

Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J.R., 3rd, and Jia, S. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Molecular cell *33*, 428-437.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science *340*, 195-199.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Weber, C.M., Henikoff, J.G., and Henikoff, S. (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nature structural & molecular biology *17*, 1500-1507.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular cell *53*, 819-830.

Wen, H., Li, J., Song, T., Lu, M., Kan, P.Y., Lee, M.G., Sha, B., and Shi, X. (2010). Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. The Journal of biological chemistry *285*, 9322-9326.

West, M.H., and Bonner, W.M. (1980). Histone 2A, a heteromorphous family of eight protein species. Biochemistry *19*, 3238-3245.

Wiedemann, S.M., Mildner, S.N., Bonisch, C., Israel, L., Maiser, A., Matheisl, S., Straub, T., Merkl, R., Leonhardt, H., Kremmer, E.*, et al.* (2010). Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. The Journal of cell biology *190*, 777-791.

Wilkins, B.J., Rall, N.A., Ostwal, Y., Kruitwagen, T., Hiragami-Hamada, K., Winkler, M., Barral, Y., Fischle, W., and Neumann, H. (2014). A cascade of histone modifications induces chromatin condensation in mitosis. Science *343*, 77-80.

Woodcock, C.L., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. Experimental cell research *97*, 101-110.

Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M.F., Xu, C., Dombrovski, L., Qiu, W., Wang, Y., and Min, J. (2011). Structural and histone binding ability characterizations of human PWWP domains. PloS one *6*, e18919.

Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G., Wei, D., and Wu, C. (2005). Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. Nature structural & molecular biology *12*, 1064-1071.

Xu, Y., Ayrapetov, M.K., Xu, C., Gursoy-Yuzugullu, O., Hu, Y., and Price, B.D. (2012). Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair. Molecular cell *48*, 723-733.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Molecular cell *2*, 851-861.

Yamada, M., Masai, H., and Bartek, J. (2014). Regulation and roles of Cdc7 kinase under replication stress. Cell Cycle *13*, 1859-1866.

Yang, J., and Everett, A.D. (2007). Hepatoma-derived growth factor binds DNA through the N-terminal PWWP domain. BMC molecular biology *8*, 101.

Yen, K., Vinayachandran, V., and Pugh, B.F. (2013). SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. Cell *154*, 1246-1256.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.

Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. Structure *16*, 166-179.

Zovkic, I.B., Paulukaitis, B.S., Day, J.J., Etikala, D.M., and Sweatt, J.D. (2014). Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. Nature *515*, 582-586.

APPENDIX

Table A.1 SK-mel147 H2A.Z interactome

SK-mel147 Exp. 1		SK-mel147 Exp. 2	
GFP-H2A.Z.1	**GFP-H2A.Z.2**	**GFP-H2A.Z.1**	**GFP-H2A.Z.2**
FDR = 0.2	FDR = 0.15	FDR = 0.08	FDR = 0.1
$S_0 = 0.8$	$S_0 = 1.0$	$S_0 = 1.0$	$S_0 = 1.0$
ACTR6	ACTR6	ACTR6	ACTR6
ASH1L	ASH1L	ASH2L	ASH2L
BAHD1	BCORL1	BAHD1	BCORL1
BCORL1	BRWD3	BRD2	BRWD3
BRWD3	CBX7	BRWD3	CUL4A
CCDC101	CCDC101	CDYL	CUL4B
CDYL	CDYL	CHD1	DDB1
CUL4A	CDYL2	CUL4A	DEK
CUL4B	CLASRP	DIDO1	DIDO1
DIDO1	CUL4A	DMAP1	DMAP1
BRD2	CUL4B	EP400	DSG1
DMAP1	BRD2	ERH	EP400
EP400	DMAP1	H2A.Z	H2A.Z
FMN1	EP400	HCFC2	HCFC2
H2A.Z	ETV6	HDAC2	HDAC2
HCFC2	H2A.Z	HIST1H1D	HIST1H2AJ
HMG20A	HCFC2	HIST1H2AJ	HMG20A
ING5	HDAC2	HMG20A	IMP4
KDM2A	HMG20A	IMP4	KDM2A
MAGEA10	HNRPLL	KDM2A	KPRP
MEN1	INIP	KDM2B	MEAF6
MIER1	INTS5	KPRP	MEN1
MLL	KANSL1	MAGEA10	MLL
MORF4L2	KDM2A	MEAF6	MTA1
MTA1	KDM6B	MEN1	MTA1-3
MTA1-3	LONP1	MIER1	MYPOP
MYPOP	LRIF1	MLL	NEDD8
ORC1	MAGEA10	MORF4L1	NES
PHF14	MBD3	MTA1	NOLC1
PHF14-2	MCRS1	MTA1-3	PCGF1
PHF2	MEN1	NEDD8	PES1
PWWP2A	MIER1	NES	PHF14
RAI1	MTA1	NOLC1	PHF14-2
RUVBL1	MTA1-3	PCGF1	PHF2
RUVBL2	MYPOP	PHF14	PHIP
SKP1	NABP2;OBFC2B	PHF14-2	PWWP2A
SPIN1	NEDD8	PHF2	RAI1
SPIN2B;SPIN2A	NR2F2	PHF20L1	RBBP4
SPIN4	NUP205	PHIP	RBBP5

SK-mel147 Exp. 1		SK-mel147 Exp. 2	
GFP-H2A.Z.1	**GFP-H2A.Z.2**	**GFP-H2A.Z.1**	**GFP-H2A.Z.2**
FDR = 0.2	FDR = 0.15	FDR = 0.08	FDR = 0.1
$S_0 = 0.8$	$S_0 = 1.0$	$S_0 = 1.0$	$S_0 = 1.0$
RCAP	ORC1	PHRF1	RBBP7
SUV420H1	PHF14	PWWP2A	RBX1
TAF3	PHF14-2	RAI1	RUVBL1
TCF20	PHF2	RBBP7	RUVBL2
TRRAP	PHF20L1	RBX1	SRCAP
VPS72	PWWP2A	RUVBL1	TCF20
WBP7	RAD18	RUVBL2	TRRAP
YEATS4	RAI1	SRCAP	VIM
ZFX;ZFY	RBBP4	TCF20	VPS72
ZMYM4	RBBP5	TRRAP	WHSC1L1
ZNF174	RBBP7	TTN	ZCCHC17
ZNF444	RUVBL1	VIM	ZNF512
ZNF512B	RUVBL2	VPS72	ZNHIT1
ZNF687	SPIN1	WBP7	
ZNF711	SPIN2B;SPIN2A	WHSC1L1	
ZNHIT1	SPIN4	ZCCHC17	
	SRCAP	ZFX;ZFY	
	TCF20	ZNF512B	
	TRRAP	ZNF579	
	UHRF2	ZNHIT1	
	VPS72		
	WBP7		
	YEATS4		
	ZFX;ZFY		
	ZNF444		
	ZNF512B		
	ZNF557		
	ZNF687		
	ZNF711		
	ZNHIT1		

Table A.2 HeLa Kyoto H2A.Z interactome

HeLa Kyoto Exp. 5		HeLa Kyoto Exp. 5	
GFP-H2A.Z.1 FDR = 0.05 $S_0 = 1.0$	**GFP-H2A.Z.2** FDR = 0.05 $S_0 = 1.0$	**GFP-H2A.Z.1** FDR = 0.05 $S_0 = 1.0$	**GFP-H2A.Z.2** FDR = 0.05 $S_0 = 1.0$
ACTL6A	ACTL6A	SRCAP	KMT2A
ACTR6	ACTR2	SSR4	LIMA1
AIF1L	ACTR3	TBL3	LIMA1
BAHD1	ACTR6	TIMM13	LMO7
C3orf17	ADD1	TNKS1BP1	MCM10
C4BPA	ADD3	TPR	MEN1
CD55	AKAP2	TRRAP	MIER1
CFL2	ARPC1A	UBA52	MTA1
CXXC1	ARPC1B	UBR5	NPAT
CYBRD1	ARPC2	VPRBP	ORC3
DMAP1	ARPC3	VPS72	PEAK1
DSTN	ARPC4	WDR1	PHF6
EEF1D	ARPC5	WDR3	PHIP
EP400	ARPC5L	YEATS4	PPP1R9A
EPB41	ATAD2		PPP1R9B
H2A.Z	BAHD1		PRKAR2A
HCFC2	BCORL1		PTBP3
HMGA1	C1orf21		PWWP2A
IFI16	CD55		RBBP4
ING5	CDCA8		RBM7
ITGB1	CDYL		RBX1
LRCH1	CGN		RUVBL1
MBD2	CTTN		RUVBL2
MEN1	CUL4B		SMARCA2
MIER1	CYBRD1		SMARCD1
NPAT	DDB1		SMARCD2
PA2G4	DMAP1		SRCAP
PDLIM7	EP400		SSFA2
PLS3	H1F0		SUV420H1
POLR1C	H2A.Z		SVIL
PTBP3	HCFC2		TFAP2D
PWWP2A	HDAC1		TNKS1BP1
RUVBL1	HDAC2		TRRAP
RUVBL2	HIST1H1C		UBA52
SCIN	HIST1H4A		VPS72
SDPR	HSPA1A		YEATS4
SETD8	HSPA5		
SHROOM3	HSPA8		
SIPA1	INF2		
SMARCD1	ING5		
SMARCD2	KDM2A		

IV
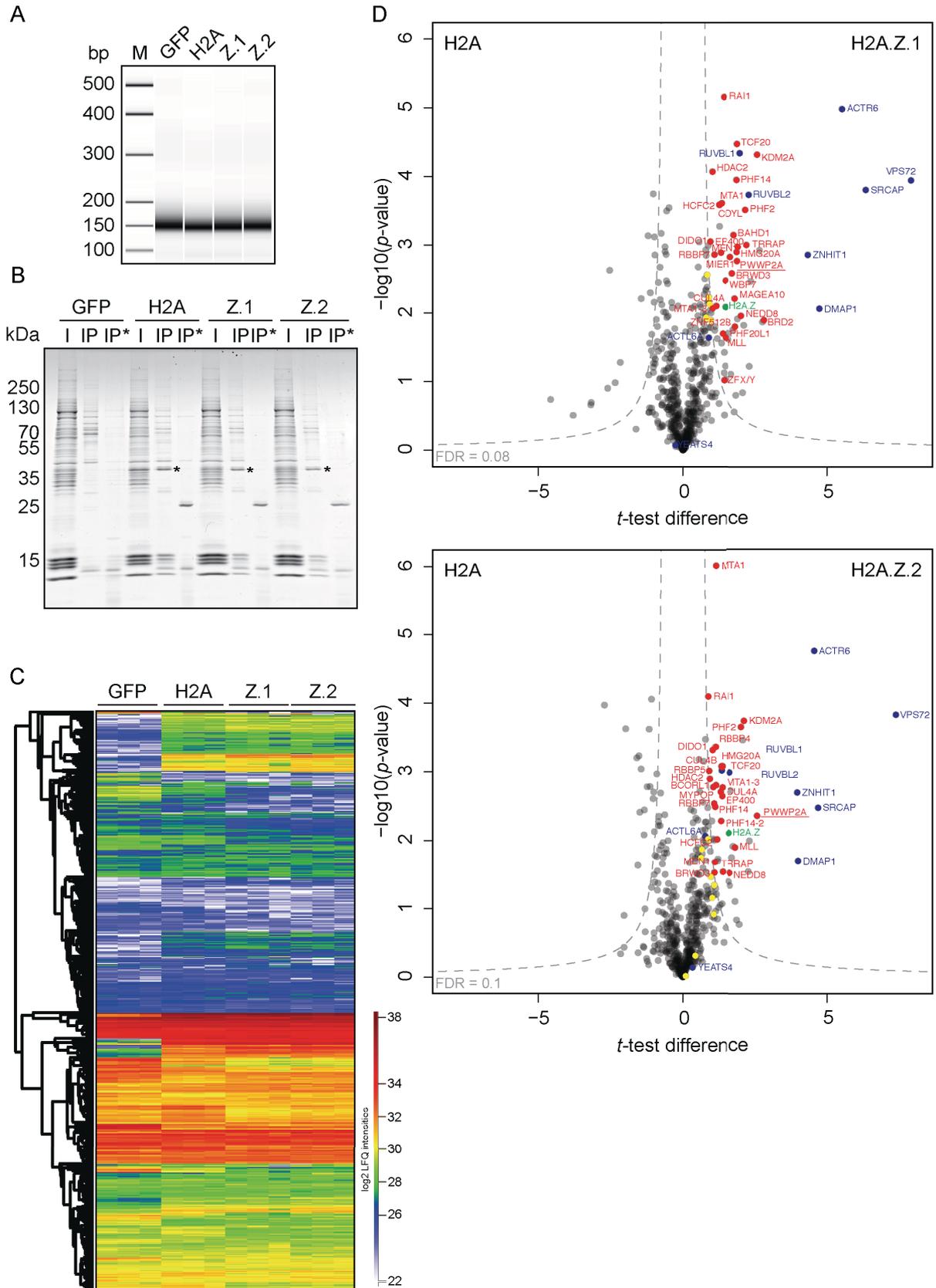
Figure A.1 Second SK-mel147 MNase ChIP experiment followed by qMS.
The second MNase ChIP experiment with chromatin from stable SK-mel147 cell lines was done as previously described in **Figure 3.1.3** and **Figure 3.1.4**. **(D)** FDR statistics and threshold line in volcano plots were adjusted: H2A.Z.1: FDR = 0.08, S_0 = 1; H2A.Z.2: FDR = 0.1, S_0 = 1.

VI

Figure A.2 PWWP2A is conserved in vertebrates. Searching the full PWWP2A amino acid sequence with the web browser-based NCBI Basic Local Alignment Search Tool (BLAST) revealed the conservation of PWWP2A among vertebrates. BLAST search was executed employing default settings and 500 aligned sequences were displayed. Hits down to a query covery of 70% (100 hits) were selected to calculate a distance tree with BLAST by pairwise alignment applying default settings.

The BLAST tree showing all hits was exported, depicting the different organisms with taxonomic names (Homo_sapiens, e.g.). The tree was converted into a circular phylogram with the CLC Main Workbench (7.0) software. Hits were colored according to their affiliation into the five different classes: fish (blue), birds (red), reptiles (green), mammals (magenta) and amphibians (orange).
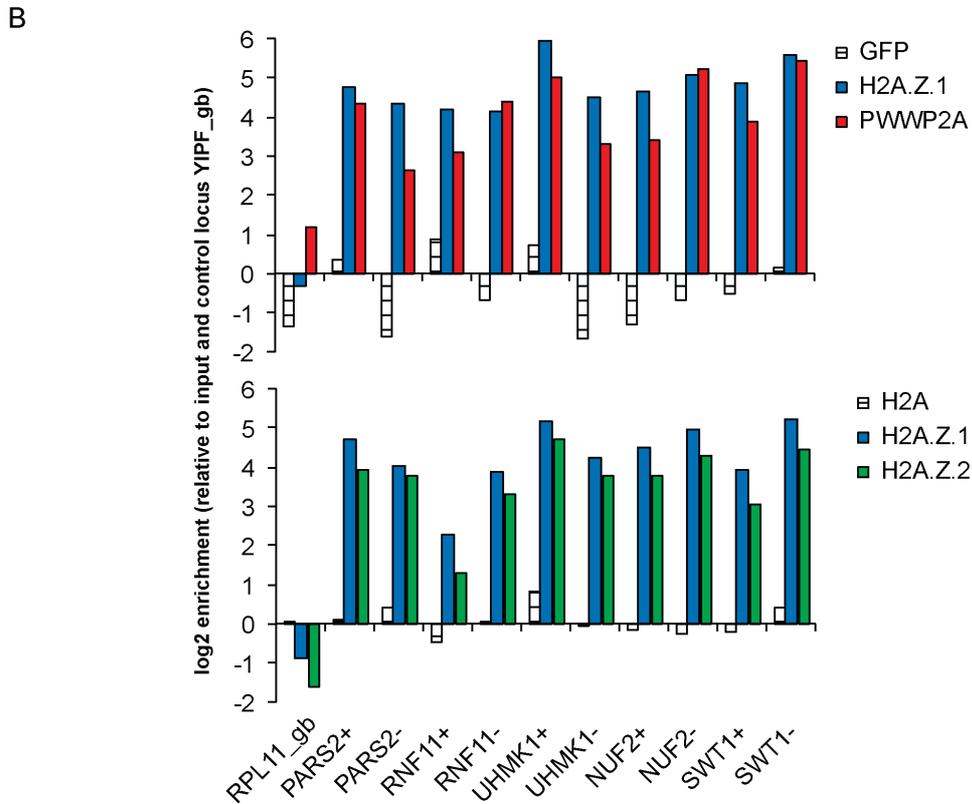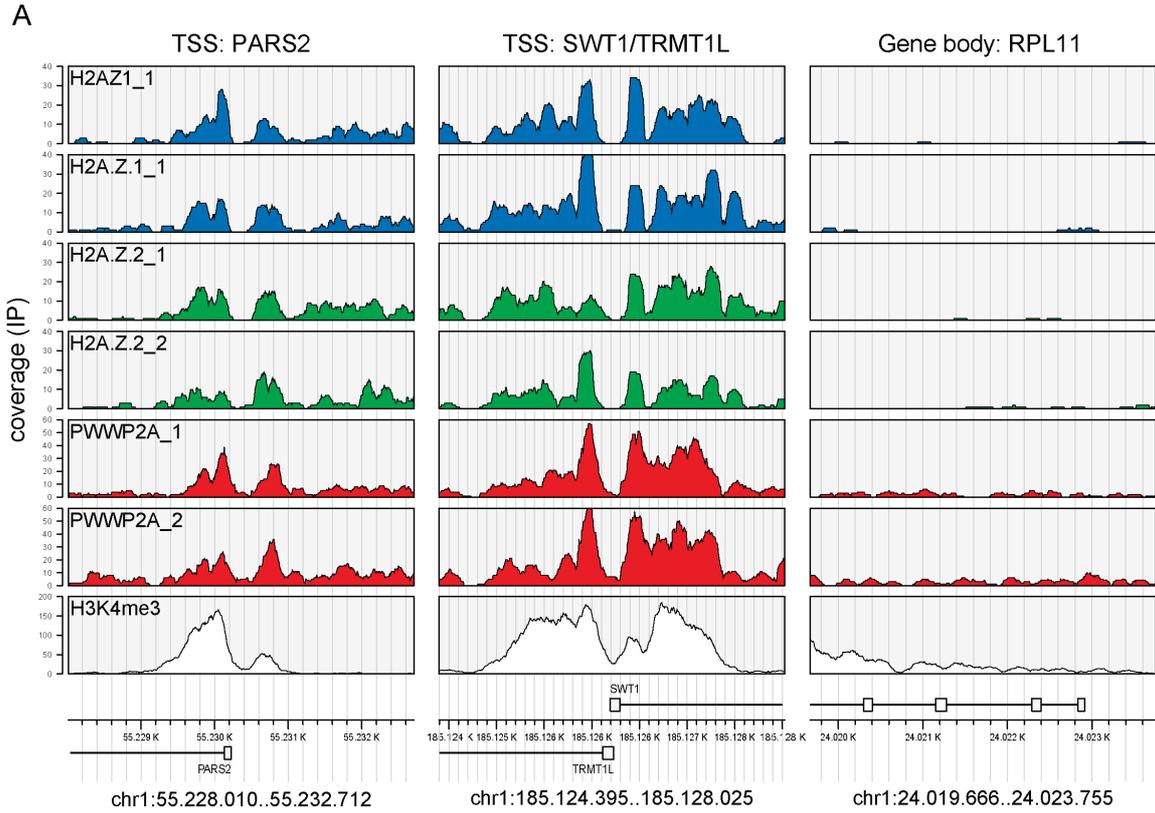
Figure A.3 Validation of MNase ChIP-sequencing results for TSS enrichments of GFP-PWWP2A, GFP-H2A.Z.1 and GFP-H2AZ.2. (A) Snapshots of the coverage of GFP-PWWP2A, GFP-H2A.Z.1, GFP-H2A.Z.2 and H3K4me3 around the TSS of two exemplary genes (PARS2, SWT1/TRMT1L) and the gene body of RPL11. Candidate genes were all encoded on human

chromosome 1; for exact genomic coordinates, please see Table 2.1.2. **(B)** Enrichment (log2) of GFP-PWWP2A, GFP-H2A.Z.1 and GFP-H2A.Z.2 relative to ChIP input DNA and a control locus (gene body of YIPF = YIPF_gb) at -1 and +1 nucleosomes of five target genes (PARS2, RNF11, UHMK1, NUF2 and SWT1/TRMT1L). Immunoprecipitated MNase-digested DNA fragments and input DNA fragments were amplified with candidate specific primer pairs spanning either the +1 or -1 nucleosome (PARS2+ and PARS2-, for example) by qPCR, followed by absolute quantification with the LC480 software utilizing one input sample with known concentration as standard. The fold change (log2) enrichments at target loci relative to input DNA and the YIPF_gb control locus were calculated in Excel and plotted as bar plots. MNase ChIP with chromatin from HeLa Kyoto cells expressing GFP-only or GFP-H2A served as negative controls with respect to ChIP input. The gene body of RPL11 (RPL11_gb) was chosen as negative control with respect to a region that showed no enrichment of GFP-PWWP2A, -H2A.Z.1 or H2A.Z.2 in MNase ChIP-sequencing.

```
                                                              P1
hs  AAVAAEAAAATAASPGEGGAGEAEPEMEPIPGSEAGTDPLPVTA--TEASVPDGET-DGQQSAPQADEPPLP-PPPP--PPGELARSPEAVGPELEAEEKLSVRVAESAAAAPQGGPELPP
bt  AAVAAEAAAATAASPGEGGAGEAEPEMEPIPGSEAGTDPLPVTA--TEASVPDGEA-DGQQSSPQADEPPLP-LPPP--PPGELARSPEAAGLELEPEEKLPARVVEPGAAAPPEGPSLPP
mm  AAVAAEAAAATAASPGEGGAGEAEPELEPIPGSEAGT-PLPVTA--TEAAVPDGEA-DGRQSAPQADEQPLP-PPPPPPPPGELADSSEAEE--------------------AKPPEPAA
gg  AAMAAEAAAATAAVPGDGGAGEAEPEMEPIPGSEAGADPPPAVTEAVESAVPDGEEADGGKGAPGEAEE---------PPPVQLARSPAGTR-------------------EPEGG-----
xt  -----------------------------------------------------------EQLPREPFLQRQPIPMILSEVTDSGIYTGLCLEEA-KF----------SSNNGSEASL
dr  --------------------------------------------------------------------------------------------------------------------

hs  SPASPPEQPPAPEEREEPPLPQPVAPALVPPAGGDSTVSQLIPGSEVRVTLDHIIEDALVVSFRFGEKLFSGVLMDLSKRFGPHGIPVTVFPKREYKDKPEA-MPLQSNTFQEGTEVKCE
bt  SPVPPPEEPPAPEERQEPPLPQPAAPALVPPAGGDSAVSQLIPGSEVRVTLDHIIEDALVVSFRLGEKLFSGVLMDLSKRFGPHGIPVTVFPKREYKDKPEA-MQLQSNTFQEGTEVKCE
mm  VPVSPPEQPPAAPEQPEDAPRPPPAPALVPPAGGDSAVSHLIPGSEVRVTLDHIIEDALVVSFRLGEKLFSGVLMDLSKRFGPHGIPVTVFPKREYKDKPDA-MQLQSTTFQEGIEVKQE
gg  ---------------------P-DRGEAEAPLLPGSEVRVTLDHIIEDALVVSFCHGNRIFSGVLMDLSKRFGPHGIPVTIFPKREYKDKPEA-MQLQSKPFQDEAQVKCE
xt  N----------YSECDNR------SVGTTEGSVVTSGIDSLIPGSEVQVTLDHIIQDALVVSFCHGNRIFSGVLMDLSKRFGPHGIPVTVVHPRREYKNKPVESIQEESKSFHEEPLVKSE
dr  ----------------------------------RLVLGSEVRVTLDHIIDDALVVSFRLGEKIFSGVLMDLSRRFGPYGIPITMFPKREYKDKPES-MQLKMEAFQSETDKGDA
                                           *: ****:*******:******* *:::*********:****:***:*:.*:****:**   :  :  *:.

    P2                                      I          conI
hs  ANGAVPDDPSPVPHPELSLAESLWTSKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRKIKRPKRKMYREEPTSIMNAIKLRPRQVLCDKCKNSVVAEKKEIRKGSSATDSSKYEDK-KR
bt  VNGAVPNDPSPVPPPELSLAESLWTSKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRKIKRPKRKMYREEPTSIMNAIKLRPRQVLCDKCKNSVVAEKKEIRKGGSASDPSKYEDK-KR
mm  VNGAVPDDLSPVPPP-----ERLWASKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRKIKRPKRKMYREEPTSIMNAIKLRPRQVLCDKCKNSVVAEKKEIRKGSS--DSSRYEDK-KR
gg  SNAAVPDDSSLTQPSEPSLAKSLWTSKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRKIKRPKRKMYREEPTSIMNAIKLRPRQVLCDKCKNSVVAEKKEIKKGGNASDSSKYEDNKKR
xt  E--NSPEDVTPIQQLENCKVQNLWTTKPPPLFHEGAPYPPPLFIRDTYNQSIPQPPPRKIKRPKKKIYREEPTSIMNAIKLRPRQVLCDKCKNNVVADKKEIRKVA--TDSYKTEEGKRR
dr  Q----GSSNAPTPQPCPNPQPNPWTSKPPPLFQEGAPYPPPLFIRDTYNQSLPQPLPRKIKRPKRRLYREEPTSIMNAIKLRPRQVLCDKCKGAVVQGDKRETRRGPISDC-RSDDAKRR
       .. :         *::******:*****************:*** ********:::***********************  ** .*. :.  *  : :: :*

hs  RNESVTTTV-NKKLKTDHKVD--GKNQNESQKRNAVV-----KVSNIAHSRGRVVKVSAQANTSKAQLSTKKVLQSKNMDHAKAREVLKIAKEKAQKKQNETSTSKNAHSKVHFTRRYQNP
bt  RSESVTTTV-NKKLKTDHKVD--GRSQNESQKRSAVV-----KVSNIPHSRGRVVKVSAQANTSKAQLNTKKVLQNKNMDHAKAREVLKIAKEKAQKQSETSTSRSAHSKVHFTRRYQNP
mm  RNDSVATV--NKKLKTDHKVD--GKNQNESQRRNTVV-----RVSSIAHSRGRVVKVSAQANTSKAQLNTKKVLQSKNMDHAKAREVLKIAKEKAQKQSETSTSKTAHAKVHFTRRYQSP
gg  RNESVTTV--NKKLKTDHKVD--GKNQNESQKRNAVV-----KVSNIAHSRGRVVKVTAQANTSKAQLNTKKVLQSKNMDHVKAREVLKMAKEKAQKQSATSSSKNAHSKVHFTRRLQNT
xt  RHETVTTTV-NKKLKTDHKVN--GKSQNESQKRTPIT-----KVTNLAHGRGKVVKVPSQTSAAKTQMHTKKVLQNKNMDHVKAREVLKMAKEKAQKQKVTTSSKNAHSKVHFTRRLQNS
dr  RNDNTTAAASKRPRSDPRSEEKGRGSDGPKRQGSAVRVSSTGSTLAHSQGV--STTTSSANSRVQLNAKKVLQSKNVDHSKAREVLKMAKAQRRQKETDTGSSGNAHQKVHFTRRLQQI
     *  :..::.  .*:  ::*  : :   *:  .:   :::    .      ::.:  *.:.   .. :.:  ::.*:  :*****.**:**  *******:**  : ::*:.  *  :*  .** *******  *.

hs  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKMRSGKMAPKP---QSRCTSTRSAGEAPSE--------NQSPSKGPEEASSEVQDTNEVHVPGD----QD--EPQTLGKKGSK
bt  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKMRSSKMAPKP---QSRCTSTRSAGEAPSE--------NQSPSTGPEEASSGVQDTTEVLVPGE----QE--EPQTLGRKGSK
mm  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKIRSGKLAPKP---QSRCTSTRSAGEAPSE--------KPSPSEGPEESAGEVQDTSRVRVPGE----QE--ELRMLGKKGSK
gg  S-----SGSLPPRLRLKPQRYRNEENDSSLK--TGLEKIRSGKMATKP---QSRCSSTRSAGEAPSE--------NQSPSKGPEEASSEVQDTSEVHVTVD----QD--EHQTLGKRGSK
xt  S-----SGTLPPRLRLKPQRYRNEENDSSLK--PGLETLRSSKIGIKP---QTRYSATRSAGETPSE--------IQSPSNGPEEISSEMQDTNVCVPPEE----QD--LQQTLGKRGSK
dr  SGGGPGSNPLPPRMRIKPQRYRNEENDSSTCKPPCLEKVPGGGRLPLPKPALPRCTSTRSSSSSCSASQVATAVEPQRPGKAIESLLSQAQPATEYREPQAEPEEQEGREEKRETRGSKA
     *       *    ****:*:**************     **.:  ..    *      * ::***:..: *          *. .*.  . * :.          *:    :     : -..

hs  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKMRSGKMAPKP---QSRCTSTRSAGEAPSE--------NQSPSKGPEEASSEVQDTNEVHVPGD----QD--EPQTLGKKGSK
bt  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKMRSSKMAPKP---QSRCTSTRSAGEAPSE--------NQSPSTGPEEASSGVQDTTEVLVPGE----QE--EPQTLGRKGSK
mm  S-----SGSLPPRVRLKPQRYRNEENDSSLK--TGLEKIRSGKLAPKP---QSRCTSTRSAGEAPSE--------KPSPSEGPEESAGEVQDTSRVRVPGE----QE--ELRMLGKKGSK
gg  S-----SGSLPPRLRLKPQRYRNEENDSSLK--TGLEKIRSGKMATKP---QSRCSSTRSAGEAPSE--------NQSPSKGPEEASSEVQDTSEVHVTVD----QD--EHQTLGKRGSK
xt  S-----SGTLPPRLRLKPQRYRNEENDSSLK--PGLETLRSSKIGIKP---QTRYSATRSAGETPSE--------IQSPSNGPEEISSEMQDTNVCVPPEE----QD--LQQTLGKRGSK
dr  SGGGPGSNPLPPRMRIKPQRYRNEENDSSTCKPPCLEKVPGGGRLPLPKPALPRCTSTRSSSSSCSASQVATAVEPQRPGKAIESLLSQAQPATEYREPQAEPEEQEGREEKRETRGSKA
     *       *    ****:*:**************     **.:  ..    *      * ::***:..: *          *. .*.  . * :.          *:    :     : -..

    S                                                                                             PWWP
hs  NNISVYMTLNQKKSDSSSASVCSIDSTDDLKSSNSECSSSESFDFPPGSMH-A---PSTSST---S--SSSKEEKKLSNSLKMKVFSKNVSKCVTPDGRTICVGDIVWAKIYGFPWWPAR
bt  SSISVYMTLNQKKSDSSSASVCSIDSTDDLKSSNSECSSSESFDFPPC-MH-A---PSTS-T---S--SSSKEEKKLSNSLKMKVFSKNVSKCVTPDGRTICVGDIVWAKIYGFPWWPAR
mm  SSISVYLTLNQETSDSSSASVCSIDSMDDLKSSNSECSSSESFVFPPGCMH-A---PSASST---STSFSSKEENSLRNSLKMKIFSKNVSKCITPDGRTICVGDIVWAKIYGFPWWPAR
gg  SNITVYMTLNQKKSDSSSASVCSSDSTDDLKSTNSECSSTESFDFPPGSMH-A---PSSSS-------SSSKEEKKLSNSLKMKVFSKKVSKCVTPDGRTICVGDIVWAKIYGFPWWPAR
xt  SNITVYMTINQKKANSSSASVCSSDSTDDMKSSHSECSSTENFDFPPGSMH-A---PSSSSS---S-SASSKEEKKLSNSLKIKFMSKNVSKCVTPDGRTICVGDIVWAKIYGFPWWPAR
dr  GNLVVYMTINPSQPDSSSTSMCSIDSADDLKSSNSECSSSTETFDFPPPGDLRSAPAPGTSSAFASASPSAPQDDKKRSKSLKAAVFSKNVTKCVALDGRTICVGDIVWAKILGFPWWPAR
     .: **:*:*  .  :***:*:** ** **:**:.*****:*.* ***      :   *.:*         : ::::.  :*** :***:*:**:: **************** ********

hs  ILTITVSRKDNGLLVRQEARISWFGSPTTSFLALSQLSPFLENFQSRFNKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFET
bt  ILTITVSRKDSGILVRQEARISWFGSPTTSFLALSQLSPFLENFQSRFNKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFES
mm  ILTITVSRKDNGLLARQEARISWFASPTTSLALSQLSPFLENFQLRFNKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFET
gg  ILTITVSRKDNGLLVRQEACISWFGSPTTSILALSQLSPFLENFQSRFNKKRKGLYRKAVTEAAKAAKQLTPEVRALLTQFET
xt  ILAITVSRKDTGLLVRQEARISWFGSPTTSFLALSQLTPFLENFQSRFNKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFET
dr  ILAITVSRKDNGLLVRQEARVSWFGSPTTSFLAITQLAPFLENFQSRFDKKRKGLYRKAITEAAKAAKQLTPEVRALLTQFET
     **:*******.*:*.**** :***.***** **::**:****** **:********:*:*******************:
```

Figure A.4 PWWP2A cross-species primary structure alignment. Amino acid sequences of PWWP2A from six organisms (hs: *Homo sapiens* [Q96N64], bt = *Bos Taurus* [F1MKS1], mm = *Mus musculus* [Q69Z61], gg = *Gallus gallus* [E1BWL2], xt = *Xenopus tropicalis* [B1H3K5], dr = *Danio rerio* [F1QXJ8]) as available from UniProt [ID] were aligned. Important regions based on the human PWWP2A protein sequence are highlighted with colored boxes and labeled with the respective names: P1 (proline-rich stretch 1) in grey, P2 (proline-rich stretch 2) in grey, I (the primary internal domain) in light green with IN separated from IC with a dotted line, conI (the highly conserved part of the internal domain) in dark green, S (serine-rich stretch) in grey and PWWP (PWWP domain) in red. Consensus symbols: fully conserved residue (*), conserved strongly similar residue (.) and conserved weakly similar residue (:). Alignments were carried out using Clustal Omega.

XII