Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation

Table of contents:

- 1. Appendix materials and methods
- 2. Appendix figures and legends S1-S5
- 3. Appendix references

1. Appendix materials and methods

Label-free quantitative mass spectrometry (LFQ-MS) and data analysis

Interacting proteins were identified by performing two-sample *t*-tests with permutation-based false discovery rate (FDR) statistics essentially as described before (Vardabasso et al, 2015). 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		
	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	0.08	1.0	
HeLa Kyoto	0.05	1.0	0.05	1.0	

First, 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). After GFP-background filtering, all remaining proteins were subjected to a second two sample *t*-tests (including FDR statistics), now comparing both GFP-H2A.Z.1 and GFP-H2A.Z.2 with GFP-H2A to identify interacting proteins specifically enriched on H2A.Z.1- or H2A.Z.2- versus H2A-containing mononucleosomes. The respective *p*-values (technical replicates) and *t*-test differences were plotted against each other in volcano plots using R (version 3.2.0). Proteins within the cutoff line (based on the FDR and S₀) of both SK-mel147

experiments were selected as promising candidates for further investigations. In addition to the volcano plots, interacting proteins were also visualized in heatmaps. To do so, the ratio (fold change) of the logarithmized and averaged triplicate LFQ intensities of GFP-histones (H2A, H2A.Z.1 and H2A.Z.2) were plotted as heatmaps using the Perseus software integrated in the MaxQuant environment (Cox & Mann, 2008).

nChIP-seq

Immunoprecipitations were carried out with S1 mononucleosomes derived from HeLaK cells stably expressing GFP or GFP-tagged H2A, H2A.Z.1, H2A.Z.2 or PWWP2A from 4 x 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 (10 mM Hepes pH 7.6, 100 mM NaCl, 1.5 mM MgCl₂, 0.5 mM EGTA, 10% glycerol (v/v), 10 mM β -glycerolphosphate, 1mM DTT, 1x protease inhibitors), added to 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 as previously described (Cuddapah et al, 2009). For GFP-PWWP2A IPs, mononucleosomes prepared from stable HeLa Kyoto GFP-PWWP2A cell line were precleared with G-protein magnetic beads (Millipore) for 1 h at 4°C and subsequently incubated over night (ON) with 2 µl of GFP antibody (kind gift of Prof. Andreas Ladurner, LMU Munich) 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. An IP with an unspecific rabbit IgG antibody served as negative control. Beads were washed as described for GFP-trap IPs, excluding the LiCl buffer wash step. For native H3K4me3 IPs and GFP-H2A.Z.1 IPs after control or PWWP2A depletion or Ch-PWWP2A overexpression, mononucleosomes from wild type HeLaK cells were prepared. The MNase digestion protocol was optimized for the use of 4 x 10^6 cells per IP. The precleared mononucleosomes were incubated with 1 µl of a rabbit H3K4me3 antibody (Diagenode) or GFP-Trap magnetic beads (ChromoTek). All further steps were carried out as described for GFP IPs. 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 μ l TE containing 0.5 M NaCl, magnetically separated and the supernatant mixed with the first supernatant. Input fractions were processed in parallel. After Phenol/chlorophorm/isoamylalcohol extracting and ethanol precipitating the DNA, the IP DNA pellet was resuspended in 12 μ l and the input DNA pellet in 32 μ l 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).

Illumina Sequencing libraries were established with the MicroPlex Library Preparation Kit (Diagenode) following the manufacturer's instructions with the following variations. The number of step 5 amplification cycles was scaled according to the amount of input material. When purifying libraries after amplification, samples were incubated 15 min with AMPure beads (Beckman Coulter) 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.

Illumina Sequencing

Next generation sequencing of nChIPs were performed by Dr. Stefan Krebs at the Laboratory of Functional Genome Analysis (LAFUGA) in Munich (Gene Center) or Dr. Gregor Gilfillan (NGS). 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.

nChIP-seq analyses

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.

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.

Quantitative PCR (qPCR)

qPCR on cDNA was carried out as previously described (Jack et al, 2013). Results were normalized to HPRT1 and GAPDH levels. For validation of PWWP2A knockdown efficiency primers recognizing the main splice product (PW 2: Fwd: 5'-GAAGTGCGGGCTTTGTTGAC-3', Rev: 5'-CTCCAATCTGGCCACGCTAT-3') predicted splice or recognizes all forms (PW all: Fwd: 5'-ACGGTGTCGCAACTGATCC-3', Rev: 5'-CCATGGGGCCCAAACCTTTT-3') were used.

GFP-H2A.Z.1 enrichments of percent input after PWWP2A depletion or overexpression were determined via ChIP-qPCR of control gene body locus (RPL11: Fwd: 5'-ACAGCTTTGGGTGATGCAGT-3', Rev: 5'-TTGTTGGACCAAAACACGGC-3') and of defined H2A.Z.1 binding sites 5'-GGGATGCAAGTGGGAAAAC-3'. 5'-(PARS2+: Fwd: Rev: PARS2-: 5'-ATTGCGGTAGGTGAACGTG-3'; Fwd: 5'-AGACGCCTTTATTACAGTGCCC-3', Rev: TCTACGTGGTAGCAGCTCAAAA-3'; NUF2+: Fwd: 5'-CACTGTAGGTGAGCGCGAGA-3', Rev: 5'-CGCTGAGCACGACGAAAACA-3'; NUF2-: 5'-GCATCTAACAAAACCCGGCAC-3', 5'-Fwd: Rev: GTCCGAGTTGAAGAGCAAAACC-3').

Recombinant expression and purification of GST proteins and in vitro binding studies

50 µl of competent E.coli (BL21) were transformed with 50 ng plasmid DNA and incubated for 30 min on ice. After heat shock (42°C, 45 sec) and cool down phase (2 min, ice), 100 µl LB (lysogeny broth) medium was added and cells were incubated for 1 h at 37°C, plated on ampicillin agar and incubated again at 37°C ON. Single clones of grown colonies were picked, inoculated in 6 ml LB medium with 0.1 mg/ml ampicillin and incubated at 37°C while shaking at 150 rpm ON. Induction with 0.3 mM IPTG (Roth) occurred after inoculating 2 ml of pre culture (OD of 600 nm) in 200 ml LB medium with 0.1 mg/ml ampicillin. After 16-18 h of induction, bacteria culture was centrifuged for 20 min, 4000 rpm at 4°C and the obtained pellet was quickly frozen with liquid nitrogen. Pellet was resuspended in 10 ml lysis buffer (1xPBS, 0.4 M NaCl, Protease Inhibitor Cocktail (Roche, Germany)) and sonicated (Branson Sonifier 250-D). After sonication, cells in lysis buffer were centrifuged for 30 min, 15,000 rpm at 4°C. Supernatant was mixed with 200 µl of 2x washed (1xPBS) glutathione sepharose beads (GE healthcare) and incubated at 4°C for 2 h. Beads were washed 3x5 min with washing buffer (1xPBS, Protease Inhibitor Cocktail (Roche, Germany)). For several experiments the elution of the protein from the beads followed by dialysis was necessary. Elution was performed 3x1 h (rotating 4°C) in elution buffer (1xPBS, 30 mM glutathione (Sigma)). Eluates were kept in dialysis membrane (MWCO 6-8.000, Spectra/Por, Roth) and 200 µl were dialyzed in 1 l dialysis buffer (1xPBS, 50 mM NaCl, 10% (v/v) glycerol) while stirring ON. Purified recombinant GST-protein was concentrated by centrifugation (20 min, 4000 rpm, 4°C) with centrifugal filter units (Millipore).

Electrophoretic Mobility Shift Assay (EMSA)

DNA shifts

A 4.5% native acryl amide gel (0.3x TBE, 4.5% acryl amide, 5% (v/v) glycerol, 0.00075% APS, 0.0001% TEMED) was used for DNA shifts. After polymerization, the gel was pre-run (30 min, 15 mA, 4°C) in running buffer (0.3x TBE, 5% (v/v) glycerol). 1 μ M Cy5-labeled R-601 DNA (Eurofins Genomics) was diluted to a concentration of 100 nM in binding buffer (1x TE pH 7.4, 50 mM NaCl, 10% (v/v) glycerol, 0.1% BSA, 0.5 mM DTT). Purified recombinant GST-protein was

centrifuged (30 min, 15.000 rpm, 4°C) and diluted to desired concentrations in binding buffer. 100 nM Cy5-labeled R-601 DNA and different protein concentrations were mixed 1:1 and incubated for 30 min at RT. 20 μ l of each sample were loaded on a native 4.5% acryl amide gel and ran (75 min, 110 V) on ice. The gel was analyzed using a Phosphoimager (Fujifilm).

Assembly of recombinant nucleosomes

Human histones (H2A, H2A.Z, H2B, H3, H4) were provided by Hataichanok Scherman (www.histonesource.com). The 187 bp long DNA fragment (pUC18 with Widom 601 sequence, gift from F. Mueller-Planitz) used for mononucleosomes with 20 bp linker DNA to both sites (20N20) was prepared by PCR using Cy5- or Cy3-labeled primers (Sigma). The PCR product was PEG- and isopropanol precipitated and resolved in 1x TE (pH 8). The assembly of histone octamers and the reconstitution of mononucleosomes were performed by salt-gradient dialysis as previously described (Luger et al, 1999; Mueller-Planitz et al, 2013). The purity of mononucleosomes was enhanced by applying glycerol gradient ultracentrifugation. Nucleosomal DNA concentration was determined by measuring the UV absorption (260 nm).

Competitive recombinant mononucleosome shifts

A 5% native acryl amide gel (0.2x TBE, 5% acryl amide, 0.00075% APS, 0.0001 % TEMED) was used for recombinant mononucleosomal shifts. After polymerization, the gel was pre-run (1h, 100V, on ice) in running buffer (0.2x TBE). Reconstituted, Cy3- (H2A.Z) or Cy5-(H2A) labeled mononucleosomes (20N20) were diluted to a final concentration of 20 nM in binding buffer (1x TE pH 7.4, 50 mM NaCl, 10% (v/v) glycerol, 0.1% BSA, 0.5 mM DTT). The recombinant GST-proteins were diluted in dilution buffer (150 mM NaCl, 25 mM HEPES pH 7.6, 2mM MgCl₂, 10% glycerol, 0.1 mM EDTA, 0.05% NP-40). For competitive EMSAs 20 nM of each recombinant mononucleosome (H2A or H2A.Z) and different protein concentrations were mixed 1:1 and incubated for 30 min on ice. 15 μ L were loaded on a native 5% acryl amide gel and ran (90 min, 100-110 V) on ice. The gel was analyzed using the Typhoon FLA9500 (GE Healthcare Life Sciences).

Fluorescence Recovery After Photobleaching (FRAP) and live cell imaging

FRAP analyses were performed 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 $8-10 \mu m$, 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 frames were recorded at maximum speed (197 ms). Postbleach frames were recorded as follows: 150 images at maximum speed and 180 images at a rate of 1 frame per second 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.

RNAi, cell cycle and proliferation analysis

Determination of cell growth after RNAi was performed as previously described (Wiedemann et al, 2010), without changing medium during the time of the experiment. Proliferation curves were generated by Excel. Determination of cell cycle stages was done by flow cytometry. Briefly, 3 days after RNAi $1x10^{6}$ HeLaK cells were washed with PBS, pelleted and fixed with 70% methanol for at least 1 h at - 20°C. After one round of PBS washing and pelleting, cells were either directly incubated with propidium iodide (PI) (details see below) or first stained with α -H3S10ph primary antibody for 1 h at RT. Cells were washed with PBS, pelleted and stained with secondary antibody coupled to AlexaFluor 488 for 1 h at RT in the dark. After PBS wash and pelleting, cells were treated with 100 µg/ml RNase A (Sigma-Aldrich) and 50 µg/ml PI for 30 min at 37°C in the dark and subjected to cell cycle analysis using a FACSCanto (BDI Bioscience) machine. PI positive cells were plotted in a linear histogram, allowing a discrimination of cells in G₁, S or G₂/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). Discrimination between cells in G_2 and M phase was achieved by specifically staining mitotic cells for H3S10ph. 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.

IF microscopy (confocal, chromosome spreads)

Single optical sections or stacks of optical sections were collected using a TCS SP5 or a SP5 II (Leica) confocal microscopes equipped with Plan Apo $63 \times /1.4$ NA oil or Plan Apo CS 63x/1.3 NA glycerol immersion objectives and lasers with excitation lines 405, 488, 543 and 594 nm. Laser intensities were adjusted using the acoustooptical tunable filters (AOTFs). Photomultiplier tubes (PMTswere used in sequential mode and a smart gain of 600 - 800V. 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' (iterations: 40, signal/noise: 15, quality threshold: 0.000005, iteration mode: optimized, bricklayout: auto). Images were cropped to 512 x 512 pixels and background intensities manually set for every image.

Nuclear area of PWWP2A RNAi or control cells was determined as number of pixels/nucleus for 100 cells per condition in three independent experiments with ImageJ.

2. Appendix figures and figure legends



Appendix Figure S1 - PWWP2A internal domains facilitate H2A.Z-specificty and nucleosome binding.

Appendix Fig. S1

A Amino acid sequence of human PWWP2A [Q96N64]. P1 and P2: proline-rich regions 1 and 2; gray; S: serine-rich, blue; PWWP: red.

B, **C** Coomassie-stained SDS-PAGE gel of GST-deletions (see also Fig. 2E). * indicates respective GST-proteins.

D Pulldowns of GST or GST-internal deletions with mononucleosomes derived from HeLaK cells followed by Coomassie staining. Notice that I, conI (evolutionary conserved region in I) and IN (see also Appendix Fig. S1C) efficiently pull-down nucleosomes as seen by histone bands, while GST control does not and IC binds only trace amounts of nucleosomes (see also Fig. 2F and G).



Appendix Figure S2 - PWWP2A is strong chromatin binder.

A Pull-downs of recombinant GST or GST-internal (I) or GST-PWWP (PWWP) with mononucleosomes derived from HeLaK cells. Precipitated recombinant GST proteins and histones are detected with Coomassie Blue staining and H3 and H2A.Z with specific antibodies in immunoblots. * indicates right size of purified and precipitated recombinant proteins. Notice the reduction in H2A.Z binding of PWWP domain in comparison to I region.

B FRAP quantification curves of average GFP signal relative to fluorescence signal prior to bleaching from transient expressed GFP (negative control), GFP-tagged histones, GFP-tagged chromatin binders (human HP1 β and mouse UHRF1)

(Darmochwal-Kolarz et al, 2003; Rottach et al, 2010) and GFP-PWWP2A (n=6-11). See also Fig. 3C for FRAP experiment with PWWP2A deletion constructs.

C, **D** FRAP experiment to evaluate chromatin association of wild type and deletion GFP-PWWP2A and GFP-H2A.Z variants using spinning disk confocal microscopy. Half nucleus of HeLaK cells transient (**C**) or stable (**D**) expressing GFP-tagged PWWP2A proteins was photobleached and the recovery of the signal was monitored for 120 seconds. For each construct, three selected time points of one exemplary time series are shown. Scale bar = 5 μ m.

E FRAP quantification curves of average GFP signal relative to fluorescence signal prior to bleaching from stable expressed GFP (negative control), GFP-tagged histones, and GFP-tagged PWWP2A mutants (n=10-11). See also Fig. 3C for FRAP experiments using transient expressed GFP-tagged proteins. Notice that transient and stable transfections show similar recovery kinetics and are therefore comparable.

Appendix Fig. S3

REVIGO treemap de-regulated genes

positive regulation of developmenta process	urogenita system developme	al regulation developme ent process	posit of regula intal of leuk s apopt proce	ive ation ocyte totic ess	nbrane arization	positive regulation of cellular component biogenesis	positive regulation of biological	cellular response to transforming growth factor beta stimulus	response to drug	e sigr transdi	al uction el	onse to r ogenous mulus v	esponse to vounding	cell communication —Cell C	small molecule metaboli	terpen metabo proce	bid blic ss ation
negative regulation of protein metabolic	positive regulatior of cell	positive regulation of cellula compone	e on g ar deve	gland elopment	regulat peptidyl- phospho	tion of -serine rylation de	digestive system evelopment	response to stress	endoplasmic reticulum stress	to inorgani substanc	e response to topologicall incorrect protein	y enzyme linke receptor protein signaling pathway	to growth hormone	single-organism cellular process	single-organis metabolic process	n of lipic metabo proces	^{on} retinoid I Iic metabolic _s process
cell projection morphogenesis	cell part morphogenesis	regulation of catabolic	ion prostate gland developme	orga subst	nic _r ance d	egulation of alcohol iosynthetic	positive regulation of lymphocyte	cellu respon90	owth fa	onse ctor l	to tra peta si	imulu	ning s	plasma Golgi t	egulation if receptor	of mag	cell rowth
protein si positiv	regulation e.regulation	process ation o	regulation	proc elopn	ess renta	process solution of tuter protein alopro	negative regulation of CCSSI	chemical	response to growth factor	regulation of cell communicatio	regulation of response to stimulus	cellular response to metal ion	tyrosine phosphorylation of STAT protein	trar naintenance		ediated of	esponse to stimulus
positive regulation of organelle	quality positive regulation of	aging	regulation o nterleukin-1 production	t of neuro	ment of ons mi	gulation f smooth uscle cell pliferation	proteolysis	pathway-restricted SMAD protein phosphorylation	wound healing	to abiotic stimulus	response to oxygen levels	to hypoxia	y yan pangan pangan Pangan pangan pangan Pangan pangan	catabo	lic sm gro	wth	signaling
regulation of protein catabolic process	ossification regulation of protein metabolic	positive regulation of endothelial cell proliferation	<u></u>	type 2 immune response	celtular composent assembly involved in morphogenes	regulatio of localizatio	on negative regulation of cellular component organization	cellular componer	protein con subuni It organizat	nplex ex t ion or	tracellular matrix ganization	of cellul compone biogenes	on ceil projection ent organization ŝis	cellular component organization or biogenesis	death	cell activati	organonitrogen compound metabolium
digestive tract development	positive regulation of cellular amide metabolic process	protein catabolic process	nterleukin-10 production	mazonokoule catabolic process	gastrulatic	organ regeneratio	regulation of megalatryocyte differentiation	regulation	lular ço	ere ex mpoi oly ^{ori}	tracellular nent as ganization	semb	yvesicle organization	developmental process	immune system process	multicellul organisma process	ar chondroitin sulfate metabolism
positive regulation of lipid biosynthetic process	positive regulation of catabolic process	regulation of cysteline-type endopeptidase activity	egulation of protein stability	muscle structure development	leukocy mediate immuni	rte ^{regulation} of catalytic ty activity	C morphogenosis	component organization	containir chromati organizati	ig ren in ce on ce	hromatin nodeling at ntromere	protein-DN complex subunit organization	macromolecular complex subunit organization	single-organism process	monosaccharide metabolism	phosphor phosphor metabolisi process	us collagen collagen metabolism process

Appendix Figure S3 - PWWP2A affects transcription of genes involved in development, cellular response and component assembly.

GO term analysis of deregulated transcripts upon PWWP2A RNAi (see also Fig. 5C for scatter plot).



Appendix Figure S4 – PWWP2A does not affect H2A.Z genomic localization A FRAP experiment to evaluate chromatin association of GFP-H2A or GFP-H2A.Z.1 two days after control (wt or Luci) or PWWP2A (PW#1, PW#2) RNAi (see Fig. 7A), performed identical as described in Appendix Fig. S2C. Scale bar = 5 μm.

B FACS sorting of lowly or highly expressed Ch-PWWP2A in stably expressing GFP-H2A.Z HeLaK cells.

C FRAP experiment to evaluate chromatin association of GFP- Δ IC (see Fig. 7C) performed identical as described in Appendix Fig. S2C. Scale bar = 5 μ m.

Homo sapiens Xenopus laevis Xenopus tropicalis	10 M A A V A A E A A A M A A	T A A S P G E G V A A A P G P G V A G A P G P G	20 GAGEAEPEME EAGESEQDSK EGGESEQDSE	30 PIPGSE TTPGEWRLGE TIPGERRLGR	40 A G T D P I V A V L L C E P M E A E L	50 D S E R Q S L I R C D S E R R S L V R C	60 E M D S D T E P C C G M E I D T	
Homo sapiens Xenopus laevis Xenopus tropicalis	A SVPDGETDC KAGPVVEQLC NAGPDIEQLC	Q Q S A P Q	90 LCECRGDGLG	100 F P A G R D D R D V G	110 E E T M Q V V D E H	120 V M A S K P V E M E	130 5 Q S C P E N F L Q	140 2 T Q
Homo sapiens Xenopus laevis Xenopus tropicalis	S S E Q S I S A F S	Q G A A V E Q E	160 CISEEHSEAS	170 H T A L H P P E P E	180 LQMYVPQLEE	190 SEFQQLQMHS	200 D V L Q P P Y T D I	210 L Q
Homo sapiens Xenopus laevis Xenopus tropicalis	MPGDVSQLPE	PDAQQLQV	230 H S E V P Q A L E P	240 ELQQLHMHGN	250 V S Q P P E P E F Q	260 Q Q E S Y V P Q P P	270 E P E V Q H Q E M Y	280 (G G
Homo sapiens Xenopus laevis Xenopus tropicalis	290 V P R P L E L E V (QQEMYGDV	300 S Q P A A L E V Q P	310 	320 I V V Q Q D M Q Q E	330 L P I S L Q V M Q V	340 	350 K P
Homo sapiens Xenopus laevis Xenopus tropicalis	360 LNGPVPEET	SGGFIAQE	370 PLSIGEVLLK	380 E A L L V E K G V S	390 GEALSVGDTK	400 V L S V G D I R S E	410 SIGQAGPEET	420 F I S
Homo sapiens Xenopus laevis Xenopus tropicalis	430	SLREEGPG	440 D T I P L F N K G F	450 N D P P S M Q K S S	460 FGEILPIQTD	470 Y A E E L L P I R E	480 N R P A E G S L S G	490 5 A L
Homo sapiens Xenopus laevis Xenopus tropicalis	500 L D Q R G N I A T E	K D T E E S P D	510 	520 5 A A G K I L S Q E E	530 L V P E I A K E Q A	540 A T E S L P V I E T	550 I S G E V L A V Q E	560
Homo sapiens Xenopus laevis Xenopus tropicalis	570 T M V S L T Q Q K (LDVTPIKQ	580 S G L L K I C A L E	590 T Q P Q G S Q K N F	600 P S D S E H L L E E	610 	620 	630 2 E A
Homo sapiens Xenopus laevis Xenopus tropicalis	640 Q N T H V P E S C 5	LPQNCIPD	650 D Q H I K E L V P E	660 C Q R T Q K P A N V	670 Q G P V S G D I L Q	680 PDKESMELYQ	690 Q H S L E F I G A H	700
Homo sapiens Xenopus laevis Xenopus tropicalis	710	SESFSETK	720 	730 	740 	750 		770
Homo sapiens Xenopus laevis Xenopus tropicalis	780	IISFEKDHE	700 SPLFIVQSDG	800 RNVSCLENKK	810 T H E L C M T V D T	820 F L K T Y P N K L G	830 E E N N T K N I T C	840
Homo sapiens Xenopus laevis Xenopus tropicalis	850	SDQLSKDM	840 A M E T I N G S D S	870 5 E N Y A F A H E L I	880 PTDKPVTEDT	890 	900 ADEPPLP 	910 P P K T K T
Homo sapiens Xenopus laevis Xenopus tropicalis	920 PP		930 P P G K <u>P P Y R G I N Q E</u> M G I N Q E	940 ELARSPEAVG RIAGENKKAR QIAPENRKAR		960 V R V A E S A A A A Q K E P T P M I L S Q R Q P I P M I L S	970 PQGGPELPPS EVRDSGVHTE EVTDSGIYTG	980 P A M C
Homo sapiens Xenopus laevis Xenopus tropicalis	SPPEQPPAPE LEEATLSSNE LEEAKFSSNE			¹ 2 ¹⁰¹⁰ T S E G S V V T S G T T F G S V V T S G	1020 V S Q L I P G S E V I G S L I P G S E V I D S L I P G S F V		1040 ALVVSFRFGE ALVVSFCHGN ALVVSFCHGN	1050 K L R I
Homo sapiens Xenopus laevis Xenopus tropicalis	1060 FSGVLMDLSF FSGVLMDLSF FSGVLMDLSF	RFGPHGIP RFGPHGIP RFGPHGIP	1070 V T V F P K R E Y K V T V H P R R E Y K V T V H P R R E Y K		1090 SNTFQEGTEV SKSFHEETLU SKSFHEEPLV			1120 S L E E E C K
Homo sapiens Xenopus laevis		PLFHEGAP PLFHEGAP	1140 Y P P P L F I R D T Y P P P L F I R D T Y P P P L F I R D T		I 1160 R K I K R P K R K M R K I K R P K K K I P K I K R P K K K I	1170 Y R E E P T S I M N Y R E E P T S I M N Y R E E P T S I M N		1190 C D C D
Homo sapiens Xenopus laevis	1200 K C K N S V V A E P K C K N N V V A D P	KEIRKGSS KEIRK - V	1210 A T D S S K Y E D - A T D S Y K T E E G A T D S Y K T E E G	1220 KKRRNESVTT KRRHETITT KBRRHETVTT	1230 V N K K L K T D H K V N K K L K T D H K V N K K L K T D H K	1240 V D G K N Q N E S Q V N G K G Q H E S Q V N G K S Q N E S Q		
Homo sapiens Xenopus laevis					1300 V L K I A K E K A Q V L K M A K E K A Q		1320 N A H S K V H F T R N A H S K V H F T R	1330 R Y R R L
Homo sapiens Xenopus laevis				1360 KTGLEKMRSG		1380 T S T R S A G E A P S A T R S A G E A P	1390 SENQSPSKGP SEIQSPSNGP	1400 E E E E
Homo sapiens Xenopus laevis		V H V P G D Q D C V P P E E Q D	1420 E P Q T L G K K G S L Q Q T L G K R G S		S 1440 N Q K K S D S S S A N Q K K A N S S S A	1450 S V C S L D S T D D S V C S S D S T D D	1460 LKSSNSECSS MKSSHSECSS	1470 ISE STE
Aenopus tropicalis Homo sapiens Xenopus laevis	1480 SFDFPPGSMF NFDFPPGSMF		1490 S S S K E E K K L S A S S K E E K K L	1500 . S N S L K M K V F S . S N S L K I K M F S	1510 K N V S K C V T P D K N V S K C V T P D	S C S S D S T D D PWWP 1520 G R T I C V G D I V G R T I C V G D I V	m n S S H S E C S S 1530 W A K I Y G F P W W W A K I Y G F P W W	1540 / P A / P A
xenopus tropicalis Homo sapiens Xenopus laevis			1560 E A R I S W F G S P E A R I S W F G S P	ISTO TTSFLALSQL	N N V S K C V T P D	GRTICVGDIV 1590 FNKKRKGLYR FNKKRKGL <u>Y</u> R	1600 KAITEAAKAA KAITEAAKAA	1610 K Q
Xenopus tropicalis Homo sapiens Xenopus laevis Xenopus tropicalis	R T L A I T V S R ■ 1620 L T P E V R A L L T L T P E V R A L L T L T P E V R A L L T	QFET QFET QFET	EÁRISWFG SP	TTSFLALSQL	TIFFLENFOSR	FNKKRKGLYR	<u>K A I T E A A K A A</u>	KQ

Appendix Figure S5 - Sequence alignment of PWWP2A *Xenopus* homologs and human protein.

Protein sequences of *Homo sapiens* [Q96N64], *Xenopus laevis* [XeXenL6RMv10002517m, Xenbase] and Xenopus tropicalis [XM_002940175, Xenbase] were aligned using MacVector (12.7.3) ClustalW multiple sequence alignment tool with default settings. Identical amino acid are highlighted in dark gray, similar amino acids in light gray and changes are set apart on white background. Important regions based on the human PWWP2A sequence are depicted in colored

boxes and labeled with respective names. P1 region is not shown, since it is not conserved between species, P2: proline-rich region 2; gray; I: internal region, green; S: serine-rich, blue; PWWP: red.

3. References

Cox J, 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

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

Darmochwal-Kolarz D, Rolinski J, Tabarkiewicz J, Leszczynska-Gorzelak B, Buczkowski J, Wojas K, Oleszczuk J (2003) Blood myeloid and lymphoid dendritic cells are stable during the menstrual cycle but deficient during mid-gestation. *Journal of reproductive immunology* **59**: 193-203

Fox MH (1980) A model for the computer analysis of synchronous DNA distributions obtained by flow cytometry. *Cytometry* **1:** 71-77

Jack AP, Bussemer S, Hahn M, Punzeler S, Snyder M, Wells M, Csankovszki G, Solovei I, Schotta G, Hake SB (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

Luger K, Rechsteiner TJ, Richmond TJ (1999) Preparation of nucleosome core particle from recombinant histones. *Methods in enzymology* **304:** 3-19

Mueller-Planitz F, Klinker H, Ludwigsen J, Becker PB (2013) The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nature structural & molecular biology* **20:** 82-89

Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic acids research* **38**: 1796-1804

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

Vardabasso C, Gaspar-Maia A, Hasson D, Punzeler S, Valle-Garcia D, Straub T, Keilhauer EC, Strub T, Dong J, Panda T, Chung CY, Yao JL, Singh R, Segura MF, Fontanals-Cirera B, Verma A, Mann M, Hernando E, Hake SB, Bernstein E (2015) Histone Variant H2A.Z.2 Mediates Proliferation and Drug Sensitivity of Malignant Melanoma. *Molecular cell* **59**: 75-88

Wiedemann SM, Mildner SN, Bonisch C, Israel L, Maiser A, Matheisl S, Straub T, Merkl R, Leonhardt H, Kremmer E, Schermelleh L, Hake SB (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