Ruprecht − Karls − University
Heidelberg

Dissertation

The histone fold protein CHRAC-14 controls CENP-A loading and gene expression in *Drosophila melanogaster*

M.Sc. Sarah Doppler
2022

Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht − Karls − University

Heidelberg

Presented by

M.Sc. Sarah Doppler

born in: Schwetzingen, Germany

Oral Examination: February 24th, 2022

The histone fold protein CHRAC-14 controls CENP-A loading and gene expression in *Drosophila melanogaster*

Referees: Prof. Dr. Elmar Schiebel

Prof. Dr. Sylvia Erhardt

SUMMARY

Centromeres are specialized chromatin domains present on each chromosome, which determine the location of kinetochore formation – the multiprotein platform for spindle microtubule attachment during mitosis. Since the underlying centromeric DNA sequence is not conserved, centromere inheritance and function rely on the conserved epigenetic marker CENP-A (also known as CID in *Drosophila*). CENP-A is highly enriched in centromeric chromatin and partially replaces canonical histone H3. Specific loading and stabilization of CENP-A at the centromere is ensured by a concert of different factors such as its dedicated chaperone called CAL1 in *Drosophila*. Whereas the primary DNA sequence is dispensable for centromere formation, CENP-A is both absolutely essential and sufficient to maintain and establish functional centromeres. When overabundant due to transcriptional upregulation for example in cancer cells, CENP-A escapes the endogenous loading pathway and hijacks other histone loading machineries, which leads to promiscuous non-centromeric CENP-A incorporation throughout the genome with detrimental effects for genome stability.

Even though ectopic CENP-A accumulation existentially threatens cellular and organismal viability, CENP-A is distributed at basal levels throughout the genome under physiological conditions. The function and mechanistic details of genome-wide CENP-A loading remain elusive, but the field speculates, that low levels of ectopic CENP-A confer epigenetic plasticity and prime chromatin for neocentromere formation in the event of ancestral centromere loss. Previous work in our lab identified the histone fold protein CHRAC-14 as an important regulator of ectopic CENP-A incorporation and as a new DNA damage factor in *Drosophila melanogaster*. CHRAC-14 depletion leads to the increase of CENP-A levels, increased CENP-A ectopic loading – possibly at telomeric DNA repair sites – and to DNA repair defects accompanied by G2/M checkpoint failure.

Since we lacked detailed mechanistic insights and clear data whether ectopic CENP-A localizes to and has a role at DNA lesions, I set out to further elucidate this pathway. Employing a portfolio of biochemical and molecular biological techniques such as CUT&Tag-Seq, RNA-Seq and immunoprecipitation combined with mass spectrometry analysis, I report in this thesis, that CHRAC-14 knockdown leads to the accumulation of CENP-A at centromeres and genome-wide. In addition, altered expression of genes associated with gene ontology categories such as ‘mitotic progression’ and ‘DNA damage response’ was observed. Furthermore, I identified candidate genes, which are mis-expressed whilst showing increased CENP-A binding. Moreover, I found, that both CHRAC-14 and CENP-A interact with Casein kinase 2 (CK2) and are phosphorylation substrates thereof. Interestingly, CK2 phosphorylation of CHRAC-14 seems to promote the turnover of a post translationally modified version of

CENP-A, whereas CK2 phosphorylation of endogenous CENP-A itself seems to be essential for protein stability.

Taken together, my study provides new insights into CHRAC-14 mediated CENP-A titration in chromatin and opens up a novel direction involving CK2 as a key regulator in this pathway.

ZUSAMMENFASSUNG

Zentromere sind spezialisierte Chromatindomänen, die auf allen Chromosomen vorhanden sind und die jene Stellen festlegen, an denen sich das Kinetochor bildet – ein Multiprotein-Komplex, welcher der Anheftung von Mikrotubuli des Spindelapparates während der Mitose dient. Da die zugrunde liegende zentromerische DNA-Sequenz nicht konserviert ist, hängen Vererbung und Funktion von Zentromeren vom konservierten epigenetischen Marker CENP-A (in *Drosophila* auch als CID bekannt) ab. CENP-A ist in zentromerischem Chromatin stark angereichert und ersetzt dort teilweise das kanonische Histon H3. Das spezifische Einbauen und Stabilisieren von CENP-A am Zentromer wird durch ein Zusammenspiel verschiedener Faktoren sichergestellt, wie beispielsweise durch sein zugehöriges Chaperon CAL1 in *Drosophila*. Während die primäre DNA-Sequenz für die Zentromerbildung vernachlässigbar ist, ist CENP-A sowohl unbedingt notwendig als auch ausreichend, um funktionsfähige Zentromere zu erhalten und zu etablieren. Wenn CENP-A aufgrund hochregulierter Transkription, beispielsweise in Krebszellen, im Überfluss vorhanden ist, kann es seinen endogenen Ladeweg umgehen. CENP-A nutzt dann Ladeverfahren anderer Histone, was zu einem promiskuitiven, nicht-zentromerischen Einbau von CENP-A im gesamten Genom führt und sich schädlich auf die Genomstabilität auswirkt.

Obwohl die Lebensfähigkeit von Zellen und Organismen durch ektopische CENP-A Akkumulation existenziell bedroht wird, ist CENP-A unter physiologischen Bedingungen dennoch in marginalen Mengen im gesamten Genom verteilt. Die Funktion und mechanistischen Details des Einbaus von genomweitem CENP-A sind bisher größtenteils unbekannt, doch das Wissenschaftsfeld spekuliert bereits, dass eine niedrige Konzentration an ektopischem CENP-A womöglich epigenetische Plastizität verleiht und Chromatin für die Bildung von Neozentromeren vorbereitet, sollte das ursprünglichen Zentromer abhandenkommen. Eine vorherige Studie aus unserem Labor identifizierten CHRAC-14, ein Protein mit Histon-Faltung, als wichtigen Regulator für den Einbau von ektopischem CENP-A und außerdem als neuen Akteur im Bereich der DNA-Schädigung in *Drosophila melanogaster*. Die Depletion von CHRAC-14 führt zu einem CENP-A Proteinanstieg, zu vermehrtem Einbau von ektopischem CENP-A – womöglich an telomerischen Regionen aktiver DNA-Reparatur – und zu DNA-Reparaturdefekten, die mit einem Versagen des G2/M Checkpoints einhergehen.

Da diesbezüglich detaillierte, mechanistische Erkenntnisse und klare Daten, ob ektopisches CENP-A an DNA-Läsionen lokalisiert und dort eine Rolle spielt, fehlen, werde ich diesen Regulationsweg hiermit eingehend erforschen. Mit einem Portfolio an biochemischen und molekularbiologischen Techniken wie CUT&Tag-Seq, RNA-Seq und Immunpräzipitation in Kombination mit Massenspektrometrie zeige ich in dieser Arbeit, dass die Reduzierung von CHRAC-14 zur Akkumulation von CENP-A am Zentromer wie auch genomweit führt. Zudem

sind Veränderungen in der Expression von Genen zu beobachten, die mit genontologischen Kategorien wie ‚mitotische Progression' und ‚DNA-Schadensantwort' in Verbindung stehen. Darüber hinaus habe ich differentiell exprimierte Genkandidaten identifiziert, die gleichzeitig eine erhöhte Bindung an CENP-A aufweisen. Außerdem konnte ich ermitteln, dass sowohl CHRAC-14 als auch CENP-A mit Casein Kinase 2 (CK2) interagieren und zu deren Phosphorylierungssubstraten gehören. Interessanterweise scheint die CK2 Phosphorylierung von CHRAC-14 den Umsatz einer posttranslational modifizierten Version von CENP-A zu fördern, während die Phosphorylierung von endogenem CENP-A selbst essentiell für dessen Proteinstabilität zu sein scheint.

Zusammenfassend liefert meine Studie neue Einblicke in die CHRAC-14-vermittelte Titration von CENP-A in Chromatin und eröffnet somit eine neue Richtung, in der CK2 als Schlüsselregulator innerhalb dieses Signalweges beteiligt ist

Table of Contents

1 Introduction

1.1 Chromatin

Eukaryotes are faced with a complex problem: On the one hand their genomic information encoded by the DNA must be accessible for transcription or replication. On the other hand, DNA is a long linear molecule (one large *Drosophila* chromosome is ~5 cm long) and must be fitted into a single nucleus with a size in the µm scale (5 µm in *Drosophila*; Schwartz and Cavalli, 2017). Moreover, DNA must be extensively ordered and compacted for faithful chromosome segregation during mitosis (Bonev and Cavalli, 2016; Schwartz and Cavalli, 2017). To meet these contrary requirements, chromatin provides a 3D scaffold to organize and compact DNA. Only if DNA polymers are structured and ordered through chromatin spatio-temporal regulation of gene expression or silencing and accurate chromosome segregation during cell division is possible. The term chromatin refers to a supramolecular, architectural complex, which, at its basic scale, consists of DNA wrapping around histone octamers forming the nucleosome (Bonev and Cavalli, 2016; Schwartz and Cavalli, 2017; Takada *et al.*, 2020)

1.1.1 Histones and their paralogs at the basis of nucleosome formation

In order to form the basic unit of chromatin - a nucleosome core particle - two histones interact with each other via their histone fold domains forming heterodimers consisting either of H2A-H2B or H3-H4. Two copies of each dimer then associate to form an octamer, which binds and wraps ~146 bp of DNA (Luger *et al.*, 1997). A (mono-)nucleosome in its complete form is often defined by the nucleosome and the histone-free linker DNA and the linker histone H1, which binds close to the dyad of the nucleosome (Brockers and Schneider, 2019; Cutter and Hayes, 2015; Parmar and Padinhateeri, 2020; Takada *et al.*, 2020).

Histones not only contain a histone fold domain for histone dimerization, they also exhibit N-terminal tails, which serve as a hub for numerous post-translational modifications such as methylation, acetylation, ubiquitinylation, SUMOylation or phosphorylation. Modified histone tails are major regulators of chromatin structure and function (Bannister and Kouzarides, 2011). For example, histone acetylation on a lysine residue by a histone acetyl transferase (HAT) neutralizes its positive charge and, therefore, weakens the interaction of the histone with DNA, promoting chromatin opening and transcription. Histone deacetylases (HDACs) are in place to reverse this modification (Bannister and Kouzarides, 2011).

Eukaryotes are also equipped with paralogue histone variants, which diverged in their amino acid sequence from the canonical histones in order to fulfil specialized functions in chromatin regulation. In contrast to canonical histones, these variants can be exchanged replication independently (Talbert and Henikoff, 2021). The described *Drosophila* histone variants are BigH1, H3.3, CenH3 (CENP-A or CID) and H2Av (Llorens-Giralt *et al.*, 2021). Since H3.3 and H2Av are relevant for this study they are discussed in further detail below. CENP-A as the main subject of this study is introduced separately in section 1.3.

The histone H3 variant H3.3

The evolutionary conserved histone variant H3.3 replaces a subset of the canonical H3 throughout the genome and is usually present at sites of elevated histone turn-over such as promoters, gene bodies or enhancers of active genes (Ahmad and Henikoff, 2002; Bano *et al.*, 2017; Shi *et al.*, 2017). In addition, H3.3 is detectable at pericentric heterochromatin or telomeres, and chromatin loading of H3.3 is mainly accomplished by its specific chaperones HIRA and DAXX-ATRX (Bano *et al.*, 2017; Drané *et al.*, 2010; Goldberg *et al.*, 2010; Shi *et al.*, 2017). Interestingly, H3.3 was shown to act as a placeholder for the centromeric H3 variant CENP-A. At the time of replication where CENP-A is diluted to both daughter strands H3.3 substitutes CENP-A until its replication uncoupled loading during M- or G_1-phase (cf. section 1.3.3). On another note, CENP-A and H3.3 seem to form heterotypic nucleosomes when CENP-A is overexpressed and ectopically localized (cf. section 1.4.1).

The histone H2A variant H2Av

H2Av (also known as H2A.V) is the only *Drosophila* H2A variant and represents the ortholog to human H2A.Z. *Drosophila* does not seem to contain an ortholog to human H2A.X but H2Av exhibits a phosphorylation motif, which is also conserved in H2A.X. H2Av is, therefore, believed to be a multifunctional H2A variant fulfilling combined tasks of both H2A.Z and H2A.X in *Drosophila* (Baldi and Becker, 2013; Giaimo *et al.*, 2019). H2A.Z/H2Av is found at promoters where it contributes to both transcriptional activation and repression but also at heterochromatic areas promoting heterochromatin formation (Baldi and Becker, 2013; Giaimo *et al.*, 2019). On the other hand, H2Av in its phosphorylated form is found at sites of DNA damage similar to its counterpart H2A.X (Rogakou *et al.*, 1998; Ismail and Hendzel, 2008). H2Av is phosphorylated by ATM and ATR during the DNA damage response and thus its phosphorylated form - known as γH2Av - marks double strand breaks (DSBs; Madigan *et al.*, 2002). Accordingly, γH2Av and γH2A.X are widely used as a DSB markers (Iacovoni *et al.*, 2010; Harpprecht *et al.*, 2019; Scacchetti and Becker, 2020).

1.1.2 Chromatin structure and function

Nucleosome occupancy

At its basis chromatin consists of DNA, which wraps around reoccurring nucleosomes connected by a stretch of linker DNA (Baldi *et al.*, 2020). Hereby, the occupancy of nucleosomes along the DNA is a key regulator for protein accessibility during transcription, replication or repair. Accordingly, active gene elements such as enhancers or promoters are devoid of nucleosomes, whereas at inactive sites the nucleosome pattern is more dense in order to limit DNA accessibility (Lai and Pugh, 2017; Klemm *et al.*, 2019).

Euchromatin and Heterochromatin

Active and inactive chromatin areas are often broadly specified as euchromatin or heterochromatin, respectively. Actively transcribed genes reside in euchromatin, whereas heterochromatin mainly contains inactivated elements such as repetitive satellites, telomere sequences or transposable elements (Murakami, 2013). Moreover, heterochromatin is further classified as 'facultative' or 'constitutive': Constitutive heterochromatin comprises areas such as the pericentromeric region or telomeres, which are stably silenced in all cell types (Allshire and Madhani, 2018). Facultative heterochromatin in contrast is regarded to be reversible, developmentally regulated and/or allele specific and a prominent example for facultative heterochromatin formation is the female mammalian X-chromosome inactivation (Liu *et al.*, 2020; Żylicz and Heard, 2020). Moreover, the developmental regulation of homeotic genes via the Polycomb and Trithorax proteins during metazoan embryo anterior-posterior patterning has been defined by certain authors as the formation of facultative heterochromatin (Kassis *et al.*, 2017; Liu *et al.*, 2020).

Both chromatin classes are characterized by prevalent histone modifications and chromatin binders. Heterochromatin generally lacks histone acetylation and additionally exhibits high levels of H3K9me3 and binding of HP1 in the case of constitutive heterochromatin whereases facultative heterochromatin typically shows H3K27me3 in the presence of Polycomb proteins (Murakami, 2013). Euchromatin on the other hand is defined by tremendous histone acetylation levels and the lack of repressive histone marks such as H3K9 or H3K27 methylation (Murakami, 2013). Instead various other histone methylation types are common such as H3K4me3 at active enhancers or H3K4me1 accumulating at transcriptional start sites, to name just two examples (Bannister and Kouzarides, 2011).

Chromatin 3D organization

In most eukaryotes chromatin fibers are found to cluster into topologically associated domains (TADs). TADs are sub-megabase, spatially adjacent regions of varying sizes

exhibiting frequent interdomain chromatin contacts (Figure 1). They assist the modular approximation or expulsion of regulatory elements (e.g. enhancer) with their target genes. Moreover, TADs are characterized by similar chromatin modifications and function (Sexton *et al.*, 2012; Szabo *et al.*, 2019; Sikorska and Sexton, 2020).

According to the present literature, most TADs consist of chromatin fiber loops, which are anchored by clustering insulators - DNA motifs, that are bound by architectural proteins such as CTCF. In mammals, a model has been favored, where insulator activity is dependent on the loop extrusion activity of Cohesin (Szabo *et al.*, 2019; Sikorska and Sexton, 2020). In *Drosophila* other insulator factors are important such as BEAF-32, M1BP, Chromator or CP190, and insulator activity in general seems to be less dependent on CTCF, rather involving transcription and active promoters (Schwartz and Cavalli, 2017; Szabo *et al.*, 2019; Llorens-Giralt *et al.*, 2021).

At the large scale, chromatin is divided into two multimegabase compartments (Figure 1), of which compartment A contains active, gene-rich regions and compartment B consist of heterochromatic genomic areas, largely devoid of genes (Lieberman-Aiden *et al.*, 2009; Llorens-Giralt *et al.*, 2021). Moreover, interphase chromosomes are not randomly residing in the nucleus but occupy distinct areas termed chromosome territories, which regulated genome organization (Fritz *et al.*, 2019).

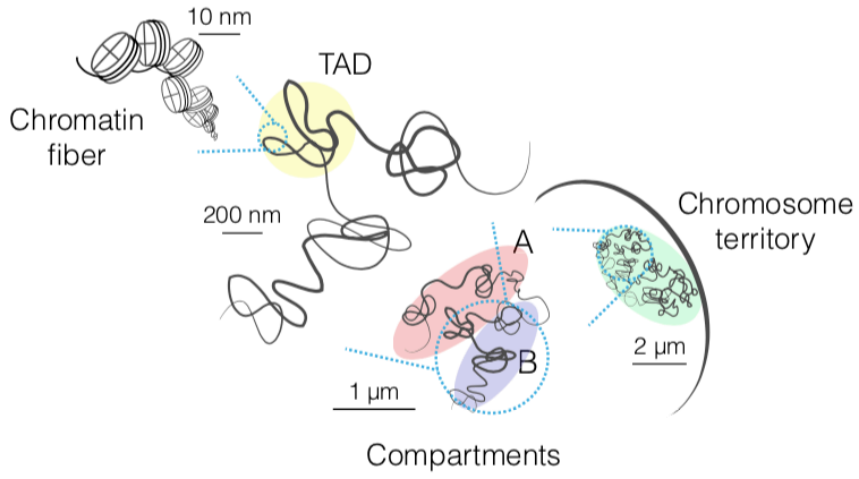


Figure 1 3D organization of the genome

The DNA is wrapped around nucleosomes and is further hierarchically organized into chromatin domains of increasing size and complexity known as TADs, A/B compartments and chromosome territories. From Szabo et al. (2019).

1.1.3 Chromatin remodeling at the heart of defining the nucleosomal composition

Chromatin remodeling is an essential prerequisite to regulate DNA accessibility for protein binding during all types of DNA transactions such as replication, transcription or repair. To avail binding sites for proteins, nucleosomes, therefore, must be slid, stripped, edited or exchanged (Figure 2; Becker and Workman, 2013; Clapier, 2021). Such events, which alter the nucleosome density, positioning and composition are taken over by different classes of chromatin remodeling complexes (CRCs), which utilize the energy from hydrolyzed adenosine triphosphate (ATP) to modulate histone DNA interactions (Clapier, 2021; Magaña-Acosta and Valadez-Graham, 2020).

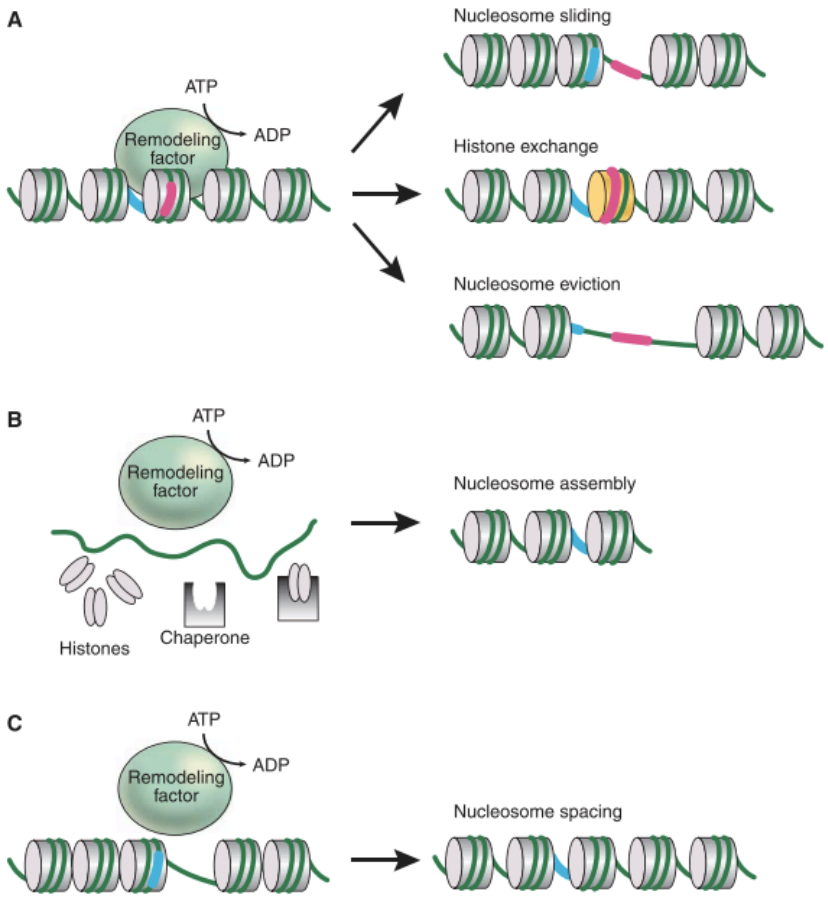


Figure 2 Chromatin remodeling activities

Scheme of chromatin remodeling modes, where remodeling factors use the energy from ATP hydrolysis to alter nucleosome position and occupancy. **(A)** Nucleosome sliding, histone exchange or nucleosome eviction regulates accessibility to regulative sites (pink DNA stretch). **(B)** Pre-nucleosomal histones are assembled onto DNA in co-operation with histone chaperones for example during replication. **(C)** Nucleosomes can be regularly spaced for example after RNA polymerase passage to promote higher order structures. From Becker and Workman (2013).

In eukaryotes four different types of chromatin remodelers exist, which are named and classified according to their ATPase component: CHD (chromodomain-helicase-DNA binding), ISWI (Imitation switch), SWI/SNF (Switch defective/Sucrose non-fermenting) and INO80 (Inositol requiring 80).

Chromatin assembly and maturation upon replication and generally nucleosome sliding and spacing is achieved by ISWI and CHD sub-families. CHRAC-14, which is a central subject of this study, is a subunit of the ISWI CRC CHRAC in *Drosophila* and its roles are discussed in greater detail in section 1.1.5. Of note, CHD remodelers also fulfil further tasks in chromatin opening by histone eviction and nucleosome editing by histone variant exchange (Clapier, 2021).

Moreover, chromatin opening is achieved through SWI/SNF remodelers, which are large macromolecular complexes with numerous accessory subunits participating not only in chromatin opening but also in nucleosome editing. They are important cell-type specific regulators during development involved in gene expression and DNA repair (Magaña-Acosta and Valadez-Graham, 2020; Ribeiro-Silva *et al.*, 2019). 20% of all human cancers exhibit mutations in SWI/SNF genes, highlighting the diverse functional implications for cell viability of this chromatin remodeling sub-family (Reyes *et al.*, 2021; Ribeiro-Silva *et al.*, 2019). A well-studied mode of action of SWI/SNF complexes is the chromatin opening at promoters, where a histone acetyl transferase such as SAGA generates locally acetylated histones, which recruit the chromatin remodeling complex via its bromodomain-containing subunit (Becker and Workman, 2013).

Lastly, INO80 CRCs mainly contribute to change nucleosome composition by the incorporation or eviction of nucleosome variants resulting in functionally specialized chromatin sections (Clapier, 2021). The SWR1 complex for example exchanges H2A-H2B with H2A.Z-H2B (Mizuguchi *et al.*, 2004). Moreover, INO80 promotes CENP-A chromatin assembly in fission yeast (Choi *et al.*, 2017; Singh *et al.*, 2020).

Importantly, CRCs rely on the cooperation with histone chaperones. This protein clade is responsible to bind and deliver histones for scheduled chromatin eviction or loading. (Hammond *et al.*, 2017). Accordingly, histone chaperones enable the regulated migration of newly translated pre-nucleosomal histones from the cytosol to the nucleus by preventing aggregation or non-specific interactions. For example Asf1 (anti-silencing function 1) binds newly synthesized H3-H4 dimers and deposits them for replication-coupled chromatin loading in co-operation with the chaperone CAF1 (chromatin assembly factor 1) , whereas the chaperone FACT (facilitates chromatin transcription) promotes H2A-H2B eviction ahead of the replication fork but also during transcription (Hammond *et al.*, 2017). In heterochromatin, the histone chaperone DAXX (death domain associated protein) allows targeted H3.3-H4

incorporation by associating with the SWI/SNF-like CRC ATRX (X-linked helicase), which specifically binds to H3K9me3 (Hammond *et al.*, 2017).

1.1.4 DNA damage response in the context of chromatin

DNA is continuously assaulted by damaging agents such as intrinsic reactive oxygen species (ROS), replication stress or exogenous irradiation and carcinogens (J. J. Kim *et al.*, 2019). In this study, we use the genotoxic chemical Neocarzinostatin (Edo and Koide, 1997) and an endonuclease to elicit DNA double strand breaks. In order to ensure genome integrity and stability, cells must respond to severe damage with immediate DNA repair and eukaryotes evolved two major pathways to deal with DSBs: Homology directed repair by homologous recombination (HR) and non-homologous end joining (NHEJ) (Kim *et al.*, 2019). In HR, the free DNA ends are resected for the pairing with a homologous sister chromatid strand, which is employed as templates for DNA lesion synthesis. This mechanism hence provides faithful restorage of the DNA but can only occur during S and G_2, when replicated sister chromatids are available (J. J. Kim *et al.*, 2019). In NHEJ the free DNA ends at the break site are stabilized and re-ligated. This pathway is error-prone since previous alterations of the strand are not restored with the help of template-directed DNA synthesis (J. J. Kim *et al.*, 2019). Generally, the DNA damage response comprises complex signaling, chromatin remodeling and DNA repair cascades at the DNA lesion: At the onset of the DNA damage response (Figure 3), the key kinases ATM or ATR are recruited to the DNA lesion. Both kinases phosphorylate H2A.X generating a local DNA damage signal hub. Moreover, ATM/ATR phosphorylate the checkpoint kinases CHK1 and CHK2, which in term initiate cell cycle arrest. Phosphorylated H2A.X (known as γH2A.X) for example mediates the recruitment of the HR complex MRN (MRE11-RAD50-NBS1; Carusillo and Mussolino, 2020).

The binding of the MRN complex to DNA lesions is one of the first steps in DSB HR repair (Figure 3). This is followed by initial DNA end resection via the MRN endonuclease activity and by CtIP, which binds to the coordination factor BRCA1. Long-range end resection is then achieved by the exonucleases EXO1 and DNA2 (Fijen and Rothenberg, 2021; Makharashvili and Paull, 2015). In that way, single strand DNA (ssDNA) stretches are generated, which are protected by the ssDNA binding protein RPA1. Next, homology search for sister chromatid strand invasion is initiated by loading of the recombinase RAD51. This process requires the mediator proteins RAD52, BRCA1, BRCA2 and PALB2 (Li and Heyer, 2008). Eventually, a homologous strand is utilized for repair via polymerases and the lesion is sealed by ligation (Fijen and Rothenberg, 2021). In NHEJ (Figure 3), the free DNA ends are first bound by Ku, which further recruits and activates DNA-PK followed by XLF, XRCC4 and DNA ligase 4

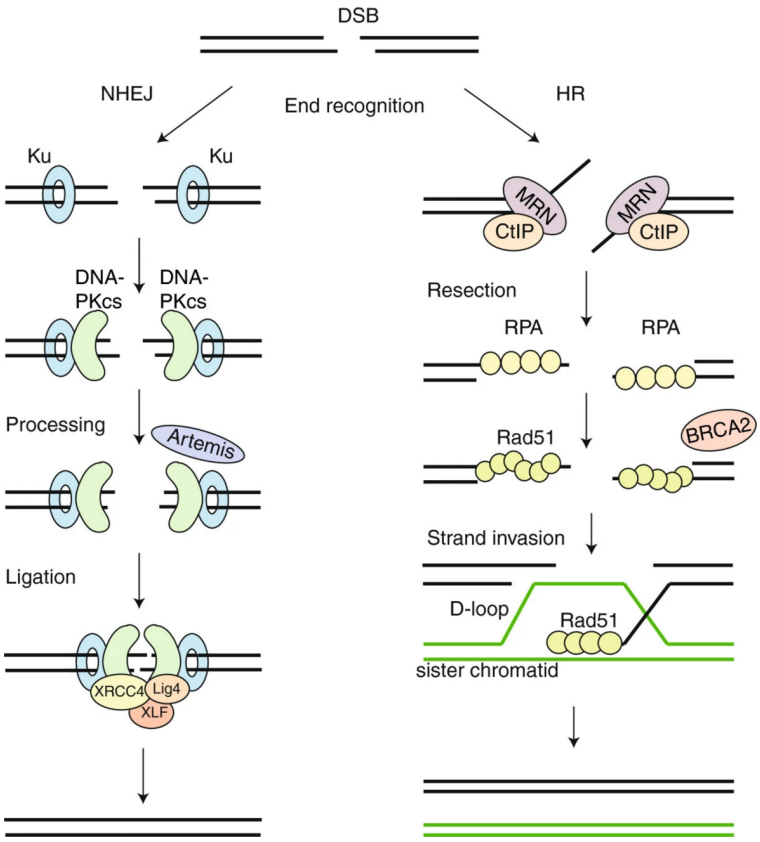


Figure 3 Mechanisms of DSB repair

After DSB induction of chromatin-wrapped DNA the lesion is repaired by two major pathways: Non-Homologous End joining (NHEJ) and Homologous Recombination (HR). From Brandsma and Gent (2012).

binding. The multi-protein complex is stabilized by PAXX and together mediates alignment and ligation of the DNA lesion (Blackford and Jackson, 2017).

In *Drosophila* DSBs are essentially processed similarly, using orthologs of the aforementioned repair proteins, however, other factors such as BRCA1 and Rad52 are missing and it is yet unclear, how *Drosophila* compensates for their functions (Song, 2005; Sekelsky, 2017).

Besides the physical repair of the DNA molecule, chromatin remodeling at the DNA lesion is a prerequisite to decondense chromatin and achieve DNA accessibility for association with the repair machinery and probably to facilitate strand movement for homology search. DSB are sites of increased histone turnover and histone eviction involving INO80 and p400 nucleosome remodeling complexes (Ferrand *et al.*, 2020). Additionally, incorporation of histone variants seems to aid DNA damage signaling and repair (Adam *et al.*, 2015). Interestingly, CENP-A recruitment to laser-induced DNA lesions in human cells and to uncapped telomeres has been observed by our lab and others (Mathew *et al.*, 2014; Zeitlin *et al.*, 2009) but the observation is in conflict with data from another study (Helfricht *et al.*, 2013). Thus, if CENP-A is indeed localizing to DSB and whether it has a functional role is still a matter

of debate and will be address experimentally within this thesis. This topic is discussed in further depth in section 1.4.5.

1.1.5 CHRAC-14 is a versatile factor with implications in DNA damage

ACF1-containing ISWI CRCs

The evolutionary conserved ACF (ATP-dependent chromatin assembly and remodeling factor) and CHRAC (chromatin accessibility complex) are part of the ISWI CRC clade and consist of the ATPase ISWI and ACF1 as the regulative subunit (Figure 4). In contrast to ACF CHRAC additionally associates with two other subunits: CHRAC-14 and CHRAC-16 (Poot *et al.*, 2000; Bouazoune and Brehm, 2006; Aydin *et al.*, 2014). CHRAC-14 and CHRAC-16 facilitate the nucleosome sliding activity within CHRAC and form a heterodimer via their histone fold domains. As a histone analogue they were suggested to bind DNA during chromatin remodeling (Hartlepp *et al.*, 2005; Kukimoto *et al.*, 2004). Moreover, CHRAC-14 and CHRAC-16 can form chimeric complexes with histones (Bolognese *et al.*, 2000; Corona *et al.*, 2000). Functions of CHRAC and ACF include regular spacing and phasing of nucleosomes contributing to global transcriptional repression (Aydin *et al.*, 2014; Baldi *et al.*, 2018; Scacchetti *et al.*, 2018). ACF has additional functions in replication through heterochromatin

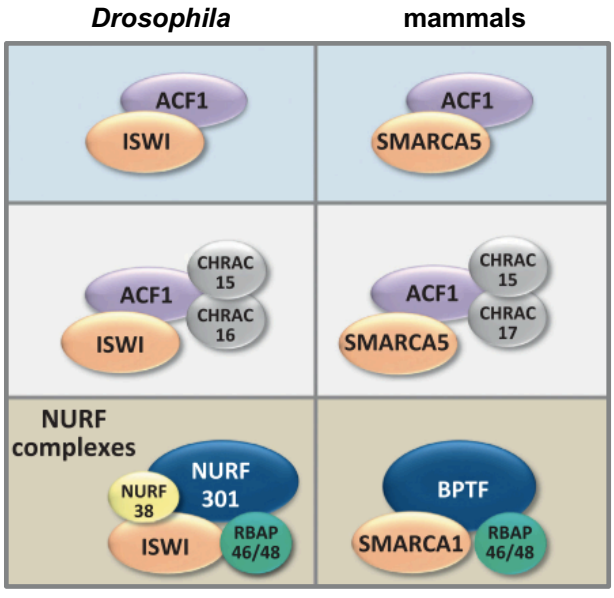


Figure 4 ISWI chromatin remodeling complexes

Depiction of ISWI chromatin remodeling complexes (orthologs are of same colour), which are conserved and present in *Drosophila* and mammals, containing orthologs of the ATPase ISWI. ACF only contains ISWI and ACF1 whereas CHRAC additionally associates with the auxiliary subunits CHRAC-14 and CHRAC-16 in *Drosophila* and their respective orthologs in mammals. The NURF (Nucleosome remodeling factor) complexes are more distinct and contain other subunits than ACF and CHRAC. Adapted from Petty and Pillus (2013).

and transcriptional regulation. Moreover, ACF contributes to chromatin remodeling during DSB repair and is involved in the G2/M checkpoint. (Aydin *et al.*, 2014).

CHRAC-14 participates in various complexes

Interestingly the subunit CHRAC-14 has been found to associate with additional complexes involved in chromatin remodeling. CHRAC-14 seems to interact with the histone chaperone PolE4 (yeast 2-hybrid, Shokri *et al.*, 2019), which is also known as Mes4, and both seem to be part of the DNA Polymerase ε in *Drosophila.* This assumption is based on data in human, showing that their orthologs POLE3 (CHRAC-14) and POLE4 (Mes4) are Pol ε subunits. (Li *et al.*, 2000; Bellelli *et al.*, 2018a, 2018b; Marygold *et al.*, 2020). CHRAC-14, furthermore, CHRAC-14 is a subunit of the histone acetyltransferase ATAC (Ada2a-containing complex) in *Drosophila,* which probably contributes to histone acetylation-linked gene expression (Suganuma *et al.*, 2008; Torres-Zelada and Weake, 2021).

CHRAC-14 in DNA damage repair

In Mathew *et al.* (2014) CHRAC-14 was identified as a factor involved in the DNA damage response in *Drosophila*. This study conducted in our lab demonstrated, that CHRAC-14 depletion via RNA interference (RNAi) in S2 cells and in deficient mutant flies impaired the DNA damage response in several aspects. First, *chrac-14* mutant larvae were hypersensitive to genotoxic stress inflicted by γ-irradiation. As a result, adult animals exhibited severe developmental defects and tumour formation. Cells and larvae, moreover, had a non-functional DNA damage checkpoint, since there was no G2/M arrest upon γ-irradiation. Importantly, ACF1 mutants maintained a functional checkpoint and arrested in G2/M, implicating a CHRAC-independent function of CHRAC-14 during DNA damage. Furthermore, in S2 cells, CHRAC-14 absence led to the accumulation of the DNA damage marker γH2Av and a significant disability of cells to repair drug-induced DNA lesions in a comet assay. Taken together this implied, that CHRAC-14 has a function in DNA repair and checkpoint signaling, which is most likely independent of CHRAC and suggests a cooperation of CHRAC-14 with other factors.

On another important notion, the authors found that CHRAC-14 plays a crucial role to limit CENP-A incorporation during physiological and genotoxic stress conditions, which is further discussed in the CENP-A-related section 1.3.

1.2 The centromere: A specialized chromatin domain

Centromeres are specialized chromatin regions, which are cytologically visible as the primary constriction on each chromosome during mitosis (Flemming, 1882; Wong *et al.*, 2020) (Figure 5). They serve as the underlying platform for kinetochore formation, a proteinaceous macromolecular complex, which binds mitotic microtubules to enable chromosome segregation (Fukagawa and Earnshaw, 2014). Acting at the heart of chromosome segregation by dictating the attachment location for microtubules, centromeres are amongst the most important cellular structures to ensure genome stability and faithful inheritance of the genomic content (Balzano and Giunta, 2020). Importantly, their inheritance and function is epigenetically determined by the centromere-specific H3 variant CENP-A (Allshire and Karpen, 2008).

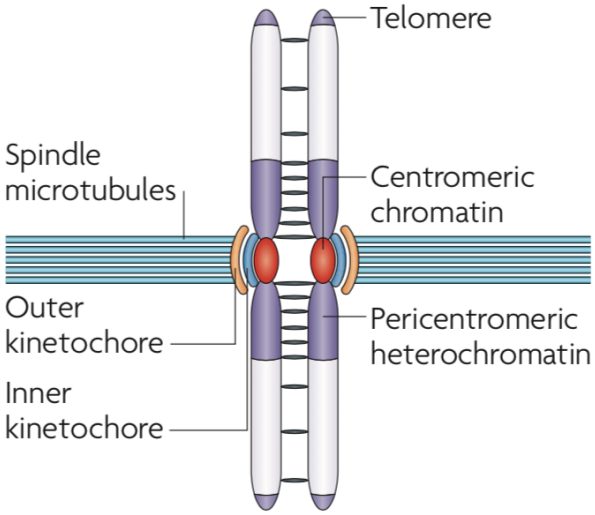


Figure 5 Centromere organization on chromosomes
Centromeres are embedded in flanking pericentromeric heterochromatin and serve as the platform for inner and outer kinetochore assembly and spindle microtubule attachment. From Allshire and Karpen (2008).

1.2.1 Centromere configuration

Despite the deep evolutionary conservation and significance for propagating organisms, centromeric DNA sequences are variable and rapidly evolving, a phenomenon that is also known as the centromere paradox (Arunkumar and Melters, 2020; Henikoff *et al.*, 2001). Nevertheless, the DNA sequence of centromeres exhibits certain characteristics, which are common amongst most eukaryotes, and a major feature is its repetitive nature. In the majority of eukaryotes centromeres consist of large blocks of AT-rich satellite tandem repeats of variable size, structure, and sequence with interspersed transposable elements (TEs; Rosin and Mellone, 2017). Repetitiveness of centromeres most likely developed through mutagenic

processes like replication slippage, recombination, DNA damage repair or transposable element insertion (Balzano and Giunta, 2020).

A second evolutionary similarity of centromeres is that they are flanked by large blocks of pericentromeric heterochromatin, which also contains satellite repeats and is characterized by hypoacetylation, H3K9 methylation and HP1 binding (Allshire and Madhani, 2018; Balzano and Giunta, 2020). Pericentromeric heterochromatin is, furthermore, enriched in Cohesin and is important for sister chromatid cohesion and repression of transcription and centromere rearrangements (Balzano and Giunta, 2020). The composition of centromeric chromatin modifications in turn is distinct from eu- or heterochromatin and varies across species. Many eukaryotes, including human and *Drosophila,* typically exhibit hypoacetylation and H4K4me2, which in human promotes the recruitment of HJURP (Holliday junction recognition protein), the human CAL1 (chromosome alignment defect 1) equivalent, and stimulates kinetochore formation (Wong *et al.*, 2020).

1.2.2 The *Drosophila melanogaster* centromere

Recent advances in the linear assembly of *Drosophila* centromere sequences through third-generation long read sequencing (PacBio and Nanopore technologies) uncovered important details on centromere composition and architecture. Chang *et al.* (2019) reported that core centromeres harbor specific islands of retroelements, each consisting of a specific combination of transposable elements, embedded in AT-rich simple tandem satellite repeats (Figure 6). The *G2/Jockey-3* LINE (long interspersed nuclear element, also known as non-LTR [non-long terminal repeat]) retrotransposon was the only element residing in all centromeres. Besides *G2/Jockey-3,* every centromere exhibited a unique combination of other LINE elements such as *Jockey* family elements, ribosomal IGS or LTR (long terminal repeat) elements like *Gypsy* and *Copia.* An important notion of this study was, that in S2 cells,

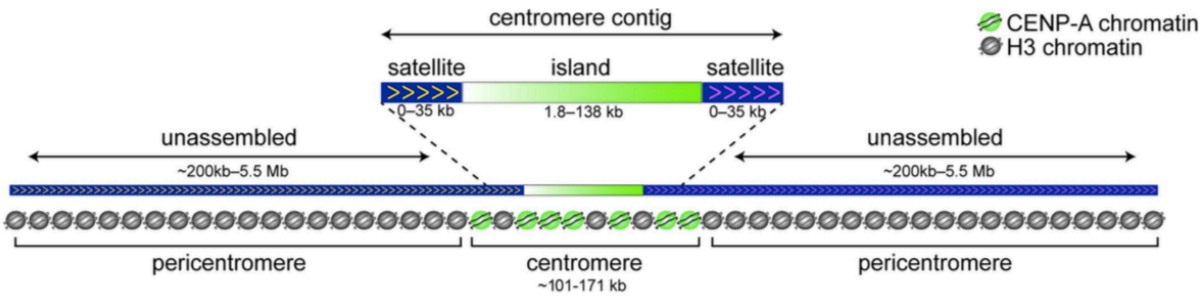


Figure 6 *Drosophila* centromere composition

Model of the current *Drosophila* centromere assembly published by Chang *et al.* (2019), consisting of CENP-A decorated retroelement islands embedded in simple satellite tandem repeats. Up to this point flanking pericentromeric simple satellite repeats remain unassembled due to the consistently low sequence complexity.

additional other transposable elements were observed, some of which are pericentric in embryos (1.688 satellite and *Responder*). As an explanation the authors suggest that centromeres in S2 cells have expanded due to structural rearrangements inherent to cultured *Drosophila* cell lines (Chang *et al.*, 2019b; Lee *et al.*, 2014).

1.3 CENP-A: Epigenetic centromere identifier

Centromere DNA sequence lacks conservation, and it is hence conceivable that centromere identity is regulated epigenetically by the incorporation of the centromere-specific histone H3 variant CENP-A (Allshire and Karpen, 2008). CENP-A is sufficient for centromere formation and essential for proper chromosome segregation, since it serves as the anchor for kinetochore assembly (Takahashi *et al.*, 2000; Blower and Karpen, 2001; Allshire and Karpen, 2008; Mendiburo *et al.*, 2011; Musacchio and Desai, 2017). When overexpressed, CENP-A is ectopically loaded at non-centromeric sites and causes functional kinetochore assembly (Heun *et al.*, 2006), highlighting its function as a key centromere determinant.

1.3.1 CENP-A loading at centromeres

Under physiological conditions, CENP-A is loaded into centromeres independent of replication in late telophase and G_1 in human (Jansen *et al.*, 2007), during anaphase in *Drosophila* syncytial embryos (Schuh *et al.*, 2007) and in metaphase in cultured *Drosophila* S2 cells (Mellone *et al.*, 2011). Specific targeting to centromeres is ensured by the dedicated CENP-A chaperones HJURP in human (Dunleavy *et al.*, 2009; Foltz *et al.*, 2009) and CAL1 in *Drosophila* (Chen *et al.*, 2014). The mechanisms of CENP-A incorporation after the onset of mitosis are still a matter of intense investigation. In human cells, the Mis18 complex is phosphorylated by Plk1 in anaphase and recruits HJURP to centromeres. Beforehand, Mis18 itself becomes dephosphorylated through a drop in CDK activity enabling the binding of Mis18 to CENP-C, a member of the constitutive centromere-associated network (CCAN) (Musacchio and Desai, 2017; Mahlke and Nechemia-Arbely, 2020). Similarly, in *Drosophila*, CAL1 is recruited via CENP-C (Dong *et al.*, 2021). Furthermore, it has been shown, that CAL1, CENP-A and CENP-C are mutually interdependent for centromere location (Erhardt *et al.*, 2008). Maturation of newly deposited CENP-A is then aided by RSF and MgcRacGap (Perpelescu *et al.*, 2009; Lagana *et al.*, 2010; Mahlke and Nechemia-Arbely, 2020).

1.3.2 Centromere transcription encourages CENP-A chromatin in multiple ways

In the past years it has been established that transcription of centromeres during mitosis is crucial for centromere function and CENP-A loading (Arunkumar and Melters, 2020).

Specifically, the transcription complex FACT is a binding partner of CAL1 and is required for transcription-couped CENP-A deposition (Chen *et al.*, 2015). In line with that, the chromatin remodeling factors Hap2-Ino80 promote transcription-coupled CENP-A incorporation (Choi *et al.*, 2017; Singh *et al.*, 2020). Other studies, moreover, directly linked the process of transcription to centromeric CENP-A chromatin maturation (Bobkov *et al.*, 2018; Shukla *et al.*, 2018). On another notion, maintenance of pre-existing CENP-A during transcription is ensured by the histone chaperone Spt6 and by CENP-I (Bobkov *et al.*, 2020; Hirai *et al.*, 2021). Interestingly, whereas transcription enforces CENP-A chromatin it also seems to be a consequence of *de novo* CENP-A seeding at neocentromeres (Murillo-Pineda *et al.*, 2021; Naughton *et al.*, 2021).

In addition to transcription coupled chromatin remodeling also the resulting cenRNA (centromeric RNA) products emerged as important centromere and CENP-A regulators and numerous roles have been suggested: Recruitment and/or stabilization of centromeric factors such as CENP-C or CENP-A, scaffolding, phase separation or signaling (Arunkumar and Melters, 2020; Corless *et al.*, 2020). A recent study in *Drosophila* reported that repetitive RNAs originating from active retrotransposons associate with pericentromeres and aid to maintain heterochromatin via the siRNA pathway and thereby contributing to proper chromosome segregation (Hao *et al.*, 2020).

1.3.3 **CENP-A is inherited by maintenance**

During replication, the CENP-A nucleosome count at centromeres is halved, as existing CENP-A is distributed between the daughter chromatids and replenished only after the onset of mitosis. Until then, empty slots are filled with the placeholder histone H3.3 (Dunleavy *et al.*, 2011; Mahlke and Nechemia-Arbely, 2020). The recycling of ancestral CENP-A behind the replication fork is accomplished by HJURP together with the helicase complex subunit MCM2 probably involving the H3-H4 chaperone CAF1 (Zasadzińska *et al.*, 2018; Mahlke and Nechemia-Arbely, 2020). Moreover, the SUMO protease SENP6 is required to maintain CENP-A throughout the cell cycle and its loss leads to hyper-SUMOylation of CENP-A-stabilizing factors and their evasion from the centromere (Fu *et al.*, 2019; Liebelt *et al.*, 2019; Mitra *et al.*, 2020). In yeast, SUMOylation of CENP-A promotes its centromeric deposition (Ohku *et al.*, 2020). In *Drosophila*, our lab established the importance of CUL3-RDX-mediated CENP-A ubiquitylation for centromere targeting and maintenance (Bade *et al.*, 2014). Also in human CENP-A maintenance is promoted by lysine 124 ubiquitinylation via CUL4A-RBX1-COPS8 (Niikura *et al.*, 2015) and later the authors propose a model where ubiquitinylated CENP-A dimerizes with nascent CENP-A for loading and inheritance (Niikura *et al.*, 2019, 2016).

1.4 Ectopic CENP-A as a genome hazard

CENP-A is highly enriched at centromeres administrating its primary function as an epigenetic centromere marker. Surprisingly, it has been shown, that endogenous CENP-A is also distributed throughout the genome at basal levels in yeast, *Drosophila* and human (Bodor *et al.*, 2014; Chang *et al.*, 2019b; Nechemia-Arbely *et al.*, 2019; Shukla *et al.*, 2018) but functional details remain elusive.

When overexpressed, large amounts of CENP-A localize outside of centromeres, representing a serious hazard for genome integrity (Shrestha *et al.*, 2017) because ectopic CENP-A enrichment potentially leads to neocentromere formation and segregation defects (Mendiburo *et al.*, 2011; Murillo-Pineda *et al.*, 2021; DeBose-Scarlett and Sullivan, 2021). Therefore, it is not surprising, that CENP-A overexpression and/or mislocalization is a hallmark of many cancer types (Amato *et al.*, 2009; Li *et al.*, 2011; Qiu *et al.*, 2013; Saha *et al.*, 2020; Tomonaga *et al.*, 2003; Wu *et al.*, 2012; Zhang *et al.*, 2016, 2020) and that nowadays, CENP-A is used as a marker for breast cancer (Cardoso *et al.*, 2016). Given the deleterious effects of mislocalized CENP-A, mechanisms and features of ectopic CENP-A loading and chromatin incorporation are a matter of extensive investigation.

1.4.1 Ectopic CENP-A loading factors

When overexpressed CENP-A seems to bypass dedicated loading pathways and hijacks loading machineries of other histones. It has been shown that in human ectopic loading of overexpressed Flag-HA-tagged CENP-A is dependent on the H3.3 chaperone DAXX and that CENP-A forms a heterotypic tetramer with H4 and H3.3-H4 (Lacoste *et al.*, 2014; Arimura *et al.*, 2014). In line with that, the deposition of innately overexpressed CENP-A in colorectal cancer cells was accomplished via ATRX and DAXX (Athwal *et al.*, 2015). In colon cancer cells CENP-A ectopic targeting was furthermore dependent on the second H3.3 chaperone HIRA and antagonized by HJURP and, in addition, cells exhibited increased DNA damage rates (Nye *et al.*, 2018). Moreover, the histone H3-H4 chaperone CAF1 seems to aid ectopic CENP-A loading in yeast (Hewawasam *et al.*, 2018). In *Drosophila* our lab revealed that overexpressed CENP-A-GFP is loaded via the NuRD (Nucleosome remodeling and deacetylase) chromatin remodeling complex (Demirdizen *et al.*, 2019).

1.4.2 Preferred sites of ectopic CENP-A loading

In studies using chromatin immunoprecipitation followed by sequencing (ChIP-Seq) CENP-A was predominantly detected at sites exhibiting high histone turnover rates such as transcriptional start sites, regulative sites or loci of CTCF binding (Lacoste *et al.*, 2014; Athwal

et al., 2015; Nechemia-Arbely *et al.*, 2019; Mahlke and Nechemia-Arbely, 2020). In studies focusing on neocentromere formation, CENP-A binding sites coincided with transposable elements (Barry *et al.*, 1999; Chueh *et al.*, 2005; DeBose-Scarlett and Sullivan, 2021; Lo, 2001; Lo *et al.*, 2001) or were near telomeric regions (Ketel *et al.*, 2009; DeBose-Scarlett and Sullivan, 2021). Sub-telomeres were also bound by ectopic CENP-A in human cancer cells (Athwal *et al.*, 2015). Neocentromere CENP-A seeding in general seems to favor heterochromatic regions in different organisms (Ishii *et al.*, 2008; Olszak *et al.*, 2011; DeBose-Scarlett and Sullivan, 2021; Murillo-Pineda *et al.*, 2021).

1.4.3 **Consequences of ectopic CENP-A localization**

As aforementioned, CENP-A overexpression and subsequent ectopic localization directly leads to chromosome instability due to segregation errors (cf. section 1.4). CENP-A ectopic localization might furthermore alter chromatin structure, since it was shown to bind at CTCF sites, suspending the insulator protein (Lacoste *et al.*, 2014). Altered chromatin structure could lead to misexpression of genes and indeed, recent studies reported that ectopic CENP-A causes misexpression of surrounding genes and transcriptional activation at neocentromeres (Jeffery *et al.*, 2020; Lacoste *et al.*, 2014; Murillo-Pineda *et al.*, 2021; Naughton *et al.*, 2021; Saha *et al.*, 2020; Shrestha *et al.*, 2021) and promiscuous gene expression can be deleterious to organisms and cells (Lee and Young, 2013). Importantly, CENP-A is not only localizing outside of the centromere when overexpressed or upon programmed neocentromere formation but was detected at basal levels genome-wide in yeast and human (cf. section 1.4). Current speculations favor a scenario where low-level, ectopic CENP-A accounts for the formation of new centromeres in case of centromere loss by faulty mitosis or DNA breaks contributing to karyotype evolution (Dong *et al.*, 2021).

1.4.4 **Mechanisms counteracting ectopic CENP-A**

It was recently published that replication seems to act as an error correction mechanism during S-Phase in HeLa cells, removing non-centromeric CENP-A without subsequent recycling (Nechemia-Arbely *et al.*, 2019). Mechanistic details regarding the removal of CENP-A are unknown but the authors suggest that replication-coupled CENP-A eviction is followed by proteolysis (Nechemia-Arbely *et al.*, 2019).

Indeed, several studies (Dong *et al.*, 2021) linked the regulation of ectopic CENP-A to ubiquitinylation and proteasomal degradation. For example, CENP-A levels in *Drosophila* are restricted with the help of the E3-ligases APC/C^{Cdh1} and SCF^{Ppa} (Huang *et al.*, 2019; Moreno-Moreno *et al.*, 2019). The study by Huang *et al.* moreover showed that phosphorylation on Serine 20 (S20) by Casein kinase 2 (CK2 or CKII) triggers ectopic CENP-A removal via SCF^{Ppa}

and the proteasome. Data from both studies indicate, that pre-nucleosomal and centromeric association with CAL1 seems to shield CENP-A from being targeted for degradation at the centromere, only rendering unprotected, ectopic CENP-A for removal. The SCF ligase also controls CENP-A levels in unperturbed yeast cells and prevents mislocalization together with Met30a and Cdc4 (Au *et al.*, 2020).

Furthermore, proteasomal degradation of ectopic CENP-A has been connected to transcription-coupled chromatin remodeling in yeast, where the FACT subunit Spt16 mediates an interaction of CENP-A to the E3-ligase Psh1, promoting CENP-A ubiquitinylation and degradation. (Deyter and Biggins, 2014). The authors suggest proline isomerization, which is needed for CENP-A to Psh1 interaction (Ohkuni *et al.*, 2014), as a regulative switch to protect centromeric CENP-A from degradation. Also HIR, the yeast ortholog of the histone H3.3 chaperone HIRA, was shown to facilitate CENP-A to Psh1 binding to mediate proteolysis and prevent ectopic loading (Ciftci-Yilmaz *et al.*, 2018). Of note, the role of Psh1 in CENP-A degradation is well established (Dong *et al.*, 2021). An important notion is that Psh1 mediated CENP-A proteolysis was shown to be dependent on CENP-A deubiquitylation (Canzonetta *et al.*, 2016).

SUMOylation of CENP-A is another mechanism in yeast to target CENP-A for ubiquitin mediated degradation (Ohkuni *et al.*, 2018, 2016). In animals, SUMOylation of CENP-A has not yet been reported (Srivastava and Foltz, 2018; Dong *et al.*, 2021).

Regarding ectopic CENP-A regulation through chromatin remodeling, depletion of the yeast SWI/SNF chromatin remodeling complex subunit Snf2 lead to ectopic CENP-A localization (Gkikopoulos *et al.*, 2011). Data from our lab has linked the histone chaperone CHRAC-14, a subunit of CHRAC, ATAC and DNA Polymerase ε (cf. section1.1.5) to ectopic CENP-A localization (Mathew *et al.*, 2014), a phenotype which is described in further detail in the following section.

1.4.5 **CHRAC-14 regulates CENP-A in DNA damage**

As mentioned in section 1.1.5 Mathew *et al.* introduced a new role for the histone chaperone CHRAC-14 in DNA damage regulation. A second striking finding was, that CHRAC-14 regulates CENP-A localization. In their study they show that CHRAC-14 post-transcriptional silencing in S2 cells or *chrac-14* gene interruption by P-element insertion in mutant embryos leads to an increase in CENP-A protein levels and to the formation of additional CENP-A foci in interphase nuclei and on mitotic chromosome spreads. This was accompanied by the appearance of dicentric chromosomes and chromosome segregation defects.

When CHRAC-14 depletion was combined with targeted DNA damage induction by telomere uncapping through co-depletion of the telomere capping protein HIPHOP, CENP-A

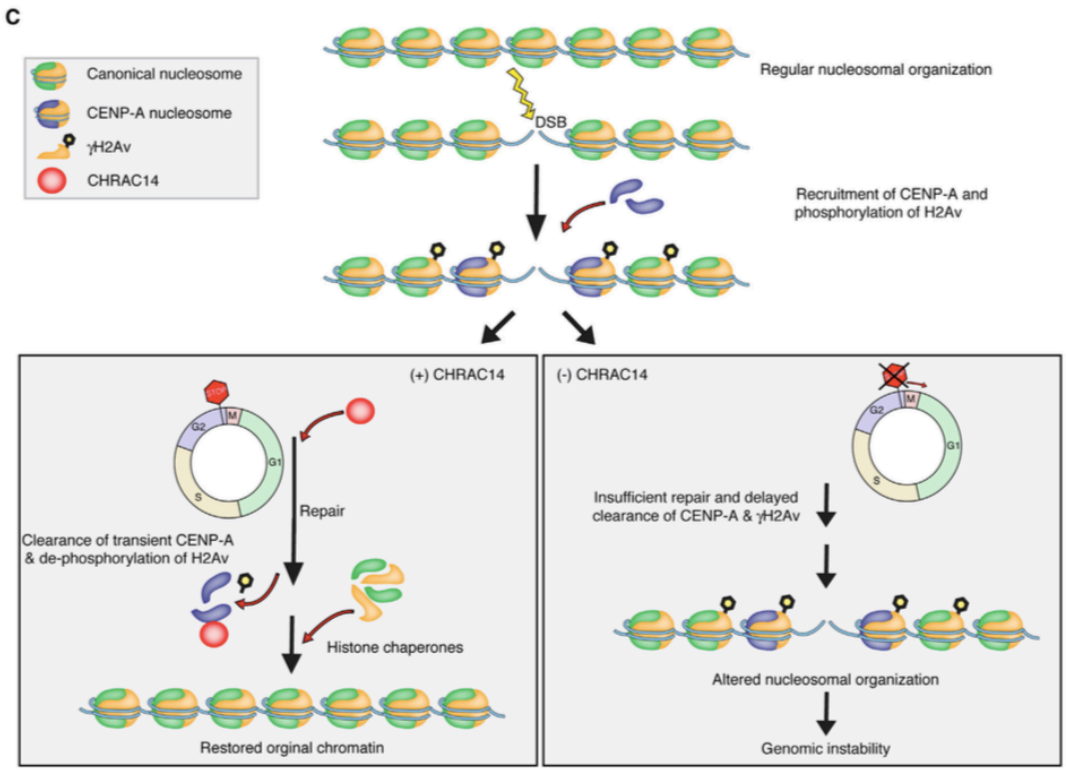


Figure 7 CHRAC-14 dependent CENP-A regulation in DNA damage

Working model of CHRAC-14 dependent CENP-A regulation in *Drosophila melanogaster*. As a histone chaperone CHRAC-14 safeguards CENP-A loading at DNA double strand breaks (DSB) during DNA repair ensuring the restoration of the original chromatin state. CHRAC-14 absence leads to the accumulation of CENP-A and γH2Av at DSBs resulting in malfunctioning DNA repair and defective DNA damage checkpoint.

was detected at telomeres – the sites of DNA damage – on mitotic chromosome spreads. Furthermore, chromosomes in the double knockdown did not show signs of DNA repair by non-homologous end joining, since in contrast to HIPHOP depletion, chromosomes were not fused at the telomeres. In interphase cells, which were subjected to drug-induced DNA damage, CENP-A localized near γH2Av marked DNA lesions in CHRAC-14 absence, strongly suggesting that CENP-A is regulated by CHRAC-14 upon DNA damage. This finding was fortified by the observation, that recombinant CENP-A and CHRAC-14 interact *in vitro*. *In vivo* both proteins show an interaction after CENP-A-GFP immunoprecipitation from extracts of irradiated embryos. Taken together, these observations lead to the following working model (Figure 7): As reported previously upon DNA damage induction (Zeitlin *et al.*, 2009), CENP-A is recruited to DSB or the vicinity thereof, where CHRAC-14 regulates its loading. (Since ectopic CENP-A was also observed without experimental DNA damage induction, CHRAC-14 might regulate its loading under physiological conditions and/or at sites of intrinsic DNA damage). Additionally, CHRAC-14 absence leads to DNA damage hypersensitivity due to inefficient repair and to a defective DNA damage checkpoint. This might be caused by promiscuous CENP-A enrichment at the DSB leading to alterations in the adjacent chromatin

environment, neocentromeres and chromosome segregation defects. However, detailed mechanistic insights and the precise localization of CENP-A in CHRAC14 depleted cells and mutant embryos is still unknown and may give new insights into how CENP-A misexpression alters gene expression, chromosome segregation and the DNA damage response pathway

2 Aim Of Thesis

Based on previous work in our lab, CENP-A chromatin incorporation in *Drosophila melanogaster* is regulated by CHRAC-14 and may be linked to the DNA damage response (Mathew *et al.*, 2014).

CENP-A recruitment to DSB remains a matter of debate due to conflicting results (Helfricht *et al.*, 2013; Mathew *et al.*, 2014; Zeitlin *et al.*, 2009) but clearly, CHRAC-14 emerged as an important regulator of CENP-A, as the absence of CHRAC-14 leads to increased CENP-A protein levels and ectopic CENP-A incorporation. As summarized in section 1.4, excess CENP-A and ectopic loading are deleterious for organisms and are hallmarks of many cancer types.

CHRAC-14 also arose as a DNA damage factor but since mechanistic details are yet to be uncovered, it is unknown if the DNA damage function of CHRAC-14 is linked to the phenotype of ectopic CENP-A after CHRAC-14 depletion. Conclusively, CHRAC-14 regulates two delicate pathways (DNA damage response and CENP-A ectopic loading) and failure within either of them is associated with disease and cancer. Conceivably, CHRAC-14 is a key factor to maintain genome integrity and prevent cancer development.

Therefore, it is of great relevance to mechanistically advance our present working model and to precisely understand, how ectopic CENP-A is regulated by CHRAC-14 and how the proteins are involved in the DNA damage response. This study thus aims, to target following open questions:

1. Is endogenous CENP-A recruited to DNA lesions?
2. How does CHRAC-14 modulate CENP-A at DNA lesions?
3. What are the co-operating factors for CHRAC-14-dependent CENP-A regulation?

3 Results

3.1 Identification of common CENP-A and CHRAC-14 interaction partners

Even though there is a clear correlation between CENP-A ectopic localization and depletion or mutation of CHRAC-14, it remains to be determined how CHRAC-14 influences CENP-A localization mechanistically during physiological and genotoxic stress conditions. We propose a model where CHRAC-14 acts as a histone adapter and mediates CENP-A interaction in combination with a chromatin remodeling complex leading to altered CENP-A loading. Alternatively, CHRAC-14 may bridge an interaction of CENP-A with another regulative entity such as a kinase or E3-ligase to initiate posttranslational modifications.

As a first step we, therefore, screened for common interaction partners of CHRAC-14 and CENP-A by immunoprecipitation followed by mass spectrometry in *Drosophila* S2 cells. To determine if a complex formation of CENP-A and CHRAC-14 with other proteins is DNA damage dependent, we additionally induced acute DNA breaks using the ionizing radiation imitating drug Neocarzinostatin (Edo and Koide, 1997). I performed duplicate anti-V5 immunoprecipitations in cell lines stably transfected with C-terminally V5-His tagged CENP-A or CHRAC-14. Anti-IgG immunoprecipitations were used as a control for unspecific binding and as a reference for differential binding analysis. Protein expression from the metallothionein promoter (pMT, Bunch *et al.*, 1988) was induced with copper sulfate ($CuSO_4$).

To prevent promiscuous localization of the proteins due to immoderate gene expression, an appropriate $CuSO_4$ concentration was determined in the beginning. V5 immunostaining of mitotic spreads from cells treated with increasing $CuSO_4$ concentrations for 10 h revealed that 10 µM $CuSO_4$ causes CENP-A-V5-His to incorporate ectopically into chromosome arms. At lower concentrations, CENP-A-V5-His was only detected at the centromeres (Figure 8 A). In line with that, protein levels increased acutely at 10 µM $CuSO_4$ induction resembling an overexpression scenario (Figure 8 C). To balance both the production of sufficient protein amounts for mass spec analysis and proper protein localization, induction at 1 µM $CuSO_4$ over 10 h was selected as an appropriate condition. This concentration was also used for CHRAC-14-V5-His cells yielding an adequate amount of protein as judged by Western blotting (Figure 8 C) and immunofluorescence (IF; Figure 8 B). Of note, since both cell lines exhibited V5 signals without induction, protein expression from the pMT promotor is leaky.

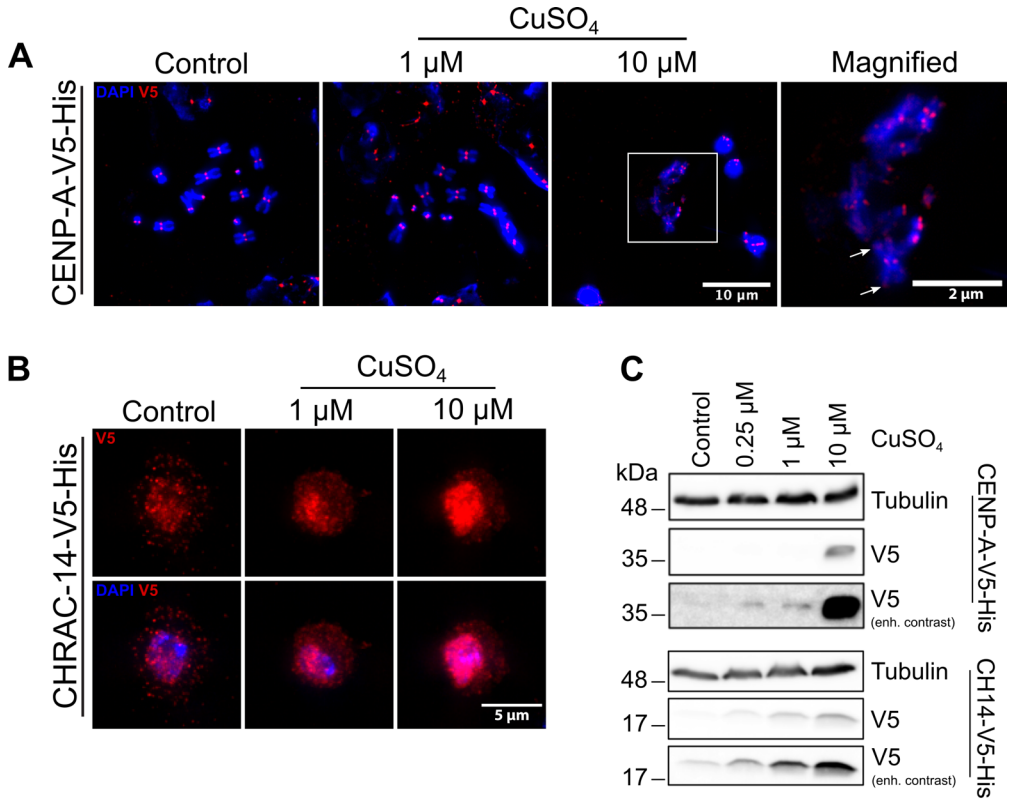


Figure 8 $CuSO_4$ titration in pMT-V5-His cell lines

In *Drosophila* S2 cells, stably transfected with indicated constructs, expression of exogenous proteins was titrated with increasing $CuSO_4$ concentrations. Protein localization and levels were analyzed by immunofluorescence widefield microscopy. **(A)** Mitotic spreads from CENP-A-V5-His cell line, stained with V5 antibody and DAPI **(B)** Normally cycling, fixed cells from CHRAC-14-V5-His cell line, stained with V5 antibody and DAPI. **(C)** Exogenous protein levels evaluated in whole cell lysates of the respective cell lines. Microscopy images are z-projections. CH14 = CHRAC-14.

Having determined a suitable expression condition, the immunoprecipitation experiments were conducted next, and the efficiency of each pulldown was assessed. As reported in Figure 9 B and E the tagged proteins were specifically and efficiently enriched in the V5 immunoprecipitations compared to the control IgG pulldowns. Noteworthy, the Western blot transfers for CHRAC-14-V5-His samples (Figure 9 D-F) were technically problematic, leading to poor signals, especially of the IP bands (Figure 9 E). In the CHRAC experiment, I additionally tested the efficiency of the large-scale cell lysis (8 x 10^8 cells) and analyzed the protein content in the debris-pellet (Figure 9 D lane 3). Obviously, cell rupture was incomplete, and a significant amount of protein remained within pelleted, non-lysed cells.

Importantly, successful DNA damage induction was verified by observing an increase of the DNA damage chromatin marker γH2Av via Western blotting as shown in Figure 9 C and F for both cell lines.

Despite of an incomplete lysis, the immunoprecipitations in each cell line and replicate were highly specific and efficient and thus the IP eluates were measured by quantitative tandem mass spectrometry and documented at the EMBL proteomics facility. With their

analysis pipeline, they normalized the detected peptides to the IgG content and scored them according to the interaction significance between replicates and relative to IgG using statistical tests.

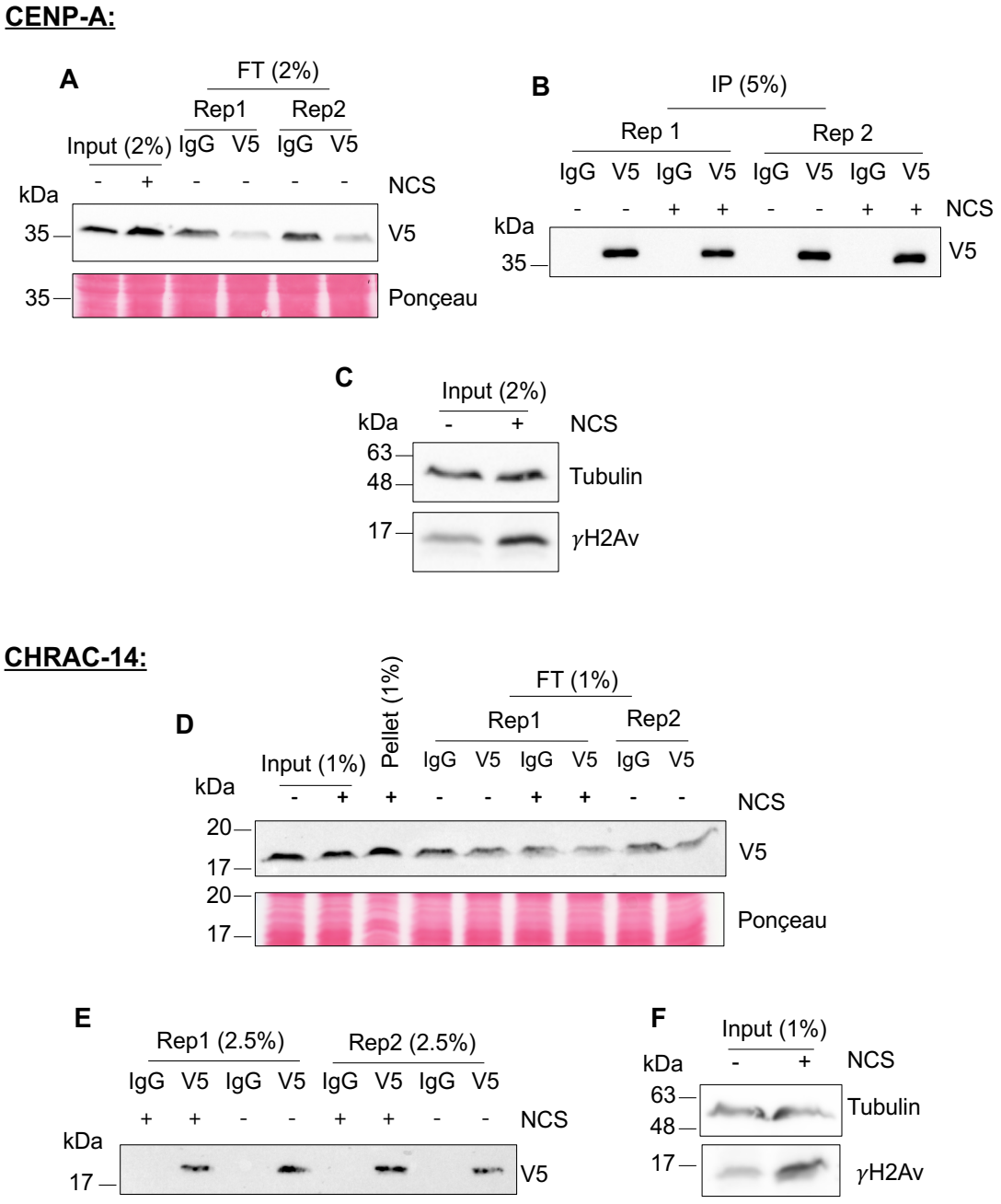


Figure 9 DNA damage immunoprecipitations for mass spectrometry in S2 cells

Mild expression of proteins was induced with 1 µM $CuSO_4$ over 10 h followed by a DNA damage treatment with NCS for 1 h and cell lysis; as a control cells were left untreated. IP and DNA damage efficiency were monitored by Western blotting. **(A)** CENP-A-V5-His input samples and IgG and V5 flow through samples of indicated replicates and conditions **(B)** CENP-A-V5-His IgG and V5 IP samples from all conditions and replicates **(C)** DNA damage induction efficiency in input samples, monitored by immunoblotting for γH2Av **(D)** CHRAC-14-V5-His input samples, one exemplary debris-pellet to assess lysis efficiency, and IgG and V5 flow through samples of indicated replicates and conditions. **(E)** CHRAC-14-V5-His IgG and IP samples from all conditions and replicates **(F)** DNA damage induction efficiency in input samples, monitored by immunoblotting for γH2Av. IP = immunoprecipitation, FT = flow through, NCS = Neocarzinostatin.

3.1.1 Interaction candidates of CHRAC-14 and CENP-A

In the following, detected interaction candidates of CENP-A and CHRAC-14 are described mentioning their main functions. This information was gathered via the protein knowledgebase uniport unless stated otherwise (Bateman *et al.*, 2021).

The strongest CENP-A and CHRAC-14 hits (Figure 10 A and Supplemental Table 1 for a complete list of significant and condition-specific interaction candidates) identified in our mass spec analysis are well-characterized interactors, such as all canonical histones for CENP-A (cf. section 1.3) or CHRAC-16 for CHRAC-14 (cf. section 1.1.5). These interaction partners are robustly present in both conditions, clearly indicating that the identification of interacting proteins was successful. Importantly, CHRAC-14 was identified as a strong interactor of CENP-A in both conditions, confirming the interaction between both proteins by mass spectrometry in S2 cells. Previously, an *in vivo* interaction was solely based on immunoprecipitation and Western blotting from *Drosophila* embryo extracts (Mathew *et al.*, 2014).

Interestingly, The CHRAC-14 interactome in this experiment seemed to be highly specific for DNA damage, since the majority of interactions were detected upon DNA damage induction (Figure 10 B), including interactors that are involved in the DNA damage response in *Drosophila* such as RPA3, and Cdk5. This confirms our initial observations where CHRAC-14 plays a role in DNA damage (Mathew *et al.*, 2014).

Intriguingly, Mes4, also known as PolE4 – the putative ortholog of human POLE4 – and subunit of human DNA polymerase ε (cf. section 1.1.5) was detected in both conditions as the strongest CHRAC-14 interaction candidate (Figure 10 A). Hitherto, the interaction of the *Drosophila* pol ε subunits Mes4 and CHRAC-14 were only observed via a yeast 2-hybrid assay (Shokri *et al.*, 2019) and the strong interaction detected by mass spectrometry in this study confirms a genuine interaction *in vivo*.

Furthermore, upon genotoxic stress, CHRAC-14 also interacted with other relevant proteins such as the SUMO protease Velo, the ubiquitin E3-ligase Bre1, the cohesion subunits Smc1 and Smc5 and the insulator factor M1BP, with transcriptional elongator subunits (TFIIA, TFIIS) and with subunits of the chromatin remodeling complexes pBAF (polybromo, Bap60, Bap111) and ATRX (XNP in *Drosophila*). In the control condition a further candidate was the repair factor Rad23.

In the case of CENP-A, the majority of candidates interact independent of genotoxic stress (Figure 10 B). It is to note, that typical centromeric interaction partners such as CAL1 or CENP-C are lacking in the CENP-A IP. However, the presence of INCENP in the control condition, an inner centromere protein of the chromosomal passenger complex (Adams *et al.*, 2001), strongly suggests that I was able to pull down centromeric CENP-A.

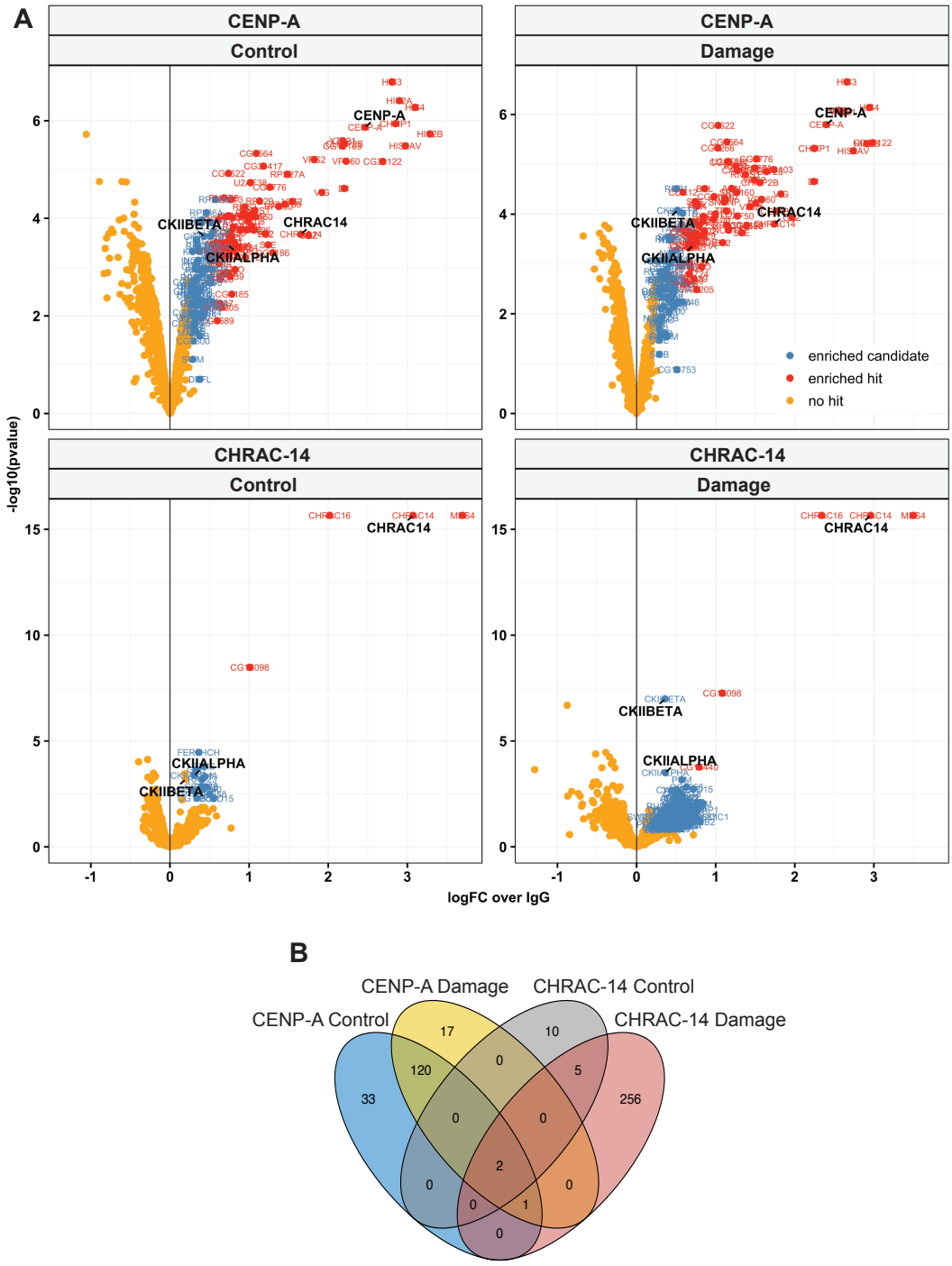


Figure 10 Overview of interaction candidates identified via mass spectrometry

(A) Volcano plots depicting interaction candidates of CENP-A-V5-His and CHRAC-14-V5-His identified by immunoprecipitation and mass spectrometry analysis in control (- NCS) and DNA damage (+ NCS). V5 data was normalized to the IgG background signals and candidates were scored for their interaction significance and strength with statistical tests (no hit, enriched candidate, enriched hit). In both conditions CHRAC-14 was detected as a CENP-A interaction partner as well as both subunits of Casein kinase 2 (CKIIALPHA = CK2α, CKIIBETA = CK2β). CK2α also interacts with CHRAC-14 in both conditions and CK2β switches from no hit to enriched candidate upon DNA damage. **(B)** Venn diagram illustrating the number of detected candidates in each condition and their overlap. CHRAC-14 and CK2α are present in all pulldowns and CK2β is only overlapping between CENP-A ± NCS and CHRAC-14 + NCS. FC = fold change, NCS = Neocarzinostatin.

Intriguingly, also for CENP-A there was a set of DNA damage-specific binding partners including the E3 ligase Snama and Fmr1, interactions, which have been observed in previous mass spec analyses in our lab (Snama: unpublished data; Fmr1 in Demirdizen *et al.*, 2019). Relevant interaction candidates common in both conditions were SUMO (Smt3 in *Drosophila*), the insulator factors Ibf1 and Ibf2, the histone chaperone DEK, the pericentromeric protein D1, the RNA-binding protein Rump, the repair endonuclease Rrp1 and the DNA damage response factor PNUTS.

Most importantly, a common interaction candidate could be discovered: Both subunits of protein kinase 2 (CK2α and CK2β) emerged as interaction partners of CENP-A and CHRAC-14 (Figure 10 A and B; Supplemental Table 1). CK2α interacted with both proteins independent of the damage treatment. Intriguingly, CK2β, the regulatory subunit of CK2 (Bandyopadhyay *et al.*, 2016), interacted with CHRAC -14 upon DNA damage more significantly.

Taken together, the mass spec analysis of CENP-A and CHRAC-14 yielded known interaction partners for both proteins and uncovered new, relevant interaction candidates. In both cases, these new interaction candidates have roles in insulator activity, protein post translational modification, chromatin remodeling and DNA damage response. Noteworthy, Mes4 could be confirmed as a strong *in vivo* interaction partner of CHRAC-14.

Above all, according to our initial aim, CK2 was revealed as the common interaction partner of CENP-A and CHRAC-14. CK2 as a kinase has the potential to regulate CENP-A by phosphorylation and is an established DNA damage factor, being involved in chromatin remodeling during DNA repair (reviewed in Montenarh, 2016). Therefore, I decided to investigate its role in CHRAC-14-dependent CENP-A regulation.

3.1.2 **Validation of CK2 interaction with CENP-A and CHRAC-14 pending**

To verify that CK2β is a genuine interaction partner of both CENP-A and CHRAC-14 and to test if DNA damage induction indeed promotes an interaction between CK2β and CHRAC-14, I established stable S2 cell lines transfected with pMT-CK2β-V5-His. Using these cells, I performed reciprocal V5 immunoprecipitations using similar pulldown and DNA damage conditions as previously in the mass spec experiments. With several repeats of this approach, I could not show an interaction with either of the proteins (data not shown). A potential reason for not detecting CHRAC-14 in the CK2β-V5-His eluates may be the low sensitivity of the CHRAC-14 antibody. I, therefore, set up an additional cell line transfected with both, pMT-CHRAC-14-Flag and pMT-CK2β-V5-His for co-expression of both proteins and to facilitate the detection of Flag-tagged CHRAC-14 with a Flag antibody instead of the CHRAC-14 antibody. However, the outcome remained unchanged and CHRAC-14-Flag was not detected - neither with the Flag antibody nor with the CHRAC-14 antibody (data not shown). Of note, I attempted

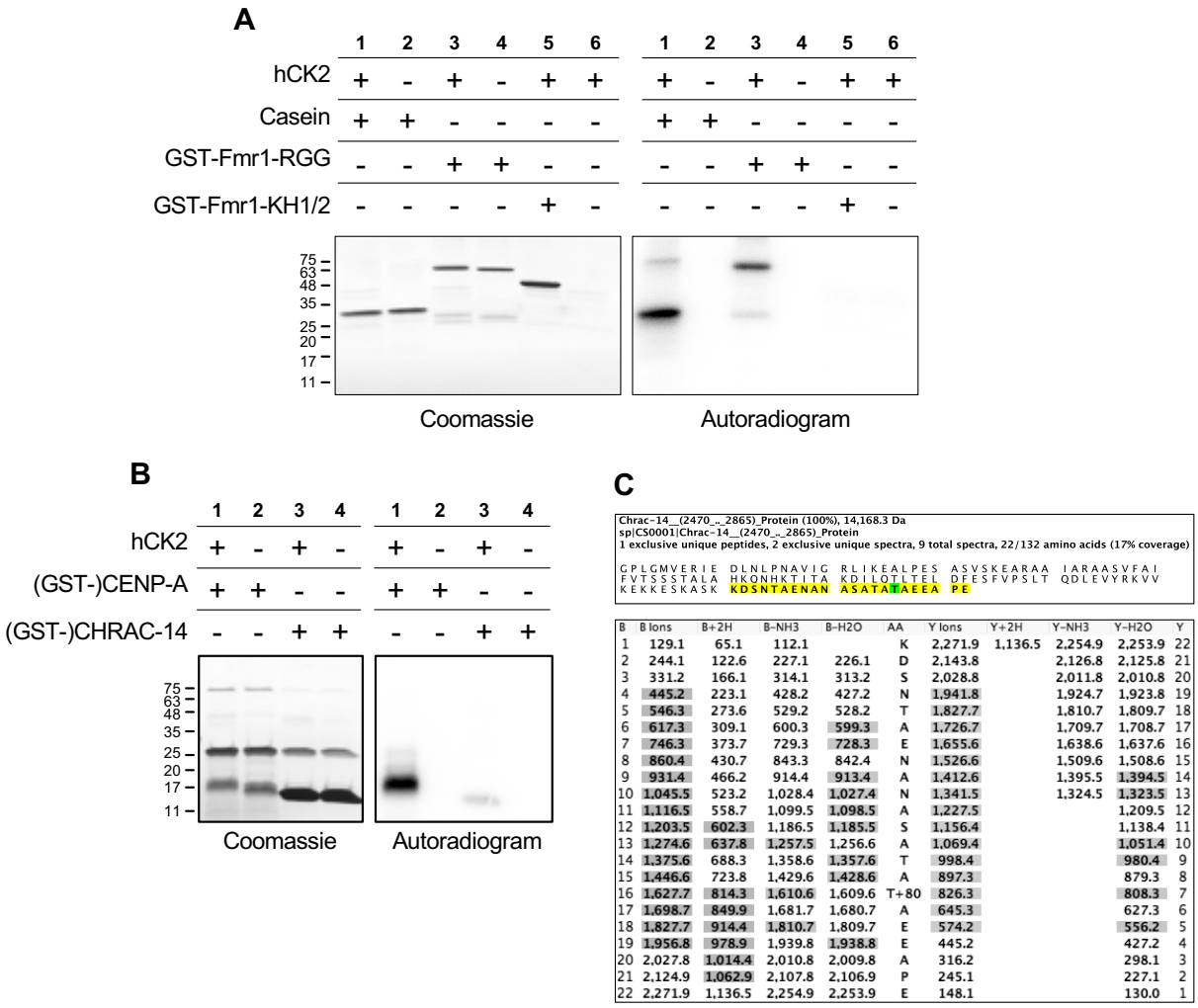
to unsuccessfully optimize the immunoprecipitation conditions by increasing the cell number, by using more antibody for the pulldown and by using non-covalently coupled beads.

3.1.3 CK2 phosphorylates CENP-A and CHRAC-14 *in vitro*

Next to the attempt of validating an interaction of CK2β with CENP-A and CHRAC-14 I aimed to establish a functional link between the proteins. Since CK2 is a kinase we speculated that CHRAC-14 and CENP-A are CK2 phosphorylation substrates. Hence, I tested our assumption with an *in vitro* radioactive kinase assay. For this, recombinant GST-tagged CHRAC-14 and CENP-A were expressed in *E. coli* BL21 cells and were affinity purified from the bacterial lysates with the ÄKTA system followed by proteolytic cleavage and removal of the GST-tag. Purified CHRAC-14 and CENP-A were then used as substrates in the kinase assay. Reactions were supplied with commercial human CK2 (hCK2) and radioactive [γ-^{32}P]ATP (Figure 11). As a negative control the Fmr1 KH1/2 domain and as a positive control the Fmr1 RGG domain were used (proteins kindly provided by J. Luitz, Siomi *et al.*, 2002). Additionally, I included Casein as a *bona fide* CK2 substrate (Kennedy, 1992). After the *in vitro* phosphorylation reaction, the samples were subjected to SDS-PAGE and incorporation of ^{32}P during the process of substrate phosphorylation was monitored with an autoradiogram.

First of all, in the presence of hCK2, the positive controls (Fmr1-RGG and Casein) show ^{32}P-labelled bands at the expected sizes when compared to the Coomassie-stained gel (Figure 11 A). The negative control (Fmr1-KH1/2) or samples lacking hCK2 did not show radioactive bands (Figure 11 A, B). hCK2 alone did also not lead to phospho-signals. Importantly, in those reactions containing hCK2, bands with ^{32}P incorporation are visible for CHRAC-14 at the expected ~14 kDa and CENP-A at ~ 17 kDa (Figure 11 B). However, the detected size of approx. 17 kDa for CENP-A is smaller than the expected full-length size of 26 kDa. Interestingly, in the Coomassie-stained gel, some of the phosphorylated bands exhibit a modest size shift due to strong phosphorylation (CENP-A and Fmr1-RGG). On another notion, the ^{32}P-labelled CHRAC-14 band is notably less intense compared to the other samples.

In order to confirm the observed phosphorylation of CHRAC-14 and CENP-A and to map the precise phosphorylation site in both proteins the reactions were performed again, and the gel was submitted for mass spectrometry analysis at the ZMBH proteomics facility. In the course of the measurement, threonine at residue 122 (T +80 m/z) was detected to be phosphorylated in a total of 9 MS2 spectra within an exclusive unique CHRAC-14 C-terminal peptide (Figure 11 C). However, no phosphorylation site could be mapped within CENP-A. According to personal communication with the proteomics facility, the tryptic digest of CENP-A produced short, unmeasurable peptides. Furthermore, GST was detected in the CENP-A sample, which indicates that the band at ~26 kDa was mistaken for CENP-A but rather



B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	129.1	65.1	112.1		K	2,271.9	1,136.5	2,254.9	2,253.9	22
2	244.1	122.6	227.1	226.1	D	2,143.8		2,126.8	2,125.8	21
3	331.2	166.1	314.1	313.2	S	2,028.8		2,011.8	2,010.8	20
4	445.2	223.1	428.2	427.2	N	1,941.8		1,924.7	1,923.8	19
5	546.3	273.6	529.2	528.2	T	1,827.7		1,810.7	1,809.7	18
6	617.3	309.1	600.3	599.3	A	1,726.7		1,709.7	1,708.7	17
7	746.3	373.7	729.3	728.3	E	1,655.6		1,638.6	1,637.6	16
8	860.4	430.7	843.3	842.4	N	1,526.6		1,509.6	1,508.6	15
9	931.4	466.2	914.4	913.4	A	1,412.6		1,395.5	1,394.5	14
10	1,045.5	523.2	1,028.4	1,027.4	N	1,341.5		1,324.5	1,323.5	13
11	1,116.5	558.7	1,099.5	1,098.5	A	1,227.5			1,209.5	12
12	1,203.5	602.3	1,186.5	1,185.5	S	1,156.4			1,138.4	11
13	1,274.6	637.8	1,257.5	1,256.6	A	1,069.4			1,051.4	10
14	1,375.6	688.3	1,358.6	1,357.6	T	998.4			980.4	9
15	1,446.6	723.8	1,429.6	1,428.6	A	897.3			879.3	8
16	1,627.7	814.3	1,610.6	1,609.6	T+80	826.3			808.3	7
17	1,698.7	849.9	1,681.7	1,680.7	A	645.3			627.3	6
18	1,827.7	914.4	1,810.7	1,809.7	E	574.2			556.2	5
19	1,956.8	978.9	1,939.8	1,938.8	E	445.2			427.2	4
20	2,027.8	1,014.4	2,010.8	2,009.8	A	316.2			298.1	3
21	2,124.9	1,062.9	2,107.8	2,106.9	P	245.1			227.1	2
22	2,271.9	1,136.5	2,254.9	2,253.9	E	148.1			130.0	1

Figure 11 *In vitro* phosphorylation of recombinant CHRAC-14 and CENP-A

(A) Recombinant proteins (as indicated) of known CK2 substrates and negative controls were subjected to a radioactive *in vitro* kinase assay with or without human recombinant CK2 (hCK2). **(B)** Similar reactions as in (A) using purified CENP-A and CHRAC-14 as test substrates. **(C)** Mass spectrometry data of the phosphorylated CHRAC-14 band from the kinase assay. Upper panel: From all eleven CHRAC-14 exclusive unique peptides (total coverage was 89%) only a peptide covering the CHRAC-14 C-terminus is depicted (shaded in yellow), in which the phosphorylated residue T122 was detected (shaded in green). Lower panel: Fragmentation table of one exemplary MS2 spectrum.

represents a mix of CENP-A and GST (the band was recognized by a CENP-A antibody in a Western blot [data not shown]). The lower CENP-A band at ~17 kDa, which was phosphorylated *in vitro*, most likely constitutes degradation products of CENP-A. Equally, a GST band at the same height was also present in the CHRAC-14 protein sample. In both cases the free GST was a remnant from incomplete tag removal after cleavage during the protein purification process.

In summary CK2 phosphorylates CENP-A and CHRAC-14 *in vitro* and for CHRAC-14 the phosphorylation was verified by mass spec and mapped at residue T122. Concerning

CENP-A, Huang *et al.* (2019) reported in parallel that CENP-A is phosphorylated *in vivo* by CK2 at serine 20 (S20) in the N-terminus. In summary their findings undermine my observations and despite the lacking proof for a physical interaction other than mass spectrometry, strongly suggest a functional connection of CENP-A, CHRAC-14 and CK2.

3.1.4 **CK2β influences endogenous CENP-A levels**

As reported in my study for CHRAC-14 and by Huang *et al.* for CENP-A, these two proteins are new phosphorylation substrates of CK2. This led us to speculate that the phosphorylation has an effect on the stability or turnover of either of the proteins. In contrast to CHRAC-14 we have an excellent antibody at hand for CENP-A and, therefore, chose to examine the influence of CK2 depletion by RNA interference on CENP-A levels. Earlier, Huang *et al.* did not report any effect of CK2 depletion on CENP-A protein levels because their study was exclusively conducted in cells overexpressing CENP-A, a non-physiological condition, which is often connected to alternative regulation pathways (cf. section 1.4). Accordingly, here, I show that endogenous CENP-A decreases upon CK2β depletion whereas the effect on exogenous CENP-A seems negligible in uninduced cells stably transfected with pMT-CENP-A-V5-His (Figure 12 A).

In order to reliably claim that the observed endogenous CENP-A decrease can be attributed to CK2β depletion and is not a technical artefact, two other CK2β-specific dsRNAs were used for RNAi treatment (Figure 12 B). CENP-A levels in wild type S2 similarly decreased with each of the dsRNAs used (dsRNA2 seems to work less efficiently compared to dsRNA1 and dsRNA3, see left plot for CENP-A Western quantification). Please note, that for dsRNA1 three replicates but for dsRNA2 and dsRNA3 only two replicates were conducted, which is why the statistical power is weaker and hence the CENP-A decrease was less significant for dsRNA2 and dsRNA3. Importantly, the RNAi approach effectively depleted CK2β as monitored by qPCR (Figure 12 B, right plot).

Moreover, I did not only observe reduced CENP-A levels upon a CK2β knockdown, the RNAi treatment with three different dsRNAs against CK2α also diminished CENP-A drastically (Figure 12 C).

Since CK2 depletion leads to diminished CENP-A levels we reasoned that this may be due to protein destabilization by proteasomal degradation. Based on work from our lab (PhD thesis of Dr. Demirdizen) we know that CENP-A proteasomal degradation can be stimulated via the ubiquitin ligase Hyperplastic discs (Hyd) in *Drosophila.* To assess if the observed CENP-A protein decrease is mediated by ubiquitinylation via Hyd, a double knockdown of CK2α and Hyd was performed (Figure 12 D). The Hyd knockdown, however, did not counteract the decreasing CENP-A levels. It is to mention, that the knockdown efficiency was not

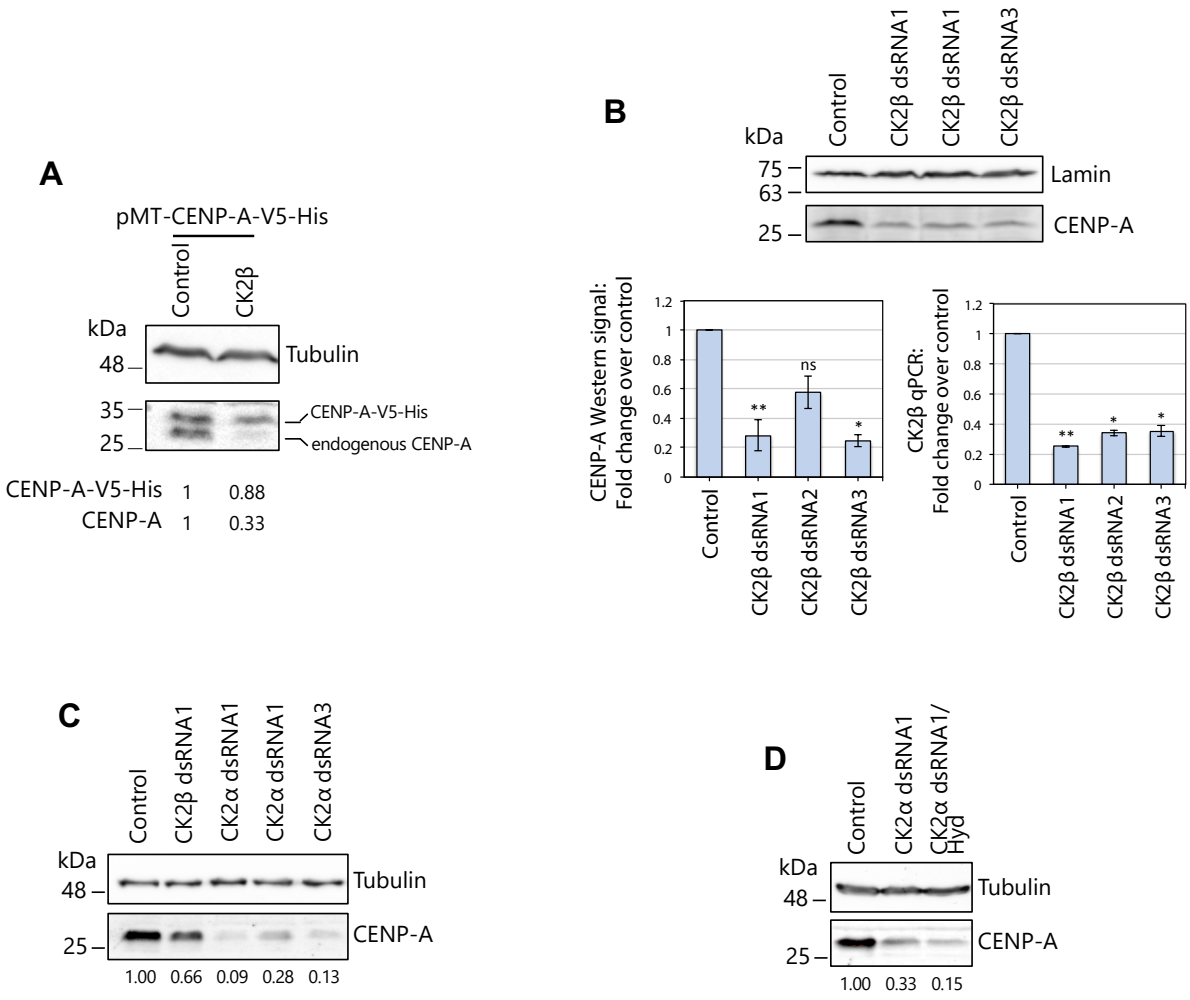


Figure 12 CK2 RNAi reduces CENP-A protein levels

S2 cells were RNAi treated with the indicated dsRNAs for 72 h and CENP-A levels were assessed by Western blotting in whole protein lysates. Brown RNAi was used as a control. Signal fold changes over the control knockdown are indicated below each blot. **(A)** The effect on CK2β RNAi on exogenous vs. stable CENP-A in pMT-CENP-A-V5 transfected cells **(B)** Western blot showing endogenous CENP-A levels after CK2 depletion with three different dsRNAs in wild type cells (top) and the respective quantification (left plot). CK2β transcript expression monitored by qPCR (right plot, CK2β levels were normalized to Gapdh1). Significance determined by t-test, dsRNA1: n=3, dsRNA2 and dsRNA3: n=2. **(C)** Wildtype cells were RNAi treated with three different CK2α dsRNAs and CK2β and Brown dsRNAs were included as a reference. **(D)** Co-depletion of CK2α and Hyd.

monitored in this experiment since an effective Hyd knockdown was previously established by Dr. Demirdizen.

Based on these results it can be concluded, that CK2 phosphorylation has a stabilizing effect on endogenously expressed CENP-A, since the independent depletion of both CK2 subunits, strongly diminished endogenous but not exogenous CENP-A levels.

3.1.5 The effect of mutated phosphosites on the interaction between CENP-A and CHRAC-14

Because phosphorylation can influence the interaction dynamics between proteins (Bhaskara *et al.*, 2019) we wondered if CK2-dependent phosphorylation could restrict or promote the interaction between CENP-A and CHRAC-14. To explore this further, I established stable S2 cell lines, transfected with CENP-A and CHRAC-14 harboring the wild type version of the CK2 phosphorylation motif, a non-phosphorylatable mutation to alanine or a mutation to aspartic acid mimicking a permanent phosphorylation. To facilitate the pulldown and the detection the proteins were additionally tagged with V5. The experimental design was set up in a way that both, pull down of CENP-A-V5 and probe for CHRAC-14 in the eluate and the other way around was possible, in order to observe changes in their interaction strength. The expression of the proteins was slightly induced over night with 1 µM $CuSO_4$ to facilitate the detection via Western blot and to marginally increase their levels for an efficient pulldown. Since the the CHRAC-14 antibody may be too insensitive for a proper CHRAC-14 detection I also established CENP-A-V5 phosphomutant cell lines that additionally carry the pMT-CHRAC-14-Flag construct. No matter in which direction the pulldowns were performed, no interaction could be observed at all. As before in section 3.1.2 I attempted to optimize the pulldown conditions by using more antibody, more cells or non-covalently coupled antibody.

Despite the drawback of not being able to examine the interaction behavior between CENP-A and CHRAC-14 in dependency of their CK2 phosphorylation status in S2 cells I detected a new CENP-A band in the cell lysate of CHRAC-14-T122A-V5 cells (Figure 13). This band runs at approximately 40 kDa; representing an approximately 15 kDa size shift compared to the native 26 kDa CENP-A band. Noteworthy, the new band is not visible in either of the other cell lines (wt or T122D). All three cell lines express CHRAC-14-V5 at different levels and the protein is barely detectable in input samples of the T122A and T122D cell lines. Only after V5 immunoprecipitation, CHRAC-14-V5 became visible by Western blot.

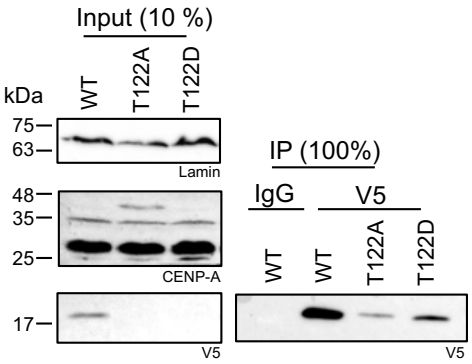


Figure 13 CHRAC-14 phosho mutant cells form a larger version of endogenous CENP-A

S2 cells were stably transfected with pMT-CHRAC-14-V5-His containing the wild type CK2 phosphosite or pMT-CHRAC-14-V5 containing the indicated CK2 motif mutations. Protein expression was induced with 1 µM $CuSO_4$ for 16 h, cells were lysed and subjected to an anti-V5 immunoprecipitation. WT = wildt ype, T122A = Threonine 122 to alanine mutation, T122D = Threonine 122 to aspartic acid mutation.

3.2 CENP-A chromatin composition after CHRAC-14 depletion and in DNA damage

3.2.1 DIvA system in *Drosophila* S2 cells

Previous results implicate that CENP-A protein levels are increased when CHRAC-14 is depleted either by RNAi in cells or in mutant flies (Mathew *et al.*, 2014). An effect that peaks in the formation of extra CENP-A spots which are visible on the microscopic level. However, we are missing detailed insights into CENP-A chromatin distribution, especially during DNA damage. In order to gain a better understanding of the CENP-A distribution in the absence of CHRAC-14 we traced endogenous CENP-A genome-wide by CUT&Tag sequencing in asynchronously cycling *Drosophila* S2 cells. To eliminate CHRAC-14 (or Brown as a control) cells were RNAi treated for 6 days. Additionally, site-specific DNA damage was induced on day 6 in order to assess if CENP-A is accumulating at DNA damage sites.

To achieve all this, we induced site-specific DNA damage in this study with the restriction enzyme AsiSI, an advantageous setup where the restriction motif and thus possible DNA damage sites can be bioinformatically mapped. The DNA damage induction via AsiSI system termed 'DIvA' has been previously established by the Legube lab (Iacovoni *et al.*, 2010). To establish the DIvA system in the lab, stable *Drosophila* S2 cell lines were created in the course of his study harboring an inducible pMT-NLS-AsiSI-mAID-GFP or -V5 construct. Any potential leaky AsiSI expression from the pMT vector and, therefore, unwanted induction of DNA damage was constantly inhibited via the attached minimal auxin inducible degron tag

(mAID) which mediates polyubiquitinyation and proteasomal degradation. Of note, the cells are co-transfected with a pMT-OsTIR1 construct, encoding for the *Oryza sativa (OsTIR1)* F-box protein TIR1, which mediates substrate binding in the presence of auxin. The auxin-based degron system is a widely used system for rapid and inducible protein depletion, which was developed in the Kanemaki lab (Nishimura *et al.*, 2009) and was implemented for the use with AsiSI by the Legube lab as well (Aymard *et al.*, 2014). Thus, in my system, only upon auxin depletion and mild induction with $CuSO_4$ AsiSI is stabilized and re-localizes to the nucleus to cut DNA at its recognition motif.

Initially, the system had to be validated in S2 cells: Figure 14 A shows a representative time course experiment of the DIvA cell lines where earliest after 3 hours a clear increase of the damage marker γH2Av was visible. Also, C-terminally tagged AsiSI showed higher levels of the DNA damage marker compared to N-terminally tagged AsiSI. Moreover, a phosphorylated H2Av increase is only visible in cells transfected with the AsiSI construct and not in control cells transfected with the empty vector. DNA damage levels and AsiSI localization were also monitored via fluorescent microscopy and one representative experiment is shown in Figure 14 B. In the control state ($AsiSI^{OFF}$) AsiSI seems to be leaky expressed, as the V5 signal is clearly visible but is predominantly restricted to the cytoplasm of almost all cells in the image. After $CuSO_4$ induction and auxin depletion ($AsiSI^{ON}$) AsiSI accumulated in the nucleus to varying extents. The nuclear γH2Av signal increased as well and occasionally formed distinct spots. Of note, γH2Av and AsiSI signals were not equally high in all the cells but rather only in a subset. Usually, cells with strong V5 nuclear signal exhibited distinct γH2Av nuclear dots.

I also tested, if CHRAC-14 can be efficiently knocked down in the $AsiSI^{OFF}$ and $AsiSI^{ON}$ conditions. A representative experiment is depicted in Figure 14 C. Depletion of CHRAC-14 upon RNAi in the DIvA system is evident but seems incomplete as a faint CHRAC-14 band is still visible. Since the desired treatments in combination with the DIvA system worked satisfactorily for our purposes I proceeded with the establishment and optimization of the CUT&Tag protocol.

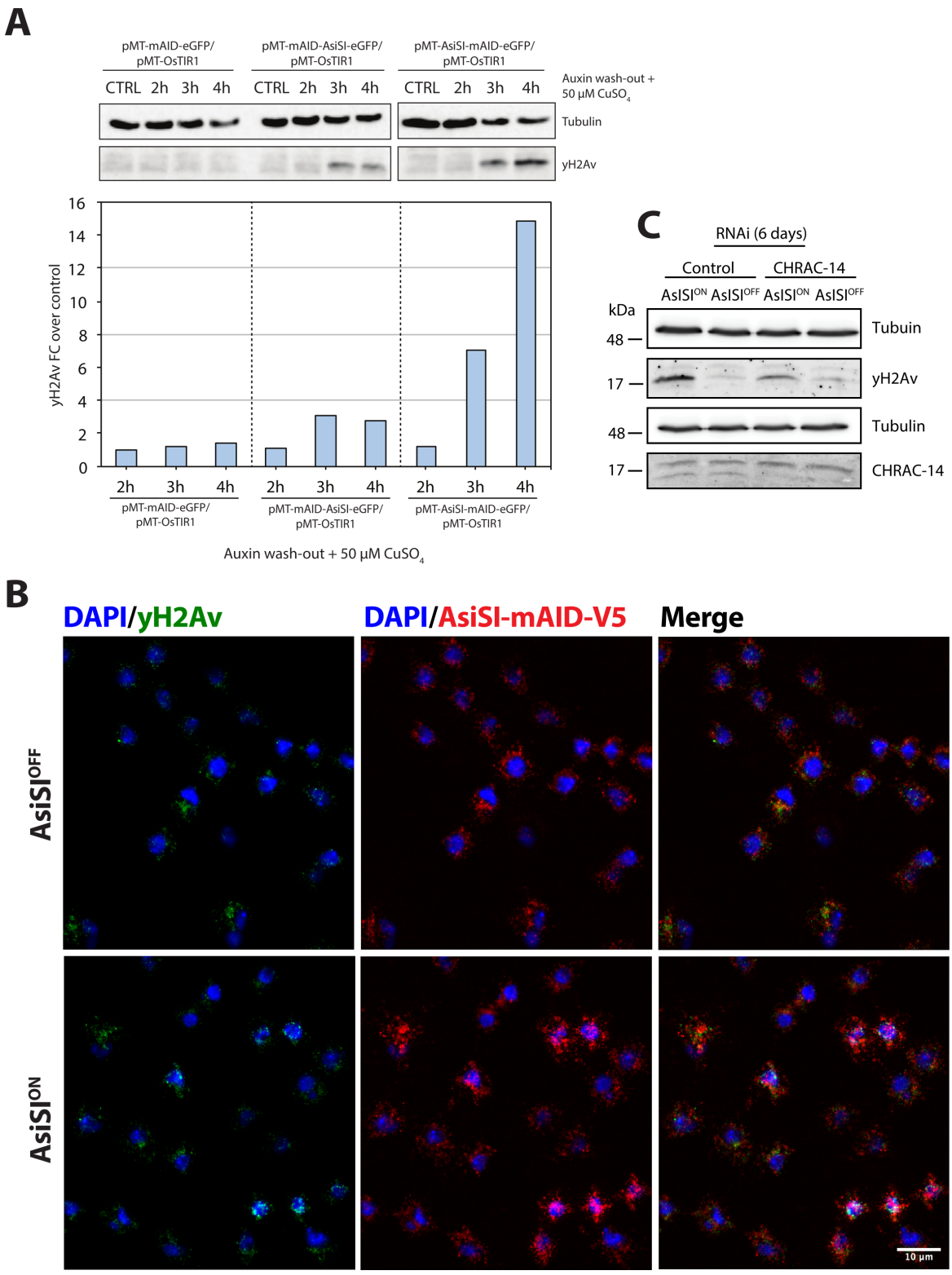


Figure 14 DIvA S2 cells

(A) S2 cell lines harboring the indicated AsiSI or control constructs next to the pMT-OsTIR1 construct were subjected to a 4 h DNA damage time course or left untreated as a control. To induce DNA damage via AsiSI, destabilizing auxin was washed out and protein expression was induced with 50 µM $CuSO_4$. The efficiency of the DNA damage treatment was monitored with an γH2Av Western blot and Tubulin was used as a loading control and for normalization during the quantifications. ▶

3.2.2 CENP-A and γH2Av CUT&Tag experimental setup

Briefly, the CUT&Tag workflow is an *in-situ* targeted sequencing approach which avoids formaldehyde fixation (Figure 15). The procedure starts with the isolation and permeabilization of nuclei and continues by incubation with primary and secondary antibodies against the desired protein target. Then, a protein A transposase fusion protein (pA-Tn5) is added and binds to the secondary antibodies via the protein A tag. Of note, with the recombinant pA-Tn5, traces of bacterial DNA are carried-over. Subsequently, DNA cleavage and simultaneous adapter tagmentation by the transposase is activated, generating footprints of the protein of interest and of the carried-over *E. coli* DNA, which is non-specifically tagmented, as well. Eventually, total DNA is extracted, tagged fragments are amplified by PCR and cleaned up with magnetic, nucleic acid-binding beads. In this thesis, CENP-A and γH2Av primary antibodies were used to track CENP-A incorporation and to monitor DNA damage hotspots, respectively. As a background control an IgG epitope control antibody was used. Initially, three replicates were prepared for sequencing, however, in one replicate the CHRAC-14 knockdown efficiency was unsatisfactory (data not shown). Therefore, we decided to continue with two instead of three replicates for this first pilot round of sequencing. Furthermore, the library preparation for the IgG control in replicate 2 failed, which is why this experiment includes only one IgG control from the first replicate. As a side note, the IgG CUT&Tag samples were prepared with wild type S2 cells because the DIvA cell count after the RNAi course was too low for additional use as negative control.

3.2.3 CENP-A and γH2Av CUT&Tag sequencing metrics

For the bioinformatic CUT&Tag data processing and analysis reads were aligned using four different reference genomes. Since centromere sequences are not fully assembled in the official BDGP *Drosophila* genome (Adams *et al.*, 2000; Hoskins *et al.*, 2015), I additionally used custom *Drosophila melanogaster* assemblies, which were recently established by the Mellone and Larracuente labs and reported in Chang *et al.*, 2020. The authors on the one hand created a centromere-only reference by assembling CENP-A ChIPtigs and on the other hand compiled a whole genome assembly with annotated centromeres and repeats (here forth referred to as PacBio assembly). This enabled the analysis of CENP-A binding at centromeres and genome-wide repeat elements. Additionally, it is necessary to align to the *E. coli* genome

(B) DIvA cells were subjected to the standard 3 h of DNA damage treatment and γH2Av and AsiSI-mAID-V5 localization was visualized by IF and widefield microscopy. Images are single z-slices. **(C)** CHRAC-14 RNAi efficiency was validated in the DIvA system. The upper band in the CHRAC-14 Western refers to recombinant MNase, which was used to digest chromatin during protein preparation.

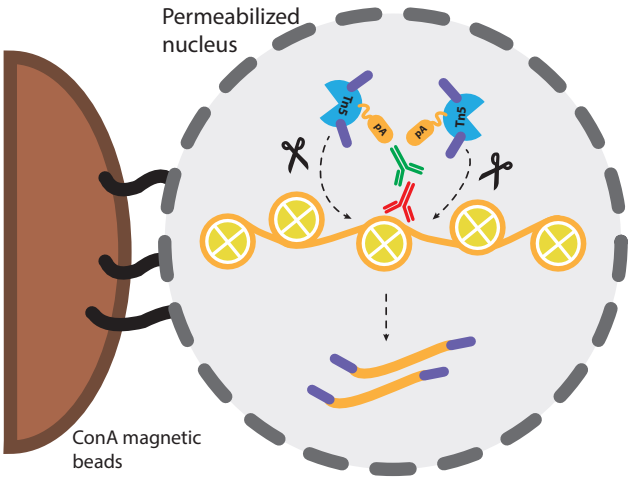


Figure 15 CUT&Tag workflow
CUT&Tag starts with the preparation of nuclei, which are permeabilized and kept at this condition throughout the experiment. Then nuclei are incubated with first and secondary antibodies allowing *in situ* target recognition. Next, a transposase (Tn5) loaded with adapters is added, which gets recruited to the antibody bound target of interest via a protein A (pA) tag. After washing out unbound Tn5, DNA fragmentation and tagmentation with small adaptors is activated. Eventually, DNA is extracted and tagged fragments are enriched in a library PCR using Illumina sequencing compatible primers.

since the reads originating from the carried-over bacterial DNA are used as a spike in control for data calibration. Moreover, different alignment settings were used depending on the employed reference genome. To analyze genome-wide CENP-A binding with the BDGP assembly, paired alignment was conducted. For exploring CENP-A repeat and centromere binding with the custom PacBio and ChIPtigs assemblies, single-end alignment was carried out, since paired alignment is rather inaccurate with repetitive sequences. To provide a technical overview of this pilot CUT&Tag sequencing approach, I briefly summarize the main sequencing metrics the following paragraphs.

The CUT&Tag sequencing approach led to an overall alignment rate of an average of 90.5 % (Figure 16 A). In general, the alignment rates differ depending on what genome and what antibody was used: With the whole genome assemblies (BDGP and PacBio) the alignment rate was consistently at around 80%. The alignment to the CENP-A ChIPtigs was between 15% to 30% and was clearly higher for the CENP-A samples than for γH2Av (29.9% for CENP-A vs. 16.8% for γH2Av), especially in CHRAC-14 depleted samples. There was also notable alignment to the *E. coli* genome (CENP-A 3.9%, γH2Av 1.3%).

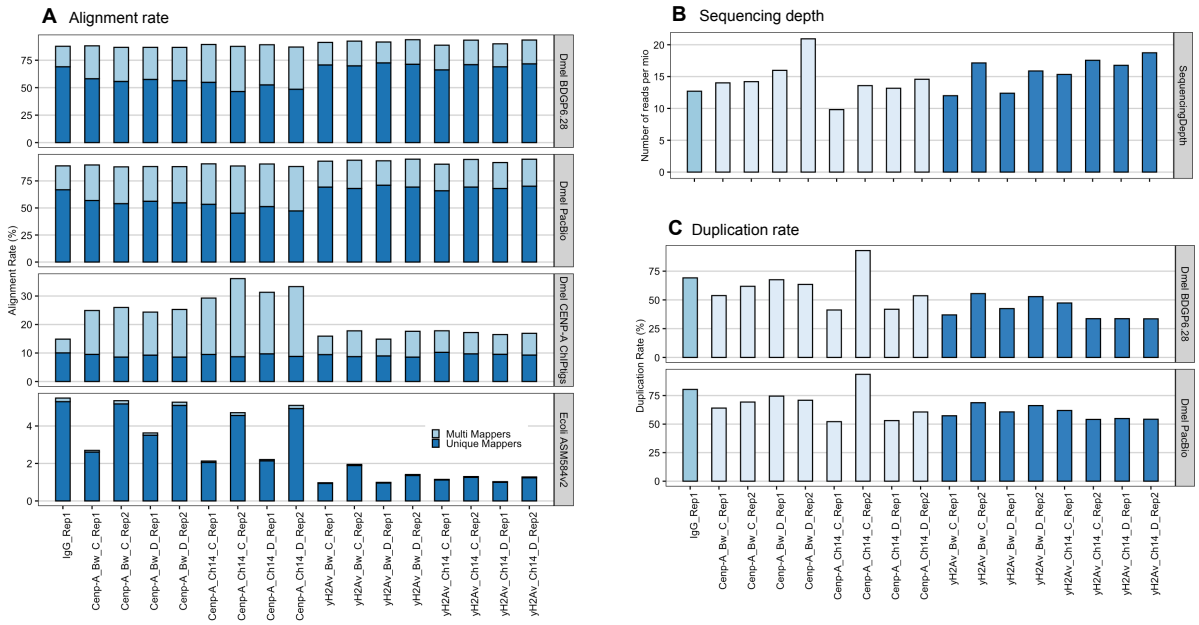


Figure 16 Technical CUT&Tag sequencing metrics

(A) Alignment statistics for the four indicated read alignment sets, partitioned into multi mappers and unique mappers of each sample. **(B)** Total sequencing depth per sample. **(C)** Duplication rates for the whole genome data sets per sample. Rep = Replicate, Bw = Brown control RNAi, Ch14 = CHRAC-14 RNAi, C = AsiSI^{OFF} (Control, no damage), D = ASIS^{ON} (Damage)

Importantly, the CENP-A alignment rate for multi mappers was higher than for γH2Av regardless of which *Drosophila* assembly was employed.

To ensure a decent mapping quality the alignment data was then filtered with SAMtools (Li *et al.*, 2009) (-q 30 for PacBio and -q 20 for BDGP) and the duplication rates were determined (Figure 16 C). Especially certain CENP-A and the IgG samples have a duplication rate of > 60% in the BDGP dataset (IgG: 69.1%, CENP-A: 59.5%, γH2Av 42.0%), so that library sizes are approximately halved after de-duplication. The average duplication rates are generally higher for the PacBio alignment data (IgG: 80.4%, CENP-A: 67.3%, γH2Av 59.8%). In both cases, specifically the Chrac-14_Damage_Rep2 sample has a very high duplication rate, which is why the majority of the reads are lost after de-duplication. Aside from that, despite of the fact that the CENP-A duplication rate is higher than for γH2Av it seems to decrease with CRACH-14 RNAi.

Next, in order to maintain hypothetical repetitive CENP-A reads only the BDGP data but not the PacBio data was depleted of duplicates. Since the PacBio dataset will be used to determine CENP-A repeat and centromere binding, de-duplication would be counterproductive rather than beneficial. Centromeres and elements such as transposable elements are repetitive and, therefore, are characterized by low sequence complexity hindering accurate distinction between true PCR duplicates and low complexity repeating units. The observed combination of increased duplication rate and large multi mapper portions in the CENP-A data

aligned to centromere annotated custom references strongly indicates that this dataset most likely contains biologically meaningful repetitive alignments, which we decided to include in our analysis.

As an additional technical survey measure, the count and length of the cleaned reads was assessed (Figure 17). For all samples the first peak in fragment length is at 65 nucleotides followed by a second peak at approximately ~170 nucleotides. The second peak is clearly stronger for the γH2Av samples. These samples also show a faint third peak at a fragment length of ~360 nucleotides. The replicates show comparable fragment sizes and counts, however, Cenp-A_Ch14_C replicate 2 is an exception and has very low counts in comparison to replicate 1, which was expected after de-duplication and the high duplication rate observed for this sample. Notably, the fragment size periodically peaks every 10 nucleotides, especially for smaller fragments, resembling a saw tooth like pattern, which is very common for CUT&Tag experiments and has been reported before (Zhu *et al.*, 2019).

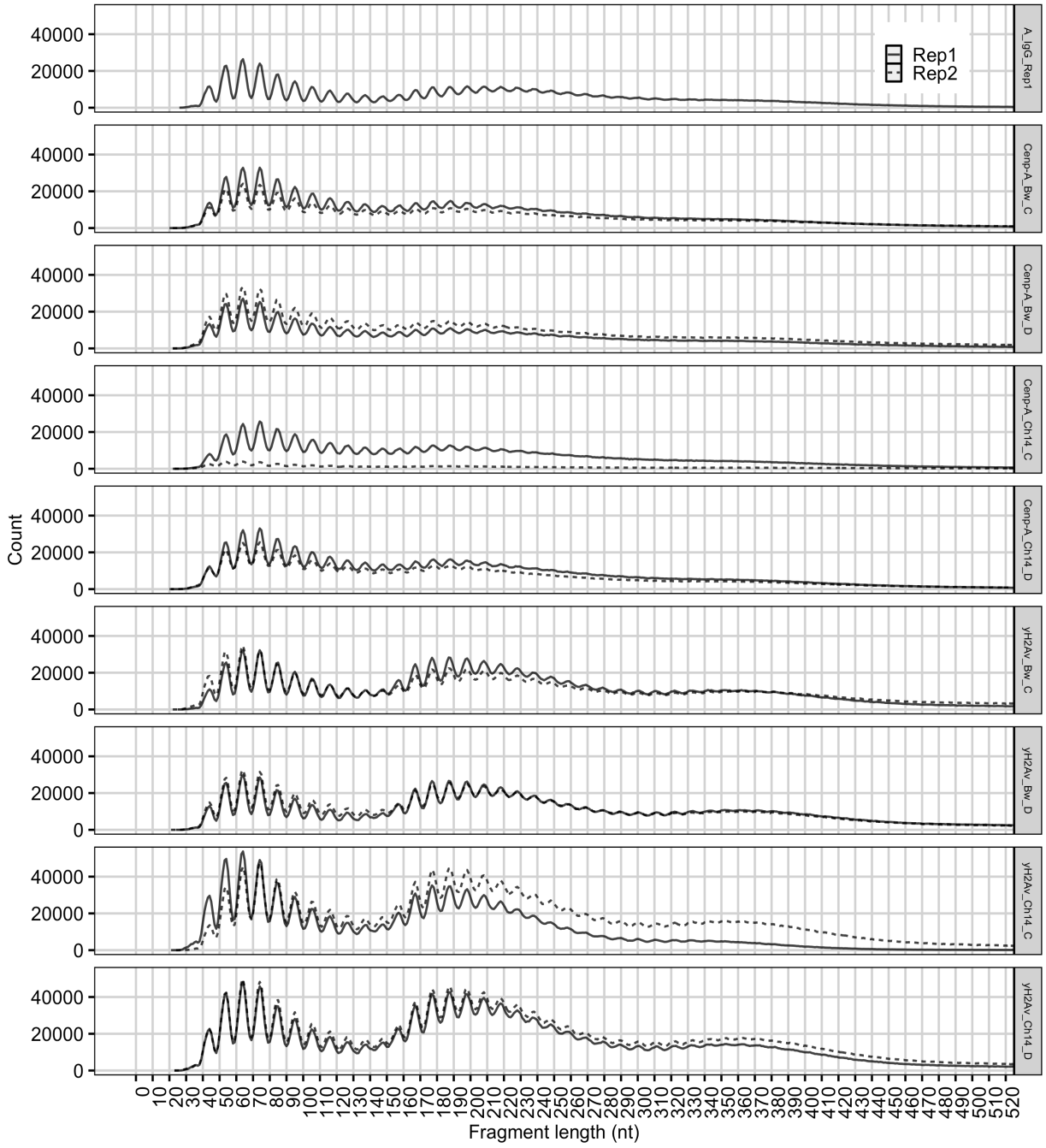


Figure 17 CUT&Tag fragment sizes

Fragment lengths and counts were calculated from the filtered (i.e. only concordant reads with a maximal size of 700 bp and with decent mapping quality) and de-duplicated sam files of the BDGP dataset and the distribution is plotted here.

3.2.4 **CENP-A CUT&Tag enriched centromeric elements**

After assessing the technical survey of the CUT&Tag data, the first step of the core analysis was to validate the ability of CENP-A CUT&Tag to enrich expected centromeric target sequences. For that, the cleaned and calibrated coverage bigwig files resulting from the PacBio alignment was utilized for peak calling with MACS2 (-p 0.01)(Zhang *et al.*, 2008) in control samples (i.e. control RNAi and AsiSI^{OFF}). The IgG background sample was used as a control for normalization and the peak data from both replicates was then overlapped to extract reproducible peaks. Then, peaks residing in centromeric contigs (chr2: tig00057289, chr3: 3R_5, chr4: Contig119, chrX: Contig79, chrY: Y_Contig26, established by Chang *et al.*) were

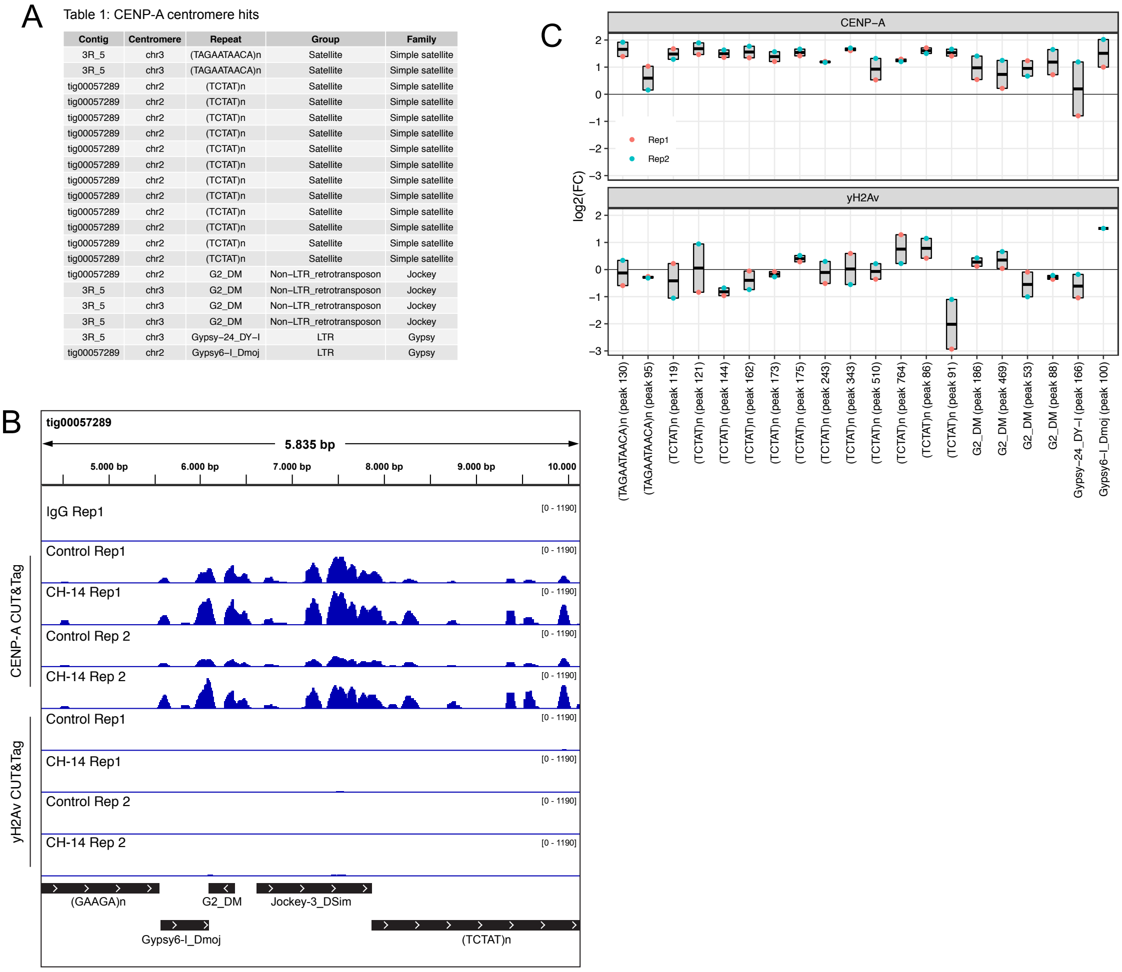


Table 1: CENP-A centromere hits

Contig	Centromere	Repeat	Group	Family
3R_5	chr3	(TAGAATAACA)n	Satellite	Simple satellite
3R_5	chr3	(TAGAATAACA)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	(TCTAT)n	Satellite	Simple satellite
tig00057289	chr2	G2_DM	Non−LTR_retrotransposon	Jockey
3R_5	chr3	G2_DM	Non−LTR_retrotransposon	Jockey
3R_5	chr3	G2_DM	Non−LTR_retrotransposon	Jockey
3R_5	chr3	G2_DM	Non−LTR_retrotransposon	Jockey
3R_5	chr3	Gypsy−24_DY−I	LTR	Gypsy
tig00057289	chr2	Gypsy6−I_Dmoj	LTR	Gypsy

Figure 18 Analysis of centromeric CENP-A

Peaks were called with MACS2 on CENP-A CUT&Tag control bigwig files (i.e. control RNAi and AsiSI^{OFF}) and the resulting data was overlapped between both replicates to extract reproducible peaks. **(A)** Centromere-associated peaks were then identified based on the contig sequence names and all hits are listed. **(B)** Snapshot of the chr 2^{Capri} centromere island with indicated bigwig tracks (equal scale) for signal comparison of CENP-A to IgG and γH2Av samples. **(C)** Average bigwig scores underlying the centromeric peak areas were computed and the fold change in CHRAC-14 over control RNAi was calculated and plotted. Horizontal line = indicates hypothetical control mean of 1, CH-14 = CHRAC-14.

extracted and listed in Figure 18 A and Supplemental Table 2. CENP-A peaks were detected in centromere 2 and 3 at the non-LTR retrotransposon G2/*Jockey-3*, at LTR retrotransposons of the Gypsy family and at the simple satellite tandem repeats (TCTAT)n and (TAGAATAAC)n. As exemplified with the chr 2^{Capri} centromere island in Figure 18 B the CENP-A signal at centromeric locations is very clear and strongly exceeds the signal of IgG and γH2Av. Next, we wondered, if increasing CENP-A levels upon CHRAC-14 RNAi, which are valid on the total protein or IF level, can be observed via CUT&Tag at centromeres. This was examined by computing the bigwig score underlying the centromeric CENP-A peaks followed by calculating the fold change (FC) of the CHRAC-14 over the control RNAi. As a control this was done for γH2Av as well. Results are plotted in Figure 18 C where the horizontal line indicates the hypothetical mean value 1 of the control RNAi. Compared to that, fold change means between replicates for many of the elements were above 1 for CENP-A but not for γH2Av. Especially for elements in centromere 2 the fold change seems to robustly increase. In comparison, γH2Av shows a tendency of negative fold changes at the centromere elements. The fold change significance was not tested here, since only two replicates are available in this study.

3.2.5 CENP-A binds non-centromeric repetitive elements

As a complementary approach I analyzed if CENP-A was binding to other repeats outside of the centromere, employing the CENP-A dataset of significant control peaks introduced in the previous section (3.2.4). To do so, the data was further filtered to eliminate centromeric peaks and peaks at simple satellite repeats, and only the top 20% of significant peaks were considered for analysis (Supplemental Table 3). As shown in Figure 19 A and B CENP-A peaks were identified at a number of different repeat families, of which some are represented also at the centromere (Jockey LINE or Gypsy LTR). Other detected repeat families include simple satellites such as *dodeca*, Helitrons (rolling-circle DNA TE), 1.688 satellites and Pao elements (LTR). Again, we were interested to know if CHRAC-14 depletion would increase CENP-A binding to the genome. As done in section 3.2.4 the bigwig scores underlying the CENP-A peaks were computed followed by calculating the fold change in the CHRAC-14 over the control RNAi. In addition, γH2Av data was used as a comparison. The results (Figure 19 B) are difficult to interpret, since on the one hand the replicates are inconsistent for many elements and show a large variance and on the other hand a clear tendency towards a positive or negative fold change was not directly obvious. Some elements seemed to lose, and some seemed to gain CENP-A signals in the CHRAC-14 RNAi; a pattern common for both histone variants. It is to note that the peaks outside of the centromere were variable in their underlying CENP-A signal intensity. As depicted in Figure 19 C, some regions

exhibited very strong signals, comparable to what is observed at centromeres (Contig102). At other γH2Av positive regions, CENP-A signals were less intense (3R_28).

3.2.6 **CENP-A incorporates genome-wide**

Since various studies have reported CENP-A ectopic incorporation upon overexpression (cf. section 1.4) I set out to explore global genome-wide CENP-A loading upon CHRAC-14 depletion. For this the BDGP alignment dataset was employed for peak calling with MACS2 (-q 0.01) (Zhang *et al.*, 2008) using the IgG sample as a background control. The resulting peak data from the control and the CHRAC-14 knockdown was overlapped to identify commonalities and differences between both conditions. Unfortunately, the CENP-A coverage in the chromosome arms was very poor for the second replicate of the CHRAC-14 control RNAi condition (cf. Figure 17), which prohibited any analysis including this sample. Therefore, the peak calling is only based on replicate 1. As separate datasets, shared peaks between both conditions and peaks, which are only present in the control or the CHRAC-14 RNAi sample were extracted. For further analyses only the top 20% most significant hits were considered in order to increase the reliability of future observations (Supplemental Table 4). Figure 20 A illustrates the overlap between the RNAi datasets of replicate 1 and unravels that CHRAC-14 depletion seemed to lead to an increase in CENP-A peak count creating a major set of extra peaks. Next the distribution pattern of CENP-A to genomic elements was analyzed (Figure 20 B). Globally CENP-A peaks most frequently localized within promoter regions with ≦ 1 kb distance to the TSS, to other introns (i.e. other than the first intron of an ORF), to distal intergenic regions, to promoter regions with 1-2 kb distance from TSS, to other exons (i.e. other than the first exon of an ORF) and to promoter regions with 2-3 kb distance from the TSS. Upon CHRAC-14 RNAi treatment there was no remarkable change in CENP-A distribution. Slightly more often, peaks seemed to localize in promoter regions with 1-2 kb distance from TSS, in other exons and in promoter regions with 2-3 kb distance from the TSS. On the other hand, there were less frequent peaks in promoter regions with ≦1 kb distance to the TSS, in other introns or in distal intergenic regions. However, this trend can also be observed in the control RNAi treatment.

From this we can conclude that CENP-A localizes to promoters and gene bodies and to distal intergenic regions. Even though RNAi treatment as such rather than CHRAC-14 depletion leads to a marginal differential distribution of these peaks, CHRAC-14 absence clearly promotes CENP-A accumulation genome-wide, which we detected as a large extra set of peaks.

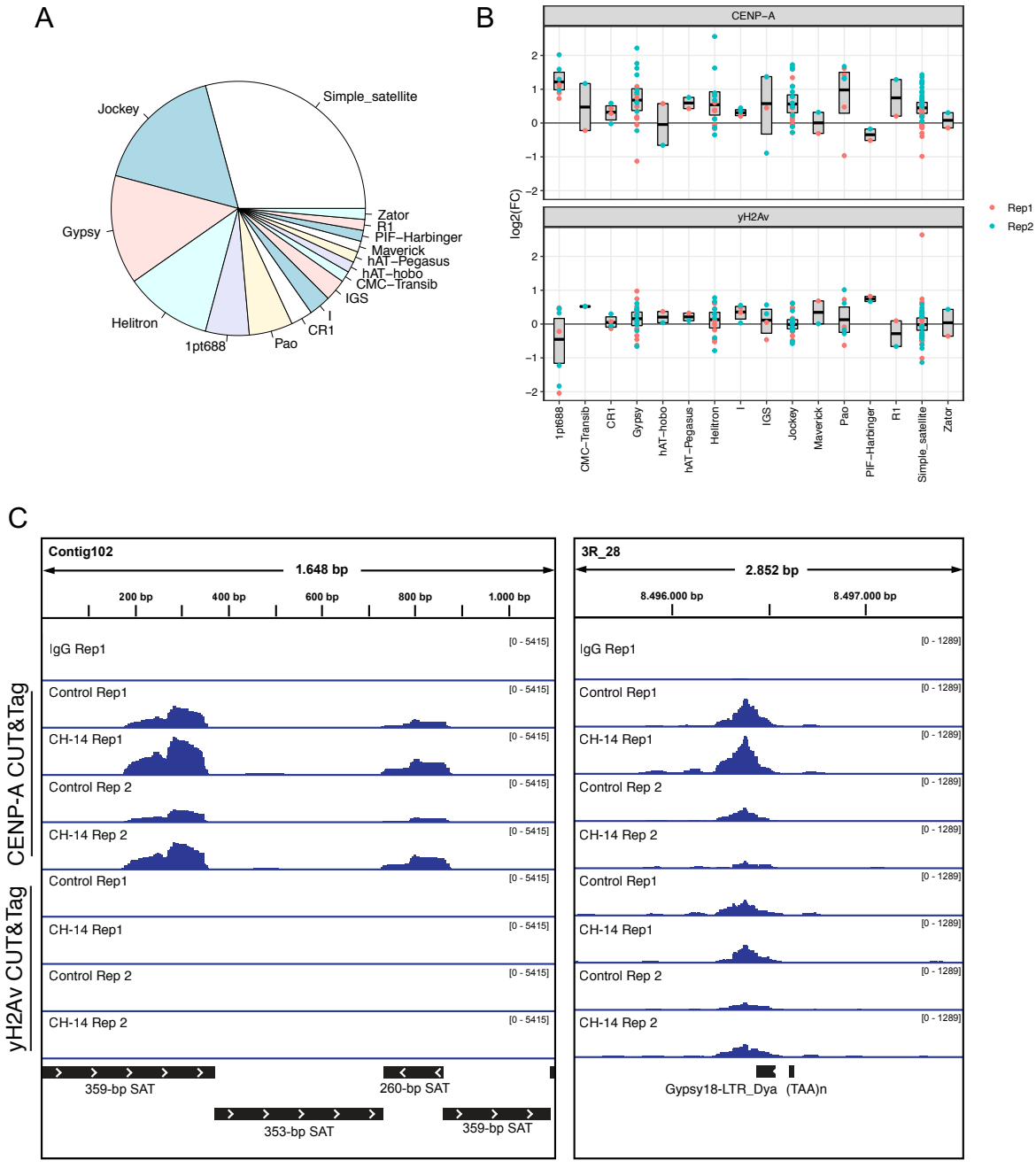


Figure 19 Analysis of non-centromeric CENP-A repeat binding

The overlapping PacBio peak data introduced in Figure 18 was further cleared of centromeric peaks and peaks at simple repeats and only the top 20% significant CENP-A hits were considered for analysis. **(A)** Overview of repeat families with CENP-A peaks outside of centromeres (top 20%). **(B)** Average bigwig scores underlying the peak areas were computed and the log2 fold change in CHRAC-14 RNAi over control RNAi was calculated and plotted. **(C)** Snapshot of two exemplary non-centromeric regions (left: Contig102, right: 3R_28) with indicated bigwig tracks.

To undermine our finding that CENP-A accumulates upon CHRAC-14 knockdown, I again calculated the bigwig score fold change at the peak areas. The plot in Figure 20 C compares how CENP-A enrichment changed in areas where peaks were found in both RNAi conditions and in areas of RNAi-specific peaks. As a reference I included γH2Av. The CENP-A signal within the shared peaks increased modestly as the mean $\log_2$ fold change at these sites was slightly above 0 (0.21 ± SD 0.33). This was also valid for γH2Av at these locations (0.16 ± SD 0.63). In contrast, the CENP-A load seemed to be more affected upon CHRAC-14 RNAi, where a mean $\log_2$ fold change of 1.08 ± SD 0.46 was reached. Importantly, such a pronounced shift was not evident for γH2Av (0.12 ± SD 0.61). Opposite to that, at control RNAi-specific peak sites, the CENP-A mean $\log_2$ fold change was slightly below 0 (-0.22 ± SD 0.45). γH2Av values were comparable to the shared peaks scenario (0.16 ± SD 0.63). Considering the standard deviation values, only the change of CENP-A values in the CHRAC-14 RNAi-specific peak areas seem significant. Lastly, in order to demonstrate that the presented bioinformatical approach enables the identification of genuine new CENP-A peaks, an example is depicted in Figure 20 D from replicate 1. CENP-A seems clearly more enriched in the exon of CG5883 on chr3L in the CHRAC-14 RNAi condition replicate 1 when compared to the control samples (red box). Only at these CHRAC-14-specific sites, CENP-A but not γH2Av binding increases notably. This confirmed our assumption, that CHRAC-14 absence causes ectopic CENP-A accumulation, and that CHRAC-14 seems to be a histone chaperone primarily dedicated to CENP-A under physiological conditions.

In summary, the location of CENP-A binding in the genome is affected by both the control and the CHRAC-14 RNAi, since in both conditions, CENP-A forms specific peaks. Nevertheless, CHRAC-14 depletion causes far more new CENP-A peaks than the control peak.

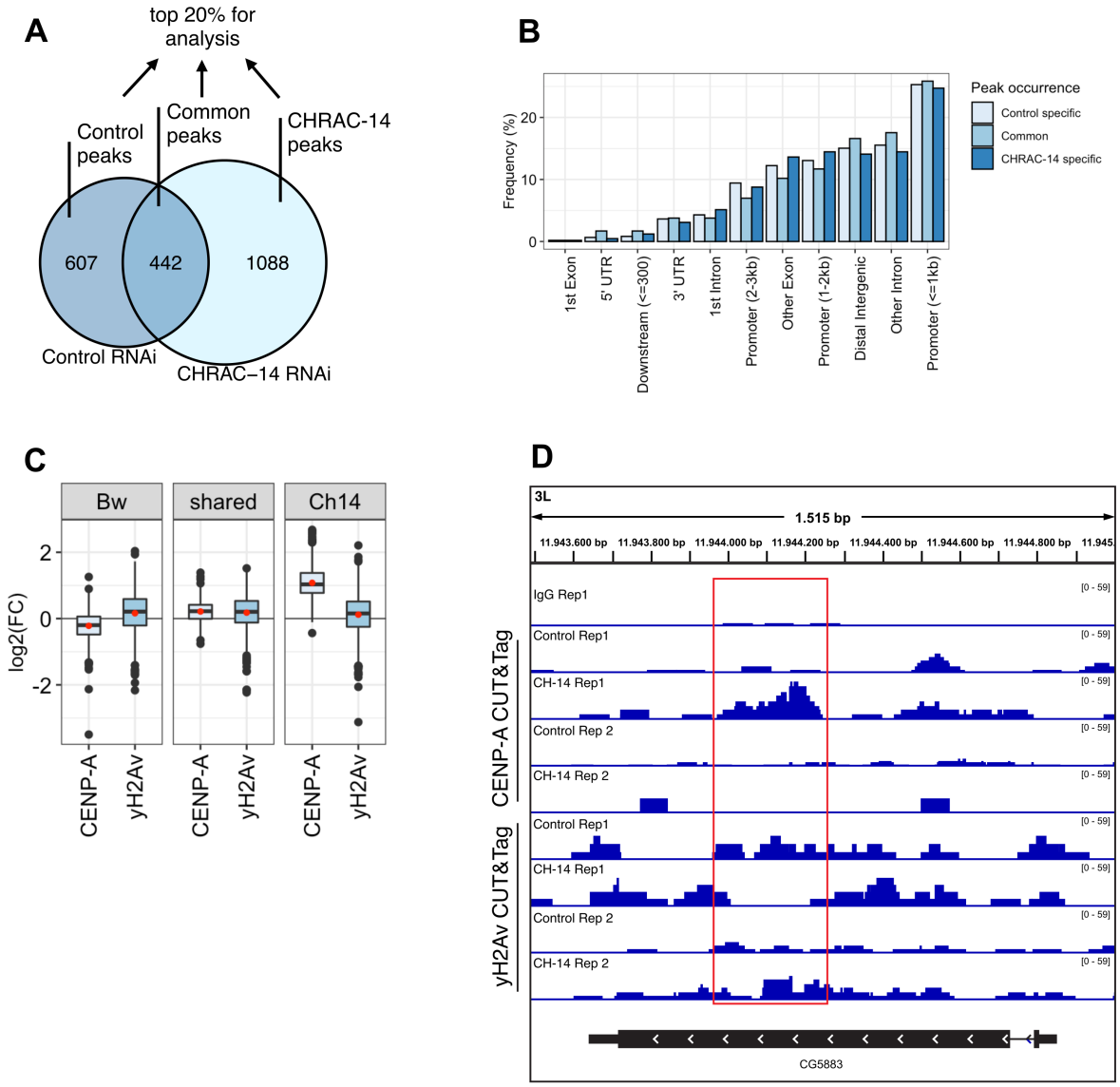


Figure 20 CENP-A accumulates genome-wide upon CHRAC-14 RNAi

(A) For replicate 1 CENP-A peaks in the control and CHRAC-14 RNAi condition were called with MACS2 (-p 0.01) normalizing to the IgG background sample. **(B)** CENP-A peak distribution to genome elements specific to the control or CHRAC-14 RNAi treatment and for peaks found in both conditions **(C)** CENP-A and γH2Av mean $\log_2$ fold changes of bigwig scores in CHRAC-14 over control RNAi samples were calculated for locations underlying common peaks or new peaks found after control or CHRAC-14 RNAi. Red dots indicate the mean, FC = fold change, Bw = Brown control RNAi, Ch14 = CHRAC-14 RNAi **(D)** Example of newly accumulated CENP-A identified via the presented bioinformatics approach.

3.2.7 CENP-A is not detected at candidate DNA damage sites

One part of our aim was to determine if CENP-A accumulates at sites of DNA damage when CHRAC-14 is knocked down and thus I set out to identify DNA damage sites, which were cut by AsiSI and are presumably marked by γH2Av. The first step to locate DNA damage sites was to call damage-specific γH2Av peaks in the AsiSI^{ON} over the AsiSI^{OFF} condition with MACS2 (only control RNAi samples were used, -q 0.01)(Zhang *et al.*, 2008). Then the peak datasets of both replicates were overlaid yielding 125 common and supposedly reproducible γH2Av peak locations (Figure 21 A). Scanning the vicinity (± 3 kb) around these regions for the presence of an AsiSI motif resulted in 35 hits, which I inspected with IGV (Robinson *et al.*, 2011) and aimed for a well recognizable increase in γH2Av upon AsiSI activation. A set of 8 preeminent regions was selected from there and the γH2Av bigwig average profiles in AsiSI^{ON} and AsiSI^{OFF} conditions are depicted in Figure 21 B and listed in Supplemental Table 5. For reference, 1211 predicted AsiSI sites exist in the human genome and in human the Legube lab mapped 174 AsiSI-cut sites (Clouaire *et al.*, 2018). In *Drosophila* AsiSI seems to cut more frequently, as there are 1721 mapped sites (predicted with HOMER [Heinz *et al.*, 2010], data not shown) even though the genome is smaller (1.68×10^8 bp vs. 3.4×10^9 bp, [Gregory, 2021]). As a negative control average profiles were also plotted over randomly distributed EcoRI sites. The γH2Av signal shows a recognizable increase in AsiSI^{ON} compared AsiSI^{OFF}. This increase in signal is more prominent in replicate 2 and does not occur at EcoRI sites. Importantly, the averaged γH2Av signal intensity at AsiSI motifs is clearly stronger compared to the IgG signal or the signal over EcoRI sites. To answer our initial question, the signal pattern of CENP-A at these γH2Av positive sites was explored next. However, CENP-A does not exhibit any significant signal intensities and the average profiles looked similar to IgG and to the profile at EcoRI sites (data not shown). Despite this drawback, Figure 21 C exemplifies that the DNA damage sites mapped via this approach leads to the discovery of γH2Av signal accumulation near AsiSI recognition motifs in AsiSI^{ON} cells. The DIvA system, which I established in the course of this study is appropriate for studying DNA damage sites when combined with deeper sequencing to generate sufficient coverage of candidate proteins at the AsiSI motif.

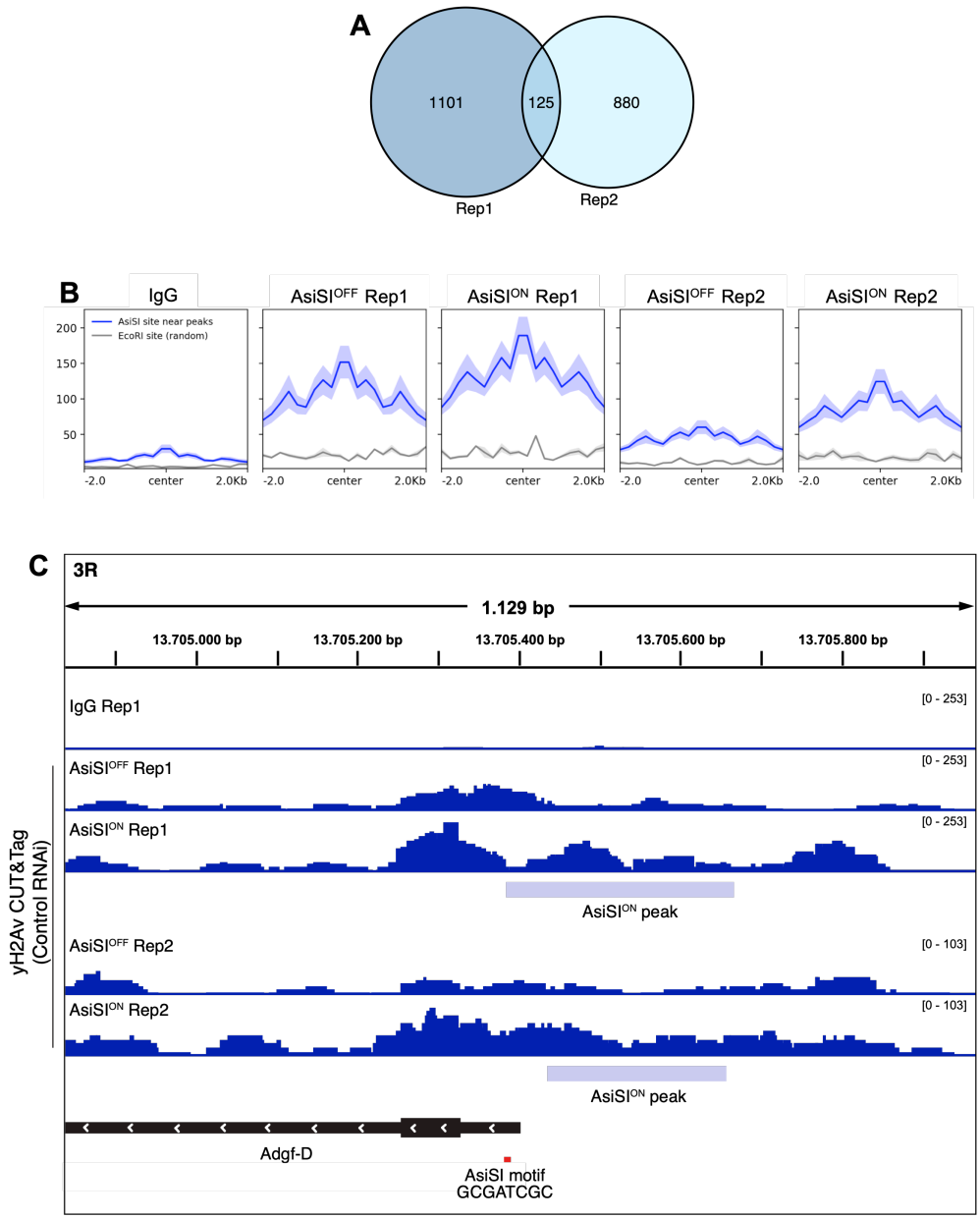


Figure 21 DNA damage analysis

(A) Overlap of γH2Av peaks called with MACS2 (-p 0.01) on AsiSI^{ON} over AsiSI^{OFF} control RNAi samples. Both replicates have an overlap of 125 DNA damage-specific peaks. **(B)** The sequence surrounding the 125 peaks from (A) were scanned for the presence of AsiSI recognition sites in a ± 3 kb vicinity yielding 35 locations. The regions were inspected with IGV to select sites of well recognizable γH2Av signal increase (8 regions) and over these candidate DNA damage sites the γH2Av bigwig score was plotted as an average profile centered on the AsiSI motif. As a negative control, average profiles were also plotted over random EcoRI recognition sites. Shaded area indicates the SE. **(C)** Example of a candidate DNA damage site, marked by increased γH2Av signals and the presence of a DNA damage-specific peak (light blue bar) in close vicinity to an AsiSI motif.

3.3 Gene expression analysis upon CHRAC-14 depletion

The CUT&Tag approach uncovered that CENP-A is increasingly incorporated at and around gene bodies or at distal intergenic regions which often consist of transposable elements or enhancers (Gilbert *et al.*, 2021; Panigrahi and O'Malley, 2021; Zhou *et al.*, 2020). Such sites underly strict regulation in regard to their accessibility status. The accessibility of genomic elements is (amongst other mechanisms) determined by the chromatin composition and structure present at these sites (cf. section 1.1.2). We, therefore, speculated that the native chromatin environment at genes or transposable elements is profoundly changed by the increased occurrence of CENP-A, which may be leading to altered gene expression patterns. In order to test this hypothesis, we performed total RNA-Seq in biological triplicates comparing gene expression patterns in CHRAC-14 depletion and control. Noteworthy, we used two different model systems to later be able to discover robust changes, which are independent of the cell type and specific for the CHRAC-14 depletion. However, this comparison due to time limitations, was not within the scope of my thesis. Nevertheless, I prepared the complete dataset by using S2 cells where CHRAC-14 was acutely depleted by RNAi and 4-6 h aged mutant CHRAC-14^{KG01051} embryos, with a disruptive P-element insertion at the *chrac-14* locus.

3.3.1 Validation of experimental conditions

Beforehand, CHRAC-14 reduction in S2 cells was validated for each replicate by Western blot (Figure 22 A, B) and qPCR (Figure 22 C). In embryos CHRAC-14 reduction could only be assessed by qPCR since our antibody works with S2 but not embryo protein extracts. The Western blot shows an average CHRAC-14 protein decrease of ~70% in S2 cells and according to qPCR CHRAC-14 was reduced by ~ 80%. In embryos, CHRAC-14 levels were reduced by ~60%. The depletion efficiency in the triplicates of each condition was consistent and significant and thus I proceeded with library reparation and sequencing.

3.3.2 Alignment and duplication statistics of total RNA-Seq experiment

The resulting reads were aligned with HISAT2 (D. Kim *et al.*, 2019; Kim *et al.*, 2015) in a stringent and splice variant sensitive manner to a rough masked version of the *Drosophila* BDGP genome, which prevents alignments to hardly quantifiable content such as low

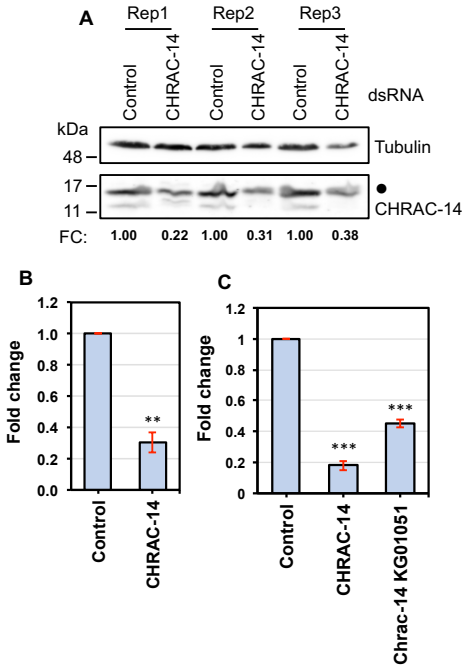


Figure 22 Validation of CHRAC-14 depletion in total RNA-Seq samples

For total RNA-Seq CHRAC-14 down-regulation was achieved by RNA interference in S2 cells (Brown as a control) and via a P-element insertion in the CHRAC-14^{KG01051} mutant fly line (Ore-R was used as a wild type control). **(A)** CHRAC-14 protein reduction was analyzed in RNAi-treated S2 cells by Western blotting. The dot marks an unspecific band representing recombinant MNase, which was used to digest chromatin during cell lysis. **(B)** Quantification of the CHRAC-14 Western blot signal shown in (A) normalized to tubulin. **(C)** RNA was extracted from RNAi treated cells and from 4-6 h staged mutant and control embryos to assess gene expression levels of CHRAC-14 with RT-qPCR using Rpl32 as an input control. ** = p-value < 0.01, *** = p-value < 0.001, n=3, one sample *t*-test.

complexity repetitive DNA or overabundant rDNA sequences (Adams *et al.*, 2000; Hoskins *et al.*, 2015; Kim *et al.*, 2015). The reported sequencing depth was appr. 15 – 25 mio reads per sample (Figure 23 A) and the overall alignment rate was below 50% (Figure 23 B). However, the multi mappers rate was low as well, indicating that repetitive and rRNA reads were successfully blocked. Next, reads were filtered for a decent mapping quality (-q 30) followed by duplicate removal. Since the duplication rate was appr. 12 – 18% (Figure 23 C) the final library size was reduced to appr. 5 – 9 mio reads per sample (Figure 23 D).

3.3.3 CHRAC-14 depletion causes differential gene expression

After the dataset was ‘cleaned’ as described above, it was subjected for transcriptome assembly and differential expression analysis with Cufflinks, Cuffdiff and cummeRbund (Goff *et al.*, 2021; Trapnell *et al.*, 2013, 2010). In the scope of this study, only the S2 cells data with

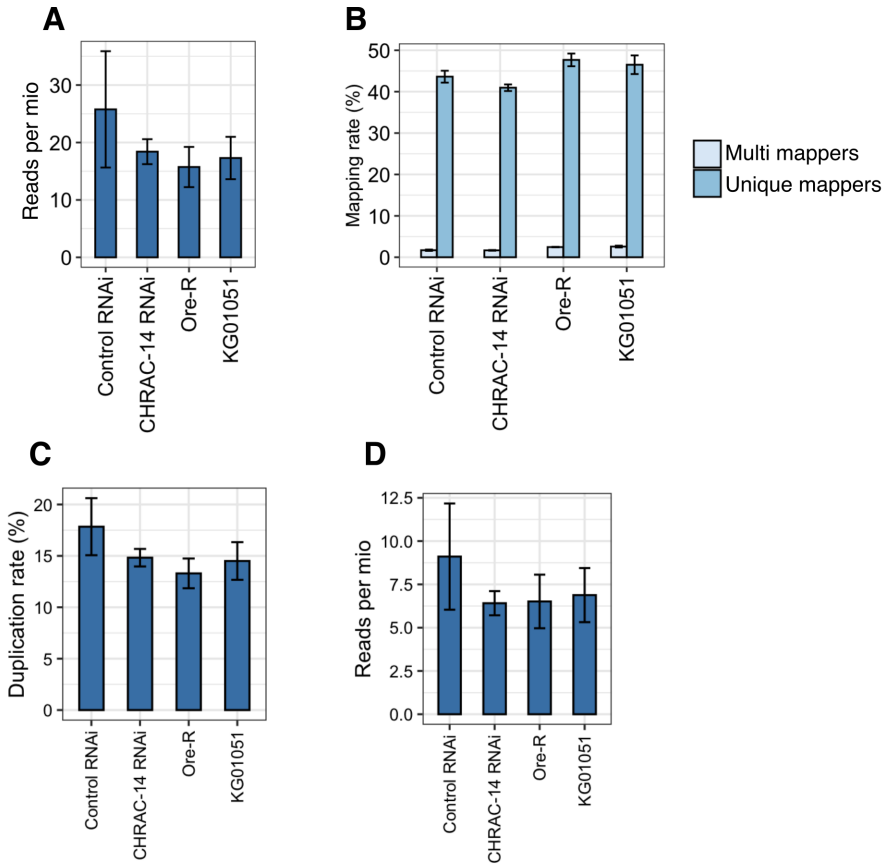


Figure 23 Technical parameters of RNA-Seq data output

Total RNA was purified from S2 cells after CHRAC-14 or Brown (control) RNAi treatment and from 4-6 h staged Ore-R or CHRAC-14 deficient embryos. rRNA was depleted and libraries were subjected for paired-end sequencing. Data is from biological triplicates. **(A)** Average sequencing depth for each condition **(B)** Average mapping rate for each condition **(C)** Average duplication rate for each condition **(D)** Final average estimated library size for each sample.

RNAi treatment was analyzed, since I primarily planned to compare the RNA-Seq data with the CUT&Tag experiments, which were conducted in S2 cells.

First, I assessed the variance of gene expression between samples to get an initial idea about the reliability of the experiment: The dendrogram depicted Figure 24 A shows that samples of the same knockdown condition cluster together as separate groups. However, CHRAC-14 RNAi samples show a high variance represented by greater distances between the respective data points in the MDA plot shown in Figure 24 B. The distance within the control group is smaller and samples seem to exhibit less variance in this condition. Figure 24 C depicts the expression levels of all detected genes and illustrates that CHRAC-14 knockdown causes various genes to be significantly differentially expressed showing a trend towards higher expression levels in the CHRAC-14 RNAi condition (above the diagonal line; Supplemental Table 6). The volcano plots in Figure 24 D and E, where significant genes are in red, also show more genes having a positive than a negative $\log_2$ fold change upon CHRAC-

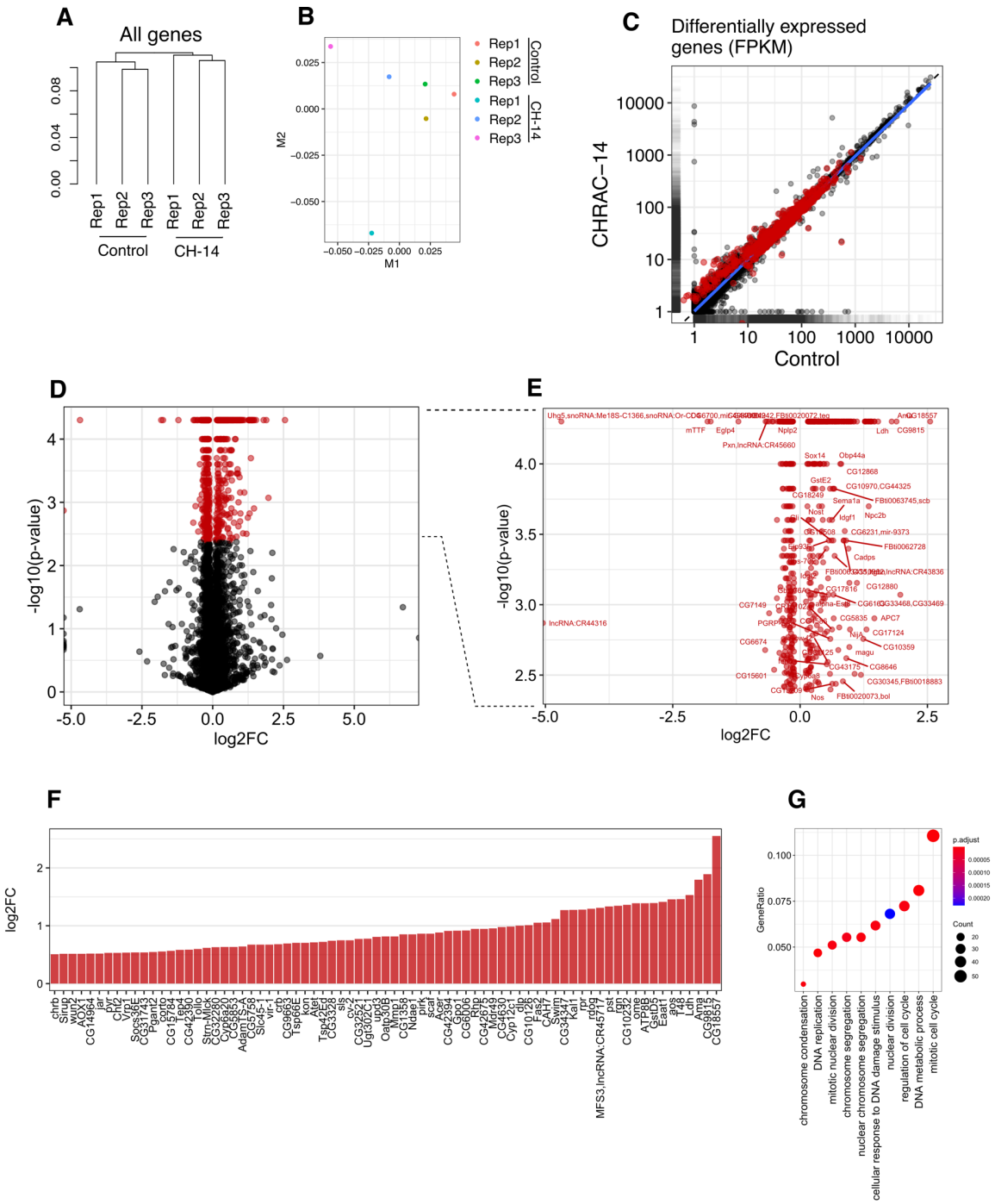


Figure 24 CHRAC-14 depletion causes differential gene expression

(A) Dendrogram of each sample considering all genes. **(B)** MDA plot of all samples considering all genes **(C)** Correlation of RNA expression levels (FPKM) comparing both RNAi conditions (significant RNAs in red). **(D)** Volcano plot depicting the correlation of significance and $\log_2$FC of all genes upon CHRAC-14 RNAi (significant genes in red). **(E)** Magnification of volcano plot shown in (D) depicting only significantly mis-expressed genes. **(F)** Histogram giving an overview of the strongest upregulated genes based on the expression level fold change upon CHRAC-14 knockdown. **(G)** Gene ontology term enrichment analysis (biological process) of differentially expressed genes. FC = fold change, CH-14 = CHRAC-14, FPKM = Fragments Per Kilobase Million.

14 RNAi. As can be deducted from the plots the fold changes show a maximum at around ± 2.5. Figure 24 E is a magnification of all significant genes to provide a brief overview of the most affected candidates. Figure 24 F ranks those genes with the largest fold change

independent of their p-value. To gain a better functional overview of the mis-expressed genes, a GO term enrichment analysis was conducted and Figure 24 G depicts the most significantly enriched biological process GO categories. The mis-expressed genes are overrepresented in pathways associated with the regulation of the mitotic cell cycle, nuclear division, chromosome segregation, DNA repair and DNA replication.

3.3.4 **A subset of misregulated genes exhibit increased CENP-A binding**

So far, the differential gene expression analysis seemed to confirm our hypothesis because CHRAC-14 reduction via RNAi caused mis-expression of genes. We then asked our main question: Are mis-expressed genes packed in chromatin with an increased number of CENP-A nucleosomes detected by CUT&Tag? To investigate this, the CUT&Tag CHRAC-14 specific peak data set (all hits, not only the top 20%) was intersected with the RNA-Seq hits to identify mis-expressed genes with elevated CENP-A levels.

Indeed, a subset of mis-expressed genes were amongst the hits that had extra CENP-A incorporated when CHRAC-14 was downregulated (Supplemental Table 7). The heatmap in Figure 25 A illustrates that those genes exhibit altered expression levels in the CHRAC-14 RNAi condition but that the replicates are not always equal. This was expected, since the MDA plot in Figure 24 B already indicated quite a variance between the replicate samples. In Figure 24 B the genes are ranked according to the fold change in gene expression and it seems that the majority of the genes are upregulated after CHRAC-14 depletion. Not only are more genes up than down regulated but also the degree of the expression change is more pronounced in the upregulated genes, which exhibit larger fold changes than the downregulated candidates. However, as we saw already for the total gene set discussed in the previous section 3.3.3 the fold changes of gene expression levels remain mild in general (below 2.5). Nevertheless, the p-values of most of the candidates are below $1e^{-3}$ and thus these hits can be considered to be mis-expressed very significantly. Interestingly, many of the candidates are also amongst the top 20% significant CUT&Tag hits showing increased CENP-A binding (genes marked in red in Figure 25 C). To demonstrate that the candidate misregulated genes globally show increased CENP-A binding the CUT&Tag $\log_2$ fold change of bigwig scores underlying peaks in these genes was calculated. Figure 25 D illustrates that the CENP-A $\log_2$ fold change in both replicates is significantly higher compared to γH2Av representing another histone variant. Of note, the result of replicate 2 is rather unreliable because – as mentioned before – the CENP-A coverage in the chromosome arms in the CHRAC-14 RNAi replicate 2 sample was very poor (cf. section 3.2.6). As exemplified in Figure 25 E the candidate gene CG9815, which shows strong correlation of p-values between CUT&Tag and RNA-Seq (Figure 25 C)

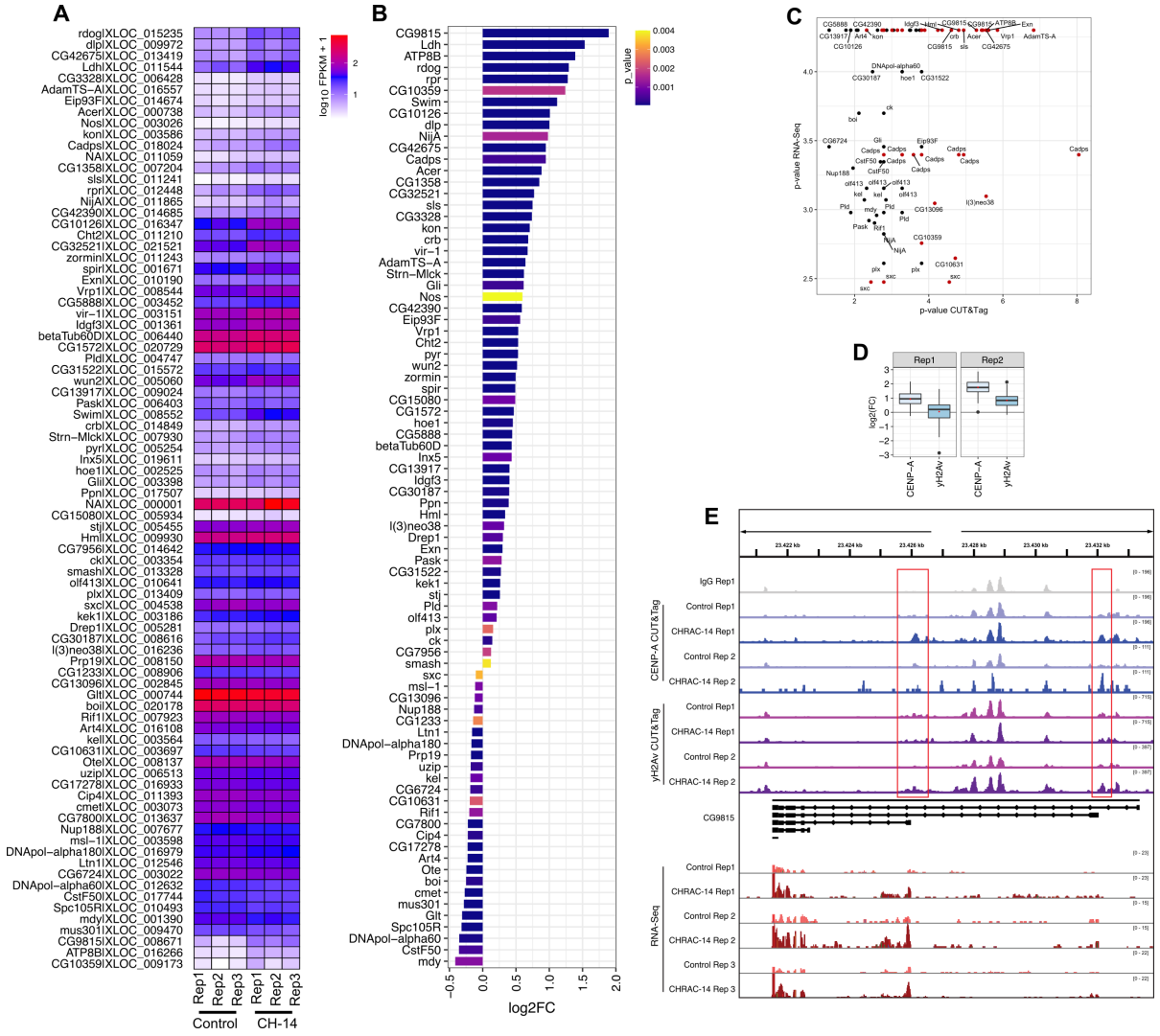


Figure 25 RNA-Seq to CUT&Tag comparison upon CHRAC-14 depletion

Differentially expressed genes upon CHRAC-14 knockdown were intersected with genes exhibiting CHRAC-14 RNAi specific peaks and common candidates were analyzed. **(A)** RNA-Seq derived expression levels (log_{10}FPKM + 1) per replicate of all differentially expressed genes overlapping with the CUT&Tag dataset. **(B)** Genes from (A) ranked according to the expression log_2 fold change with color coded p-value indication. **(C)** Scatterplot to visualize the correlation between the RNA-Seq vs. CUT&Tag p-values highlighting most affected genes in both datasets. Genes which are amongst the top 20% CUT&Tag candidates are marked as red dots. **(D)** Average log_2 fold change of CUT&Tag bigwig scores at CHRAC-14 RNAi specific CENP-A peaks in differentially expressed genes. **(E)** Example of a gene (CG9815) with CENP-A CUT&Tag peaks arising at transcriptional start sites and increased RNA-Seq read coverage after CHRAC-14 depletion. FC = fold change, CH-14 = CHRAC-14, FPKM = Fragments Per Kilobase Million.

accumulated CENP-A at two transcription start sites (red boxes) upon CHRAC-14 RNAi treatment (blue tracks) and at the same time the level of mRNA expression increased as detected by RNA-Seq (red tracks).

3.3.5 Validation of gene misexpression

In order to validate the RNA sequencing result, the expression levels of the most significant up and down regulated genes (Figure 25 B) showing an appropriate CUT&Tag score (Figure 25 C) were assessed in two additional biological CHRAC-14 RNAi replicates (Brown RNAi was used as a control). The CHRAC-14 knockdown in both replicates was efficient, however, the replicates of most other genes show great expression variances. Since only two replicates were generated, in combination, the statistical power is low and further replicates are needed, to determine a clear tendency of the genes to be up or down regulated or unaffected. Despite the low significance (* or n.s.), the tendency observed by qPCR matches with the tendency detected via RNA-Seq. Spc105 and mus301 indeed seem to be downregulated whereas CG9815 seems upregulated. For ATP8B and AdamTS-A only in replicate 2 an upregulation could be observed. Of note, replicate 2 showed a more efficient knockdown of CHRAC-14 (rep1 = 0.14 vs. rep 2 = 0.05), which might account for stronger misregulation of the genes in replicate 2.

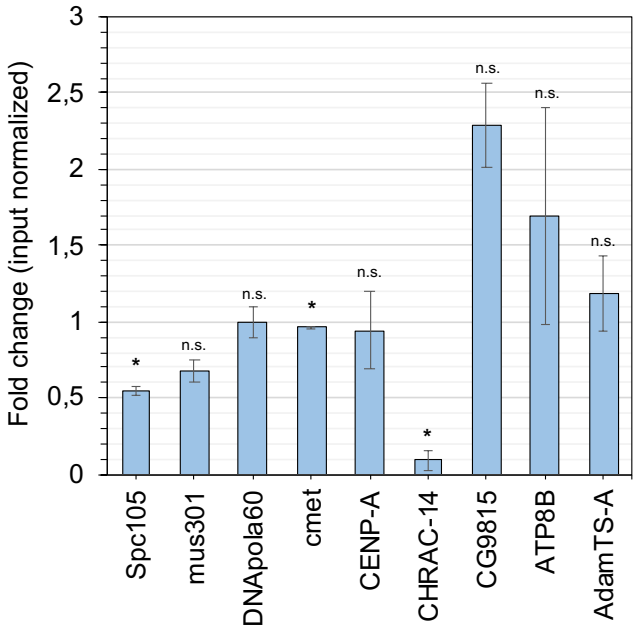


Figure 26 qPCR validation of mis-expressed genes

CHRAC-14 or Brown as a control were depleted in S2 cells by RNAi, total RNA was extracted and expression levels of indicated genes were quantified by qPCR. Values were normalized to Rpl32 and fold changes were calculated over the control RNAi values. One sample *t*-test, n.s. = not significant, * = p-value < 0.05, n=2

3.3.6 CHRAC-14 knockdown causes upregulation of transposable elements

My analysis so far suggests that certain genes, which bind more CENP-A upon CHRAC-14 knockdown, seem to be mis-regulated. Since several transposable elements also showed increased CENP-A binding by CUT&Tag, I asked if they could be mis-regulated as well. To investigate this, I started a preliminary analysis where I aligned the RNA-Seq reads to the custom PacBio assembly introduced earlier in section 3.2.3 (Chang *et al.*, 2019b). The resulting alignment data was then filtered, duplicates were removed and differential TE expression was examined with the TEtranscripts and DESeq2 software (Jin *et al.*, 2015; Love *et al.*, 2014). Indeed, the approach yielded a list of transposable elements with differential expression upon CHRAC-14 RNAi (Supplemental Table 8). The most frequent misregulated elements belong to similar repeat families, which were preferentially bound by CENP-A according to CUT&Tag: Gypsy, Jockey and Pao (Figure 27 A). Despite of matching repeat families, the elements as such are not the same ones compared to the top 20% CUT&Tag hits. Moreover, the degree of misregulation is rather mild as can be judged from the low $\log_2$ fold change values (Figure 27 B). Of note, Gypsy-12_DVir-LTR represents an exception since it seems to be strongly downregulated ($\log_2$FC = -6.25) whereas all the other elements are mildly upregulated (except for BEL_I-int, which is the second of two downregulated elements). A subset of the observed misregulated repeats can be found in centromere contigs (red annotations in Figure 27 C) and *HeT-A* was detected as a telomeric misregulated element (green annotation in Figure 27 C).

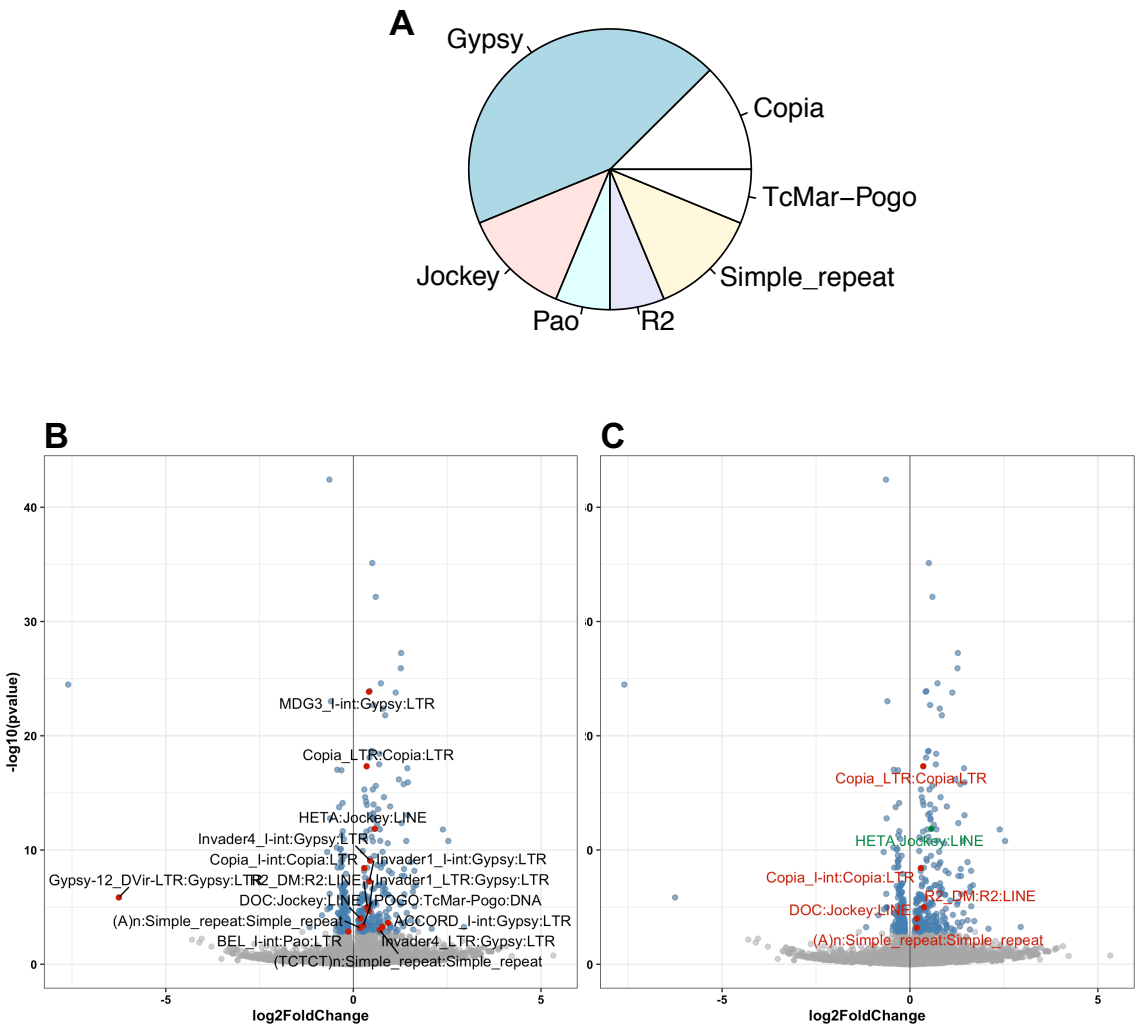


Figure 27 Misregulation of repeats upon CHRAC-14 knockdown

RNA-Seq reads were aligned to the centromere islands-containing PacBio assembly and expression levels of repeats were analyzed. **(A)** Overview of differentially expressed repeat families **(B)** Depiction of all differentially expressed repeats (red dots) in the context of the complete dataset (significant genes in blue, non-significant genes in grey). **(C)** Volcano plot from (B) highlighting differentially expressed centromeric (red dots) and telomeric (green dot) repeats.

4 Discussion

4.1 Identification of new CHRAC-14 and CENP-A interactors

4.1.1 Technical evaluation of the mass spectrometry approach

Immunoprecipitation followed by mass spectrometry analysis of CENP-A and CHRAC-14 identified a range of interaction candidates including known interacting factors (Figure 10 and Supplemental Table 1). The important CENP-A interaction partners CENP-C and CAL1 (cf. section 1.3) were not detected in this experiment. CENP-C was also missing from another mass spectrometry approach in our lab (Demirdizen *et al.*, 2019) whilst it was detected by Barth *et al.* (2015). These discrepancies are likely caused by technical differences of preparing the lysate for the immunoprecipitations: Bath and colleagues extracted nuclei and used a different protocol for chromatin solubilization with MNase than what has been done in this study. Their protocol most likely leads to a more efficient nucleosome isolation and increased recovery of very low abundant chromatin-associated proteins such as CENP-C. In Demirdizen *et al.* no MNase digest was used, which then probably led to equally weak CENP-C enrichment. Moreover, Demirdizen *et al.* analyzed CENP-A interaction candidates by crosslink-immunoprecipitation and the fixation step probably improved the isolation of CAL1 together with CENP-A. Conceivably, since in my study, I employed a different lysis and chromatin solubilization protocol and conducted native immunoprecipitations, certain interaction partners of low abundance or transient binding were not efficiently enriched in contrast to other strongly associated proteins such as canonical histones.

The CHRAC-14 pulldown successfully enriched its best-known interacting partner CHRAC-16 but other known interaction partners of the multi-subunit chromatin remodeling complexes (CRC) ATAC or CHRAC were not detected. From personal communication with the histone mass spectrometry specialist Prof. Axel Imhof (LMU Munich) we know, that the interaction between CHRAC-14 and its partner CRCs is far from robust and very variable, when analyzed by immunoprecipitation and mass spectrometry. The reason why CHRAC-14 is such a volatile interacting protein is unknown, but there is the possibility that this reflects a genuine biological property of CHRAC-14 to interact with different CRCs depending on the cellular context. Indeed, DNA damage treatment in this study, caused CHRAC-14 to interact with a multitude of DNA damage-specific factors (Figure 10 B).

In summary, my approach successfully led to the identification of well-known, strongly enriched interaction candidates and the DNA damage-specific set of CHRAC-14 interactions strengthens our assumption, that CHRAC-14 is a protein with DNA damage functions.

4.1.2 CHRAC-14 and CENP-A interact with each other independently of DNA damage

Importantly, I confirmed here by mass spectrometry, that CHRAC-14 is interacting with CENP-A in S2 cells and in addition the interaction was independent of DNA damage (Figure 10). This broadens our existing knowledge about this interaction, which until now, has been solely observed by immunoprecipitations from embryo extracts upon DNA damage or with recombinant proteins *in vitro* (Mathew *et al.*, 2014). We can now conclude, that CHRAC-14 is a universal CENP-A regulator, which is in line with data showing elevated CENP-A levels caused by CHRAC-14 depletion already under physiological conditions (Mathew *et al.*, 2014). It is an open question, why in Mathew *et al.* and in my study (cf. section 3.1.5) the interaction of CHRAC-14 and CENP-A is difficult to detect without mass spectrometry. It might be that the interaction is too transient and thus hardly detected by a less sensitive method such as Western blotting. Other than that, maybe only a subset of CHRAC-14 is interacting with CENP-A or the other way around, which also leads to a low number of targets in the eluates, falling below the detection limit of the immunoblot. Moreover, in Mathew *et al.* 4 mg of embryo extract from CENP-A-GFP containing animals was used for the GFP-Trap immunoprecipitation. Embryos contain a lot of maternally deposited histones for nuclei amplification during early development and only cycle between S and M-phase, which requires constant CENP-A loading and regulation on newly replicated centromeres (Horard and Loppin, 2015). This system thus might provide sufficient material for an immunoprecipitation and might be more suitable to detect a presumably faint interaction between CENP-A and CHRAC-14. That in embryos, the interaction is detected upon DNA damage only might reflect biological differences of early embryos and cultured S2 cells. It will be important for future experimental designs, to recognize this difference and carefully interpret aspects of CENP-A regulation by CHRAC-14 in dependency of the system used.

Lasty, a central problem of detecting CHRAC-14 is, that us and other groups could not manage to generate a reliable and well working antibody, which additionally hampers an efficient CHRAC-14 detection in Western blots or immunostainings.

4.1.3 CHRAC-14 - an ‘all-rounder’ CRC subunit for CENP-A regulation?

Polymerase ε

With the mass spectrometry screen, I could confirm a robust interaction of CHRAC-14 with CHRAC-16 and Mes4, which is independent of DNA damage, as well. As aforementioned, a Mes4/PolE4 with CHRAC-14 interaction in *Drosophila* has only been documented via an *in vitro* yeast 2-hybrid screen (Shokri *et al.*, 2019) and here we verify this interaction in *Drosophila* S2 cells. Hence, this strengthens the assumption that, like in human, both proteins are interacting and are subunits of pol ε in *Drosophila* (Marygold *et al.*, 2020). This notion could be

fortified, when MES4 was studied in greater detail in the scope of a Master thesis by Pia Freidel in our lab. She confirmed by reciprocal immunoprecipitations, that Mes4, indeed, interacts with CHRAC-14 and showed that Mes4 RNAi in S2 cells led to reduced replication efficiency and increased cytokinesis failure, probably due to persistent, unresolved replicative structures (Master Thesis Pia Freidel, data not shown). In my mass spec screen CHRAC-14 strongly interacted with the chaperones CHRAC-16 and Mes4 independent of the experimental conditions, suggesting that CHRAC-14 prevalently acts within the CHRAC and Pol ε complexes.

XNP/ATRX

Only upon DNA damage induction, interactions with other chromatin remodeling factors (and also other proteins) was detected. This might indicate that CHRAC-14 associates with different CRCs depending on the cellular chromatin remodeling 'demand' (replication in S-phase, chromatin remodeling during transcription etc.). DNA damage or generally stress-specific chromatin remolding complexes containing CHRAC-14 detected in this screen include XNP/ATRX and pBAF, and both have not yet been associated with CHRAC-14 in the literature.

XNP is the *Drosophila* ATRX ortholog, which is responsible for H3.3 loading at telomeres and pericentric heterochromatin (cf. section 1.1.1). Interestingly, ATRX-mediated incorporation of H3.3 seems to be limited by the chromatin protein and histone chaperone DEK (Ivanauskiene *et al.*, 2014), and DEK in turn was interacting with CENP-A in my assay independent of DNA damage induction. In line with that, XNP and DEK have been previously detected as CENP-A interaction partners in Demirdizen *et al.*, which fortifies a genuine connection between the proteins. Intriguingly, next to CHRAC-14 depletion also DEK depletion was found to lead to increased CENP-A loading (PhD Thesis of Dr. V. Mathew), implicating that DEK not only limits H3.3 but also CENP-A incorporation and might cooperate with CHRAC-14 and XNP.

Moreover, both ATRX and DEK have been shown to promote DNA repair (Juhász *et al.*, 2018; Kavanaugh *et al.*, 2011) and this suggests, that XNP/DEK/CHRAC-14 regulate CENP-A especially during DNA damage or generally during stress conditions.

pBAF

Furthermore, as mentioned above, three subunits of the SNF/SWI pBAF chromatin remodeling complex were found in the damage-specific CHRAC-14 IP and pBAF seems to contribute to DNA damage repair by limiting transcription at DSBs (Kakarougkas *et al.*, 2015, 2014). Given that these chromatin remodeling complexes are DNA damage factors, it seems feasible, that at least ATRX and pBAF subunits occurred as damage-specific CHRAC-14

interaction partners and this might suggest that ATRX-CHRAC-14 or pBAF-CHRAC-14 together with DEK regulate CENP-A chromatin occupancy during DNA damage.

So far, CHRAC-14 has only been described as a subunit of CHRAC, Pol ε and ATAC and it would be very intriguing to test by immunoprecipitation and Western blotting, if CHRAC-14, indeed, interacts with XNP/ATRX or pBAF upon DNA damage or upon stress in general leading to the discovery of new CHRAC-14 containing CRCs. Other than that, one could examine if the depletion of any of these factors causes CENP-A accumulation at DSBs similar to what has been observed upon CHRAC-14 RNAi (Mathew *et al.*, 2014).

4.1.4 **Other candidate CHRAC-14-associated CENP-A regulators**

Not only chromatin remodeling complexes were interacting with CHRAC-14 but also other factors with regulatory potential such as the E3-ubiquitin ligase Bre1 and the SUMO protease Velo. Since CENP-A levels and loading is well known to be regulated by ubiquitinylation (cf. sections 1.3.3 and 1.4.4) it is intriguing to speculate that CHRAC-14 might cooperate with Bre1 to mediate CENP-A ubiquitinylation upon DNA damage and thereby regulating CENP-A abundance in damaged and/or native chromatin. This would be conceivable because Bre1 is a DNA replication and repair factor and mediates ubiquitinylation of H2B, which in turn acts as an important signaling residue (Liu *et al.*, 2021; Zheng *et al.*, 2018).

Similar to ubiquitinylation, SUMOylation of CENP-A has been reported in yeast (Ohku *et al.*, 2020; Ohkuni *et al.*, 2018, 2016). Generally, SUMO-mediated centromere regulation is an established pathway, since multiple studies reported that deSUMOylation of centromere and kinetochore factors (but not of CENP-A) by SENP6 is important for centromere function (Fu *et al.*, 2019; Liebelt *et al.*, 2019; Mitra *et al.*, 2020). Intriguingly, here and in Demirdizen *et al.* Smt3/SUMO has been detected to interact with CENP-A in *Drosophila*. The fact that upon genotoxic stress CHRAC-14 interacts with the SUMO protease Velo leads us to speculate that deSUMOylation could be a regulative mechanism for CENP-A loading during DNA damage.

However, since these factors are not associated with CENP-A and CHRAC-14 at the same time, it seemed more promising to focus on the only factor, which commonly interacted with both proteins and which was promoted upon DNA damage: Casein kinase 2.

4.1.5 **CK2 interaction with CHRAC-14 and CENP-A**

CK2α was binding to CENP-A and CHRAC-14 with and without DNA damage treatment in contrast to CK2β, which bound CHRAC-14 damage-specifically (interacting with CENP-A in both conditions). With the detection of CK2 we successfully addressed our initial aim to identify

a common interaction partner of CHRAC-14 and CENP-A during DNA damage. Hereby, CK2β seemed to interact with CHRAC-14 DNA damage-specifically, which we interpreted as a potential DNA damage-specific regulation opportunity of CHRAC-14 towards CENP-A.

To investigate this further, I initially attempted to corroborate the observed interaction of CK2β-V5-His with CENP-A and CHRAC-14 in control and damage induction by V5 reciprocal immunoprecipitations. Unfortunately, I could not show an interaction with that approach. An explanation could be, that the interaction of CK2β with either of the proteins is transient and – in contrast to mass spectrometry – a Western blot is too insensitive to detect minor amounts of CENP-A or CHRAC-14 in the eluates. Indeed, the $\log_2$ fold changes (Supplemental Table 1) and thus the interaction strength of CK2β was only between 0.3 and 0.5 and was considerably weak in comparison to strong interactors such as CHRAC-14 ($\log_2$FC ~ 2.0). Of note, CK2 was amongst the interaction candidates in Demirdizen *et al.,* which further indicates that the proteins are interacting. Moreover, CENP-A was shown to be regulated by CK2 recently (Huang *et al.*, 2019). Despite their findings, it is it still important to optimize the reciprocal immunoprecipitation to detect a physical interaction with CENP-A and CHRAC-14 (which is also missing in the study of Huang *et al.* for CENP-A) in order to back up our mass spectrometry results and CK2 related data. One possibility is to test, if different tags are more successful. Second, one could try to use a CK2 antibody and enrich for endogenous CK2, since the exogenous protein might be non-functional or not interacting with CENP-A and CHRAC14. It could be also tested, if an interaction with CK2α instead of CK2β is generally more successful, since we have no insight into which subunits or domain mediates an interaction to CENP-A and CHRAC-14. Lastly, an endogenously tagged CK2β fly line was designed and generated by Qidong Fangjing Biological Technology and immunoprecipitations from embryo or fly extracts could be tested for an interaction.

4.2 CHRAC-14 and CENP-A as CK2 substrates

The *in vitro* kinase assays followed by mass spectrometry analysis of the phosphorylated proteins confirmed our assumption that CENP-A and CHRAC-14 are previously unknown CK2 substrate *in vitro*. This was an important finding, since it implicates that CENP-A and CHRAC-14, indeed, interact with CK2 as new substrates.

CHRAC-14

CHRAC-14 showed a clear phosphorylation signal in the kinase assay (Figure 11 B). In comparison to the positive controls CENP-A was fainter despite equal protein amounts. This suggests that an efficient phosphorylation by CK2 might be dependent on additional factors such as modifications or auxiliary proteins, which are missing in a recombinant protein assay.

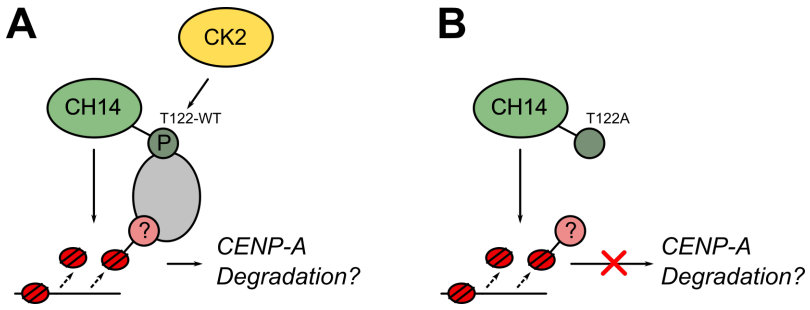


Figure 28 CHRAC-14-T122P might promote CENP-A degradation

(A) CHRAC-14 phosphorylation at T122 by CK2 might promote the degradation of modified CENP-A after chromatin eviction by mediating the binding to a third regulative factor such as a SUMO protease **(B)** Lacking T122 phosphorylation leads to the accumulation of evicted and modified CENP-A.

The signal, nevertheless, was strong enough to reveal residue T122 as the CK2 phosphorylation site on CHRAC-14 by mass spectrometry. Interestingly, the phospho-site pT122-A-E-E perfectly fits the reported minimum CK2 consensus motif (S/T)-x-x-(E/D/pS/pT) (Meggio and Pinna, 2003) and implicates that CHRAC-14 is a genuine CK2 substrate. It is to note, that the motif does not seem to be conserved in orthologs of other species (data not shown) so it remains to be addressed if CK2 phosphorylates CHRAC-14 orthologs in other species than *D. melanogaster.* Generally, up to this point, it is unknown, if the phosphorylation of CHRAC-14 takes place *in vivo.* So far, my study indirectly suggests, that CK2 phosphorylation of CHRAC-14 is evident in *Drosophila* cells and has a function in regulating CHRAC-14 and CENP-A. Here I showed, that the phosphorylation status of CHRAC-14 at T122 might influence CENP-A post translational modification.

As presented in Figure 13, in cells that express a non-phosphorylatable CHRAC-14 mutant (T122A) a new CENP-A band is formed. Even though this finding is to be confirmed with further replicates it might indicate that a post translationally modified CENP-A version is stabilized when CHRAC-14 remains unphosphorylated, mimicking absence of CK2 activity. The rather large size shift of approximately 15 kDa suggests that CENP-A is modified with a peptide like SUMO or Ubiquitin. Previously, work from our lab demonstrated that ubiquitinylation of CENP-A in *Drosophila* leads to a larger apparent size shift of ~25 kDa than the here observed 15 kDa shift (Bade *et al.*, 2014). Therefore, it might be more likely, that CENP-A is SUMOylated instead of ubiquitinylated. This is especially imaginable, since as aforementioned, in my study and in Demirdizen *et al.,* Smt3/SUMO was detected as a CENP-A interacting protein and in yeast CENP-A is SUMOylated, as well (Ohku *et al.*, 2020; Ohkuni *et al.*, 2018, 2016)

It is tempting to speculate (Figure 28), that CK2 phosphorylates CHRAC-14 at T122, which then causes CHRAC-14 to mediate an interaction of CENP-A with the SUMO protease Velo, that was detected in this study. In the absence of T122-phosphorylated CHRAC-14, CENP-A-SUMO might be stabilized due to lacking deSUMOylation and becomes detectable as shown in the Western blot in Figure 13. It is possible that SUMOylation of CENP-A blocks its degradation by counteracting ubiquitinylation, a pathway which has been described in different studies (Cartier *et al.*, 2019; Ramachandran *et al.*, 2015; Rott *et al.*, 2017). In line with that is the study by Ohku *et al.* (2020), which showed that in yeast SUMOylation of CENP-A promotes its centromeric deposition, indicating that SUMO could act as a stabilizing modification of chromatin-bound CENP-A. On the contrary, other studies showed, that CENP-A SUMOylation mediated ubiquitin directed proteolysis (Ohku *et al.*, 2020; Ohkuni *et al.*, 2016, 2018). This suggests, that SUMOylation of CENP-A probably has divers roles.

To gain a better understanding, it is important to characterize the modification by mass spectrometry analysis of the modified CENP-A band. Next, it could be tested, if a depletion of the SUMO protease Velo in wild type cells would lead to the formation of a similar shifted CENP-A band like in CHRAC-14-T122A cells. Furthermore, it would be interesting to characterize by IF and microscopy combined with CUT&Tag, where this modified version of CENP-A is localizing (centromere vs. chromosome arms vs. nuclear lumen) to locally pinpoint its biological function. However, this would require the generation of a modification-specific CENP-A antibody. In summary, the detection of larger, presumably modified version of CENP-A in a CHRAC-14 cell line carrying a non-phosphorylatable CK2 motif is an important finding of this study and probably provides a central regulative handle, connecting CK2, CHRAC-14 and CENP-A.

CENP-A

Like CHRAC-14, CENP-A is a CK2 substrate. In my study a truncated version of CENP-A showed a ^{32}P signal in the kinase assay (Figure 11) and in line with that Huang *et al.* shortly after published, that CK2 phosphorylates CENP-A at the N-terminal residue S20. The authors find that this phosphorylation leads to the proteasomal degradation of exogenously expressed pre-nucleosomal or ectopic, but not of centromeric CENP-A. Experiments in this thesis, however, yielded contradicting results. Whilst Huang *et al.* find no effect on the CENP-A protein levels upon CK2 depletion, in my approach, CENP-A levels are drastically decreased (Figure 12). This result was highly reproducible and specifically caused by the depletion of CK2 via RNAi, which was conducted using different dsRNAs. We can thus reliably claim, that CK2 phosphorylation of CENP-A has a stabilizing effect and does not lead to proteasomal degradation. I propose that the opposing outcomes of CENP-A phosphorylation by CK2 arise because we used different cell lines. In my experiments I analyzed endogenous CENP-A levels

in wild type S2 cells, whereas Huang *et al.* used a cell line with stably transfected pMT-GFP-CENP-A. Therefore, their experiments rely on exogenous CENP-A. When I depleted CK2 in my CENP-A-V5 transgenic cell line, only endogenous CENP-A was decreasing (Figure 12), which clearly demonstrates that exogenous and endogenous CENP-A are regulated differentially by CK2. Nevertheless, the model proposed by Huang *et al.* might still apply: CK2 promotes ectopic CENP-A-S20P depletion and centromeric CENP-A-S20P is not affected since it is shielded from the degradation machinery through the association with proteins such as CAL1. Since Huang *et al.*, according to personal reassurance, did not monitor effects of endogenous CENP-A-S20P in wild type cells, my study rather complements than contradicts their findings: Centromeric GFP-CENP-A-S20P is resistant to proteasomal degradation and in addition, endogenous CENP-A depends on S20 phosphorylation for its stability. Nevertheless, exogenous and endogenous CENP-A levels seem to be affected differently by CK2 phosphorylation, which indicates that different regulative pathways are at play. Indeed it is known, that over-expressed CENP-A in *Drosophila* is loaded by the NuRD complex (Demirdizen *et al.*, 2019) and endogenous CENP-A relies on CAL1 (Chen *et al.*, 2014) and this already clearly implicates that endogenous and exogenous CENP-A are handled by different pathways.

As a last point it is to mention, that the reduction of endogenous CENP-A upon CK2 depletion could be an indirect effect and might not be caused by the lack of phosphorylation. CK2 is a pleiotropic kinase targeting a vast set of substrates involved in many different pathways including cell division, chromatin organization and histone ubiquitinylation (Rusin *et al.*, 2017). It cannot be ruled out, that disturbing these processes by CK2 knockdown results in CENP-A decrease. In the future, the effect of lacking CENP-A S20 phosphorylation should be directly addressed by examining the protein stability of endogenously mutated CENP-A-S20 to a non-phosphorylatable alanine residue.

As a side note, we initially expected, that CK2 regulates CENP-A and CHRAC-14 during DNA damage. However, I could not verify a DNA damage-specific interaction and furthermore, CK2 effects on CHRAC-14 and CENP-A are evident without DNA damage. This led me to assume, that CK2 is a general regulator of both proteins.

4.3 CENP-A profiling in CHRAC-14 depleted DIvA cells

4.3.1 DIvA system as a molecular tool for CUT&Tag

Another overarching aim of this study was to trace CENP-A genome-wide upon CHRAC-14 RNAi and during DNA damage. For that, a stable cell line, in which DNA damage can be induced at defined loci via the expression of the restriction enzyme AsiSI was generated. As demonstrated in Figure 14, the DIvA system was functional. Monitoring γH2Av

via Western blotting verified, that DNA damage can be efficiently induced after 3 h of auxin depletion combined with mild AsiSI induction with 50 µM $CuSO_4$. Importantly, the damage induction was directly caused by restriction digestion through AsiSI, since the control cell line, which carries an empty vector, shows no DNA damage onset even at later time points. Other than that, it was important to note, that AsiSI DNA damage induction is more effective in the presence of a C-terminal tag, probably because the spatial conformation and bioactivity is less impaired when tagged at this position.

As expected from the functional DNA damage response in whole protein lysates, fluorescence microscopy confirmed, that AsiSI is stabilized and re-localizes to the nucleus in the $AsiSI^{ON}$ condition, where it is usually accompanied with strong γH2Av signals. A limitation to this system is, that nuclear AsiSI localization co-occurring with strong γH2Av signal is only triggered in a subset of cells (Figure 14 B). This could be due to an incomplete transfection efficiency so that only some cells contain the DIvA machinery. However, since all cells show AsiSI signal caused by leaky expression in the cytoplasm, it seems that all the cells do express AsiSI. Possibly, the timing of the auxin depletion response and $CuSO_4$ induction depends on the cell cycle stage of the cells and is, therefore, asynchronous. In that case, an asynchronous DNA damage response is expected to 'blur' bulk CUT&Tag sequencing results. It could be thus considered to conduct single cell sequencing in order to gain well resolved data in the future.

Moreover, I validated, that CHRAC-14 can be efficiently depleted by RNAi in the DIvA cells. As shown in Figure 14 C the knockdown is strong but yet incomplete, since the CHRAC-14 band was still faintly visible in the CHRAC-14 knockdown samples. This represents another limitation to our experimental approach, since it probably leads to weak phenotypes because a small amount of functional CHRAC-14 is still present in the cells. Using CRISPR KO cell lines or a rapid inducible protein depletion system instead of RNAi interference could be advantageous in the future. Nevertheless, the DNA damage response was fully functional and CHRAC-14 levels were robustly reduced by RNAi, so we moved forward with the system and employed the DIvA cells for the CUT&Tag CENP-A profiling.

4.3.2 **CUT&Tag data quality**

The alignment rates, which were at average 90.5% for the whole genome assemblies (Figure 16 A, BDGP and PacBio) can be considered to be rather on the lower limit, since alignment rates of 95% - 100% are desirable in order to make full use of the read output. The reason for the lower rates observed here was most likely the stringent alignment parameters used, where only concordant pairs and fragments below 700 nts are retained. Other than that, the S2 cells might have undergone genomic alterations such as chromosomal rearrangements,

which is a typical feature of cultured S2 cells (Lee *et al.*, 2014) and which limits the mapping of concordant pairs to the original, unaltered BDGP genome.

Next it is to mention, that a robust alignment rate for the IgG control sample was expected. IgG just like every protein binds non-specifically to other proteins to a certain degree. Moreover, pA-Tn5 in the absence of a chromatin bound antibody has the capability to bind and cut DNA non-specifically. Hence the IgG control at the same time monitors ATAC-Seq-like (assay for transposase-accessible chromatin sequencing) un-tethered pA-Tn5 background activity at accessible chromatin sites. Both scenarios yield tagged and sequence-able fragments, which will align successfully to the refences (Kaya-Okur *et al.*, 2019). It is necessary in the future to optimize the pA-Tn5 dosage, in order to minimize background tagmentation.

Similarly, as expected (cf. section 3.2.2), we see unspecific tagmentation of *E. coli* DNA yielding reads that map to the *E. coli* reference. We can only speculate that this could be a typical characteristic for working with low abundant epitopes like CENP-A or γH2Av. Low abundant epitopes are quickly saturated with pA-Tn5 leading to an excess of unbound Tn5 in the reaction. Unbound Tn5 might then bind chromatin non-specifically as mentioned above, and tags random DNA loci at a higher rate than observed with more abundant epitopes such as H3K27me3. Again, adjusting the pA-Tn5 concentrations might reduce high background cleavage of *E. coli* DNA. It could be advantageous to sequence a more abundant epitope control in parallel to CENP-A and γH2Av samples and compare the alignment rates to the *E. coli* genome in order to test this.

Lastly, it makes biological sense that the CENP-A data contains more multi mappers than γH2Av and IgG since the prime CENP-A binding sites are repetitive sequences at the centromere (Allshire and Karpen, 2008). These are usually multi mappers because it is impossible to precisely distinguish, which unit within the repetitive element served as the original template. It was intriguing to note, that the multi mapper rate increased in the CENP-A CHRAC-14 RNAi samples. This could suggest an increased CENP-A binding rate at the centromere in this condition, which in concordance with our CHRAC-14 phenotype of elevated CENP-A protein levels and IF signals and provides a first indication of successful CENP-A CUT&Tag.

The fact that CENP-A and γH2Av are low abundant epitopes might also be the cause of the high duplication rates reflected in Figure 16 C. It is possible that CENP-A and γH2Av CUT&Tag from 2 mio nuclei generated only a sparse number of tagged fragments and since 14 amplification cycles were used in the subsequent library preparation, the input was probably overamplified, yielding many PCR duplicates. On the other hand, it is expected to a certain extent, that libraries based on targeted fragmentation by restriction digestion through endonucleases, contain similar fragments at a higher rate than approaches using random

fragmentation such as sonication (FastQC online help platform https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3AnalysisModules/). Such biological duplicates can be misinterpreted as PCR duplicates whereas they represent accurate footprints of the proteins, the Tn5 was recruited to. Interestingly, the duplication rate decreases in CENP-A CHRAC-14 RNAi samples and it is tempting to speculate that increased CENP-A incorporation into the genome yields more input fragments for the library PCR, which might reduce the level of PCR duplicates in these samples.

The high duplication rates observed in this experiment is probably caused by a combination of the above-described circumstances and could be reduced by optimizing and lowering the amplification cycle number during the library preparation. However, if the experiment is further and well optimized in the future and a high rate of technical PCR duplicates can be ruled out, it is accepted in the field to keep duplicates for the analysis assuming that they represent biologically similar fragments rather than technical duplicates (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3AnalysisModules/). For the data mapped to the custom PacBio genome, which serves as the basis for repeat binding analysis we thus kept the duplicates since at least a certain proportion probably reflects biological duplicates of repetitive fragments. It is to mention, that after filtering and deduplication at average 9 mio reads were left, which is an amount well suitable for my analysis and is comparable with the read numbers in the CUT&Tag bench mark publication (~8 mio reads, Kaya-Okur *et al.*, 2019).

In addition, the observed CUT&Tag fragment size distribution (Figure 17) met our expectations and demonstrates that pA-Tn5 fragmented the chromatin successfully. The fragment lengths peak at ~170 nt and ~360 nt most likely representing CENP-A and γH2Av nucleosomal ladders (including linker DNA) and is comparable to a pattern in other publications such as Li *et al.* (2021). The first strong peak at ~65 nt is prominent because the Tn5 is known to also cut at the nucleosome surface or at linker DNA (Zheng *et al.*, 2020), yielding approximately half of a nucleosome footprint. The periodicity of 10-nt fragment increments reflects typical Tn5 cutting pattern and successful tagmentation (Zheng *et al.*, 2020). Also, IgG exhibits a similar profile to the histone variants, again indicating that unspecific binding of IgG or pA-Tn5 to nucleosomes occurred.

The CUT&Tag metrics implicate, that the technique worked successfully but certainly, the data interpretation of this CUT&Tag approach requires solid validation with more replicates, since only two replicates were included in this study. This is aggravated by the fact that one important sample (CENP-A_C_CH14_Rep2) exhibits almost no coverage particularly outside of the centromere (Figure 17 and Figure 20 D), which serves as the primary CENP-A binding

site. Thus, for the genome-wide CENP-A binding analysis only replicate 1 could be used, limiting the reliability of the presented results. Nevertheless, we do observe commonalities with the RNA-Seq data set (cf. section 3.3.4) and with CENP-A ChIP data from Chang *et al.* (2019a), providing at least a considerable level of confidence (discussed in the next section).

In summary we are aware that our experimental approach had technical limitations of various sources: Asynchronous DNA damage response, incomplete CHRAC-14 depletion and low abundant protein targets for CUT&Tag yielding high duplication rates and tagged *E. coli* DNA. Moreover, the lack of a sufficient number of replicates requires cautious data interpretation. Nonetheless, I could successfully generate CUT&Tag datasets that are conform to the current standard of this technique and of sufficient read depth, allowing us to address our aim of CENP-A chromatin tracking.

4.3.3 **CENP-A binds transposable elements at the centromere and genome-wide**

The first important observation was that via CUT&Tag, CENP-A was detected at the recently established *Drosophila* centromere-resident non-LTR/LINE retroelement *G2/Jockey-3* (Chang *et al.*, 2019b)(Figure 18 A). Significant CENP-A peaks were found within centromere 2 and 3, where simple satellites, LTR retrotransposons of the Gypsy family and non-LTR elements of the Jockey family were bound by CENP-A. In this dataset, which was generated by overlapping the CENP-A control RNAi replicates, no significant peaks in centromeres of chromosome 4 or the sex chromosomes were detected. Possibly, each replicate individually contains also peaks at these other centromeres but were not reported here because I selected only the most reproducible hits by overlapping both replicates. The centromeric region of chr2^{Capri} in Figure 18 B, which contains the *G2/Jockey-3* and *Gypsy6-I_Dmoji* retroelements, clearly demonstrates that centromeric regions exhibit very strong and specific CUT&Tag CENP-A signals. Of note, CENP-A binding to *Gypsy6-I_Dmoji* in chr2^{Capri} was not reported by Chang *et al.*, which might be due to methodological differences or a different data analysis approach in my study. The centromere islands composition reported by the authors is based on embryo ChIP data, whereas here, we used S2 cells. Indeed, Chang *et al.* point out, that in ChIPs from S2 cells, CENP-A enriched at additional other elements such as 1.688 satellite and *Responder*, which are pericentromeric in embryos. The authors reason, that in S2 cells the CENP-A bound centromere region may expanded into the pericentromere. My data supports this notion, since many peaks (Figure 19 A) covered members of the 1.688 satellite family (260bp_SAT, 353bp_SAT and 359bp_SAT), most of which are pericentromeric in *Drosophila* (Usakin *et al.*, 2007). Moreover, these strong, presumably non-centromeric CENP-A peaks in embryos showed no signal in IgG or γH2Av (Figure 18 C, left plot), indicating that

here, CENP-A is maximally enriched compared to other histone variants or the control and this suggests, that these sites indeed represent centromeric regions in S2 cells.

In my dataset, I observed that γH2Av signals were co-occurring at transposable elements with weaker CENP-A peaks (Figure 19 C, right plot), and I assume that these sites represent genuine non-centromeric CENP-A, since another histone variant is binding as well. Generally, we found CENP-A enrichment at a whole list of other non-centromeric retroelements (Figure 19 A) and most of these elements were also detected in Chang *et al.* (supplemental table 4 and 8 in Chang *et al.*). The authors also categorize these elements as non-centromeric and furthermore note, that CENP-A binding to many of them was weak. This is why they believe these peaks might be non-specific. The other way around, in my study binding to telomeric transposable elements was low and therefore not reported here, but Chang *et al.* point out, that they observed binding to the telomere-associated element *TART-A*. Thus, both our studies showed that CENP-A binds to such non-centromeric or telomeric transposable elements even though with varying strength. Moreover, in my dataset, the IgG control, which monitors unspecific background signal is almost absent. Collectively, this strongly indicates, that CENP-A targets transposable elements not only at the centromere but also ectopically at basal levels.

Based on this data, we are increasingly confident that CENP-A CUT&Tag in our hands worked successfully and allowed us to reliably profile CENP-A binding genome-wide, since expected CENP-A centromeric repeat targets were enriched as the most significant binding sites, which was in accordance with the dataset of Chang *et al.* As in Chang *et al.*, with CUT&Tag we even detected non-centromeric transposable elements as CENP-A targets. Regarding this qualitative overlap between our studies, I suggest that these locations are genuine, centromeric and non-centromeric CENP-A binding sites in S2 cells.

4.3.4 Genome-wide CENP-A profile

In the previous section I elaborated, that CENP-A seems to bind transposable elements outside of the centromere. Adding to this notion, the alignment of CENP-A reads to the current *Drosophila* BDGP genome release and subsequent gene annotation of the peaks revealed genome-wide CENP-A binding. This observation lines up with the accumulating evidence of CENP-A being distributed genome-wide at basal levels (cf. section 1.4). CENP-A, as in previous studies, most frequently located to sites of active chromatin such as TSS, promoters and gene bodies but also to introns and intergenic distal sequences (Figure 20 B). CENP-A at intergenic distal sequences or in other words at non-coding DNA and in introns most likely reflects CENP-A binding to enhancers or transposable elements, which often occur at such

sites (Zhou *et al.*, 2020, Gilbert *et al.*, 2021; Panigrahi and O'Malley, 2021). Transposable elements were not annotated in the BDGP dataset, but it is an important future task, to examine if intergenic and intronic peaks correspond to transposable elements as it was observed with the repeat annotated PacBio data (section 4.3.3). This could be done by annotating the BDGP mapped reads/peaks using an annotation approach that identifies transposable elements such as RepeatMasker (https://www.repeatmasker.org/). Furthermore it would be interesting, to characterize the chromatin environment at CENP-A peak locations at non-coding DNA to determine if there is colocalization with enhancer typic histone marks such as H3K4me1 and H3K27ac (Calo and Wysocka, 2013), with polycomb responsive elements (PRE) or with insulator factors such as BEAF-32 (cf. section 1.1.2). Intriguingly, the mass spec analysis of CENP-A binding partners revealed an association with insulator factors Ibf1 and Ibf2 (Supplemental Table 1) and neocentromere formation was detected at chromatin boundaries (Olszak *et al.*, 2011), which in summary indicates CENP-A binding to insulators.

4.3.5 **CHRAC-14 depletion causes CENP-A chromatin accumulation**

So far, we hold proof that CHRAC-14 absence leads to an increase of CENP-A protein levels and IF signals (Mathew *et al.*, 2014) and we were excited to discover via CUT&Tag, that excess CENP-A in this condition seems to be indeed incorporated into chromatin in physiological conditions (Figure 20). Based on my data, CHRAC-14 depletion seems to cause CENP-A accumulation on the centromere itself, since here we saw a clear tendency of increased CENP-A binding. Hereby, CENP-A predominantly increased at simple repeats (Figure 18 C). Since we did not specifically orient our analysis towards simple repeat analysis i.e. by including a kmer analysis with k-Seek (https://github.com/weikevinhc/k-seek; Wei *et al.*, 2018) like Chang and colleagues (Chang *et al.*, 2019), mapping of our reads to simple repeat locations might not be as accurate and hampers quantitative claims. However, we can conclude that the CENP-A increase is generally prominent at simple satellite repeats independent of if they were mapped correctly or not based on the quantitative increase of the bigwig score to simple repeats and also based on the increased multi mapper rate (Figure 16 A).

An increase of CENP-A was not as clear and prominent at non-centromeric repeats, where only some elements showed accumulating CENP-A (Figure 19 B). If a non-centromeric transposable element exhibits increased CENP-A or not might depend on its genomic location and its chromatin state i.e. if it is buried in stably silenced chromatin or if it is active and accessible for chromatin remodeling and CENP-A loading (discussed in 4.4.1). Notably, amongst the non-centromeric repeats, 1.688 satellites showed the most consistent fold changes and as previously discussed (cf. section 4.3.3), 1.688 satellites can be considered to

be centromeric in S2 cells. This again supports the notion that predominantly centromeric repeats are subject of increased CENP-A binding in the absence of CHRAC-14.

In addition, we found that CHRAC-14 depletion causes an excess of genome-wide CENP-A peaks, which were annotated with regular gene annotation file without repetitive annotations (Figure 20 A). Assessment of the CENP-A signal at these locations disclosed a significant CENP-A increase upon CHRAC-14 knockdown compared to the control RNAi and compared to γH2Av peaks (Figure 20 C). Hence, we propose that CENP-A is not only increasing at the centromere but also in chromosome arms upon CHRAC-14 knockdown, whereby it seems to maintain its localization pattern to distal intergenic sequences (likely transposable elements and enhancers) or active chromatin such as promoters and gene bodies (Figure 20 B).

4.3.6 CENP-A is not detected at DNA damage sites in DIvA cells

The outcome of our approach to map CENP-A during DNA damage in DIvA cells, is ambiguous. Eight candidate DSB sites were identified by an increase of γH2Av in the vicinity (± 3 kb) around an AsiSI motif but then no CENP-A could be detected in none of the conditions. On the one hand, this might implicate that in contradiction to our hypothesis, CENP-A is not accumulating at DNA damage sites, neither in control nor in CHRAC-14 RNAi conditions.

However, the foreseen problems with the DIvA system in S2 cells eventually prove true. As discussed previously (section 4.3.1), my DIvA system has the disadvantage of an asynchronous DNA damage response. Additionally, not every AsiSI motif is equally accessible in every cell at a given timepoint. Consequently, the data most likely reflects a mixture of different AsiSI-cut motifs and γH2Av CUT&Tag signals at cell-specific DNA lesions might be averaged out in the bulk cell population. In accordance with that, we did not find sharp damage-specific γH2Av peaks at AsiSI motifs but rather a mild accumulation as depicted in Figure 21 B. Adding up to the problem, AsiSI^{OFF} samples already showed robust CUT&Tag γH2Av signal and thus γH2Av localized throughout the genome in physiological conditions. Indeed, *Drosophila* phosphorylated H2Av is a histone variant with different roles. Next to acting as a DNA damage marker (Madigan *et al.*, 2002), γH2Av is also involved in other processes such as transcriptional regulation (Kotova *et al.*, 2011) and plays a role in insulator function (Simmons *et al.*, 2021). Consequently, γH2Av localizes throughout the genome and is not exclusive to DSB. To circumvent such non-damage signals, another DNA damage marker could be used in the future, which is more specific for DSB than γH2Av. We chose to use γH2Av in the beginning because (i) it serves as an internal reference for histone variant dynamics and (ii) of the lack of ChIP grade antibodies for *Drosophila* DNA damage markers.

An endogenously tagged maker in combination with a ChIP grade epitope antibody such as V5 could be considered in the future. Moreover, there are dedicated methods to map DNA damage sites such as BLESS, which could be considered for upcoming experiments, to precisely map DSBs in *Drosophila* DIvA cells (Crosetto *et al.*, 2013; Clouaire *et al.*, 2018). Other than that, employing a different DNA damage induction strategy may be beneficial as well. Spatio-temporal well refined DSBs could be induced by targeting Cas9 nuclease to one specified site in order to locally focus the accumulation of proteins of interest (Vítor *et al.*, 2020).

On another note, the CENP-A coverage turned out to be low in the chromosome arms in our experiment, which raises the possibility that the CENP-A signal at the DSBs was not covered sufficiently and thus remained undetected. Hence, the read output should be increased next time. Also, the knockdown does not lead to a complete loss of CHRAC-14 and hence CENP-A accumulation at DSBs upon CHRAC-14 reduction might has been inefficient and hindered a clear CENP-A signal at DSBs.

In summary, in our DNA damage system, which is functional in general, CENP-A was not observed at DSBs and this would be line with results of Helfricht *et al.* (2013). The reason that in Mathew *et al.* CENP-A was detected at uncapped telomeres as sites of DNA damage repair upon CHRAC-14 knockdown might be, that telomeres in contrast to random AsiSI DNA lesions seem to be preferred sites of CENP-A loading (cf. section 1.4.2; [Chang *et al.*, 2019a]).

However, the technical limitations of our DNA damage approach such as inefficient DSB induction and detection, undetectable CENP-A accumulation due to the low sequencing coverage and/or insufficient CHRAC-14 depletion, rule out any definite interpretation of the data. This leads us to admit, that thorough optimization of the approach and further replicates are necessary in the future in order to revisit the question, if CENP-A is localizing at DSBs or not and if CHRAC-14 is regulating this process.

4.3.7 **CHRAC-14 depletion causes upregulation of genes enriched in CENP-A**

Given that CENP-A seems to accumulate at promoters, TSSs, gene bodies and presumably at enhancers led us to speculate, that excess CENP-A could disturb the native chromatin environment and alter the expression of associated genes. According to the literature CENP-A ectopic incorporation seems to be linked to the activation of underlying genes (Murillo-Pineda *et al.*, 2021; Naughton *et al.*, 2021; Saha *et al.*, 2020) and supports our assumption. We set out to answer this question, by performing total RNA-Seq in CHRAC-14 depleted cells and it became apparent, that CHRAC-14 depletion causes vast misregulation of many genes (Figure 24), with a tendency towards an upregulation. Importantly, a gene ontology term enrichment analysis revealed that genes involved in the DNA damage response,

the mitotic cell cycle and chromosome segregation are differentially expressed (Figure 24 G). This is an important new finding and could in part serve to explain CHRAC-14 phenotypes such as a compromised DNA damage response, a defective G2/M checkpoint and chromosome segregation defects (Mathew *et al.*, 2014). Of course this phenotype could be attributed to the fact, that CHRAC-14 is involved in chromatin remodeling and its depletion automatically leads to chromatin changes and expression alterations. Nevertheless, we asked, if genes that are misregulated upon CHRAC-14 RNAi also exhibit increased CENP-A, addressing the possibility, that altered gene expression could be caused by CENP-A accumulation. I found an overlap of 84 transcripts/genes (Figure 25), which was remarkable, given that CUT&Tag was performed in DIvA and RNA-Seq in wild type cells. Thus, the observed overlap of both datasets probably reflects robust and reliable hits. The majority of these genes are upregulated supporting the previous notion that CENP-A accumulation is associated with gene activation.

By qPCR I next tested, if the differential expression of candidate genes can be reproduced in two additional biological replicates. Up to this point, the gene expression changes mostly showed clear and similar tendencies like in RNA-Seq but are not or only minor significant, because the statistical power with two replicates is too weak (Figure 26). Generating a few more replicates will increase the statistical power and enable a reliable conclusion, whether the observed changes for these candidates are biologically relevant.

In addition to misregulated genes, we also found RNAs from transposable elements to be differentially expressed (Figure 27). Amongst these elements were centromeric and telomeric RNAs and TE classes of which other representatives were bound by CENP-A in my CUT&Tag data (telomeric TE biding fell below the threshold and was not reported). In summary this suggests that similarly to genes, those TE classes, which are preferably bound by CENP-A potentially show differential expression upon CHRAC-14 RNAi.

In conclusion, up to this point, I could show that CHRAC-14 depletion *per se* causes gene and TE misregulation and that the upregulation of genes is partly accompanied by increased CENP-A binding. It is a key task for the future to test our hypothesis and assess if CENP-A accumulation locally and directly causes the upregulation of genes and TEs. For this, it could for example be tested, if a co-depletion of CHRAC-14 and CENP-A reverses the upregulation of the genes. Hereby it is crucial, to only mildly reduce CENP-A levels in order to counteract its accumulation and simulate endogenous CENP-A levels during a CHRAC-14 knockdown. Next to assessing the gene expression levels, it would be an important control, to precisely monitor CENP-A incorporation in that condition by CUT&Tag.

4.4 Assembling puzzle pieces: CHRAC-14 titrates CENP-A chromatin and preserves genome function

4.4.1 Ectopic CENP-A loading might be promoted by transcription-coupled chromatin remodeling

It is well established that, in yeast and human, endogenous CENP-A is sparsely distributed throughout the entire genome (Bodor *et al.*, 2014; Nechemia-Arbely *et al.*, 2019; Shukla *et al.*, 2018). Even though genome-wide CENP-A incorporation is not a major focus in Chang *et al.* (2019), CENP-A was mapped in chromosome arms in their study. In my thesis, I now confirm their results and show that CENP-A is incorporated genome-wide in *Drosophila*. Similar to other studies (cf. section 1.4.2), I identified CENP-A binding at promoters, transcriptional start sites, exons, transposable elements and distal intergenic regions (presumably enhancers and TEs). Two central questions arising from this observation are, why CENP-A intrinsically localizes to these sites in physiological conditions and what mechanism regulates this behavior.

All of the 'CENP-A-bound genome elements have in common that they are subject of transcription and it seems that transcriptionally active sites simply provide the easiest and most pervasive opportunity for CENP-A loading because the process is accompanied by chromatin remodeling. Indeed, centromeric CENP-A loading and inheritance has been associated with transcription and the FACT complex (cf. section 1.3.2), supporting the assumption that also genome-wide, transcription has the potential to promote CENP-A incorporation.

Obviously, active elements such as promoters, transcriptional start sites and gene bodies are subject of transcription and, therefore, it is conceivable, that most often CENP-A localizes here.

On the contrary, it is puzzling, how CENP-A loading at presumably inactivated transposable elements could occur through transcription. Interestingly, many transposable elements are active in *Drosophila* (Luo and Lu, 2017). The Gypsy TE family has been shown to be actively transcribed and produce repeat RNAs, which are fed into the siRNA pathway for pericentric heterochromatin maintenance in *Drosophila* S2 cells (Hao *et al.*, 2020). Generally, heterochromatin formation at the pericentromere through transcription and the siRNA pathway is well established in yeast (Corless *et al.*, 2020) and repeat derived piRNAs promote retrotransposon silencing in *Drosophila* germline cells (Halic and Moazed, 2009). Moreover, mammals have adapted the regulatory region of transposable elements as transcription factor binding sites or *cis*-regulatory elements, which need chromatin remodeling to avail their binding motifs (Sundaram and Wysocka, 2020).

In summary, transposable elements might show strong CENP-A binding because they are transcribed but also because chromatin remodeling at their regulatory region permits CENP-A loading.

Moreover, CENP-A seems to bind telomeric regions: It was found to localize as basal levels to telomere TEs in Chang *et al.* (2019) and also in my study (but not amongst the top 20% since less significant, data not shown). Also, in Mathew *et al.* (2014), CENP-A localized to unprotected telomeres and less frequently even under physiological conditions (unpublished data from Dr. Mathew). Furthermore, when overexpressed in cancer, CENP-A can be found at sub-telomeric regions (Athwal *et al.*, 2015). In accordance to my hypothesis, CENP-A loading at telomeres could be promoted by transcription-coupled chromatin remodeling as well, because in stem cells and cancer cells chromatin is newly established each cell cycle with the help of the telomere-transcribed RNA TERRA (Tardat and Déjardin, 2018).

Interestingly, insulators at TAD boundaries are often transcribed in *Drosophila* (cf. section 1.1.2) and were observed to favor neocentromere formation and CENP-A binding (Lacoste *et al.*, 2014; Olszak *et al.*, 2011), which further supports the notion, that CENP-A is attracted by transcriptionally active sites.

Lastly, it was proposed, that CENP-A localizes to double strand breaks (Mathew *et al.*, 2014; Zeitlin *et al.*, 2009). Even though this is a matter of debate and in my study, I found no indication of this, DSB could be conceivable CENP-A binding sites, since DSB repair also involves chromatin remodeling (cf. section 1.1.4).

Collectively, I come to conclude that CENP-A binding to genomic elements reported in this study exploits chromatin remodeling processes that are coupled to transcription and *cis*-regulatory element 'activation' leading to sparse CENP-A distribution genome-wide in *Drosophila*. The purpose of basal CENP-A levels might be to prime the genome for neocentromere formation in case of centromere loss due to DNA breakages contributing to karyotype evolution and species specification (Dong *et al.*, 2021). Indeed, neocentromere formation has been observed at chromosomal break points (DeBose-Scarlett and Sullivan, 2021; Hasson *et al.*, 2011). When CHRAC-14 is depleted, CENP-A vastly accumulates at sites of previous low-level incorporation, indicating, that CHRAC-14 is involved in balancing CENP-A loading during omnipresent chromatin remodeling. This is an important task, as excessive CENP-A accumulation leads to neocentromere formation and genome instability.

4.4.2 **Possible mechanisms of CHRAC-14 mediated CENP-A titration**

Collectively, based on the presented data I propose a model, where CHRAC-14 balances CENP-A incorporation via transcription or other types of chromatin remodeling such as DNA repair or telomere maintenance so that low basal levels of CENP-A are established

genome-wide (Figure 29). We do not have any indication yet how cells sense and control the level of ectopic CENP-A and how this information is passed on to CHRAC-14 for adequate CENP-A titration. Interestingly, CK2 not only is a regulator of CENP-A levels (Huang *et al.;* this study) but also phosphorylates CHRAC-14, as shown in this thesis. CK2 phosphorylation of CHRAC-14 seemed important for the elimination of a post translationally modified version of CENP-A (presumably SUMOylated). Despite lacking mechanistic details, this CK2-dependent modification axis provides an interesting basis for further mechanistic investigations.

Regarding the pathway of CENP-A titration via CHRAC-14, two mechanisms are imaginable:

1. CHRAC-14 sequesters CENP-A by dimerization, completely shielding it from nucleosome assembly and mediating its degradation when not presented to its rate-limiting loading factor CAL1 or other genome-wide loading factors. However, the fact that CHRAC-14 has been exclusively described as a subunit of chromatin remodeling complexes argues against this theory.

2. As a replication independent histone variant CENP-A could be promiscuously loaded into chromatin genome-wide by transcription-coupled chromatin remodeling involving factors, which have been associated with ectopic CENP-A loading in the literature such as DAXX/ATRX and NuRD and at the centromere by CAL1 (Figure 29). CHRAC-14 evicts previously deposited excessive CENP-A levels via replication-coupled chromatin remodeling, since CHRAC-14 together with Mes4 is likely to be part of Pol ε in *Drosophila*. This mechanism is very conceivable, since there are evidences for replication-coupled CENP-A eviction in human cells (Nechemia-Arbely *et al.*, 2019). Up to now, the field lacks any insights on how CENP-A is evicted, and CHRAC-14 may be the missing link. CHRAC-14 may act as the receiving CENP-A chaperone at the replication fork, mediating CENP-A elimination by proteasomal degradation.

Upon stress such as DNA damage, however, CHRAC-14 might associate with other stress-specific chromatin remodeling complexes such as pBAF, in order to balance CENP-A incorporation for example at DNA damage sites.

Importantly, CENP-A removal by replication is a mechanism, which could be implemented at centromeres as well. As I showed in my study, CENP-A not only accumulates genome-wide but also at the centromere when CHRAC-14 is depleted. Since the replication machinery is also passing at the centromere, it could serve as a ubiquitous CENP-A titration machinery in co-operation with CHRAC-14.

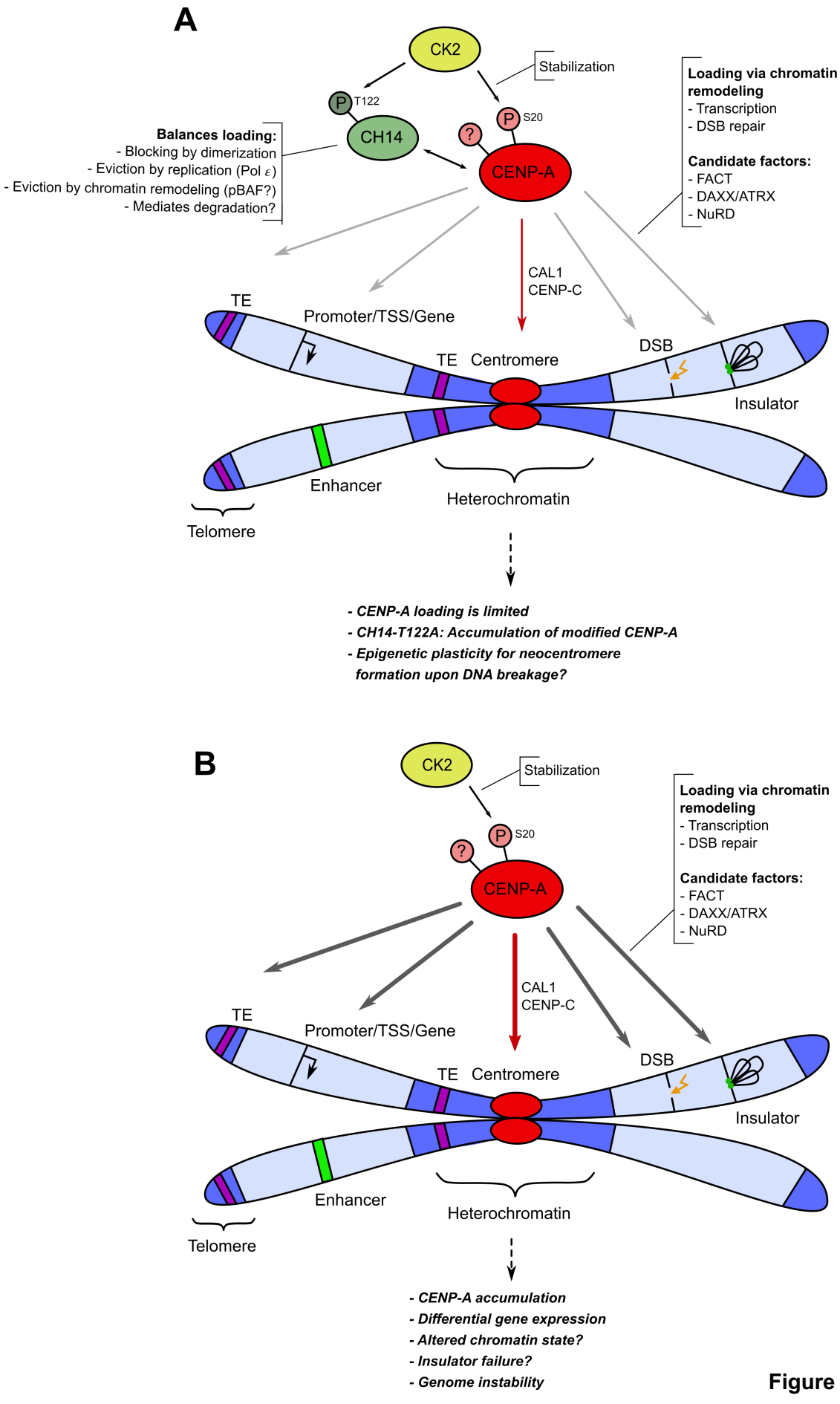


Figure legend ▶

Regarding this model we cannot help but think, that CENP-A seems to behave like the histone variant H3.3. H3.3 is loaded by DAXX/ATRX at similar locations like CENP-A (cf. section 1.1.1) and CENP-A in turn can be incorporated by DAXX/ATRX in humans. It will be a fascinating future task, to explore if CENP-A has a centromere-independent role in genome regulation, just like the H3 variant H3.3. As a first step into this direction, one could assess by IF or sequencing, if both variants are colocalizing or if H3.3 depletion causes CENP-A to take its place in chromatin. This is an attractive hypothesis, since H3.3 is deposited as a CENP-A placeholder during replication at centromeres until CENP-A loading later on (Dunleavy *et al.*, 2011), illustrating the intertwining relationship of both histone H3 variants.

It is important to keep in mind that endogenous CENP-A is available for chromatin loading during canonical transcription or other reasons of chromatin remodeling during G_1-S phase, since CENP-A seems to available in these cell cycle stages in *Drosophila* (Lidsky *et al.*, 2013; Moreno-Moreno *et al.*, 2019).

Taken together, according to this model (Figure 29) CENP-A seems to have the intrinsic capacity of localizing to transcriptionally active sites and for this might exploit H3.3 loading machineries or NuRD at sites of chromatin remodeling and histone turnover. This is then counteracted during replication by CHRAC-14, which most likely is the CENP-A binding subunit of Pol ε. In the absence of CHRAC-14 CENP-A is increasingly loaded ectopically but also at the centromere. This leads to detrimental problems and genomic instability, since accumulated CENP-A interferes with the bound element so that for example CTCF is displaced at insulators or the altered chromatin environment causes transcriptional upregulation. In addition, extensive CENP-A accumulation might peak in neocentromere formation.

Figure 29 Schematic model of CENP-A titration via CHRAC-14

(A) CENP-A chromatin loading is mediated via chaperones and chromatin remodeling during transcription or DSB repair and is balanced by CHRAC-14 via different possible pathways. CK2 phosphorylation of CENP-A has a stabilizing effect. CHRAC-14 phosphorylation by CK2 might have no effect on CENP-A titration but might mediate CENP-A degradation afterwards. **(B)** In the absence of CHRAC-14 active loading of CENP-A is not balanced leading to CENP-A accumulation and detrimental effects eventually resulting in genome instability.

5 Conclusion and Perspectives

CHRAC-14 is a crucial regulator of CENP-A loading and in my thesis, I aimed to gain a better understanding of the underlying mechanism in *Drosophila melanogaster*. I identified CK2 as an interaction partner of CHRAC-14 and CENP-A and showed that CK2 phosphorylates both proteins *in vitro*. My data implicates, that the phosphorylation of CHRAC-14 at residue T122 is important to eliminate a post translationally modified version of CENP-A. Furthermore, my study led to the conclusion that CK2 activity promotes endogenous CENP-A stability. Moreover, my screen revealed other hitherto unknown CHRAC-14 interactors, which are involved in chromatin remodeling.

In the second part of this thesis, I set out to profile CENP-A chromatin and loading upon CHRAC-14 depletion and during site-directed DNA damage induction with the restriction enzyme AsiSI by targeted CENP-A CUT&Tag sequencing. With that I confirmed, that endogenous CENP-A is distributed throughout the genome at basal levels at transcriptionally active sites and that CHRAC-14 depletion causes CENP-A accumulation at the centromere and in chromosome arms. In addition, my data showed, that CHRAC-14 depletion causes mis-expression of genes primarily involved in mitotic processes such as chromosome segregation and mitotic cell cycle regulation or in the DNA damage response. This might implicate that CHRAC-14-depended chromatin remodeling in general is important for accurate gene expression and an adequate DNA damage response. It remains to be determined if CENP-A accumulation upon CHRAC-14 depletion is causing gene mis-expression but the fact that many upregulated genes show CENP-A enrichment supports this notion.

I propose a model, where CENP-A is loaded at the centromere by the canonical CAL1-mediated pathway and ectopically at transcriptionally active sites by factors such as DAXX/ATRX or NuRD, which are linked to ectopic CENP-A loading in the literature (cf. section 1.4.1). During stress such as DNA damage, which creates an additional opportunity for chromatin remodeling-coupled CENP-A loading at DSBs but we did not find direct evidence for this. CHRAC-14 might limit this process in co-operation with other stress-specific chromatin remodeling factors such as pBAF or ATRX, factors which emerged as damage-specific CHRAC-14 interactors in this study. In the contrary, during physiological conditions excess CENP-A might be removed during replication-coupled chromatin remodeling and polymerase ε as suggested in human cells (Nechemia-Arbely *et al.*, 2019) using CHRAC-14 as the CENP-A binding subunit. It is tempting to speculate that CHRAC-14 then mediates CENP-A degradation, a well-known mechanism to limit ectopic CENP-A loading (cf. section 1.4.4).

Even though we gained new insights, open questions are persisting: Up to now it remains to be determined with which post-translational modification CENP-A is decorated in CHRAC-14-T122A cells to cause a shift in protein size and if this form of CENP-A is degraded in CHRAC-14 wild type cells. As a first step, the CENP-A modification should be characterized by mass spectrometry. Then it would be interesting to screen whether modified CENP-A is stabilized upon RNAi-mediated depletion of certain factors, which are involved in proteasomal degradation such as the herein detected E3 ligases Bre1 and Snama or the SUMO protease Velo.

It also remained elusive if CENP-A is localizing to DNA lesions. In order to re-tackle this question, it would be advantageous to perform single cell sequencing in DIvA cells in the future. This would enable a precise analysis of CENP-A localization in individual cells with a well-refined temporal resolution. In addition, it would help to determine, if CENP-A in individual cells can accumulate to an extent that causes neocentromere formation, since such sporadic events remain undetected in a bulk sequencing approach.

Furthermore, it is an important next step, to determine if CENP-A accumulation at genes or TEs upon CHRAC-14 depletion directly causes enhanced transcription thereof. For this, increasing CENP-A levels and, hence, its accumulation after CHRAC-14 RNAi could be prevented with a co-knockdown of CENP-A to test, if this rescues the gene expression pattern.

Moreover, it is unknown, if an upregulation of genes upon CENP-A accumulation is accompanied by changes of the surrounding chromatin state. It has been reported, that neocentromere formation coincided with the loss of heterochromatin marks (Murillo-Pineda *et al.*, 2021). Hence it would be intriguing, to determine by IF or targeted sequencing, whether the histone mark composition changes at sites of CENP-A accumulation. In general, it is important to determine by colocalization IF experiments or sequencing, which chromatin makeup favors CENP-A binding and if CENP-A colocalizes with genomic elements such as insulators, which has been observed for overexpressed CENP-A (Lacoste *et al.*, 2014). If chromatin state changes are valid in our pathway, this potentially leads an altered spatial organization of the genome (cf. section 1.1.2). Thus it could be tested by Hi-C, a sequencing-based method that captures chromatin contacts (van Berkum *et al.*, 2010), if CENP-A accumulation is accompanied by such spatial changes. Indeed, unpublished data from our lab indicates that CENP-A affects chromatin compaction.

Lastly, it is crucial to test our model where CENP-A is titrated by replication-coupled removal in collaboration with CHRAC-14. For this, replication could be inhibited *in vivo* in order to monitor, if CENP-A is accumulating similarly to when CHRAC-14 is knocked down. However, tackling replication as one of the most integral cellular mechanisms, probably causes vast pleiotropic side effects and artefacts. Therefore, an *in vitro* replication assay using

reconstituted CENP-A chromatin and recombinant CHRAC-14 might be a better suitable system (Svitin and Chesnokov, 2010).

In conclusion, with the experiments conducted in this study, I disclosed a regulative signaling axis involving CHRAC-14, CK2 and a yet uncharacterized CENP-A post translational modification. Moreover, I showed, that CHRAC-14 balances CENP-A loading at centromeres and in chromosome arms implicating CHRAC-14 as the missing link, that couples replication to CENP-A titration. These findings provide an important basis for further investigations targeting the pathway of CHRAC-14-mediated CENP-A loading, contributing to a better understanding of this vital mechanism, which protects faithful chromosome segregation and genome stability.

6 Materials

6.1 Chemicals

All chemicals used in this study were mostly purchased from Sigma, Merck, Roth, AppliChem, Invitrogen or Roche. A detailed overview of the most important chemicals is shown in the following table:

Table 1 Chemicals

Name	Provider
2-Propanol	AppliChem
Acetic acid	Roth
Agar bacteriology grade	AppliChem
Agarose Ultra Pure	Sigma
Ammonium persulfate (APS)	AppliChem
Ampicillin Sodium salt	AppliChem
Auxin (IAA)	Sigma
Bromphenol blue	AppliChem
BSA (Bovine serum albumin)	AppliChem
Calcium chloride ($CaCl_2$)	AppliChem
Chloroform	Roth
Coomassie brilliant blue R125	AppliChem
Copper sulphate ($CuSO_4$)	Th. Geyer
DAPI	Sigma
Digitonin	EMD Millipore
Dimethyl pimelimidate (DMP)	Sigma
Dimethyl sulfoxide (DMSO)	AppliChem
Disodium hydrogen phosphate (Na_2HPO_4)	AppliChem
Dithiothreitol (DTT)	Sigma

Egtazic acid (EGTA)	AppliChem
Ethanol	Roth
Ethidium Bromide (EtBr)	AppliChem
Ethylenediaminetetraacetic acid (EDTA)	Roth
Glutathione	Sigma
Glycerol	Roth
Glycine	Sigma
Grape juice	Rewe
Heparin	Sigma
HEPES	Sigma
Hydrochloric acid (HCL 37%)	Roth
Hygromycin B	Sigma
Imidazole	Roth
Isopropyl β-d-1-thiogalactopyranoside (IPTG)	Roth
Kanamycin sulfate	AppliChem
Magnesium chloride ($MgCl_2$)	AppliChem
Manganese(II) chloride ($MnCl_2$)	Sigma
Methanol	ZMBH
N-Ethylmaleimide (NEM)	Sigma
Neocarzinostatin (NCS)	Sigma
Nonidet P-40 (NP-40)	AppliChem
Paraformaldehyde (PFA)	AppliChem
Penicillin/Streptomycin (100x)	Capricorn Scientific
Phenol-chloroform-isoamyl alcohol 25:24:1 (PCI)	Roth
Phenylmethylsulfonylfluoride (PMSF)	Sigma
Ponçeau S solution	AppliChem
Potassium chloride (KCl)	AppliChem
Potassium dihydrogen phosphate (KH_2PO_4)	AppliChem

Puromycin	Sigma
Roche Complete Protease Inhibitor EDTA-free tablets	Sigma
Schneider's Medium	Gibco
Skim milk powder	Sigma
Sodium azide	AppliChem
Sodium chloride (NaCl)	Sigma
Sodium citrate	Sigma
Sodium dodecyl sulfate pellets (SDS)	Serva
Sodium fluoride (NaF)	Sigma
Sodium hydroxide pellets (NaOH)	AppliChem
Sodium hypochlorite	Roth
Sodium orthovanadate	Sigma
Sodium tetraborate	Merck
Spermidine	Sigma
Synperonic PE(R)/F68	Sigma
Tetramethylethylenediamine (TEMED)	AppliChem
Trichloroacetic acid (TCA)	AppliChem
Tris (tris(hydroxymethyl)aminomethane hydrochloride)	AppliChem
TRIsure	Bioline
Triton X-100	Merck
Trypan Blue	Sigma
Tween 20	AppliChem
Yeast	Rewe
β-Mercaptoethanol	AppliChem

6.2 Materials and consumables

Table 2 Materials and consumables

Name	Provider
Agarose gel electrophoresis tank	ZMBH workshop
Agarose gel trays and combs	ZMBH workshop
ÄKTA Pure and appendant equipment	GE Healthcare
Aqua Polymount	Polysciences
Balance	Sartorius
Binocular microscope	Zeiss
Bioanalyzer 2100	Agilent
Biorupter Plus	Diagenode
Blotting paper, extra thick 2.45 mm	Bio-Rad
Blotting paper, thick 1.0 mm	Roth
Blotting paper, thin 0.2 mm	Roth
Cell counting system, LUNA	Logos Biosystems
Cell culture 6-well plate	Greiner Bio-One
Cell culture flasks (T25, T75, T125)	TPP
Cell culture incubator	Heraeus
Centricon, Amicon	Sigma
Centrifuge bottle, polypropylene	Nalgene
Centrifuge tube, polypropylene	Nalgene
Centrifuge, 5810 R	Eppendorf
Centrifuge, RC 6 Plus	Sorvall
Centrifuge, table top	Eppendorf
Coplin staining jar	Neolab
Coverslips	Neolab
Cytology funnel	Thermo
Cytology funnel clip	Fisher Scientific
Cytospin	Thermo
Deltavision microscope	Olympus/GE Healthcare
Fine Balance	Sartorius
Fly cages	ZMBH workshop

Fly plastic vials, small	Stein
Fly plastic vials, big	Greiner
Fly vials plug, small	Kisker
Fly vials plug, big	KTK
Fragment Analyzer 5200	Agilent
Freezer, -20°C	Liebherr
Freezer, -80°C	Heraeus
Fridge, 4°C	Liebherr
Gel Doc XR+ System	Bio-Rad
Gel Dryer, Model 583	Biorad
GyroMini nutating mixer	Labnet
Lab tape	Zentrallager Heidelberg
LAS-4000	Fujifilm Life Science
LightCycler 480 Multiwell Plate 384, white	Roche
LightCycler 480 Multiwell Sealing Foil	Roche
LightCycler 480 System, 384-well version	Roche
Liquid nitrogen portable container	KGW Isotherm
Magnetic stand	ZMBH workshop, GE Healthcare
MaXtract High Density Phase-lock tubes	Qiagen
Microfilter	Roth
Microfluidizer, Emulsiflex-C5	Avestin
Micropipettes	Gilson
Microwave	Sharp
Milli-Q filter system	Merck
Mini centrifuge	Nippon Genetics
Mini Trans-Blot cell	Bio-Rad
Mini-PROTEAN Tetra Vertical Electrophoresis	Bio-Rad
Mini-PROTEAN Tetra handcast system	Bio-Rad
Mr. Frosty	Thermo
Multichannel Pipet	Thermo
Nanodrop A260	Nanodrop
NextSeq550 Sequencer	Illumina
Nitrocellulose membranes	Amersham Biosciences

Nunc Cryo Tubes	Sigma
Parafilm	Bemis
PD-10 column	GE Healthcare
pH-meter	Sartorius, Kern AG
Pipette tips	Sarstedt, TipOne, Avant Guard,
Plastic pipettes, sterile	VWR
Plastic syringe	BD
PowerPac power supply	Bio-Rad
precast Gels	Bio-Rad
Qubit 3 Fluorometer	Invitrogen
Reaction tube, 0.2 ml	Sarstedt
Reaction tube, 0.5 ml	Eppendorf
Reaction tube, 1.5 ml	Sarstedt
Reaction tube, 1.5 ml, protein low binding	Eppendorf
Reaction tube, 15 ml falcon	Sarstedt
Reaction tube, 2 ml	Sarstedt
Reaction tube, 50 ml falcon	Sarstedt
ReadyPrep Mini Grinders	Bio-Rad
Rotation wheel	Labinco BV
Semi dry blotting machine	ZMBH workshop
Shaker, Mini-100 Orbital-Genie	Scientific Industries
Shaking incubator	Infors
Storage phosphor screen and cassette	VWR
Superfrost Plus Slides	Thermo
Thermo-Gradientencycler	Nippon Genetics
Thermomixer	Eppendorf
Tube roller	IDL
Typhoon FLA 7000 laser scanner	GE Healthcare
Ultracentrifugation tube, polypropylene	Beckman Coulter
Ultracentrifuge	Beckman Coulter
Vortex	Scientific industries
Water bath	Memmert, Roth, ZMBH workshop

6.3 Buffers

Double distilled or Milli-Q filtered water was used to prepare buffers.

Table 3 Buffers

Name	Composition
Bacteria lysis buffer	1X PBS 2 mM PMSF 1 mM DTT
CUT&Tag 5% Digitonin solution	50 mg dissolved in 1 ml boiling water by vortexing (2 ml reaction tube); store at RT for max 2 days (redissolve at 85°C if necessary)
CUT&Tag annealing buffer	10 mM Tris pH8 50 mM NaCl 1 mM EDTA
CUT&Tag antibody buffer	20 mM HEPES pH 7.5 150 mM NaCl 0.5 mM Spermidine 2mM EDTA 0.1% BSA 0.05% Digitonin 1X Protease inhibitor cocktail
CUT&Tag binding buffer	20 mM HEPES pH 7.5 10 mM KCl 1 mM $CaCl_2$ 1 mM $MnCl_2$
CUT&Tag Dig-300 buffer	20 mM HEPES pH 7.5 300 mM NaCl 0.5 mM Spermidine 0.01% Digitonin 1X Protease inhibitor cocktail

Name	Composition
CUT&Tag Dig-wash buffer	20 mM HEPES pH 7.5 150 mM NaCl 0.5 mM Spermidine 0.05% Digitonin 1X Protease inhibitor cocktail
CUT&Tag NE1 buffer	20 mM HEPES pH 7.9 10 mM KCl 0.5 mM Spermidine 0.1% Triton X-100 20% Glycerol 1X Protease inhibitor cocktail
CUT&Tag tagmentation buffer	20 mM HEPES pH 7.5 300 mM NaCl 0.5 mM Spermidine 10 mM $MgCl_2$ 0.01% Digitonin 1X Protease inhibitor cocktail
CUT&Tag TE buffer	10 mM Tris-HCl pH 8.0 1 mM EDTA
CUT&Tag wash buffer	20 mM HEPES pH 7.5 150 mM NaCl 0.5 mM Spermidine 1X Protease inhibitor cocktail
Coomassie staining solution	50% Methanol 10% Acetic Acid 0.1% Coomassie brilliant blue R250
Crosslink buffer	200 mM Sodiumtetraborate pH 9.0
Destaining solution	5 % Acetic Acid 16.5% Ethanol
DMP-crosslink buffer	200 mM Sodiumtetraborate pH 9.0 20 mM DMP

Name	Composition
Grape juice plates	1.75% Agar 25% grape juice
GST protein elution buffer	50 mM Tris-HCl pH 8.0 final 20 mM Glutathione
IP buffer	1X PBS 0.5% TX-100 200 mM final NaCl 20 mM NEM 10 mM NaF 1 mM Na-Orthovanadate 1 µg/ml Aprotinin/Leupeptin 2 mM PMSF 0.5 µg/ml Pepstatin 1X Protease inhibitor cocktail
Laemmli buffer (4X)	50 mM Tris-HCl pH 6.8 10% Glycerol 2% SDS 0.5% β-Mercaptoethanol 0.02% Bromphenolblue
Membrane blocking buffer	1X TBST 5% Milk powder
MS lysis buffer	1X PBS 200 mM final NaCl 0.1 % Triton X-100 20 mM NEM 10 mM NaF 1 mM Na-Orthovanadate 1 µg/ml Aprotinin/Leupeptin 2 mM PMSF 0.5 µg/ml Pepstatin 1X Protease inhibitor cocktail

Name	Composition
Phosphate-buffered saline (1X PBS)	137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4
PreScission protease cleavage buffer	50 mM Tris-HCL pH 7.0 0.01 % TX-100 150 mM NaCl 1 mM EDTA 1 mM DTT
S2 freezing medium	10% DMSO in conditional medium
SDS-PAGE running buffer	25 mM Tris 190 mM Glycine 0.1% SDS
Separation gel (15%)	375 mM Tris-HCl pH 8.8 15% Acrylamide/Bisacrylamide 30:0.8% 0.1% SDS 0.05% APS 0.05% TEMED
Squish buffer	10 mM Tris-HCL pH 8.0 1 mM EDTA 25 mM NaCl 0.2 mg/ml ProteinaseK
Stacking gel (4.4%)	123 mM Tris-HCl pH 6.8 4.4% Acrylamide/Bisacrylamide 30:0.8% 0.1% SDS 0.03% APS 0.1% TEMED

Name	Composition
Standard Fly food (by ZMBH fly food cooking service)	7.2% (w/v) maize 2.4% molasses 7.2% (w/v) malt 0.88% (w/v) soya 1.464% (w/v) yeast acid mix (1% propionic acid + 0.064% orthophosphoric acid)
TBS (10X)	24 g/l Tris 88 g/l NaCl pH 7.5
TBST	1X TBS 0.1% Tween
Tris-acetate-EDTA (TAE, 50X)	242 g/l Tris-HCl 18.6 g/l EDTA pH 7.7 adjusted with acetic acid
Western blotting buffer	25 mM Tris 190 mM Glycine 0.1% SDS 20% Methanol

6.4 Reagents and kits

Table 4 Reagents and kits

Name	Provider
1 kb Plus DNA Ladder	NEB
100 bp DNA Ladder	NEB
ATP	Thermo
Basemuncher	Expedeon
Bioanalyzer High Sensitivity DNA Kit	Agilent
Cellfectin II	Thermo

Colcemid (10µg/ml)	Capricorn Scientific
Concanavalin A (ConA)-coated magnetic beads	Bangs Laboratories
dNTPs	NEB
Dynabeads	Invitrogen
Dynabeads protein G/A	Thermo
ECL	Thermo
ECL Femto	Thermo
FBS	Capricorn Scientific
Gel Loading Dye, Purple (6X)	NEB
Gibson Assembly mix	lab-made
HS NGS Fragment Kit (1-6000bp)	Agilent
HS RNA Kit	Agilent
Human Casein kinase II (hCK2) with 10X reaction buffer	NEB
Ionic Detergent Compatibility Reagent	Thermo
LightCycler® 480 SYBR Green I Master	Roche
Lysozyme	AppliChem
MEGAscript® RNAi Kit	Thermo Fisher
Micrococcal Nuclease (MNase)	Thermo
NEBNext HiFi 2x PCR Master mix	NEB
NextSeq 500/550 High-Output v2.5 Kit (150 cycles)	Illumina
Nucleobond kits	Machery-Nagel
Pierce™ 660nm Protein Assay Kit	Thermo
PreScission Protease	GE Healthcare
Protein A–Tn5 (pATn5) fusion protein (2mg/ml)	- Kind gift from Henikoff lab - EMBL
Proteinase K (5 mg/ml)	lab-made
Q5 Polymerase reaction kit	NEB
QuantiTect Reverse Transcription Kit	Quiagen

Qubit 1X dsDNA HS Assay Kit	Invitrogen
Qubit RNA HS Assay Kit	Invitrogen
Quick Ligase	NEB
RedTaq reaction kit	Jena Bioscience
Restriction enzymes	NEB
RNA Clean & Concentrator-5	Zymo Research
SPRI paramagnetic beads (Agencourt AMPure XP)	Beckman Coulter
SUMO Protease	Invitrogen
Taq Polymerase reaction kit	NEB
Universal Plus Total RNA-Seq with NuQuant	TECAN
[γ-^{32}P]ATP	ZMBH isotope lab service
α-Casein, dephosphorylated	Sigma

6.5 Antibodies

Table 5 Primary antibodies

Antibody	Raised in	Dilution	Source
α-CENP-A	Rabbit	WB 1:1000	Active motif (39713)
α-CENP-A	Rabbit	CaT 1:100	Active motif (39719)
α-CHRAC-14	Rabbit	WB 1:100	Erhardt lab (74A)
α-Flag	Mouse	WB 1:1000	Sigma (F1804)
α-H3K27me3	Rabbit	CaT 1:100	Cell Signaling Technology (9733)
α-Lamin	Mouse	WB 1:5000	DSHB (ADL84.12-c)
α-Tubulin (α)	Mouse	WB 1:5000	Sigma (T5168)
α-V5	Mouse	WB 1:5000 IF 1:1000	Invitrogen (R960-25)
α- γH2Av	Rabbit	WB 1:1000 CaT 1:100	Rockland (600-401-914)

Antibody	Raised in	Dilution	Source
α- γH2Av	Rabbit	IF 1:500	Erhardt lab

Table 6 Isotype control antibodies

Antibody	Raised in	Dilution	Source
α-IgG_{2a}	Mouse	-	Sigma (M5409)
α-IgG	Rabbit	CaT 1:100	Sigma (15006)

Table 7 Secondary antibodies

Antibody	Raised in	Dilution	Source
α-mouse polyclonal IgG HRP	Goat	WB 1:10000	Sigma
α-rabbit polyclonal IgG HRP	Goat	WB 1:5000	Sigma
α-rabbit IgG Alexa 488	Goat	IF 1:500	Thermo
α-mouse IgG Alexa 647	Goat	IF 1:500	Thermo
α-rabbit IgG	Guinea pig	CaT 1:100	antikörper-online.de

6.6 DNA constructs

Table 8 DNA vector constructs

Name	Source
pGEX6P1-CHRAC-14	This study
pGEX6P1-CENP-A	Dr. S. Acharya
pMT-SV40-mAID-eGFP-Puro	This study
pMT-SV40-mAID-AsiSI-eGFP-Puro	This study
pMT-SV40-AsiSI-mAID-eGFP-Puro	This study
pMT-OsTIR1-mCherry-Hygro	This study
pMT-SV40-mAID-V5-Puro	This study

pMT-SV40-mAID-AsiSI-V5-Puro	This study
pMT-SV40-AsiSI-mAID-V5-Puro	This study
pMT-OsTIR1-Hygro	This study
pMT-CHRAC14-V5-His-Hygro	This study
pMT-CID-V5-His-Hygro	This study
pMT-CK2β-V5-His-Hygro	This study
pMT-CENP-A(wt)-V5-His-Hygro	This study
pMT-CENP-A(S20A)-V5-His-Hygro	This study
pMT-CENP-A(S20D)-V5-His-Hygro	This study
pMT-CHRAC14(T122A)-V5-Puro	This study
pMT-CHRAC14(T122D)-V5-Puro	This study
pMT-CHRAC14-Flag-Puro	P. Freidel

6.7 Oligos and Primers

All Oligos and primers were ordered and synthesized at Sigma/Merck.

Cloning

All primers used for molecular cloning are specified in the Erhardt lab primer register.

CUT&Tag

Oligos for Tn5 loading and primers for library amplification were from in Kaya-Okur *et al.,* 2019 and Kaya-Okur and Henikoff, 2020. Library primer sequences are specified in supplementary Table1 in Buenrostro *et al.*, 2015.

Table 9 RNA interference T7 primers

Name	Sequence (5' - 3')
CHRAC-14 Fw	TAATACGACTCACTATAGGGCGCATCGAGGATCTTAACC
CHRAC-14 Rv	TAATACGACTCACTATAGGGACTCGGGGGCTTCCTCT
Brown Fw	TAATACGACTCACTATAGGGAGCTCTCCTTCGTGCCCGT

Name	Sequence (5' - 3')
Brown Rv	TAATACGACTCACTATAGGGATCAATAGTAACCACTGCGGTGAAT
Hyd Fw	CTAATACGACTCACTATAGGGAGAGCGACCGAATAAGTCCAGAG
Hyd Rv	CTAATACGACTCACTATAGGGAGAGCCACACGACCAGAGGTTATC
CENP-A Fw	TAATACGACTCACTATAGGGACCGTGCAGCAGGAAAG
CENP-A Rv	TAATACGACTCACTATAGGGCCCCGGTCGCAGATGTA
CK2β dsRNA 1 DRSC20230 Fw	TAATACGACTCACTATAGGGAAGCTCGAGGACAATCCAC
CK2β dsRNA 1 DRSC20230 Rv	TAATACGACTCACTATAGGGAGGCTGACTTTCACAGTAGAC
CK2β dsRNA 2 DRSC33296 Fw	TAATACGACTCACTATAGGGACGAGGCAATGGTTAAGACCT
CK2β dsRNA 2 DRSC33296 Rv	TAATACGACTCACTATAGGGAGTAGGACGCTTGGGACGATA
CK2β dsRNA 3 DRSC33297 Fw	TAATACGACTCACTATAGGGAGTACCCAACTATCGGCAAGC
CK2β dsRNA 3 DRSC33297 Rv	TAATACGACTCACTATAGGGACCGAAATCGCCAGTTTGATA
CK2α dsRNA 1 DRSC1194 Fw	TAATACGACTCACTATAGGGAACTTGCGTGGAGGAACTAAT
CK2α dsRNA 1 DRSC1194 Rv	TAATACGACTCACTATAGGGAGTAGAGTTCTTCGGTGCCCA
CK2α dsRNA 2 DRSC34542 Fw	TAATACGACTCACTATAGGGAACCACCAACCGACTGAGAAC
CK2α dsRNA 2 DRSC34542 Rv	TAATACGACTCACTATAGGGATAACAACGCACTTTTCCGTG
CK2α dsRNA 3 DRSC34543 Fw	TAATACGACTCACTATAGGGACGATCCAAGATTTCACGACA
CK2α dsRNA 3 DRSC34543 Rv	TAATACGACTCACTATAGGGATGGCTTCTATCCCAAACTCG

Table 10 qPCR primers

Primer name	Sequence (5' - 3')
CK2β Fw	AGGCCGCCGAGATGCTCTAC
CK2β Rv	TGTCCGACAGACCCAATGGCA
CHRAC-14 Fw	TACAGACCCTCACCGAGCTA
CHRAC-14 Rv	TCGGCAGTGTTGGAATCCTT
RpL32 Fw	CGGATCGATATGCTAAGCTGT
RpL32 Rv	GCGCTTGTTCGATCCGTA

Gapdh1 Fw	CCATCGCAGCGCCATTCTCC
Gapdh1 Rv	CCGATGCGGCCAAATCCGTT
CG9815 Fw	TGATCGACAGCATTTTCGAGATTC
CG9815 Rv	TCCAGCTGCCACGGATTTTT
ATP8B Fw	TAAAGCGTGCCAAAAACGCA
ATP8B Rv	TGGAGAGGACAGCCTGAAGA
AdamTS-A Fw	ACAATTCCCGTGTCAGCGAT
AdamTS-A Rv	CTGCCAGGGCATTAGGTGAT
cmet Fw	AGGGTCCCTTAACAACGGAT
cmet Rv	CGCACTCCACCTCAAAATGC
mus301 Fw	ACAAGTTTCCCAGCAAGGGC
mus301 Rv	ACTCGTCCTGCCATTCATAAAG
Spc105 Fw	GTAAGGTGCCGACCAAGGAA
Spc105 Rv	GTGTTTACGAATTCCCTCACGG
DNApola60 Fw	TGTCCTCCTATGGCCTGTCA
DNApola60 Rv	CGACGGGTGAATTGTGGGTA
CENP-A Fw	ATACCCAAGCTGCCGTTCTC
CENP-A Rv	CCGCTGCGTCAAGTACATCT

6.8 Organisms

Table 11 *E. coli* strains

Name	Source	Genotype
DH5α	Erhardt lab	F- Phi80dlacZ DeltaM15 Delta(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rK-mK+)phoA supE44 lambda- thi-1
BL21 (DE3)	Erhardt lab	F- ompT hsdSB(rB-, mB-) gal dcm (DE3)

Table 12 *Drosophila melanogaster* fly stains

Gene	Source	Genotype
Oregon-R (Ore-R)	Erhardt lab	wild type
CHRAC-14 KG01051	Bloomington 13190	y^1 w^{67c23}; P{SUPor- P}mus201^{KG01051}Chrac-14^{KG01051}

Table 13 *Drosophila melanogaster* S2 cell lines

Name	Source
wild type	Erhardt lab
pMT-SV40-mAID-eGFP-Puro / pMT-OsTIR1-mCherry-Hygro	This study
pMT-SV40-mAID-AsiSI-eGFP-Puro / pMT-OsTIR1-mCherry-Hygro	This study
pMT-SV40-AsiSI-mAID-eGFP-Puro / pMT-OsTIR1-mCherry-Hygro	This study
pMT-SV40-mAID-V5-Puro / pMT-OsTIR1-Hygro	This study
pMT-SV40-mAID-AsiSI-V5-Puro / pMT-OsTIR1-Hygro	This study
pMT-SV40-AsiSI-mAID-V5-Puro / pMT-OsTIR1-Hygro	This study
pMT-CHRAC14-V5-His-Hygro	This study
pMT-CID-V5-His-Hygro	This study
pMT-CID-V5-His-Hygro / pMT-CHRAC14-Flag-Puro	This study
pMT-CK2β-V5-His-Hygro	This study
pMT-CK2β-V5-His-Hygro / pMT-CHRAC14-Flag-Puro	This study
pMT-CENP-A(wt)-V5-His-Hygro	This study
pMT-CENP-A(S20A)-V5-His-Hygro	This study
pMT-CENP-A(S20D)-V5-His-Hygro	This study

Name	Source
pMT-CENP-A(wt)-V5-His-Hygro / pMT-CHRAC14-Flag-Puro	This study
pMT-CENP-A(S20A)-V5-His-Hygro / pMT-CHRAC14-Flag-Puro	This study
pMT-CENP-A(S20D)-V5-His-Hygro / pMT-CHRAC14-Flag-Puro	This study
pMT-CHRAC14(T122A)-V5-Puro	This study
pMT-CHRAC14(T122D)-V5-Puro	This study

Table 14 Software

Name	Company
Illustrator CS6	Adobe
Mendeley Desktop	Mendeley
Microsoft Office	Microsoft
Photoshop CS6	Adobe
Scaffold4	Proteome Software
SeqMan Pro	DNASTAR
SnapGene Viewer	GSL Biotech
Sublime Text	Sublime HQ

7 Methods

7.1 Fruit fly methods

7.1.1 *Drosophila melanogaster* fly culture

Drosophila melanogaster flies were cultured in small or big plastic vials containing standard fly food with a 12 hour day-night cycle at 18°C or 25°C. Every third week flies were flipped into fresh vials to prevent mite reproduction. Old vials were frozen at -20°C for 24 h and discarded.

7.1.2 Embryo collection

Newly hatched flies from one to two big vials were moved into a small egg-laying cage mounted on a grape juice agar plate with yeast paste at the center. The flies were kept in the cages for two days to let them adapt to their new environment. Agar plates were changed once a day during maintenance to provide fresh yeast and to avoid hatching larvae. Embryos were then collected either from an overnight plate (non-staged) or in increments of 2 or more hours depending on the anticipated age of the embryo (staged). To achieve proper embryo staging the first plate in the morning was discarded because it contained fertilized eggs, which were retained by the females. To harvest embryos, yeast was removed from the plate with a spatula and embryos were detached in distilled water using a fine brush. The water containing the embryos was then drained in a collection tube fitted with a fine net and were washed thoroughly with water. For dechorionation, the embryos were incubated in 5% sodium hypochlorite for 1.5 min and then washed twice in distilled water. After that, embryos were collected into a pre-weighted reaction tube and the weight of the embryo pellet was determined. The tubes were then snap frozen in liquid nitrogen and stored at -80°C.

7.1.3 Virgin collection

Female *Drosophila melanogaster* virgin flies were collected 8 to 16 hours after removing adults from the vials. Virgin flies were selected based on their distinctive morphology: light body color, bigger body size, a dark spot in the transparent abdomen and unfolded wings. Before mating virgins were maintained for 4 to 5 days to ensure that they are unfertilized.

7.2 *Drosophila melanogaster* cell culture

7.2.1 S2 cells cultivation

Drosophila melanogaster Schneider S2 cells were obtained from late embryos as an immortalized cell line by I. Schneider (Rogers and Rogers, 2008; Schneider, 1972). The semi-adherent cell line was cultured at 25°C in the dark in Schneider Medium supplied with 10% heat inactivated fetal bovine serum (FBS) and 2% antibiotics (penicillin/streptomycin). Transgenic cell lines were kept under constant selection with hygromycin B (250 µg/ml) or puromycin (10 µg/ml). DIvA cell lines were cultured in medium supplemented with 200 µM auxin.

7.2.2 S2 cells freezing

Per cryotube ~ 2 x 10^6 cells were resuspended in 1 ml of freezing medium and placed into Mr Frosty freezing container. To gently cool down and freeze the cells, the container was put to -80°C for 24 h and cell lines were then transferred to a liquid nitrogen tank for long term storage.

7.2.3 S2 cells thawing

Vials were removed from liquid nitrogen, thawed in a 37°C water bath and subsequently transferred into a T-25 cell culture flask containing 4 ml of medium. After cells were settled (max. 15 min) medium was exchanged, and cells were left until full recovery before splitting them. Until full recovery was reached, medium was exchanged with conditional medium once a week.

7.2.4 S2 cells transfection

To generate transgenic cell lines, cells were transfected with a vector carrying the desired coding sequence and a drug resistance gene, which allowed constant selection with Hygromycin B or Puromycin. Transfection was performed in 6-well plates using Cellfectin II. For this, 1.5 x 10^6 cells per well were plated in 2 ml medium one day prior to treatment. For the transfection, 5 µg plasmid DNA and 4 µl well vortexed Cellfectin II reagent were diluted separately in 200 µl medium. The diluted Cellfectin II reagent was then mixed with the diluted plasmid dropwise and the solution was incubated for 45 min at RT with occasional mixing. Meanwhile, cells were washed three times with serum-free medium. To start the transfection, 600 µl serum-free medium was added dropwise to the 400 µl of Cellfectin II/plasmid mix yielding 1 ml of total transfection suspension. Washed cells were then incubated in the

transfection mix for 4 h and after that 1 ml of 20% serum medium was added. After 3 days Hygromycin B or Puromycin was added to select for transgenic cells and as soon as the wells were confluent, cells were resuspended and transferred to T-25 flask, adding 2 ml of fresh medium.

7.2.5 Induction of gene expression in transgenic S2 cells

Exogenous gene expression in transgenic cell lines transfected with vectors carrying the metal inducible metallothionein promoter (pMT) expression cassette (Bunch *et al.*, 1988) was induced by the addition of $CuSO_4$ to the medium.

7.2.6 DNA damage treatment in S2 cells

To induce acute DNA damage, cells were treated with 0.1 µg/ml Neocarzinostatin for 1 h. In order to cause site-specific DNA double strand breaks, expression of the restriction enzyme AsiSI was induced in the transgenic DIvA cell line. For this, auxin-containing medium was exchanged to 50 µM $CuSO_4$-containing medium and cells were incubated for 3 h to allow AsiSI expression and *in vivo* DNA restriction digestion.

7.2.7 RNA interference (RNAi)

Cells were seeded in a 6-well plate at a density of 1.5 x 10^6 cells per well one day prior to RNAi treatment. To introduce dsRNA into the cells, wells were washed thrice with 1 ml of serum-free medium and the cells were then incubated in 1 ml serum-free medium containing 15-20 µg dsRNA. After 1 h, 1 ml of 20% serum-containing medium was added and cells were incubated for 3 more days (day of RNAi treatment was day 0). For prolonged incubations up to 6 days, cells were resuspended in the well, transferred to a T-25 flask, volume was adjusted to 3 ml and another 15-20 µg of dsRNA was added. The amount of dsRNA and the treatment time span was dependent on the target protein's stability.

7.3 Nucleic acids methods and molecular cloning

7.3.1 Polymerase chain reaction (PCR)

PCRs were carried out using different polymerases (Q5, Taq, Taq mastermix) applying the polymerase-specific protocol. The number of amplification cycles typically ranged between 15 and 40 depending on the purpose of the experiment and the annealing temperature and

elongation time was dependent on the primer pair and the size of the anticipated fragment. Generally, the reaction followed a standard program:

1.	93-95°C	initial denaturation	10 sec – 5 min
2.	93-95°C	denaturation	10 – 30 sec
3.	50-72°C	primer annealing	10 – 30 sec
4.	68-72°C	elongation	10 sec – 2 min
5.	Cycle through steps 2 - 4 x15 - 40 times		
6.	68-72°C	final elongation	2 - 5 min
7.	4°C	hold	∞

7.3.2 Restriction digest

Digestion of DNA plasmids or fragments was done using restriction enzymes following the enzyme-specific protocol. Typically, in 25 – 50 µl total reaction volume, 0.5 – 1 µl enzyme, 2.5 – 5 µl buffer and 1-5 µg DNA were combined, and the mixtures were incubated for 1 h at the appropriate temperature.

7.3.3 Plasmid construction

To generate custom plasmids, Gibson assembly or Quick Ligase was used according to the supplier's recommendations. For Gibson assembly. 0.06 pmol insert was combined with 0.02 pmol restriction digested backbone and 7.5 µl Gibson assembly master mix to a total of 15 µl reaction volume and the samples were incubated at 50°C for 15 min. For ligation with Quick Ligase, 0.06 pmol insert was combined with 0.02 pmol restriction digested backbone, 10 µl Quick Ligase Reaction Buffer (2X) and 1 µl Quick Ligase to a total of 20 µl reaction volume. The mixtures were incubated at RT for 5 min.

7.3.4 Transformation of bacteria

E. coli cells were stimulated for plasmid DNA uptake by standard heat shock transformation. The chemically competent bacteria were thawed on ice and mixed with 1-5 µl of plasmid DNA by carefully stirring with the pipet tip. The mix was incubated on ice for 30 min and cells were heat shocked by shifting them to 42°C for 45 sec followed by an incubation on ice for 2 min. Subsequently, 1 ml of pre-warmed LB medium was added, and the culture was incubated for 45 min under constant shaking at 37°C. To reduce the culture volume for proper plating, the cells were quick spun down and liquid was withdrawn leaving 50 µl behind in which

the pellet was resuspended. This cell suspension was then spread on a pre-warmed LB agar plate and incubated over night at 37°C.

7.3.5 Isolation of plasmid DNA from *E. coli* cultures

Plasmid DNA purification from bacterial small or medium sized cultures was done with the NucleoSpin Plasmid or NucleoBond Xtra Midi kit following the manufacturer's protocols. 5 or 100 ml of LB supplied with selective antibiotics was inoculated with one bacterial colony and grown over night on a shaker at 37°C beforehand. The cultures were harvested, lysed and then loaded on the column for binding and washing of the relieved plasmid DNA. Elution was done in 50 or 100 µl of elution buffer and the yield was quantified using a Nanodrop spectrophotometer.

7.3.6 Colony PCR

To screen bacterial colonies for the presence of the correct plasmid, one colony was dissolved in 20 µl of water and 1-5 µl of this mix was subjected to a standard PCR using the Red Load Taq Master Mix. The remaining bacterial solution was used to inoculate overnight cultures in LB medium.

7.3.7 Agarose gel electrophoresis

Nucleic acid fragments were separated according to their size using agarose gel electrophoresis. The percentage of the agarose gel was dependent on the size of the fragment to be visualized (typically 1-3%). The respective amount of agarose was dissolved in 50 or 150 ml of 1X TAE buffer by boiling in a microwave and after cooling down ethidium bromide was added to a final concentration of 0.001%. After the agarose was poured into a tray for solidification the gel was placed into an electrophoresis tank and samples containing 1X gel loading dye next to an appropriate marker were loaded into the wells. The electrophoresis was run at 100 V for 30-60 min. To visualize DNA bands the gel was exposed to UV light in a Gel Doc XR+ apparatus.

7.3.8 RNA extraction

RNA was extracted from tissue or cells with TRIsure by following the supplier's recommendations. Briefly, snap frozen samples were lysed in 1 ml of TRIsure and in case of fly tissue or embryos a small pistil was used to crush the material. After incubation at RT for 5 min 200 µl of chloroform was added and tubes were shaken by hand for 15 sec. After a 3 min

incubation at RT the phases were separated by centrifugation at 12,000 x g for 15 min at 4°C and the aqueous phase was transferred into a fresh tube. RNA was precipitated by mixing with 500 µl isopropyl alcohol, incubation at RT for 10 min and subsequent centrifugation at 12,000 x g for 10 min at 4°C. The RNA pellet was then washed with 1 ml of freshly prepared 75% ethanol by vortexing and centrifuged again at 7,500 x g for 5 min at 4°C. Ethanol was removed and remaining liquid traces were collected by a quick spin and removed with a small pipet tip. The RNA pellet was then air dried for 5 min and resuspended in sterile 25-50 µl RNase free water and stored at -80°C. If used for qPCR, RNA was further purified and depleted of contaminating genomic DNA with the RNA Clean & Concentrator-5 kit, following the manufacturer's instructions.

7.3.9 Extraction of genomic DNA from flies

One male and one female snap frozen fly were grinded up with a pistil in separate tubes each containing 50 µl of squish buffer with proteinase K. Proteins were then digested at 37°C for 30 min and the enzyme was inactivated at 95°C for 2 min. The debris was separated by centrifugation for 7 min at max speed and the supernatants of both samples were mixed in a fresh tube and stored at 4°C.

7.4 Biochemical methods

7.4.1 Antibody to dynabeads crosslinking

For one standard IP from 2 mio S2 cells, 20 µl magnetic protein-A or -G coupled dynabeads slurry was crosslinked to 3 µg antibodies. For this, all solutions were prepared freshly, and all incubations steps were carried out for 30 min at RT on a rotation wheel. The appropriate volume of bead slurry was washed thrice in 1X PBS and mixed with the antibodies in 500 µl 1X PBS for binding. Then, beads with captured antibodies were equilibrated in crosslink buffer and crosslinking was induced by resuspension in DMP-crosslink buffer. This step was repeated and afterwards, the crosslinking was quenched by washing the beads in 100 mM glycine pH 7.0 . To remove uncoupled antibody, the beads were then incubated in 100 mM glycine pH 2.5. Finally, the beads were briefly washed in 1X PBS and resuspended in 1X PBS supplemented with 0.01% sodium azide and stored at 4°C.

7.4.2 Protein quantification

Proteins were quantified using the Pierce 660nm Protein Assay kit in combination with the Ionic Detergent Compatibility reagent for samples in RIPA or Laemmli Buffer. The manufacturer's instructions were followed.

7.4.3 Radioactive kinase assay

In a 20 µl reaction, 2 µg substrate was combined with 0.5 U hCK2, 10X kinase assay buffer (included with commercial hCK2), 200 µM ATP, 10 µCi [γ-^{32}P]ATP and deionized water. Samples were incubated at 30°C for 30 min in a thermo mixer and reactions were terminated by the addition of 7 µl 4X Laemmli buffer and boiling at 95°C for 5 min. Proteins were then resolved on a SDS gel by gel electrophoresis and afterwards the gel was stained with Coomassie staining buffer for 10 minutes and destained for 1 h in detaining solution. Then the gels were fitted on a piece of thin blotting paper, covered in cling foil and the pocket was sealed with lab tape along the sides. The gel was then dried in a gel dryer at 80°C for 2 h followed by exposure to a storage phosphor screen for 5 minutes. Subsequently, the autoradiograph was visualized with a laser scanner.

7.4.4 GST affinity purification of recombinant proteins

A pre-culture was started by inoculating 100 ml LB medium with bacteria and the appropriate antibiotics were added. The culture was grown over night at 37°C under constant shaking. On the next day, the pre-culture was expanded by a 1:10 dilution into 1-2 l of fresh LB medium and bacteria were grown until a density of OD_{600} = 0.6-0.9 was reached. Next, the cultures were cooled down and protein expression was induced with 0.5 mM IPTG over night at 18°C under constant shaking. On the following day cells were harvested at 3000 rpm for 30 min and the pellet was washed once in ice-cold 1X PBS. Then the pellet was snap frozen in liquid nitrogen and stored at -80 °C until use. On the day of purification, the pellet was defrosted on ice and resuspended in 25 ml bacteria lysis buffer and 1 mg/ml Lysozyme and 2 µl Basemuncher per g of cell pellet was added and incubated for 15 min on a roller at 4°C. Subsequently, cell walls were disrupted with a microfluizider using 5-7 cycles and the lysate was cleared by centrifugation at max speed for 30 min at 4°C. Immediately afterwards, the supernatant was separated from debris, filtered through a 0.45 µM syringe filter, and prepared for injection into the equilibrated ÄKTA purification system via a 50 ml super loop. The proteins were bound to the column affinity matrix (GSTrap HP 1ml) and washed with 25 column volumes of 1X PBS at a flow rate of 0.5 - 1 ml/min. Eventually, the purified proteins were released from the column using a gradient elution with Glutathione at a flow rate of 0.5 - 1 ml/min and 0.5 ml fractions were collected. The fractions with high protein content previously determined by SDS-PAGE were pooled and the buffer was exchanged to PreScission protease cleavage buffer with a PD-10 column. Afterwards the proteins were concentrated with a Centricon and submitted for overnight cleavage at 7°C according to the manufacturer's instructions. The liberated tags and the proteases were then eliminated by reapplying the reactions to the ÄKTA system mounted with a GSTrap column and collecting the flow through.

Again, proteins were concentrated with a Centricon, quantified and analyzed by SDS-PAGE. For storage the samples were then snap frozen in 50% glycerol and held at -80°C.

7.4.5 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gels were hand casted according to the composition shown in the materials section 6.3 using Bio-Rad gel pouring stations. To prevent gel leakage, the glass plates were sealed at the bottom with a layer of 300 µl separation gel mix supplemented with 3 µl of TEMED. Denatured protein samples containing 1X Laemmli buffer were then separated according to their apparent weight using the Bio-Rad Mini-PROTEAN Tetra Cell with 1X running buffer under constant 100-120 V until the desired level of protein separation was reached (1-2 h). The stacking gel and sealing layer were then clipped off and the gel was subjected to blotting or Coomassie staining.

7.4.6 Immunoprecipitation (IP)

All incubations and steps were performed at 4°C on a rotation wheel unless specified otherwise and magnetic beads were separated from liquid with a magnetic stand. Protein low binding 1.5 ml reaction tubes were used. For one pull down sample typically 2 x 10^7 cells were settled by centrifugation (800 x g, 5 min) and lysed in 200 µl IP lysis buffer by pipetting up and down and followed by a 5 min incubation step. Cells were then further ruptured by sonication with a Bioruptor (settings: 5 cycles of 30 sec sonication followed by 30 sec pause; 'HIGH' function toggled). Optionally, protein isolation was facilitated by enzymatically digesting the chromatin with 75 U MNase per 1 x 10^7 cells activated by the addition of 2 mM $CaCl_2$ for 30 min at 26°C and 650 rpm on a thermoshaker. After inactivating MNase with 10 mM EGTA, cell debris was eliminated by centrifugation at 16,000 x g for 30 min, and the cleared supernatant was incubated with 30 µl of equilibrated magnetic protein-G or protein-A magnetic dynabeads crosslinked to 3 µg antibody for 2 h. To account for unspecific binding, an isotype IgG control was always included. Captured protein complexes were then washed three times for 10 min with fresh lysis buffer and were eluted from the beads by incubation with 30 µl 1X Laemmli buffer at 95°C for 5 min. To assess the pulldown efficiency, typically 7.5% of the input lysate and the flow through and 50 - 100% of the eluate was analyzed by SDS-PAGE and Western blotting. Samples were stored at -20°C.

7.4.7 Immunoprecipitation for mass spectrometry

Cenp-A-V5-His and Chrac-14-V5-His were mildly expressed in stably transfected S2 cells via induction of the *Drosophila* metallothionein gene promoter (pMT, [Bunch *et al.*, 1988])

by addition of 1 µM $CuSO_4$ for 10 h. All further steps were performed on ice or at 4°C unless specified otherwise. For four pulldown samples per cell line (V5-Control, V5-Damage, IgG-Control, IgG-Damage) a total of 8 x 10^8 cells were settled by centrifugation (800 x g, 20 min), washed three times in 1X PBS and lysed for 1 h on a rotation wheel in 6 ml MS lysis buffer. Cells were further ruptured by douncing and sonication (Bioruptor, settings: 5 cycles of 30 sec sonication followed by 30 sec pause; “HIGH” function toggled). To facilitate protein isolation, the chromatin was enzymatically digested with 3,750 U MNase for 30 min at 26°C and 650 rpm on a thermoshaker. After cell debris was eliminated by centrifugation at 16,000 x g for 1 h, the cleared supernatant was split into four equal samples (2 x 10^8 cells per sample and incubated with 8.75 µg V5 antibody crosslinked to 40 µl magnetic protein-G magnetic beads for 2 h on a rotation wheel. To account for unspecific binding, the two remaining samples were incubated with an isotype IgG control equally crosslinked to magnetic beads. Captured protein complexes were then washed six times for 10 min with fresh lysis buffer lacking triton X-100 on a rotation wheel and were eluted by incubation with 30 µl Laemmli buffer at 95°C for 5 min. The pulldowns were conducted in duplicates. To identify interaction candidates immunoprecipitate eluates were subjected to quantitative LC-MS/MS. For this, samples were submitted to the EMBL proteomics core facility where an in-solution tryptic digest was conducted followed by sample labelling with TMT10plex, multiplexing in a 1:1 ratio and analysis on a Q Exactive Plus instrument. The protein.txt output files of IsobarQuant were analyzed using the R programming language (https://www.r-project.org). For the analysis only proteins with a unique peptide count of at least two were considered. Briefly, summed TMT reporter ion intensities (signal_sum columns) were cleaned for batch effects using *limma* (Ritchie *et al.*, 2015) and normalized using vsn (variance stabilization normalization). V5 antibody and IgG control pulldowns were normalized separately to maintain the abundance difference between both groups. Proteins were tested for differential expression using the *limma* package (Ritchie *et al.*, 2015) in order to score them according to their interaction strength. A protein was interpreted as a hit when the false discovery rate (fdr) was smaller than 5 % and the fold-change was at least 100 %. With a fdr below 20 % and a fold-change of at least 50 % proteins were interpreted as candidates.

7.4.8 **Immunoblotting (Western blotting)**

Immunological detection of proteins separated according to their apparent weight on a SDS-PAGE gel was done by Western blotting. Gels were briefly rinsed in wet blotting buffer and a wet transfer was conducted using the Bio-Rad transfer system with the following setup: Submerged in buffer, one foam pad, one thick blotting paper, one 0.2 µm nitrocellulose membrane, one SDS-PAGE gel, one thick blotting paper and one foam pad were assembled within the transfer cassette, which then was placed into the transfer tank together with a cooling

unit to prevent overheating and a stirring bar was added. The transfer was conducted for 1.5 – 2 h under constant 100 V at 4°C under constant stirring. For a semi-dry protein transfer, the set up was as follows: pre-soaked in blotting buffer, one extra thick Whatman paper, 0.2 µm nitrocellulose membrane, SDS-PAGE gel and one extra thick Whatman paper were stacked onto the transfer machine and blotting was done at 23-24 constant Volt for 45 min. Following transfer, the membrane was washed in distilled water and briefly stained in Ponceau-S to check the transfer efficiency. Subsequently, the membrane was blocked for 30 min in 5% blocking buffer and then incubated 5% milk in 1X TBST supplied with appropriate primary antibody overnight on a roller at 4°C. Excess antibody was washed off by washing thrice in 1X TBST for 10 min, followed by a 1 h incubation in 5% milk in 1X TBST supplied with secondary antibodies at RT on a roller. To reduce unspecific binding, the membrane was again washed three times in 1X TBST for 10 min each. The proteins of interest were then visualized by incubating the membrane in ECL substrate for 2 min followed by chemiluminescent detection with a LAS imaging system.

7.5 Molecular Biology methods

7.5.1 Complementary DNA synthesis

The QuantiTect Reverse Transcription kit was used to prepare complementary DNA (cDNA) from 1 µg fly or S2 cell RNA and the manufacturer's instructions were followed. Genomic DNA (gDNA) remains were first digested in a total reaction volume of 14 µl and incubation at 42°C for 10 min. The reaction volume was then divided and subjected to both a reverse transcription reaction and a control reaction by mixing 0.5 µl RT primer mix, 2 µl RT buffer and 0.5 µl reverse transcriptase or sterile water with 7 µl of gDNA-digested RNA. The reactions were incubated at 42°C for 30 min and then inactivated at 95°C for 3 minutes. cDNA was usually diluted with sterile water 1:20 and stored at -20°C.

7.5.2 Quantitative PCR

In order to quantify transcript abundance, I used LightCycler 480 SYBR Green I Master and generally adhered to the appendant protocol with a downscaled the reaction volume. In a 15 µl reaction, 0.75 µl forward and 0.75 µl reverse primers (0.5 µM final), 5 µl diluted cDNA and 7.5 µl master mix were mixed and dispersed into a LightCycler 480 multiwell plate 384. The plate was sealed with LightCycler 480 multiwell sealing foil and liquid was collected at the bottom of the plate by centrifugation for 1 min at 2,000 x g. The reactions were always performed in technical triplicates and non-reverse transcribed control cDNA for each sample

was included as well as a water control. The reactions were subjected to a standard qPCR program:

1.	95°C	Initial denaturation	5 min
2.	95°C	Denaturation	10 sec
3.	55°C	Primer annealing	10 sec
4.	72°C	Elongation	10 sec
5.	Cycle through steps 2 – 4 x40 times		
6.	95°C	Melting curve	5 sec
7.	65°C		1 min
8.	97°C		amplicon dependent
9.	40°C	Cooling	10 sec

After the run, the product specificity was assessed by evaluation of the melting curves and transcript levels were analyzed by averaging the Ct values of the triplicates and normalization to the housekeeping gene expression using the following formula:

$$\Delta Ct = 2^{-(Ct\ target\ -\ Ct\ house\ keeping)}$$

7.5.3 Double-stranded RNA production

For the preparation of double-stranded RNA (dsRNA) for RNAi experiments the MEGAscript RNAi kit was used according to the manufacturer's instructions. The template for *in vitro* transcription was generated beforehand by PCR using cDNA with primers flanked with the T7 promoter. The PCR product was column purified, eluted in 25 µl sterile water and 8 µl were mixed with the T7 *in vitro* transcription components. The reaction was incubated at 37°C for max 16 h and dsRNA was subsequently annealed in a PCR machine by heating to 75°C for 5 min and ramping down the temperature to 25°C, over the time course of 1 h. Thereafter, remaining single stranded RNA (ssRNA) and plasmid DNA were digested at 37°C for 1 h and the dsRNA was column purified according to the kit protocol. Elution from the column was done by pipetting 100 µl elution buffer in the column, heating to 65°C for 2 min and subsequent centrifugation. This was done twice, and the eluates were pooled and quantified using a Nanodrop spectrometer. To prevent repeated freeze-thaw cycles, dsRNA was stored as 15 or 20 µg aliquots at -20°C.

7.6 Microscopy

7.6.1 Metaphase arrest and chromosome spreads

2 x 10^6 cells were seeded in a 6-well plate one day prior to the experiment. On the next day, 2 x 10^5 exponentially growing cells per sample were arrested in metaphase with 1:3 colcemid for 1-2 hours at room temperature on a rotating wheel. Subsequently, cells were harvested at 800 x g for 5 min and were resuspended in 500 µl of hypotonic sodium citrate solution (0.5% Na-citrate in ddH_2O). After a 7-10 min incubation the swollen cells were spun onto positively charged microscopy glass slides mounted onto cytology funnels in a cytospin centrifuge at 900 rpm for 10 min. Then, chromosomes were fixed in 4% paraformaldehyde in 1X PBS for 10 min at RT, briefly washed in 1X PBS and subjected to IF, continuing with the cell permeabilization step.

7.6.2 Immunofluorescence (IF)

For IF 1.5 x 10^6 cells were seeded in a 6-well plate one day before fixation. Then 4 x 10^5 cells per slide were harvested by centrifugation for 5 min at 800 x g and re-suspend in 50 µl 1X PBS. The cell solution was spotted onto a positively charged glass slide and cells were allowed to settle for 10 minutes. After fixing the cells in 4% paraformaldehyde in 1X PBS for 10 minutes at RT, cells were briefly washed in 1X PBS. All washes from here were done in coplin jars on a shaker unless stated otherwise. The cells were then permeabilized with 0.1% TX-100 in 1X PBS for 5 min at room temperature and washed thrice for 10 minutes in 1X PBS. Subsequently, cells were blocked for 30 min in 50 µl 4% BSA, 1X PBS at RT before primary antibodies diluted in 50 µl 4% BSA, 1X PBS were added. The spotted cells were covered with a piece of parafilm and incubated over night at 4°C in a wet chamber. Afterwards cells were washed thrice with 1X PBS for 10 min and were incubated with fluorophore-coupled secondary antibodies diluted in 50 µl 4% BSA ,1X PBS for 2 hours at RT in the dark in a wet chamber, again covered with parafilm. After washing cells thrice for 10 min in 1X PBS, cells were counterstained for 5 min with 1 µg/µl DAPI in diluted in 1X PBS at RT. Then cells were briefly washed three times with 1X PBS and mounted using Aqua Polymount and coverslips were sealed with nail polish. After leaving the slides at RT in the dark overnight they were stored at 4℃ in the dark until imaging.

7.6.3 Image acquisition

For image acquisition a wide field fluorescence microscope with a DeltaVision Core system (Applied Precision) and a charge-coupled device camera (CoolSNAP HQ2, Photometrics) in combination with the softWoRx v.5.5 suite as operation software was used.

Images were acquired with 100x UPlan- SApochromat (NA 1.4), 60x Plan-Apochromat N (NA 1.42) or 40x UApo (NA 1.35) objective lenses and with a binning of 1X 1 or 2 x 2. Images were usually acquired as z-stacks (0.2 or 0.3 µm distance).

7.6.4 **Image analysis**

Acquired images with the Olympus system were usually deconvolved (conservative ratio, 10 cycles), maximum projected or kept as single stacks, cropped and exported as TIFF images using the softWoRx 6.5.2 suite software. Further image processing and analysis was then done in Fiji (Schindelin *et al.*, 2012). Background signals were subtracted using a rolling ball radius. Brightness and contrast were adjusted on concatenated images, to achieve comparable signals between images and conditions.

7.7 NGS methods

7.7.1 **CENP-A and yH2Av CUT&Tag**

The CUT&Tag transposase Tn5 was loaded with annealed adapter oligos as described online in Bryson and Henikoff, 2019. The single stranded oligos were resuspended to 200 µM in annealing buffer and 100 µl ME-rev was combined separately with both 100 µl ME-A and 100 µl ME-B yielding equimolar mixtures. The oligos were then annealed by heating to 95°C for 2 minutes, temperature increment decrease (5 min incubation steps lowering the temperature 5°C each time) eventually reaching an incubation temperature of 25°C for 5 minutes. 0.8 µl of both annealed oligos were then mixed with 10 µl 3xFlag-pATn5 (2 mg/ml in 50% glycerol) and binding was allowed for 1h at RT. The Tn5-adapter complex was stored at -20°C.

CHRACH-14 or Brown as a control were knocked down in S2 cells for 6 days via RNAi and on the last day, DNA damage was induced for 3 h by the addition of 50 µM $CuSO_4$ and depletion of auxin from the medium. After the treatment cells were harvested and directly subjected to CUT&Tag.

CUT&Tag for S2 cell nuclei was essentially performed as published by the Henikoff lab in Kaya-Okur *et al.*, 2019 with certain adjusted steps for S2 cells and CENP-A. A magnetic stand was used in all steps to separate magnetic beads from liquid. All incubation steps were carried out on a nutator unless stated otherwise. 20 µl magnetic Concavalin A (ConA) beads per sample were transferred into 1.6 ml binding buffer. Liquid was withdrawn with a magnetic stand and beads were washed in 1.5 ml binding buffer and resuspended in 20 µl binding buffer

per sample and held until use. Cells were harvested by centrifugation at 800 x g for 5 min and after washing the cell pellet in 500 µl 1X PBS S2 nuclei were extracted by incubation in 500 µl ice-cold NE1 buffer for 10 min on ice. After centrifugation at 1,300 x g at 4°C for 4 min the remaining pellet was washed in 500 µl of 1X PBS and centrifuged again at 1,300 x g at 4°C for 4 min. For binding of the nuclei to the ConA beads, the pellet was resuspended in 500 µl wash buffer in a 2 ml tube, the equilibrated ConA bead slurry was added dropwise, and binding was allowed for 10 min at RT on a rotation wheel. Successful binding was indicated by clumping of the beads. The bead-bound nuclei were then dispersed into separate tubes according to the experimental sample setup. After withdrawing the liquid, beads were resuspended in 100 µl primary antibody containing (1:100) rabbit was diluted 1:100 in 100 µl Dig-wash buffer per sample and beads were incubated at RT for 1 h. The beads were then washed thrice by brief resuspension in 1 ml of Dig-wash buffer. For binding of the pA-Tn5 adapter complex, pA-Tn5 was diluted 1:250 in 100 µl Dig-300 buffer per sample and beads were incubated with the tagmentase for 1 h at RT and afterwards were washed thrice with 1 ml of Dig-300 buffer. To activate the tagmentation beads were then resuspended in 300 µl tagmentation buffer and were transferred into 0.5 ml tubes. The samples were then incubated for 1 h in a 37°C room. Tagmentation was stopped and DNA was solubilized by the addition of 10 µl 0.5 M EDTA and 3 µl 10% SDS and full speed vortexing for several seconds. Before DNA extraction, proteins were degraded by the addition of 50 mg ProteinaseK and incubation at 50°C in a thermomixer for 1h. To isolate total DNA, 300 µl phenol/chloroform was added and the reactions were mixed by full speed vortexing for several seconds. The reactions were transferred into a Phase-lock tube and the liquid phase was separated via centrifugation for 3 min at 16 000 x g for 30 sec. 300 µl chloroform was added to the tube and mixed by inverting 10 times and the tubes were centrifuged again for 3 min at 16,000 x g for 30 sec. The aqueous layer was withdrawn, and DNA was precipitated by 750 µl 100% ethanol and centrifugation for 15 min at 4°C at 16,000 x g. The DNA pellets were washed in 1 ml 100% ethanol and after spinning for 5 min at 4°C t 16,000 x g ethanol was withdrawn, and pellets were air dried. Remaining RNA was digested by 1/400 RNaseA (10 mg/ml) in 25 µl TE buffer per sample at 37°C for 10 min. To generate libraries from the tagged DNA fragments, 21 µl of the extracted CaT DNA was mixed with 2 µl universal i5 primer (10 µM), 2 µl sample specific barcoded i7 primer (10 µM) and 25 µl NEBNext 2x master mix and subjected to a library PCR:

72 °C for 5 min (gap filling)

98 °C for 30 sec

98 °C for 10 sec

63 °C for 10 sec

Repeat cycle 3. + 4. 13 times (14 total)

72 °C for 1 min

Hold at 8 °C

After library PCR a two-sided size selection (0.4 – 1.1) was performed using Ampure XP beads according to the manufacturer's instructions to achieve a PCR product size range of 150 – 1000 bp. Libraries were quantified with a Qubit (DNA HS assay) and by Céline Schneider from the Kaessmann lab (ZMBH) using a Fragment Analyser and the HS NGS Fragment kit. From this analysis the average size was assessed, and the libraries were then pooled to a total of 4 nM. The pool was again purified with Ampure XP beads and analyzed with a Qubit and on a Fragment Analyser, as described above. Sample preparation and paired end sequencing was performed by Kai Fenzl in the Bukau lab (ZMBH) on a NextSeq550 (Illumina) system using the 150 cycles NextSeq 550 System High-Output Kit (Illumina). For this 1.8 pM diluted library was loaded on the flow cell.

7.7.2 **Total RNA Sequencing**

In triplicates, total RNA was extracted from 2 x 10^6 S2 cells depleted of Brown or CHRAC-14 by RNAi and from staged Oregon-R or CHRAC-$14^{KG01051}$ mutant embryos (4-6 h, 5 days old parents) with TRISure. The RNA was then additionally column purified with the RNA Clean & Concentrator-5 kit including an on-column DNAse treatment. RNA was eluted in 25 µl. Libraries from 0.5 µg pure RNA were then prepared with the Universal Plus Total RNA-Seq with NuQuant kit using 12 amplification cycles. The library molarity was afterwards measured with NuQuant and the size profile was assessed by Céline Schneider from the Kaessmann lab (ZMBH) using a Fragment Analyser and the HS NGS kit. Then, 11 nM library stocks were prepared according to the NuQuant measurements in 20 µl water and from these a pool of 10 nM in 200 µl was set up. This pool was cleaned up again with Ampure XP beads (x 1.2) and eluted in 200 µl 10 mM Tris-HCl pH 8.5. This pool was again quantified to determine the exact molarity after the clean-up using the Qubit. The pool was finally diluted to 4 nM, quantified again with the Qubit and finally submitted for sequencing. Sample preparation and paired end sequencing was performed by Anja Schubert and Jaro Schmitt in the Bukau lab (ZMBH) on a NextSeq550 system using the 150 cycles NextSeq 550 System High-Output kit.

7.8 Bioinformatics

For bioinformatic analysis of sequencing datasets, UNIX-shell scripting was used, and data was stored and processed in the bwForCluster MLS&WISO (Production) cluster, partly utilizing the software embedded in the environment.

7.8.1 CUT&Tag bioinformatics analysis

Generally, the analysis until read calibration was tightly oriented towards the published CUT&Tag analysis online tutorial (Zheng *et al.*, 2020).

In the beginning, acquired reads were trimmed from adapters with TrimGalore 0.6.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using the built in FastQC quality assessment option in order to validate sufficient quality of the reads before proceeding with the analysis.

```
trim_galore \
--phred33 \
--fastqc \
--dont_gzip \
--a adaptersequence \
--paired read1 read2 \
```

Subsequently, reads were aligned to the current *Drosophila* genome and to new assemblies from the Mellone and Larracuente labs, which contain annotated *Drosophila* centromeres (PacBio and ChIPtigs) (Table 15). Reads were also aligned to the *E. coli* genome (Table 15). For the alignment bowtie2 v. 2.3.0 (Langmead and Salzberg, 2012) was used. Hereby, single-end mode for Mellone and Larracuente assemblies and paired-end mode for the normal *Drosophila* reference was applied. The *E. coli* alignment was done in both modes. The phred quality filter was set to 33 and maximal fragment length cut-off was set to 700 bp. Genome assembly files used in this analysis are specified in the table below.

Table 15 Genome data and files for CUT&Tag read alignment

Genome	File	Online Source
Current Berkeley *Drosophila* Genome Project (BDGP) assembly (Adams *et al.*, 2000; Hoskins *et al.*, 2015)	`Drosophila_melanogaster.BDGP6.28.dna.toplevel.fa`	Ensembl genome browser (Yates *et al.*, 2020)
Escherichia coli strain K-12 substrain MG1655 (Blattner *et al.*, 1997; Freddolino *et al.*, 2012)	`Escherichia_coli_str_k_12_substr_mg1655.ASM584v2.dna.toplevel.fa from`	Ensembl Bacteria genome browser Howe *et al.*, 2020
Cenp-A *de novo* ChIPtigs (Chang *et al.*, 2019b)	`File.S6.Chang_et_al.fasta`	Dryad repository of Chang *et al.*, 2019
PacBio assembly (Chang *et al.*, 2019b)	`File.S8.Chang_et_al.fasta`	Dryad repository of Chang *et al.*, 2019

To generate index files for the genome assemblies prior to the alignment following command was used:

```
bowtie2-build \
-genome \
-outname \
```

Commands used for the alignments were as follows:

BDGP:

```
bowtie2 \
--local \
--very-sensitive \
--no-mixed \
--no-discordant \
--phred33 \
-I 10 \
-X 700 \
-x index \
--1 read1.fq \
--2 read2.fq \
-S outname.sam &> outname_statistics.txt
```

PacBio and ChIPtigs:

```
bowtie2 \
--local \
--very-sensitive \
--no-mixed \
--no-discordant \
--phred33 \
-I 10 \
-X 700 \
-x index \
-U read1.fq \
-S outname.sam &> outname_statistics.txt
```

Escherichia coli:

```
bowtie2 \
--local \
--very-sensitive \
--no-overlap \
--no-dovetail \
--no-mixed \
--no-discordant \
--phred33 \
-I 10 \
-X 700 \
-x index \
--1 read1.fq \
--2 read2.fq \
-S outname.sam &> outname_statistics.txt
```

(Alternatively, '-U read1.fq' for unpaired alignments instead of specifying '--1' and '--2' was used)

Aligned data was then quality filtered and sorted with SAMtools v. 1.3.1 (Danecek *et al.*, 2021; Li *et al.*, 2009). Here '-F 0x04' excludes unmapped reads, '-f 0x02' keeps proper pairs in paired-end alignments and '-q [0-44]' defines the mapping quality of the output. For ChIPtigs q 10', for PacBio '-q 30' and for BDGP '-q 20' was used. For the *E. coli* data respective variations were run to generate matching files for subsequent calibration of the different other data sets. A command example is given below.

```
samtools view \
-q 20 \
-F 0x04 \
-f 0x02 \
-h \
-O SAM file.sam |

samtools sort \
-o file_q20filtered_sorted.sam \
-O SAM
```

After that, duplicates were identified in all datasets and removed only from PacBio and *E. coli* data using Picard v. 2.20.0 (http://broadinstitute.github.io/picard/).

```
java -jar /opt/bwhpc/common/bio/picard/2.20.0-java_jdk-1.8.0/lib/picard.jar \
MarkDuplicates \
I=file.sam \
REMOVE_DUPLICATES=true \
O=file_dupRm.sam \
METRICS_FILE=file_dupRm_report.txt
```

Next, files were converted to bed format (including '-bedpe' for paired-end data), fragments above 1000 bp were filtered out and for the upcoming spike-in calibration the length of each read was calculated and pasted as a new tab separated column. SAMtools v. 1.3.1 (Danecek *et al.*, 2021; Li *et al.*, 2009) and BEDTools v. 2.26 (Quinlan and Hall, 2010) and the following command pipe were used:

```
samtools sort -n file.sam |
samtools view -Sb |
bedtools bamtobed -bedpe|
bedtools sort |
awk '$1==$4 && $6-$2 < 1000 {print $0}' |
cat | awk -v OFS='\t' '{len = $3 - $2; print $0, len }' > file_clean.bed
```

Finally, the data were calibrated with the corresponding *E. coli* aligned files. In the course of the script, which essentially employs the BEDTools command 'genomecov' (Quinlan and Hall, 2010), files are also converted to bedgraph format. The script I used is part of the CUT&Tag analysis tutorial and can be separately downloaded from Github (https://github.com/Henikoff/Cut-and-Run/blob/master/spike_in_calibration.csh). For the calibration it is necessary to provide a record of the sizes of all the chromosomes present in the file to be calibrated. This chromosome length file was generated using a custom code by extracting and formatting the header of any of the aligned .sam files as follows (example given for a BDGP file):

```
samtools view \
-H \
-o chrom_legth_dm6.sam \
Cenp-A_Bw_C_Rep1_S5_dm6.sam

grep "SN" chrom_legth_dm6.sam |
cut -f2,3 |
sed 's/SN://g'$1 |
sed 's/LN://g'$2 > chrom_length_dm6
```

The cshell script for the calibration was then run with the command given below, specifying seven arguments: 1. file to be calibrated (genome.bed), 2. spike-in file (spike_genome.bed), 3. scale factor, 4. output format for genomecov, 5. chromosome length file, 6. minimum fragment length, 7. maximum fragment length.

```
spike_in_calibration.csh \
genome.bed \
Ecoli_file.bed \
1000000 \
bg \
chrom_length_file \
1 \
1000
```

For further analysis purposes, the bedgraph files were converted into bigwig files with a UCSC Genome Browser software (Kent *et al.*, 2010):

```
bedGraphToBigWig \
file.bg \
chrom_length_file \
outname.bw
```

The calibrated bedgraph files were used for peak calling with MACS2 v. 2-2.2.6 (Zhang *et al.*, 2008).

```
macs2 callpeak \
-t file.bg \
-c controlfile.bg \
-p 0.01 \
-f BED \
-g dm \
--nomodel \
--outdir outdir \
-n outname 2> outname_report.txt
```

For peak annotation, overlapping replicates, peak coordinate extraction, data plotting, calculation of fold changes and further sequence analyses the R programming language v. 4.0.3 (https://www.r-project.org/) and RStudio (http://www.rstudio.com/) was used utilizing basic R packages (R base, ggplot etc.) and Bioconductor packages (e.g. ChIPpeakAnno) (Gentleman *et al.*, 2004; Zhu *et al.*, 2010). To annotate centromeric peaks or to extract coordinates of repeats a custom repeat annotation file from the Mellone and Larracuente labs was used (File.S9.Chang_et_al.gff from Chang *et al.*, 2019).

deepTools v. 3.5.0 (Ramírez *et al.*, 2016) was used to compute average peak profiles across defined regions with ‘computeMatrix’ and ‘plotProfile’. To calculate coverage summaries over regions of interest ‘multiBigwigSummary’ was used.

```
computeMatrix reference-point \
-S file1.bigwig file2.bigwig file3.bigwig \
   file4.bigwig file5.bigwig \
-R coordinates1.bed coordinates2.bed \
--outFileNameMatrix outname.txt \
-a 3500 \
-b 3500 \
--binSize 350 \
--samplesLabel “label1” “label2” “label3” “label4” \
   “label5” \
-out outname_mat.gz

plotProfile \
-m outname_mat.gz \
--startLabel "start" \
--endLabel "end" \
--plotWidth 6 \
```

```
--plotType se \
--legendLocation upper-left \
--colors blue grey \
--regionsLabel "label1" "label2" \
-out outplot.png"

multiBigwigSummary BED-file \
--bwfiles file1.bw file2.bw file3.bw file4.bw file5.bw \
--BED region.bed \
--labels IgG Bw-control Ch14-control Bw-damage Ch14-damage \
-o outname.npz \
--outRawCounts outname.tab
```

Mapping of AsiSI recognition motifs:

HOMER (Heinz *et al.*, 2010) was used to generate endonuclease motif files with 'seq2profile.pl'. Then AsiSI restriction sites near γH2Av damage peaks were identified with 'annotatePeaks.pl':

```
seq2profile.pl motif_sequence 0 motifname > motifname.motif

annotatePeaks.pl \
       peakregions.bed \
       dm6 \
       -gtf annotations.gtf \
       -size -3000,3000 \
       -m AsiSI.motif \
       -mdist \
       -mbed outfile.bed > outfile-report.txt
```

Genome-wide AsiSI motifs were mapped using:

```
scanMotifGenomeWide.pl \
motifname.motif \
dm6 > outfile.bed
```

7.8.2 Total RNA-Seq bioinformatic analysis

To map reads and annotate transcripts, following genome and annotation data was used:

Table 16 Genome data and files for RNA-Seq analysis

Data	File	Online Source
Current Berkeley *Drosophila* Genome Project (BDGP) assembly (Adams *et al.*, 2000; Hoskins *et al.*, 2015)	`Drosophila_melanogaster.BDGP6.32.dna_rm.toplevel.fa`	Ensembl genome browser Yates *et al.*, 2020
Current Berkeley *Drosophila* Genome Project (BDGP) gene annotation (Adams *et al.*, 2000; Hoskins *et al.*, 2015)	`Drosophila_melanogaster.BDGP6.32.104.gtf`	Ensembl genome browser Yates *et al.*, 2020
PacBio assembly (Chang *et al.*, 2019b)	`File.S8.Chang_et_al.fasta`	Dryad repository of Chang *et al.*, 2019
PacBio gene annotation (Chang *et al.*, 2019b)	`File.S10.Chang_et_al.gff`	Dryad repository of Chang *et al.*, 2019
PacBio repeat annotation (Chang *et al.*, 2019b)	`File.S9.Chang_et_al.gff.txt`	Dryad repository of Chang *et al.*, 2019

Analysis of gene expression:

In the beginning, read quality was monitored with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, reads were aligned to a rough masked (rm) version of the BDGP assembly with HISAT2 (D. Kim *et al.*, 2019; Kim *et al.*, 2015).

```
hisat2-build \
genome.fa DromeHisat2_6.32_rm_Index
```

```
hisat2 \
-a --no-mixed --no-discordant --phred33 --rna-strandness FR \
--un-conc Unconc.sam \
--summary-file Alignmentreport.txt \
-x DromeHisat2_6.32_rm_Index \
-1 R1_sample.fastq \
-2 R2_sample.fastq \
-S Aligned.sam
```

Afterwards, the aligned data was filtered and sorted with SAMtools v. 1.3.1 (Danecek *et al.*, 2021; Li *et al.*, 2009), duplicates were removed with Picard v. 2.20.0 (http://broadinstitute.github.io/picard/) and files were converted to bam format.

```
samtools view file.sam -q 30 -F 0x04 -f 0x02 -h |
samtools sort -O SAM -o filtered_sorted.sam

java -jar /opt/bwhpc/common/bio/picard/2.20.0-java_jdk-1.8.0/lib/picard.jar \
MarkDuplicates I=file.sam REMOVE_DUPLICATES=true \
O=file_dupRm.sam METRICS_FILE=dupRm_report.txt

samtools view file_dupRm.sam -O BAM -o file_dupRm.bam
```

Then, the Cufflinks software v. 2.2.1 was used to generate data-specific GTF files with ‘cufflinks’ and ‘cuffmerge’ (Trapnell *et al.*, 2010). The custom GTF file was then employed for differential gene expression analysis with ‘cuffdiff’ (Trapnell *et al.*, 2013). The mask file used in these commands refers to a GTF file containing all annotation entries for CHRAC-14 and Brown, which is used to exclude these transcripts from the data analysis.

```
cufflinks    \
--frag-bias-correct genome.fa \
--GTF-guide annotation.gtf \
--mask-file mask.gtf \
--library-type fr-secondstrand \
--multi-read-correct \
-o out_name \
file.bam &> Cufflinks_report.txt

cuffmerge    \
--ref-sequence genome.fa \
--ref-gtf annotation.gtf \
assemblies.txt

cuffdiff \
```

```
-b genome.fa \
--library-type fr-secondstrand --multi-read-correct \
-o out_dir \
Merged.gtf \
file1.bam, file2.bam, <...> \
file11.bam, file12.bam, <...> &> Cuffdiff_report.txt
```

For further analysis and data plotting the R programming language v. 4.0.3 (https://www.r-project.org/) and RStudio (http://www.rstudio.com/) was used utilizing basic R packages (R base, ggplot etc.) and the Bioconductor package CummeRbund (Gentleman *et al.*, 2004; Goff *et al.*, 2021).

Analysis of transposable element expression:

To examine transposable element expression, reads were aligned to the custom PacBio assembly (Table 16) using HISAT2 (D. Kim *et al.*, 2019; Kim *et al.*, 2015):

```
hisat2-build \
PacBio.fa \
DromeHisat2_Mellone_Index

hisat2 \
-q --no-mixed --no-discordant --phred33 --rna-strandness FR \
--summary-file Alignmentreport.txt \
-x Path/to/DromeHisat2_Mellone_Index \
-1 R1 sample.fastq \
-2 R2_sample.fastq \
-S Aligned.sam
```

Afterwards, the aligned data was filtered and sorted with SAMtools v. 1.3.1 (Danecek *et al.*, 2021; Li *et al.*, 2009), duplicates were removed with Picard v. 2.20.0 (http://broadinstitute.github.io/picard/) and files were converted to bam format.

```
samtools view file.sam -q 10 -F 0x04 -f 0x02 -h |
samtools sort -O SAM -o filtered_sorted.sam

java -jar /opt/bwhpc/common/bio/picard/2.20.0-java_jdk-1.8.0/lib/picard.jar \
MarkDuplicates  I=file.sam REMOVE_DUPLICATES=true \
O=file_dupRm.sam METRICS_FILE=dupRm_report.txt

samtools view file_dupRm.sam -O BAM -o file_dupRm.bam
```

Then, TEtranscripts software (Jin *et al.*, 2015) was used to compute the differential expression data. For this, both the gene annotation GTF file and a formatted version of the repeat annotation file (instructions can be found in the TEtranscripts manual) of the Mellone and Larracuente labs was used (Table 16).

```
TEtranscripts \
--format BAM \
-t CHRAC14RNAi_Rep1.bam CHRAC14RNAi_Rep2.bam CHRAC14RNAi_Rep3.bam\
-c BwRNAi_Rep1.bam BwRNAi_Rep2.bam BwRNAi_Rep3.bam\
--TE TE_annotation.gtf \
--GTF PacBio_annotation.gtf \
--mode multi \
--stranded forward \
--project RNAi \
--sortByPos
```

For further analysis and data plotting the R programming language v. 4.0.3 (https://www.r-project.org/) and RStudio (http://www.rstudio.com/) was used utilizing basic R packages (R base, ggplot etc.). The Bioconductor package DeSeq2 was used for differential expression analysis (Gentleman *et al.*, 2004; Love *et al.*, 2014) and clusterProfiler for GO term analysis (Yu *et al.*, 2012).

8 Supplemental Material

Supplemental Table 1 Mass spectrometry data of interaction analysis

This table is additionally stored on the attached compact disc data storage device.

Gene_name	CENP-A-V5-His						CHRAC-14-V5-His						Condition	Category
	Control			Damage			Control			Damage				
	logFC	pvalue	hit_class	logFC	pvalue	hit_class	logFC	pvalue	hit_class	logFC	pvalue	hit_class		
CKIIBETA	0,431930	0,000235	enriched candidate	0,517772	0,000084	enriched candidate	NA	NA	NA	0,360672	0,000000	enriched candidate	CA; CH14_D	1
CHRAC-14	1,655981	0,000211	enriched hit	1,745133	0,000157	enriched hit	NA	NA	NA	NA	NA	NA	ALL	2
CHRAC-14-V5-His	NA	NA	NA	NA	NA	NA	3,073851	0,000000	enriched hit	2,959915	0,000000	enriched hit	ALL	2
CKIIALPHA	0,736962	0,000354	enriched hit	0,711802	0,000429	enriched hit	0,305436	0,000423	enriched candidate	0,364441	0,000313	enriched candidate	ALL	2
AAC11	0,397231	0,000125	enriched candidate	0,279242	0,000877	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
ACN	1,001715	0,000087	enriched hit	1,200740	0,000031	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
AGO2	0,894274	0,000168	enriched hit	0,854227	0,000218	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
B52	1,211892	0,000208	enriched hit	1,090429	0,000375	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
BEL	0,712106	0,000093	enriched hit	0,861476	0,000031	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
BX42	0,800295	0,000149	enriched hit	0,843500	0,000111	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CAPER	0,919329	0,000089	enriched hit	1,274918	0,000013	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CAPR	0,412756	0,005041	enriched candidate	0,725228	0,000267	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CAZ	1,752362	0,000225	enriched hit	1,966144	0,000117	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CDK12	0,327590	0,000962	enriched candidate	0,587362	0,000037	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CENP-A-V5-His	2,466285	0,000001	enriched hit	2,400305	0,000002	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG10077	0,411741	0,001373	enriched candidate	0,380761	0,002054	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CG10103	2,180751	0,000003	enriched hit	1,735464	0,000013	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG10333	0,585389	0,000481	enriched hit	0,670577	0,000227	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG1316	0,297244	0,002422	enriched candidate	0,276192	0,003488	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CG14215	0,326995	0,003955	enriched candidate	0,431378	0,000965	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CG15019	0,770924	0,000436	enriched hit	0,669943	0,000931	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG1622	0,738942	0,000012	enriched hit	1,030043	0,000002	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG1646	0,576564	0,000377	enriched candidate	0,627138	0,000236	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG2186	1,299704	0,000508	enriched hit	1,778082	0,000088	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG2199	0,755381	0,000197	enriched hit	0,979407	0,000045	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG30122	2,688087	0,000007	enriched hit	2,983610	0,000004	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG32409	0,810658	0,000456	enriched hit	0,601670	0,002202	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG34417	1,179985	0,000009	enriched hit	1,166956	0,000009	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG3689	0,597932	0,012569	enriched hit	1,390233	0,000172	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG3800	0,293813	0,032779	enriched candidate	0,403975	0,008910	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CG42232	0,674882	0,000415	enriched hit	0,792296	0,000170	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG4266	0,548755	0,000180	enriched candidate	1,026662	0,000005	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG42724	0,869010	0,000393	enriched hit	0,667659	0,001608	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG4511	0,480796	0,001092	enriched candidate	0,508915	0,000807	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CG5726	1,060263	0,000175	enriched hit	1,641152	0,000014	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG7185	0,782767	0,003592	enriched hit	1,392505	0,000168	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG7564	1,090719	0,000005	enriched hit	1,141589	0,000004	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG8108	0,988728	0,000116	enriched hit	0,973800	0,000126	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG9775	0,529385	0,001118	enriched candidate	0,658394	0,000344	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG9776	1,261481	0,000023	enriched hit	1,515038	0,000008	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CHIC	0,337825	0,000276	enriched candidate	0,274387	0,000860	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
CHMP1	2,851501	0,000001	enriched hit	2,247419	0,000005	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CHMP2B	2,205724	0,000003	enriched hit	1,558572	0,000023	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
D1	2,204056	0,000024	enriched hit	2,241743	0,000022	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
DEK	0,716981	0,001289	enriched hit	0,523193	0,006156	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
ENS	0,305396	0,015022	enriched candidate	0,470760	0,001984	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
GALPHAI	0,289781	0,003122	enriched candidate	0,371444	0,000870	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
GLO	0,822475	0,001120	enriched hit	0,821363	0,001128	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HIS1	0,625700	0,000322	enriched hit	0,284369	0,016015	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
HIS2A	2,899357	0,000000	enriched hit	2,557310	0,000001	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HIS2AV	2,974850	0,000003	enriched hit	2,740206	0,000005	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HIS2B	3,286262	0,000002	enriched hit	2,910291	0,000004	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HIS3	2,805639	0,000000	enriched hit	2,657806	0,000000	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HIS4	3,096679	0,000001	enriched hit	2,942932	0,000001	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HNRNP	0,293843	0,002356	enriched candidate	0,269779	0,003601	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
HRB27C	0,312500	0,003662	enriched candidate	0,385824	0,001259	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
HRB98DE	0,757823	0,000379	enriched hit	0,762341	0,000367	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
HYX	0,587061	0,000312	enriched hit	0,764742	0,000070	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
IBF1	0,401966	0,010717	enriched candidate	0,413538	0,009454	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
IBF2	0,303773	0,015279	enriched candidate	0,324449	0,011537	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
LARK	0,715771	0,000310	enriched hit	0,662238	0,000476	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
MAGO	0,268994	0,022427	enriched candidate	0,305574	0,013285	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
MAP205	0,652538	0,006717	enriched hit	0,759401	0,003255	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
NITO	0,266212	0,003345	enriched candidate	0,413569	0,000330	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
NOCTE	0,953227	0,000630	enriched hit	1,163475	0,000210	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
NON3	0,470949	0,000459	enriched candidate	0,447510	0,000606	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
NONA	0,618209	0,000855	enriched hit	0,747308	0,000304	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
P32	1,209723	0,000212	enriched hit	1,506717	0,000061	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
PGD	0,563360	0,000528	enriched candidate	0,336477	0,007175	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
PNUTS	0,683417	0,000038	enriched hit	0,583249	0,000096	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
PRP38	0,681803	0,001337	enriched hit	0,601628	0,002536	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
QKR58E	0,517547	0,001630	enriched candidate	0,657299	0,000456	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RBP1	0,405229	0,000108	enriched candidate	0,502927	0,000031	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RBP2	0,289047	0,002489	enriched candidate	0,293922	0,002288	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
REF1	0,665276	0,000602	enriched hit	0,461503	0,003965	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RM62	0,943510	0,000058	enriched hit	1,255675	0,000011	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RNPS1	1,073047	0,000069	enriched hit	1,377525	0,000016	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RPL11	0,342249	0,000797	enriched candidate	0,284423	0,002094	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPL26	0,630647	0,001771	enriched hit	0,471611	0,007298	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPL27A	0,529814	0,000123	enriched candidate	0,449234	0,000311	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPL29	1,132217	0,000044	enriched hit	0,825840	0,000267	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RPL34B	0,576484	0,000041	enriched candidate	0,304991	0,001413	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPL37A	0,662467	0,000600	enriched hit	0,477599	0,003275	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPL39	0,761681	0,001553	enriched hit	0,726259	0,001983	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RPL5	0,419192	0,001100	enriched candidate	0,342804	0,003078	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPS23	0,611145	0,000107	enriched hit	0,569927	0,000159	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RPS26	0,976018	0,000170	enriched hit	0,683232	0,001192	enriched hit	NA	NA	NA	NA	NA	NA	CA	3

RPS27A	1,486483	0,000013	enriched hit	1,499495	0,000012	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RPS30	1,372820	0,000057	enriched hit	1,139957	0,000165	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
RPS8	0,311031	0,008696	enriched candidate	0,385299	0,003162	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RRP1	0,346192	0,004866	enriched candidate	0,311387	0,007968	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
RUMP	0,709595	0,000616	enriched hit	1,116206	0,000048	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SAF	0,538595	0,000341	enriched candidate	0,740925	0,000056	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SC35	0,642808	0,000568	enriched hit	0,493529	0,002277	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SF1	0,814183	0,000311	enriched hit	0,997005	0,000099	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SF2	0,929737	0,000078	enriched hit	0,838422	0,000141	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SLE	1,239813	0,000350	enriched hit	1,330165	0,000237	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SMB	0,377178	0,025759	enriched candidate	0,288979	0,065119	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SMD3	0,502891	0,002083	enriched candidate	0,427517	0,004645	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SME	0,326575	0,020862	enriched candidate	0,287619	0,033895	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SMT3	0,772528	0,000041	enriched hit	0,590499	0,000192	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SNRNP	1,019094	0,000092	enriched hit	1,107351	0,000057	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SPT5	0,369440	0,000458	enriched candidate	0,401990	0,000287	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SRM160	1,080545	0,000093	enriched hit	1,268844	0,000037	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SRP54	0,538234	0,000748	enriched candidate	0,726394	0,000143	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SSX	0,738026	0,000088	enriched hit	0,777695	0,000065	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
SWM	0,288851	0,077990	enriched candidate	0,388911	0,028714	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
SYP	0,510673	0,002121	enriched candidate	0,421920	0,005410	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
TANGO1	0,518028	0,000228	enriched candidate	0,530885	0,000199	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
TM9SF4	0,435447	0,008626	enriched candidate	0,470924	0,006022	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
TOP1	0,380687	0,002928	enriched candidate	0,378274	0,003021	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
TOP2	0,777172	0,000242	enriched hit	0,753419	0,000288	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
TRAL	1,033683	0,000146	enriched hit	1,133434	0,000086	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
U2AF38	1,016656	0,000019	enriched hit	1,154563	0,000009	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
U2AF50	1,452056	0,000052	enriched hit	1,274532	0,000109	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
VHAM8	0,267418	0,010677	enriched candidate	0,391370	0,001722	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
VIG	1,919758	0,000030	enriched hit	1,825679	0,000040	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
VIG2	1,546764	0,000045	enriched hit	1,430459	0,000071	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
VPS2	1,823087	0,000006	enriched hit	1,489943	0,000021	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
VPS24	0,624092	0,000394	enriched hit	0,386475	0,004711	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
VPS60	2,226238	0,000007	enriched hit	1,580194	0,000051	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
VSX1	0,759076	0,000253	enriched hit	0,757979	0,000255	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
X16	0,480928	0,000617	enriched candidate	0,494399	0,000531	enriched candidate	NA	NA	NA	NA	NA	NA	CA	3
YT521	2,185405	0,000003	enriched hit	2,605646	0,000001	enriched hit	NA	NA	NA	NA	NA	NA	CA	3
CG11148	0,266252	0,014191	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
CG11583	0,276778	0,008537	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
CG18178	0,281253	0,002020	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
CG2691	0,339742	0,005484	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
DEFL	0,378443	0,199322	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
DOS	0,267221	0,009098	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
FIB	0,334556	0,003037	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
INCENP	0,356342	0,000722	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
KCC	0,284075	0,000480	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
NOP5	0,315885	0,001908	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL10	0,303673	0,002706	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL13	0,343492	0,003955	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL14	0,416589	0,009397	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL17	0,631844	0,005521	enriched hit	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL18	0,441601	0,001060	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL18A	0,459721	0,006590	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL21	0,447491	0,005074	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL23A	0,344891	0,001626	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL24	0,545522	0,005404	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL27	0,406066	0,006161	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL3	0,282022	0,005587	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL35	0,787821	0,000098	enriched hit	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL35A	0,426592	0,007086	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL36	0,594238	0,006194	enriched hit	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL36A	0,460163	0,000077	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL7	0,318395	0,006164	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL7A	0,367165	0,004405	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPL8	0,310577	0,018369	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPS11	0,310687	0,011085	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPS2	0,406388	0,005576	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPS6	0,346613	0,000371	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
RPS9	0,341978	0,001131	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
VIR	0,304510	0,014099	enriched candidate	NA	NA	NA	NA	NA	NA	NA	NA	NA	CA_C	4
ART8	NA	NA	NA	0,284184	0,002030	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CG16753	NA	NA	NA	0,509136	0,131822	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CG2982	NA	NA	NA	0,305139	0,004672	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CG5746	NA	NA	NA	0,583243	0,005836	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CG9977	NA	NA	NA	0,291426	0,004057	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CIN	NA	NA	NA	0,270785	0,014910	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CSTF	NA	NA	NA	0,320880	0,027309	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
FMR1	NA	NA	NA	0,343877	0,001617	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
HRB87F	NA	NA	NA	0,375783	0,000832	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
LAM	NA	NA	NA	0,329841	0,002890	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
LIG	NA	NA	NA	0,365186	0,024732	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
MOE	NA	NA	NA	0,639845	0,002433	enriched hit	NA	NA	NA	NA	NA	NA	CA_D	5
NOP60B	NA	NA	NA	0,307544	0,012159	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
NOPP140	NA	NA	NA	0,473878	0,002553	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
PSQ	NA	NA	NA	0,268900	0,028792	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
SNAMA	NA	NA	NA	0,568871	0,001100	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
SRP	NA	NA	NA	0,285308	0,020570	enriched candidate	NA	NA	NA	NA	NA	NA	CA_D	5
CG15098	NA	NA	NA	NA	NA	NA	1,010106	0,000000	enriched hit	1,083282	0,000000	enriched hit	CH14	6
CG4658	NA	NA	NA	NA	NA	NA	0,502250	0,003228	enriched candidate	0,632380	0,001486	enriched candidate	CH14	6
CHRAC-16	NA	NA	NA	NA	NA	NA	2,019346	0,000000	enriched hit	2,342151	0,000000	enriched hit	CH14	6

MES4	NA	NA	NA	NA	NA	NA	3,700378	0,000000	enriched hit	3,500275	0,000000	enriched hit	CH14	6
TBC1D15	NA	NA	NA	NA	NA	NA	0,552908	0,005178	enriched candidate	0,717090	0,001891	enriched candidate	CH14	6
CG10340	NA	NA	NA	NA	NA	NA	0,410237	0,001915	enriched candidate	NA	NA	NA	CH14_C	7
CG11858	NA	NA	NA	NA	NA	NA	0,340637	0,005027	enriched candidate	NA	NA	NA	CH14_C	7
CG13630	NA	NA	NA	NA	NA	NA	0,300860	0,002305	enriched candidate	NA	NA	NA	CH14_C	7
CG5886	NA	NA	NA	NA	NA	NA	0,396124	0,001392	enriched candidate	NA	NA	NA	CH14_C	7
FER1HCH	NA	NA	NA	NA	NA	NA	0,362776	0,000035	enriched candidate	NA	NA	NA	CH14_C	7
GBB	NA	NA	NA	NA	NA	NA	0,331760	0,000221	enriched candidate	NA	NA	NA	CH14_C	7
P47	NA	NA	NA	NA	NA	NA	0,443618	0,000481	enriched candidate	NA	NA	NA	CH14_C	7
RAD23	NA	NA	NA	NA	NA	NA	0,408303	0,000640	enriched candidate	NA	NA	NA	CH14_C	7
SHRB	NA	NA	NA	NA	NA	NA	0,459601	0,001490	enriched candidate	NA	NA	NA	CH14_C	7
TFIIS	NA	NA	NA	NA	NA	NA	0,434117	0,000158	enriched candidate	NA	NA	NA	CH14_C	7
ADF1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,377101	0,033543	enriched candidate	CH14_D	8
ADK3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,472406	0,076597	enriched candidate	CH14_D	8
AGO1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,547751	0,020209	enriched candidate	CH14_D	8
ALC	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,456416	0,043606	enriched candidate	CH14_D	8
ALG	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,536916	0,065823	enriched candidate	CH14_D	8
APP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,470988	0,041281	enriched candidate	CH14_D	8
ARFGAP1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,772252	0,017846	enriched candidate	CH14_D	8
ARL1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,555855	0,053014	enriched candidate	CH14_D	8
ARP10	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,476639	0,005023	enriched candidate	CH14_D	8
ATPSYNE	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,425183	0,072960	enriched candidate	CH14_D	8
BAP111	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,429021	0,045132	enriched candidate	CH14_D	8
BAP60	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,435877	0,048166	enriched candidate	CH14_D	8
BLOS4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,426177	0,127528	enriched candidate	CH14_D	8
BRE1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,435739	0,045124	enriched candidate	CH14_D	8
BSF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,602394	0,017951	enriched candidate	CH14_D	8
BUN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,482334	0,086782	enriched candidate	CH14_D	8
BYS	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,467239	0,039125	enriched candidate	CH14_D	8
CANA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454140	0,023514	enriched candidate	CH14_D	8
CANB2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,803382	0,065893	enriched candidate	CH14_D	8
CAR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,448529	0,111469	enriched candidate	CH14_D	8
CDASE	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,613994	0,008350	enriched candidate	CH14_D	8
CDK5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,515650	0,039426	enriched candidate	CH14_D	8
CG10376	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,490436	0,031669	enriched candidate	CH14_D	8
CG10638	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,481822	0,040286	enriched candidate	CH14_D	8
CG10932	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,577440	0,008710	enriched candidate	CH14_D	8
CG11127	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,519612	0,044668	enriched candidate	CH14_D	8
CG11134	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,363595	0,106918	enriched candidate	CH14_D	8
CG11138	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,521223	0,096155	enriched candidate	CH14_D	8
CG11178	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,572751	0,003141	enriched candidate	CH14_D	8
CG11448	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,644697	0,012839	enriched candidate	CH14_D	8
CG11771	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,312695	0,120759	enriched candidate	CH14_D	8
CG11779	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,425022	0,069865	enriched candidate	CH14_D	8
CG11883	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,550996	0,051711	enriched candidate	CH14_D	8
CG11964	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,541899	0,041637	enriched candidate	CH14_D	8
CG12096	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,551995	0,037231	enriched candidate	CH14_D	8
CG1218	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,539738	0,062013	enriched candidate	CH14_D	8
CG12262	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,547512	0,005212	enriched candidate	CH14_D	8
CG12264	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,513332	0,024494	enriched candidate	CH14_D	8
CG12321	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,567392	0,043052	enriched candidate	CH14_D	8
CG1291	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,530683	0,141272	enriched candidate	CH14_D	8
CG13090	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,445396	0,091226	enriched candidate	CH14_D	8
CG13117	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,393040	0,151563	enriched candidate	CH14_D	8
CG13185	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,459181	0,064142	enriched candidate	CH14_D	8
CG13366	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454900	0,050050	enriched candidate	CH14_D	8
CG13638	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,572801	0,009357	enriched candidate	CH14_D	8
CG13901	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,471544	0,041207	enriched candidate	CH14_D	8
CG14232	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,403022	0,060879	enriched candidate	CH14_D	8
CG14407	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,396246	0,153038	enriched candidate	CH14_D	8
CG14544	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,449155	0,095437	enriched candidate	CH14_D	8
CG14894	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,618242	0,044631	enriched candidate	CH14_D	8
CG15014	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,556849	0,009537	enriched candidate	CH14_D	8
CG1513	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,570457	0,042971	enriched candidate	CH14_D	8
CG15440	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,790522	0,000169	enriched hit	CH14_D	8
CG1550	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,468519	0,047557	enriched candidate	CH14_D	8
CG15717	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,422544	0,057641	enriched candidate	CH14_D	8
CG15735	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,545242	0,051035	enriched candidate	CH14_D	8
CG16935	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,449337	0,019884	enriched candidate	CH14_D	8
CG17078	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,420296	0,088415	enriched candidate	CH14_D	8
CG17294	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,544654	0,101660	enriched candidate	CH14_D	8
CG17333	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,522396	0,065685	enriched candidate	CH14_D	8
CG1749	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,539246	0,013027	enriched candidate	CH14_D	8
CG17746	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,611472	0,044935	enriched candidate	CH14_D	8
CG18661	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,387492	0,088430	enriched candidate	CH14_D	8
CG1885	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,591690	0,022446	enriched candidate	CH14_D	8
CG1968	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,471046	0,094549	enriched candidate	CH14_D	8
CG1969	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,445408	0,051297	enriched candidate	CH14_D	8
CG2100	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,528465	0,035791	enriched candidate	CH14_D	8
CG2118	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,557225	0,027656	enriched candidate	CH14_D	8
CG2974	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,354265	0,109710	enriched candidate	CH14_D	8
CG31064	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,368257	0,069828	enriched candidate	CH14_D	8
CG31122	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,542525	0,037648	enriched candidate	CH14_D	8
CG32758	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,483936	0,069084	enriched candidate	CH14_D	8
CG33158	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,538338	0,026081	enriched candidate	CH14_D	8
CG3402	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,579146	0,004474	enriched candidate	CH14_D	8
CG3529	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,372704	0,145753	enriched candidate	CH14_D	8
CG3847	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,504966	0,031808	enriched candidate	CH14_D	8
CG3909	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,466568	0,050628	enriched candidate	CH14_D	8
CG42788	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454733	0,017511	enriched candidate	CH14_D	8

CG4300	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,464287	0,028224	enriched candidate	CH14_D	8	
CG4572	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,423638	0,055629	enriched candidate	CH14_D	8	
CG4858	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,538987	0,040150	enriched candidate	CH14_D	8	
CG4951	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,517106	0,017064	enriched candidate	CH14_D	8	
CG4968	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,607007	0,014371	enriched candidate	CH14_D	8	
CG5044	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,594401	0,022189	enriched candidate	CH14_D	8	
CG5110	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,609924	0,055694	enriched candidate	CH14_D	8	
CG5220	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,500604	0,035029	enriched candidate	CH14_D	8	
CG5377	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,465740	0,110748	enriched candidate	CH14_D	8	
CG5382	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,571714	0,037928	enriched candidate	CH14_D	8	
CG5599	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,511133	0,043624	enriched candidate	CH14_D	8	
CG5745	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,560584	0,004720	enriched candidate	CH14_D	8	
CG5757	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,442700	0,067441	enriched candidate	CH14_D	8	
CG5902	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,507187	0,038801	enriched candidate	CH14_D	8	
CG6299	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,429065	0,072777	enriched candidate	CH14_D	8	
CG6422	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,566634	0,012775	enriched candidate	CH14_D	8	
CG6983	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,530696	0,070448	enriched candidate	CH14_D	8	
CG7322	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,595796	0,055988	enriched candidate	CH14_D	8	
CG7332	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,650089	0,007604	enriched candidate	CH14_D	8	
CG7791	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,546693	0,015643	enriched candidate	CH14_D	8	
CG7857	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,434820	0,096308	enriched candidate	CH14_D	8	
CG7889	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,546069	0,041994	enriched candidate	CH14_D	8	
CG8173	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,477948	0,058943	enriched candidate	CH14_D	8	
CG8176	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,449325	0,052607	enriched candidate	CH14_D	8	
CG8199	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,389874	0,037415	enriched candidate	CH14_D	8	
CG8525	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,523846	0,039658	enriched candidate	CH14_D	8	
CG8549	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,536940	0,027039	enriched candidate	CH14_D	8	
CG8613	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,389996	0,120908	enriched candidate	CH14_D	8	
CG9186	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,440282	0,040682	enriched candidate	CH14_D	8	
CG9286	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,623529	0,013559	enriched candidate	CH14_D	8	
CG9578	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,414518	0,046705	enriched candidate	CH14_D	8	
CG9629	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,550479	0,023006	enriched candidate	CH14_D	8	
CIAPIN1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,357035	0,092895	enriched candidate	CH14_D	8	
CIP4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,531598	0,013628	enriched candidate	CH14_D	8	
CKIALPHA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,615319	0,013066	enriched candidate	CH14_D	8	
CLIP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,404468	0,043829	enriched candidate	CH14_D	8	
CPSF160	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,563187	0,054742	enriched candidate	CH14_D	8	
CPT2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,361306	0,121808	enriched candidate	CH14_D	8	
CRK	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,510300	0,010198	enriched candidate	CH14_D	8	
CRP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,432949	0,056083	enriched candidate	CH14_D	8	
CSL4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,478973	0,045924	enriched candidate	CH14_D	8	
CT	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,476439	0,070605	enriched candidate	CH14_D	8	
CTSB1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,396097	0,125468	enriched candidate	CH14_D	8	
CYP12A4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,433428	0,071061	enriched candidate	CH14_D	8	
CYP12D1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,503467	0,002047	enriched candidate	CH14_D	8	
D2HGDH	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,498567	0,048642	enriched candidate	CH14_D	8	
DAP160	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,354905	0,017866	enriched candidate	CH14_D	8	
DBP80	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,527372	0,025580	enriched candidate	CH14_D	8	
DCP1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,537504	0,010826	enriched candidate	CH14_D	8	
DGT6	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,490931	0,025472	enriched candidate	CH14_D	8	
DPH5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,456016	0,032388	enriched candidate	CH14_D	8	
DREDD	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,297724	0,150355	enriched candidate	CH14_D	8	
EIF2D	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,503588	0,058303	enriched candidate	CH14_D	8	
EIF6	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,545627	0,059842	enriched candidate	CH14_D	8	
ELM	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,836977	0,009240	enriched candidate	CH14_D	8	
ELOB	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,522271	0,112875	enriched candidate	CH14_D	8	
ELP2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,478804	0,042467	enriched candidate	CH14_D	8	
ENDOB	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,558201	0,030018	enriched candidate	CH14_D	8	
ETF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,637169	0,012880	enriched candidate	CH14_D	8	
FBL	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,797463	0,008112	enriched candidate	CH14_D	8	
FLII	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,446068	0,041316	enriched candidate	CH14_D	8	
FPPS	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,655363	0,010301	enriched candidate	CH14_D	8	
GCC185	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454477	0,056837	enriched candidate	CH14_D	8	
GK1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,512014	0,043242	enriched candidate	CH14_D	8	
GMAP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,628644	0,054809	enriched candidate	CH14_D	8	
GMD	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,432168	0,047582	enriched candidate	CH14_D	8	
GOLGIN245	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,335565	0,039150	enriched candidate	CH14_D	8	
GSTO2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,489740	0,079669	enriched candidate	CH14_D	8	
GSTO3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,545869	0,034573	enriched candidate	CH14_D	8	
GSTT4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,440807	0,106761	enriched candidate	CH14_D	8	
HBS1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,467164	0,097577	enriched candidate	CH14_D	8	
HEX	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,573748	0,016786	enriched candidate	CH14_D	8	
HEXO2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,501582	0,043580	enriched candidate	CH14_D	8	
HO	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,523311	0,025817	enriched candidate	CH14_D	8	
HOOK	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,554436	0,026077	enriched candidate	CH14_D	8	
HSP26	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,523801	0,019256	enriched candidate	CH14_D	8	
HSP27	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,522403	0,027428	enriched candidate	CH14_D	8	
JAFRAC1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,370145	0,104124	enriched candidate	CH14_D	8	
KAY	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,321912	0,057383	enriched candidate	CH14_D	8	
KEY	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,552969	0,042969	enriched candidate	CH14_D	8	
LANA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,316726	0,092484	enriched candidate	CH14_D	8	
LANB2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,273567	0,101300	enriched candidate	CH14_D	8	
LID	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,656480	0,028937	enriched candidate	CH14_D	8	
LSM	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,365093	0,094555	enriched candidate	CH14_D	8	
LSM1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,405559	0,112565	enriched candidate	CH14_D	8	
LSM7	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,440703	0,066087	enriched candidate	CH14_D	8	
M1BP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,558029	0,040713	enriched candidate	CH14_D	8	
MAB	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,459824	0,075236	enriched candidate	CH14_D	8	
MAD1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,600098	0,015818	enriched candidate	CH14_D	8	
MADM	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,526332	0,039770	enriched candidate	CH14_D	8	

MBS	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,608503	0,028948	enriched candidate	CH14_D	8	
MESK2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,511643	0,038658	enriched candidate	CH14_D	8	
MESR6	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,505815	0,046878	enriched candidate	CH14_D	8	
MLF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,572106	0,009376	enriched candidate	CH14_D	8	
MON2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,596932	0,021655	enriched candidate	CH14_D	8	
MOP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,456698	0,055082	enriched candidate	CH14_D	8	
MTP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,449103	0,067817	enriched candidate	CH14_D	8	
MTR3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,548117	0,026142	enriched candidate	CH14_D	8	
MUTED	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,532154	0,037306	enriched candidate	CH14_D	8	
MXT	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,381000	0,088654	enriched candidate	CH14_D	8	
NCLB	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,508731	0,020774	enriched candidate	CH14_D	8	
NITFHIT	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,484244	0,017652	enriched candidate	CH14_D	8	
NOT3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,524649	0,016044	enriched candidate	CH14_D	8	
NS3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,658132	0,024299	enriched candidate	CH14_D	8	
NTR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,387591	0,112521	enriched candidate	CH14_D	8	
NUP154	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,645017	0,019156	enriched candidate	CH14_D	8	
OR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,458578	0,048365	enriched candidate	CH14_D	8	
OX	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,529256	0,064359	enriched candidate	CH14_D	8	
P24	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,481007	0,069126	enriched candidate	CH14_D	8	
P5CR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,623834	0,022469	enriched candidate	CH14_D	8	
PAK	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,280803	0,149349	enriched candidate	CH14_D	8	
PAPLA1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,582312	0,021785	enriched candidate	CH14_D	8	
PLL	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,515374	0,003713	enriched candidate	CH14_D	8	
PN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,410100	0,066311	enriched candidate	CH14_D	8	
POLY	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,626476	0,016924	enriched candidate	CH14_D	8	
POLYBROMO	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,367342	0,085058	enriched candidate	CH14_D	8	
PP1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,532862	0,034464	enriched candidate	CH14_D	8	
PP4	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,496203	0,075562	enriched candidate	CH14_D	8	
PPAT	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,471130	0,059906	enriched candidate	CH14_D	8	
PUM	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,571997	0,000680	enriched candidate	CH14_D	8	
QM	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,351867	0,092451	enriched candidate	CH14_D	8	
RAB14	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,471454	0,039676	enriched candidate	CH14_D	8	
RABEX	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,529463	0,035912	enriched candidate	CH14_D	8	
RAGA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,508842	0,066541	enriched candidate	CH14_D	8	
RBCN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,483942	0,066160	enriched candidate	CH14_D	8	
RBPN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,269763	0,161840	enriched candidate	CH14_D	8	
RGA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,545290	0,006292	enriched candidate	CH14_D	8	
RHOL	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,539746	0,040352	enriched candidate	CH14_D	8	
RPA3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,407281	0,047484	enriched candidate	CH14_D	8	
RPII33	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,360550	0,101534	enriched candidate	CH14_D	8	
RPS15AA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,518508	0,041640	enriched candidate	CH14_D	8	
RUDHIRA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,385068	0,011764	enriched candidate	CH14_D	8	
SAP47	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,338866	0,123511	enriched candidate	CH14_D	8	
SBF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,560238	0,018627	enriched candidate	CH14_D	8	
SBR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,479784	0,056380	enriched candidate	CH14_D	8	
SCAMP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454349	0,052345	enriched candidate	CH14_D	8	
SCPX	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,427358	0,091590	enriched candidate	CH14_D	8	
SCU	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,547979	0,011692	enriched candidate	CH14_D	8	
SDHAF3	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,553927	0,019966	enriched candidate	CH14_D	8	
SEC16	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,370886	0,067757	enriched candidate	CH14_D	8	
SEC5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,478189	0,037767	enriched candidate	CH14_D	8	
SEC71	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,537708	0,076632	enriched candidate	CH14_D	8	
SERCA|CA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,532002	0,100704	enriched candidate	CH14_D	8	
SHARK	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,453378	0,020230	enriched candidate	CH14_D	8	
SIP2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,386794	0,102686	enriched candidate	CH14_D	8	
SLS	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,558975	0,009609	enriched candidate	CH14_D	8	
SMC1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,455991	0,106870	enriched candidate	CH14_D	8	
SMC5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,424441	0,103445	enriched candidate	CH14_D	8	
SNAPIN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,475717	0,112863	enriched candidate	CH14_D	8	
SNI	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,521874	0,086849	enriched candidate	CH14_D	8	
SNRPG	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,366499	0,081109	enriched candidate	CH14_D	8	
SPN42DA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,457560	0,074733	enriched candidate	CH14_D	8	
SPN42DC	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,483315	0,054254	enriched candidate	CH14_D	8	
SPN55B	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,469834	0,063805	enriched candidate	CH14_D	8	
SUCB	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,565655	0,038965	enriched candidate	CH14_D	8	
SW	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,525900	0,038336	enriched candidate	CH14_D	8	
TAILOR	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,549409	0,074209	enriched candidate	CH14_D	8	
TBCE	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,500619	0,024498	enriched candidate	CH14_D	8	
TEX	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,574966	0,010391	enriched candidate	CH14_D	8	
TFIIA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,454369	0,139363	enriched candidate	CH14_D	8	
TGT	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,309169	0,023040	enriched candidate	CH14_D	8	
THOC5	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,530832	0,073653	enriched candidate	CH14_D	8	
TINA	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,389902	0,130753	enriched candidate	CH14_D	8	
TWF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,443547	0,002280	enriched candidate	CH14_D	8	
TWIN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,523110	0,024717	enriched candidate	CH14_D	8	
TZN	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,388979	0,080444	enriched candidate	CH14_D	8	
U2A	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,362912	0,018742	enriched candidate	CH14_D	8	
UBC2	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,553836	0,020054	enriched candidate	CH14_D	8	
UBP64E	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,539389	0,033234	enriched candidate	CH14_D	8	
UPF1	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,484848	0,016959	enriched candidate	CH14_D	8	
USP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,507576	0,061344	enriched candidate	CH14_D	8	
VELI	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,334675	0,048080	enriched candidate	CH14_D	8	
VELO	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,526738	0,038922	enriched candidate	CH14_D	8	
VHA14	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,483753	0,039560	enriched candidate	CH14_D	8	
VIAF	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,419431	0,059417	enriched candidate	CH14_D	8	
VPS45	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,447340	0,035221	enriched candidate	CH14_D	8	
WAC	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,469596	0,070751	enriched candidate	CH14_D	8	
XNP	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,538806	0,029670	enriched candidate	CH14_D	8	

All NGS datasets from CUT&Tag-Seq and RNA-Seq experiments are stored as supplemental tables on the attached compact disc data storage device.

Supplemental Table 2 CUT&Tag centromeric repeat hits (PacBio dataset)
Supplemental Table 3 CUT&Tag non-centromeric repeat hits (PacBio dataset)
Supplemental Table 4 CUT&Tag genome-wide hits (BDGP dataset)
Supplemental Table 5 CUT&Tag DNA damage sites
Supplemental Table 6 Differentially expressed genes upon CHRAC-14 RNAi
Supplemental Table 7 Differentially expressed genes with enriched CENP-A CUT&Tag signal
Supplemental Table 8 Differentially expressed repeats upon CHRAC-14 RNAi

9 List of Abbreviations

°C	degree Celsius
3D	three-dimensional
aa	amino acid
ac	acetyl
ACF	ATP-dependent chromatin assembly and remodeling factor
Asf1	Anti-silencing function 1
ATAC	Ada2a-containing complex
ATAC-Seq	Assay for transposase-accessible chromatin sequencing
ATP	Adenosine triphosphate
ATRX	X-linked helicase
BDGP	Berkeley *Drosophila* genome project
bp	base pair
Bw	Brown
C-terminal	carboxy-terminal
CAF1	Chromatin assembly factor 1
CAL1	Chromosome alignment defect 1
CaT	CUT&Tag (cutting under target and tagmentation)
CCAN	Constitutive centromere-associated network
cDNA	complementary DNA
CenH3	Centromere-specific histone H3
CENP-A	Centromere protein A
CENP-C	Centromere protein C
cenRNA	centromeric RNA
Ch14, CH14, CH-14	CHRAC-14
CHD	Chromodomain helicase DNA-binding
ChIP-Seq	Chromatin immunoprecipitation followed by sequencing

chr	chromosome
CHRAC	Chromatin accessibility complex
CID	Centromere identifier
CK2, CKII	Casein kinase 2
cm	centimeter
CRC	Chromatin remodeling complex
$CuSO_4$	copper sulphate
d	day(s)
DAXX	Death domain associated protein
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
DSB	double strand break
dsDNA	double stranded DNA
dsRNA	double stranded RNA
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
FACT	facilitates chromatin transcription
FC	fold change
FPKM	Fragments per kilobase million
FT	flow through
fw	forward
g	gram
gDNA	genomic DNA
GFP	Green fluorescent protein
GO term	gene ontology term
h	hour(s)
H1	Histone 1
H2A	Histone 2A

H2Av	Histone 2Av
H2B	Histone 2B
H3	Histone 3
H3.3	Histone 3.3
H4	Histone 4
HAT	histone acetyl transferase
hCK2	human Casein kinase 2
HDAC	histone deacetylase
HJURP	Holliday junction recognition protein
HR	homologous repair
HRP	horse radish peroxidase
Hyd	hyperplastic discs
IF	immunofluorescence
IGV	Integrative genomics viewer
INO80	Inositol requiring 80
IP	immunoprecipitation
ISWI	Imitation switch
kb	kilobase
KD	knockdown
kDa	kilodalton
l	liter
LC-MS/MS	liquid chromatography tandem mass spectrometry
LINE	long interspersed nuclear elements
LTR	long terminal repeat
M	molar
m/z	mass-to-charge ratio
mAID	mini auxin inducible degron
mass spec, MS	mass spectrometry

max	maximal
Mbp	megabase pair
MDA	mixture discriminant analysis
me	methyl
mg	milligram
min	minute(s)
mio	million
ml	milliliter
mM	millimolar
MNase	Micrococcal nuclease
MRN	MRE11–RAD50–NBS1
mRNA	messenger RNA
n =	sample size =
N-terminal	amino-terminal
n.s,	not significant
ncRNA	non-coding RNA
NCS	Neocarzinostatin
ng	nanogram
NGS	next generation sequencing
NHEJ	non-homologous end joining
NLS	nuclear localization signal
nm	nanometer
nM	nanomolar
nt	nucleotide
NuRD	Nucleosome remodeling and deacetylase
NURF	Nucleosome remodeling factor
o/n	overnight
Ore-R	Oregon-R

pA-Tn5	Protein A-Transposase5
PAGE	poly acrylamide gel electrophoresis
pBAF	Polybromo-associated BAF complex
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
piRNA	Piwi-interacting RNAs
pM	picomolar
pMT	metallothionein gene promoter
Pol ε	Polymerase ε
PRE	Polycomb responsive element
qPCR	quantitative polymerase chain reaction
rep	replicate
RNA	ribonucleic acid
RNAi	RNA interference
RNase	Riboluclease
ROS	reactive oxygen species
rpm	revolutions per minute
rRNA or rDNA	ribosomal RNA or ribosomal DNA
RT	room temperature
rv	reverse
S2	*Drosophila* Schneider 2 cell line
S20	Serine 20
S20P	phosphorylated serine 20
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	standard error
sec	second(s)
siRNA	small interfering RNA

ssDNA	single stranded DNA
ssRNA	single stranded RNA
SUMO	Small ubiquitin-like modifier
SWI/SNF	SWItch/Sucrose non-fermentable
T122	Threonine 122
T122A	Threonine 122 to alanine mutation
T122D	Threonine 122 to aspartic acid mutation
TAD	topologically associating domain
TE	transposable element
Tris	tris(hydroxymethyl)aminomethane hydrochloride
TSS	transcriptional start site
WB	Western blot
wt, WT	wild type
x g	times gravity
γH2Av	phosphorylated Histone 2Av
µg	microgram
µl	microliter
µm	micrometer
µM	micromolar

10 List of Figures

11 List of Tables

12 Bibliography

Adam S, Dabin J, Polo SE. 2015. Chromatin plasticity in response to DNA damage: The shape of things to come. *DNA Repair (Amst)* **32**:120–126. doi:10.1016/j.dnarep.2015.04.022

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Venter JC, *et al.* 2000. The Genome Sequence of Drosophila melanogaster. *Science (80-)* **287**:2185–2195. doi:10.1126/science.287.5461.2185

Adams RR, Maiato H, Earnshaw WC, Carmena M. 2001. Essential roles of Drosophila inner centromere protein (INCENP) and aurora B in histone H3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation. *J Cell Biol* **153**:865–879. doi:10.1083/jcb.153.4.865

Ahmad K, Henikoff S. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**:1191–1200. doi:10.1016/S1097-2765(02)00542-7

Allshire RC, Karpen GH. 2008. Epigenetic regulation of centromeric chromatin : old dogs , new tricks ? *Nat Rev Genet* **9**:923–937. doi:10.1038/nrg2466

Allshire RC, Madhani HD. 2018. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**:229–244. doi:10.1038/nrm.2017.119

Amato A, Schillaci T, Lentini L, Di Leonardo A. 2009. CENPA overexpression promotes genome instability in pRb-depleted human cells. *Mol Cancer* **8**:1–14. doi:10.1186/1476-4598-8-119

Andrews S. n.d. FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arimura Y, Shirayama K, Horikoshi N, Fujita R, Taguchi H, Kagawa W, Fukagawa T, Almouzni G, Kurumizaka H. 2014. Crystal structure and stable property of the cancer-associated heterotypic nucleosome containing CENP-A and H3.3. *Sci Rep* **4**:1–7. doi:10.1038/srep07115

Arunkumar G, Melters DP. 2020. Centromeric transcription: A conserved swiss-army knife. *Genes (Basel)* **11**:1–23. doi:10.3390/genes11080911

Athwal RK, Walkiewicz MP, Baek S, Fu S, Bui M, Camps J, Ried T, Sung MH, Dalal Y. 2015. CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics and Chromatin* **8**:1–23. doi:10.1186/1756-8935-8-2

Au WC, Zhang T, Mishra PK, Eisenstatt JR, Walker RL, Ocampo J, Dawson A, Warren J, Costanzo M, Baryshnikova A, Flick K, Clark DJ, Meltzer PS, Baker RE, Myers C, Boone C, Kaiser P, Basrai MA. 2020. Skp, cullin, F-box (SCF)-Met30 and SCF-Cdc4- mediated proteolysis of CENP-A prevents mislocalization of CENP-A for chromosomal stability in budding yeast, PLoS Genetics. doi:10.1371/journal.pgen.1008597

Aydin ÖZ, Vermeulen W, Lans H. 2014. ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle* **13**:3016–25. doi:10.4161/15384101.2014.956551

Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP, Legube G. 2014. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**:366–374. doi:10.1038/nsmb.2796

Bade D, Pauleau AL, Wendler A, Erhardt S. 2014. The E3 Ligase CUL3/RDX Controls Centromere Maintenance by Ubiquitylating and Stabilizing CENP-A in a CAL1-Dependent Manner. *Dev Cell* **28**:508–519. doi:10.1016/j.devcel.2014.01.031

Baldi S, Becker PB. 2013. The variant histone H2A.V of Drosophila - Three roles, two guises. *Chromosoma* **122**:245–258. doi:10.1007/s00412-013-0409-x

Baldi S, Jain DS, Harpprecht L, Zabel A, Scheibe M, Butter F, Straub T, Becker PB. 2018. Genome-wide Rules of Nucleosome Phasing in Drosophila. *Mol Cell* **72**:661-672.e4. doi:10.1016/j.molcel.2018.09.032

Baldi S, Korber P, Becker PB. 2020. Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nat Struct Mol Biol* **27**:109–118. doi:10.1038/s41594-019-0368-x

Balzano E, Giunta S. 2020. Centromeres under pressure: Evolutionary innovation in conflict with conserved function. *Genes (Basel)* **11**:1–28. doi:10.3390/genes11080912

Bandyopadhyay M, Arbet S, Bishop C, Bidwai A. 2016. Drosophila Protein Kinase CK2: Genetics, Regulatory Complexity and Emerging Roles during Development. *Pharmaceuticals* **10**:4. doi:10.3390/ph10010004

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**:381–395. doi:10.1038/cr.2011.22

Bano D, Piazzesi A, Salomoni P, Nicotera P. 2017. The histone variant H3.3 claims its place in the crowded scene of epigenetics. *Aging (Albany NY)* **9**:602–614. doi:10.18632/aging.101194

Barry AE, Howman E V., Cancilla MR, Saffery R, Choo KHA. 1999. Sequence analysis of an 80 kb human neocentromere. *Hum Mol Genet* **8**:217–227. doi:10.1093/hmg/8.2.217

Barth TK, Schade GOM, Schmidt A, Vetter I, Wirth M, Heun P, Imhof A, Thomae AW. 2015. Identification of Drosophila centromere associated proteins by quantitative affinity purification-mass spectrometry. *Data Br* **4**:544–550. doi:10.1016/j.dib.2015.07.016

Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Zhang J, *et al.* 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**:D480–D489. doi:10.1093/nar/gkaa1100

Becker PB, Workman JL. 2013. Nucleosome remodeling and epigenetics. *Cold Spring Harb Perspect Biol* **5**. doi:10.1101/cshperspect.a017905

Bellelli R, Belan O, Pye VE, Clement C, Maslen SL, Skehel JM, Cherepanov P, Almouzni G, Boulton SJ. 2018a. POLE3-POLE4 Is a Histone H3-H4 Chaperone that Maintains Chromatin Integrity during DNA Replication. *Mol Cell* **72**:112-126.e5. doi:10.1016/j.molcel.2018.08.043

Bellelli R, Borel V, Logan C, Svendsen J, Cox DE, Nye E, Metcalfe K, O'Connell SM, Stamp G, Flynn HR, Snijders AP, Lassailly F, Jackson A, Boulton SJ. 2018b. Polε Instability Drives Replication Stress, Abnormal Development, and Tumorigenesis. *Mol Cell* **70**:707-721.e7. doi:10.1016/j.molcel.2018.04.008

Bhaskara GB, Wong MM, Verslues PE. 2019. The flip side of phospho-signalling: Regulation of protein dephosphorylation and the protein phosphatase 2Cs. *Plant Cell Environ* **42**:2913–2930. doi:10.1111/pce.13616

Blackford AN, Jackson SP. 2017. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell* **66**:801–817. doi:10.1016/j.molcel.2017.05.015

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science (80-)* **277**:1453–1462. doi:10.1126/science.277.5331.1453

Blower MD, Karpen GH. 2001. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat Cell Biol* **3**:730–739. doi:10.1038/35087045

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

Bobkov GOM, Huang A, van den Berg SJW, Mitra S, Anselm E, Lazou V, Schunter S, Feederle R, Imhof A, Lusser A, Jansen LET, Heun P. 2020. Spt6 is a maintenance factor for centromeric CENP-A. *Nat Commun* **11**:2919. doi:10.1038/s41467-020-16695-7

Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, Cleveland DW, Black BE, Shah J V, Jansen LE. 2014. The quantitative architecture of centromeric chromatin. *Elife* **3**:1–26. doi:10.7554/eLife.02137

Bolognese F, Imbriano C, Caretti G, Mantovani R. 2000. Cloning and characterization of the histone-fold proteins YBl1 and YCL1. *Nucleic Acids Res* **28**:3830–3838. doi:10.1093/nar/28.19.3830

Bonev B, Cavalli G. 2016. Organization and function of the 3D genome. *Nat Rev Genet* **17**:661–678. doi:10.1038/nrg.2016.112

Bouazoune K, Brehm A. 2006. ATP-dependent chromatin remodeling complexes in Drosophila. *Chromosom Res* **14**:433–449. doi:10.1007/s10577-006-1067-0

Brandsma I, Gent DC. 2012. Pathway choice in DNA double strand break repair: Observations of a balancing act. *Genome Integr* **3**:1. doi:10.1186/2041-9414-3-9

Brockers K, Schneider R. 2019. Histone H1, the forgotten histone. *Epigenomics* **11**:363–366. doi:10.2217/epi-2019-0018

Bryson T, Henikoff S. 2019. 3XFlag-pATn5 Protein Purification and MEDS-loading (5x scale, 2L volume). *protocols.io*. doi:dx.doi.org/10.17504/protocols.io.8yrhxv6

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**:486–490. doi:10.1038/nature14590

Bunch TA, Grinblat Y, Goldstein LSB. 1988. Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells. *Nucleic Acids Res* **16**:1043–1061. doi:10.1093/nar/16.3.1043

Calo E, Wysocka J. 2013. Modification of Enhancer Chromatin: What, How, and Why? *Mol Cell* **49**:825–837. doi:10.1016/j.molcel.2013.01.038

Canzonetta C, Vernarecci S, Iuliani M, Marracino C, Belloni C, Ballario P, Filetici P. 2016. SAGA DUB-Ubp8 deubiquitylates centromeric histone variant Cse4. *G3 Genes, Genomes, Genet* **6**:287–298. doi:10.1534/g3.115.024877

Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, Pierga J-Y, Brain E, Causeret S, DeLorenzi M, Glas AM, Golfinopoulos V, Goulioti T, Knox S, Matos E, Meulemans B, Neijenhuis PA, Nitz U, Passalacqua R, Ravdin P, Rubio IT, Saghatchian M, Smilde TJ, Sotiriou C, Stork L, Straehle C, Thomas G, Thompson AM, van der Hoeven JM, Vuylsteke P, Bernards R, Tryfonidis K, Rutgers E, Piccart M. 2016. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* **375**:717–729. doi:10.1056/nejmoa1602253

Cartier E, Garcia-Olivares J, Janezic E, Viana J, Moore M, Lin ML, Caplan JL, Torres G, Kim YH. 2019. The SUMO-conjugase Ubc9 prevents the degradation of the dopamine transporter, enhancing its cell surface level and dopamine uptake. *Front Cell Neurosci* **13**:1–23. doi:10.3389/fncel.2019.00035

Carusillo A, Mussolino C. 2020. DNA Damage: From Threat to Treatment. *Cells* **9**:1–20. doi:10.3390/cells9071665

Chang C-H, Chanvan A, Palladino J, Wei X, Martins NMC, Santinello B, Chen C-C, Erceg J, Wu C-T, Larracuente AM, Mellone BG. 2019a. Data from: Islands of retroelements are major components of Drosophila centromeres. *Dryad*. https://doi.org/10.5061/dryad.rb1bt3j

Chang C-H, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, Chen C-C, Erceg J, Beliveau BJ, Wu C-T, Larracuente AM, Mellone BG. 2019b. Islands of retroelements are major components of Drosophila centromeres. *PLOS Biol* **17**:e3000241. doi:10.1371/journal.pbio.3000241

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

Chen C-C, Dechassa ML, Bettini E, Ledoux MB, Belisario C, Heun P, Luger K, Mellone BG. 2014. CAL1 is the Drosophila CENP-A assembly factor. *J Cell Biol* **204**:313–329. doi:10.1083/jcb.201305036

Choi ES, Cheon Y, Kang K, Lee D. 2017. The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3. *Nat Commun* **8**. doi:10.1038/s41467-017-00704-3

Chueh AC, Wong LH, Wong N, Choo KHA. 2005. Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum Mol Genet* **14**:85–93. doi:10.1093/hmg/ddi008

Ciftci-Yilmaz S, Au WC, Mishra PK, Eisenstatt JR, Chang J, Dawson AR, Zhu I, Rahman M, Bilke S, Costanzo M, Baryshnikova A, Myers CL, Meltzer PS, Landsman D, Baker RE, Boone C, Basrai MA. 2018. A genome-wide screen reveals a role for the HIR histone chaperone complex in preventing mislocalization of budding yeast CENP-A. *Genetics* **210**:203–218. doi:10.1534/genetics.118.301305

Clapier CR. 2021. Sophisticated conversations between chromatin and chromatin remodelers, and dissonances in cancer. *Int J Mol Sci* **22**. doi:10.3390/ijms22115578

Clouaire T, Rocher V, Lashgari A, Arnould C, Aguirrebengoa M, Biernacka A, Skrzypczak M, Aymard F, Fongang B, Dojer N, Iacovoni JS, Rowicka M, Ginalski K, Côté J, Legube G. 2018. Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell* 1–13. doi:10.1016/j.molcel.2018.08.020

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

Corona DF, Eberharter A, Budde A, Deuring R, Ferrari S, Varga-Weisz P, Wilm M, Tamkun J, Becker PB. 2000. Two histone fold proteins, CHRAC-14 and CHRAC-16, are developmentally regulated subunits of chromatin accessibility complex (CHRAC). *EMBO J* **19**:3049–59. doi:10.1093/emboj/19.12.3049

Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K, Pasero P, Rowicka M, Dikic I. 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**:361–365. doi:10.1038/nmeth.2408

Cutter AR, Hayes JJ. 2015. A brief review of nucleosome structure. *FEBS Lett* **589**:2914–2922. doi:10.1016/j.febslet.2015.05.016

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**:1–4. doi:10.1093/gigascience/giab008

DeBose-Scarlett EM, Sullivan BA. 2021. Genomic and Epigenetic Foundations of Neocentromere Formation. *Annu Rev Genet* **55**:1–18. doi:10.1146/annurev-genet-071719-020924

Demirdizen E, Spiller-Becker M, Förtsch A, Wilhelm A, Corless S, Bade D, Bergner A, Hessling B, Erhardt S. 2019. Localization of Drosophila CENP-A to non-centromeric sites depends on the NuRD complex. *Nucleic Acids Res* **47**:11589–11608. doi:10.1093/nar/gkz962

Deyter GMR, Biggins S. 2014. The FACT complex interacts with the E3 ubiquitin ligase Psh1 to prevent ectopic localization of CENP-A. *Genes Dev* **28**:1815–1826. doi:10.1101/gad.243113.114

Dong Q, Yang J, Gao J, Li F. 2021. Recent insights into mechanisms preventing ectopic centromere formation. *Open Biol* **11**:210189. doi:10.1098/rsob.210189

Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. 2010. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* **24**:1253–1265. doi:10.1101/gad.566910

Dunleavy EM, Almouzni G, Karpen GH. 2011. H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G 1 phase. *Nucleus* **2**:146–157. doi:10.4161/nucl.2.2.15211

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

Edo K, Koide Y. 1997. Neocarzinostatin Chromophore: Structure and Mechanism of DNA Cleavage. *Neocarzinostatin* 23–45. doi:10.1007/978-4-431-66914-2_3

Erhardt S, Mellone BG, Betts CM, Zhang W, Karpen GH, Straight AF. 2008. Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation. *J Cell Biol* **183**:805–818. doi:10.1083/jcb.200806038

Ferrand J, Plessier A, Polo SE. 2020. Control of the chromatin response to DNA damage: Histone proteins pull the strings. *Semin Cell Dev Biol* 1–13. doi:10.1016/j.semcdb.2020.07.002

Fijen C, Rothenberg E. 2021. The evolving complexity of DNA damage foci: RNA, condensates and chromatin in DNA double-strand break repair. *DNA Repair (Amst)* **105**:103170. doi:10.1016/j.dnarep.2021.103170

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

Foltz DR, Jansen LET, Bailey AO, Yates III JR, Bassett EA, Wood S, Black BE, Cleveland DW. 2009. Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137**:472–484. doi:http://dx.doi.org/10.1016/j.cell.2009.02.039

Freddolino PL, Amini S, Tavazoie S. 2012. Newly identified genetic variations in common Escherichia coli MG1655 stock cultures. *J Bacteriol* **194**:303–306. doi:10.1128/JB.06087-11

Fritz AJ, Sehgal N, Pliss A, Xu J, Berezney R. 2019. Chromosome territories and the global regulation of the genome. *Genes Chromosom Cancer* **58**:407–426. doi:10.1002/gcc.22732

Fu H, Liu N, Dong Q, Ma C, Yang J, Xiong J, Zhang Z, Qi X, Huang C, Zhu B. 2019. SENP6-mediated M18BP1 deSUMOylation regulates CENP-A centromeric localization. *Cell Res* **29**:254–257. doi:10.1038/s41422-018-0139-y

Fukagawa T, Earnshaw WC. 2014. The centromere: Chromatin foundation for the kinetochore machinery. *Dev Cell* **30**:496–508. doi:10.1016/j.devcel.2014.08.016

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**. doi:10.1186/gb-2004-5-10-r80

Giaimo BD, Ferrante F, Herchenröther A, Hake SB, Borggrefe T. 2019. The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* **12**:37. doi:10.1186/s13072-019-0274-9

Gilbert C, Peccoud J, Cordaux R. 2021. Transposable Elements and the Evolution of Insects. *Annu Rev Entomol* **66**:355–372. doi:10.1146/annurev-ento-070720-074650

Gkikopoulos T, Singh V, Tsui K, Awad S, Renshaw MJ, Scholfield P, Barton GJ, Nislow C, Tanaka TU, Owen-Hughes T. 2011. The SWI/SNF complex acts to constrain distribution of the centromeric histone variant Cse4. *EMBO J* **30**:1919–1927. doi:10.1038/emboj.2011.112

Goff L, Trapnell C, Kelley D. 2021. cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data.

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, DeKelver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD. 2010. Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140**:678–691. doi:10.1016/j.cell.2010.01.003

Gregory TR. 2021. Animal Genome Size Database.

Halic M, Moazed D. 2009. Transposon Silencing by piRNAs. *Cell* **138**:1058–1060. doi:10.1016/j.cell.2009.08.030

Hammond CM, Strømme CB, Huang H, Patel DJ, Groth A. 2017. Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol* **18**:141–158. doi:10.1038/nrm.2016.159

Hao Y, Wang D, Wu S, Li X, Shao C, Zhang P, Chen JY, Lim DH, Fu XD, Chen R, He S. 2020. Active retrotransposons help maintain pericentromeric heterochromatin required for faithful cell division. *Genome Res* **30**:1570–1582. doi:10.1101/gr.256131.119

Harpprecht L, Baldi S, Schauer T, Schmidt A, Bange T, Robles MS, Kremmer E, Imhof A, Becker PB. 2019. A Drosophila cell-free system that senses DNA breaks and triggers phosphorylation signalling. *Nucleic Acids Res* **47**:7444–7459. doi:10.1093/nar/gkz473

Hartlepp KF, Fernández-Tornero C, Eberharter A, Grüne T, Müller CW, Becker PB. 2005. The histone fold subunits of Drosophila CHRAC facilitate nucleosome sliding through dynamic DNA interactions. *Mol Cell Biol* **25**:9886–9896. doi:10.1128/MCB.25.22.9886-9896.2005

Hasson D, Alonso A, Cheung F, Tepperberg JH, Papenhausen PR, Engelen JJM, Warburton PE. 2011. Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. *Chromosoma* **120**:621–632. doi:10.1007/s00412-011-0337-6

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**:576–589. doi:10.1016/j.molcel.2010.05.004

Helfricht A, Wiegant WW, Thijssen PE, Vertegaal AC, Luijsterburg MS, Attikum H Van. 2013. Remodeling and

spacing factor 1 (RSF1) deposits centromere proteins at DNA double-strand breaks to promote non-homologous end-joining. *Cell Cycle* **12**:3070–3082. doi:10.4161/cc.26033

Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science (80-)* **293**:1098–1102. doi:10.1126/science.1062939

Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH. 2006. Mislocalization of the Drosophila Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. *Dev Cell* **10**:303–315. doi:10.1016/j.devcel.2006.01.014

Hewawasam GS, Dhatchinamoorthy K, Mattingly M, Seidel C, Gerton JL. 2018. Chromatin assembly factor-1 (CAF-1) chaperone regulates Cse4 deposition into chromatin in budding yeast. *Nucleic Acids Res* **46**:4440–4455. doi:10.1093/nar/gky169

Hirai H, Shogaki Y, Sato M. 2021. Mis6/CENP-I maintains CENP-A nucleosomes against centromeric non-coding transcription during mitosis. *bioRxiv* 2021.11.03.466203. doi:10.1101/2021.11.03.466203

Horard B, Loppin B. 2015. Histone storage and deposition in the early Drosophila embryo. *Chromosoma* **124**:163–175. doi:10.1007/s00412-014-0504-7

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, Krzywinski M, Schein J, Accardo MC, Damia E, Messina G, Méndez-Lago M, De Pablos B, Demakova O V., Andreyeva EN, Boldyreva L V., Marra M, Carvalho AB, Dimitri P, Villasante A, Zhimulev IF, Rubin GM, Karpen GH, Celniker SE. 2015. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res* **25**:445–458. doi:10.1101/gr.185579.114

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Flicek P, *et al.* 2020. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res* **48**:D689–D695. doi:10.1093/nar/gkz890

Huang A, Kremser L, Schuler F, Wilflingseder D, Lindner H, Geley S, Lusser A. 2019. Phosphorylation of Drosophila CENP-A on serine 20 regulates protein turn-over and centromere-specific loading. *Nucleic Acids Res* 1–17. doi:10.1093/nar/gkz809

Iacovoni JS, Caron P, Lassadi I, Nicolas E, Massip L, Trouche D, Legube G. 2010. High-resolution profiling of γH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* **29**:1446–1457. doi:10.1038/emboj.2010.38

Ishii K, Ogiyama Y, Chikashige Y, Soejima S, Masuda F, Kakuma T, Hiraoka Y, Takahashi K. 2008. Heterochromatin Integrity Affects Chromosome Reorganization After Centromere Dysfunction. *Science (80-)* **321**:1088–1091. doi:10.1126/science.1158699

Ismail IH, Hendzel MJ. 2008. The γ-H2A.X: Is It Just a Surrogate Marker of Double-Strand Breaks or Much More? *Environ Mol Mutagen* **49**:73–82. doi:10.1002/em

Ivanauskiene K, Delbarre E, McGhie JD, Kuntziger T, Wong LH, Collas P. 2014. The PML-associated protein DEK regulates the balance of H3.3 loading on chromatin and is important for telomere integrity. *Genome Res* **24**:1584–1594. doi:10.1101/gr.173831.114

Jansen LET, Black BE, Foltz DR, Cleveland DW. 2007. Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol* **176**:795–805. doi:10.1083/jcb.200701066

Jeffery D, Podsypanina K, Gatto A, Landete RP, Bonneville L, Dumont M, Fachinetti D, Almouzni G. 2020. CENP-A overexpression drives distinct cell fates depending on p53 status. *Commun Biol* 1–18. doi:10.1101/2020.07.21.213496

Jin Y, Tam OH, Paniagua E, Hammell M. 2015. TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**:3593–3599. doi:10.1093/bioinformatics/btv422

Juhász S, Elbakry A, Mathes A, Löbrich M. 2018. ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. *Mol Cell* **71**:11-24.e7. doi:10.1016/j.molcel.2018.05.014

Kakarougkas A, Downs JA, Jeggo PA. 2015. The PBAF chromatin remodeling complex represses transcription and promotes rapid repair at DNA double-strand breaks. *Mol Cell Oncol* **2**:e970072. doi:10.4161/23723548.2014.970072

Kakarougkas A, Ismail A, Chambers AL, Riballo E, Herbert AD, Künzel J, Löbrich M, Jeggo PA, Downs JA. 2014. Requirement for PBAF in Transcriptional Repression and Repair at DNA Breaks in Actively Transcribed Regions of Chromatin. *Mol Cell* **55**:723–732. doi:10.1016/j.molcel.2014.06.028

Kassis JA, Kennison JA, Tamkun JW. 2017. Polycomb and Trithorax Group Genes in Drosophila. *Genetics* **206**:1699–1725. doi:10.1534/genetics.115.185116

Kavanaugh GM, Wise-Draper TM, Morreale RJ, Morrison MA, Gole B, Schwemberger S, Tichy ED, Lu L, Babcock GF, Wells JM, Drissi R, Bissler JJ, Stambrook PJ, Andreassen PR, Wiesmüller L, Wells SI. 2011. The human DEK oncogene regulates DNA damage response signaling and repair. *Nucleic Acids Res* **39**:7465–7476. doi:10.1093/nar/gkr454

Kaya-Okur HS, Henikoff S. 2020. Bench top CUT&Tag. protocols.io. *protocols.io*. https://dx.doi.org/10.17504/protocols.io.bcuhiwt6

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**:1930. doi:10.1038/s41467-019-09982-5

Kennedy EP. 1992. Sailing to byzantium. *Annu Rev Biochem* **61**:1–28. doi:10.5422/fordham/9780823254644.003.0005

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics* **26**:2204–2207. doi:10.1093/bioinformatics/btq351

Ketel C, Wang HSW, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J. 2009. Neocentromeres form efficiently at multiple possible loci in candida albicans. *PLoS Genet* **5**. doi:10.1371/journal.pgen.1000400

Kim D, Langmead B, Salzberg SL. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods* **12**:357–360. doi:10.1038/nmeth.3317

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**:907–915. doi:10.1038/s41587-019-0201-4

Kim JJ, Lee SY, Miller KM. 2019. Preserving genome integrity and function: the DNA damage response and histone modifications. *Crit Rev Biochem Mol Biol* **54**:208–241. doi:10.1080/10409238.2019.1620676

Klemm SL, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**:207–220. doi:10.1038/s41576-018-0089-8

Kotova E, Lodhi N, Jarnik M, Pinnola AD, Ji Y, Tulin A V. 2011. Drosophila histone H2A variant (H2Av) controls poly(ADP-ribose) polymerase 1 (PARP1) activation. *Proc Natl Acad Sci U S A* **108**:6205–6210. doi:10.1073/pnas.1019644108

Kukimoto I, Elderkin S, Grimaldi M, Oelgeschläger T, Varga-Weisz PD. 2004. The Histone-Fold Protein Complex CHRAC-15/17 Enhances Nucleosome Sliding and Assembly Mediated by ACF. *Mol Cell* **13**:265–277. doi:10.1016/S1097-2765(03)00523-9

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

Lagana A, Dorn JF, De Rop V, Ladouceur AM, Maddox AS, Maddox PS. 2010. A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. *Nat Cell Biol* **12**:1186–1193. doi:10.1038/ncb2129

Lai WKM, Pugh BF. 2017. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat Rev Mol Cell Biol* **18**:548–562. doi:10.1038/nrm.2017.47

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359. doi:10.1038/nmeth.1923

Lee H, McManus CJ, Cho DY, Eaton M, Renda F, Somma MP atrizi., Cherbas L, May G, Powell S, Zhang D, Zhan L, Resch A, Andrews J, Celniker SE, Cherbas P, Przytycka TM, Gatti M, Oliver B, Graveley B, MacAlpine D.

2014. DNA copy number evolution in Drosophila cell lines. *Genome Biol* **15**:R70. doi:10.1186/gb-2014-15-8-r70

Lee TI, Young RA. 2013. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **152**:1237–1251. doi:10.1016/j.cell.2013.02.014

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Li R, Grimm SA, Wade PA. 2021. CUT&Tag-BS for simultaneous profiling of histone modification and DNA methylation with high efficiency and low cost. *Cell Reports Methods* 100118. doi:10.1016/j.crmeth.2021.100118

Li X, Heyer WD. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res* **18**:99–113. doi:10.1038/cr.2008.1

Li Y, Pursell ZF, Linn S. 2000. Identification and cloning of two histone fold motif-containing subunits of HeLa DNA polymerase epsilon. *J Biol Chem* **275**:23247–52. doi:10.1074/jbc.M002548200

Li Y, Zhu Z, Zhang S, Yu D, Yu H, Liu L, Cao X, Wang L, Gao H, Zhu M. 2011. Shrna-targeted centromere protein a inhibits hepatocellular carcinoma growth. *PLoS One* **6**:19–23. doi:10.1371/journal.pone.0017794

Lidsky P V., Sprenger F, Lehner CF. 2013. Distinct modes of centromere protein dynamics during cell cycle progression in Drosophila S2R+ cells. *J Cell Sci* **126**:4782–4793. doi:10.1242/jcs.134122

Liebelt F, Jansen NS, Kumar S, Gracheva E, Claessens LA, Verlaan-de Vries M, Willemstein E, Vertegaal ACO. 2019. The poly-SUMO2/3 protease SENP6 enables assembly of the constitutive centromere-associated network by group deSUMOylation. *Nat Commun* **10**:1–18. doi:10.1038/s41467-019-11773-x

Liu G, Yan J, Wang Xuejie, Chen J, Wang Xin, Dong Y, Zhang S, Gan X, Huang J, Chen X. 2021. RPA-mediated recruitment of Bre1 couples histone H2B ubiquitination to DNA replication and repair. *Proc Natl Acad Sci U S A* **118**:1–10. doi:10.1073/pnas.2017497118

Liu J, Ali M, Zhou Q. 2020. Establishment and evolution of heterochromatin. *Ann N Y Acad Sci* **1476**:59–77. doi:10.1111/nyas.14303

Llorens-Giralt P, Camilleri-Robles C, Corominas M, Climent-Cantó P. 2021. Chromatin Organization and Function in Drosophila. *Cells* **10**:2362. doi:10.3390/cells10092362

Lo AWI. 2001. A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J* **20**:2087–2096. doi:10.1093/emboj/20.8.2087

Lo AWI, Magliano DJ, Sibson MC, Kalitsis P, Craig JM, Choo KHA. 2001. A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. *Genome Res* **11**:448–457. doi:10.1101/gr.GR-1676R

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:1–21. doi:10.1186/s13059-014-0550-8

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**:251–260. doi:10.1038/38444

Luo S, Lu J. 2017. Silencing of Transposable Elements by piRNAs in Drosophila: An Evolutionary Perspective. *Genomics, Proteomics Bioinforma* **15**:164–176. doi:10.1016/j.gpb.2017.01.006

Madigan JP, Chotkowski HL, Glaser RL. 2002. DNA double-strand break-induced phosphorylation of Drosophila histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res* **30**:3698–3705. doi:10.1093/nar/gkf496

Magaña-Acosta M, Valadez-Graham V. 2020. Chromatin Remodelers in the 3D Nuclear Compartment. *Front Genet* **11**:1–22. doi:10.3389/fgene.2020.600615

Mahlke MA, Nechemia-Arbely Y. 2020. Guarding the Genome: CENP-A-Chromatin in Health and Cancer. *Genes (Basel)* **11**:810. doi:10.3390/genes11070810

Makharashvili N, Paull TT. 2015. CtIP: A DNA damage response protein at the intersection of DNA metabolism. *DNA Repair (Amst)* **32**:75–81. doi:10.1016/j.dnarep.2015.04.016

Marygold SJ, Attrill H, Speretta E, Warner K, Magrane M, Berloco M, Cotterill S, McVey M, Rong Y, Yamaguchi M. 2020. The DNA polymerases of Drosophila melanogaster. *Fly (Austin)* **14**:49–61. doi:10.1080/19336934.2019.1710076

Mathew V, Pauleau A-L, Steffen N, Bergner A, Becker P, Erhardt S. 2014. The Histone-Fold Protein CHRAC14 Influences Chromatin Composition in Response to DNA Damage. *Cell Rep* **7**:321–330. doi:10.1016/j.celrep.2014.03.008

Meggio F, Pinna LA. 2003. One-thousand-and-one substrates of protein kinase CK2? *FASEB J* **17**:349–368. doi:10.1096/fj.02-0473rev

Mellone BG, Grive KJ, Shteyn V, Bowers SR, Oderberg I, Karpen GH. 2011. Assembly of drosophila centromeric chromatin proteins during mitosis. *PLoS Genet* **7**. doi:10.1371/journal.pgen.1002068

Mendiburo MJ, Padeken J, Fülöp S, Schepers A, Heun P. 2011. Drosophila CENH3 Is Sufficient for Centromere Formation. *Science (80-)* **334**:686–690. doi:10.1126/science.1206880

Mitra S, Bodor DL, David AF, Abdul-Zani I, Mata JF, Neumann B, Reither S, Tischer C, Jansen LET. 2020. Genetic screening identifies a SUMO protease dynamically maintaining centromeric chromatin. *Nat Commun* **11**:1–15. doi:10.1038/s41467-019-14276-x

Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. 2004. ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science (80-)* **303**:343–348. doi:10.1126/science.1090701

Montenarh M. 2016. Protein kinase CK2 in DNA damage and repair. *Transl Cancer Res* **5**:49–63. doi:10.3978/j.issn.2218-676X.2016.01.09

Moreno-Moreno O, Torras-Llort M, Azorin F. 2019. The E3-ligases SCFPpa and APC/CCdh1 co-operate to regulate CENP-ACID expression across the cell cycle. *Nucleic Acids Res* **47**:3395–3406. doi:10.1093/nar/gkz060

Murakami Y. 2013. Encyclopedia of Systems Biology, Encyclopedia of Systems Biology. New York, NY: Springer New York. doi:10.1007/978-1-4419-9863-7

Murillo-Pineda M, Valente LP, Dumont M, Mata JF, Fachinetti D, Jansen LET. 2021. Induction of spontaneous human neocentromere formation and long-term maturation. *J Cell Biol* **220**. doi:10.1083/JCB.202007210

Musacchio A, Desai A. 2017. A Molecular View of Kinetochore Assembly and Function. *Biology (Basel)* **6**:5. doi:10.3390/biology6010005

Naughton C, Huidobro C, Catacchio CR, Buckle A, Grimes GR, Nozawa R-S, Purgato S, Rocchi M, Gilbert N. 2021. Human centromere formation activates transcription and opens chromatin fibre structure. *bioRxiv* 1–27. doi:https://doi.org/10.1101/2021.08.01.454615

Nechemia-Arbely Y, Miga KH, Shoshani O, Aslanian A, McMahon MA, Lee AY, Fachinetti D, Yates JR, Ren B, Cleveland DW. 2019. DNA replication acts as an error correction mechanism to maintain centromere identity by restricting CENP-A to centromeres. *Nat Cell Biol* **21**:743–754. doi:10.1038/s41556-019-0331-4

Niikura Y, Kitagawa R, Kitagawa K. 2019. CENP-A ubiquitylation contributes to maintaining the chromosomal location of the centromere. *Molecules* **24**:10–14. doi:10.3390/molecules24030402

Niikura Y, Kitagawa R, Kitagawa K, Diseases B. 2016. CENP-A Ubiquitylation Is Inherited through Dimerization between Cell Divisions. *Cell Rep* **15**:61–76. doi:10.1016/j.celrep.2016.03.010.CENP-A

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

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**:1–30. doi:10.1017/CBO9781107415324.004

Nye J, Sturgill D, Athwal R, Dalal Y. 2018. HJURP antagonizes CENP-A mislocalization driven by the H3.3

chaperones HIRA and DAXX. *PLoS One* **13**:1–21. doi:10.1371/journal.pone.0205948

Ohku K, Suva E, Au W-C, Walker RL, Levy-Myers R, Meltzer PS, Baker RE, Basrai MA. 2020. Deposition of centromeric histone H3 variant CENP-A/Cse4 into chromatin is facilitated by its C-terminal sumoylation. *Genetics*.

Ohkuni K, Abdulle R, Kitagawa K. 2014. Degradation of centromeric histone H3 variant Cse4 requires the Fpr3 peptidyl-prolyl cis-trans isomerase. *Genetics* **196**:1041–1045. doi:10.1534/genetics.114.161224

Ohkuni K, Levy-Myers R, Warren J, Au WC, Takahashi Y, Baker RE, Basrai MA. 2018. N-terminal sumoylation of centromeric histone H3 variant Cse4 regulates its proteolysis to prevent mislocalization to non-centromeric chromatin. *G3 Genes, Genomes, Genet* **8**:1215–1223. doi:10.1534/g3.117.300419

Ohkuni K, Takahashi Y, Fulp A, Lawrimore J, Au WC, Pasupala N, Levy-Myers R, Warren J, Strunnikov A, Baker RE, Kerscher O, Bloom K, Basrai MA. 2016. SUMO-targeted ubiquitin ligase (STUbL) Slx5 regulates proteolysis of centromeric histone H3 variant Cse4 and prevents its mislocalization to euchromatin. *Mol Biol Cell* **27**:1500–1510. doi:10.1091/mbc.E15-12-0827

Olszak AM, Van Essen D, Pereira AJ, Diehl S, Manke T, Maiato H, Saccani S, Heun P. 2011. Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol* **13**:799–808. doi:10.1038/ncb2272

Panigrahi A, O'Malley BW. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biol* **22**:1–30. doi:10.1186/s13059-021-02322-1

Parmar JJ, Padinhateeri R. 2020. Nucleosome positioning and chromatin organization. *Curr Opin Struct Biol* **64**:111–118. doi:10.1016/j.sbi.2020.06.021

Perpelescu M, Nozaki N, Obuse C, Yang H, Yoda K. 2009. Active establishment of centromeric cenp-a chromatin by rsf complex. *J Cell Biol* **185**:397–407. doi:10.1083/jcb.200903088

Petty E, Pillus L. 2013. Balancing chromatin remodeling and histone modifications in transcription. *Trends Genet* **29**:621–629. doi:10.1016/j.tig.2013.06.006

Poot RA, Dellaire G, Hülsmann BB, Grimaldi MA, Corona DF V, Becker PB, Bickmore WA, Weisz PDV. 2000. HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J* **19**:3377–3387. doi:10.1093/emboj/19.13.3377

Qiu JJ, Guo JJ, Lv TJ, Jin HY, Ding JX, Feng WW, Zhang Y, Hua KQ. 2013. Prognostic value of centromere protein-A expression in patients with epithelial ovarian cancer. *Tumor Biol* **34**:2971–2975. doi:10.1007/s13277-013-0860-6

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033

R: A Language and Environment for Statistical Computing. n.d. https://www.r-project.org/

Ramachandran H, Herfurth K, Grosschedl R, Schäfer T, Walz G. 2015. SUMOylation blocks the ubiquitin-mediated degradation of the nephronophthisis gene product Glis2/NPHP7. *PLoS One* **10**:1–16. doi:10.1371/journal.pone.0130275

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165. doi:10.1093/nar/gkw257

Reyes AA, Marcum RD, He Y. 2021. Structure and Function of Chromatin Remodelers. *J Mol Biol* **433**:166929. doi:10.1016/j.jmb.2021.166929

Ribeiro-Silva C, Vermeulen W, Lans H. 2019. SWI/SNF: Complex complexes in genome stability and cancer. *DNA Repair (Amst)* **77**:87–95. doi:10.1016/j.dnarep.2019.03.007

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**:e47–e47. doi:10.1093/nar/gkv007

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative

genomics viewer. *Nat Biotechnol* **29**:24–26. doi:10.1038/nbt.1754

Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**:5858–5868. doi:10.1074/jbc.273.10.5858

Rogers SL, Rogers GC. 2008. Culture of Drosophila S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nat Protoc* **3**:606–11. doi:10.1038/nprot.2008.18

Rosin LF, Mellone BG. 2017. Centromeres Drive a Hard Bargain. *Trends Genet* **33**:101–117. doi:10.1016/j.tig.2016.12.001

Rott R, Szargel R, Shani V, Hamza H, Savyon M, Elghani FA, Bandopadhyay R, Engelender S. 2017. SUMOylation and ubiquitination reciprocally regulate α-synuclein degradation and pathological aggregation. *Proc Natl Acad Sci U S A* **114**:13176–13181. doi:10.1073/pnas.1704351114

RStudio: Integrated Development Environment for R. 2020. http://www.rstudio.com/

Rusin SF, Adamo ME, Kettenbach AN. 2017. Identification of Candidate Casein Kinase 2 Substrates in Mitosis by Quantitative Phosphoproteomics. *Front Cell Dev Biol* **5**:1–16. doi:10.3389/fcell.2017.00097

Saha AK, Contreras-Galindo R, Niknafs YS, Iyer M, Qin T, Padmanabhan K, Siddiqui J, Palande M, Wang C, Qian B, Ward E, Tang T, Tomlins SA, Gitlin SD, Sartor MA, Omenn GS, Chinnaiyan AM, Markovitz DM. 2020. The role of the histone H3 variant CENPA in prostate cancer. *J Biol Chem* **295**:8537–8549. doi:10.1074/jbc.RA119.010080

Scacchetti A, Becker PB. 2020. Loss of nucleosome remodelers CHRAC/ACF does not sensitize early Drosophila embryos to X-rays. *microPublication Biol* **2020**. doi:10.17912/micropub.biology.000287

Scacchetti A, Brueckner L, Jain D, Schauer T, Zhang X, Schnorrer F, van Steensel B, Straub T, Becker PB. 2018. CHRAC/ACF contribute to the repressive ground state of chromatin. *Life Sci Alliance* **1**:1–12. doi:10.26508/lsa.201800024

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

Schneider I. 1972. Cell lines derived from late embryonic stages of Drosophila melanogaster. *J Embryol Exp Morphol* **27**:353–65.

Schuh M, Lehner CF, Heidmann S. 2007. Incorporation of Drosophila CID/CENP-A and CENP-C into Centromeres during Early Embryonic Anaphase. *Curr Biol* **17**:237–243. doi:10.1016/j.cub.2006.11.051

Schwartz YB, Cavalli G. 2017. Three-dimensional genome organization and function in Drosophila. *Genetics* **205**:5–24. doi:10.1534/genetics.115.185132

Sekelsky J. 2017. DNA Repair in Drosophila: Mutagens, Models, and Missing Genes. *Genetics* **205**:471–490. doi:10.1534/genetics.116.186759

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**:458–472. doi:10.1016/j.cell.2012.01.010

Shi L, Wen H, Shi X. 2017. The Histone Variant H3.3 in Transcriptional Regulation and Human Disease. *J Mol Biol* **429**:1934–1945. doi:10.1016/j.jmb.2016.11.019

Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, Gisselbrecht SS, Dainese R, Bischof J, Furger E, Feuz J, Basler K, Deplancke B, Bulyk ML. 2019. A Comprehensive Drosophila melanogaster Transcription Factor Interactome. *Cell Rep* **27**:955-970.e7. doi:10.1016/j.celrep.2019.03.071

Shrestha RL, Ahn GS, Staples MI, Sathyan KM, Karpova TS, Foltz DR, Basrai MA. 2017. Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. *Oncotarget* **8**:46781–46800. doi:10.18632/oncotarget.18108

Shrestha RL, Rossi A, Wangsa D, Hogan AK, Zaldana KS, Suva E, Chung YJ, Sanders CL, Difilippantonio S, Karpova TS, Karim B, Foltz DR, Fachinetti D, Aplan PD, Ried T, Basrai MA. 2021. CENP-A overexpression

promotes aneuploidy with karyotypic heterogeneity. *J Cell Biol* **220**. doi:10.1083/JCB.202007195

Shukla M, Tong P, White SA, Singh PP, Reid AM, Catania S, Pidoux AL, Allshire RC. 2018. Centromere DNA Destabilizes H3 Nucleosomes to Promote CENP-A Deposition during the Cell Cycle. *Curr Biol* **28**:3924-3936.e4. doi:10.1016/j.cub.2018.10.049

Sikorska N, Sexton T. 2020. Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated. *J Mol Biol* **432**:653–664. doi:10.1016/j.jmb.2019.12.006

Simmons JR, An R, Amankwaa B, Zayac S, Kemp J, Labrador M. 2021. A Phosphorylated Histone H2A Variant Displays Properties of Chromatin Insulator Proteins in Drosophila. *bioRxiv* 2021.02.23.432395.

Singh PP, Shukla M, White SA, Lafos M, Tong P, Auchynnikava T, Spanos C, Rappsilber J, Pidoux AL, Allshire RC. 2020. Hap2-Ino80-facilitated transcription promotes de novo establishment of CENP-A chromatin. *Genes Dev* **34**:226–238. doi:10.1101/gad.332536.119

Siomi MC, Higashijima K, Ishizuka A, Siomi H. 2002. Casein Kinase II Phosphorylates the Fragile X Mental Retardation Protein and Modulates Its Biological Properties. *Mol Cell Biol* **22**:8438–8447. doi:10.1128/mcb.22.24.8438-8447.2002

Smit A, Hubley R, Glusman G. n.d. Repeat Masker. https://www.repeatmasker.org/

Song Y-H. 2005. Drosophila melanogaster: a model for the study of DNA damage checkpoint response. *Mol Cells* **19**:167–179. doi:828 [pii]

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

Suganuma T, Gutiérrez JL, Li B, Florens L, Swanson SK, Washburn MP, Abmayr SM, Workman JL. 2008. ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol* **15**:364–372. doi:10.1038/nsmb.1397

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc B Biol Sci* **375**. doi:10.1098/rstb.2019.0347

Svitin A, Chesnokov I. 2010. Study of DNA replication in Drosophila using cell free in vitro system. *Cell Cycle* **9**:815–819. doi:10.4161/cc.9.4.10730

Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* **5**. doi:10.1126/sciadv.aaw1668

Szenker E, Ray-Gallet D, Almouzni G. 2011. The double face of the histone variant H3.3. *Cell Res* **21**:421–434. doi:10.1038/cr.2011.14

Takada S, Brandani GB, Tan C. 2020. Nucleosomes as allosteric scaffolds for genetic regulation. *Curr Opin Struct Biol* **62**:93–101. doi:10.1016/j.sbi.2019.11.013

Takahashi K, Chen ES, Yanagida M. 2000. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science (80-)* **288**:2215–2219. doi:10.1126/science.288.5474.2215

Talbert PB, Henikoff S. 2021. Histone variants at a glance. *J Cell Sci* **134**:1–10. doi:10.1242/jcs.244749

Tardat M, Déjardin J. 2018. Telomere chromatin establishment and its maintenance during mammalian development. *Chromosoma* **127**:3–18. doi:10.1007/s00412-017-0656-3

Tomonaga T, Matsushita K, Yamaguchi S, Oohashi T, Shimada H, Ochiai T, Yoda K, Nomura F. 2003. Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* **63**:3511–3516.

Torres-Zelada EF, Weake VM. 2021. The Gcn5 complexes in Drosophila as a model for metazoa. *Biochim Biophys Acta - Gene Regul Mech* **1864**:194610. doi:10.1016/j.bbagrm.2020.194610

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**:46–53. doi:10.1038/nbt.2450

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**:511–515. doi:10.1038/nbt.1621

Usakin L, Abad JP, Vagin V V., De Pablos B, Villasante A, Gvozdev V a. 2007. Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in Drosophila melanogaster ovaries. *Genetics* **176**:1343–1349. doi:10.1534/genetics.107.071720

van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. 2010. Hi-C: A method to study the three-dimensional architecture of genomes. *J Vis Exp* 1–7. doi:10.3791/1869

Vítor AC, Huertas P, Legube G, de Almeida SF. 2020. Studying DNA Double-Strand Break Repair: An Ever-Growing Toolbox. *Front Mol Biosci* **7**:1–16. doi:10.3389/fmolb.2020.00024

Wei KH-C, Grenier JK, Barbash DA, Clark AG. 2014. Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster. *Proc Natl Acad Sci* **111**:18793–18798. doi:10.1073/pnas.1421951112

Wei KHC, Lower SE, Caldas I V., Sless TJS, Barbash DA, Clark AG. 2018. Variable rates of simple satellite gains across the drosophila phylogeny. *Mol Biol Evol* **35**:925–941. doi:10.1093/molbev/msy005

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D. 2002. Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Mol Biol Cell* **13**:1977–2000. doi:10.1091/mbc.02-02-0030

Wong CYY, Lee BCH, Yuen KWY. 2020. Epigenetic regulation of centromere function. *Cell Mol Life Sci* **77**:2899–2917. doi:10.1007/s00018-020-03460-8

Wu Q, Qian YM, Zhao XL, Wang SM, Feng XJ, Chen XF, Zhang SH. 2012. Expression and prognostic significance of centromere protein A in human lung adenocarcinoma. *Lung Cancer* **77**:407–414. doi:10.1016/j.lungcan.2012.04.007

Yates AD, Achuthan P, Akanni W, Allen James, Allen Jamie, Flicek P, *et al.* 2020. Ensembl 2020. *Nucleic Acids Res* **48**:D682–D688. doi:10.1093/nar/gkz966

Yu G, Wang LG, Han Y, He QY. 2012. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol* **16**:284–287. doi:10.1089/omi.2011.0118

Zasadzińska E, Huang J, Bailey AO, Guo LY, Lee NS, Srivastava S, Wong KA, French BT, Black BE, Foltz DR. 2018. Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP. *Dev Cell* **47**:348-362.e7. doi:10.1016/j.devcel.2018.09.003

Zeitlin SG, Baker NM, Chapados BR, Soutoglou E, Wang JYJ, Berns MW, Cleveland DW. 2009. Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci U S A* **106**:15762–15767. doi:10.1073/pnas.0908233106

Zhang W, Mao JH, Zhu W, Jain AK, Liu K, Brown JB, Karpen GH. 2016. Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat Commun* **7**. doi:10.1038/ncomms12619

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Shirley XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**. doi:10.1186/gb-2008-9-9-r137

Zhang Y, Yang L, Shi J, Lu Y, Chen X, Yang Z. 2020. The Oncogenic Role of CENPA in Hepatocellular Carcinoma Development: Evidence from Bioinformatic Analysis. *Biomed Res Int* **2020**. doi:10.1155/2020/3040839

Zheng S, Li D, Lu Z, Liu G, Wang Meng, Xing P, Wang Min, Dong Y, Wang X, Li J, Zhang S, Peng H, Ira G, Li G, Chen X. 2018. Bre1-dependent H2B ubiquitination promotes homologous recombination by stimulating histone eviction at DNA breaks. *Nucleic Acids Res* **46**:11326–11339. doi:10.1093/nar/gky918

Zheng Y, Ahmad K, Henikoff S. 2020. CUT&Tag Data Processing and Analysis Tutorial. *protocols.io*. doi:dx.doi.org/10.17504/protocols.io.bjk2kkye

Zhou W, Liang G, Molloy PL, Jones PA. 2020. DNA methylation enables transposable element-driven genome expansion. *Proc Natl Acad Sci U S A* **117**:19359–19366. doi:10.1073/pnas.1921719117

Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, Green MR. 2010. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**. doi:10.1186/1471-2105-11-237

Zhu Q, Liu N, Orkin SH, Yuan GC. 2019. CUT and RUNTools: A flexible pipeline for CUT and RUN processing and footprint analysis. *Genome Biol* **20**:1–12. doi:10.1186/s13059-019-1802-4

Żylicz JJ, Heard E. 2020. Molecular Mechanisms of Facultative Heterochromatin Formation: An X-Chromosome Perspective. *Annu Rev Biochem* **89**:255–282. doi:10.1146/annurev-biochem-062917-012655

ACKNOWLEDGEMENTS

First and foremost, I sincerely thank my doctoral supervisor Prof. Sylvia Erhardt for the opportunity to pursue my scientific work in her lab, for her countless advices and for critically reading and grading this work. I greatly appreciate her valuable scientific guidance, which advanced my project at all times.

I gladly thank my graduate school HBIGS and the members of my thesis advisory committee Prof. Dr. Schiebel and Prof. Dr. Hofmann who supported this study with great input and helpful directions throughout the years. I am especially thankful to Prof. Schiebel who agreed to be my second referee and for reading and grading my thesis. I would also like to thank Prof. Dr. Rippe and Prof. Dr. Bukau who consent to be part of my defense committee.

I would like to express special thanks to Prof. Dr. Rippe for generously hosting me in his lab in BioQuant during demanding times when our own institute was shut down and for the much appreciated scientific and personal advice.

Very heartfelt thanks goes to my present and former colleagues Janina and Mukta. Thank you for your vital friendship, scientific and non-scientific support and indispensable help throughout my PhD, constantly discussing ideas and issues over coffee or pizza during our countless Botanik sessions. Many thanks to Alex and all my other great current and alumni lab co-workers for the scientifically stimulating discussions, your help in the lab and the warm and friendly atmosphere, which made work life so much more pleasant.

My sincere thanks also go to the technicians and researchers of the Schiebel, Rippe, Bukau, Kaessmann and Joeazeiro labs for their kind help, technical support and for sharing materials and protocols whenever needed.

I highly appreciate and thank the EMBL and ZMBH proteomics facilities and the ZMBH sequencing unit in the Bukau lab for conducting my screens in the most helpful and scientifically knowledgeable way.

I gladly acknowledge support by the state of Baden-Württemberg through the German Research Foundation (DFG) for financing my project through the SFB1036 grant, for bwHPC and the grant INST 35/1134-1 FUGG providing the opportunity of conducting my data analysis in the BwForCluster MLS&WISO (Production) cluster.

Last but most importantly, very special thanks and infinite gratitude is dedicated to my outstanding and priceless support system comprising my lovely family, Mum, Dad and my brother Christoph, my wonderful partner Gustav and my magnificent friends. Thank you so much, you never failed to encourage and empower me throughout this journey.