Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften an der Fakultät für Biologie der Ludwig-Maximilians-Universität München

IDENTIFICATION AND CHARACTERIZATION OF NOVEL *DROSOPHILA MELANOGASTER* CENTROMERE-ASSOCIATED PROTEINS

Vorgelegt von

Teresa Barth

Aus Neumarkt i.d.OPf.

München, August 2014

Dissertation eingereicht am:	26.08.2014
Mündliche Prüfung am:	19.11.2014
1. Gutachter:	Prof. Dr. Peter Becker
2. Gutachter:	Prof. Dr. Dirk Eick
3. Gutachter:	Prof. Dr. Angelika Böttger
4. Gutachter:	Prof. Dr. John Parsch
5. Gutachter:	Prof. Dr. Michael Boshart
6. Gutachter:	PD Dr. Anna Friedl

Wesentliche Teile dieser Arbeit sind veröffentlicht in:

Proteomics, 2014 May 19, DOI 10.1002/pmic.201400052
Identification of novel Drosophila centromere-associated proteins.
Barth TK*, Schade GO*, Schmidt A, Vetter I, Wirth M, Heun P, Thomae AW, Imhof A.
*gleiche Beteiligung

Mitwirkungen:

Statistische Datenanalyse für Abbildung 2.10 durchgeführt von Dr. Andreas Schmidt (ZfP LMU München).
Abbildungen 2.14, 2.15 und 2.17 erstellt in Zusammenarbeit mit Georg Schade (MPI Freiburg).
Abbildungen 2.16, 2.18, 2.25 und 2.26 erstellt von Georg Schade.
Stabile Zelllinien etabliert von Dr. Patrick Heun (CID-GFP, H3.2-GFP, H3.3-GFP, MPI Freiburg), Georg Schade (CG2051-GFP, CG14480-GFP, CG11076-GFP, CG1091-GFP und CG3548-GFP), Weiguo Zhang (SNAP-CID/CID-GFP, UC Berkeley/Gary Karpen Labor).
CellProfiler-Pipeline entwickelt von Dr. Andreas Thomae (LMU München).

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde.

München, den _____ _____

Teresa Barth

Erklärung

Hiermit erkläre ich, dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, den _____ _____

Teresa Barth

TABLE OF CONTENTS

SUMMARY

It is vital for organisms to undergo cell division in order to survive, grow and produce offspring. Errors in this process can lead to uncontrolled cell division and diseases like cancer. During each cell cycle, duplicated DNA has to be equally distributed to the prospective daughter cells. The proteinaceous kinetochore structure is crucial for the attachment of microtubules from opposite spindle poles and thereby correct partitioning of the sister chromatids. Centromeres are epigenetically defined chromosomal regions which serve as platforms for the assembly of the kinetochore during mitosis.

In humans, many efforts have been made to identify proteins present in centromeres and kinetochores. Also *Drosophila melanogaster* has proven its value in unraveling functional aspects of centromere biology. In this model organism, insights about centromeres are more readily obtained and can frequently be transferred to higher organisms. Still, only a very limited number of centromeric proteins is known in *Drosophila*, potentially owing to the lack of sequence conservation among functional orthologs.

This study set out to identify novel proteins present at *Drosophila* centromeres. For this purpose, a constitutive centromere-bound protein, the centromeric histone H3 variant CID (centromere identifier) fused to a GFP-tag, was used as bait for affinity purification followed by mass spectrometry analysis. Thereby, 94 potentially centromeric proteins were identified that co-purified with GFP-tagged CID. Subsequent immunolocalization studies of 32 factors revealed centromeric localizations for nine factors and fourteen additional proteins localizing to the nucleus. The depletion of three factors, namely CG2051, a histone acetyltransferase, CG14480 and Hyd, in *Drosophila* cells by RNAi led to elevated mitotic errors. Unlike the known centromere proteins CENP-C and Cal1, neither of the investigated factors influenced centromere loading of CID as assessed by quantitative microscopy. On the contrary, depletion of three factors, CG6227, REG and Hyd, caused elevated levels of centromeric CID-GFP, pointing to an involvement in inhibition of CID deposition. This, together with the finding that CID mono-ubiquitination was lost when *hyd* amounts were reduced, suggested that Hyd, a putative E3 ubiquitin ligase, might be implicated in control of CID stability. Cell cycle analysis indicated that reduction of known centromeric factors like CENP-C, Cal1 and CID arrested cells in S-phase while none of the other tested factors led to severe alterations. In sum, this study provides a comprehensive analysis of *Drosophila melanogaster* centromeric protein composition.

ZUSAMMENFASSUNG

Zellteilung ist ein essentieller Prozess für Organismen, um überleben, wachsen und sich fortzupflanzen zu können. Fehler, die dabei auftreten, können zu unkontrollierter Zellteilung und somit Krankheiten wie Krebs führen. In jedem Zellzyklus muss die duplizierte DNA gleichmäßig auf die künftigen Tochterzellen verteilt werden. Die proteinhaltige Kinetochorstruktur ist wichtig für die korrekte Aufteilung der Schwesterchromatiden, da Mikrotubuli von den gegenüberliegenden Spindelpolen dort binden. Zentromere sind epigenetisch definierte Chromosomenregionen, die während der Mitose die Basis für den Kinetochoraufbau bilden.

In humanen Zellen wurden diverse Studien durchgeführt, um Zentromer- und Kinetochorproteine zu identifizieren. Auch *Drosophila melanogaster* hat sich als gutes Modellsystem bestätigt, um funktionelle Prozesse der Zentromerbiologie zu entdecken. Erkenntnisse durch Genprodukt-reduzierungen können hier einfacher gewonnen und oft auch auf höher entwickelte Organismen übertragen werden. In *Drosophila* ist jedoch nur eine sehr begrenzte Anzahl an Zentromerproteinen bekannt, vermutlich wegen einer fehlenden Konservierung der Orthologen.

In dieser Studie sollten neue Zentromerproteine von *Drosophila* identifiziert werden. Zu diesem Zweck wurde eine Affinitätsaufreinigung der GFP-getaggten zentromerischen Histon H3-Variante CID (centromere identifier) mit anschließender massenspektrometrischer Analyse durchgeführt. Dabei wurden 94 Proteine im Komplex mit CID-GFP gefunden. Anschließend wurden 32 dieser Faktoren mittels Immunfluoreszenz auf ihre Zelllokalisation überprüft und dabei neun zentromerische Lokalisationen sowie 14 kernlokalisierende Faktoren gefunden. Bei Depletion dreier Faktoren, nämlich CG2051, einer Histonacetyltransferase, CG14480 und Hyd, durch RNA-Interferenz reicherten sich mitotische Fehler in Zellen an. Quantitative Mikroskopie zeigte, dass Reduktion der Zentromerproteine CENP-C und Cal1 zu vermindertem CID-GFP-Einbau in Zentromeren führt. Kein anderer Faktor zeigte diesen Effekt, jedoch resultierte CG6227-, REG- und Hyd-Verminderung in vermehrtem CID-GFP an Zentromeren. Normalerweise sind diese Faktoren also offensichtlich an der Eindämmung von CID-Einbau beteiligt. Reduzierung von *hyd*, einer putativen E3-Ubiquitinligase, führte zu Rückgang von CID-Monoubiquitinierung. Diese Erkenntnisse deuten auf eine Regulation der CID-Stabilität durch Hyd hin. Verringerung von CENP-C, Cal1 und CID führte zu S-Phase-Arrest der Zellen, während bei keinem anderen getesteten Kandidaten eine Zellzyklusbeeinflussung beobachtet wurde. Zusammengenommen bietet diese Studie eine umfassende Analyse des Proteinaufbaus von *Drosophila melanogaster*-Zentromeren.

DANKSAGUNG

Zuallererst möchte ich Prof. Axel Imhof für die Aufnahme in seine Gruppe für meine Diplomarbeit und anschließend die Doktorarbeit danken. Du hast mich über die Jahre gefördert und mir vieles ermöglicht (UCI, Konferenzen, Vorträge), was nicht selbstverständlich ist und mich als Wissenschaftlerin weitergebracht hat. Danke für deine stets offene Tür und die Vermittlung deiner wissenschaftlichen Interessen!

Außerdem danke ich Prof. Peter Becker dafür, dass ich die letzten Jahre an seinem Institut verbringen konnte. Die Atmosphäre, die du hier geschaffen hast, ist inspirierend, kollegial und freundschaftlich. Ich habe die zahlreichen, qualitativ hochwertigen Seminare im Institut und auch die Retreats, Chromatin Days und BBQs sehr genossen!

Des Weiteren möchte ich den Mitgliedern meines Thesis Advisory Committees, PD Dr. Aloys Schepers, Dr. Ignasi Forné und Prof. Axel Imhof für konstruktive Diskussionen danken und dafür, dass sie sich immer die Zeit für meine Arbeit und meine Anliegen genommen haben.

Für Beteiligung an meiner Forschungsarbeit danke ich vor allem Dr. Andreas Thomae dafür, dass er sich für mich verantwortlich gefühlt und sich stets gekümmert hat. Dein Projektdesign und die Planung im Verlauf waren immer sehr motivierend und hilfreich. Außerdem danke ich Dr. Andreas Schmidt, Irene Vetter und Marc Wirth für Hilfe mit den Daten. Danke auch an Dennis Sadic für Klonierungsberatungen.

Natürlich auch großen Dank an unsere hervorragenden Kollaborationspartner in Freiburg/Edinburgh, Georg Schade und Dr. Patrick Heun. Die Arbeit mit euch war immer sehr aufschlussreich, zielstrebig und unverzüglich. Wir haben von eurem Zentromerwissen sehr profitiert und gemeinsam eine tolle Arbeit geliefert!

Vielen Dank an die ZfP-Truppe für die Analyse unzähliger MS-Proben und schnelle Bearbeitung von "high priority samples" sowie ein stets offenes Ohr für MS-basierte Nachfragen: Pierre Schilcher, Marc Wirth, Dr. Andreas Schmidt und Dr. Ignasi Forné.

Danke an Edith Müller und Caroline Brieger für ständige schnelle und unkomplizierte Unterstützung im Sekretariat.

Danke an Andi, Ignasi und Andreas für das Korrekturlesen dieser Arbeit.

Thanks also to the Elitenetzwerk Bayern and the program "Protein Dynamics in Health and Disease". All the seminars, courses and retreats improved my scientific performance. Since then, I thrive under pressure! I thank Prof. Peter Becker for the possibility to join this excellent program. Thanks to all the members who taught me also other aspects of biology beyond chromatin research as well as for our discussions about science, our future and various mixed topics!

Thanks to all the Imhofs and members of the whole institute, past and present. You made my PhD time fun and I enjoyed all short chats on the corridor, lunch and coffee in the kitchen, spontaneously ordered dinners and one or two shared beers at different times and occasions - especially to my peers and closest colleagues Viola Sansoni, Simone Vollmer, Miriam Pusch, Henrike Klinker, Silvia Dambacher and Gabriele Wagner – for five years, we went through ups and downs together! I thank Irene Vetter, the heart and soul of the lab, for always knowing everything and permanent willingness to help. Thanks to my mentor Annette Scharf for the nice setup in the lab, both scientifically and socially, and continuous career council. A special thanks goes to Andi and Ignasi for investing so much time and nerves into educating my mind.

Zu guter Letzt danke ich meinem Vater für alles, was er mir immer ermöglicht hat und dies weiterhin tut. Deine ständige, bedingungslose Unterstützung hat mir den Weg hierher geebnet!

1 INTRODUCTION

1.1 Cell division and aberrations

More than 100 years ago, Theodor Boveri, a developmental biologist working with sea urchin in Würzburg, observed that chromosomal aneuploidies have deleterious effects to cells (Boveri 1902). These incorrect chromosome numbers in the two generated cells can result from missegregation of the two sister chromatids in mitosis or meiosis. In 1914, he proposed that these abnormalities might lead to establishment of malignant cancers (Boveri 1914). These facts show that the process of cell division in general and chromosome distribution and its participating factors in particular has been intensively studied since decades as it is a crucial process in biology. Since Boveri's first proposal of cancerogenesis due to chromosomal aneuploidies, it has been confirmed in many cases (reviewed in Fang and Zhang 2011). Nevertheless, the molecular mechanisms leading to these aberrations are still not fully understood. Unraveling these functions can lead to hints about how to control the deleterious chromosome missegregations.

During mitotic cell division, one mother cell divides into two daughter cells. With the few exceptions of asymmetric divisions, these daughters are identical to the mother cell. To maintain this identity, the DNA content of the mother cell has to be duplicated before division. The process of cell division is separated into different stages of a so-called cell cycle (figure 1.1). Replication of DNA happens in the synthesis or S-phase of the cell cycle. Before and after DNA synthesis, gap phases are taking place, called G1- and G2-phase. All three phases together are termed interphase. After G2-phase, cells enter mitosis (M-phase), which is divided into five different stages: prophase, prometaphase, metaphase, anaphase and telophase. Cytokinesis follows mitosis, where the cytoplasm is separated. If cells are differentiated and do not undergo division anymore, their status is termed G0.

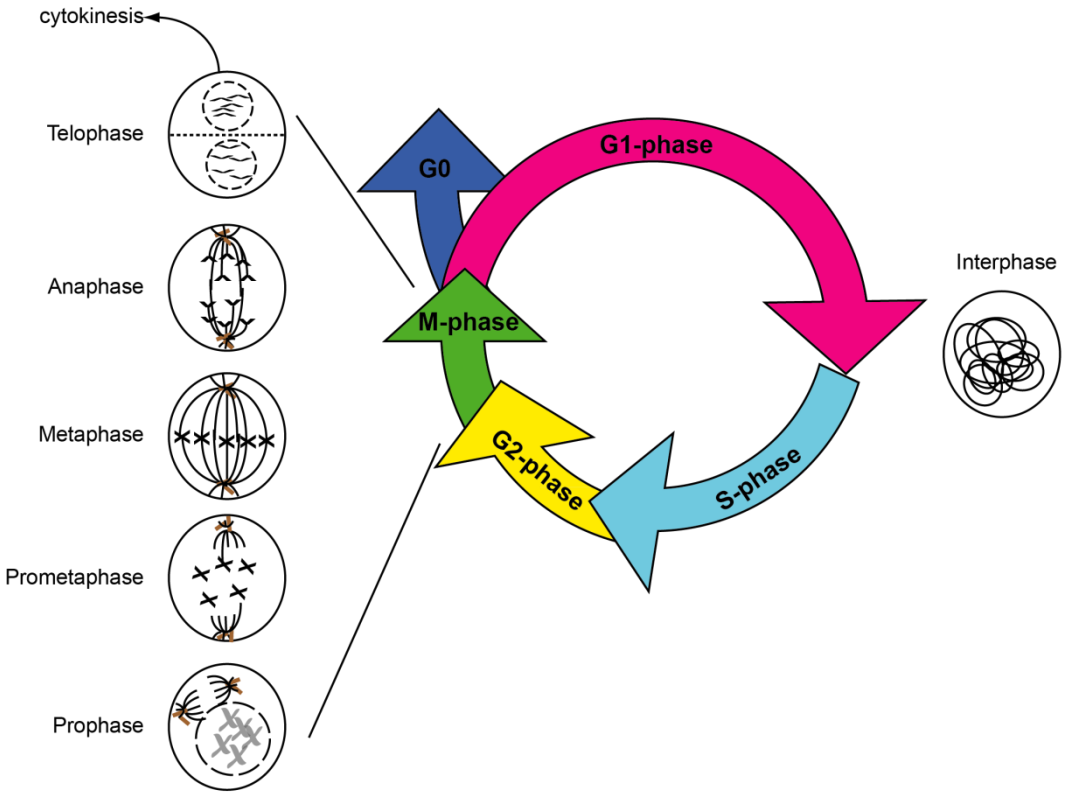

Figure 1.1: Schematic of a cell cycle. Different phases with approximate durations are indicated by arrows. G1/G2/G0-phase: gap phases. S-phase: synthesis phase. M-phase: mitosis. Mitosis is subdivided in five different phases, shown with names and schematics in the figure. Brown: Centrosomes/Microtubule organizing centers. G1, S, and G2-phase together are also termed interphase.

In prophase of mitosis, chromosomes are starting to condensate and the nuclear membrane is dissolving. Centrosomes, which are microtubule organizing centers (MTOCs), are duplicated and move to opposite sides of the cell where the future spindle poles will be located. In prometaphase, microtubules (MTs) emerging from the centrosomes are either attaching to the kinetochore part (see chapter 1.2) of chromosomes (kinetochore microtubules), to a MT from the opposite pole (polar microtubules) or helping to anchor the MTOCs to the cell membrane (astral microtubules). In metaphase, condensed chromosomes are aligned in the metaphase plate such that the two kinetochores face the two opposite spindle poles. In anaphase, the two sister chromatids of one chromosome are actually getting separated and torn towards opposite spindle poles. This process is assisted by polar microtubules that are pushing the spindle poles further apart and kinesins at the kinetochore that help move the chromosomes towards the spindle pole along microtubules. In telophase, the nuclear membrane is forming again and the cleavage furrow between the prospective daughter

cells is established. The cytoplasm is separated to two distinct cells by the process of cytokinesis, thus ending cell division.

As mentioned, cell division is a critical process in cells and has to be well controlled. Therefore, different checkpoints exist during the cell cycle that can only be passed if all prerequisites for continuation are fulfilled. There are checkpoints at G1/S-phase transition, intra S-phase, at G2/M-transition and in M-phase of the cell cycle. If DNA is damaged, cells will arrest in late G1-, S- or G2-phase to allow for repair. The checkpoint in mitosis is also called spindle assembly checkpoint (SAC). Proteins of the SAC make sure that sister chromatids are only separated when all kinetochores are bound by microtubules, thereby preventing chromosomal aberrations (reviewed in Musacchio and Salmon 2007). Constituents of the kinetochore like PLK (Polo-like kinase) and Zwilch are also members of the spindle assembly checkpoint.

Mitosis is a cell cycle phase regulated by posttranslational modifications of proteins: for example, (de)phosphorylation of cyclin dependent kinases (CDKs) and cyclins. The CDK1/Cyclin B complex phosphorylates Aurora B kinase, Polo-like kinase (PLK) and APC/C (anaphase promoting complex/cyclosome) during mitosis. Once phosphorylated, APC/C associates with Cdc20 to become an active E3 ubiquitin ligase, which targets its substrates for proteasome-mediated degradation. In this way, the APC/C ubiquitinates Cyclin B after the spindle assembly checkpoint is passed and thus indirectly terminates the mitotic phosphorylations. Other APC/C dependent ubiquitination events together with PLK-dependent phosphorylations are required to inactivate Securin and Shugoshin. Thereby, Separase becomes active to cleave Cohesin, which triggers chromatid separation (reviewed in Kim and Yu 2011, Jeong and Yang 2013). Additionally, acetylation of Cyclin A has been determined as a prerequisite for its ubiquitination by the APC/C complex and its subsequent degradation by the proteasome (Mateo, Vidal-Laliena et al. 2010), adding acetylations to the list of posttranslational modifications implicated in control of mitotic progression.

Since each cell originates from another cell (*"omnis cellula e cellula",* Rudolf Virchow), cell division is a process important for an organism to stay alive, reproduce, grow, specialize and inherit traits. Each cell division is prone to errors. These can appear on different layers of cell division: either during DNA replication, which is a tightly controlled mechanism, or when chromatids of chromosomes are separated to the two cell poles. Errors in the latter process result in chromosomal aberrations which can cause diseases like cancer if cells start to proliferate in an unregulated manner. Therefore, detailed

understanding of cell division processes is crucial and prominent players in cell division are centromeres and kinetochores.

1.2 Centromeres and kinetochores

Centromeres have different appearances. In the most abundant one, when chromosomes are metacentric and monocentric, they can be seen as a primary constriction in the middle section of condensed mitotic chromosomes, making condensed chromosomes adapt an X-shaped structure. Centromeres of monocentric chromosomes can be located at different positions on the chromosome: if they are placed more towards one end of a chromatid, they are termed acrocentric; if they are placed right at the end, chromosomes are telocentric. The model organism used in this work, the fruit fly *Drosophila melanogaster*, possesses four monocentric chromosomes with the Y, 2nd and 3rd chromosome being metacentric and the X and 4th chromosome telocentric.

Three different kinds of centromeres are described. The chromosomes of the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) have point centromeres, which are constituted by one positioned nucleosome spanning around 120 bp of centromere-identity conferring DNA (Furuyama and Biggins 2007). Most other organisms harbor regional centromeres that comprise larger areas of chromatin. Regional centromeres are not very similar in length of underlying DNA: it can reach from 40 to 100 kb in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) to around 400 kb in *Drosophila melanogaster* or 10 Mb in humans (reviewed in Cleveland, Mao et al. 2003). A third type of centromeres can be found on holocentric chromosomes detected in a wide range of organisms found throughout the animal and plant kingdom (Melters, Paliulis et al. 2012) where centromeres are spread over the whole chromosome.

Centromeres build the platform for assembly of kinetochores, a proteinaceous structure only present throughout mitosis. Kinetochores are the site on chromosomes where microtubules attach during mitosis. Electron microscopy analysis and molecular studies revealed that kinetochores are made up of four different protein layers on both sister chromatids (figure 1.2). First, and innermost, there are proteins of the inner centromere, also termed the chromosomal passenger complex (CPC), colored in orange in figure 1.2. In mammalian cells, the CPC is comprised of the Aurora B kinase, INCENP, Survivin and Borealin. The second layer is represented by centromeric chromatin. It typically consists of the core histones H3, H4, H2A and H2B (see chapter 1.3), interspersed by nucleosomes containing the centromere-specific

histone H3 variant CENP-A (centromeric protein A) (Blower, Sullivan et al. 2002). Proteins of the centromeric chromatin are colored in violet in figure 1.2. The third layer is formed by proteins of the so-called constitutive centromere-associated network (CCAN) or the inner kinetochore. In human, 16 different components of the CCAN are known that are also termed CENPs (summarized in Westermann and Schleiffer 2013). They are marked in bluish colors in figure 1.2. Forth, the outermost structure is composed by proteins of the outer kinetochore or KMN-network. KMN stands for Knl1 complex, Mis12 complex and Ndc80 complex (reviewed in Varma and Salmon 2012). The different components are depicted in figure 1.2 in red/pink. Mostly the components of the Ndc80 complex interact with kinetochore microtubules.

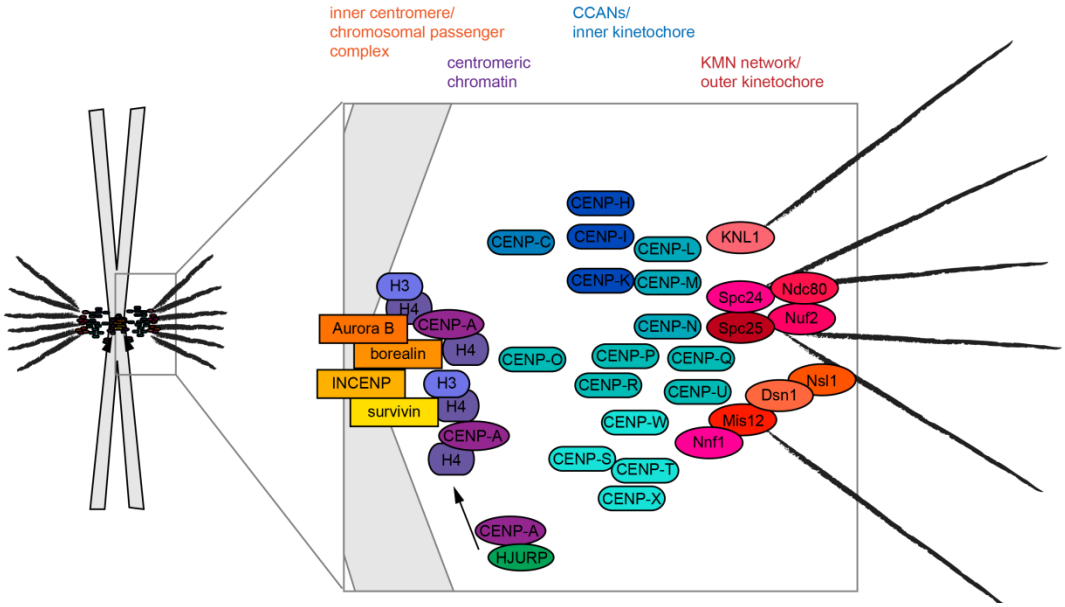

Figure 1.2: Four layers of a human kinetochore. On the left, a schematic mono- and metacentric chromosome is depicted with kinetochores on both sister chromatids. On the right, an amplification of the kinetochore layers can be seen. Different layers with their components are indicated in different colors and the names of the layers mentioned on top. Black lines indicate kinetochore microtubules attaching to kinetochores in mitosis.

The four layers of the kinetochore are differently regulated during the cell cycle. The proteins of the chromosomal passenger complex and the outer kinetochore are only present at centromere regions during M-phase of the cell cycle. Constitutive centromere-associated network proteins as well as centromeric chromatin stay bound throughout the whole cell cycle. Since during cell division, histone content needs to be doubled for the two prospective daughter cells, synthesis of the majority of histone H3 and H4 is tightly regulated and coordinated with S-phase, when DNA is replicated (Heintz, Sive et al. 1983). During centromere replication, CENP-A is evenly

distributed to the two new daughter strands of DNA. It is, however, only in the following G1-phase, after cells already divided, that the CENP-A pools are filled up to the amount before cell division (Jansen, Black et al. 2007). Also, a CENP-A specific chaperone termed HJURP (Holliday junction recognition protein) is required for deposition (Foltz, Jansen et al. 2009).

CENP-A has been discovered by staining cells with an antibody isolated from CREST (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias) syndrome patients (Earnshaw and Rothfield 1985). The obtained signal was centromeric, thus the stained proteins were termed as CENPs A, B and C. Later it was found that CENP-A is part of nucleosomes, thereby a histone and a centromere-specific variant of histone H3 (Palmer, O'Day et al. 1987). Interestingly, this histone H3 variant specific for centromeres is present in all eukaryotes investigated so far (Henikoff and Dalal 2005). In this regard, CENP-A is an exception since CCAN proteins show very limited conservation compared to proteins of the chromosomal passenger complex or the outer kinetochore that are only temporarily bound to kinetochores during mitosis. Furthermore, the underlying DNA of centromeres is hardly conserved, sometimes even differing between individual chromosomes.

These findings have led to the suggestion that centromere identity and function are not determined by DNA, but by protein composition. This is further underlined by experiments demonstrating that CENP-A (or centromere identifier (CID) in *Drosophila*) incorporation at ectopic sites creates fully functional centromeres onto which kinetochores assemble (Olszak, van Essen et al. 2011). These neocentromere kinetochores preferably form at heterochromatin borders, mirroring the situation of endogenous centromeres that are usually surrounded by pericentromeric heterochromatin (more on heterochromatin in chapter 1.3). Furthermore, if CID is targeted to a non-centromeric chromatin site, it is both necessary and sufficient for centromere and kinetochore establishment and maintenance over many cell divisions (Mendiburo, Padeken et al. 2011). Neocentromere formation proofs that there is no dependency on underlying DNA sequence for establishment of functioning centromeres and kinetochores (reviewed in Burrack and Berman 2012). Rather, it is believed that chromatin factors are responsible.

1.3 Chromatin and epigenetics

Chromatin describes the entity of DNA and associated factors like proteins or RNAs. DNA in a nucleus is always tightly associated with proteins. The most

abundant chromatin proteins are histone proteins. So-called "core" histones H2A, H2B, H3 and H4 are organized in octamers with two copies of each histone. When DNA is wrapped around this histone octamer 1.65 times (146 bp of DNA) (Luger, Mader et al. 1997), the resulting structure is called a nucleosome. Nucleosomes make up the first level of chromatin compaction and organization. A DNA fiber packed in nucleosomes forms a "beads-on-a-string"-like structure with a diameter of 10 nm *in vitro* (Olins and Olins 1974). Another histone, histone H1, termed a "linker" histone since it binds DNA in between two nucleosomes (linker DNA), helps to condense chromatin to higher structures (reviewed in Happel and Doenecke 2009). In order to obtain condensed mitotic chromosomes, more levels of compaction are applied whose precise molecular appearance is still under debate (reviewed in Hansen 2012).

Thus, chromatin helps packing DNA during interphase as well as in condensed metaphase chromosomes, but is also an important mediator of gene regulation. Almost each cell in a multicellular organism bears the same DNA, but their appearance is quite different, suggesting that genes are differentially regulated in diverse cell types. Already in 1928, two distinct chromatin states were described based on microscopic observations (Heitz 1928): the less tightly packed euchromatin and more densely packed heterochromatin, while heterochromatin is subgrouped in constitutive and facultative heterochromatin. In general, DNA located in euchromatin is more easily accessible for regulatory factors like transcription factors or RNA polymerases and thus more active than the one in heterochromatin.

Regulation of gene activity can be controlled by different so-called "epigenetic" mechanisms, with "epigenetic" describing all aspects of gene regulation that are not encoded in the DNA sequence but are inherited to the next generation during mitosis or meiosis. Figure 1.3 gives an overview about different epigenetic mechanisms.

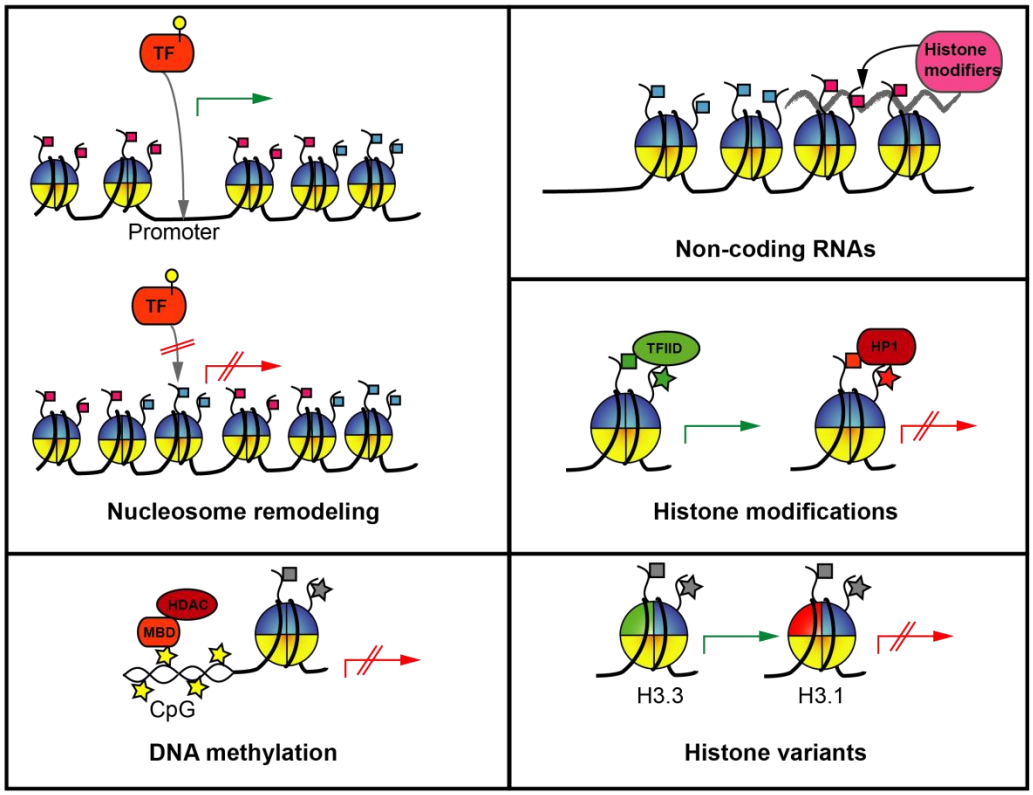

Figure 1.3: "Epigenetic" mechanisms depicted schematically. Blue/yellow: Histone octamer. TF: transcription factor. MBD: Methyl binding domain protein. HDAC: Histone deacetylase. Green arrow: Active gene transcription. Red arrow: Inactive gene. Grey line: RNA. TFIID: Transcription factor IID. HP1: Heterochromatin protein 1. Yellow stars: DNA methylation. Stars and rectangles: histone modifications. Modified after Hahn, Dambacher et al. 2010.

Multiple regulatory factors were shown to mediate epigenetic processes. First, the position of nucleosomes can be shifted on DNA by nucleosome remodelers, which are often ATP-dependent (reviewed in Clapier and Cairns 2009), making DNA more or less accessible for chromatin proteins. Second, DNA can be methylated on cytosines by DNA methyltransferases, leading to 5-methylcytosine (5-mC), a "fifth DNA base". Methylated cytosines in CpG-motifs located within promoters can cause binding of repressor proteins and thereby lead to inactivation of the corresponding gene. Additionally, 5-hydroxymethylcytosine is generated from 5-mC, which can be converted to 5-formylcytosine and 5-carboxycytosine, by action of ten-eleven-translocation (TET) enzymes (reviewed in Delatte and Fuks 2013). Third, non-coding RNAs (ncRNAs) are implicated for example in inactivation of the mammalian female X-chromosome by recruiting the histone modifying complex PRC2 (Zhao, Sun et al. 2008). Forth, histone proteins carry a multitude of different posttranslational modifications (PTMs), especially the N-terminal tails of histone H3 and H4 (see figure 1.4). These modifications, mainly acetylations, methylations, phosphorylations and ubiquitinations, are set by so-called

"writer" proteins, removed by "erasers" and recognized by "readers". Histone PTMs translate into different activity states: on the one hand, they activate the underlying gene, like in general acetylations or trimethylation of histone 3 on lysine 4 (H3K4me3) present on promoters, by recruiting RNA polymerase II via transcription factor IID (TFIID) (Vermeulen, Mulder et al. 2007). On the other hand, they can repress gene activity for example by helping to enhance chromatin compaction through the recruitment of repressors like the polycomb repressive complexes PRC1 and PRC2. The latter mono-, di- and trimethylates histone 3 on lysine 27 (H3K27me1/2/3) by its subunits Ezh1 or Ezh2 (enhancer of zeste), and this mark can recruit repressive readers (reviewed in Simon and Kingston 2009). The existence of a "histone code" has been postulated, implying that combinations of various histone modifications lead to a certain regulatory outcome, rather than single modifications (Strahl and Allis 2000). Fifth, different chromatin compartments can contain distinct, specialized histone variants, a fact made use of in this work (chapter 1.6). Mammalian histone H3 has four main variants, H3.1, H3.2, H3.3 and CENP-A. The two replication-dependent variants H3.1 and H3.2 are only incorporated in chromatin during S-phase of the cell cycle when the demand for histone proteins is high. These canonical variants differ in only one amino acid. The replication-independent variant called replacement variant H3.3 is incorporated by a different mechanism throughout the cell cycle and differs by five amino acids to H3.1. In contrast, the sequence of the centromeric variant CENP-A substantially differs from the other H3 variants (reviewed in Biterge and Schneider 2014). Histone variants can alter the features of chromatin domains where they are incorporated, for example by carrying a distinct pattern of histone PTMs, destabilization of the nucleosome or by engaging in unique protein-protein interactions.

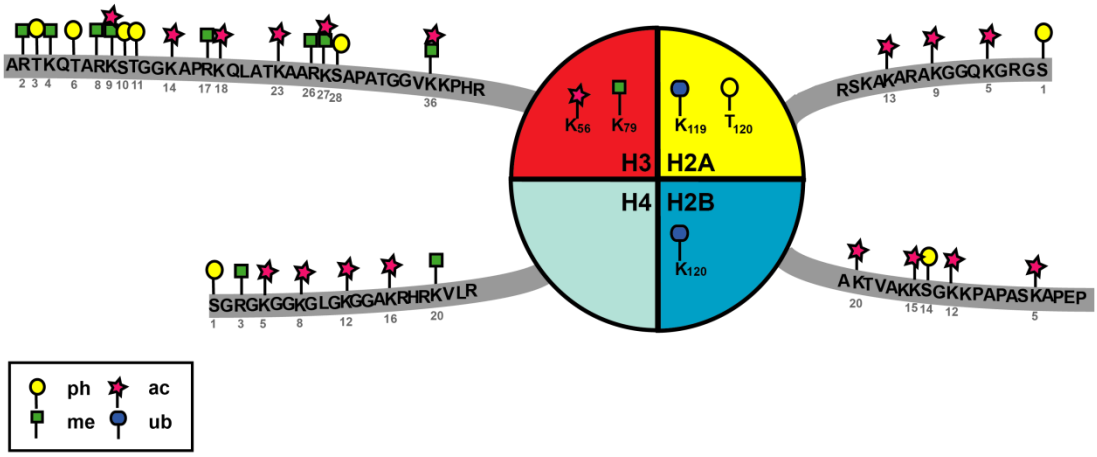

Figure 1.4: Schematic of abundant posttranslational modifications on histone tails. Not included are modifications like crotonylation, proline isomerization, formylation, hydroxylation, deamination, ADP-ribosylation, β-N- acetylglucosamine, sumoylation.

As introduced before, also centromeres and kinetochores execute their function and inheritance mostly by epigenetic mechanisms that are not encoded by the underlying DNA (with the exception of point centromeres). Centromeric chromatin has a distinct composition and is embedded in pericentromeric heterochromatin. Centromeres are in fact marked by a special pattern of histone modifications with high levels of the activating H3K4me2 mark and a low level of acetylated H3 and H4, which is usually found at repressed chromatin. Pericentromeric heterochromatin surrounding centromeric chromatin is characterized by repressive H3K9me2/3 and hypoacetylated H3 and H4 (reviewed in Dunleavy, Pidoux et al. 2005). Recently, it has been shown that H4K20me1 is present in the CENP-A containing nucleosome in chicken and human cells and that it is necessary for kinetochore assembly (Hori, Shang et al. 2014).

In this work, the centromeric histone variant CENP-A (or CID in *Drosophila*) was used as a bait to identify centromeric proteins of *Drosophila melanogaster* (see chapter 1.6). Therefore, the current status of research on *Drosophila* centromere proteins is now described.

1.4 Centromeres in the model organism *Drosophila melanogaster*

Centromeric DNA in different organisms is not well conserved and fast evolving (Henikoff, Ahmad et al. 2001). Therefore, also centromeric proteins differ substantially. Proteins of the chromosomal passenger complex as well as the outer kinetochore, which mediates binding to microtubules, however, are conserved throughout evolution. The H3 variant CENP-A and proteins of the constitutive centromere-associated network (CCAN) in contrast harbor

considerable differences between organisms that complicate sequence-homology based identification of the orthologs. Centromeric proteins known to date in *Drosophila* are hence limited to the more conserved proteins of the inner centromere (CPC) and outer kinetochore (KMN network). However, only one CCAN protein, CENP-C, was identified (table 1.1).

Table 1.1: Centromere and kinetochore proteins identified in different species. Names of the histone H3 variant are highlighted in red. Proteins from humans and *Drosophila* are shaded in blue for better comparison. Modified after Perpelescu and Fukagawa 2011.

Homo sapiens	***Gallus gallus***	***Xenopus laevis***	***C. elegans***	***Drosophila melanogaster***	***S. pombe***	***S. cerevisiae***
CENP-A	CENP-A	CenH3	HCP3	CID	Cnp1	Cse4
CENP-B					Abp1/Cbh1/ Cbh2	
CENP-C	CENP-C	CENP-C	HCP4	CENP-C	Cnp3	Mif2
CENP-E	CENP-E	CENP-E		CENP-meta/ CENP-ana		Mcm16
CENP-F		CENP-F	HCP1/2			
CENP-I	CENP-I				Mis6	Ctf3
CENP-K						
CENP-L	CENP-L				Fta1	
CENP-M	CENP-M				Mis17	Iml3
CENP-N	CENP-N				Mis15	Chl4
CENP-O	CENP-O				Mal2	Mcm21
CENP-P	CENP-P				Fta2	Ctf19
CENP-Q	CENP-Q				Fta7	
CENP-R	CENP-R					
CENP-S	CENP-S	Apitd1			SPBC2D10 .16	YOL86-A
CENP-T	CENP-T	CENP-T			SPBC800/ Cnp20	
CENP-U	CENP-50				Fta4	
CENP-W	CENP-W					
CENP-X	CENP-X					
CENP-Y						
HJURP		HJURP			Scm3	Scm3
Mis18α	Mis18α				Mis18	
Mis18β	Mis18β				Mis18	
M18BP1	Knl2		knl-2			
RbAp48		rbbp4-a	rba-1	RbAp48 = CAF-1	Mis16	Msi1

Homo sapiens	***Gallus gallus***	***Xenopus laevis***	***C. elegans***	***Drosophila melanogaster***	***S. pombe***	***S. cerevisiae***
RbAp46		Rbbp7			Mis16	Msi1
hMis12	Mis12		MIS-12	CG18156	Mis12	Mtw1
DSN1	Dsn1		KNL3		Dsn1/Mis13	Dsn1
NNF1	Nnf1		KBP-1	CG13434/ CG31658	Nnf1	Nnf1
NSL1	Nsl1		KBP-2	CG1558	Nsl1/Mis14	Nsl1
NDC80	Ndc80/ Hec1		NDC-80	CG9938-PA	Ndc80	Ndc80
NUF2	Nuf2		HIM-10	CG8902	Nuf2	Nuf2
SPC24	Spc24	spc24	KBP-4		Spc24	Spc24
SPC25	Spc25	spc25	KBP-3	CG7242	Spc25	Spc25
KNL1	Knl1		KNL1	CG11451	Spc7	Spc105

Recently, the homolog of the CENP-A chaperone HJURP in *Drosophila melanogaster* was identified to be chromosome alignment defect 1 (Cal1) (Chen, Dechassa et al. 2014). If targeted to a non-centromeric chromosomal region, Cal1 can recruit CID and lead to establishment of functioning kinetochores. These kinetochores can also be inherited. Furthermore, as mentioned earlier, the centromeric H3 variant is termed CID in *Drosophila melanogaster*, which stands for centromere identifier. It only bears around 50 % similarity with the human CENP-A. Taken together, the knowledge about *Drosophila* centromere proteins is very limited as compared to human centromeres (figure 1.5).

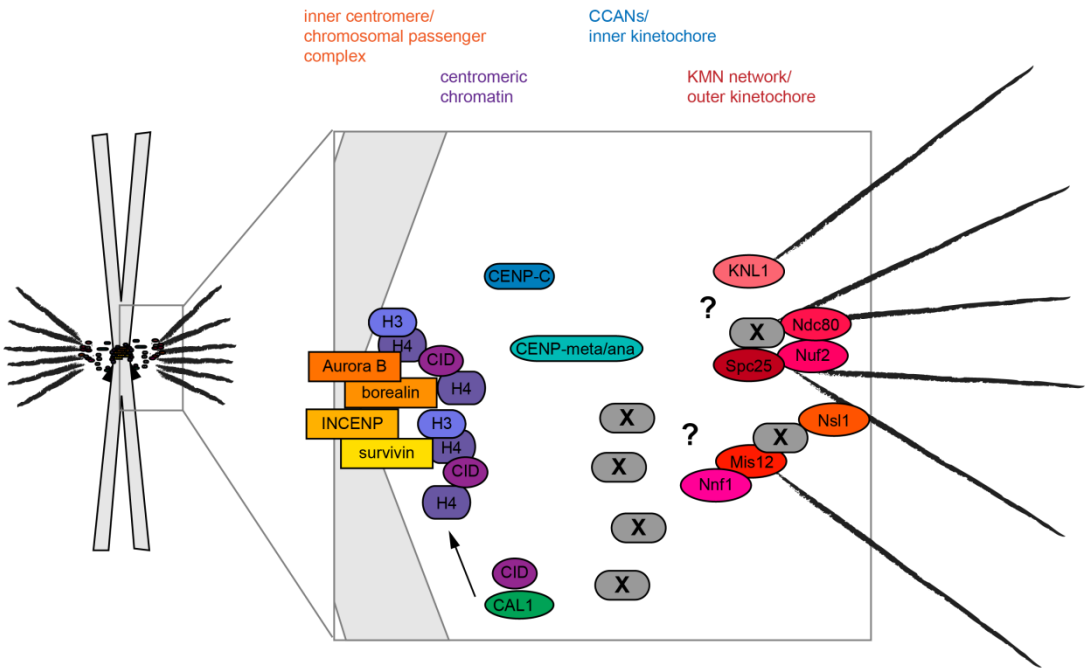

Figure 1.5: Schematic of centromere and kinetochore proteins in *Drosophila melanogaster*. On the left, a schematic mono- and metacentric chromosome is depicted with kinetochores on both sister chromatids. On the right, an amplification of the kinetochore layers can be seen. Different layers with their components are indicated in different colors and the names of the layers mentioned on top. Black lines indicate kinetochore microtubules attaching to kinetochores in mitosis.

Consistent with its heterochromatic environment, *Drosophila* centromeric DNA is replicated in late S-phase (Sullivan and Karpen 2001). However, newly synthesized CID is only incorporated in the G1-phase of the cell cycle (Lidsky, Sprenger et al. 2013). CENP-C and Cal1 are two proteins required for CID deposition (Erhardt, Mellone et al. 2008). Cal1 is binding to centromeres before newly synthesized CID is incorporated (Mellone, Grive et al. 2011). Although new CID is transiently accumulating at centromeres during M-phase, the full levels of centromeric CID present before cell division are only restored in G1-phase. These findings indicate that reconstitution of centromeric chromatin after replication is more complicated than presumed.

Another interesting finding in *Drosophila* centromere biology was a model for regulation of centromere clustering (Padeken, Mendiburo et al. 2013). Centromere clustering is a phenomenon observed in various cell types: in theory, each chromosome contains one centromere or even two when its DNA has been duplicated during S-phase. As the Schneider cell subclone used in the study, L2-4 cells, harbors 13 chromosomes in total (two X, four 2nd, four 3rd, two 4th and one 2L chromosome), 13 or 26 centromeres should

be distinguishable, respectively. However, usually only between 4-6 centromere spots can be detected by microscopy. Centromeres in nuclei are thus clustering together. This clustering involves nucleoplasmin-like protein (NLP), Modulo and CTCF. Modulo is a nucleolar protein and anchors centromere clusters to the nucleolus. The role of this centromere clustering is not exactly clear; however, the study also showed that induction of centromere declustering by knockdown of any of the involved factors destabilizes pericentromeric heterochromatin. This causes re-activation of transposable elements usually silenced by pericentromeric heterochromatin and results in genomic instability such as DNA double-strand breaks, lagging chromosomes and anaphase bridges. Hence, centromere clustering seems to play an important role in chromatin regulation.

Histone modifications on centromeric chromatin have been determined for *Drosophila* (Sullivan and Karpen 2004). For this, chromosomes were spread as fibers and incubated with modification-specific antibodies. It was found that, similar to human centromeres, CID-containing regions were enriched for H3K4me2 and some H3K9me2, but no acetylations on H3 or H4 or H3K4me3/H3K9me3 were detected. Thus, also in *Drosophila*, centromeric chromatin seems to be marked by both active and repressive histone modifications.

1.5 Previous analyses on chromatin composition of centromeres

The huge number of centromeric proteins known in humans partly derives from one affinity purification – mass spectrometry (AP-MS) study (Foltz, Jansen et al. 2006). The study made use of a TAP-tagged CENP-A expressing HeLa cell line on which purifications and subsequent identification of the enriched proteins by mass spectrometry analysis was performed. By that means, ten proteins could be newly designated to centromeres, termed CENP-K to CENP-T. Additionally, another AP-MS study, performed with an antibody directed against CENP-A in human HeLa cells, identified 34 proteins specifically associating with CENP-A (Obuse, Yang et al. 2004).

In *Drosophila*, a screen was performed to determine factors involved in the establishment of the mitotic spindle (Goshima, Wollman et al. 2007). In this genome-wide high-throughput RNAi screen, spindle phenotypes were monitored by automated microscopy. Since also chromosome structure and alignment was one of the criteria screened for, centromeric proteins like CID and CENP-C were detected in the screen. Cal1 was identified for the first time here. Also proteins of the outer kinetochore layers resulted in erroneous

phenotypes, leading to first description of the homologs in *Drosophila*. Additionally, TAP-tagged, soluble *Drosophila* CID has been purified in order to identify the specific chaperone on unincorporated CID (Furuyama, Dalal et al. 2006). The resulting complex was very simple: only CID, histone H4 and CAF-1 were determined, suggesting that CAF-1 might be the responsible chaperone for CID/H4-dimers. To date, no proteomic screen for centromeric composition in *Drosophila* is published.

1.6 Aims of this thesis

The fruit fly *Drosophila melanogaster* is a widely used model organism to unravel basic biological principles. The embryonically derived Schneider cell line is easy to manipulate by RNA interference (RNAi) (Elbashir, Harborth et al. 2001), thus enabling efficient screening for phenotypes. Some kinetochore proteins, like Aurora B kinase of the chromosomal passenger complex, have first been identified in *Drosophila* (Glover, Leibowitz et al. 1995). It is now clear that this protein has broad functions also in higher organisms including humans (Hochegger, Hegarat et al. 2013).

Still, since centromeric DNA is fast evolving, members of the constitutive centromere-associated network are not conserved among different organisms. Information about these proteins lacks almost completely in *Drosophila* (chapter 1.4). Therefore, the goal of this work was to increase knowledge about the protein composition of centromeres in the model organism *Drosophila melanogaster*. In order to gain more insight into *Drosophila* centromere biology, the following questions were asked:

1. What is the composition of the proteome present at *Drosophila* centromeres?
2. Do the proteins interacting with CID localize to centromeres? Is their binding regulated in a cell cycle dependent fashion?
3. What are possible functions of proteins present at *Drosophila* centromeres?

To answer the first question, an AP-MS strategy, similar to the work used by Foltz *et al.* that discovered novel CENPs in human cells (Foltz, Jansen et al. 2006), was chosen. For this purpose, CID was tagged with GFP. As introduced in chapter 1.2, this centromeric histone H3 variant is centromere-bound throughout the whole cell cycle and therefore constitutes the ideal bait.

Interactors determined by the AP-MS strategy were then confirmed by an independent assay. Immunofluorescence analysis was selected as it

determines localization of a factor in individual cells. Thereby, the localization can be easily monitored during different stages of the cell cycle. As introduced before (chapter 1.2), proteins belonging to different layers of the kinetochore structure show different binding behaviors during the cell cycle. Assessing if a factor is centromere-bound during interphase or specifically only during mitosis can help to closer identify its functions.

Centromere-associated factors can have several potential functions, of which some should be tested for interesting factors here. First, centromere maintenance could be compromised when the amount of a respective factor is diminished by RNAi. Second, halftime of CID or other centromeric or cell cycle proteins could be impacted. Third, the cell cycle progression could be alternated. Fourth, it can be envisioned that deletion of centromeric proteins causes mitotic errors. Fifth, the establishment of the kinetochore layers during mitosis could be hampered.

Answering these questions should help to make a big step forward to identify and characterize novel centromere-associated proteins in *Drosophila melanogaster* and thereby gain insight into processes that can lead to chromosomal aberrations.

2 RESULTS

The goal of this work was to identify novel centromeric proteins in the model organism *Drosophila melanogaster*. Therefore, an affinity purification – mass spectrometry (AP-MS) strategy was applied (chapter 2.1-2.4). Selected candidates determined by this assay were re-analyzed by immunofluorescence analysis to check for centromeric localizations (chapter 2.5). Furthermore, RNAi-mediated loss-of-function studies were conducted, addressing the potential requirement of centromere-localizing factors for CID incorporation in centromeres (chapter 2.6, 2.7), cell cycle progression (chapter 2.8) and centromeric architecture and function (chapter 2.9, 2.10). By this means, this work offers a comprehensive description of *Drosophila melanogaster* centromere composition and biology.

2.1 Workflow for centromeric protein identification by AP-MS

Figure 2.1 gives an overview of the experimental approach to isolate and identify centromeric proteins in *D. mel*. More detailed descriptions and explanations can be found in the respective chapters.

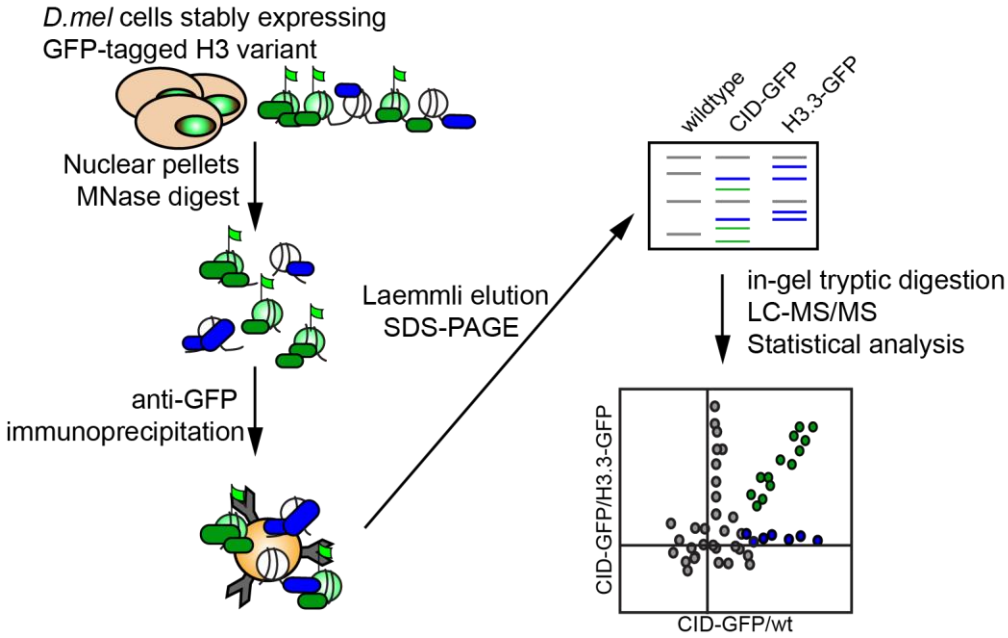

Figure 2.1: Schematic workflow for AP-MS. Cells expressing GFP-tagged histone variants (bright green flag) were used to identify novel *Drosophila* centromeric proteins. Dark green ellipsoids, bands and circles constitute potential centromeric proteins. Blue symbols represent factors generally associated with chromatin as deduced from AP-MS experiments performed on H3.3-GFP expressing cell lines. Factors that are exclusively recovered in chromatin purifications from wildtype cells or that associate with the GFP-affinity matrix irrespective of the source of the soluble chromatin are depicted in grey.

The proteomes of centromeric chromatin versus general chromatin were isolated in order to compare their compositions. For this purpose, *D. mel* Schneider cell lines expressing GFP-fusion proteins of the three *Drosophila* histone H3 variants CID, H3.2 and H3.3, respectively, were generated (chapter 2.2). CID is the centromere-specific H3 variant and its associated proteins can be compared to the ones in purifications of other H3 variants or wildtype chromatin. By that means, specific CID-GFP interacting proteins were assigned. Micrococcal nuclease (MNase) solubilized chromatin from large-scale cultures of the different cell lines was subjected to anti-GFP affinity purification (chapter 2.3). The isolated proteins were size-separated by denaturing gel electrophoresis. Histone H3 and H4 were excised individually and analyzed for posttranslational modifications (PTMs) (chapter 2.4.1). All other proteins contained in one lane were identified and quantified as described in chapter 2.4.2. Statistical analysis was utilized as a means to extract proteins that are most specific to centromeric chromatin (CID-GFP co-purifying factors). This analysis revealed 94 potential centromere associated proteins (chapter 2.4.2). One third of these factors was checked for their cellular localization (chapter 2.5) and some also in other assays linked to centromere biology (chapter 2.6-2.10).

2.2 Characterization of stable cell lines

Before starting immunoprecipitations (IPs) and subsequent MS analysis, single steps of the workflow were set up, tested and optimized, starting with generation of cell lines to perform affinity purifications. To allow for comparable immunoprecipitation conditions, the three different histone variants CID, H3.2 and H3.3 of *D. mel* were fused to the same tag. A GFP tag was used as it provides a good means for IPs by commercially available GFP-specific antibodies immobilized on beads (GFP-Traps). Moreover, this tag enables direct detection by fluorescence microscopy analysis.

The stable cell lines were checked for protein levels of the exogenously expressed, tagged histone variants by Western blot (figure 2.2). CID-GFP levels were below the ones of endogenous CID, minimizing the risk of identifying binding partners of tagged variants as a result of overexpression (compare signal intensities for CID-GFP and endogenous CID in figure 2.2, lane one, lower panel). Because of the incomparable, low H3.2-GFP expression, results with this cell line were eliminated from downstream analysis.

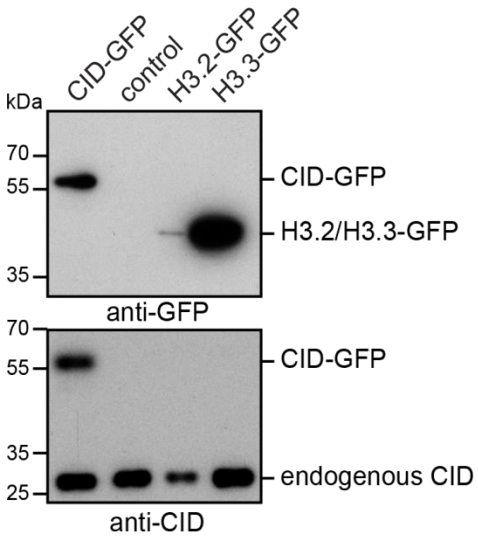

Figure 2.2: Expression levels of tagged histone variants in stable cell lines used in this work. Western blots of MNase solubilized chromatin used for purifications. The appropriate sizes of the tagged or untagged proteins are indicated.

Furthermore, localization of the tagged variants in the stable cell lines was determined by fluorescence microscopy (figure 2.3, 2.4). As expected, CID-GFP signals showed a strong overlap with endogenous CID as determined by CID antibody staining. The CID antibody used here (7A2 subclone) does not detect the GFP-tagged CID protein, but only the endogenously expressed protein (figure 4.3). H3.2-GFP and H3.3-GFP showed the expected homogenous nuclear distribution. Figure 2.3 gives an impression of GFP-tagged variant fluorescence levels, since here the settings are comparable between different panels. Confirming the Western blot results with a GFP antibody (figure 2.2), H3.3-GFP levels were higher than the ones of H3.2-GFP. CID-GFP expression being less than the one of H3.3-GFP mirrors the *in vivo* situation, since CID is only present in the confined region of centromeres whereas H3.3-GFP can be incorporated throughout the genome. The cell lines therefore represent a good experimental system to identify proteins interacting with CID.

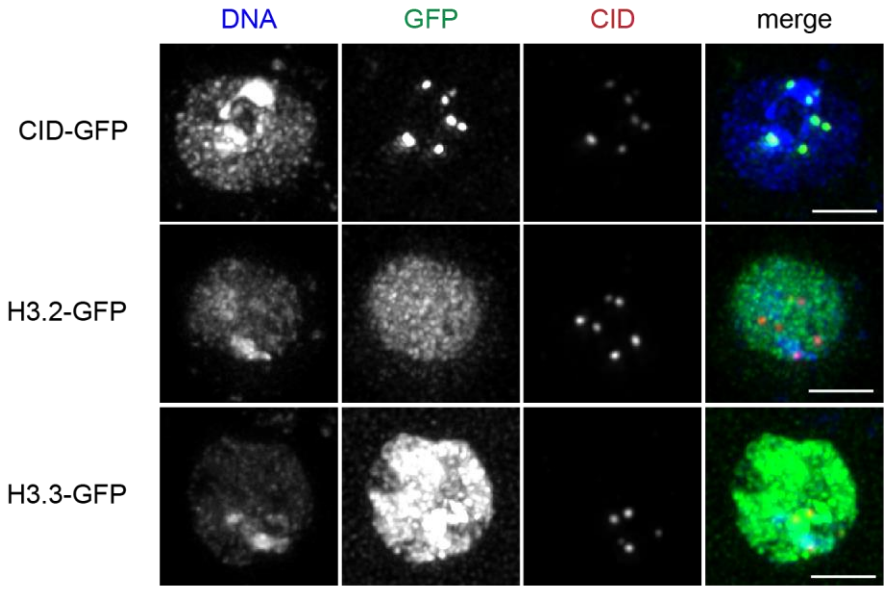

Figure 2.3: Image acquisition with identical hardware and software settings for comparing intensities confirms quantitative differences in GFP-variant expression levels. Scale bar corresponds to 3 µm. Maximum intensity projections of deconvolved images are shown. Blue: DNA stained with DAPI, green: GFP, red: CID.

Since expression levels are so different between the GFP-tagged histone variants, microscopic settings were adapted in order to avoid overexposure like for H3.3-GFP in figure 2.3. Figure 2.4 better demonstrates localization patterns of all variants.

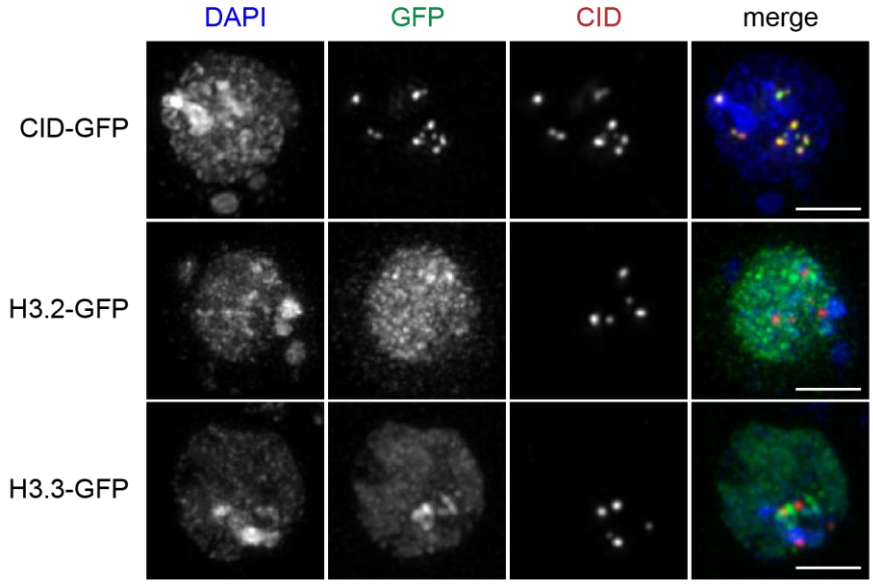

Figure 2.4: Localization of GFP-tagged histone variants in stable cell lines. Scale bar corresponds to 3 µm. Maximum intensity projections of deconvolved images are shown. Blue: DNA stained with DAPI, green: GFP, red: CID.

2.3 Optimization of conditions for immunoprecipitation

After confirming the suitability of the cell lines for affinity purifications, the right ratio of affinity-matrix amount to proteins was determined in small scale experiments (chapter 4.2.4) in order to advance purification conditions and to gain optimal results. Subsequently, the actual immunoprecipitations for MS analysis were upscaled and chromatin prepared from around $5*10^9$ cells. For immunoprecipitation, soluble input chromatin was incubated with the GFP-Trap affinity resin for 2 hours. After extensive washes to remove unspecifically associating proteins, co-precipitated material was eluted from beads by boiling ten minutes in 2x Laemmli buffer. The prolonged heating time as well as the high concentration of SDS (4 % end-concentration) were chosen to ensure a complete protein release from the resin. To evaluate whether the immunoprecipitation reaction successfully enriches for GFP-fusion proteins, a Western blot analysis with anti-GFP antibody was carried out of both MNase solubilized chromatin input material and proteins bound to the affinity matrix. Figure 2.5 shows this analysis with samples from one of the biological replicate.

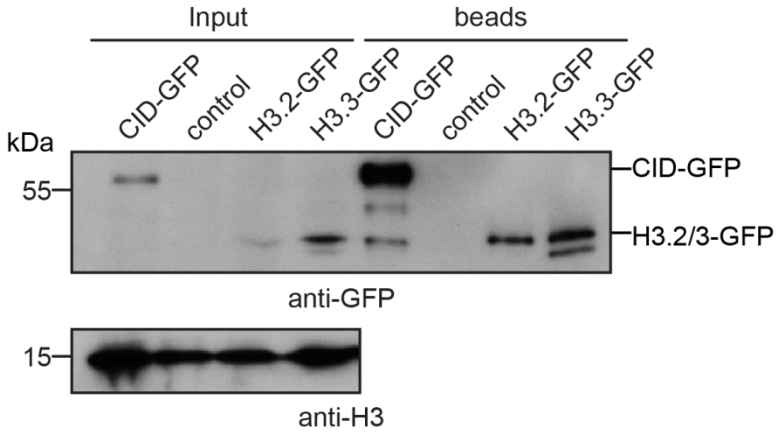

Figure 2.5: Enrichment of GFP-tagged histone variants on GFP-Traps. Western Blots of input fractions before affinity purification and proteins eluted from beads after IP. The appropriate size of tagged variants is indicated. An amount corresponding to 4% (CID), 6% (H3.2) or 8% (H3.3) of the total material used for IP reactions is loaded as input. Samples were taken from the replicate SID1392.

A similar GFP-variant abundance in input material is detected as in figure 2.2 and the Western blot signal intensity of GFP-tagged variants in the eluate fraction (beads) is indicative of a successful affinity chromatography reaction.

For the mass spectrometry analysis of immunoprecipitated material, the eluted proteins were separated on a denaturing 15 % polyacrylamide gel. The gel was then stained with either Coomassie Brilliant Blue or silver nitrate to

check for protein amounts and quality of the affinity purification. Finally, the individual lanes were evenly cut in eight fractions for downstream mass spectrometric analysis. Figure 2.6 represents one example of a silver stained SDS-PAGE from such a purification. The histone proteins (H3, H2A, H2B and H4, highlighted in figure 2.6) are abundant in the immunoprecipitated fractions, indicating that chromatin is enriched on the beads. The GFP-tagged variants, as marked by asterisks, can be identified due to their sample-specific appearance at the expected running height in the gel. The identity of the proteins corresponding to these bands was also confirmed by MS analysis (figure 1 of the appendix for CID, figure 2.7 for H3.2/H3.3).

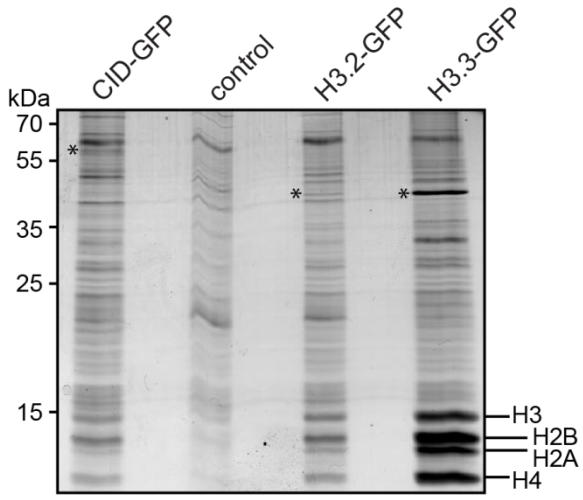

Figure 2.6: Protein gel of proteins eluted from beads in different immunoprecipitations. A representative, silver stained gel of affinity purifications from the different cell lines is shown. Asterisks mark the GFP-tagged variants (confirmed by MS analysis). H3: histone 3, H2B: histone 2B, H2A: histone 2A, H4: histone 4.

The intensities of the histone variants H3.2 and H3.3 present in the band of tagged protein from figure 2.7 can be determined by MS analysis. These two H3 variants only differ in four residues: position 31, 87, 89 and 90. With the treatment applied for analyzing histone modifications by mass spectrometry (explanation in methods chapter 4.2.5.3, chapter 2.4.1), the positions 87, 89 and 90 lie within a peptide ranging from residue 84-116, which is not well detectable by MS analysis. Therefore, the peptide containing amino acids 27 to 40 of histone H3 was used for discrimination of the two isoforms. H3.2 harbors an alanine residue and H3.3 a serine on position 31, differing by 16 atomic mass units (amu). The ionized peptides can thus be distinguished by mass spectrometry due to their different mass-to-charge ratios (m/z ratios). Figure 2.7 shows that, as expected, in the band cut from H3.2-GFP expressing cells, H3.2 is contained and vice versa for H3.3, proving that the proteins in excised bands from figure 2.6 indeed correspond to the H3.2-GFP or H3.3-GFP variants, respectively.

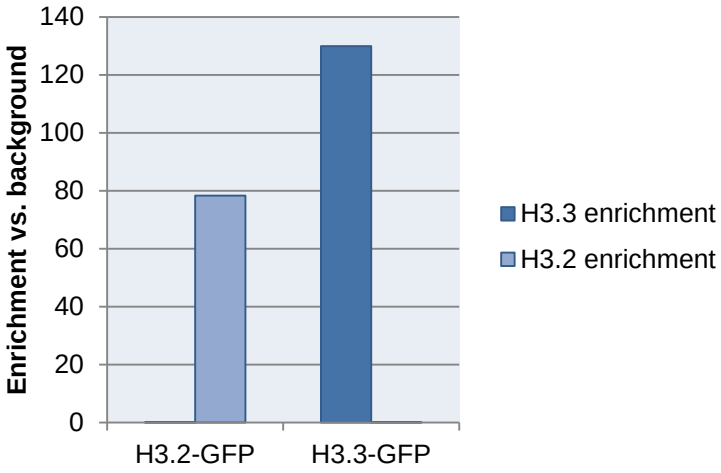

Figure 2.7: Enrichment of histone variant H3.2 and H3.3 in the isolated protein bands marked with asterisks in figure 2.6 as determined by mass spectrometry analysis. The values of unmodified peptide 27-40, which differs in mass for both variants, were taken for calculation.

With the established workflow, it is possible to precipitate chromatin containing different GFP-tagged H3 variants. As the correct GFP-tagged variants are enriched on beads, (figure 2.5, 2.6 and 2.7), MS analysis of the chromatin proteins co-purifying with different variant proteins was carried out.

2.4 Analysis of mass spectrometry data

2.4.1 Histone posttranslational modifications of H3 variant chromatin as determined by LC-MS/MS analysis

Centromeric chromatin consists of CID-containing nucleosomes as CID is the centromeric histone H3 variant. CID and H3 are incorporated as dimers with histone H4 and centromeric chromatin is also interspersed by H3-containing nucleosomes. Earlier studies of histone PTMs using antibody-based detection indicated that histone H3 and H4 are hypoacetylated in centromeric regions (chapter 1.4).

The presence of the characteristic pattern of core histones on protein gels of the different IPs enabled MS-based analysis of histone posttranslational modifications (PTMs) on cutout histone bands (figure 2.6). Histone H4 can be acetylated on four different residues in its N-terminus: lysine 5, 8, 12 and 16 (Sung and Dixon 1970). The treatment for MS analysis generates one peptide ranging from amino acids four to seventeen spanning all four of these lysines. It was therefore interesting to see whether this peptide carries different modification patterns in purifications associated with the respective H3 variants.

Histones co-purifying with chromatin from each IP were excised from the SDS-polyacrylamide gel and analyzed individually with a different protocol

specific for histone PTM mass spectrometry analysis (chapter 4.2.5.3). Figure 2.8 shows the analysis of a histone H4 peptide (residues 4 to 17) containing four lysines that can be acetylated: lysine 5, 8, 12 and 16. For display, values from all five different modification states (completely unmodified peptide or peptide carrying between 1 and 4 acetylations) were added up to 100 % and percentages of the respective modification in either of the affinity purifications are depicted.

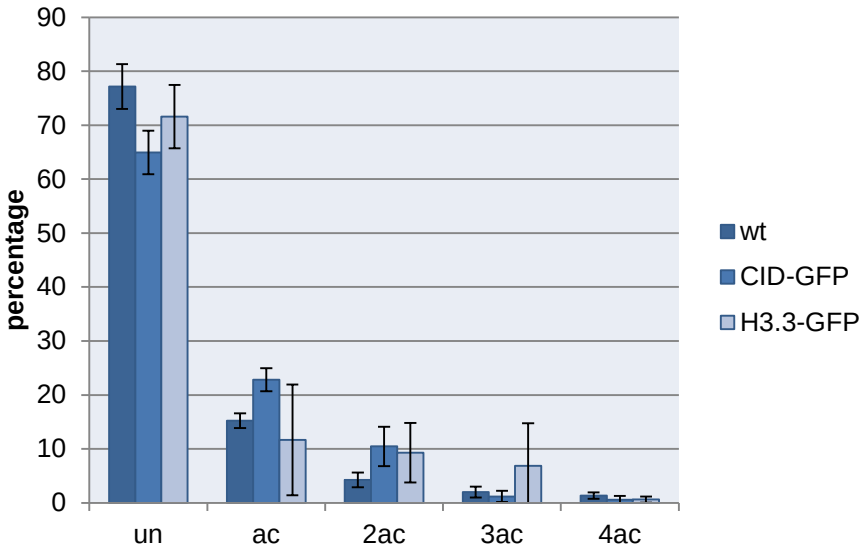

Figure 2.8: Modification status of histone H4 peptide 4-17 co-purified in different affinity purifications. Percentages of the occurrence of different modifications are displayed. Un: unmodified peptide, ac: peptide carrying one acetylation, 2ac: peptide carrying two acetylations, 3ac: peptide carrying 3 acetylations, 4ac: peptide carrying 4 acetylations. Values derived from 3 (wildtype, H3.3-GFP) or 4 (CID-GFP) biological experiments, respectively. Error bars represent standard deviations.

H4 of CID-GFP containing chromatin contains more monoacetylation and almost 2.5-fold more diacetylation, whereas the H4 molecule co-purifying with H3.3-GFP does not show a clear trend. In contrast to previous reports (Sullivan and Karpen 2004), these results indicate a slightly higher H4 acetylation in nucleosomes containing CID-GFP.

2.4.2 Identification and quantification of proteins in purified fractions by LC-MS/MS analysis

The application of mass spectrometry analysis to the immunoprecipitated material as produced for figure 2.6 results in identification and quantification of factors associated with the respective GFP-tagged histone H3 variants. The

raw files of the MS runs are deposited in ProteomeXchange with identifier PXD000758 (http://proteomecentral.proteomexchange.org/dataset/PXD000758) and a MaxQuant output list of all identified proteins is published (http://onlinelibrary.wiley.com/doi/10.1002/pmic.201400052/suppinfo, table S1). An extended description about how protein abundances can be measured quantitatively by MS analysis can be found in chapter 4.2.5.2. Briefly, the peptide sequences detected in the spectra are compared to a database containing all theoretical tryptic peptides of the *Drosophila* proteome. If an identified peptide is unique, it can be assigned to a certain protein and its abundance can be determined. Abundance values come from extracted ion chromatograms (XICs, figure 4.7). The software MaxQuant provides an iBAQ value, which stands for intensity based absolute quantification. A feature of the value is taking into account that larger proteins may generate more tryptic peptides that can be detected. In the case of larger proteins, this also translates into higher total intensities. As the iBAQ value is corrected for this issue, its value correlates well with the abundance of a protein contained in one sample irrespective of its size.

To detect differences in protein abundance in the immunoprecipitated proteins, the result files of biological replicates of control purifications (wildtype L2-4 cells, experiments SID1275, SID1341, SID1392), CID-GFP purifications (SID1275, SID1341, SID1392) and H3.3-GFP purifications (SID1341, SID1392, SID1508) were analyzed. Proteins only identified in CID-GFP IPs were defined as specific to CID-GFP containing chromatin. Figure 2.9 shows the number of proteins for which an iBAQ value could be calculated in at least one of the three biological replicates and the overlap of the respective proteins between the different immunoprecipitations. 149 proteins were identified exclusively in at least one CID-GFP purification and in none of the six other purifications (3x wildtype, 3x H3.3-GFP). A list with names of these potentially interesting factors is attached in table 1 in the appendix. Control centromeric proteins like CENP-C and Cal1 are contained therein, demonstrating the specificity of this purification.

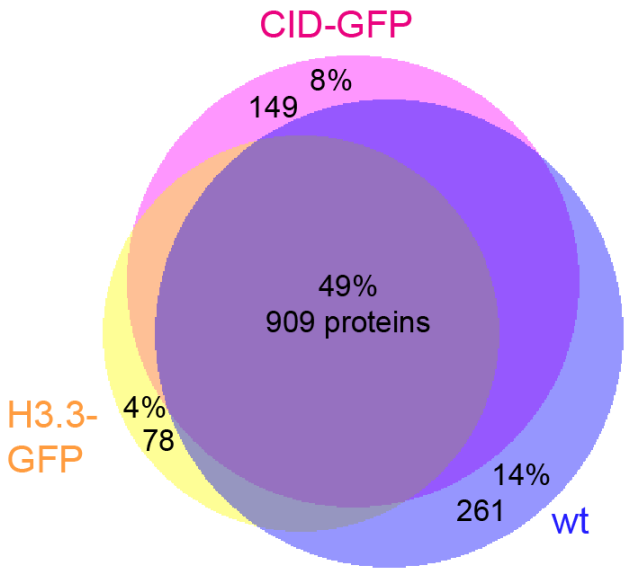

Figure 2.9: Venn diagram giving proportions of all identified proteins and their overlap in different purifications. Proteins are counted if they got identified at least in one of the three biological replicates. Percentages and numbers indicate the portion of the total proteins that were either identified in all different purifications (middle) or in one specific IP. Diagram created with BioVenn (http://www.cmbi.ru.nl/cdd/biovenn/).

Since the analysis provided quantitative values for all different affinity purifications, it allowed for the calculation of enrichments of factors in CID-GFP versus wildtype IP and CID-GFP versus H3.3-GFP IP, which eliminates proteins binding to the affinity matrix and general chromatin binding factors, respectively. This strategy should retain the most specific factors of centromeric chromatin. Statistical analysis was performed by Dr. Andreas Schmidt.

The log2-transformed iBAQ values obtained by MaxQuant analysis exhibited a Gaussian distribution, which is a prerequisite to many statistical tests (chapter 4.2.5.5, figure 4.9), thus enabling statistical analysis of this dataset. Therefore, iBAQ values of identified proteins in the CID-GFP purification and corresponding values in the control or H3.3-GFP sample were transformed by applying the logarithm to the basis of two. Zero values, occurring if a protein is not identified in a given sample, were replaced by imputation with values taken from a random distribution centered around one third of their lowest value. The imputation was repeated three times and the average value employed for subsequent calculations. These repetitions help reducing false positive or false negative assignations by imputation. ANOVA (analysis of variance) test was applied to compare protein abundances between different

affinity-purification experiments in the three biological replicates and calculate p-values (see table 2, appendix). P-values were corrected for multiple hypothesis testing by the method of Benjamini-Hochberg (Benjamini and Hochberg 1995) which is the preferred method for large proteomics datasets. The average enrichment in protein abundance of all three biological replicates is plotted in figure 2.10. Proteins with a more than 16-fold enrichment over both controls (log2 ratios of more than four) were considered as potential CID-GFP interactors (green dots, protein names included in inlet). A complete list of candidate proteins including average ratios and their standard deviations as well as p-values for both controls is included in table 2 (appendix). This table furthermore includes information about expression vectors for *Drosophila* cells created and used in this work as well as a summary of the cytological experiments that were conducted.

Among the candidate proteins were the two different isoforms of CAF-1, CAF-1-PA and CAF-1-PB. The PA-isoform contains one additional amino acid, an alanine in position 8. Apart from that, the two isoforms are completely identical. Since the peptide covering the differing amino acid was identified by MS analysis for both isoforms, it is possible to distinguish them by mass spectrometry (figure 2.10, table 1 and 2, appendix). MaxQuant analysis to determine iBAQ values was set to only quantitate unique peptides – in this case, only the one spanning amino acid 8 (4-18 in CAF-1-PA). Therefore, individual quantification of the both isoforms is indicated for example in figure 2.10 and tables in the appendix.

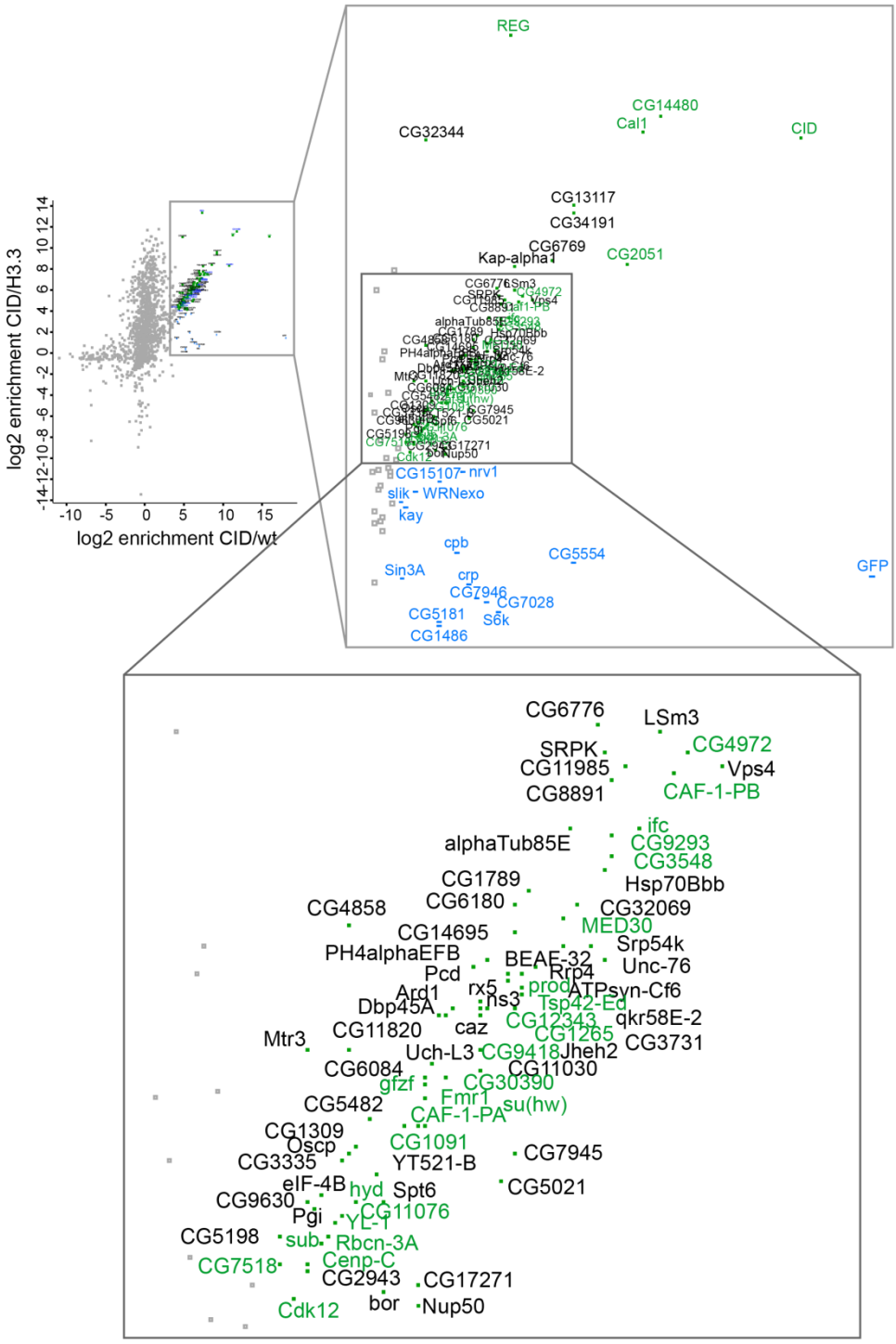

Figure 2.10: Plot of all proteins identified by MS analysis. Shown are log2 values of average ratios of iBAQ values in CID-GFP purifications versus control purifications (wt, x-axis) or versus H3.3-GFP purifications (y-axis). Each dot represents one identified protein. The inset on the right shows a magnification of the upper right quadrant and the lower one another magnification as indicated by boxes in the figure. Blue dots and protein names: Proteins only enriched over the wildtype control. Green dots and green and black protein names: average log2 value in both controls bigger than 4. Green protein names (see also table 2, appendix): Candidates further tested for cellular localization (chapter 2.5).

In total, 1871 proteins were identified in all experiments. Grey dots in figure 2.10 represent factors whose enrichment in CID-GFP purifications was less than 16-fold (log2 ratio < 4) compared to both controls. These proteins were therefore classified as unspecific affinity-matrix binders (background binders). Only 85 proteins passed the arbitrary threshold (table 2, appendix). They are labeled with green dots and their names are written in black and green in figure 2.10. As expected, the bait protein CID is the most enriched protein and GFP is highly enriched only over the wildtype control, but not over H3.3-GFP (thus labeled in blue in figure 2.10). CG30390 is a factor that is enriched in CID-GFP IP versus both H3.3-GFP and wildtype precipitation and thus locates to the upper right quadrant in figure 2.10. Hence, it is one of the 85 enriched factors that are also listed in table 2 (appendix). Its enrichment in CID-GFP purifications was also verified by Western blot analysis (figure 2.11). Similar to figure 2.5, input and bead material from another replicate was loaded on a 15 % SDS-polyacrylamide gel and analyzed by anti-GFP antibody, indicating that the right GFP-tagged H3 variant is present on the beads. Subsequently, the membrane was probed with CG30390-specific antibody. It only gives a signal in purification of the CID-GFP containing cells, meaning that CG30390 interacts with CID-GFP.

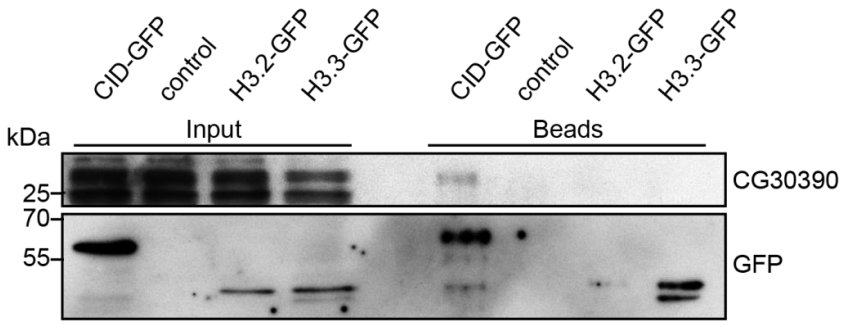

Figure 2.11: CG30390 specifically associates with CID-GFP containing chromatin. Western blot probed with a CG30390 specific antibody and GFP antibody. Loaded are 0.3 % of the input and 80 % of the immunoprecipitated material.

Reanalysis of the primary data uncovered that some factors, which repeatedly and exclusively co-purified with CID-GFP chromatin, did not pass the initial filtering. Consequently, the process of imputation might create false negative assignments by replacing zero for higher values in non-CID chromatin purifications. Thereby, the calculated enrichment of bona fide centromere associated proteins might drop below the threshold. Thus, based on their specificity of interaction with CID, additional factors were included that could also be involved in centromere biology. For these factors, at least 90 % of all iBAQ values from the nine replicates (3x wildtype, 3x CID-GFP, 3x H3.3-GFP)

must have originated from the three CID-GFP purifications, reflecting their specific presence in CID-containing chromatin. Thereby, nine proteins were included (Hcf, asp, CG6227, CG32343, CG1399, RhoGAP54D, CG8478, Top3beta, Ge-1), elevating the number of potential centromeric proteins to 94 (85 passing the threshold plus nine additional factors, see table 2, appendix).

After determining 94 potential centromere proteins (table 2, appendix), the current literature was searched for prior knowledge about these factors. This can help to categorize the factors in order to define future directions of research. The STRING database (Franceschini, Szklarczyk et al. 2013) provides a collection of information on proteins and genes from different curated databases (MINT, HPRD, BIND, DIP, BioGRID, KEGG, Reactome, IntAct, EcoCyc, NCI-Nature Pathway Interaction Database, GO) and model organisms and can automatically connect a number of factors via these data. The 94 candidates were fed into the algorithm and information with at least medium confidence (confidence score 0.4) about coexpression, experiments, databases and text mining in *Drosophila melanogaster* and other species was included in the analysis. Figure 2.12 represents the output of the STRING database analysis.

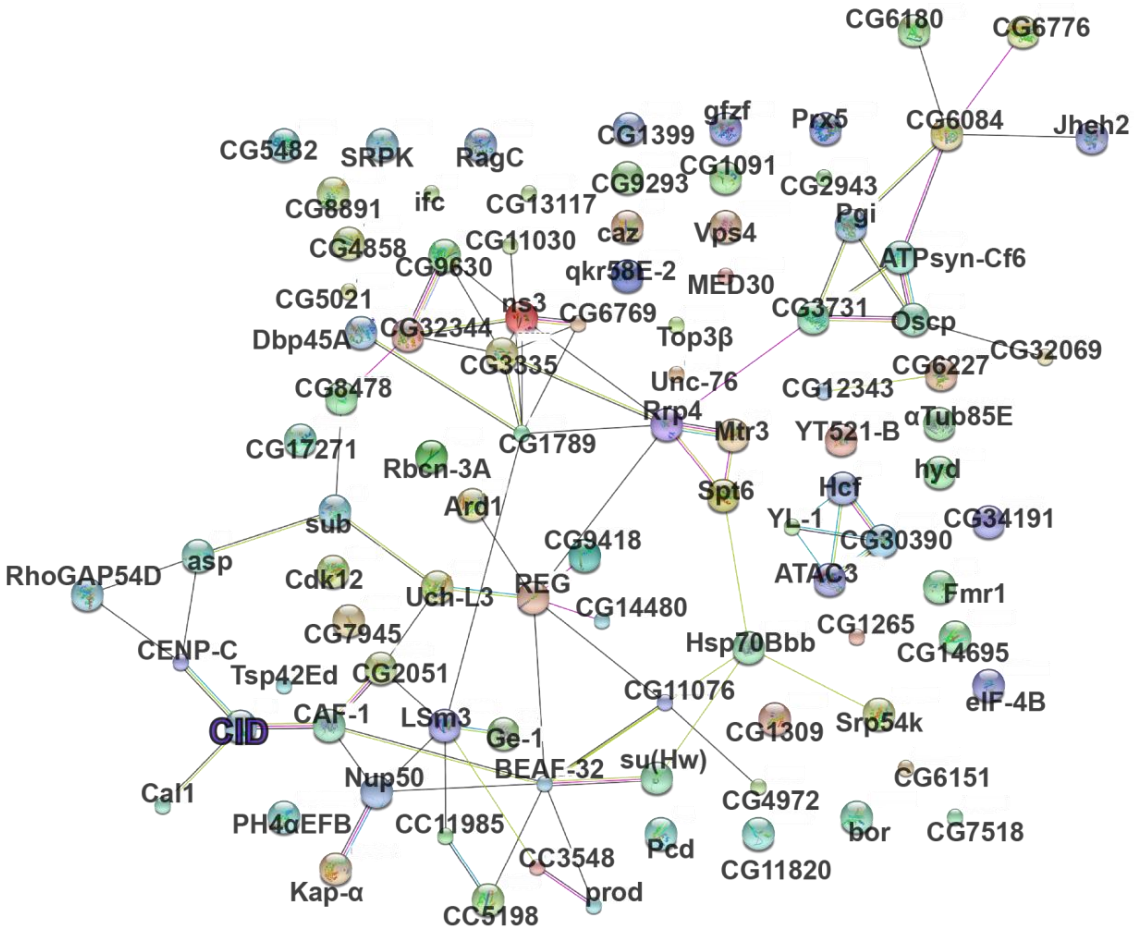

Figure 2.12: STRING database network indicating known relations of proteins that were enriched with CID-GFP. Black lines: Coexpression in *D. mel* or other species. Pink lines: Interactions based on experimental data. Blue lines: Information gained from curated databases. Green lines: Information obtained from text mining.

Different clusters can be recognized (for example, centromeres: Cal1, CID, CENP-C, CAF-1, bottom left; HATs: Hcf, YL-1, CG30390, CG32343, middle right; RNA-associated clusters like CG3335, CG9630, CG32344, CG6769, top left; REG, CG14480, Ard1 and CG9418, middle; Rrp4, Spt6 and Mtr3, middle; potential cytoplasmic factors: CG6084, CG6776, CG6180, Jheh2, upper right) indicating the co-purification of complex protein assemblies. However, many factors have not been referred together with any other of these proteins. This could be due to a lack of centromere proteome studies in *D. mel* prior to this work.

To confirm the centromeric association of candidate factors, immunolocalization was used as a complementary technology allowing single cell analysis of physiological states. Candidates to test by immuno-fluorescence were chosen from the pool of 94 proteins (table 2, appendix) that were at least identified in two out of three CID-GFP biological replicate purifications and show a high specificity for CID-GFP purifications. Of the top 40 factors fulfilling these criteria, 32 were selected and further tested (chapter 2.5). They are labeled in green in table 2 (appendix) and in figure 2.10. The remaining eight factors were not investigated since expression constructs were not available and molecular cloning from cDNA templates failed (CG32344, SRPK, qkr58E-2, CG32069, CG6151, CG13117, RhoGAP54D, and CG8478).

2.5 Cellular localization analysis of selected candidate proteins by fluorescence microscopy

As explained in chapter 2.4.2, 32 proteins enriched in CID-GFP containing chromatin were selected for further analysis. Whether they truly localize to centromeres *in vivo* was investigated by immunolocalization experiments. Therefore, colocalization of candidate proteins with CID was investigated in cells that were either transiently or stably transfected (indicated in the figure legends) with expression constructs encoding for either GFP- or FLAG/HA-tag fusion proteins. Immunofluorescence experiments were carried out by both Georg Schade and me.

For two of the most enriched factors on CID-containing chromatin, CG2051 and CG14480 (figure 2.10, table 2, appendix), stably transfected cell lines were established by selecting cells with hygromycin for at least four weeks. Stable cell lines are preferred to transiently transfected cells as they frequently display more moderate transgene expression levels. In the cases of monitoring fusion protein localization in transiently transfected cells, only low

level expressing cells were chosen to avoid overexpression-driven mislocalization. Figure 2.13 shows localization of the GFP-tagged proteins in the stable cell lines either as maximum intensity projections of all optical sections or single optical sections (indicated as "z="). This figure emphasizes the need of assessing potential colocalization of CID and the putative centromere protein in single section images. In maximum intensity projections of all sections, spatial information about localization gets lost. Signals of both CG2051-GFP and CG14480-GFP are present in euchromatic areas in the whole nucleus. In single sections, however, presence at the centromere is clearly visible.

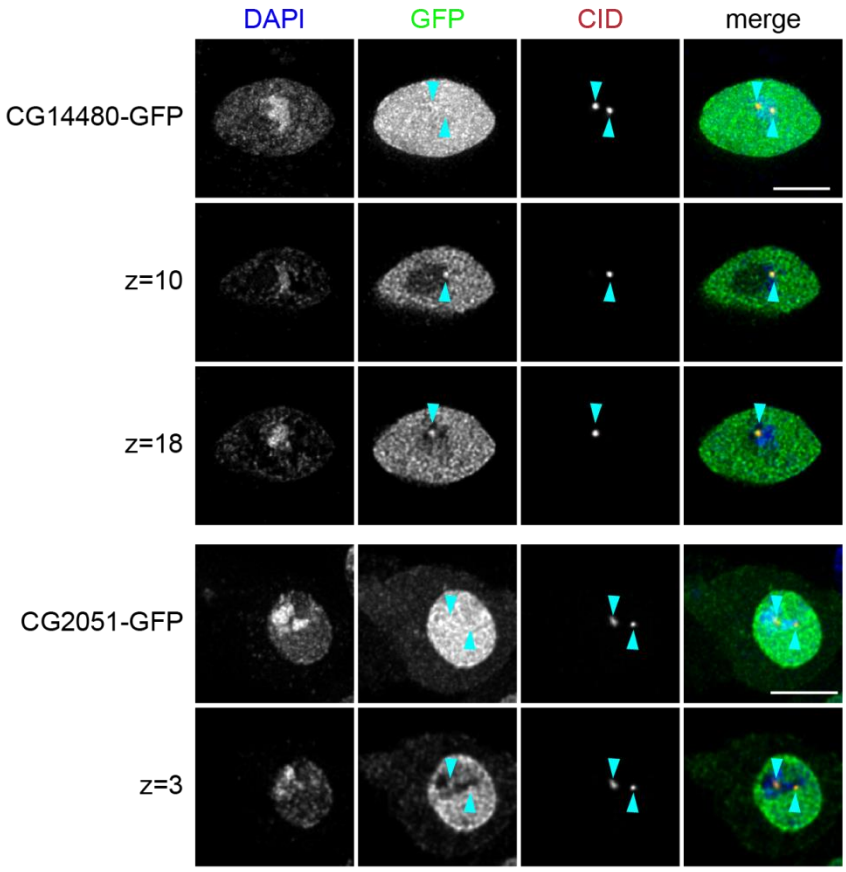

Figure 2.13: Localization of CG14480-GFP or CG2051-GFP, respectively, in stable cell lines. The pictures are maximum intensity projections unless the single optical section displayed (z) is indicated. Pictures were deconvoled. Scale bar represents 5 µm. Blue: DAPI, green: GFP signal, red: CID antibody stain.

Figure 2.14 summarizes all factors found to colocalize with CID in this analysis. Single sections were chosen to highlight colocalizations as illustrated in figure 2.13. For better visibility of signal distributions in different channels, line profiles are added in figure 2.14 and 2.15. Here, lines can be drawn in the merged picture and a software (RGB profiler) plots the intensities

of the different channels (y-axis) versus the distance on the drawn line (x-axis). The centromere signals (red peaks) can then be overlaid with information from the channel corresponding to the detection of tagged candidate protein (green signal).

In this work, the proteins CG2051, CG14480, CG3548, CG9293, CG9418, CG6227, gfzf, MED30 and REG were shown for the first time to localize to centromeres. CG3548 is found on centromeres only during M-phase of the cell cycle and not during interphase (figure 2.14). Colocalization of Subito (sub) and CID has already been demonstrated in larval brain tissue (Cesario, Jang et al. 2006).

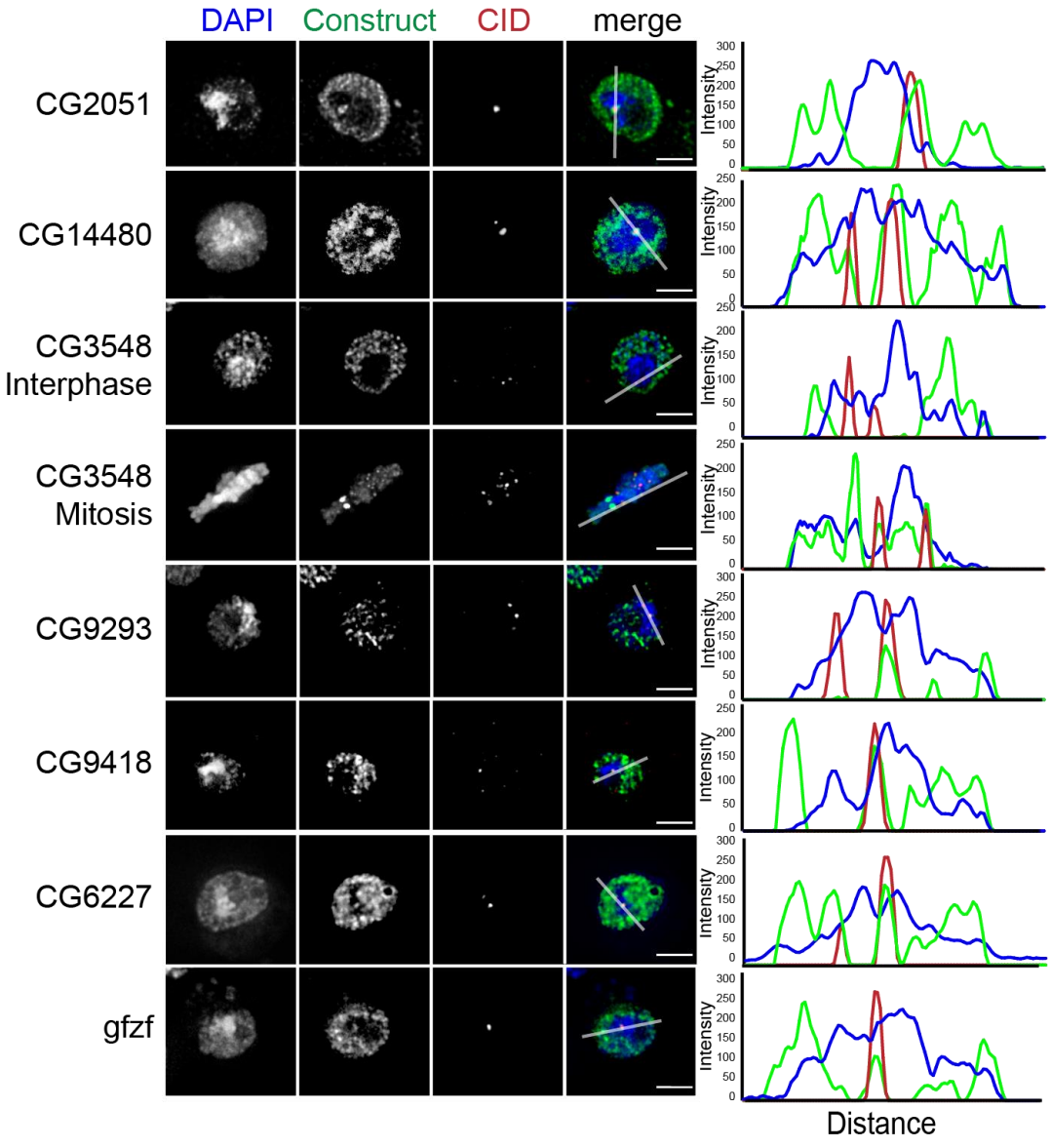

Figure 2.14, part 1. Continued on next page.

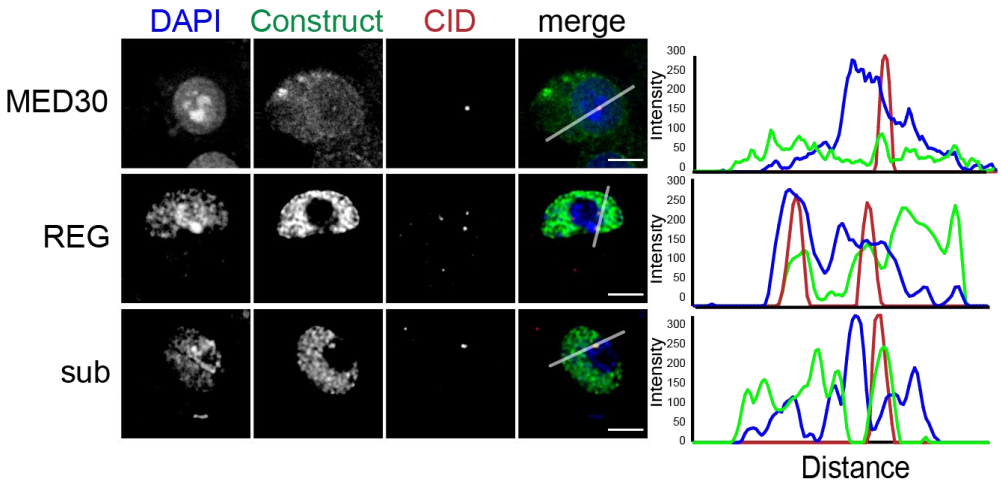

Figure 2.14: Candidates colocalizing with CID. Single optical sections of deconvolved images are shown. Scale bars represent 3 µm. The line profile plots depict the distribution of signals in the three channels along the line drawn in the merge panel. Blue: DNA stained with DAPI, green: epitope-tagged candidate protein, red: CID. Images for CG2051, CG14480 and CG3548 are obtained from stably transfected cell lines, the others from transient transfections. Immunofluorescence images have been acquired by both Georg Schade and me.

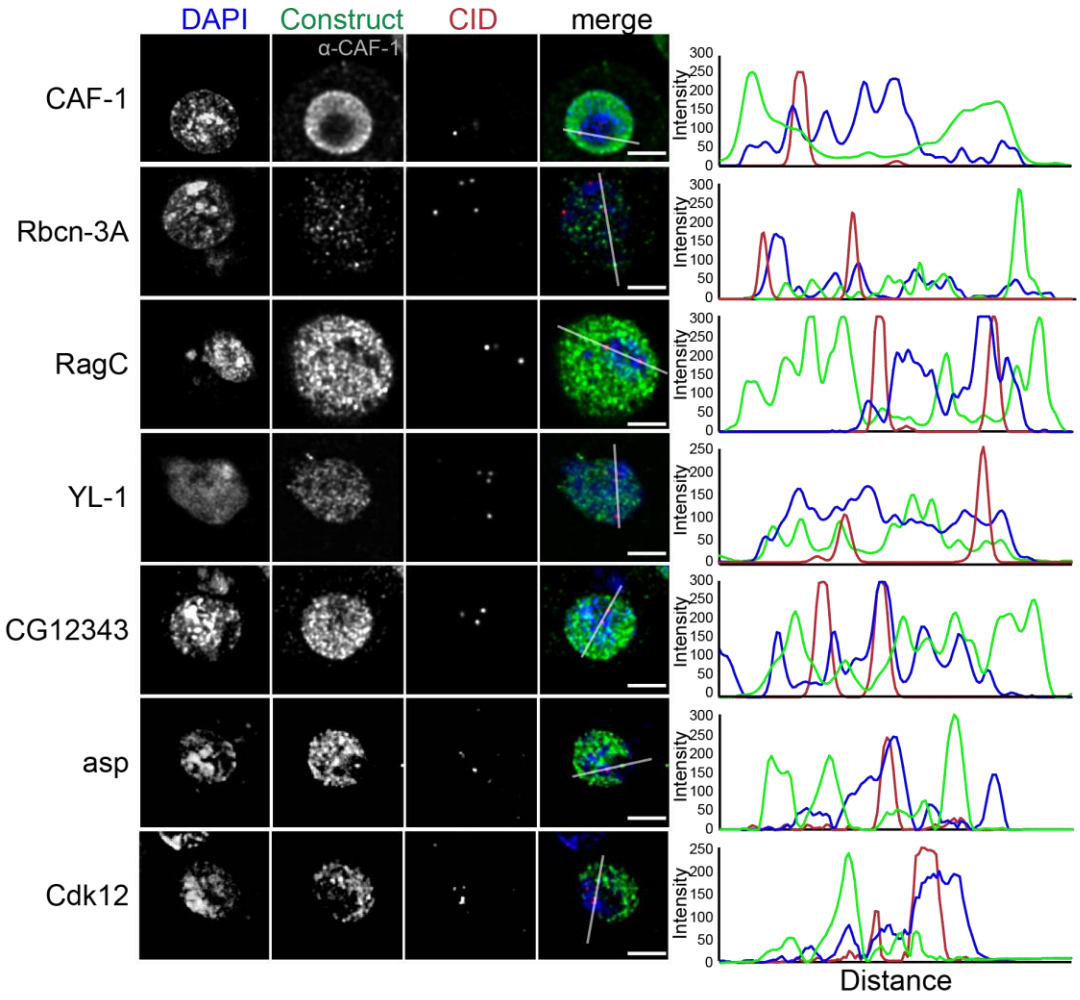

Figure 2.15, part 1. Continued on next page.

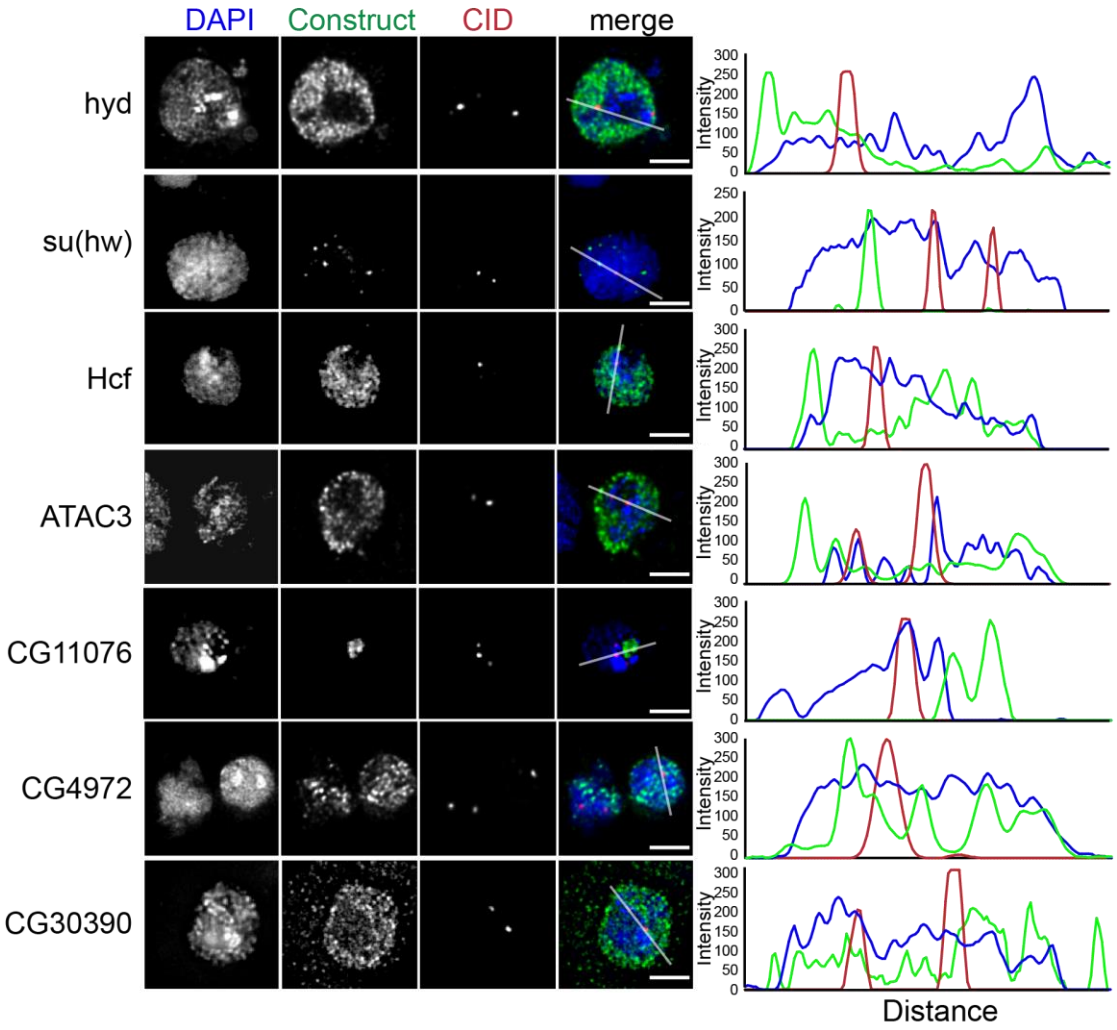

Figure 2.15: Candidates exhibiting nuclear localization. Single optical sections of deconvolved images are shown. Scale bars represent 3 µm. The line profile plots show the distribution of signals in the three channels on the location of the line drawn in the merge panel. Blue: DNA stained with DAPI, green: epitope-tagged candidate protein except CAF-1 (anti-CAF-1 antibody signal), red: CID. Images recorded on transiently transfected cells except CG11076-GFP (stable cell line). Immunofluorescence images have been acquired by both Georg Schade and me.

Despite their biochemical association with CID-containing chromatin, a number of factors exhibited nuclear localization without enrichment at centromeres (figure 2.15). One of these factors is CAF-1. As an antibody was available, no overexpression had to be performed. CAF-1 localizes to euchromatin and the line profile reveals no enrichment at CID foci. Still, CAF-1 has been shown to bind to CID (Furuyama, Dalal et al. 2006). Thus, the category of nucleus-localizing proteins can potentially also be implicated in centromere biology. Other factors of this class comprise Rbcn-3A, RagC, YL-1, CG12343, asp, Cdk12, hyd, su(hw), Hcf, ATAC3, CG11076, CG4972

and CG30390. As CG11076 displayed a distinct localization to regions of low DAPI intensity, its colocalization with the nucleolar factor Fibrillarin was tested (figure 2.16). Indeed, a good overlap of the two signals can be observed, indicating that CG11076 localizes to nucleoli.

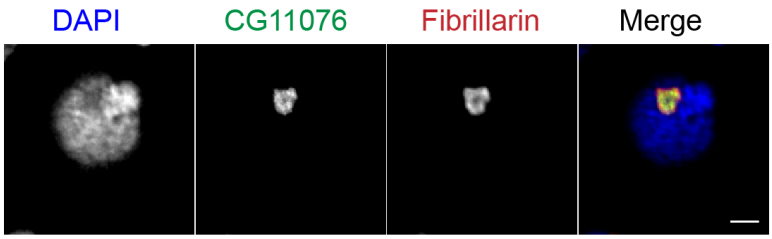

Figure 2.16: CG11076 colocalizes with Fibrillarin. Single optical sections of deconvolved images are shown. Blue: DNA stained with DAPI, green: CG11076-GFP, red: Fibrillarin. Scale bar represents 3 µm. A stable cell line expressing CG11076-GFP was used. Figure created by Georg Schade.

Thus, 24 out of 32 tested factors were found to localize to nuclei. Other eight proteins showed signals outside of nuclei (figure 2.17): either in the cytoplasm, like Fmr1, CG7518, infertile crescent (Ifc), CG1265, CG1091, and Hsp70Ab, or in the cell membrane, like CG1399 and Tsp42Ed. Hsp70Bbb is the heat shock protein isoform contained in the 94 enriched proteins; however, Hsp70Ab and Hsp70Bbb share 99 % of identity on protein level (five different amino acids in 641 amino acids). Both proteins were identified exclusively in CID-GFP purifications (table 1, appendix); but only Hsp70Bbb passed the threshold of an enrichment bigger than 16 (figure 2.10; table 2, appendix). For reasons of construct availability, figure 2.17 shows immunolocalization of Hsp70Ab only. Presumably, localization of Hsp70Bbb is similar to the one of Hsp70Ab.

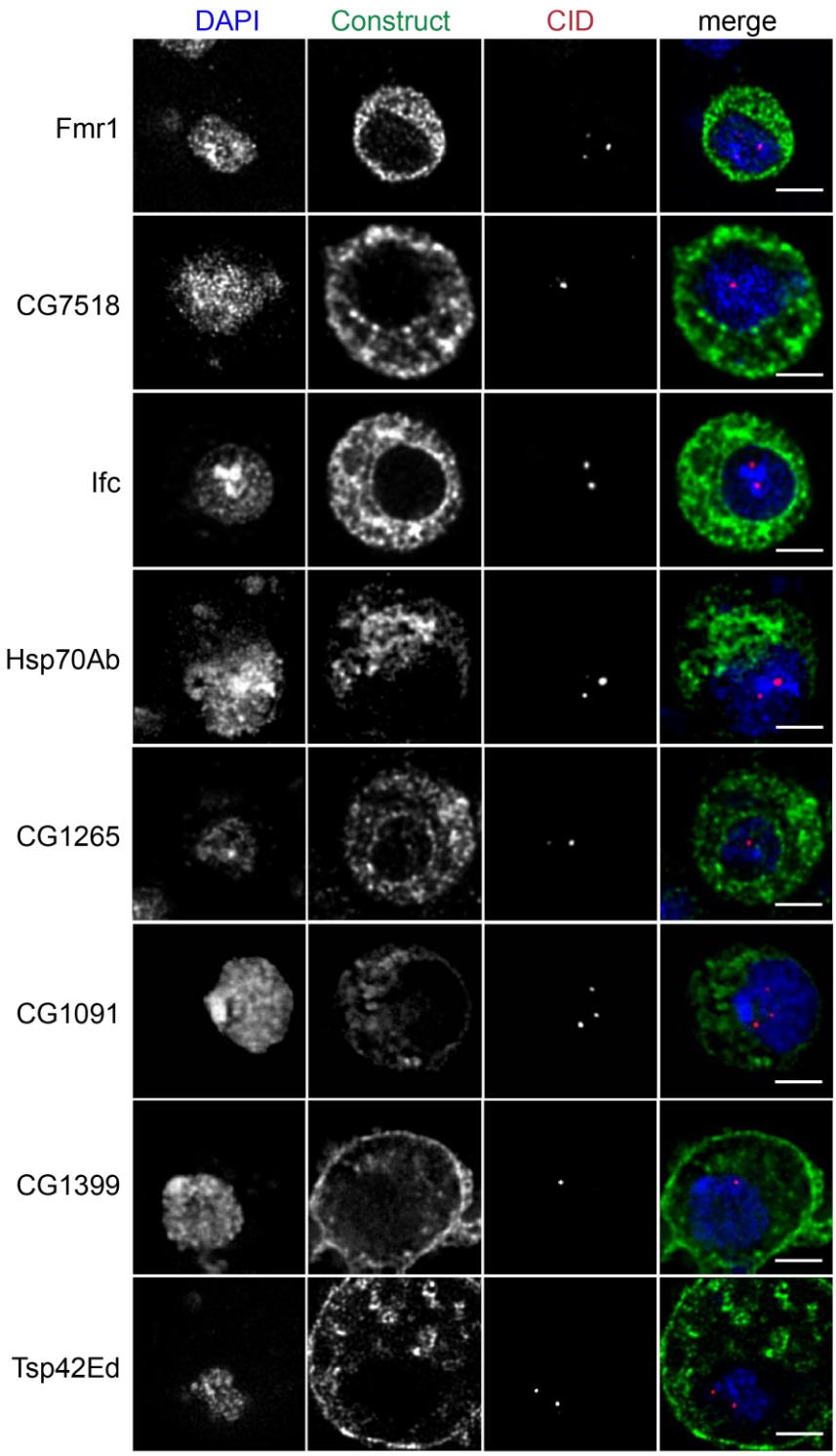

Figure 2.17: Candidates exhibiting localizations outside nuclei. Single optical sections are shown, images were deconvolved. Scale bars represent 3 µm. Blue: DNA stained with DAPI, green: epitope-tagged candidate protein, red: CID. Images obtained from transiently transfected cells except CG1091 (stable cell line). Immunofluorescence images have been acquired by both Georg Schade and me.

It was surprising to detect not only centromeric or nuclear proteins, but additionally also some that were located in the cytosol. This might be due to mixing of cytosolic and nuclear fractions during preparation of extracts (see also chapter 3.2). Alternatively, centromere association of candidate proteins may be cell cycle regulated, which was not investigated systematically for all 32 candidates. Nevertheless, different localization patterns were observed for the factor CG3548 (figure 2.14) which did not overlap with CID-stained regions in interphase, but on metaphase chromosomes. Therefore, its localization was also investigated on mitotic chromosome spreads (figure 2.18). As observed in asynchronous, fixed cells, CG3548 colocalizes with CID on mitotically condensed chromosomes with some prominent additional extra-centromeric foci. CG2051 and CG14480 are also bound to metaphase chromosomes, but are mostly distributed on chromosome arms without specific centromere enrichment (figure 2.18).

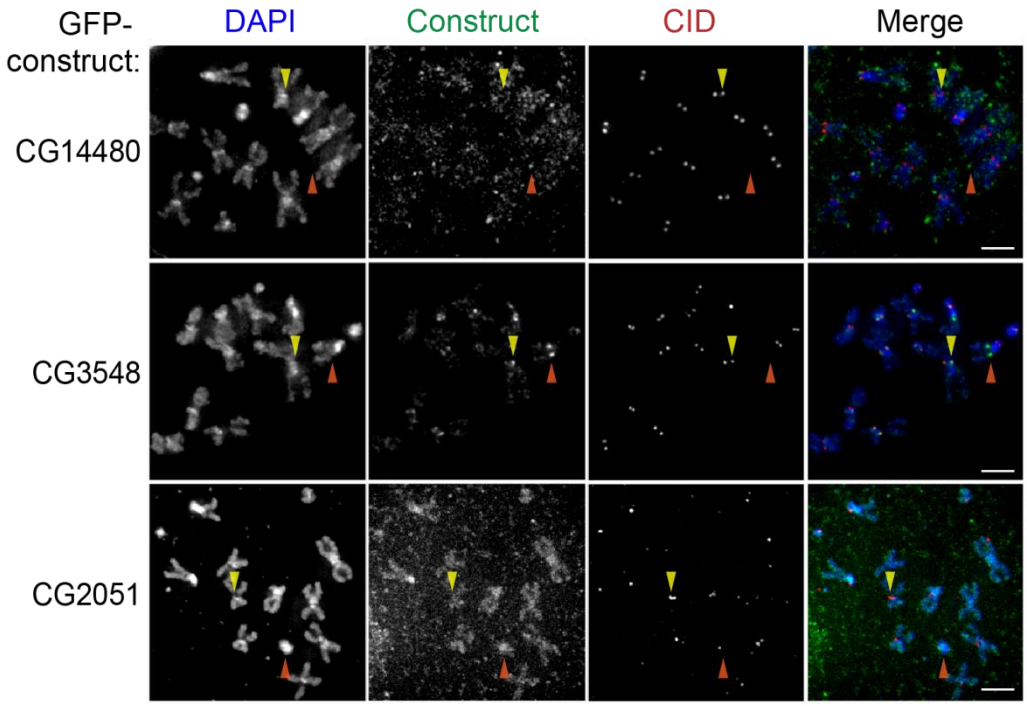

Figure 2.18: Mitotic chromosome spreads of CG2051, CG14480 and CG3548-GFP constructs in stable cell lines. Blue: DNA stained with DAPI, green: GFP signal from tagged candidate protein, red: CID. Yellow arrows point out bright red spots (CID signal), orange arrows green ones (GFP signal). Scale bar represents 3 µm. Images are deconvolved and maximum intensity projections are shown. Figure created by Georg Schade.

In sum, CG3548 is present at centromeres only during mitotic phase of the cell cycle (figure 2.14, 2.18), indicating that it possibly constitutes a kinetochore component. Adversely, the two highly enriched factors CG14480

and CG2051 can be found in euchromatin and centromeres during interphase and not during mitosis (figure 2.13, 2.14, 2.18). Other six factors were found as colocalizing with CID in interphase: CG9293, CG9418, CG6227, gfzf, MED30 and REG (figure 2.14). Fourteen more factors are localizing to the nucleus (figure 2.15) and could be involved in centromere biology just like CAF-1.

2.6 Influence of candidate proteins on CID-GFP incorporation in centromeres

One possible function of a centromeric protein is to promote the incorporation of CID at each round of cell division. To test if this hypothesis is true for some of the determined factors, the amount of candidate proteins was reduced by RNAi-mediated knockdown and the impact on centromeric CID amount was analyzed.

2.6.1 Influence of factors on centromeric incorporation of newly synthesized CID

An elegant way to measure incorporation of newly synthesized CID into centromeres is the use of the so-called SNAP technology (Keppler, Gendreizig et al. 2003). This technology is based on a suicide reaction of the enzyme O^6-alkylguanine-alkyltransferase, usually responsible for DNA damage repair involving guanine adducts. The SNAP-tag is a 20 kDa mutant form of this enzyme that reacts specifically with the substrates benzylguanine and benzylchloropyrimidine. An ether bond in the substrate is cleaved and a covalent bond with the enzyme is generated, with the reaction being irreversible. If the substrate is coupled to a fluorophore, the dye is conjugated to the SNAP-tag, making the technology suitable for fluorescence microscopy (Keppler, Pick et al. 2004). The scheme for the experiment is depicted in figure 2.19. *D.mel* Kc167 cells, a cell line derived from a female *Drosophila* embryo, coexpressing GFP-tagged CID as well as CID fused to the SNAP-tag, were obtained from the group of Gary Karpen (UC Berkeley). The CID-SNAP fusion can be reacted to the cell-permeable TMR-Star substrate (tetramethylrhodamine fluorophore coupled to benzylchloropyrimidine). Thereby, SNAP-CID can be visualized by fluorescence microscopy (figure 2.19, panel A). Alternatively, the cell-permeable drug BTP (bromothenylpteridine) can be added to cells expressing SNAP-CID. It will also be conjugated to SNAP-CID, but does not give a signal when analyzed by fluorescence microscopy (panel B). Hence, BTP treatment of cells will render all present SNAP-CID molecules unavailable for the conjugation of a

fluorophore, a process referred to as quenching. However, as soon as BTP is removed (chase), newly synthesized SNAP-CID can again react with a fluorophore (pulse). Quench-chase-pulse experiments were performed using the following scheme (panel C): Cells in different knockdown situations expressing SNAP-CID were incubated with BTP to block all SNAP enzymes (quench). After 24 hours of chase, corresponding to approximately one complete cell cycle in *Drosophila* Kc cells, TMR was added (pulse) and thereby exclusively SNAP-CID synthesized during the last 24 hours was detected by confocal microscopy. If this workflow is combined with RNAi treatment of cells before the quench, it allows screening for effects on incorporation of newly synthesized CID in the absence of a given centromeric factor.

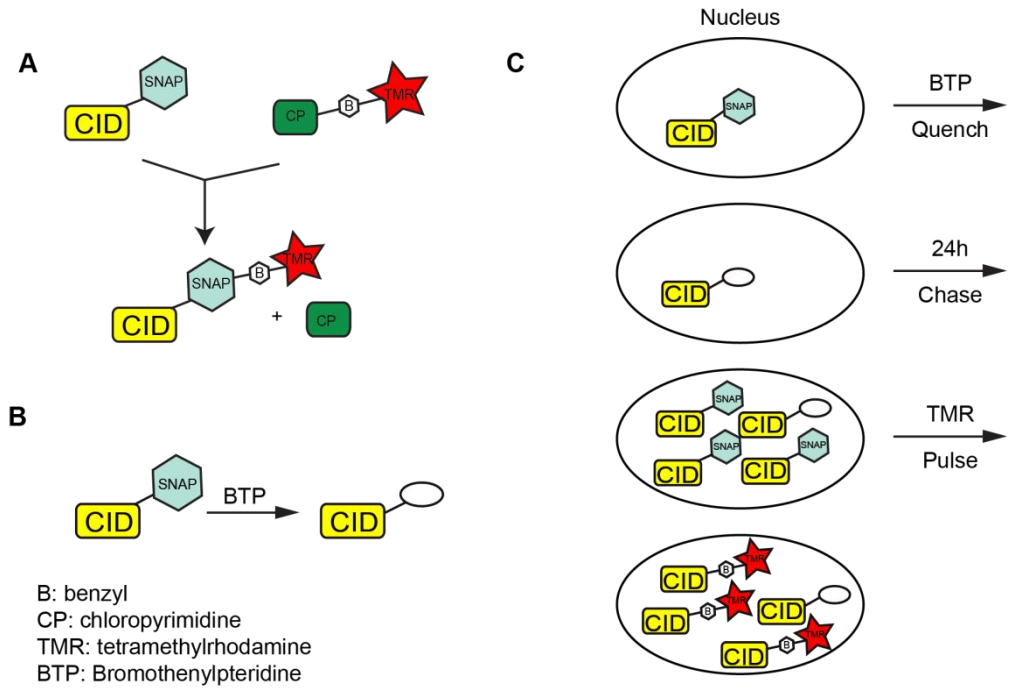

Figure 2.19: Schematic presentation for SNAP quench-chase-pulse experiment as performed in figure 2.21. A: CID is expressed with a SNAP tag. This enzyme can conjugate the benzyl and fluorophore part of an added TMR-Star substrate. B: Alternatively, the SNAP enzyme can also conjugate BTP to itself. C: For quench-chase-pulse experiments, cells in a knockdown situation are subjected to BTP, thereby quenching all preexisting SNAP tags. After 24 hours chase, TMR is added (pulse). The fluorophore then marks all SNAP-CIDs synthesized in the last 24 hours of the chase.

Figure 2.20 displays a TMR-stain of unblocked Kc cells (upper panel), demonstrating that the TMR signal colocalizes with the CID signal as expected. A quench-chase-pulse experiment with control (*GST*) knockdown is displayed in the lower panel.

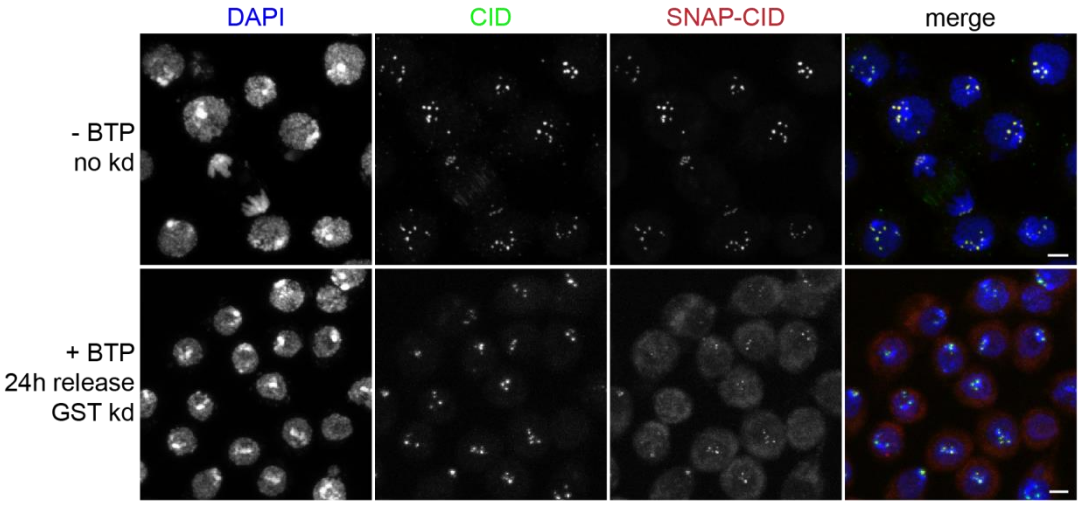

Figure 2.20: Stain of SNAP-CID in a stable cell line by TMR with or without BTP block. Blue: DNA stained with DAPI, green: CID (upper panel: antibody stain, lower panel: CID-GFP signal), red: SNAP-CID. Maximum intensity projections, images not deconvolved. Scale bar represents 3 µm.

Although colocalization of TMR-labeled SNAP-CID with CID-GFP is observed after BTP block and a 24 hour release, the high background of the TMR-Star signal in cytoplasm and nucleus hampered reliable measurement of CID-TMR signals in centromeres.

Figure 2.21 shows one quench-chase-pulse experiment series using the SNAP technology combined with knockdowns of known centromeric factors as well as selected candidate factors. An additional control was included to demonstrate that the BTP quench was successful (uppermost panel). Therefore, upon quenching and removal of BTP, cells were directly pulsed with TMR-Star without a chase. As expected, the CID-GFP signal is still visible, but no TMR-Star signal, since the cells were not allowed to newly synthesize SNAP-CID available for TMR-Star conjugation. This demonstrates that the BTP block was successful. A knockdown of *cid* itself as well as knockdown of *cal1,* which encodes the CID deposition chaperone (Mellone, Grive et al. 2011), served as positive controls, RNAi against *GST* as negative control. While control experiments produced the anticipated results (for reduction of CID and CID-GFP levels in *cid* knockdown, see also figure 2.23 lanes 9 and 10), *REG* and *CG14480* knockdowns do not lead to obvious defects in centromere loading of newly synthesized CID (figure 2.21, lowest two panels).

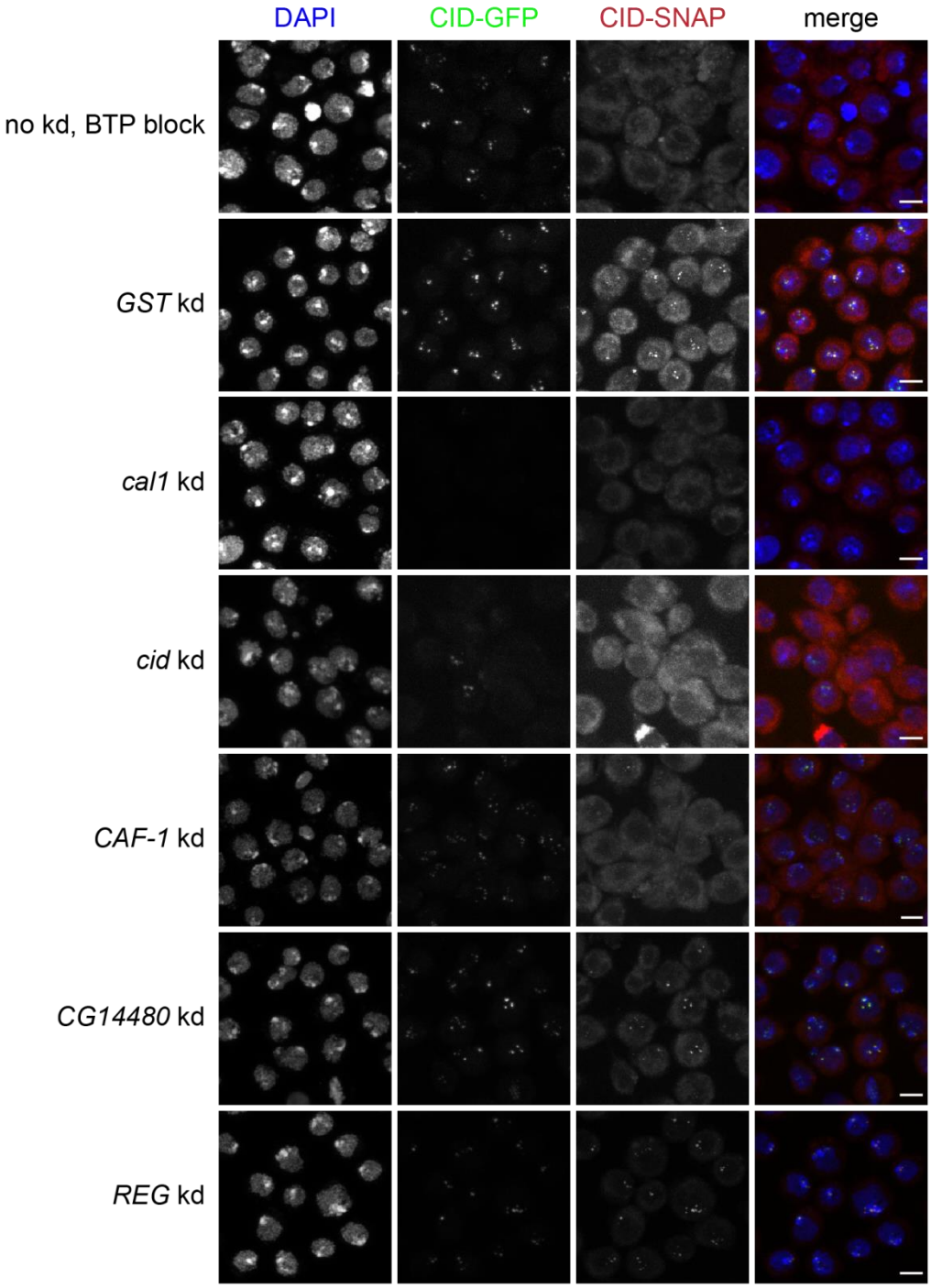

Figure 2.21: SNAP quench-chase-pulse experiment. Cells were treated with different dsRNAs (kd) for five days before the experimental scheme as indicated in figure 2.19C was applied (total knockdown time: 6 days). Blue: DNA stained with DAPI, green: CID-GFP, red: TMR-Star labeled SNAP-CID. Maximum intensity projections of non-deconvolved images are shown. Scale bars represent 5 µm.

When looking closer to the *CAF-1* knockdown images, there is a signal visible for TMR-Star labeled SNAP-CID at each centromere as determined by the CID-GFP channel. However, such faint signals can get masked by the high nuclear and cytosolic background staining. If a knockdown influences CID incorporation at centromeres, the signal in both the CID-GFP as well as the SNAP-CID channel are decreased. This, together with a problem of variable TMR-Star staining backgrounds, complicates automatic detection of centromeres and reliable quantification of SNAP-CID intensity therein. A quantitative statement about incorporation of newly synthesized CID in centromeric regions by automated analysis was thus not possible with this method. Therefore, CID-GFP fluorescence in centromeres upon candidate knockdown was measured to study their effect on CID deposition.

2.6.2 Quantification of CID-GFP signal intensities in centromeres

Knockdowns of different factors in the CID-GFP cell line also used for the AP-MS experiments were performed for six days before cells were analyzed by quantitative fluorescence microscopy. For this, fixed cells were imaged with constant hardware settings on a confocal laser scanning microscope. Deconvolved maximum intensity projections of these images were analyzed automatically with a pipeline in the software CellProfiler. The pipeline was developed by Dr. Andreas Thomae (see chapter 4.2.6.3 for a more detailed description). First, nuclei are detected using the DAPI staining. Second, centromeres are defined by a GFP signal in the nuclei. Third, average centromeric GFP intensity per cell was determined and multiplied with the detected number of centromeres to obtain the total centromeric CID-GFP intensity per cell. This value is plotted in the upper panel of figure 2.22. The pipeline also detects the number of centromere clusters present in one nucleus, an interesting variable since centromere declustering phenotypes are observed when knocking down another recently described centromere factor termed *NLP* (Padeken, Mendiburo et al. 2013) (see chapter 1.4). The number of centromere foci per cell is plotted in the lower panel of figure 2.22.

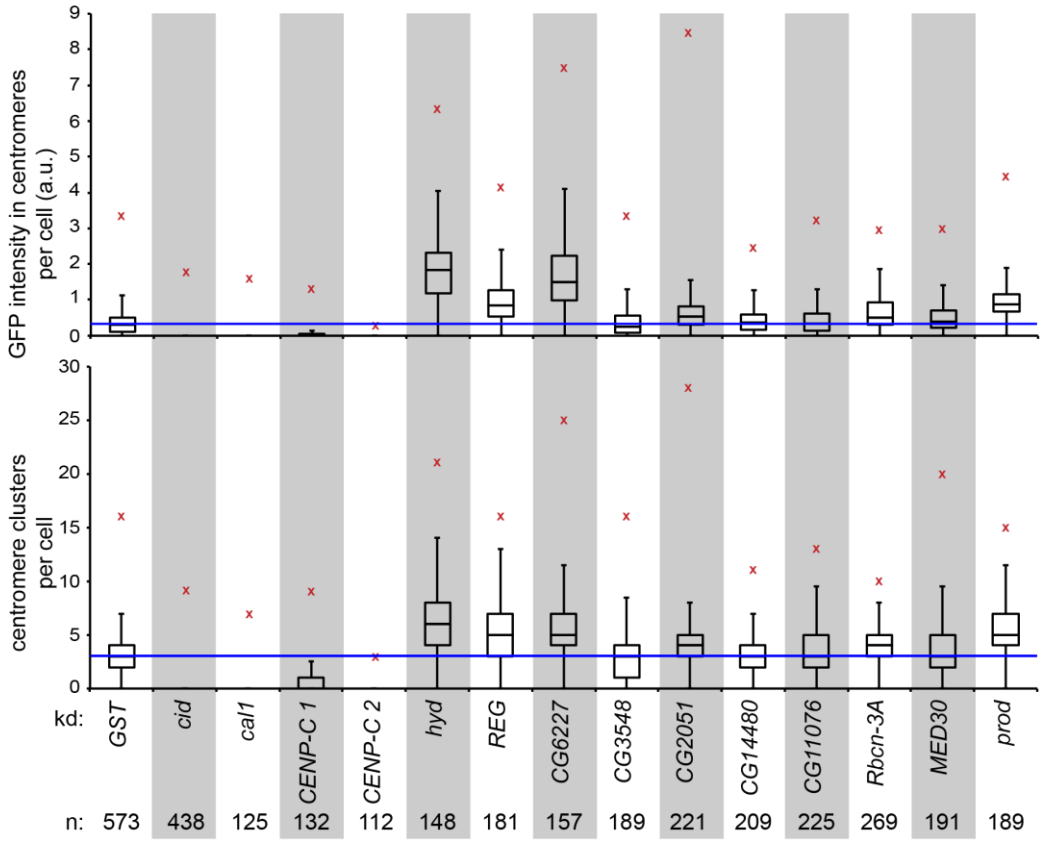

Figure 2.22: Quantification of CID-GFP in stably transfected cells in different knockdown conditions. Upper panel: Intensity of GFP signal in centromeres, arbitrary units. Lower panel: number of centromere foci measured per cell. The results are plotted in a box plot. The blue line indicates the median value of *GST* knockdown as a comparison. n: number of analyzed cells. Kd: knockdown of different genes. Red crosses show maximal outliers. Box plots created with Vertex42^{TM} Box and Whisker Plot Template. The median value is indicated by a horizontal black line surrounded by an upper and lower quartile, each comprising 25% of all values. Whiskers correspond to a maximum of 1.5 times of the inter-quartile range, meaning length of the box; maximum outliers are marked by a red cross.

As expected, control knockdowns, like *cid*, *cal1* and *CENP-C* (two different dsRNA constructs, 1 and 2) lead to a decrease in intensity of GFP in centromeres and less centromere clusters per cell as compared to a *GST* control knockdown. No other knockdown resulted in a decrease of intensity, but surprisingly, hyd, REG and CG6227 depletion in cells led to increased CID-GFP intensity in centromeres and higher centromere cluster numbers. Whether more CID-GFP results in more centromere clusters or the other way round cannot be deduced from this analysis. All other proteins analyzed here did not show any effect, indicating that they do not influence incorporation or maintenance of CID at centromeres.

2.7 Hyd or REG loss does not result in a global increase of CID-GFP

The RNAi-mediated reduction of *hyd*, *REG* and *CG6227* levels resulted in an increase of centromeric CID amounts (chapter 2.6.2). Hyd is an E3 ubiquitin ligase and REG a proteasome activator, suggesting that these factors might be involved in regulation of CID turnover through the ubiquitin proteasome system. Therefore, the effect of *hyd* and *REG* knockdowns on CID levels and ubiquitination were investigated by Western blot analysis. In order to stabilize ubiquitination, *Drosophila* L2-4 cells stably expressing CID-GFP were treated with NEM (*N*-ethylmaleimide) and/or MG-132 six hours before harvest and both inhibitors were always included in the protein extraction buffers. NEM irreversibly inhibits deubiquitinases whereas MG-132 inhibits the 26S proteasome. Treatment with one or both inhibitors was performed to investigate CID ubiquitination and proteasomal degradation by anti-CID antibody immunoblotting (figure 2.23). As opposed to the microscopy-based assay, no clear effect on CID levels upon knockdown of *hyd* as compared to *GST* control RNAi was observed (figure 2.23, lanes 1 and 2). However, when treating cells for six hours with MG-132, a slower migrating anti-CID antibody reactive band, which could represent mono-ubiquitinated CID (CID-ub), is appearing. This band is weaker in a *hyd* knockdown (compare lanes 3 and 4). The same is true when the MG-132 incubation is combined with NEM (lane 5 and 6), even though the CID-ub band is less pronounced here. With NEM treatment only, a difference between control and *hyd* knockdown is not detected, suggesting that MG-132 is the effective agent to visualize the shifted band and that it could indeed be ubiquitinated CID (lanes 7 and 8). *cid* knockdown reduces the levels of endogenous CID as well as GFP-tagged CID, as expected (lanes 9 and 10) and as observed earlier by fluorescence microscopy analysis (figure 2.21, 2.22). Furthermore, in *cid* knockdowns, the slower migrating band is also lost, underlining that this band corresponds to a modified form of CID. Comparable to the *hyd* knockdown, *REG* knockdown does not lead to a global increase in CID levels; however, here, the size-shifted CID band that could correspond to CID-ub is also present (lanes 11 and 12, compare to lane 3 and 5). Whether or not REG has an influence on the turnover of ubiquitinated CID cannot be deduced from this experiment. Since REG is an activator of the proteasome, it would be interesting to investigate if MG-132 addition is still required to stabilize ubiquitinated CID in a *REG* knockdown situation. To conclude, this preliminary experiment suggests an involvement of Hyd in CID ubiquitination, potentially thereby influencing its stability.

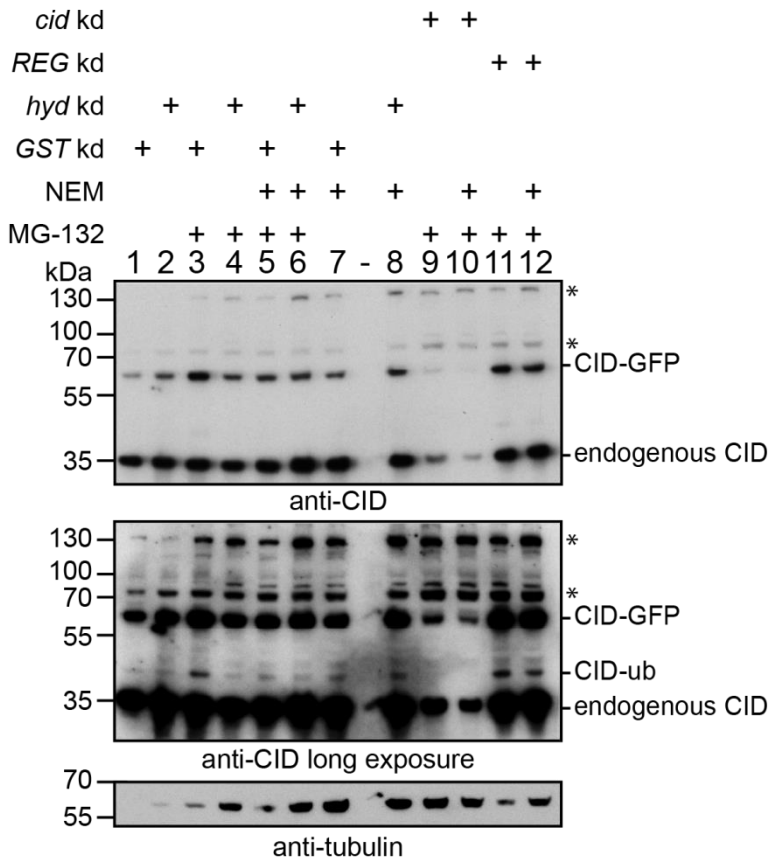

Figure 2.23: Western blot analysis of CID levels in CID-GFP expressing cell lines upon knockdown of *hyd, REG* or *cid* in combination with NEM and/or MG-132 treatment as indicated by the scheme. Anti-CID (short and long exposure) and anti-tubulin western blots are shown. Asterisks indicate unspecific bands. CID-ub: band probably corresponding to mono-ubiquitinated CID.

2.8 Influence of selected candidate proteins on cell cycle progression

The correct establishment of the kinetochore is dependent on centromeric CID and is a key step for cell cycle progression in mitosis. Hence, cell cycle state distribution analysis represents an additional readout to investigate whether a factor contributes to centromere functionality. Therefore, perturbations in cell cycle distribution upon factor knockdown were assessed by flow cytometry. After six days of knockdown, the DNA content in cells was measured by fluorescence-activated cell sorting (FACS) analysis of propidium iodide signals. Figure 2.24 comprises representative plots. G1-phase corresponds to a DNA content of 1 chromatid per DNA copy, G2 and M-phase have double the DNA amount and S-phase is the fraction between those two states. Panel A reassures that *GST* knockdown does not lead to a cell cycle defect as compared to untreated asynchronous cells. Interestingly, *cal1* and

cid knockdown lead to an increased amount of cells in S-phase (panel D). This effect is even more pronounced when CENP-C levels are reduced (two different dsRNA constructs: CENP-C 1/2, panel D). However, reduction of the two proteins strongly enriched in centromeric chromatin, CG2051 and CG14480 (figure 2.10), does not influence the cell cycle distribution (panel E). Depletion of Prod, a protein known to bind to pericentromeric heterochromatin on chromosome 2 and 3 of *D. mel*, and *gfzf* knockdown lead to slightly decreased number of cells in G2/M-phase (panel G and J). *CG6227* knockdown shows an enrichment of cells in G2/M-phase (panel K), an effect that is even more pronounced in *CAF-1* knockdown (panel H, 2 different dsRNA constructs). Nevertheless, none of the centromeric factors as determined by the biochemical purification assay shows similar cell cycle phenotypes as knockdown of *CENP-C*, the only known CCAN protein in *Drosophila*.

As histone H3 is known to be phosphorylated on serine 10 (H3S10ph) during mitosis by Aurora B kinase (Hsu, Sun et al. 2000), H3S10ph staining was used as a mitotic marker in addition to the propidium iodide staining of DNA and analyzed by FACS. This analysis reveals that in asynchronously cycling Schneider L2-4 cells, around 1.8 % of cells are in M-phase (Figure 2.24B and C). In the cases of *CAF-1* and *CG6227* knockdowns, where G2/M-phase is increased, the amount of cells in mitosis actually is decreased (panel I and L), indicating that in these knockdowns, cells are arresting in G2-phase. *CG2051* and *CG14480* knockdowns show no changes in portion of mitotic cells (only shown for CG2051, panel F. Value for CG14480 was 1.76 %.). Mitotic indexes for *prod* and *gfzf* knockdowns were not determined.

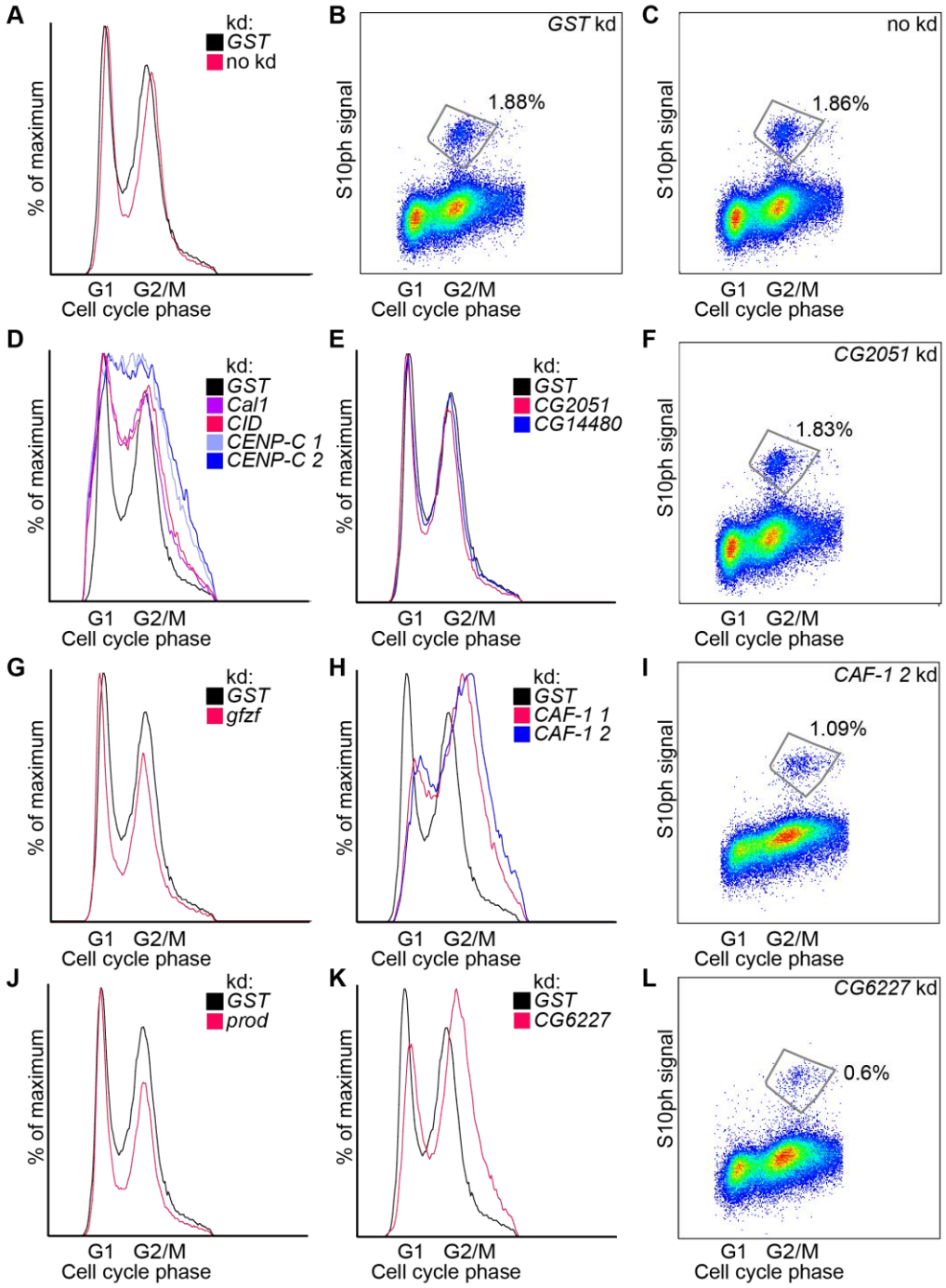

Figure 2.24: DNA content and number of mitotic cells in asynchronous cells after different knockdowns. Kd: knockdown. The x-axis displays measured DNA amount representing different cell cycle phases as indicated in the figure. Either DNA content is plotted versus cell number as histogram (percent of maximum to compare different cell numbers from different knockdowns) or DNA content versus signal from H3S10ph stain as density plot to distinguish mitotic cells. Percentages in panels B, C, F, I and L indicate the number of living, single cells positive for the mitotic marker H3S10ph.

In sum, depletion of none of the factors interacting with CID-GFP (chapter 2.4.2) induces defects as severe in cell cycle progression as observed upon knocking down the centromeric factors *CENP-C*, *cal1* and *cid*. This finding is in line with the missing effects of candidate knockdowns on CID-GFP incorporation in centromeres (chapter 2.6.2). Finally, the observed S-phase arrest of cells with reduced levels of *cid*, *cal1* and *CENP-C* is an interesting new finding.

2.9 Depletion of selected candidates leads to mitotic defects upon knockdown

The main function of the centromere and kinetochore is to ensure correct microtubule attachment and chromosome segregation during mitosis. Any failure in assembling these vital structures can therefore directly be assessed by monitoring mitotic defects. It was thus screened for lagging chromosomes, multipolar spindles and multinucleated cells in cells with reduced levels of proteins of interest by fluorescence microscopy (figure 2.25). Knockdown efficiency as measured by qPCR on mRNA level is indicated in the figure and always above 70 %, thus leading to a significant decrease of the target gene expression. A control knockdown targeting the *white* gene results in 15-20 % of mitotic errors in L2-4 cells, indicated with a red line in the figure. Thus, triggering an RNAi response in Schneider L2-4 cells alone leads to a quite high number of mitotic errors. *Prod* depletion, a protein localizing to pericentric heterochromatin, involved in centric chromatin condensation and cell proliferation and implicated in centromere function (Torok, Harvie et al. 1997), increases the number of cells displaying mitotic errors to 36 %. *CG2051*, *CG14480* and *hyd* knockdowns are in the same range. CAF-1 reduction, however, shows the strongest effect with 54 % of errors, potentially mirroring the observed cell cycle defects (figure 2.24H and I). Thus, some of the strongest enriched candidates of the biochemical screen for centromeric proteins are indeed required for proper mitotis.

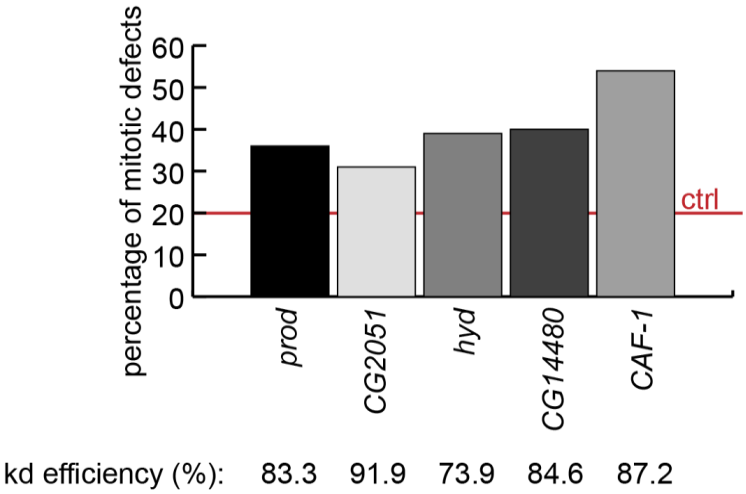

Figure 2.25: RNAi-mediated depletion of centromere associated proteins causes mitotic defects. Given is the frequency of mitotic defects (multipolar spindles, multinucleated cells and lagging chromosomes) as analyzed by fluorescence microscopy. The horizontal red line indicates the maximal percentage of defects observed in control knockdowns (*white* RNAi). Data collected by Georg Schade.

2.10 General centromere architecture is not impaired by CG2051 and CG14480 knockdown

CG2051 and CG14480 reduction results in elevated mitotic errors (figure 2.25) while CID incorporation in centromeres is not influenced (figure 2.22). Another possible consequence of their knockdown might be erroneous incorporation of other centromere and kinetochore proteins leading to defective centromere architecture. Therefore, localization of the proteins CENP-C, Rod, INCENP, Ndc80 and Polo was investigated by fluorescence microscopy in a knockdown situation of the factors *CG2051* and *CG14480* (figure 2.26). CENP-C is the only known *Drosophila* CCAN protein; INCENP is part of the inner centromere layer; Ndc80, Polo and Rod are located in the outer kinetochore and thus only present at centromeres in mitosis. Their proper localization was assessed by verifying their colocalization with or their proximity to CID foci in knockdowns of *CG14480*, *CG2051* and control knockdowns targeting the *white* gene.

However, no differences of localizations between control and factor knockdowns were observed. This, together with the fact that neither *CG2051* nor *CG14480* knockdown affects CID-GFP levels in centromeres, suggests that these factors are not involved in the establishment and maintenance of the basic centromere architecture. In addition, no influence on the cell cycle distribution was observed upon their depletion. (figure 2.24E). It will be

interesting to unravel their functions in *Drosophila* centromere biology in future experiments.

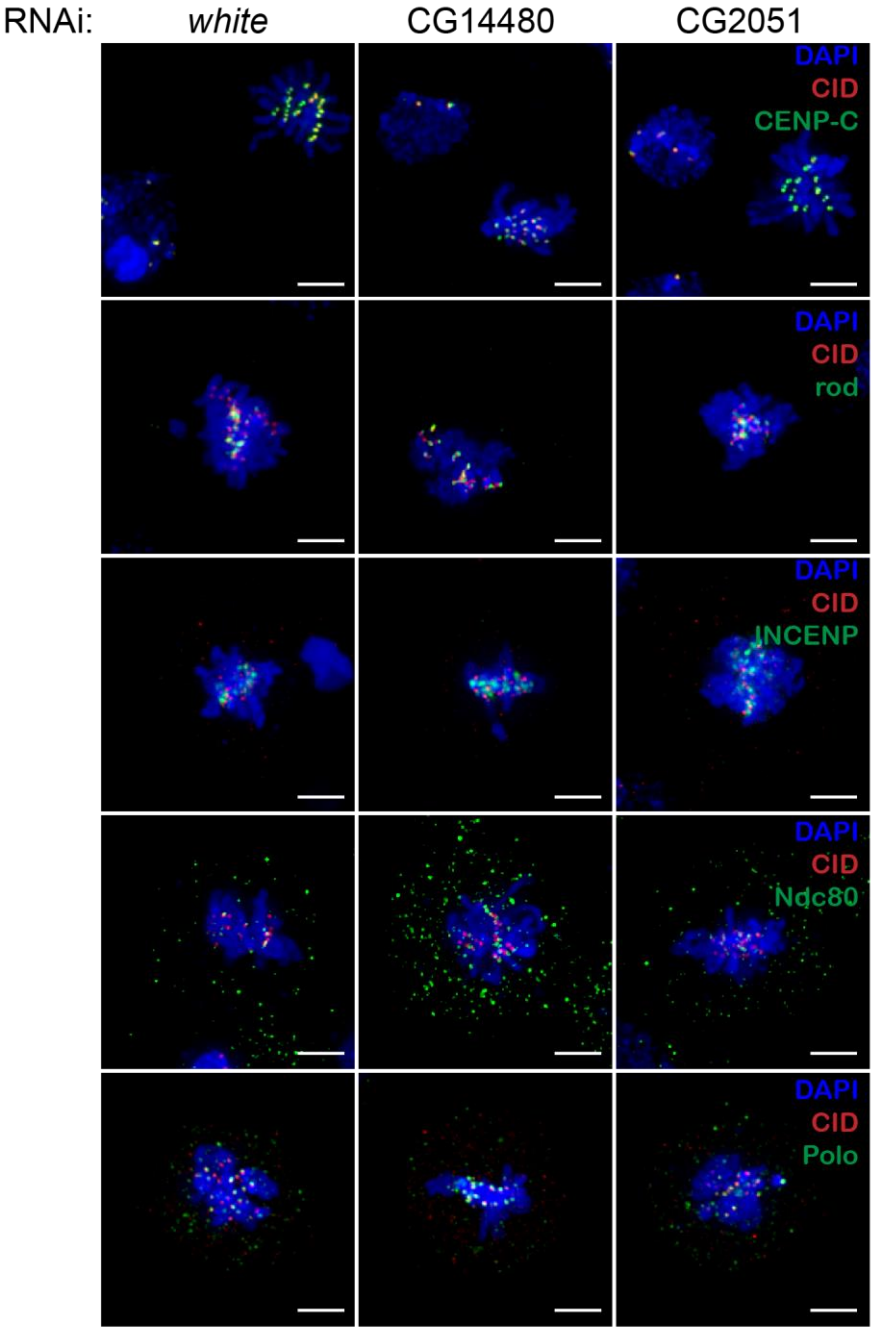

Figure 2.26: Loss of CG2051 and CG14480 does not impair the localization of known centromere and kinetochore proteins. Experiment performed by Georg Schade. Blue: DNA stained with DAPI, red: CID, green: centromere/kinetochore protein. Maximum intensity projections of deconvolved images are shown. Scale bar represents 3 µm.

3 DISCUSSION

The aim of this work was to better understand centromere composition and function in *Drosophila melanogaster*. It comprises the first published proteomic screen to identify novel *Drosophila* centromeric proteins. Immunolocalization analysis of selected candidate factors was performed as an independent assay to investigate the biochemically predicted centromeric association *in vivo*. By that means, at least nine novel factors were found to localize to centromeres. Further assays were performed to unravel possible functions in *Drosophila* centromere biology. In the following sections, the rationale of the methodology, the obtained datasets as well as the strategy for the analysis will be discussed. Some of the newly identified centromeric factors like acetyltransferases are able to modify proteins or are involved in protein turnover whereas the molecular function of others is completely unknown. The significance of these findings is considered below. Last, this work offers a first description of cell cycle defects upon knockdowns of centromeric proteins. The surprising finding of an intra S-phase block and its possible role in centromere biology is also discussed below.

3.1 (Dis)advantages of the applied AP-MS strategy

In order to determine centromeric proteins, a strategy was needed to isolate and enrich centromeric chromatin. Since centromeres are not a distinct compartment in a cell, this is not possible via centrifugation or size isolation strategies. Furthermore, centromeric chromatin needs to be solubilized from the confining nucleus and then affinity-purified. This could be done via the underlying DNA sequence by using the PICh (proteomics of isolated chromatin segments) method (Dejardin and Kingston 2009). However, as the exact DNA sequence is not known for *Drosophila* centromeres, this was not a feasible approach. Thus, a remaining strategy involved usage of a centromere-specific protein as a bait to fish for interactors.

As described in the introduction (chapter 1.4), only a few centromere proteins are known in *Drosophila*. Cal1 is the homolog of the centromeric H3 chaperone HJURP and therefore is also binding to CID in soluble nuclear fractions (Mellone, Grive et al. 2011). CENP-C as well as CID are constitutively binding to centromeres, but CID is present in the layer of centromeric chromatin and thought to confer identity to centromeric regions (chapter 1.2). Therefore, it is the best bait for identifying centromeric proteins. Some CID-specific antibodies exist; however, in order to provide a sufficiently high specificity, the centromeric chromatin has to be compared to chromatin

present in other regions of chromosomes. Therefore, the H3 variant H3.3 present in active, transcribing chromatin, was used to filter out generic chromatin binding factors in the proteomic screen. The possibility to compare the protein content associated with both H3 variants is only given if the same antibody is used for both. Thus, the variants were tagged so that they could be purified in the same manner. GFP was the tag of choice since it can be visualized by microscopy and GFP-Trap agarose resins are commercially available and can be used for immunoprecipitation.

The workflow described in chapter 2.1 was modeled after the methodology used by Foltz *et al.* who successfully applied AP-MS to identify ten novel proteins implicated in centromere biology in human HeLa cells (Foltz, Jansen et al. 2006). In order to prepare chromatin for affinity purification, it needs to be solubilized. This could be done by shearing chromatin using sonication. It is a quick way to obtain chromatin fragments that can be extracted from the nuclei. Sonication, however, may cause protein degradation and results in undefined DNA fragment sizes ranging from 100-500 basepairs. The MNase digestion applied in this work yields soluble chromatin with fragments corresponding to mainly mono- and dinucleosomes (figure 4.4), similar to the previous study in human cells. Upon MNase digestion, chromatin still needs to be separated from the unsoluble nuclear material. When different ways to release chromatin fragments were tested, neither detergent nor EDTA treatment or mild sonication resulted in detectably released DNA (figure 4.5). The only means of extraction was addition of 300 mM sodium chloride, which is consistent to the data from Foltz *et al.* (Foltz, Jansen et al. 2006). Subsequently, the IPs were also performed in these high salt conditions. This might lead to disruption of certain protein interactions and a loss of bona fide centromere proteins during the affinity purification steps. This might explain why some centromeric proteins of *Drosophila* like HMR, LHR, NLP or Umbrea (Ross, Rosin et al. 2013, Thomae, Schade et al. 2013) were not recovered in this screen. To stabilize their interactions with centromere chromatin, crosslinking with formaldehyde (FA) can be performed before chromatin extraction. However, incomplete reversal of such crosslinks can hamper MS detection and analysis, as it leads to mass shifts of 12 or 30 Daltons on crosslinked amino acids like mainly lysine or tryptophan (Sutherland, Toews et al. 2008). Furthermore, trypsin cleavage at crosslinked lysines and arginines could be inhibited thereby impeding the quantitative MS analysis.

A common problem of affinity purifications is the co-enrichment of proteins that unspecifically bind to the affinity matrix. Affinity purifications using the same affinity matrix on chromatin samples not containing the bait protein help

to identify those background binding proteins. In this work, mock GFP-Trap purifications on chromatin extracts from cells that do not express a GFP-tagged protein served to identify these proteins (figure 2.6, wildtype). Furthermore, in this analysis, the washing steps after immunoprecipitation were carried out without addition of detergent in order to preserve as many CID-GFP containing chromatin interactions as possible. In future screens, washing could be more stringent, thereby reducing the number of detected proteins – both specific and unspecific ones. Another way to reduce the number of unspecifically binding proteins in the first place would be to apply native elution conditions that specifically release the bait and its co-associated proteins from the affinity matrix. FLAG-tagged proteins for example can be eluted from immobilized anti-FLAG M2 antibody agarose beads by FLAG peptide competition circumventing the need for denaturing elution conditions like the use of Laemmli buffer (an example is the work of Thomae *et al.* (Thomae, Schade et al. 2013)). By that means, a larger fraction of proteins will be recovered due to a specific association with the bait.

Apart from unspecific matrix binders, general chromatin proteins may also be enriched in CID-GFP chromatin purifications compared to mock purifications. Therefore, quantitative comparisons of purifications of CID-GFP versus H3.3-GFP containing chromatin were performed to discriminate centromere-enriched factors from general chromatin components. However, proteins that are present in both centromeric as well as H3.3-containing chromatin would be filtered out by this calculation. Whether this type of filtering indeed increases the success rate or creates unreasonably frequent false negative assignments remains to be tested.

Surprisingly, only a limited number of kinetochore components were recovered in this study. This may relate to the fact that asynchronous cells were used for preparing the input material. Only around 2 % of asynchronous *Drosophila* Schneider cells are in mitosis (figure 2.24), decreasing the probability to detect kinetochore proteins that are only temporarily associating with centromeres during mitosis. If kinetochore proteins are to be targeted, cells can be accumulated in mitosis for example by treatment of cells with colcemid, as performed by Blower *et al.* (Blower and Karpen 2001). In the work presented here, the focus lied more on proteins binding to centromeres during interphase of the cell cycle, as CENP-C was the only known constitutive centromere-binding protein.

In summary, there are many possibilities to modify the workflow to potentially improve its overall performance. Still, the protocol used here already yielded new insights into the composition of the *Drosophila* centromere.

3.2 A critical view to the obtained AP-MS datasets

The described workflow was repeated in triplicates starting from around $5*10^9$ cells as CID and also CID-GFP in the stable cell line are only expressed at very low levels. Since the localization of this variant is restricted to centromeric chromatin, which covers only a small region of chromosomes, upscaling is necessary and therefore, the replicate experiments have been conducted on different days. Before discussing the means of statistical analysis, the triplicate datasets are examined in more detail.

At first glance, the protein content of the different affinity purifications is surprisingly similar. Almost half of the proteins (49 %) are at least once identified in all three conditions (wildtype, CID-GFP, H3.3-GFP) (figure 2.9). This underlines the common problem of unspecific binding in AP-MS approaches. Still, 149 proteins are identified exclusively in at least one of the CID-GFP purifications. These factors were considered as bona fide centromere proteins. However, there is an even larger number of proteins (261) exclusively identified in mock purifications. This larger number may result from the absence of a bait protein that competes for available binding sites on the affinity matrix resulting in a random binding of non-chromatin associated protein. Indeed, the similarity of the three replicates of mock purifications is less than in the CID-GFP precipitations (31.73 % versus 40.94 %). Although this overlap of identified proteins in the different replicates seems to be low, the percentages of reproducibility obtained are typical for MS experiments (35-60 %, Tabb, Vega-Montoto et al. 2010) and might be challenged even more if, like in this study, replicates were generated and analyzed by MS on different days.

When statistical analysis was performed as described in the results and methods section (chapter 2.4.2, 4.2.5.5), 94 potential centromere factors were determined and IF performed for 32 of them. Despite the fact that chromatin was used as input material, eight out of the 32 tested candidates enriched in CID-GFP containing chromatin showed a cytosolic localization (figure 2.17). Since chromatin isolation involved a disruption of cellular membranes, nuclear and cytosolic proteins might mix and form unspecific interactions. Another study also detected interactions of proteins that are separated by the compartmentalization in the cell and that are located in the nucleus and cytoplasm, for example latheo, a component of the origin recognition complex (ORC3), and Fmr1 (Guruharsha, Rual et al. 2011, cytoplasmic localization of Fmr1 was confirmed in figure 2.17). It is thus a common risk of biochemical isolation and purification procedures to detect non-physiological interactions. For that reason, immunolocalization studies are required to independently test

potential centromere association of candidates picked up by the CID-GFP interactome MS analysis. Another advantage of localization studies is that interactions can be monitored on single cell levels and capture distinct physiological states. The combined use of these complementary techniques is the strength of this study that aimed at finding novel *Drosophila* centromeric proteins.

3.3 Statistical analysis allows a better selection of candidate centromeric proteins

The MS dataset obtained in this work contains quantitative information about interactors of CID-GFP. The gained output is a long list of proteins with different iBAQ values. Therefore, a good way to extract the most confident candidates of true *Drosophila* centromeric proteins is needed. As mentioned in the results section (chapter 2.4.2, table 1, appendix), an easy way to extract factors from the identified proteins is to only consider proteins that were exclusively present in CID-GFP purifications. However, CID itself has been identified in three out of six control purifications (3x H3.3-GFP, 3x mock IPs), albeit to a lower level. Regarding only proteins never identified in any control purification would thus result in a false negative assignment of CID as a non-centromeric protein. Therefore, and for reasons of better comparability between replicates and quantitative ratios, statistical analysis of the data was performed to extract proteins likely to be enriched in centromeric chromatin (described in chapter 2.4.2 and 4.2.5.5). Even though not many centromeric proteins are known so far in *Drosophila*, the most reliable controls Cal1 and CENP-C are present in the list of 85 proteins passing the set threshold (chapter 2.4.2, table 2, appendix). This underlines the feasibility of the approach to identify centromeric proteins. GFP is only enriched in CID-GFP affinity purification when compared to mock purifications (as cells do not express any GFP protein), but not when compared to H3.3-GFP purifications, since here another H3 variant is tagged with GFP. These features serve as quality control for the outcome of the screen.

The statistical analysis resulted in 85 proteins with a rather stringent enrichment of 16-fold compared to the two different controls (wildtype- and H3.3-GFP purification). However, when manually revisiting the data, some factors that were identified repeatedly and almost exclusively in CID-GFP purifications were not included in this list as their enrichment factors were lower than sixteen. Therefore, nine additional factors were included in the list of centromeric factors (Ge-1, asp, CG6227, Hcf, Top3beta, CG8478, RhoGAP54D, ATAC3, CG1399) that were below the threshold due to the

imputation of zero values in control purifications needed for the statistical analysis (marked in grey in table 2, appendix). This generates a more complete picture of proteins present at *Drosophila* centromeres. Indeed, CG6227 was detected on centromeres (figure 2.14) and asp, Hcf and ATAC3 were at least demonstrated to be nuclear (figure 2.15). Without inclusion of the nine additional factors, the centromeric localization of CG6227 would have gone unnoticed in this screen. Also, imputation may well be the reason why only Hsp70Bbb, and not the isoform Hsp70Ab, was sustained in the final 94 candidates (chapter 2.5), even though they share most of the peptide sequences, since only unique peptides are taken into account. In the following chapters, CG6227 and other interesting factors will be discussed.

3.4 Involvement of posttranslational modifications on centromeric proteins

Cell cycle processes need to be well controlled, since the correct timing of events minimizes errors in DNA duplication that might otherwise be deleterious for a cell or even a whole organism. They often involve chronological cascades of protein complexes that need to interact in order to activate or inhibit each other (chapter 1.1). Posttranslational modifications like ubiquitinations and acetylations are involved in cell cycle regulation and it is thus worthwhile to screen potential centromeric proteins for protein modifiers. In this work, a number of acetyltransferases as well as proteins potentially involved in protein degradation were assigned to the *Drosophila* centromere proteome.

3.4.1 Localization of acetyltransferases at centromeres in Drosophila

One of the most strongly enriched factors identified in this screen is CG2051 (figure 2.10, table 2, appendix). It is the homolog of the human histone acetyl transferase 1 (HAT1). HAT1 acetylates histone H4 on lysines 5 and 12 in the cytoplasm immediately after synthesis (Sobel, Cook et al. 1995) and is in a complex with the histone H3/H4 dimer and the chaperone Asf1 for this purpose (Alvarez, Munoz et al. 2011). Additionally, it can acetylate H2A on lysine 5 (Verreault, Kaufman et al. 1998, Tafrova and Tafrov 2014). Once imported into the nucleus, the histone H3/H4 dimers are transferred to the CAF-1 chaperone as a $(H3/H4)_2$ tetramer, thus joining two dimers on CAF-1 (Winkler, Zhou et al. 2012). CAF-1 is responsible for loading $(H3/H4)_2$ tetramers on replicating DNA (Kaufman, Kobayashi et al. 1995). Histone H4 carries the two acetyl marks on lysine 5 and 12 when it is incorporated in chromatin; the marks are then rapidly removed (Annunziato and Seale 1983).

Despite the fact that the *Drosophila* HAT1 homolog is a nuclear protein, it is tempting to speculate that CG2051 cooperates with CAF-1 to ensure the correct acetylation status of H4 on CID/H4 dimers before or during their incorporation into centromere chromatin. Analysis of histone modifications associated with CID-GFP containing chromatin indeed showed a slight enrichment of acetylation marks on histone H4 as compared to control purification (figure 2.8). This finding is in contrast to earlier reports in which centromeric H4 was described to be devoid of acetylation (Sullivan and Karpen 2004); however, in contrast to this MS-based analysis, the mentioned study was performed using antibodies for different acetylations on histone H3 and H4. Recently, it became clear that these antibodies might not be specific for distinct modifications, but rather bind preferably to multiply acetylated H4 (Rothbart, Lin et al. 2012) which might explain the observed discrepancy between the two studies.

Immunofluorescence analysis demonstrated that CG2051 is found in the nucleus and with a clear presence at centromeres in many cases (figure 2.13, 2.14). For figure 2.13 and 2.14, prelysis was performed on cells before fixing them. By that means, most cytoplasmic proteins are washed out of the cell, increasing visibility of nuclear localization patterns. It can thus not be excluded that CG2051 is also present in the cytoplasmic fraction. However, cytoplasmic signal was also not detected in immunolocalization experiments performed without prelysis (figure 2.22, raw data not shown), suggesting that CG2051 might acetylate histone H4 in the nucleus.

Although the acetylation of K5 and K12 on newly synthesized histone H4 is highly conserved, it does not play a major role in chromatin assembly since its mutation or the total lack of H4-tails harboring K5 and K12 do not influence cell proliferation (reviewed in Parthun 2012). This might explain why CG2051 depletion has no major influence on centromere architecture or cell cycle distribution (figures 2.22, 2.24 and 2.26) even though an increase in mitotic defects is observed (figure 2.25). Since asynchronous cells were analyzed, it is possible that CG2051 is only enriched at centromeres at the time of cell cycle when centromeric chromatin is replicated. Also, a redundant acetyltransferase might exist in *D. mel* cells, making it impossible to detect effects with knockdowns of just one HAT. It might therefore be interesting to combine knockdowns of different HATs and re-screen for cell cycle effects or increases in mitotic errors.

Furthermore, the dataset contained members of different acetyltransferase complexes: for example the human YL-1 homolog is a member of the TRRAP/Tip60 complex (Cai, Jin et al. 2005) known to acetylate histone H2A

and H4 and to be involved in a number of cellular processes (reviewed in Squatrito, Gorrini et al. 2006). Additionally, the PHD (plant homeodomain) finger protein encoded by CG9293 that localizes to centromeres (figure 2.14) is the homolog of human ING5 (inhibitor of growth). ING5 is member of two different HAT complexes: p300/CBP and MYST-HATs (reviewed in Tallen and Riabowol 2014) which acetylate either H3 or H4 or the tumor suppressor p53, involving ING5 in cancer. It induces cell cycle arrest and associates with MCM (minichromosome maintenance) proteins that are part of the replicative helicase. Therefore, both of the HAT complex members might indeed directly contribute to centromere function or cell cycle control.

Moreover, three components of the ATAC complex were enriched in centromeric chromatin, namely ATAC3, CG30390 and Hcf (see figure 3.1, Suganuma, Gutierrez et al. 2008). ATAC complex is named after AdaTwoA containing complex and comprises two histone acetyltransferase subunits: Gcn5 and ATAC2 (Suganuma, Gutierrez et al. 2008). The mammalian ATAC complex has been shown to acetylate Cyclin A/Cdk2, resulting in its degradation and thus being implicated in progression through M-phase (Orpinell, Fournier et al. 2010). This together with findings of this work indicates that ATAC complex might have a function in cell cycle regulation.

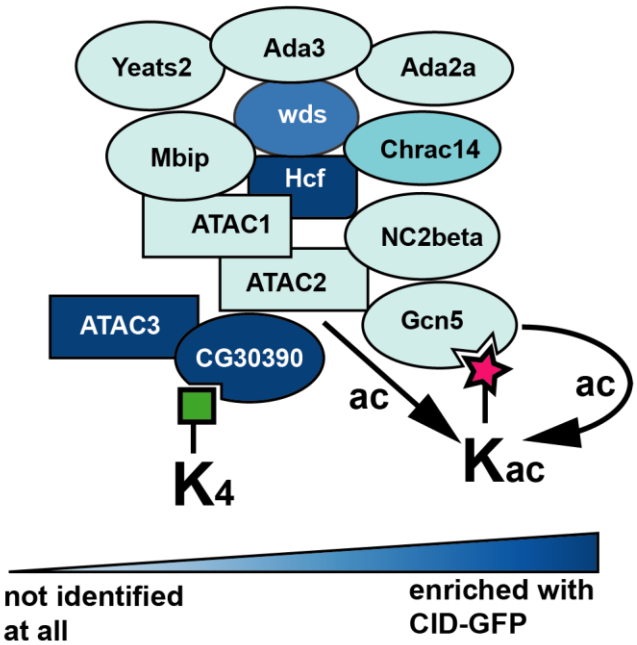

Figure 3.1: Schematic of ATAC complex component quantifications in this work. The color code indicates the specificity of detection in CID-GFP samples – more blue means more specific for CID-GFP immuno-precipitations. Green square: Trimethylation. Pink star: Acetylation. Modified after Spedale, Timmers et al. 2012).

Another study performed in human HeLa cells suggests an implication for histone acetylation and deacetylation in deposition of newly synthesized CENP-A. It was demonstrated that treatments with the histone deacetylase inhibitor Trichostatin A influences centromere priming (Fujita, Hayashi et al.

2007): A complex of three proteins, hMis18α, hMis18β, and M18BP1 binds to centromeres in telophase/G1-phase of the cell cycle, before the reduction of CENP-A levels caused by the previous round of DNA replication is adjusted. If levels of hMis18α are reduced by RNAi, incorporation of newly synthesized CENP-A is hampered. This effect can be suppressed by inhibiting HDACs by Trichostatin A in a concentration-dependent manner, indicating that histone acetylation is important for CENP-A incorporation to centromeres.

Figure 3.2 sums up all potential implications of acetyltransferases on either centromere proteins in *Drosophila* or acetylations of proteins, like Cyclin A, important for cell cycle progression. Question marks indicate hypotheses made based on this work.

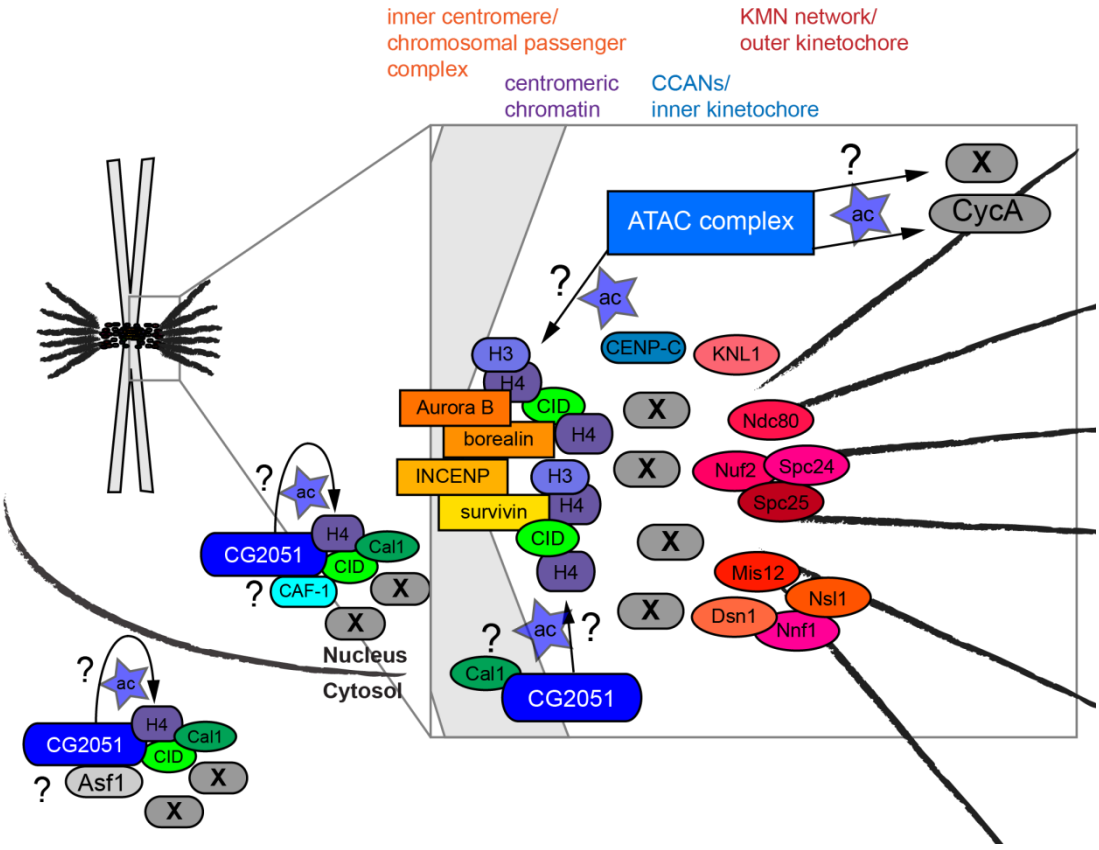

Figure 3.2: Scheme of possible acetylation events at the centromere. CG2051 and ATAC are acetyltransferases that might be involved in acetylating histones associated with CID before chromatin incorporation or at the centromere; or non-histone proteins involved in cell cycle regulation as shown for Cyclin A (CycA). A possible cytosolic function of CG2051 is also illustrated since a potential minor cytosolic localization could not be fully excluded.

Taken together, the AP-MS dataset contains several acetyltransferases or acetyltransferase-associated proteins potentially regulating centromere biology in *Drosophila melanogaster*. This might be due to acetylation of histone proteins in the cytosol or nucleus before chromatin incorporation. Alternatively, histones incorporated in centromeres might get acetylated. Figure 2.8 however only shows a low increase of histone H4 acetylation in comparison to histones originating from the mock IP. All histones detected in mock IPs are unspecifically binding to the resin. It can therefore be assumed that a high amount of background-binding histones is also present in the CID-GFP IPs. That is why a huge difference in histone modification patterns could not be detected. The acetylations of histone H3 peptides were not analyzable. Another possibility for the here identified histone acetyltransferases is that non-histone proteins are their targets.

3.4.2 CID turnover might be regulated by Hyd, a ubiquitin ligase

Additional to the importance of ubiquitination as degradation signal for cell cycle regulatory proteins as mentioned in chapter 1.1, this posttranslational modification might also be implicated directly in balancing levels of CID and/or other centromeric proteins. If CID is overexpressed in cells, it also covers the arms of condensed chromosomes (Olszak, van Essen et al. 2011). This ectopic incorporation can be deleterious to cells since CID alone can recruit a functioning kinetochore (Heun, Erhardt et al. 2006, Olszak, van Essen et al. 2011) and needs to be prevented. The existence of a mechanism involving the proteasome to degrade erroneously incorporated CID has been found (Moreno-Moreno, Torras-Llort et al. 2006). Partner of paired (Ppa), which is a subunit of the E3 ubiquitin ligase SCF, was identified as responsible for CID level regulation in *Drosophila* (Moreno-Moreno, Medina-Giro et al. 2011). Here, Ppa was never identified in a MS run, suggesting that it is not well detectable by MS analysis or only low abundant. As supernumerary kinetochores on a chromosome can lead to chromosomal aberrations, a tightly controlled system to avoid aberrant CID incorporation must exist. Therefore, it is possible that additional factors besides Ppa regulate CID levels *in vivo*. Interestingly, a set of regulatory proteins found in this screen was predicted to be implicated in protein degradation: Hyd (hyperplastic discs), which is an E3 ubiquitin ligase, as well as REG, a proteasome regulator.

Drosophila REG has been shown to increase the trypsin-like digestion activity of the 20S proteasome, whereas it decreases its chymotrypsin-like activity as determined by different fluorophoric peptides used as substrates (Masson,

Andersson et al. 2001). REG colocalizes with CID as determined by immunofluorescence analysis (figure 2.14). Its reduction leads to higher amounts of CID-GFP at centromeres (figure 2.22), which could be due to an accumulation of non-degraded CID-GFP in accordance with REG's stimulating function of the proteasome. Interestingly, the function of REG to control CID levels seems to be restricted to centromeres, as no alterations in global CID levels upon *REG* RNAi were observed in Western blot analysis (figure 2.23). Notably, STRING analysis (figure 2.12) suggests an interaction of the human homologs of REGγ and CG14480 (human homolog: FAM192A, 41.9 % identity) (Doueiri, Anupam et al. 2012). Thus, the strong CID-GFP interactor CG14480 (figure 2.10, table 2, appendix), of which no information is published so far, may cooperate with REG to control CID levels. It remains to be established whether this hypothetical function could explain the observed increase in mitotic errors upon *CG14480* RNAi.

As mentioned earlier, Hyd is another predicted ubiquitin ligase found enriched with CID-GFP (table 2, appendix). Hyd is a putative E3 ubiquitin ligase and is required for regulation of cell proliferation in imaginal discs of *Drosophila* (Mansfield, Hersperger et al. 1994). In addition, *Drosophila hyd* mutant animals are sterile due to germ cell defects. In males, sterility may be caused by chromosome condensation and spindle attachment defects and thus may be linked to an impairment of centromere function or regulation (Pertceva, Dorogova et al. 2010). Hyd is also involved in the Hedgehog signaling pathway as its mutation causes accumulation of Cubitus interruptus, Hedgehog and Decapentaplegic in *Drosophila* eye discs (Lee, Amanai et al. 2002). Allelic imbalances of EDD, the human ortholog of Hyd, are found in a number of human tumors (Clancy, Henderson et al. 2003). In flies, *hyd* reduction leads to elevated mitotic errors (figure 2.25) that might be due to a function in CID amount regulation at centromeres: In a *hyd* knockdown situation, more CID-GFP can be detected at centromeric foci (figure 2.22). Similar to *REG* loss, this increase of CID-GFP is not reflected in global changes of CID levels, arguing for a locally restricted mode of action (figure 2.23). However, experiments employing RNAi-mediated depletion of *hyd* combined with proteasome inhibition revealed that Hyd is responsible for mono-ubiquitination of CID (figure 2.23). Mono-ubiquitination per se does not target proteins for proteasomal degradation, but it can be a substrate for another ubiquitin ligase in charge of adding additional ubiquitin residues. Thus, it is unclear how exactly Hyd could be involved in controlling CID amounts at centromeres.

Figure 3.3 displays a model of how turnover of centromeric proteins might be regulated by factors detected in this work. Hyd was found as a CID-GFP interactor; its localization was determined to be nuclear, not especially overlapping with CID signals (figure 2.15). Therefore, it might also be implicated in degradation of CID that was incorporated outside of centromeres by mistake. Other possibilities are that it also regulates CID levels at centromeres or it ubiquitinates some so far unknown factors besides CID that are implicated in centromere biology.

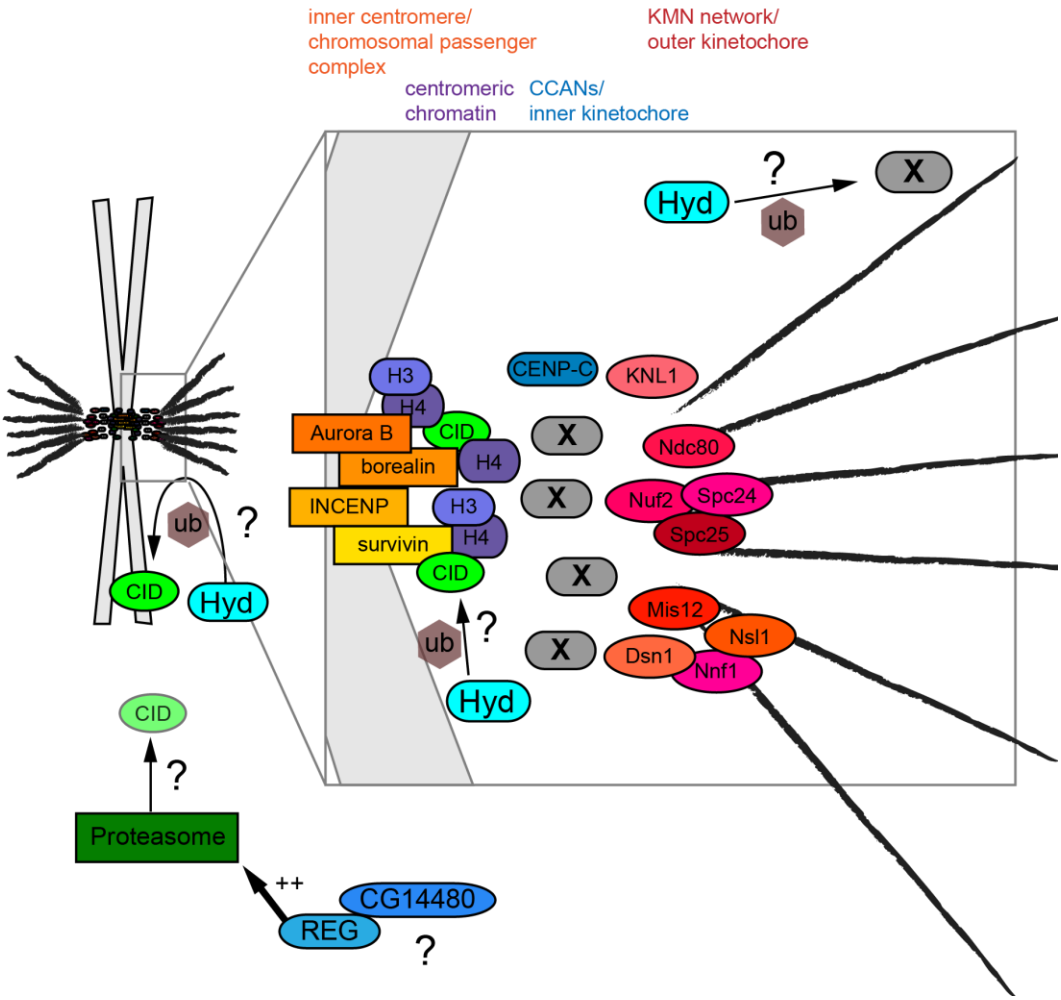

Figure 3.3: Model of protein degradation factors involved in centromere biology. REG might interact with CG14480 as in humans and was found to stimulate the proteasome, maybe for degradation of CID protein. Hyd might be an ubiquitin ligase specific for CID mono-ubiquitination and thereby trigger poly-ubiquitination and subsequent degradation of CID. This might regulate CID levels at ectopic localizations on chromosome arms or also at centromeres. Alternatively, Hyd might also ubiquitinate other proteins involved in centromere biology.

3.5 Immunolocalization proofs presence of additional proteins at *Drosophila* centromeres

3.5.1 CAF-1, Subito and Proliferation disruptor were identified as centromeric proteins as expected

Several proteins fished with the AP-MS strategy had already been reported to interact with CID. CAF-1, for example, was co-purified in *Drosophila* together with TAP-tagged CID (Furuyama, Dalal et al. 2006), suggesting that CAF-1 is responsible for centromeric histone deposition. CAF-1 is a chaperone involved in assembling nucleosomes on newly replicated DNA during S-phase of the cell cycle or DNA repair (Smith and Stillman 1989, reviewed in Loyola and Almouzni 2004). Indeed, *CAF-1* knockdown provokes mitotic errors (figure 2.25) as well as a G2-phase block in the cell cycle (figure 2.24). CAF-1, however, is also implicated in other cellular processes besides chaperoning newly synthesized histones: It is part of the Polycomb group complex PRC2 that methylates H3 on lysine 27, a repressive histone modification. Furthermore, CAF-1 is present in the HAT1 complex, HDAC1 complex and NuRD nucleosome remodeler complex (reviewed in Suganuma, Pattenden et al. 2008). Thus, the phenotypes observed in the assays performed here might originate from other functions of CAF-1.

Another protein enriched in CID-GFP containing chromatin is Subito (table 2, appendix). Immunolocalization showed its colocalization with CID as well as binding to euchromatin (figure 2.14). Subito has been described before as a kinesin motor protein that colocalizes with CID in early metaphase (Cesario, Jang et al. 2006). Together with the chromosomal passenger complex, it is needed for the bipolar orientation of homologous chromosomes (Radford, Jang et al. 2012). The colocalization of CID and Subito was also confirmed here, together with a euchromatic signal in nuclei (figure 2.14). This proofs that also factors like Sub, which might not be interacting with CID during the whole cell cycle, can be enriched with the applied experimental strategy.

Furthermore, Proliferation disruptor (Prod) is enriched on CID-GFP containing chromatin (table 2, appendix). It is here regarded as control since it was published to be localized at centric heterochromatin of the second and third *Drosophila* chromosome (Torok, Harvie et al. 1997). It is implicated in condensation of centric heterochromatin. When determining mitotic errors in a *prod* knockdown, as expected, the values are increasing (figure 2.25), yet not as strong as in *CAF-1* knockdown. This argues that *CAF-1* knockdown, as discussed above, might comprise also centromere biology-unspecific effects.

Prod might also have additional effects as it has been found to bind to telomeric retrotransposons in *Drosophila* (Torok, Benitez et al. 2007). Apart from binding centric heterochromatin, it has around 400 additional binding sites on euchromatin of polytene chromosomes (Torok, Harvie et al. 1997).

3.5.2 CG3548 is a potential component of the outer kinetochore

As discussed in chapter 3.1, the AP-MS workflow was not specially set to trap proteins transiently bound to kinetochores in M-phase. Yet, CG3548 seems to localize to centromeres in a cell cycle dependent manner, as can be seen in figure 2.14: CG3548 exhibits euchromatic localization during interphase and is bound to centromeres during mitosis (figure 2.18). In addition to the centromeric foci, it is also enriched at a few spots on the chromosome arms, arguing for kinetochore-independent functions of CG3548. Based on the very strong focal enrichment, CG3548 might bind and regulate highly repetitive DNA sequences on chromosome arms.

CG3548 has been found to interact with Prod (Guruharsha, Rual et al. 2011), a centromeric protein discussed in chapter 3.5.1. The CG3548 protein contains a Myb/SANT domain (Tweedie, Ashburner et al. 2009) which is predicted to be implicated in nucleosome remodeling as well as reading histone posttranslational modifications (Boyer, Latek et al. 2004), giving a hint about CG3548 functions in chromatin biology. Furthermore, evidence has been found in mouse that M18BP1, a protein required for centromere priming before its regeneration during the cell cycle, also contains such a SANT domain that mediates its interaction with CENP-C (Dambacher, Deng et al. 2012). Unlike loss of M18BP1, CG3548 RNAi does not change CID-GFP signals at centromeres (figure 2.22), suggesting that CG3548 function in *Drosophila* is not related to centromere priming.

3.5.3 CG9418, GFZF, MED30, CG6227 and CG11076 could be involved in centromere biology in Drosophila

Apart from proteins discussed so far, other centromeric localizations have been determined in this work by the combination of AP-MS and immunolocalization analysis (figure 2.14). One of these factors is CG9418, also known as high mobility group protein 2 (HMG-2). It contains a high mobility group box domain, which usually mediates DNA binding or protein-protein interactions (Stros, Launholt et al. 2007). In high-throughput screens, CG9418 has been found to interact with RhoGAP54D (Giot, Bader et al. 2003), a protein contained in the list of 94 centromeric proteins listed in table

2, appendix, as well as with REG, which is also enriched at centromeres seen by both AP-MS (table 2, appendix) and immunolocalization (figure 2.14, discussed in chapter 3.4.2) (Guruharsha, Rual et al. 2011). It might therefore build a complex with REG and CG14480.

GFZF stands for GST-containing FLYWCH zinc finger protein. It contains four FLYWCH domains at its N-terminus and shares a 46 % homology with GST at its C-terminus (Dai, Sun et al. 2004). FLYWCH domains are zinc finger domains responsible for binding nucleic acids or for protein-protein interactions. They were originally found in the Mod(mdg4) proteins that interact with Su(hw) insulator proteins (Dorn and Krauss 2003). The localization of GFZF has been previously determined as mostly cytoplasmic in *Drosophila* Schneider cells and embryos (Dai, Sun et al. 2004); however, the protocol used in this work, as opposed to the earlier results, contains a prelysis step designed to pre-extract the soluble cytoplasmic pool thereby revealing a potential nuclear localization of the protein. Thus, most of the GFZF molecules might reside in the cytoplasm while a small fraction localizes to centromeres in *Drosophila* Schneider cells (figure 2.14). Its depletion by RNAi reduces the number of cells in G1-phase (figure 2.24). The insulator protein Su(hw) also biochemically co-purified with centromere chromatin (table 2, appendix), however, it localizes at distinct foci in the nucleus that do not overlap with centromeric foci (figure 2.15) (Roseman, Pirrotta et al. 1993). The implication of insulator elements to centromere biology is appealing as insulators could make up the border between centromeric and pericentromeric chromatin domains. Whether GFZF is involved in this process remains to be investigated.

Another protein colocalizing with centromere foci in immunofluorescence analysis is MED30 (figure 2.14). MED30 is a subunit of the mediator complex needed for transcription regulation (reviewed in Carlsten, Zhu et al. 2013). Transcription at centromeres is a feature found in many organisms such as *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, mouse and humans (reviewed in Chan and Wong 2012). Studies in *S. pombe* unraveled that subunits of the mediator complex at centromeric regions are in fact required for incorporation of CENP-A via regulation of heterochromatin formation by inhibiting transcription of non-coding RNAs (Carlsten, Szilagyi et al. 2012). In Schneider cells, reducing amounts of *MED30* by RNAi does not seem to induce CID incorporation defects (figure 2.22). Still, the presence of MED30 at centromeres of Schneider cells might be the first indication of transcriptional regulation of (peri)centromeric chromatin in *Drosophila melanogaster*.

In analogy to loss of *hyd* and *REG*, *CG6227* knockdowns translate into higher centromeric CID-GFP levels (figure 2.22). The human homolog of CG6227 is DDX46, a DEAD box protein, so it contains the amino acid sequence aspartate-glutamate-alanine-aspartate (DEAD). DEAD box proteins are usually RNA helicases (Linder 2006) and DDX46 was found to act in pre-mRNA splicing (Will, Urlaub et al. 2002). It is thus a component of the spliceosome. Interestingly, in *S. pombe*, components of the spliceosome are implicated in centromere biology since they regulate formation of siRNA from centromeric repeats needed to establish centromeric heterochromatin to silence underlying DNA repeats (Bayne, Portoso et al. 2008). Spliceosome proteins interact with both members of the RNAi machinery as well as centromeric DNA repeats. No such connection has been made to date in *Drosophila*; however, chromosome segregation defects result if the amount of splicing genes is reduced (Somma, Ceprani et al. 2008). Therefore, CG6227 might be a splicing factor in *Drosophila* implicated in regulation of centromeric (hetero)chromatin.

CG11076, a factor enriched with CID-GFP as determined by AP-MS (table 2, appendix), did not exhibit centromeric localization in immunofluorescence analysis (figure 2.15). Rather, when costaining CG11076-GFP transfected cells with antibodies against Fibrillarin, a typical protein of nucleoli (Ochs, Lischwe et al. 1985), a colocalization of both signals was observed (figure 2.16). The nucleolus is involved in centromere biology since centromere clustering and tethering is taking place at its surface. It has recently been shown that the Nucleoplasmin-like protein (NLP), Modulo, a nucleolar protein, as well as the insulator protein CTCF are required for proper establishment of centromere clusters (Padeken, Mendiburo et al. 2013). It might be possible that CG11076 is involved in this process. However, reducing protein amounts does not lead to a centromere declustering (figure 2.22).

3.6 Only previously identified centromeric proteins cause cell cycle defects

Since the centromere is the assembly platform for the kinetochore, its integrity should be vital for cell cycle progression. It can be envisioned that disturbance of any of the involved factors leads to differences in the cell cycle distribution. Therefore, FACS analysis was employed which measures the DNA content in individual cells that linearly correlates with the different cell cycle phases.

Interestingly, RNAi-mediated depletion of the centromeric factors CID or Cal1 led to an increase in cells with an S-phase like DNA content (figure 2.24,

panel D). This effect is even stronger when levels of the constitutive centromere-associated network protein CENP-C are reduced (figure 2.24, panel D). The fact that Schneider cells depleted of centromeric factors arrest in S-phase is surprising, since these proteins fulfill their functions in mitosis. The findings here suggest the existence of an intra-S phase checkpoint that monitors centromere protein levels or centromere integrity. Biologically, the existence of such a checkpoint is plausible as it prevents the amplification of aneuploid cells which otherwise might lead to cancer. Of course, the precise nature and key molecules contributing to this pathway need to be resolved. It will be interesting to see whether CENP-C functions upstream of CID and Cal1 given the stronger S-phase arrest phenotype its loss causes.

RNAi-mediated depletion of selected novel centromere associated factors identified in this work did not result in a similar S-phase arrest phenotype. However, as knockdown of neither of the factors causes reduction of CID levels at centromeres (figure 2.22), this result is not unexpected. In this respect, it is probable that no additional bona fide CCAN protein was found in this screen. It remains unclear whether this is due to technical challenges or if *Drosophila* harbors a less complex centromere with the little number of already known proteins taking over tasks that are necessary for functional centromeres. As discussed above, this question can potentially be resolved by employing alternative experimental settings. Hence, the future challenges will be to further improve the AP-MS strategy in order to find more centromere or kinetochore proteins in *Drosophila*; and to unravel the precise centromeric functions of the novel centromeric factors discovered in this work. As the classical assays to study centromere biology failed to reveal mechanistic details, their task is not related to CCAN maintenance or kinetochore formation. Still, the list of proteins co-purifying with CID-GFP determined in this work offers a repository for scientists investigating centromere biology in *Drosophila.* As mentioned earlier, findings from *Drosophila* have already been transferred also to other model organisms and thus can lead to future insights into novel centromeric functions. It will be interesting to follow future discoveries based on the results obtained here.

4 MATERIAL AND METHODS

4.1 Material

4.1.1 Stable cell lines and vectors

4.1.1.1 Stable cell lines expressing GFP-tagged H3 variants

The histone H3 variants CID, H3.2 and H3.3 tagged with GFP were transfected in L2-4 cells, a derivative line of *Drosophila* Schneider SL2 cells (Schneider 1972) (karyotype: two X-chromosomes, four 2nd chromosomes plus 2L, four 3rd chromosomes, two 4th chromosomes; phenotype male) and transfectants were selected based on the hygromycin resistance encoded by the vector (figure 4.1). Stable cell lines possess similar expression levels of tagged proteins and thus offer more comparable conditions than transiently transfected cells. The tagged variants are under control of a metallothionein promoter (pMT) with basal expression that can be induced 30 to 100 fold by addition of copper sulfate (Bunch, Grinblat et al. 1988). For all mass spectrometry and immunofluorescence experiments, the basal promoter activity without copper sulfate induction was used to avoid overexpression and thereby the formation of potential unspecific interactions. Figure 4.3 shows that upon copper sulfate induction of the CID-GFP transfected cell line, CID-GFP in many cells localizes throughout the whole nucleus instead of being restricted to the characteristic centromeric foci usually seen (for example, figure 2.3). Instead of inducing CID-GFP expression in cells, a larger amount of cells was taken for immunoprecipitations.

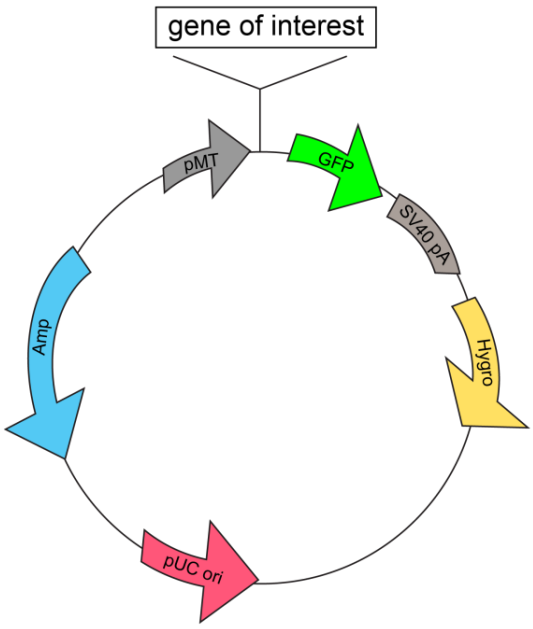

Figure 4.1: Outline of vector used to establish stable cell lines of histone H3 variants with GFP-tags. The gene of interest expression is driven by an inducible pMT promoter. The GFP tag is added on the C-terminus. Hygro: Hygromycin resistance gene. pUC ori: origin of replication. Amp: Ampicillin resistance gene. SV40 pA: polyadenylation signal from SV40 virus.

4.1.1.2 Other GFP-fusion constructs

Based on statistical analysis, 94 factors were assigned centromeric and 32 of these were checked additionally by immunolocalization (chapter 2.5). To do so, their cDNA sequences were cloned in an expression vector suitable for *Drosophila* tissue culture expression and supplied with a tag. Three different tags were used: Either a GFP-tag as for the stable H3-variant cell lines (chapter 4.1.1.2), N-terminal (chapter 4.1.1.3) or C-terminal FLAG/HA-tag (chapter 4.1.1.4), or v5-tag (for CG6227 and CG1399). All coding sequences of vectors utilized in this work were confirmed by sequencing. Table 4.1 lists GFP-tagged factors and also indicates if the construct was used to develop a stable cell line. In that case, all figures shown in the results section were obtained from the stable cell lines. The plasmid used here is indicated as “GFP Invitrogen” in table 2 of the appendix. Stable cell lines were established by Georg Schade (chapter 4.2.1).

Table 4.1: Other GFP-fusion constructs.

Construct	Stable cell line
CG1091-GFP	available
CG11076-GFP	available
CG14480-GFP	available
CG2051-GFP	available
CG3548-GFP	available
Gfzf-GFP	
Hcf-GFP	
MED30-GFP	
su(hw)-GFP	

4.1.1.3 N-terminal FLAG/HA-tagged candidates

Figure 4.2 illustrates the schematic organization of the expression vector into which cDNAs of different factors were cloned. This vector was established by Dr. Andreas Thomae. It allows expression of an N-terminally tagged protein in *Drosophila* cells. Table 4.2 gives an overview about potential centromeric factors cloned in this vector for this work. The hygromycin resistance gene contained in the vector allowed for selection of transfected cells while for transient transfections in immunolocalization studies, the pMT promoter could be induced by addition of copper sulfate, as indicated above. This vector is denoted as “pMT-hyg-FLAG-HA-Nterm” in table 2 of the appendix. Immunolocalization was then performed using an anti-FLAG antibody.

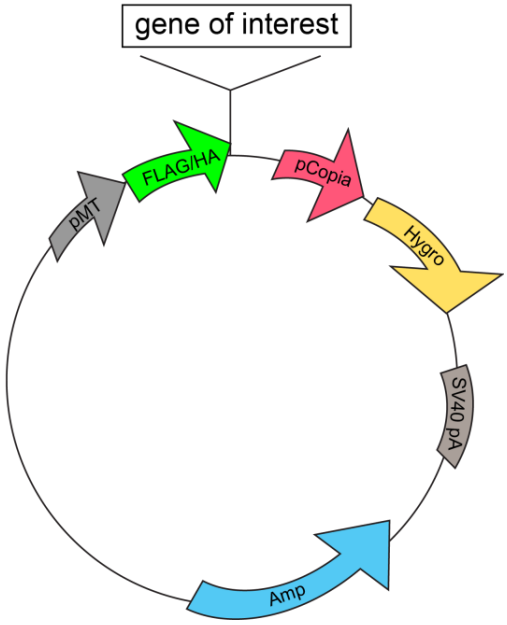

Figure 4.2: Vector map of FLAG/HA(FH)-tag expression vector. The vector was established by Dr. Andreas Thomae. The expression of the gene of interest is driven by an inducible pMT promoter. A FLAG/HA-tag is added at the N-terminus. Amp: Ampicillin resistance. Hygro: Hygromycin resistance, driven by pCopia promoter. SV40 pA: polyadenylation signal from SV40 virus.

Table 4.2: Constructs with N-terminal FLAG/HA-tag.

Construct	Construct
FH-hyd	FH-Cdk12
FH-CG4972	FH-CG7518
FH-Hsp70Ab	FH-asp
FH-CG30390	FH-ATAC3
FH-Rbcn-3A	

4.1.1.4 Expression vectors from BDGP

If available, cDNAs were ordered in expression vectors from the Berkeley Drosophila Genome Project (BDGP). They were then integrated in the pMK33-CFH-BD expression vector which encodes the coding sequence plus a C-terminal FLAG/HA(FH)-tag. Table 4.3 gives an overview about ordered factors used in this work. Immunolocalization was performed with anti-FLAG antibody.

Table 4.3: Constructs ordered from BDGP containing a C-terminal FLAG/HA-tag.

Construct	Construct
CG14480-FH	CG12343-FH
REG-FH	CG9418-FH
Ifc-FH	Fmr1-FH
CG9293-FH	RagC-FH
Tsp42Ed-FH	YL-1-FH
CG1265-FH	Sub-FH

The following cDNAs (table 4.4) were received from BDGP in other vectors and recloned to expression vectors as indicated above. CG1399 and CG6227 were cloned into a pIB-v5 expression vector (Invitrogen). Also here, all coding sequences were verified by sequencing.

Table 4.4: Non-expression vectors ordered from BDGP.

Construct	Vector
CG1091	pFlc-1
Su(hw)	pBS SK-
CG6227	pOT2
gfzf	pOT2
CG3548	pOT2
CG1399	pFlc-1
MED30	pOT2

4.1.2 Reagents

Reagent	Provider
1 kb DNA ladder	New England Biolabs
100 bp DNA ladder	New England Biolabs
ACN (acetonitrile)	Roth
Ambion® MEGAscript® RNAi kit	Life Technologies
Ammoniumhydrogencarbonate	Roth
Ampicillin	Roth
Aprotinin	Genaxxon
BSA (bovine serum albumin)	Sigma
Coomassie G250	Serva
$CuSO_4$	Sigma
DAPI	Life Technologies
DMSO (dimethyl sulfoxide)	Sigma
DTT (dithiothreitol)	Roth
FCS (fetal calf serum)	Sigma
Formaldehyde methanol –free	VWR
Glycine	VWR
HEPES	VWR
Hygromycin B	Invitrogen
Image-iT FX signal enhancer	Invitrogen
In-Fusion® HD cloning plus	Clontech
Leupeptin	Genaxxon
NGS (normal goat serum)	Dianova
Nitrocellulose membrane (Amersham® Protran)	GE Healthcare
Penicillin/Streptomycin	PAA
Pepstatin A	Genaxxon
PMSF	Sigma
Propidium iodide	Sigma
Propionic anhydride	VWR

Protein assay Dye Reagent Concentrate (Bradford assay)	Biorad
Protein marker V	Peqlab
Proteinase K	Roche
PVP (Polyvinylpyrrolidone)	Sigma
QiaQuick Gel Extraction Kit	Qiagen
RNase A	Sigma
Schneider cell medium	Life technologies
Silver nitrate	Sigma
SlowFade	Invitrogen
SNAP-cell® TMR-star	New England Biolabs
TFA	VWR
TopTip Carbon	Glygen
Triton X-100	Sigma
Vectashield mounting medium	Vector Labs
X-tremeGENE HP DNA transfection reagent	Roche

4.1.3 Enzymes

Enzyme	Provider
AvrII	New England Biolabs
Bsp120L	Thermo Scientific
EcoRV	New England Biolabs
HindIII	New England Biolabs
KpnI	New England Biolabs
MNase	Sigma
NotI	New England Biolabs
Q5® high-fidelity DNA polymerase	New England Biolabs
SuperScript® III reverse transcriptase	Life technologies
T4 DNA ligase	New England Biolabs
Taq polymerase	VWR
Trypsin	Promega
XbaI	New England Biolabs

4.1.4 Antibodies and beads

Antibody	Provider	Dilution	Comments
Primary:			
Chicken anti-CID	Patrick Heun	IF: 1:100	Raised against full-length CID fused to a his-tag
Chicken anti-Ndc80	Tom Mareska	IF: 1:100	
GFP-Trap® agarose beads	ChromoTek		IP
Mouse anti-GFP	Roche	WB: 1:1000	Clones 7.1 and 3.1
Mouse anti-Polo	Claudio Sunkel	IF: 1:50	

Mouse anti-Tubulin	Sigma	IF: 1:100	
Protein A-coupled sepharose beads	GE Healthcare		Preclear Sepharose 4, fast flow
Rabbit anti-CAF-1	James Kadonaga	IF: 1:250	
Rabbit anti-CENP-C	Christian Lehner and Stefan Heidmann	IF: 1:100	
Rabbit anti-CG30390	Jerry Workman	WB: 1:1000	
Rabbit anti-CID	Active Motif	IF: 1:250	Immunoblot grade
		WB: 1:500	
Rabbit anti-FLAG	Sigma	IF: 1:500	Polyclonal
Rabbit anti-H3	Abcam	WB: 1:5000	
Rabbit anti-H3S10ph	Abcam	FACS: 1:2000	
Rabbit anti-ROD	Roger Karess	IF: 1:100	
Rat anti-CID 7A2 and 4F8	Elisabeth Kremmer	IF 7A2: 1:100	Rat IgG2a
Rat anti-INCENP	Patrick Heun	IF: 1:100	N-terminal fragment cloned by Mar Carmena
Secondary:			
Donkey anti-rabbit Alexa 647	Jackson Immuno Research	IF: 1:1000	Preabsorbed
Donkey anti-rat Alexa 488	Jackson Immuno Research	IF: 1:200	Not preabsorbed
Donkey anti-rat Cy3	Jackson Immuno Research	IF: 1:500	Preabsorbed
Goat anti-mouse Alexa 488	Molecular Probes	IF: 1:2000	Preabsorbed
Goat anti-mouse Alexa 647	Life technologies	IF: 1:500	Preabsorbed
ECL anti-mouse IgG, HRP-linked whole antibody	Amersham	WB: 1:5000	
ECL anti-rabbit IgG, HRP-linked whole antibody	Amersham	WB: 1:5000	

To test the different subclones of the rat-anti-CID antibodies provided by Elisabeth Kremmer from the antibody facility at the Helmholtz Center in Großhadern/Munich, cells stably transfected with CID-GFP were used (figure 4.3). Expression of CID-GFP controlled by the pMT-promoter was induced by adding 250 µM $CuSO_4$ for 24 hours before fixation of cells and performing the immunolocalization. Figure 4.3 indicates that the 7A2 antibody specifically binds to untagged endogenous CID, and not to CID-GFP, since only centromeric patterns are visible even when CID-GFP is overexpressed and

exhibits localization throughout the nucleus (second panel). This is not true for the 4F8 subclone which gives a similar localization pattern to GFP and also seems to be less sensitive.

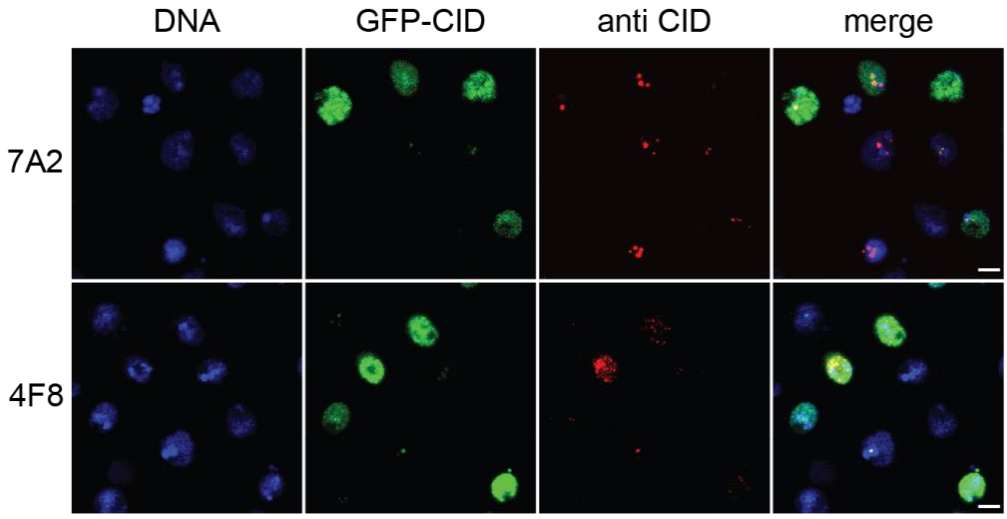

Figure 4.3: Test of reactivity of monoclonal rat antibodies upon copper-induction of stable CID-GFP expressing cells. Either the monoclonal with the number 7A2 or 4F8 was used as primary antibody and the same secondary antibody in both cases. Single plane images are shown, pictures not deconvolved. Scale bar represents 3 µm.

4.1.5 Primers for generating dsRNAs for knockdowns

The website SnapDragon (www.flyrnai.org/cgi-bin/RNAi_find_primers.pl) was used for assistance with primer design. dsRNA length was defined to be around 500 bp. Custom oligos were ordered via Sigma-Aldrich.

Oligo specificity	Sequence
CAF-1 1 fw (CID loading/FACS)	TTAATACGACTCACTATAGGGAGACCGTTCCTGTACGATCTGGTCAT
CAF-1 1 rev (CID loading/FACS)	TTAATACGACTCACTATAGGGAGAATCAGAGGCTGACAGCAAATAGCC
CAF-1 2 fw (CID loading/FACS)	TTAATACGACTCACTATAGGGAGAGACGTGGCTTGGCATCTGCTAC
CAF-1 2 rev (CID loading/FACS)	TTAATACGACTCACTATAGGGAGAGAGAAATCGCTAATCTTGGCAGTGTG
CAF-1 fw (mitotic defects)	TAATACGACTCACTATAGGGCGCGTGATCAATGAGGAGTA
CAF-1 rev (mitotic defects)	TAATACGACTCACTATAGGGAACGCTGGCGATGAGTAGAT
Cal1 fw	TTAATACGACTCACTATAGGGAGACGGTGGTGGACGAGGAAACACT
Cal1 rev	TTAATACGACTCACTATAGGGAGACGACGACTCTGCTGGATGTAGCG

CENP-C 1 fw	TTAATACGACTCACTATAGGGAGACAAGCTTGCCGAAAATAAGCCGG
CENP-C 1 rev	TTAATACGACTCACTATAGGGAGATTTCTCTGTGCAAGGTGTGCTGCTTATTTC
CENP-C 2 fw	TTAATACGACTCACTATAGGGAGAATCCCTTGGCCTGAGTACCTTGACGTG
CENP-C 2 rev	TTAATACGACTCACTATAGGGAGATTCGCTTGTTTCATGCTACGTTTTTGGTATG
CG11076 fw	TTAATACGACTCACTATAGGGAGAGAGCGAGGCACTGGGCTATTTG
CG11076 rev	TTAATACGACTCACTATAGGGAGAGAATTAAATTTTTTCTTGCGCATTTTG
CG14480 fw	TTAATACGACTCACTATAGGGAGAATGAGTTCGGGCTTTGTGACTGAAGC
CG14480 rev	TTAATACGACTCACTATAGGGAGAGTCTTCTGAGTGCTACGCCCGACC
CG2051 fw (CID loading)	TTAATACGACTCACTATAGGGAGAACACCACCGTATACGAGTACTACGCCTATC
CG2051 rev (CID loading)	TTAATACGACTCACTATAGGGAGACAGGCGAGCACGCAGCATTTC
CG2051 fw (mitotic defects)	TAATACGACTCACTATAGGGTCCACGAACCACAAGATGAA
CG2051 rev (mitotic defects)	TAATACGACTCACTATAGGGCATCTACTTGGGCGTCGATT
CG3548 fw	TTAATACGACTCACTATAGGGAGAGAAACGCTCCCTTAACTGGGAG
CG3548 rev	TTAATACGACTCACTATAGGGAGACTCGCTGGCGTTCGTAGAGTTG
CG6227 fw	TTAATACGACTCACTATAGGGAGAGTCGATAAGCAGGAGAACGCTGAC
CG6227 rev	TTAATACGACTCACTATAGGGAGAGATGTCCTCCTCATCATCGGAATC
CID fw	TTAATACGACTCACTATAGGGAGACCGTCGGCGAACAACTCAAAGT
CID rev	TTAATACGACTCACTATAGGGAGAAGCCGCTGCGTCAAGTACATCTC
Gfzf fw	TTAATACGACTCACTATAGGGAGACATCAATGTGGACTTTTGCGCC
Gfzf rev	TTAATACGACTCACTATAGGGAGAATGTGTGACATATCCGGTGGGTTC
GST fw	TTAATACGACTCACTATAGGGAGAAGTTTGAATTGGGTTTGGAGTTTCC
GST rev	TTAATACGACTCACTATAGGGAGATCGCCACCACCAAACGTGG
hyd fw	TTAATACGACTCACTATAGGGAGACCATGCATTCTTTGCTCGATCTG
hyd rev	TTAATACGACTCACTATAGGGAGACTATGATAACCTCTGGTCGTGTGGC
MED30 fw	TTAATACGACTCACTATAGGGAGACTATCTCGCAGCAAAATCCTCACAAG
MED30 rev	TTAATACGACTCACTATAGGGAGACAGCATGGTGTTGATCTCCCAG
prod fw (mitotic defects)	TAATACGACTCACTATAGGGGCTGCACAGTACAAAGCTGG
prod rev (mitotic defects)	TAATACGACTCACTATAGGGGCTAATGCTGATCCTTTCGC
prod fw (CID loading/FACS)	TTAATACGACTCACTATAGGGAGAACGAGGACATGATGGAGGAG
prod rev (CID loading/FACS)	TTAATACGACTCACTATAGGGAGATATAAGGACGGCGGATCGTA
Rbcn-3A fw	TTAATACGACTCACTATAGGGAGACTAGCACACAAAGTGCTTCGCAAT
Rbcn-3A rev	TTAATACGACTCACTATAGGGAGAGTTGTCTCACATCGAAGACGCAA
REG fw	TTAATACGACTCACTATAGGGAGACGTTGATCCTCAAGGCAGAGC
REG rev	TTAATACGACTCACTATAGGGAGAATAATCATCGATGTGCGGATACTTG
White fw	TTAATACGACTCACTATAGGGACTGCTCAATGGCCAACCTGTGGAC
White rev	TTAATACGACTCACTATAGGGCCTCGGCCATCAGAAGGATCTTGTC

4.2 Methods

4.2.1 Tissue culture

Cell maintaining:
If not mentioned otherwise in the text, the *Drosophila* Schneider S2 cell line subclone L2-4 has been used in this work (exception: Kc167 cells for SNAP experiments). Cells were maintained in Schneider medium (Life technologies) supplemented with 10 % FCS (Sigma) and Penicillin and Streptomycin (PAA) in order to avoid bacterial contaminations. Cells were kept in Greiner flasks at 26 °C and split to 1-2*10^6 cells per milliliter every two to three days. For harvesting cells, they were spun down at 1000 g for 20 minutes.

Cell freezing and thawing:
To freeze cells, logarithmically growing cultures at 100 % confluency were spun at 170 g for six minutes and diluted in 50 % FCS/10 % DMSO/40 % Schneider cell medium. The solution was put in cryovials (Thermo Scientific) in freezing racks (Fisher Scientific) with a cooling rate of 1 °C per minute over night at -80 °C and then transferred to liquid nitrogen tanks for long-term storage. Thawing was performed quickly in 15 mL Schneider medium, cells seeded in a Greiner flask and after settling of the cells (approx. 1 hour), the medium was replaced with fresh one in order to avoid cell damage by DMSO.

Knockdowns:
For performing knockdowns, 1*10^6 cells in log-phase were seeded in six well plates in 1 mL serum-free medium. 10 µg dsRNA were added. Plates were carefully shaken for 10 minutes at room temperature and subsequently incubated 50 minutes at 26 °C. Each well was then supplied with 2 mL serum-containing medium. Knockdowns were generally carried out for six days total incubation time. dsRNA was generated using the Ambion MEGAscript® RNAi kit according to manufacturer's instructions.

Transient transfections and stable cell lines:
For transfections, 3*10^6 cells were seeded in six well plates in 1.5 mL Schneider medium. 2 µg of plasmid DNA diluted in 100 µL medium were pre-mixed with 5 µL transfection reagent (X-tremeGENE HP DNA Transfection Reagent, Roche) in an Eppendorf tube, incubated for 10 minutes, 500 µL of Schneider medium added and carefully pipetted under the surface of the medium. Transfections were incubated for 24 hours before 250 µM copper sulfate was added to induce transcription of the transfected gene. The protocol for immunofluorescence was performed 24 hours later. To create stable cell lines, cells were selected after transfection and without induction

for four weeks in Schneider medium supplemented with 100 µg/mL Hygromycin B (Invitrogen).

4.2.2 DNA methods

Cloning:
The open reading frames of hyd, CG7518, Rbcn-3A and asp were PCR amplified from *Drosophila* cDNA and cloned into the expression vector described in chapter 4.1.1.3 using the in-Fusion HD cloning kit (Clontech). Since they comprise very long sequences, they were first subcloned in two pJet1.2 vectors by the in-Fusion strategy, excised with respective restriction enzymes (see below) and ligated into the expression vector. Additionally, Cdk12, ATAC3 and CG4972 were PCR amplified from cDNA, cloned into pJet1.2 vector by in-Fusion reaction and inserted in the vector by restriction and ligation. CG30390 and Hsp70Ab were directly cloned into the vector after PCR amplification from cDNA.

Restriction sites added by PCR reaction to the different cDNAs for cloning:

Target gene	Added restriction sites
Asp	1st *Bsp120L*/*HindIII*, 2nd *HindIII*/*XbaI*
ATAC3	*XbaI*/*NotI*
Cdk12	*XbaI*/*NotI*
CG30390	*XbaI*/*NotI*
CG4972	*XbaI*/*NotI*
CG7518	1st *Bsp120L*/*KpnI*, 2nd *KpnI*/*XbaI*
Hsp70Ab	*NotI*/*AvrII*
Hyd	1st *NotI*/*XbaI*, 2nd *XbaI*
Rbcn-3A	1st *NotI*/*HindIII*, 2nd *HindIII*/*XbaI*

1st and 2nd denote insertions in two individual pJet1.2 intermediate vectors

Other constructs were ordered from BDGP and served as template for PCR-based amplification and subsequent cloning into the pMT-hyg-FLAG-HA-Nterm, pIB-v5 or a pIB-GFP expression vector (see tables 4.1/4.2/4.3/4.4). All sequences obtained by cloning or ordering were verified by sequencing. Cloning into pIB-GFP/v5 vectors was performed by Georg Schade. The vectors containing H3.2-GFP, H3.3-GFP and CID-GFP used to generate stable cell lines were cloned by Patrick Heun (H3.2-GFP and CID-GFP cell lines described in Heun, Erhardt et al. 2006).

Generation of cDNA from *Drosophila* L2-4 cell mRNA:
cDNA was generated using SuperScript III reverse transcriptase (Life technologies) with random as well as oligo-dT primers according to manufacturer's instructions in 40 µL reaction volume. RNase H digestion was performed afterwards.

In-Fusion reaction:
For in-Fusion reactions, the pJet1.2 vector was linearized using *EcoRV* enzyme and supplied with the PCR amplified inserts in a ratio determined by the Clontech ratio calculator. The exact concentration of inserts as well as pJet1.2 vector was determined by gel electrophoresis and comparison with the 2log-ladder (New England Biolabs, see below). The end volume of the in-Fusion reaction was 5 µL. After the in-Fusion reaction, the resulting vectors were transformed into competent Stellar cells (Clontech): 20 µL of competent cell solution plus DNA were kept on ice for 30 minutes, heat shocked for 45 seconds at 42 °C in a waterbath, 250 µL of SOC medium added and kept at 37 °C shaking before plating on an ampicillin-containing medium plate.

Restriction digests:
For restriction digests of DNA, the DNA was mixed with water and 10x buffer 4 of New England Biolabs, BSA added to a concentration of 1 mg/mL and the ideal unit amount for the enzymes calculated. The reaction was carried out at the required temperature (mostly 37 °C) until 2–10 fold overdigestion as determined by calculation was achieved.

Ligations:
For ligation of an insert into a digested vector, 40 ng of the plasmid were used and three times more moles of the insert added. The ligation was performed for 30 minutes at room temperature using 320 U T4 DNA ligase (New England Biolabs) and supplemented with the required volume of 10x ligation buffer.

PCR protocol:
PCRs were performed with the Q5 polymerase from New England Biolabs since it offers a low error rate.
The reaction was separated in three aliquots of 25 µL and run in a gradient thermocycler at different annealing temperatures between 60 °C and 66 °C (step 3 in the PCR conditions)

For difficult amplifications from the whole cDNA-pool, the following exemplary protocol was applied:

15 µL	5x Q5 reaction buffer
1.5 µL	10 mM dNTPs (final concentration: 200 µM)
3.75 µL	forward primer 10 µM, final concentration 0.5 µM
3.75 µL	reverse primer 10 µM, final concentration 0.5 µM
1.25 µL	template cDNA, corresponding to 250 ng dsRNA used for cDNA synthesis
0.75 µL	Q5 high-fidelity DNA polymerase (final concentration 0.02U/µL)
15 µL	5x Q5 high GC enhancer
34 µL	nuclease-free water
final volume:	75 µL

PCR conditions:
1. 98 °C 30 seconds
2. 98 °C 10 seconds
3. 60 ° to 66 °C 20 seconds
4. 72 °C 40 seconds per kb
5. 72 °C 5 minutes
6. 4 °C

Steps 2-4 were repeated 35 times.

Amplified DNA was size separated by electrophoresis using a 1 % (w/v) agarose gel (Bio&SELL) in TBE (0.9 M Tris, 0.9 M Boric acid, 32 mM EDTA) buffer. The DNA band at the right size was excised from the gel and purified using the Qiagen QiaQuick Gel extraction kit.

DNA quantification:
DNA quantification was either carried out with a Nanodrop device (ND-1000 UV Spectrophotometer, Peqlab) at a wavelength of 260 nm or 1 µL of the DNA was applied on a 1 % agarose gel and 10 µL of the 2log-ladder (New England Biolab) which enables good estimation of DNA amounts.

Preparation of DNA for gel electrophoresis:
For RNA digestion, 50 µg/mL RNase A (Sigma) and 0.5 % (w/v) SDS were added to nuclear pellets for 30 minutes at 37 °C and subsequently, proteins were digested by adding 0.5 mg/mL proteinase K (Roche) for two hours at 65 °C. For ethanol precipitation, 1/10 volume of 3 M sodium acetate pH 5.2 was added followed by 2.5 volumes of ice-cold ethanol. Precipitation was performed at -20 °C for at least 30 minutes. Precipitated DNA was collected in a pellet by 30 minutes of centrifugation at maximum speed. The pellet was washed in 70 % ethanol and the DNA finally redissolved in TE (10 mM Tris/1 mM EDTA) buffer.

4.2.3 Protein methods

SDS-PAGE:
To distinguish proteins by size, SDS-PAGE analysis was performed. Either pre-cast polyacrylamide gradient gels from 4-20 % PAA (Expedeon, RunBlue) were used or self-cast 15 % PAA gels for MS analyses (polyacrylamide solution: Rotiphorese, 37.5:1 acrylamide/bisacrylamide (Roth), separation gel buffer (375 mM Tris/HCl pH 8.8), stocking gel buffer (125 mM Tris/HCl pH 6.8), 0.05 % TEMED: Roth, 0.1 % APS: Roth). Protein samples were mixed with Laemmli buffer (final concentration: 50 mM Tris/HCl, pH 6.8, 2 % SDS, 10 % glycerol, 0.05 % bromophenol blue, 0.15 M β-mercaptoethanol; except for elution from GFP-Traps for which all concentrations were doubled) and boiled for 5 minutes (elution: 10 minutes) at 95 °C before application for electrophoresis. Gel chambers were XCell Sure Lock from Invitrogen and gels were run at 130 V to 180 V for varying times.
To stain proteins with Coomassie Brilliant Blue, the gels were shaken for 20 minutes in staining solution (50 % methanol, 10 % acetic acid, 0.25 % Coomassie G-250 (Serva)) directly after size separation electrophoresis. Destaining was performed in 10 % acetic acid at room temperature until protein bands were visible. Silver staining was performed according to the protocol of Blum *et al*. (Blum, Beier et al. 1987).

Western Blotting:
Western blot analysis was performed using a wet blot chamber (Mini-Protean Tetra System, Biorad). Proteins were transferred to a nitrocellulose membrane (Protran®, Amersham, GE Healthcare) at 300 mA for two hours at 4 °C in a transfer buffer containing 15 % (v/v) methanol, 0.02 % (w/v) SDS, 25 mM Tris, 192 mM glycine. Before incubation with the primary antibody, the membrane was blocked with 5 % (w/v) milk powder (Heirler) in PBS for 30 minutes. Incubation with the primary antibody was performed at 4 °C over night. The membrane was then washed with PBS/0.1 % (v/v) Tween 20 (Sigma) three times for 10 minutes and blocked again in 5 % (w/v) milk/PBS for 30 minutes before the secondary antibody was added (usually diluted in 1 % (w/v) milk/PBS or 3 % (w/v) BSA/PBS). The secondary antibody was coupled with horseradish peroxidase. It was used in a 1:5000 dilution in 1 % (w/v) milk/PBS and incubated for one hour at room temperature. The membrane was washed in PBS/0.1 % (v/v) Tween 20 again three times for 10 minutes, incubated in enhanced chemiluminescence solution (Amersham ECL Prime Western Blotting Detection Reagent, GE Healthcare) for five minutes and chemiluminescence signals were visualized with Fuji X-ray films (Röntgen Bender) on an Agfa Curix 60 developer machine.

Protein quantification:
To determine concentration of proteins in a solution, a Bradford assay was conducted using the Protein assay Dye Reagent Concentrate (Biorad). For this purpose, 800 µL of ddH_2O were mixed well with 1.5 µL of protein solution and 200 µL of the dye added. After five minutes of incubation, absorption at 595 nm was measured on a UV Visible Spectrophotometer (Ultrospec 2000, Pharmacia Biotech) and compared to a standard curve of five known BSA concentrations.

Protease inhibitors and reducing agent:

Name	Stock	Working concentration
Aprotinin	1 mg/mL in H2O	1 µg/mL
Leupeptin	1 mg/mL in H2O	1 µg/mL
Pepstatin	0.7 mg/mL in ethanol	0.7 µg/mL
PMSF	0.2 M in isopropanol	0.2 mM
DTT	1 M in 10 mM sodium acetate pH 5.2	1 mM

4.2.4 GFP affinity purifications

A purification of centromeric chromatin has to be performed on soluble chromatin material as an input. Digestion of chromatin by MNase is a well-established method to generate such chromatin fragments. The aim was to not overdigest chromatin during MNase treatment, but still release as much CID-GFP as possible. Overdigestion might lead to loss of interactions as nucleosomal proteins and their binding partners might fall apart. The packing degree and structure of centromeric chromatin is not yet clear, so the MNase conditions had to be adjusted for this work. Therefore, different conditions for the MNase digest were assayed on the nuclear pellet of CID-GFP expressing cells, resulting in a nucleosomal array of mostly mononucleosomes to higher chromatin sizes (figure 4.4A). The respective fractions were tested for CID-GFP amounts in either the supernatant after MNase digestion and centrifugation or the remaining non-soluble protein in the pellet (figure 4.4B). No strong difference of released CID-GFP was observed with different concentrations of MNase. An MNase concentration corresponding to 1x in figure 4.4B was chosen, as small chromatin fragments of mostly mono- and dinucleosomes were generated here (figure 4.4A). This may also increase the specificity of purifying centromere chromatin, as larger chromatin fragments bear the risk of reaching into neighbouring non-centromeric chromatin. Still, one experiment with longer arrays of nucleosomes (SID1508), prepared with an MNase concentration corresponding to 1/10 dilution (figure 4.4B), was performed to compare MS results to the ones of shorter arrays. However, the

most enriched factors of CID-GFP versus wildtype purification were the same in both MNase conditions, indicating that the MNase digestion degree does not influence detection of these CID-GFP interactors.

Since there seemed to be a degradation of CID-GFP in the titration experiment (figure 4.4B), all chromatin preparations actually used for mass spectrometry experiments were checked by GFP Western blot (figure 4.4C). No degradation in the experiments used for identification and quantitation was detected. The conclusion was that the degradation did not result from the applied MNase conditions. Figure 4.4D shows the DNA fragment size distribution of solubilized chromatin from one actual replicate experiment, SID1392, where chromatin is mostly mononucleosomal.

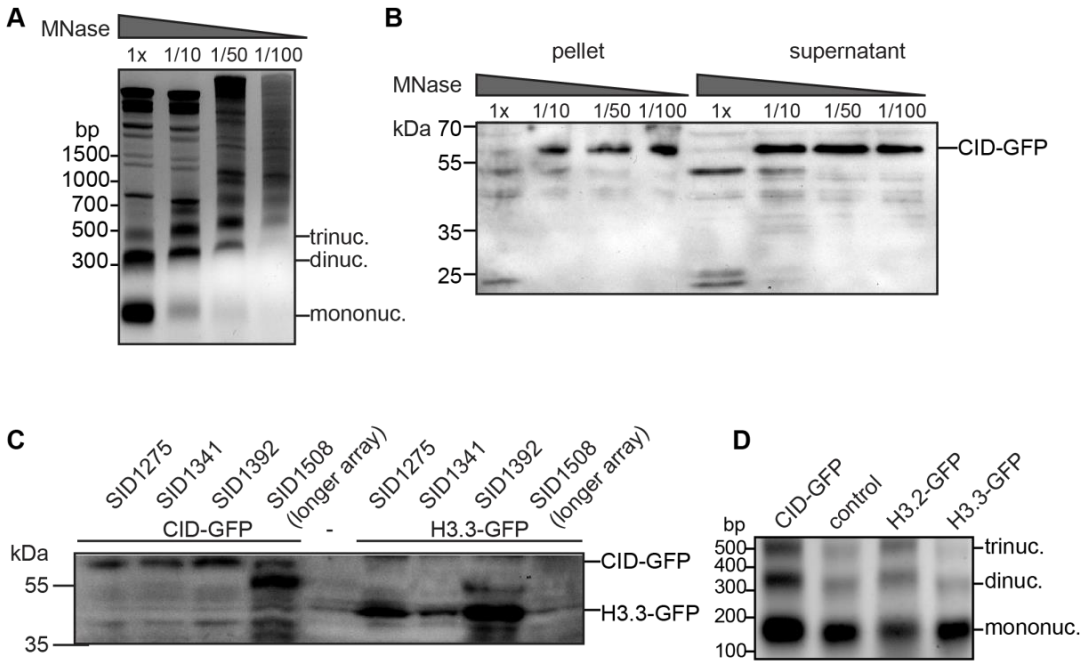

Figure 4.4: Preparation of chromatin for immunoprecipitation. A: Ethidium bromide stained agarose gel demonstrating the digestion degree of chromatin prepared from CID-GFP expressing cells using different concentrations of MNase. B: Corresponding Western blot depicting levels of CID-GFP released by different degrees of MNase digestion (supernatant) and insoluble levels remaining in the pellet. C: Western blot of input samples analyzed by mass spectrometry. D: Representative DNA gel showing lengths of chromatin elements used for immunoprecipitations. Appropriate sizes of mono-, di- and trinucleosomes (mono-, di- and trinuc.) are indicated.

MNase digestion was performed in Ex100 buffer containing 100 mM sodium chloride (NaCl). The standard protocol for extracting MNase digested chromatin from nuclei is to apply a so-called resuspension buffer for one hour.

This buffer consists of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.4 mM K_2HPO_4) plus 150 mM sodium chloride, 2 mM EDTA and 0.1% (v/v) Triton X-100 to efficiently extract chromatin. However, such high concentrations of salt can disrupt interactions of complexes and might lead to less efficient immunoprecipitation as well as loss of binding partners. Therefore, it was tested whether also single conditions contained in the resuspension buffer were sufficient to extract chromatin (figure 4.5). Chromatin was either not digested with MNase (first lane) or it was digested with MNase (same batch for lanes 2-6) and afterwards incubated for one hour with either the complete resuspension buffer (lane 6) that was known to release digested chromatin from nuclear pellets, or with PBS plus only 300 mM salt conditions (lane 2), 3 mM EDTA (lane 3), PBS and sonication (lane 4, Branson Sonifier 250-D, sonication for two times 60 seconds at 22 % power) or PBS plus 0.1 % (v/v) Triton X-100 (lane 5). After extraction and centrifugation at 5000 g for 10 minutes to remove unsoluble fractions, RNA of the supernatant was removed by RNase A treatment (chapter 4.2.2), proteins by proteinase K digestion and the DNA precipitated by addition of ethanol. The DNA gel picture (figure 4.5) indicates that the most efficient way to extract chromatin is indeed by high salt conditions as only there, the bands corresponding to mono- and dinucleosomes and larger chromatin sizes are visible. Thus, chromatin in this workflow was extracted in around 300 mM salt solution and subsequent affinity purification performed in the same conditions.

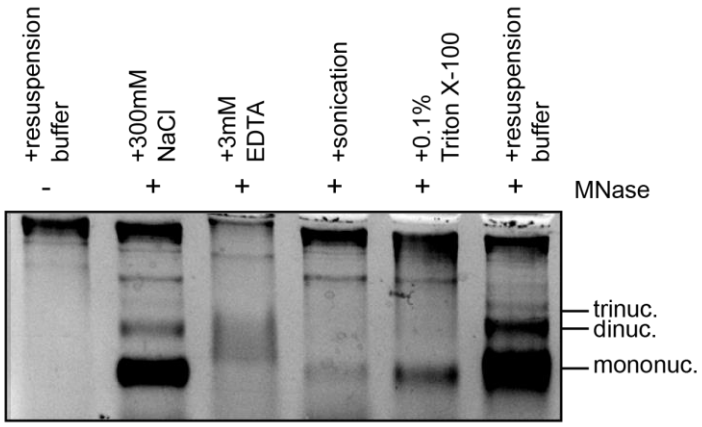

Figure 4.5: DNA gel indicating effectiveness to release digested chromatin from nuclei. Except in the first lane, material from the same batch of MNase digested L2-4 nuclear pellets was extracted for one hour on a rotating wheel at 4°C with the conditions indicated on top. Resuspension buffer contains roughly 300 mM NaCl, 2 mM EDTA and 0.1 % triton X-100.

To determine amounts of GFP-Traps required to quantitatively precipitate the GFP-fusion proteins from a given amount of starting material, a titration was performed using increasing volumes of affinity matrix (beads). Increased depletion of CID-GFP or H3.3-GFP from input samples could be observed

after affinity purification when using increasing bead volumes (figure 4.6). For further experiments, a volume of beads was used that would correspond to between 1-5 µL of bead volume from the experiment shown in figure 4.6. Under these conditions, inputs were considerably depleted of the GFP-tagged protein, but no full saturation of the beads was obtained yet. This should ensure the recovery of interactors of lower abundance and at the same time limit the association of factors that unspecifically bind to the affinity matrix.

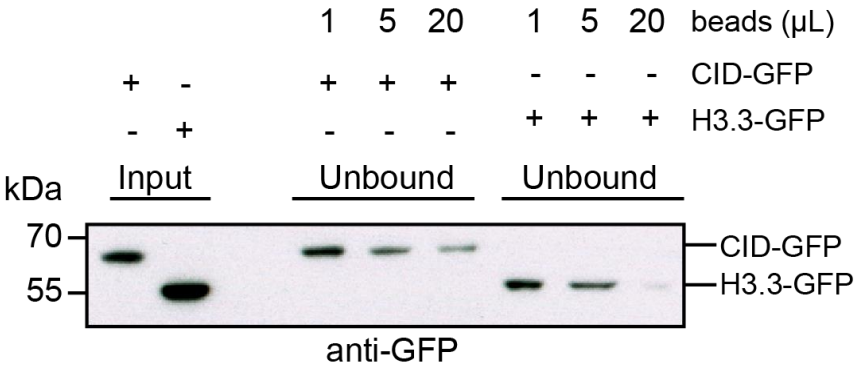

Figure 4.6: Titration of GFP-Trap amounts for immunoprecipitation. Western blot with inputs before binding to beads and unbound fractions after incubation with beads. The same amount of total protein between CID-GFP and H3.3-GFP immunoprecipitations as determined by Bradford assay is loaded.

The whole protocol for affinity purifications was performed as follows: For GFP affinity purifications, stable cell lines expressing H3.2-GFP, H3.3-GFP or CID-GFP were upscaled in roller bottles to around $5*10^9$ cells per cell line. Cells were harvested by centrifugation at 1000 g for 15 minutes at 15 °C in a Heraeus centrifuge. Cell pellets were washed once with 10 mL PBS per 10^9 cells, centrifuged at 1000 g for 10 minutes and the pellet then resuspended in hypotonic buffer (10 mM HEPES-KOH pH7.6, 15 mM NaCl, 1.5 mM $MgCl_2$, 0.1 mM EDTA, plus freshly added protease inhibitors plus DTT). Cells were left swelling in the hypotonic buffer on ice for 20 minutes. To disrupt cell membranes, the same volume of hypotonic buffer plus 0.2 % (v/v) Triton X-100 was added and cells spun at 4 °C for 10 minutes. Volumes of hypotonic buffer were chosen such that the end volume including cell volume plus hypotonic buffer plus hypotonic buffer supplemented with Triton X-100 reached 2 mL per 10^9 cells. One tenth of this volume of 1 M NaCl was added after preparing nuclei in order to reestablish more physiological salt conditions and thereby stabilize interactions that only exist under these conditions. Nuclear pellets resulting from that treatment were washed once in 4 mL per 10^9 cells of EX100 buffer (10 mM HEPES-KOH pH7.6, 100 mM NaCl, 1.5 mM

$MgCl_2$, 0.1 mM EDTA, plus freshly added protease inhibitors plus DTT). Nuclei were then resuspended in 1-2 mL per 10^9 cells of EX100 for MNase digestion. $CaCl_2$ of a 100 mM stock solution was added to a final concentration of 2 mM as calcium is a requirement for MNase activity. 2000 Becker units of MNase per 10^9 cells were added and the reaction incubated at 26 °C for 20 minutes. Becker units result if 500 units of MNase from Sigma are diluted in 850 µL EX50 buffer – the resulting solution contains 50 Becker units per microliter. The reaction was stopped on ice by adding the calcium chelator EGTA to 10 mM final concentration. Chromatin was further extracted by adding NaCl to an end-concentration of 300 mM and homogenized with ten strokes in a Dounce homogenizer with a tight fit pestle. Afterwards, the solution was rotated at 4 °C for one hour, followed by centrifugation at 5000 g to remove debris and harsh centrifugation at top speed for 20 minutes to further clear the protein solution. Aliquots were kept as input samples.

Before performing affinity purification, this solution was further pre-cleared by rotation at 4 °C for 30 minutes using equilibrated Sepharose beads coupled with protein A (GE Healthcare). At the same time, GFP-Traps, after equilibrating them in EX100 containing 300 mM NaCl end-concentration, were pre-blocked by adding 0.5 % (w/v) PVP/0.5 % (w/v) BSA in EX100 also for 30 minutes rotating at 4 °C. The pre-cleared MNase digested chromatin was added to the GFP-Traps and incubated with rotation at 4 °C for two hours. After immunoprecipitation, beads were washed three times with EX100 containing 300 mM NaCl for 5 minutes while rotating at 4 °C. Beads were then diluted in 2x Laemmli solution (final concentration: 100 mM Tris/HCl, pH 6.8, 4 % (w/v) SDS, 20 % (v/v) glycerol, 0.1 % (w/v) bromophenol blue, 0.3 M β-mercaptoethanol) and boiled at 95 °C for 10 minutes. The proteins released from beads under these conditions were then directly applied for SDS-PAGE.

4.2.5 Mass spectrometry

4.2.5.1 In-gel tryptic digestion

After separating bead-binding proteins by size by SDS-PAGE and Coomassie staining, each lane was evenly divided in eight subfractions using a disposable gridcutter (Gel company). Each of the eight fractions was subjected to in-gel tryptic digestion and then loaded individually on the mass spectrometer (chapter 4.2.5.2). In-gel tryptic digestion was performed by Marc Wirth. To identify and measure proteins in different elution fractions of immunoprecipitation experiments, bottom-up, "shotgun" mass spectrometry analysis was carried out. In bottom-up analysis, peptides instead of proteins are measured as they bear smaller masses, occupy lower numbers of charge

states upon ionization and produce less complex fragment spectra. Trypsin is commonly employed for this proteolysis as it specifically cleaves peptide bonds C-terminal to arginines and lysines.

All chemicals used for MS treatment were MS/HPLC grade. The excised bands were first washed two times with water and two times with 20 mM ammonium bicarbonate for pH adjustment. To remove the Coomassie dye, which could cause problems during subsequent MS measurement, gel pieces were incubated with 10 mM ammonium bicarbonate/50 % (v/v) acetonitrile (ACN) for 60 minutes at 37 °C. Afterwards, the gel pieces were again washed three times with 20 mM ammonium bicarbonate. Gel pieces were dehydrated by application of ACN until they appeared white. They were rehydrated with 10 mM DTT in 20 mM ammonium bicarbonate and incubated for one hour at room temperature to reduce bisulfite bonds between cysteines. Cysteines were then alkylated by adding 55 mM iodoacetamide (IAA) in 20 mM ammonium bicarbonate for 30 minutes at room temperature in the dark. Gel pieces were then washed with 20 mM ammonium bicarbonate once and dehydrated by addition of ACN until they appear white. Gel pieces were rehydrated in trypsin solution: 25 ng/µL trypsin (Promega) in 20 mM ammonium bicarbonate were added and the supernatant not aspirated by the gel pieces removed after 30 to 45 minutes incubation. 20 mM ammonium bicarbonate was added to cover the gel pieces and digestion was performed over night at 37 °C. Before loading the resulting peptide solution on the mass spectrometer, it was acidified with trifluoroacetic acid (TFA) to an end-concentration of 0.1 % (v/v), a prerequisite for peptide ionization.

4.2.5.2 Analysis by LC-MS/MS

The peptides were fractionated by reversed phase - liquid chromatography (RP-LC) where peptides are bound to a hydrophobic C18 resin column and eluted by a gradient of increasing acetonitrile, a hydrophobic compound. In RP-LC, more hydrophilic peptides are eluting earlier. By applying a high voltage in a process called electrospray ionization (ESI), peptides are changing from liquid phase into gas phase and subsequently enter the mass spectrometer at different times during the upstream chromatography step (retention time). This reduces complexity of the sample and thus increases the number of identified peptides.

In the mass spectrometer, first a survey spectrum to detect abundant signals and their mass-to-charge (m/z) ratios at a given retention time is measured (MS1 spectrum). Based on the intensity of the signals in the MS1 spectrum, the six most intense peaks (Top6-method) are subjected to fragmentation. In peptide fragmentation, the peptide is dissociated upon collision with an inert

gas which in most cases breaks a single peptide bond within the peptide sequence. Bond breakage relies on the dissociation energy of the peptide bond. By breakages at different positions within a peptide, fragment products after ideally each single amino acid of the parental peptide result. Thus, the fragmentation spectrum (MS2 or MS/MS spectrum) can be used to read the amino acid sequence of the peptide and to determine the location of posttranslational modifications, if present. Raw data of MS1 and MS2 spectra over the retention time of the whole LC run thus comprise m/z values of survey as well as several fragmentation spectra.

MS1 overview spectra are recorded every few seconds, leading to several measurements during the retention time of one peptide (peak width 0.5-2 minutes). In the meantime, the six MS2 spectra are obtained (figure 4.7). When searching the MS data for the m/z ratio of a certain peptide in form of an extracted ion chromatogram (XIC, see figure 4.7), it is first eluting to a low intensity, which increases over time and decreases thereafter.

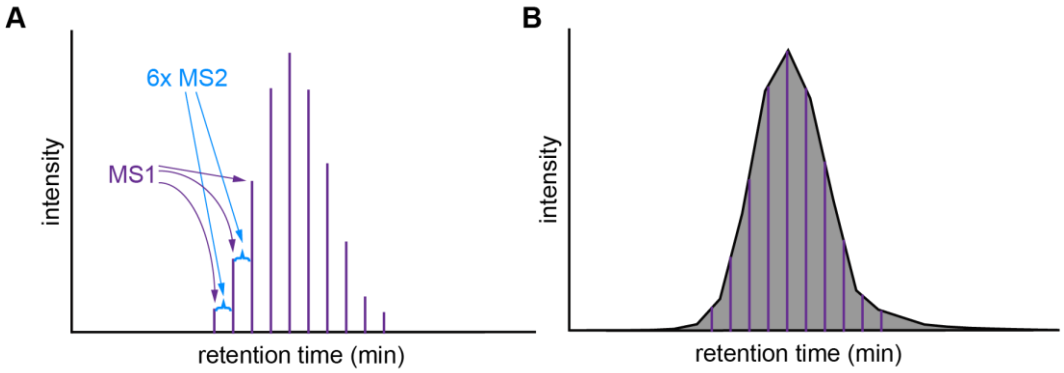

Figure 4.7: Quantification of protein intensities by mass spectrometry. A: Extracted ion chromatogram (XIC) of a certain m/z-ratio. Every few seconds, a MS1 overview spectrum is generated (violet, vertical line), in the time between the MS1 spectra, six MS2 spectra are recorded. The peptide of interest is starting to elute from the column at a certain retention time (x-axis), its eluted amount is increasing over time and then again decreasing. The intensity of detection in the mass spectrometer mirrors this phenomenon. B: In order to obtain quantitative values of peptide intensities, the MS1 intensities are combined to form a peak and the area under the peak is calculated (grey).

The area under such a peak can be integrated and this gives quantitative information about the abundance of a protein in a sample. This absolute intensity though can differ between different peptides due to factors like feasibility of ionization or binding to the column. To improve comparison between different proteins, the MaxQuant software used here (chapter

4.2.5.4) calculates a value called iBAQ (intensity based absolute quantification, (Schwanhausser, Busse et al. 2011). This value puts the theoretical detection probability for different proteins into account – the bigger a protein is, the higher the probability to measure one of its peptides, thus minimizing detection differences and facilitating comparison of protein abundances. In general: a high iBAQ value stands for high protein abundances.

To analyze samples by LC-MS/MS, 50 % of the in-gel digested peptides of one fraction of a lane were injected into an Ultimate 3000 HPLC system (Thermo Fisher Scientific). Samples were desalted on-line by a C18 micro trap column (5 mm x 300 µm inner diameter, packed with C18 PepMapTM, 5 µm, 100 Å, Thermo Fisher Scientific), and peptides were separated by a gradient from 5 % to 60 % (v/v) acetonitrile in 0.1% (v/v) formic acid over 40 minutes at 300 nL/minute on a C18 analytical column (10 cm x 75 µm, packed in house with C18 PepMapTM, 3 µm, 100 Å, Thermo Fisher Scientific). The effluent from the HPLC was directly infused into the LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) via a nano-electrospray ion source. The MS instrument was operated in the data-dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS1 spectra (m/z 350-2000) were acquired in the Orbitrap with resolution 60.000 at m/z 400. For all measurements with the Orbitrap detector, three lock-mass ions from ambient air (m/z=371.10123, 445.12002, 519.13882) were used for internal calibration as described (Olsen, de Godoy et al. 2005). The six most intense peptide signals with charge states between two and five were sequentially isolated applying a 1 Dalton (Da) window centered on the most abundant isotope to a target value of 10.000 and fragmented in the linear ion trap by collision-induced dissociation. Fragment ion spectra were recorded in the linear trap of the instrument. Typical mass spectrometric conditions were: spray voltage 1.4 kV; no sheath and auxiliary gas flow; heated capillary temperature 200°C; activation time 30 ms and normalized collision energy 35 % for collision-induced dissociation in the linear ion trap.

4.2.5.3 Protocol to analyze histone modifications

As described in chapter 1.3 and 2.4.1, histone proteins are vastly modified mainly on their N-termini, which consist of a high number of lysines and arginines. Trypsin digests after lysines and arginines which for histones results in very small peptides that are not convenient for MS analysis (figure 4.8). That is why in order to detect as many peptides carrying histone posttranslational modifications as possible, the trypsin digestion is combined

with a propionylation treatment that modifies unmodified and monomethylated lysines so trypsin only cuts after arginines – that is, because trypsin anyways cannot digest after di- or trimethylated or acetylated lysines. It results in the peptides indicated in figure 4.8.

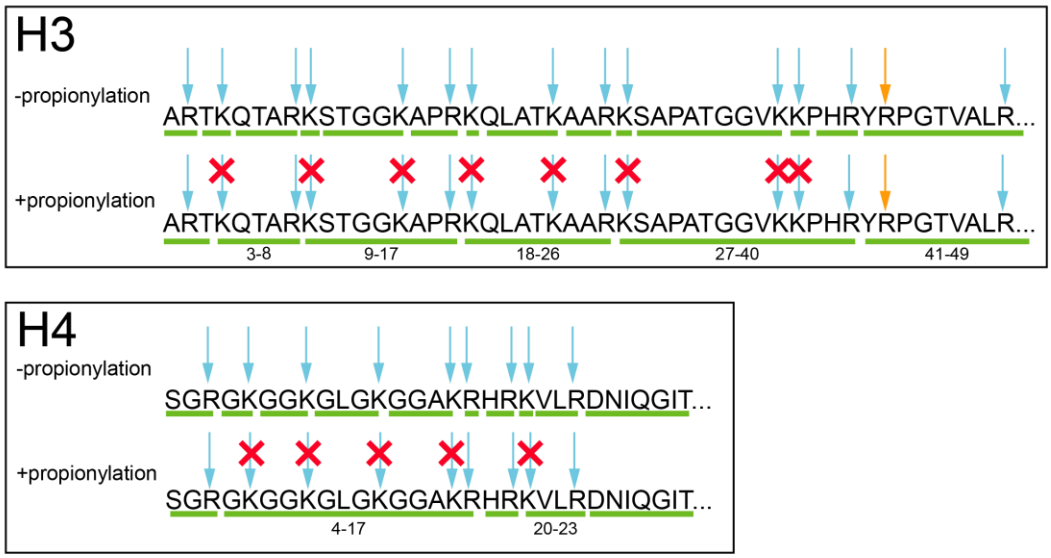

Figure 4.8: Propionylation inhibits trypsin digestion after lysines and therefore creates typical tryptic peptides of histone H3 and H4. Blue arrow: trypsin cutting site. Red cross: cutting site inhibited by propionylation. Orange arrow: residue protected from trypsin cleavage because of the following proline. Numbers indicate amino acids of analyzed tryptic peptides.

Histone proteins at a size between 11 and 17 kDa were excised from gels and in-gel tryptic digestion was performed. First, the bands were destained from Coomassie essentially as described for proteomics analysis in chapter 4.2.5.1. Instead of treatment with DTT and IAA, 1 µL propionic anhydride in 10 µL 100 mM ammonium bicarbonate was added to the gel pieces followed by 30 µL 1 M ammonium bicarbonate, to keep the pH during the 30 minutes incubation at room temperature between 7 and 8. Gel pieces were washed carefully afterwards three times with 100 mM ammonium bicarbonate to avoid over-propionylation events that result in unspecific mass shifts. Afterwards, trypsin digestion over night was performed similar to the description in chapter 4.2.5.1. Here, 200 ng of trypsin were used per sample. To get a high sequence coverage, histone peptides were desalted using Carbon ZipTips (Glygen) according to manufacturer's instructions and the desalted peptides loaded directly onto the LC column without the use of a trap column. Otherwise, LC-MS/MS analysis was carried out essentially as described in the previous chapter for proteomics analyses.

4.2.5.4 MaxQuant analysis for protein identification and quantification

A large amount of data is acquired in the eight LC-MS/MS runs resulting from the eight chopped gel pieces containing information about identity and abundance of detected proteins. This information should be compared to a theoretical database of a *Drosophila* proteome. MaxQuant is a software that searches peptide sequences obtained from MS2 spectra versus an annotated proteome as in this case the one from the flybase database for *Drosophila melanogaster*. The software correlates the measured sequences with theoretical sequences with a certain probability. Actually measured peptide sequences that can be assigned to a protein or protein isoforms from the database then results in identification of that protein in the MS run.

For protein identification, the raw data were analyzed with the Andromeda algorithm of the MaxQuant protein analysis package (version 1.2.2.5) against the Flybase dmel-all-translation-r5.24.fasta database including reverse sequences to determine false discovery rates and contaminants. Trypsin/P was selected as protease and maximum two missed cleavages allowed. Fixed modifications were carbamidomethylation of cysteines resulting from IAA treatment, while methionine oxidation and N-terminal acetylation of proteins were set as variable modifications. The mass tolerance of the initial search was 20 ppm; after recalibration, 6 ppm mass error were applied for the main search. Fragment ions were searched with a mass tolerance of 0.5 Da using the six most intense signals within 100 Da. Searching for secondary peptide hits within already assigned MS/MS spectra was enabled. The search results were filtered with a peptide and protein false discovery rate of 0.01 and a minimum peptide length of six amino acids. Protein identifications with at least one unique peptide were accepted. For quantification, the intensity based absolute quantification (iBAQ) values were calculated from peptide intensities and the protein sequence information of unmodified and modified unique peptide species with a minimum of two identified peptides per protein.

4.2.5.5 Statistical analysis to determine centromeric proteins

Statistical analysis was carried out by Dr. Andreas Schmidt. The output of the MaxQuant search was used for further statistical analysis. First, protein identifications from reverse sequences or contaminants as well as identifications without MS1 quantification values were removed. IBAQ values were log2-transformed. If a protein did not get identified in one out of nine samples (3x mock purification, 3x CID-GFP and 3x H3.3-GFP IPs), the iBAQ value in this sample is zero. Since for ratios, division by zero errors would occur, the zeros need to be replaced. This was done by imputing values by a random distribution centered at 1/3*log2 of the lower distribution of the complete iBAQ dataset. In order to avoid huge deviations by imputation, it

was repeated three times and the average value was further used. ANOVA (analysis of variance) was applied in DanteR (vs 0.2, PNNL, Richland, USA) as statistical means to calculate protein enrichment factors and p-values (see table 2, appendix) and obtained p-values were corrected for multiple hypothesis testing by the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

An important prerequisite for performing ANOVA analysis is that the obtained experimental values show Gaussian distribution. Figure 4.9 displays on one of the analyzed replicates that this is the case here: The log2-value of all iBAQ values of one run (from eight gel fractions) is plotted. A high log2-value probably stands for specific enrichment of one protein and consistently, especially in the purifications with CID-GFP and H3.3-GFP, a slight second peak in the area of log2 of twenty and more is appearing.

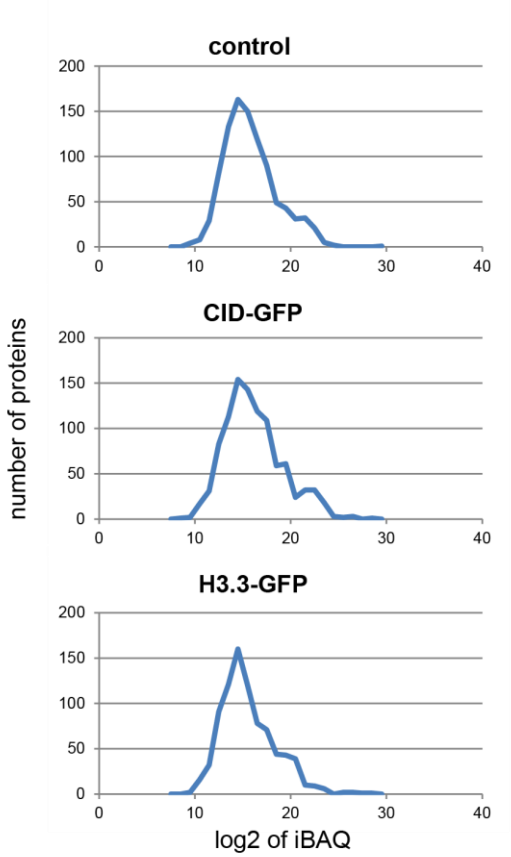

Figure 4.9: Distribution of log2-transformed iBAQ values of experiment SID1341. X-axis: all log2 iBAQ values of one experiment. Y-axis: number of proteins harboring a given log2 iBAQ value. The Gaussian distribution allows subsequent statistical analysis.

A cutoff of a 16 times enrichment (log2 of four) over both H3.3-GFP samples and mock purifications was chosen in order to determine potential centromeric proteins. The outcome was 85 proteins plus nine proteins that were added because of their specificity of detection in CID-GFP purifications (see table 2, appendix): 90 % of all their determined iBAQ values had to arise specifically in CID-GFP IPs. Furthermore, they had to be identified in at least two out of

three CID-GFP purifications. The nine proteins with the highest average iBAQ values were added. To select around one third of factors to be checked by immunofluorescence analysis, the resulting 94 proteins were sorted by the following criteria: 1. The protein had to be detected in at least two out of three CID-GFP purifications. 2. The sum of iBAQ values originating from the three CID-GFP purifications out of the nine purifications (including 3x mock and 3x H3.3-GFP purifications) had to exceed 90 %. 3. Factors were sorted according to their average iBAQ values in the three CID-GFP purifications with the highest iBAQ values on top. The top third of proteins were chosen for IF analysis. Because of missing construct availability, the following proteins were not checked by IF: CG32344, SRPK, qkr58E-2, CG32069, CG6151 (fwe), CG13117, RhoGAP54D, and CG8478.

4.2.6 Microscopy methods

4.2.6.1 Methods for microscopic analyses

Immunolocalization protocol:
For immunolocalization in Schneider L2-4 cells, transiently transfected and copper-induced cells (or uninduced, stable cell lines as indicated in figure legends) were settled on poly-L-lysine (Sigma) coated coverslips for minimum one hour, prelysed with PBS/0.1 % (v/v) Triton X-100 (PBT) for 40 seconds and gently washed with PBS before fixation with PBS/3.7 % (v/v) formaldehyde (10 min at room temperature). The prelysis step removes soluble proteins of the nucleoplasm which might otherwise cover centromeric signals. Cells were permeabilized for six minutes on ice with PBS/0.25 % (v/v) Triton X-100, subsequently washed twice with PBS for ten minutes and blocked for 45 minutes with Image-iT® FX signal enhancer (Invitrogen). Primary antibodies were diluted in PBS/5 % (v/v) normal goat serum (NGS) (Dianova) and incubated with the coverslips over night at 4 °C. After two ten minute washes with PBT, fluorophore coupled-secondary antibodies in diluted in 5 % (v/v) NGS/PBS were added for 45 min to one hour at room temperature. Cells were washed two times for ten minutes with PBT and then stained with PBS/DAPI (200 ng/mL) for six minutes, washed again with PBS and mounted in Vectashield (Vector Labs) or SlowFade (Invitrogen).

Mitotically condensed chromosome analysis:
For cytospins, performed by Georg Schade, 1-2*10^5 stably transfected cells were enriched in metaphase by colcemid treatment (1 µg/mL) for 30 minutes. Cells were resuspended in 500 µl 0.5 % sodium citrate, incubated for seven minutes and then spun using a single-chamber cytospin funnel in a Shandon

Cytospin 4 (10 minutes; 900 g; high acceleration). Cells were fixed for 9 minutes in 3.7 % (v/v) formaldehyde in PBS solution, washed once with PBT for 5 minutes, then blocked with Image-iT® FX signal enhancer (Invitrogen) for 1 hour. Subsequent steps for staining DNA and mounting cells were performed as described above.

Analysis of mitotic defects:
The protocol to determine mitotic defects was carried out by Georg Schade. RNAi experiments were essentially performed as described above (chapter 4.2.1). Tubulin stainings in order to visualize the mitotic spindle and detect aberrant phenotypes and staining against the mitotic marker H3S10ph (see also chapter 2.8) were conducted. Two replicates of the experiment with 50 imaged cells in each replicate were carried out.

SNAP protocol:
Kc167 cells obtained from Gary Karpen's lab (UC Berkeley) expressing SNAP-CID and CID-GFP were used for the SNAP protocol. The SNAP protocol is modeled after the publication from this lab by Mellone *et al.* (Mellone, Grive et al. 2011). RNAi was performed as described above (chapter 4.2.1). Conditioned medium was produced by growing cells in Schneider medium for 2-3 days, spinning down the cells for five minutes at 600 g and the supernatant mixed 1:1 with fresh Schneider medium.

To block SNAP-CID with BTP (quench), cells were treated with 12 µM BTP in conditioned medium for 30 minutes, washed three times with 1 mL of fresh medium for three minutes, once with conditioned medium for 30 minutes and afterwards released to progress in cell cycle (0h timepoint). The chase was performed for 24 hours. Then, cells were stained with 3 µM TMR in conditioned medium for 45 minutes in the dark (pulse). Cells were washed three times for five minutes with fresh medium, once 30 minutes with conditioned medium and twice for ten minutes with conditioned medium. After the quench-chase-pulse method, cells were treated as described above (fixation, permeabilization, DAPI stain and embedding).

4.2.6.2 Analysis by confocal microscopy

The microscopy setup used by me was a Leica TCS SP5 II confocal microscope with a 63x glycerol immersion objective (NA = 1.3). Z-stacks were deconvolved using the Huygens Essential Software (SVI) and further analysed with ImageJ software. Georg Schade used a DeltaVision RT microscope with 100x oil immersion objective. Pictures were deconvolved and analysed using SoftWorx Explorer Suite (Applied Precision). Line profiles for

both cases were produced using RGB profiler plugin for ImageJ (http://rsb.info.nih.gov/ij/plugins/rgb-profiler.html)

4.2.6.3 Quantitative determination of GFP amounts in centromeres

The CellProfiler pipeline described here was established by Dr. Andreas Thomae. Maximum intensity projections of deconvolved immunolocalization images were loaded into a pipeline of the CellProfiler software. First, nuclei were identified as primary objects with Otsu Global two classes thresholding method with the option to manually remove mitotic cells or misrecognized nuclei to avoid false quantifications. CID-GFP images were then enhanced as speckles. The enhanced CID-image was masked with nuclei to later only identify nuclear CID-GFP intense foci (centromeres). Next, centromeres were identified as primary objects with Otsu Global two classes thresholding method using the enhanced and masked CID-GFP image. Centromere objects (child objects) were related to the respective nucleus (parent object), and the per-parent mean integrated centromeric CID-GFP intensity was determined. The per-parent mean integrated intensity value was multiplied by the number of centromeres determined per nucleus to calculate total centromeric CID-GFP intensities per nucleus. Data were plotted using the Vertex42^{TM} Box and Whisker Plot Template. Box plot shows upper and lower quartile and median values. Whiskers indicate upper and lower boarders of the 1.5-fold IQR (interquartile range) and contain more than 90 % (or more than 80 % in case of CENP-C 1 knockdown) of data points.

4.2.7 FACS analysis

Around 5*10^6 cells were harvested, washed with PBS and fixed in 70 % methanol/30 % PBS. Fixed cell pellets were kept at -20 °C for at least one hour and maximum one week. Pellets were then washed twice with PBS and then supplemented with 100 µg/mL RNase A in 200 µL PBS and incubated for ten minutes on ice. 10 µL of propidium iodide stock solution (1 mg/mL, Sigma) with 2 µL of EDTA in PBS were added and the samples directly measured on the FACSCanto device. Cell appearance was measured with forward and side scatter (FSC/SSC) indicating cell size or granularity, respectively. Propidium iodide was measured in the PE channel and the signal corresponding to Alexa 488 in the FITC channel. Per condition, 100.000 cell events were measured. Results were analyzed using the FlowJo software. Gating on cells was performed as shown in figure 4.10: Cells were separated from debris by a plot of FSC versus SSC. Single cells were obtained by plotting PE-A (area) versus PE-W (width). In FACS analysis, the DNA content per cell is determined and if two cells in G1-phase stick together, the DNA content resembles the one of

cells in G2/M-phase. Therefore, this second gating event is applied. If an event harbors the same area, but double the width, it probably results from two cells sticking together. The histograms and density plots shown in figure 2.24 result from single cells.

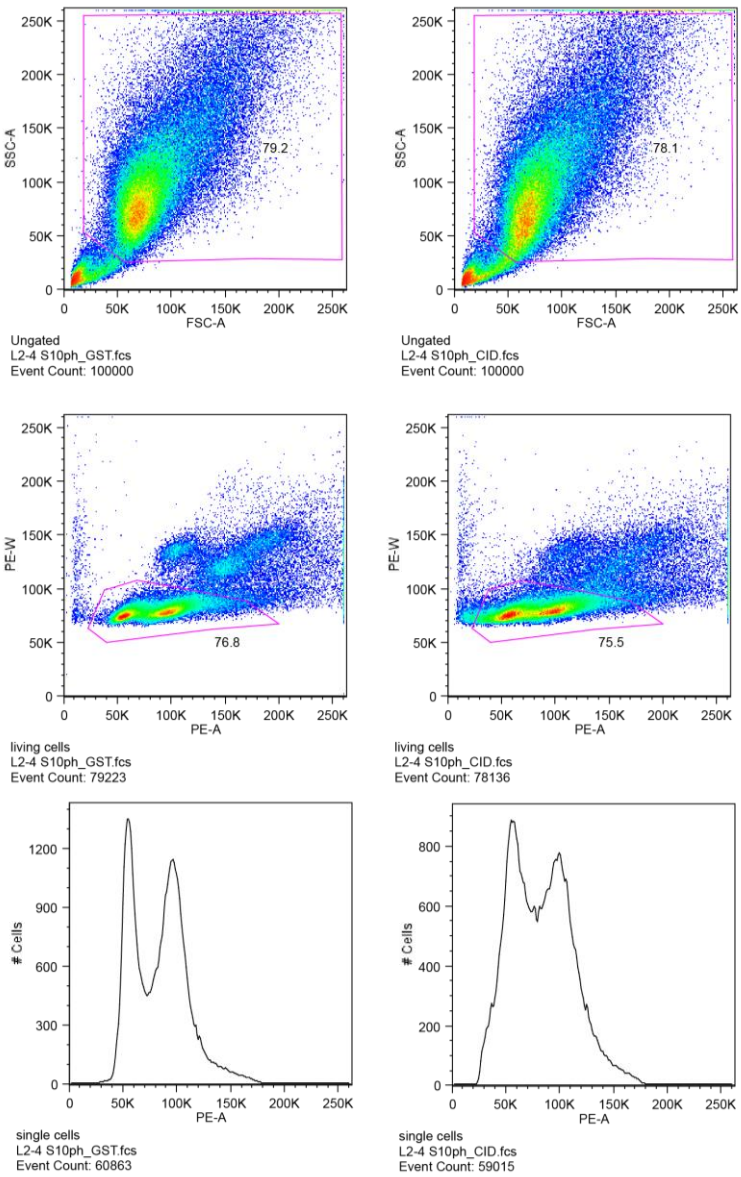

Figure 4.10: Gating of raw FACS data. Pink shapes are self-drawn gates that are kept the same for all conditions. Upper panel: forward scatter versus side scatter (FSC vs. SSC) density plot identifies living cells for GST knockdown (left) or CID knockdown (right). Middle panel: Single cells are determined by plotting PE-A versus PE-W. Numbers in plots indicate percentages contained in gates. Lower panel: Histograms obtained from only single cells of PE-A signal versus cell number as shown in the results part (figure 2.24).

APPENDIX

Table 1: Proteins only identified in CID-GFP pulldowns. Listed are proteins that were identified at least in one of the three biological replicates of CID-GFP pulldown and never in control pulldowns (wildtype, H3.3-GFP). Blue: Proteins present in final list of CID-GFP enriched proteins after statistical analysis (see table 2).

asp	CG17187	CG4289	CHKov1	MED4	Src42A
Baldspot	CG17202	CG4679	CLIP-190	MEP-1	Stim
Bap60	CG17293	CG4699	coil	mge	su(Hw)
Brd8	CG17544	CG4887	CycB	mip120	Su(var)3-7
CAF-1-PB	CG17660	CG4972	dream	mit(1)15	sub
Cal1	CG17768	CG5004	ec	Mlf	swm
Cdk12	CG2186	CG5009	Edem1	mRpL12	Synd
CENP-C	CG2469	CG5664	eIF2B-gamma	NAT1	Taf12
CG10131	CG30390	CG6151	enok	ncm	Tango1
CG10320	CG3056	CG6227	Ets97D	Orc1	Tango5
CG10420	CG31048	CG6230	Fatp	Orc2	tex
CG10600	CG31510	CG6525	Ge-1	pch2	Top3beta
CG1091	CG31650	CG6766	gfzf	pea	torp4a
CG1092	CG32069	CG6904	Gint3	phr	tou
CG11076	CG3209	CG7338	hay	prod	Trap1
CG11200	CG32243	CG7518	Hsp70Ab	qkr58E-2	Trp1
CG11577	CG32343/ATAC3	CG8478	Hsp70Bbb	Rae1	Tsp42Ed
CG12104	CG32549	CG8525	hyd	RagC	v(2)k05816
CG12343	CG32554	CG9293	ifc	Rbcn-3A	VhaM9.7-2
CG1265	CG32699	CG9302	Jheh1	Rcd5	vimar
CG1399	CG33691	CG9418	Klp3A	RhoGAP54D	Vsx1
CG14722	CG3548	CG9601	l(1)G0230	Rlb1	wapl
CG14480	CG3625	CG9609	LSm1	RN-tre	YL-1
CG15012	CG3714	CG9776	MED17	SF1	ZnT63C
CG17153	CG41519	Chd1	MED30	Spn4	

Table 2: Proteins enriched in CID-GFP pulldown as determined compared to wildtype and H3.3-GFP pulldowns by statistical analysis. The list is sorted due to enrichment of CID-GFP over control pulldown with the highest enriched proteins on top. Names, average log2 enrichment values of three biological replicate experiments plus standard deviations and p-values, experimental evidence from this work and constructs created for analyses in this work are indicated. Factors tested by immunolocalization analysis in this work are highlighted in green.

name	CID-GFP/control		CID-GFP/H3.3-GFP		Experimental evidence	Construct
	log2	p-value	log2	p-value		
CID	15.9 ±1.1E-01	1,60E-02	11.1 ±4.5E-05	5,80E-02	Control	
CG14480	11.8 ±8.7E-02	1,70E-05	11.6±7.4E-02	1,40E-05	Centromeric and euchromatic, on chromosome arms during mitosis, elevated mitotic errors in knockdown	pMK33-CFH-BD
Cal1	11.3 ±9.1E-02	1,60E-05	11.2 ±7.6E-02	1,70E-05	Control	
CG2051	10.8 ±1.0E-05	5,40E-02	8.3 ±3.1E-07	1,10E-01	Centromeric and euchromatic, on chromosome arms during mitosis, elevated mitotic errors in knockdown	GFP Invitrogen
CG13117	9.2 ±1.1E-01	7,60E-04	9.6 ±2.6E-02	6,50E-04		
CG34191	9.2 ±1.8E-01	3,50E-04	9.5 ±1.3E-01	3,00E-04		
CG6769	8.6 ±9.2E-02	2,20E-04	8.4 ±7.2E-02	2,60E-04		
Vps4	8.0 ±4.3E-01	9,20E-02	7.5 ±4.0E-01	1,00E-01		
CG4972	7.7 ±1.6E-01	6,20E-05	7.6 ±4.7E-02	5,30E-05	Nuclear	pMT-hyg FLAG-HA-Nterm
CAF-1-PB	7.6 ±8.6E-03	1,90E-04	7.5 ±1.1E-02	1,70E-04	Control, elevated mitotic errors in knockdown, nuclear (antibody stain)	
LSm3	7.5 ±7.6E-02	3,90E-04	7.7 ±5.4E-02	2,50E-04		
Kap-alpha1	7.5 ±4.4E-01	1,10E-02	8.3 ±2.5E-01	4,60E-03		
REG	7.3 ±1.1E-04	4,00E-01	13.4 ±3.8E-02	6,40E-02	Centromeric and euchromatic, higher CID-GFP levels in cells after knockdown	pMK33-CFH-BD

name	CID-GFP/control		CID-GFP/H3.3-GFP		Experimental evidence	Construct
	log2	p-value	log2	p-value		
ifc	7.3 ±9.8E-02	2,20E-04	7.1 ±6.8E-02	2,60E-04	Cytoplasmic	pMK33-CFH-BD
CG11985	7.2 ±7.5E-02	7,10E-04	7.5 ±6.7E-02	3,50E-04		
CG9293	7.1 ±1.7E-01	3,70E-05	7.1 ±6.2E-02	4,30E-05	Centromeric and euchromatic	pMK33-CFH-BD
CG8891	7.1 ±1.2E-01	7,70E-03	7.4 ±1.2E-01	5,00E-03		
CG3548	7.0 ±1.0E-01	2,00E-04	7.0 ±4.9E-02	2,50E-04	At centromeres and euchromatin in mitosis, nuclear in interphase	GFP Invitrogen
Hsp70 Bbb	7.0 ±2.1E-02	1,40E-03	6.9 ±6.0E-02	1,50E-03	Hsp70Ab: cytoplasmic	pMT-hyg FLAG-HA-Nterm
SRPK	7.0 ±6.1E-01	2,60E-02	7.6 ±2.8E-01	1,40E-02		
Unc-76	7.0 ±9.5E-02	8,80E-03	6.3 ±8.1E-02	1,20E-02		
CG6776	6.9 ±2.1E-01	8,40E-02	7.8 ±9.1E-02	4,30E-02		
Srp54k	6.9 ±5.0E-01	1,90E-01	6.4 ±5.3E-01	1,30E-01		
CG32069	6.7 ±1.7E-01	1,60E-04	6.6 ±6.5E-02	1,70E-04		
alphaTub 85E	6.7 ±7.2E-02	1,80E-03	7.1 ±1.0E-01	9,20E-04		
MED30	6.6 ±7.5E-02	1,40E-04	6.6 ±6.6E-02	1,80E-04	Centromeric and euchromatic	GFP Invitrogen
CG6151	6.6 ±9.6E-02	1,60E-04	6.4 ±8.1E-02	1,30E-04		
Rrp4	6.4 ±1.1E-01	1,70E-03	6.3 ±6.4E-02	1,30E-03		
CG1789	6.3 ±2.1E-02	1,10E-03	6.8 ±2.7E-02	7,30E-04		
qkr58E-2	6.3 ±9.7E-02	1,80E-04	6.1 ±6.8E-02	2,10E-04		
Prod	6.3 ±1.4E-01	1,20E-04	6.2 ±1.1E-01	1,50E-04	Control, elevated mitotic errors in knockdown	
Tsp42Ed	6.3 ±1.0E-01	2,00E-04	6.1 ±8.1E-02	1,90E-04	Membrane protein	pMK33-CFH-BD
CG6180	6.3 ±1.4E-01	9,20E-04	6.6 ±2.8E-02	6,70E-04		
CG7945	6.3 ±4.2E-01	1,80E-01	5.1 ±1.0E+00	2,80E-01		
CG14695	6.2 ±1.9E-01	2,30E-03	6.5 ±1.1E-01	1,80E-03		
CG3731	6.2 ±6.5E-02	4,00E-04	6.0 ±5.0E-02	5,00E-04		
ATPsyn-Cf6	6.2 ±2.2E-01	1,20E-05	6.2 ±5.0E-02	1,30E-05		
BEAF-32	6.1 ±5.9E-02	5,30E-04	6.2 ±7.5E-02	5,70E-04		
CG5021	6.1 ±6.2E-01	1,90E-01	4.9 ±1.2E+00	3,10E-01		
PH4alpha EFB	6.0 ±1.0E-01	1,70E-03	6.3 ±9.6E-02	1,50E-03		

name	CID-GFP/control		CID-GFP/H3.3-GFP		Experimental evidence	Construct
	log2	p-value	log2	p-value		
ns3	6.0 ±9.2E-02	2,60E-01	6.0 ±7.1E-02	2,10E-01		
CG1265	6.0 ±1.5E-01	9,70E-04	6.0 ±5.0E-02	1,10E-03	Cytoplasmic	pMK33-CFH-BD
CG12343	6.0 ±6.5E-02	1,20E-04	6.0 ±6.7E-02	1,10E-04	Nuclear	pMK33-CFH-BD
CG11030	5.9 ±3.8E-01	3,00E-01	5.6 ±4.3E-01	2,40E-01		
CG9418	5.9 ±8.6E-02	1,80E-04	5.7 ±9.0E-02	2,50E-04	Centromeric and euchromatic	pMK33-CFH-BD
Prx5	5.9 ±1.2E-01	2,00E-02	6.1 ±4.6E-02	1,60E-02		
Jheh2	5.9 ±6.5E-02	8,40E-04	5.9 ±1.1E-02	8,20E-04		
Pcd	5.9 ±6.6E-02	4,70E-03	6.3 ±1.1E-01	3,50E-03		
Caz	5.7 ±1.3E-01	2,10E-05	6.0 ±6.1E-02	8,70E-06		
CG30390	5.7 ±1.9E-01	1,80E-04	5.6 ±8.5E-02	2,40E-04	Nuclear	pMT-hyg FLAG-HA-Nterm
Ard1	5.6 ±4.8E-02	1,80E-03	6.0 ±1.4E-01	1,40E-03		
Dbp45A	5.6 ±2.3E-01	1,30E-03	6.0 ±3.1E-02	1,00E-03		
Uch-L3	5.5 ±1.4E-01	7,70E-04	5.7 ±6.9E-02	6,00E-04		
Fmr1	5.5E±4.5E-02	2,00E-01	5.4 ±1.2E-01	1,70E-01	Cytoplasmic	pMK33-CFH-BD
Gfzf	5.5 ±9.0E-02	2,70E-04	5.5 ±1.1E-01	2,50E-04	Centromeric and euchromatic	GFP Invitrogen
su(Hw)	5.4 ±8.7E-02	6,00E-04	5.3 ±8.2E-02	4,10E-04	Non-centromeric foci in nucleus and cytoplasm	GFP Invitrogen
CG6084	5.4 ±9.2E-02	7,80E-03	5.6 ±4.8E-02	6,70E-03		
Nup50	5.4 ±2.8E-02	1,10E-01	4.1 ±3.3E-01	2,10E-01		
RagC	5.4 ±1.5E-01	5,40E-04	5.3 ±1.6E-01	4,00E-04	Nuclear	pMK33-CFH-BD
CG17271	5.4 ±6.9E-01	2,50E-01	4.3 ±1.2E+00	1,80E-01		
CAF-1-PA	5.3 ±1.7E-07	7,20E-03	5.3 ±6.8E-07	6,50E-03	Control, elevated mitotic errors in knockdown, nuclear (antibody stain)	
CG1091	5.3 ±1.2E-01	7,20E-04	5.3 ±1.3E-01	9,30E-04	Cytoplasmic	GFP Invitrogen
Bor	5.1 ±1.2E-01	2,80E-02	4.2 ±4.1E-01	6,80E-02		
Spt6	5.1 ±3.9E-02	1,90E-01	4.8 ±7.6E-02	1,90E-01		
YT521-B	5.0 ±3.5E-01	3,50E-01	5.0 ±9.7E-02	3,20E-01		
CG5482	5.0 ±9.0E-02	3,40E-04	5.3 ±5.3E-02	2,20E-04		
CG1309	4.9 ±7.4E-03	8,90E-03	5.1 ±3.7E-02	4,00E-03		

name	CID-GFP/control		CID-GFP/H3.3-GFP		Experimental evidence	Construct
	log2	p-value	log2	p-value		
Hyd	4.9 ±1.6E-01	1,80E-03	4.8 ±4.4E-02	2,30E-03	Nuclear, excluded from heterochromatin higher CID levels in knockdown, elevated mitotic errors in knockdown	pMT-hyg FLAG-HA-Nterm
CG32344	4.9 ±5.8E-06	4,20E-01	11.1 ±4.1E-02	3,10E-02		
CG11820	4.9 ±2.9E-01	6,90E-02	5.8 ±2.5E-01	3,30E-02		
Oscp	4.8 ±1.0E-01	2,30E-02	5.1 ±7.6E-02	1,90E-02		
CG4858	4.8 ±1.5E-01	6,70E-01	6.5 ±5.8E-02	3,30E-01		
CG3335	4.8 ±9.8E-02	5,80E-03	5.0 ±7.8E-02	4,20E-03		
CG11076	4.8 ±4.3E-02	2,30E-03	4.7 ±1.0E-01	2,10E-03	Nucleolar	GFP Invitrogen
YL-1	4.7 ±8.9E-02	6,90E-04	4.7 ±5.9E-02	7,10E-04	Nuclear	pMK33-CFH-BD
Sub	4.6 ±1.0E-01	5,30E-03	4.6 ±1.3E-01	6,70E-03	Control	pMK33-CFH-BD
Rbcn-3A	4.6 ±1.4E-01	3,60E-04	4.5 ±7.8E-02	4,10E-04	Nuclear	pMT-hyg FLAG-HA-Nterm
eIF-4B	4.6 ±8.0E-02	1,90E-02	4.8 ±8.2E-02	1,30E-02		
Pgi	4.5 ±9.7E-02	9,70E-03	4.7 ±1.6E-02	7,20E-03		
CG9630	4.5 ±7.3E-02	2,80E-03	4.8 ±1.6E-01	1,20E-03		
CENP-C	4.5 ±1.3E-01	3,00E-03	4.4 ±7.0E-02	2,90E-03	Control	
Mtr3	4.5 ±1.1E-01	5,20E-01	5.8 ±6.4E-02	2,70E-01		
CG2943	4.5 ±7.7E-02	6,10E-01	4.4 ±5.5E-02	5,10E-01		
Cdk12	4.4 ±2.2E-01	5,00E-04	4.2 ±7.5E-02	5,70E-04	Nuclear, excluded from heterochromatin	pMT-hyg FLAG-HA-Nterm
CG7518	4.3 ±1.2E-01	5,40E-04	4.4 ±7.4E-02	6,00E-04	Cytoplasmic	pMT-hyg FLAG-HA-Nterm
CG5198	4.2 ±3.8E-02	3,20E-02	4.6 ±1.2E-01	1,10E-02		
CG1399	3.8 ±8.4E-02	1,73E-03	3.8 ±5.0E-02	1,64E-03	Cytoplasmic	pIB-v5
CG6227	4.0 ±2.8E-02	3,29E-03	4.0 ±9.8E-02	2,29E-03	Centromeric and euchromatic	GFP Invitrogen
asp	1.9 ±5.0E-02	7,49E-02	1.8 ±8.8E-02	6,59E-02	Nuclear	pMT-hyg FLAG-HA-Nterm
CG32343 /ATAC3	3.8 ±1.5E-01	7,13E-03	3.7 ±6.5E-02	8,38E-03	Nuclear	pMT-hyg FLAG-HA-Nterm

name	CID-GFP/control		CID-GFP/H3.3-GFP		Experimental evidence	Construct
	log2	p-value	log2	p-value		
RhoGAP 54D	3.5 ±1.8E-01	1,51E-03	3.4 ±4.7E-02	9,81E-04		
CG8478	3.8 ±8.1E-02	1,85E-02	3.6 ±3.3E-02	1,81E-02		
Top3beta	3.4 ±6.5E-02	3,40E-02	3.4 ±9.4E-02	3,17E-03		
Ge-1	3.0 ±7.0E-02	2,53E-02	2.8 ±7.1E-02	2,03E-02		
Hcf	3.5 ±2.7E-02	2,44E-01	2.7 ±3.1E-02	3,21E-01	Nuclear	GFP Invitrogen

Grey: factors not contained in 85 initially identified CID-chromatin co-purifiers and selected based on their high and CID-GFP specific iBAQ values

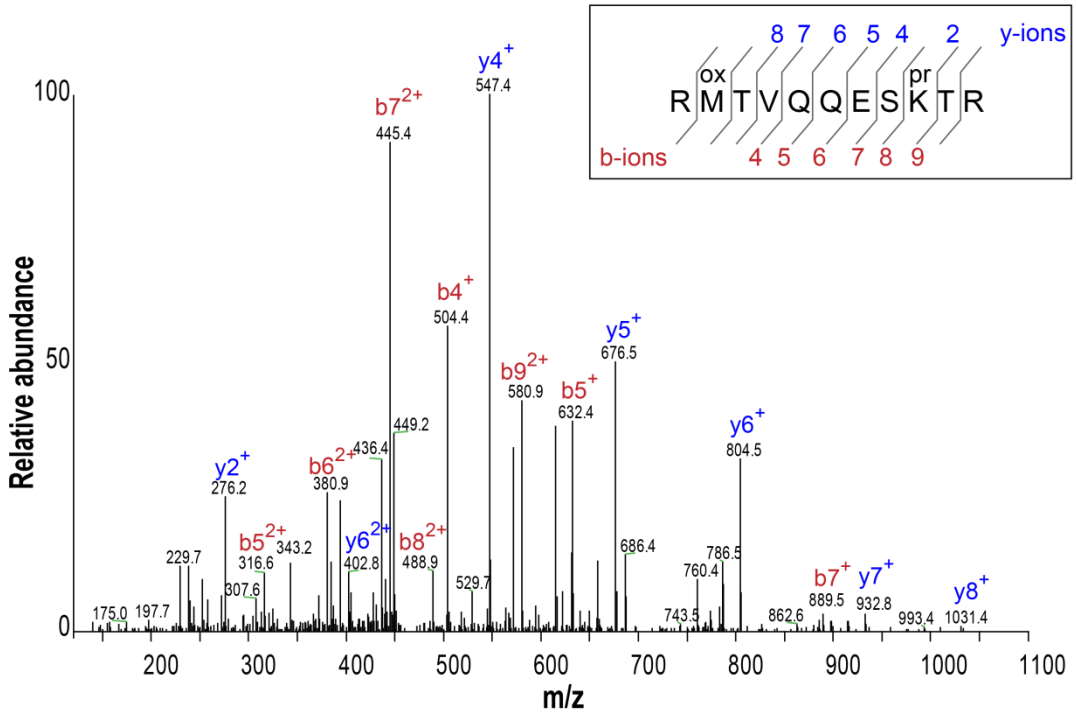

Figure 1: Identification of CID-GFP in the band excised from the gel in figure 2.6. MS2 spectrum of a proteotypic peptide of CID (parent mass: 479.24948, z=3). The peptide sequence is shown in the inlet. ox: oxidated methionine. pr: propionylated lysine. The b- and y-ions assigned in the spectrum are indicated in the inlet. The complete b- and y-ions series was detected by the software Proteome Discoverer.

ABBREVIATIONS

5-mC	5-methylcytosine
ACN	Acetonitrile
amu	Atomic mass unit
ANOVA	Analysis of variance
AP-MS	Affinity purification – mass spectrometry
APC/C	Anaphase promoting complex/cyclosome
Asf1	Anti-silencing function protein 1
BDGP	Berkeley Drosophila Genome Project
bp	Basepair(s)
CAF-1, p55	Chromatin assembly factor 1
Cal1	Chromosome alignment defect 1
CCAN	Constitutive centromere-associated network
Cdc20	Cell division cycle protein 20
CDK	Cyclin-dependent kinase
CENP	Centromeric protein
CID	Centromere identifier
CID-ub	Monoubiquitinated band of CID
CPC	Chromosomal passenger complex
CREST	calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasias
D. mel	*Drosophila melanogaster*
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
DTT	Dithiothreitol
ESI	Electrospray ionization
Ezh1/2	Enhancer of zeste homolog ½
FA	Formic acid
FACS	Fluorescence-activated cell sorting
fw	Forward
GFP	Green fluorescent protein
H1	Histone 1
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3S10ph	Histone 3 serine 10 phosphorylation
H4	Histone 4
HAT	Histone acetyl transferase
HJURP	Holliday junction recognition protein
IAA	Iodoacetamide
iBAQ	Intensity based absolute quantification
IF	Immunofluorescence
ING	Inhibitor of growth

IP	Immunoprecipitation
kb	Kilobase(s)
kDa	Kilodalton
KMN	Knl1 complex, Mis12 complex and Ndc80 complex
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
m/z	Mass-to-charge
Mb	Megabase(s)
MCM	Minichromosome maintenance
min	Minute(s)
mL	Milliliter
MNase	Micrococcal nuclease
MS	Mass spectrometry
MT	Microtubule
MTOC	Microtubule organizing center
NaCl	Sodium chloride
ncRNA	Non-coding RNA
NEM	N-ethylmaleimide
NLP	Nucleoplasmin-like protein
ORC3	Origin recognition complex
PAGE	Polyacrylamide gel electrophoresis
PHD	Plant homeodomain
PLK	Polo-like kinase
pMT	Metallothionein promoter
ppa	Partner of paired
ppm	Parts per million
PRC1/2	Polycomb repressive complex
prod	Proliferation disruptor
PTM	Posttranslational modification
rev	Reverse
RNA	Ribonucleic acid
RNAi	RNA interference
RP-LC	Reversed phase – liquid chromatography
SAC	Spindle assembly checkpoint
SDS	Sodium dodecyl sulfate
SID	Sample identifier
siRNA	Small interfering RNA
TAP	Tandem affinity purification
TET	Ten-eleven-translocation
TFA	Trifluoroacetic acid
TFIID	Transcription factor II D
wt	Wildtype
XIC	Extracted ion chromatogram

REFERENCES

Alvarez, F., F. Munoz, P. Schilcher, A. Imhof, G. Almouzni and A. Loyola (2011). "Sequential establishment of marks on soluble histones H3 and H4." J Biol Chem **286**(20): 17714-17721.

Annunziato, A. T. and R. L. Seale (1983). "Histone deacetylation is required for the maturation of newly replicated chromatin." J Biol Chem **258**(20): 12675-12684.

Bayne, E. H., M. Portoso, A. Kagansky, I. C. Kos-Braun, T. Urano, K. Ekwall, F. Alves, J. Rappsilber and R. C. Allshire (2008). "Splicing factors facilitate RNAi-directed silencing in fission yeast." Science **322**(5901): 602-606.

Benjamini, Y. and Y. Hochberg (1995). "Controllling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.

Biterge, B. and R. Schneider (2014). "Histone variants: key players of chromatin." Cell Tissue Res **356**(3): 457-466.

Blower, M. D. and G. H. Karpen (2001). "The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions." Nat Cell Biol **3**(8): 730-739.

Blower, M. D., B. A. Sullivan and G. H. Karpen (2002). "Conserved organization of centromeric chromatin in flies and humans." Dev Cell **2**(3): 319-330.

Blum, H., H. Beier and H. J. Gross (1987). "Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels." Electrophoresis **8**(2): 93-99.

Boveri, T. (1902). "Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns." Verh. d. phys.-med. Ges. Würzburg **35**: 67-90.

Boveri, T. (1914). "Zur Frage der Entstehung maligner Tumoren." Gustav Fischer Verlag.

Boyer, L. A., R. R. Latek and C. L. Peterson (2004). "The SANT domain: a unique histone-tail-binding module?" Nat Rev Mol Cell Biol **5**(2): 158-163.

Bunch, T. A., Y. Grinblat and L. S. Goldstein (1988). "Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells." Nucleic Acids Res **16**(3): 1043-1061.

Burrack, L. S. and J. Berman (2012). "Neocentromeres and epigenetically inherited features of centromeres." Chromosome Res **20**(5): 607-619.

Cai, Y., J. Jin, L. Florens, S. K. Swanson, T. Kusch, B. Li, J. L. Workman, M. P. Washburn, R. C. Conaway and J. W. Conaway (2005). "The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes." J Biol Chem **280**(14): 13665-13670.

Carlsten, J. O., Z. Szilagyi, B. Liu, M. D. Lopez, E. Szaszi, I. Djupedal, T. Nystrom, K. Ekwall, C. M. Gustafsson and X. Zhu (2012). "Mediator promotes

CENP-a incorporation at fission yeast centromeres." Mol Cell Biol **32**(19): 4035-4043.

Carlsten, J. O., X. Zhu and C. M. Gustafsson (2013). "The multitalented Mediator complex." Trends Biochem Sci **38**(11): 531-537.

Cesario, J. M., J. K. Jang, B. Redding, N. Shah, T. Rahman and K. S. McKim (2006). "Kinesin 6 family member Subito participates in mitotic spindle assembly and interacts with mitotic regulators." J Cell Sci **119**(Pt 22): 4770-4780.

Chan, F. L. and L. H. Wong (2012). "Transcription in the maintenance of centromere chromatin identity." Nucleic Acids Res **40**(22): 11178-11188.

Chen, C. C., M. L. Dechassa, E. Bettini, M. B. Ledoux, C. Belisario, P. Heun, K. Luger and B. G. Mellone (2014). "CAL1 is the Drosophila CENP-A assembly factor." J Cell Biol **204**(3): 313-329.

Clancy, J. L., M. J. Henderson, A. J. Russell, D. W. Anderson, R. J. Bova, I. G. Campbell, D. Y. Choong, G. A. Macdonald, G. J. Mann, T. Nolan, G. Brady, O. I. Olopade, E. Woollatt, M. J. Davies, D. Segara, N. F. Hacker, S. M. Henshall, R. L. Sutherland and C. K. Watts (2003). "EDD, the human orthologue of the hyperplastic discs tumour suppressor gene, is amplified and overexpressed in cancer." Oncogene **22**(32): 5070-5081.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annu Rev Biochem **78**: 273-304.

Cleveland, D. W., Y. Mao and K. F. Sullivan (2003). "Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling." Cell **112**(4): 407-421.

Dai, M. S., X. X. Sun, J. Qin, S. M. Smolik and H. Lu (2004). "Identification and characterization of a novel Drosophila melanogaster glutathione S-transferase-containing FLYWCH zinc finger protein." Gene **342**(1): 49-56.

Dambacher, S., W. Deng, M. Hahn, D. Sadic, J. Frohlich, A. Nuber, C. Hoischen, S. Diekmann, H. Leonhardt and G. Schotta (2012). "CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin." Nucleus **3**(1): 101-110.

Dejardin, J. and R. E. Kingston (2009). "Purification of proteins associated with specific genomic Loci." Cell **136**(1): 175-186.

Delatte, B. and F. Fuks (2013). "TET proteins: on the frenetic hunt for new cytosine modifications." Brief Funct Genomics **12**(3): 191-204.

Dorn, R. and V. Krauss (2003). "The modifier of mdg4 locus in Drosophila: functional complexity is resolved by trans splicing." Genetica **117**(2-3): 165-177.

Doueiri, R., R. Anupam, M. Kvaratskhelia, K. B. Green, M. D. Lairmore and P. L. Green (2012). "Comparative host protein interactions with HTLV-1 p30 and HTLV-2 p28: insights into difference in pathobiology of human retroviruses." Retrovirology **9**: 64.

Dunleavy, E., A. Pidoux and R. Allshire (2005). "Centromeric chromatin makes its mark." Trends Biochem Sci **30**(4): 172-175.

Earnshaw, W. C. and N. Rothfield (1985). "Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma." Chromosoma **91**(3-4): 313-321.

Elbashir, S. M., J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-498.

Erhardt, S., B. G. Mellone, C. M. Betts, W. Zhang, G. H. Karpen and A. F. Straight (2008). "Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation." J Cell Biol **183**(5): 805-818.

Fang, X. and P. Zhang (2011). "Aneuploidy and tumorigenesis." Semin Cell Dev Biol **22**(6): 595-601.

Foltz, D. R., L. E. Jansen, A. O. Bailey, J. R. Yates, 3rd, E. A. Bassett, S. Wood, B. E. Black and D. W. Cleveland (2009). "Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP." Cell **137**(3): 472-484.

Foltz, D. R., L. E. Jansen, B. E. Black, A. O. Bailey, J. R. Yates, 3rd and D. W. Cleveland (2006). "The human CENP-A centromeric nucleosome-associated complex." Nat Cell Biol **8**(5): 458-469.

Franceschini, A., D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering and L. J. Jensen (2013). "STRING v9.1: protein-protein interaction networks, with increased coverage and integration." Nucleic Acids Res **41**(Database issue): D808-815.

Fujita, Y., T. Hayashi, T. Kiyomitsu, Y. Toyoda, A. Kokubu, C. Obuse and M. Yanagida (2007). "Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1." Dev Cell **12**(1): 17-30.

Furuyama, S. and S. Biggins (2007). "Centromere identity is specified by a single centromeric nucleosome in budding yeast." Proc Natl Acad Sci U S A **104**(37): 14706-14711.

Furuyama, T., Y. Dalal and S. Henikoff (2006). "Chaperone-mediated assembly of centromeric chromatin in vitro." Proc Natl Acad Sci U S A **103**(16): 6172-6177.

Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, Jr., K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant and J. M. Rothberg (2003). "A protein interaction map of Drosophila melanogaster." Science **302**(5651): 1727-1736.

Glover, D. M., M. H. Leibowitz, D. A. McLean and H. Parry (1995). "Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles." Cell **81**(1): 95-105.

Goshima, G., R. Wollman, S. S. Goodwin, N. Zhang, J. M. Scholey, R. D. Vale and N. Stuurman (2007). "Genes required for mitotic spindle assembly in Drosophila S2 cells." Science **316**(5823): 417-421.

Guruharsha, K. G., J. F. Rual, B. Zhai, J. Mintseris, P. Vaidya, N. Vaidya, C. Beekman, C. Wong, D. Y. Rhee, O. Cenaj, E. McKillip, S. Shah, M. Stapleton, K. H. Wan, C. Yu, B. Parsa, J. W. Carlson, X. Chen, B. Kapadia, K. VijayRaghavan, S. P. Gygi, S. E. Celniker, R. A. Obar and S. Artavanis-Tsakonas (2011). "A protein complex network of Drosophila melanogaster." Cell **147**(3): 690-703.

Hahn, M., S. Dambacher and G. Schotta (2010). "Heterochromatin dysregulation in human diseases." J Appl Physiol (1985) **109**(1): 232-242.

Hansen, J. C. (2012). "Human mitotic chromosome structure: what happened to the 30-nm fibre?" EMBO J **31**(7): 1621-1623.

Happel, N. and D. Doenecke (2009). "Histone H1 and its isoforms: contribution to chromatin structure and function." Gene **431**(1-2): 1-12.

Heintz, N., H. L. Sive and R. G. Roeder (1983). "Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle." Mol Cell Biol **3**(4): 539-550.

Heitz, E. (1928). "Das Heterochromatin der Moose." 1. Jahrb. wiss. Bot.(69): 762-818.

Henikoff, S., K. Ahmad and H. S. Malik (2001). "The centromere paradox: stable inheritance with rapidly evolving DNA." Science **293**(5532): 1098-1102.

Henikoff, S. and Y. Dalal (2005). "Centromeric chromatin: what makes it unique?" Curr Opin Genet Dev **15**(2): 177-184.

Heun, P., S. Erhardt, M. D. Blower, S. Weiss, A. D. Skora and G. H. Karpen (2006). "Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores." Dev Cell **10**(3): 303-315.

Hochegger, H., N. Hegarat and J. B. Pereira-Leal (2013). "Aurora at the pole and equator: overlapping functions of Aurora kinases in the mitotic spindle." Open Biol **3**(3): 120185.

Hori, T., W. H. Shang, A. Toyoda, S. Misu, N. Monma, K. Ikeo, O. Molina, G. Vargiu, A. Fujiyama, H. Kimura, W. C. Earnshaw and T. Fukagawa (2014). "Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly." Dev Cell **29**(6): 740-749.

Hsu, J. Y., Z. W. Sun, X. Li, M. Reuben, K. Tatchell, D. K. Bishop, J. M. Grushcow, C. J. Brame, J. A. Caldwell, D. F. Hunt, R. Lin, M. M. Smith and C. D. Allis (2000). "Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes." Cell **102**(3): 279-291.

Jansen, L. E., B. E. Black, D. R. Foltz and D. W. Cleveland (2007). "Propagation of centromeric chromatin requires exit from mitosis." J Cell Biol **176**(6): 795-805.

Jeong, A. L. and Y. Yang (2013). "PP2A function toward mitotic kinases and substrates during the cell cycle." BMB Rep **46**(6): 289-294.

Kaufman, P. D., R. Kobayashi, N. Kessler and B. Stillman (1995). "The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication." Cell **81**(7): 1105-1114.

Keppler, A., S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson (2003). "A general method for the covalent labeling of fusion proteins with small molecules in vivo." Nat Biotechnol **21**(1): 86-89.

Keppler, A., H. Pick, C. Arrivoli, H. Vogel and K. Johnsson (2004). "Labeling of fusion proteins with synthetic fluorophores in live cells." Proc Natl Acad Sci U S A **101**(27): 9955-9959.

Kim, S. and H. Yu (2011). "Mutual regulation between the spindle checkpoint and APC/C." Semin Cell Dev Biol **22**(6): 551-558.

Lee, J. D., K. Amanai, A. Shearn and J. E. Treisman (2002). "The ubiquitin ligase Hyperplastic discs negatively regulates hedgehog and decapentaplegic expression by independent mechanisms." Development **129**(24): 5697-5706.

Lidsky, P. V., F. Sprenger and C. F. Lehner (2013). "Distinct modes of centromere protein dynamics during cell cycle progression in Drosophila S2R+ cells." J Cell Sci **126**(Pt 20): 4782-4793.

Linder, P. (2006). "Dead-box proteins: a family affair--active and passive players in RNP-remodeling." Nucleic Acids Res **34**(15): 4168-4180.

Loyola, A. and G. Almouzni (2004). "Histone chaperones, a supporting role in the limelight." Biochim Biophys Acta **1677**(1-3): 3-11.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Mansfield, E., E. Hersperger, J. Biggs and A. Shearn (1994). "Genetic and molecular analysis of hyperplastic discs, a gene whose product is required for regulation of cell proliferation in Drosophila melanogaster imaginal discs and germ cells." Dev Biol **165**(2): 507-526.

Masson, P., O. Andersson, U. M. Petersen and P. Young (2001). "Identification and characterization of a Drosophila nuclear proteasome regulator. A homolog of human 11 S REGgamma (PA28gamma)." J Biol Chem **276**(2): 1383-1390.

Mateo, F., M. Vidal-Laliena, M. J. Pujol and O. Bachs (2010). "Acetylation of cyclin A: a new cell cycle regulatory mechanism." Biochem Soc Trans **38**(Pt 1): 83-86.

Mellone, B. G., K. J. Grive, V. Shteyn, S. R. Bowers, I. Oderberg and G. H. Karpen (2011). "Assembly of Drosophila centromeric chromatin proteins during mitosis." PLoS Genet **7**(5): e1002068.

Melters, D. P., L. V. Paliulis, I. F. Korf and S. W. Chan (2012). "Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis." Chromosome Res **20**(5): 579-593.

Mendiburo, M. J., J. Padeken, S. Fulop, A. Schepers and P. Heun (2011). "Drosophila CENH3 is sufficient for centromere formation." Science **334**(6056): 686-690.

Moreno-Moreno, O., S. Medina-Giro, M. Torras-Llort and F. Azorin (2011). "The F box protein partner of paired regulates stability of Drosophila centromeric histone H3, CenH3(CID)." Curr Biol **21**(17): 1488-1493.

Moreno-Moreno, O., M. Torras-Llort and F. Azorin (2006). "Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres." Nucleic Acids Res **34**(21): 6247-6255.

Musacchio, A. and E. D. Salmon (2007). "The spindle-assembly checkpoint in space and time." Nat Rev Mol Cell Biol **8**(5): 379-393.

Obuse, C., H. Yang, N. Nozaki, S. Goto, T. Okazaki and K. Yoda (2004). "Proteomics analysis of the centromere complex from HeLa interphase cells: UV-damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase." Genes Cells **9**(2): 105-120.

Ochs, R. L., M. A. Lischwe, W. H. Spohn and H. Busch (1985). "Fibrillarin: a new protein of the nucleolus identified by autoimmune sera." Biol Cell **54**(2): 123-133.

Olins, A. L. and D. E. Olins (1974). "Spheroid chromatin units (v bodies)." Science **183**(4122): 330-332.

Olsen, J. V., L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning and M. Mann (2005). "Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap." Mol Cell Proteomics **4**(12): 2010-2021.

Olszak, A. M., D. van Essen, A. J. Pereira, S. Diehl, T. Manke, H. Maiato, S. Saccani and P. Heun (2011). "Heterochromatin boundaries are hotspots for de novo kinetochore formation." Nat Cell Biol **13**(7): 799-808.

Orpinell, M., M. Fournier, A. Riss, Z. Nagy, A. R. Krebs, M. Frontini and L. Tora (2010). "The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates." EMBO J **29**(14): 2381-2394.

Padeken, J., M. J. Mendiburo, S. Chlamydas, H. J. Schwarz, E. Kremmer and P. Heun (2013). "The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus." Mol Cell **50**(2): 236-249.

Palmer, D. K., K. O'Day, M. H. Wener, B. S. Andrews and R. L. Margolis (1987). "A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones." J Cell Biol **104**(4): 805-815.

Parthun, M. R. (2012). "Histone acetyltransferase 1: More than just an enzyme?" Biochim Biophys Acta **1819**(3-4): 256-263.

Perpelescu, M. and T. Fukagawa (2011). "The ABCs of CENPs." Chromosoma **120**(5): 425-446.

Pertceva, J. A., N. V. Dorogova, E. U. Bolobolova, O. O. Nerusheva, S. A. Fedorova and L. V. Omelyanchuk (2010). "The role of Drosophila hyperplastic discs gene in spermatogenesis." Cell Biol Int **34**(10): 991-996.

Radford, S. J., J. K. Jang and K. S. McKim (2012). "The chromosomal passenger complex is required for meiotic acentrosomal spindle assembly and chromosome biorientation." Genetics **192**(2): 417-429.

Roseman, R. R., V. Pirrotta and P. K. Geyer (1993). "The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects." EMBO J **12**(2): 435-442.

Ross, B. D., L. Rosin, A. W. Thomae, M. A. Hiatt, D. Vermaak, A. F. de la Cruz, A. Imhof, B. G. Mellone and H. S. Malik (2013). "Stepwise evolution of essential centromere function in a Drosophila neogene." Science **340**(6137): 1211-1214.

Rothbart, S. B., S. Lin, L. M. Britton, K. Krajewski, M. C. Keogh, B. A. Garcia and B. D. Strahl (2012). "Poly-acetylated chromatin signatures are preferred epitopes for site-specific histone H4 acetyl antibodies." Sci Rep **2**: 489.

Schneider, I. (1972). "Cell lines derived from late embryonic stages of Drosophila melanogaster." J Embryol Exp Morphol **27**(2): 353-365.

Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.

Simon, J. A. and R. E. Kingston (2009). "Mechanisms of polycomb gene silencing: knowns and unknowns." Nat Rev Mol Cell Biol **10**(10): 697-708.

Smith, S. and B. Stillman (1989). "Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro." Cell **58**(1): 15-25.

Sobel, R. E., R. G. Cook, C. A. Perry, A. T. Annunziato and C. D. Allis (1995). "Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4." Proc Natl Acad Sci U S A **92**(4): 1237-1241.

Somma, M. P., F. Ceprani, E. Bucciarelli, V. Naim, V. De Arcangelis, R. Piergentili, A. Palena, L. Ciapponi, M. G. Giansanti, C. Pellacani, R. Petrucci, G. Cenci, F. Verni, B. Fasulo, M. L. Goldberg, F. Di Cunto and M. Gatti (2008). "Identification of Drosophila mitotic genes by combining co-expression analysis and RNA interference." PLoS Genet **4**(7): e1000126.

Spedale, G., H. T. Timmers and W. W. Pijnappel (2012). "ATAC-king the complexity of SAGA during evolution." Genes Dev **26**(6): 527-541.

Squatrito, M., C. Gorrini and B. Amati (2006). "Tip60 in DNA damage response and growth control: many tricks in one HAT." Trends Cell Biol **16**(9): 433-442.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Stros, M., D. Launholt and K. D. Grasser (2007). "The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins." Cell Mol Life Sci **64**(19-20): 2590-2606.

Suganuma, T., J. L. Gutierrez, B. Li, L. Florens, S. K. Swanson, M. P. Washburn, S. M. Abmayr and J. L. Workman (2008). "ATAC is a double

histone acetyltransferase complex that stimulates nucleosome sliding." Nat Struct Mol Biol **15**(4): 364-372.

Suganuma, T., S. G. Pattenden and J. L. Workman (2008). "Diverse functions of WD40 repeat proteins in histone recognition." Genes Dev **22**(10): 1265-1268.

Sullivan, B. and G. Karpen (2001). "Centromere identity in Drosophila is not determined in vivo by replication timing." J Cell Biol **154**(4): 683-690.

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

Sung, M. T. and G. H. Dixon (1970). "Modification of histones during spermiogenesis in trout: a molecular mechanism for altering histone binding to DNA." Proc Natl Acad Sci U S A **67**(3): 1616-1623.

Sutherland, B. W., J. Toews and J. Kast (2008). "Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions." J Mass Spectrom **43**(6): 699-715.

Tabb, D. L., L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, A. J. Ham, D. M. Bunk, L. E. Kilpatrick, D. D. Billheimer, R. K. Blackman, H. L. Cardasis, S. A. Carr, K. R. Clauser, J. D. Jaffe, K. A. Kowalski, T. A. Neubert, F. E. Regnier, B. Schilling, T. J. Tegeler, M. Wang, P. Wang, J. R. Whiteaker, L. J. Zimmerman, S. J. Fisher, B. W. Gibson, C. R. Kinsinger, M. Mesri, H. Rodriguez, S. E. Stein, P. Tempst, A. G. Paulovich, D. C. Liebler and C. Spiegelman (2010). "Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry." J Proteome Res **9**(2): 761-776.

Tafrova, J. I. and S. T. Tafrov (2014). "Human histone acetyltransferase 1 (Hat1) acetylates lysine 5 of histone H2A in vivo." Mol Cell Biochem **392**(1-2): 259-272.

Tallen, G. and K. Riabowol (2014). "Keep-ING balance: Tumor suppression by epigenetic regulation." FEBS Lett **588**(16): 2728-2742.

Thomae, A. W., G. O. Schade, J. Padeken, M. Borath, I. Vetter, E. Kremmer, P. Heun and A. Imhof (2013). "A pair of centromeric proteins mediates reproductive isolation in Drosophila species." Dev Cell **27**(4): 412-424.

Torok, T., C. Benitez, S. Takacs and H. Biessmann (2007). "The protein encoded by the gene proliferation disrupter (prod) is associated with the telomeric retrotransposon array in Drosophila melanogaster." Chromosoma **116**(2): 185-195.

Torok, T., P. D. Harvie, M. Buratovich and P. J. Bryant (1997). "The product of proliferation disrupter is concentrated at centromeres and required for mitotic chromosome condensation and cell proliferation in Drosophila." Genes Dev **11**(2): 213-225.

Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. McQuilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, H. Zhang and C. FlyBase (2009). "FlyBase: enhancing Drosophila Gene Ontology annotations." Nucleic Acids Res **37**(Database issue): D555-559.

Varma, D. and E. D. Salmon (2012). "The KMN protein network--chief conductors of the kinetochore orchestra." J Cell Sci **125**(Pt 24): 5927-5936.

Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann and H. T. Timmers (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." Cell **131**(1): 58-69.

Verreault, A., P. D. Kaufman, R. Kobayashi and B. Stillman (1998). "Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase." Curr Biol **8**(2): 96-108.

Westermann, S. and A. Schleiffer (2013). "Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans." Trends Cell Biol **23**(6): 260-269.

Will, C. L., H. Urlaub, T. Achsel, M. Gentzel, M. Wilm and R. Luhrmann (2002). "Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein." EMBO J **21**(18): 4978-4988.

Winkler, D. D., H. Zhou, M. A. Dar, Z. Zhang and K. Luger (2012). "Yeast CAF-1 assembles histone (H3-H4)2 tetramers prior to DNA deposition." Nucleic Acids Res **40**(20): 10139-10149.

Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." Science **322**(5902): 750-756.