**Université de Montréal**

# Interpretation of the centromere epigenetic mark to maintain genome stability

**Par Valérie De Rop**

**Unité académique de biologie moléculaire option biologie des systèmes**
**Institut de Recherche en Immunologie et en Cancérologie (IRIC)**
**Département de pathologie et biologie cellulaire**
**Faculté de Médecine**

Thèse présentée à la Faculté de médecine en vue de l'obtention du grade de Ph.D.
en biologie moléculaire option biologie des systèmes

Avril, 2014

## RÉSUMÉ

Le centromère est la région chromosomique où le kinétochore s'assemble en mitose. Contrairement à certaines caractéristiques géniques, la séquence centromérique n'est ni conservée entre les espèces ni suffisante à la fonction centromérique. Il est donc bien accepté dans la littérature que le centromère est régulé épigénétiquement par une variante de l'histone H3, CENP-A. KNL-2, aussi connu sous le nom de M18BP1, ainsi que ces partenaires Mis18α et Mis18β sont des protéines essentielles pour l'incorporation de CENP-A nouvellement synthétisé aux centromères. Des évidences expérimentales démontrent que KNL-2, ayant un domaine de liaison à l'ADN nommé Myb, est la protéine la plus en amont pour l'incorporation de CENP-A aux centromères en phase G1. Par contre, sa fonction dans le processus d'incorporation de CENP-A aux centromères n'est pas bien comprise et ces partenaires de liaison ne sont pas tous connus.

De nouveaux partenaires de liaison de KNL-2 ont été identifiés par des expériences d'immunoprécipitation suivies d'une analyse en spectrométrie de masse. Un rôle dans l'incorporation de CENP-A nouvellement synthétisé aux centromères a été attribué à MgcRacGAP, une des 60 protéines identifiées par l'essai. MgcRacGAP ainsi que les protéines ECT-2 (GEF) et la petite GTPase Cdc42 ont été démontrées comme étant requises pour la stabilité de CENP-A incorporé aux centromères. Ces différentes observations ont mené à l'identification d'une troisième étape au niveau moléculaire pour l'incorporation de CENP-A nouvellement synthétisé en phase G1, celle de la stabilité de CENP-A nouvellement incorporé aux centromères. Cette étape est importante pour le maintien de l'identité centromérique à chaque division cellulaire.

Pour caractériser la fonction de KNL-2 lors de l'incorporation de CENP-A nouvellement synthétisé aux centromères, une technique de microscopie à haute

résolution couplée à une quantification d'image a été utilisée. Les résultats générés démontrent que le recrutement de KNL-2 au centromère est rapide, environ 5 minutes après la sortie de la mitose. De plus, la structure du domaine Myb de KNL-2 provenant du nématode *C. elegans* a été résolue par RMN et celle-ci démontre un motif hélice-tour-hélice, une structure connue pour les domaines de liaison à l'ADN de la famille Myb. De plus, les domaines humain (HsMyb) et *C. elegans* (CeMyb) Myb lient l'ADN *in vitro*, mais aucune séquence n'est reconnue spécifiquement par ces domaines. Cependant, il a été possible de démontrer que ces deux domaines lient préférentiellement la chromatine CENP-A-YFP comparativement à la chromatine H2B-GFP par un essai modifié de SIMPull sous le microscope TIRF. Donc, le domaine Myb de KNL-2 est suffisant pour reconnaître de façon spécifique la chromatine centromérique.

Finalement, l'élément reconnu par les domaines Myb *in vitro* a potentiellement été identifié. En effet, il a été démontré que les domaines HsMyb et CeMyb lient l'ADN simple brin *in vitro*. De plus, les domaines HsMyb et CeMyb ne colocalisent pas avec CENP-A lorsqu'exprimés dans les cellules HeLa, mais plutôt avec les corps nucléaires PML, des structures nucléaires composées d'ARN. Donc, en liant potentiellement les transcrits centromériques, les domaines Myb de KNL-2 pourraient spécifier l'incorporation de CENP-A nouvellement synthétisé uniquement aux régions centromériques.

## MOTS CLÉS

Centromère • CENP-A • MgcRacGAP • KNL-2 • Domaine Myb • Chromatine centromérique • ARN • RMN • Microscopie à haute résolution • TIRF

## ABSTRACT

Centromeres are chromosomal loci that direct kinetochore assembly in mitosis. Unlike genetic features, centromere DNA sequence is not conserved through phylogeny nor is it sufficient for centromere function. Therefore, it is commonly accepted that centromeres are epigenetically defined; a process mediated by the histone H3 variant CENP-A. KNL-2, also called M18BP1, together with its partners Mis18α and Mis18β is essential for newly synthesized CENP-A incorporation at centromeres in *C. elegans* and in humans. Evidence from the literature suggests KNL-2, having a predicted Myb DNA binding domain, as the most upstream player for CENP-A loading in G1. However, its actual function for CENP-A incorporation at the centromere and its binding partners required for the incorporation of CENP-A remained elusive.

New binding partners of KNL-2 were identified by immunoprecipitation experiments followed by mass spectrometry analysis. MgcRacGAP is one of the 60 hits identified and its role in the newly synthesized CENP-A incorporation pathway, together with the GEF ECT-2 and the small GTPase Cdc42, was confirmed by shRNA depletion. More interestingly, those proteins are required for the stability and the maintenance of incorporated CENP-A at centromeres. These observations lead to the identification of a third step in the CENP-A incorporation pathway important for the centromere identity maintenance over subsequent cell divisions.

To characterize the function of KNL-2 in the newly synthesized CENP-A incorporation pathway, high-resolution microscopy coupled to image quantification shows a rapid recruitment of KNL-2 at the centromere in early G1. Also, the predicted *C. elegans* KNL-2 Myb (CeMyb) domain structure was solved by NMR. This revealed an expected helix-loop-helix structure; which is the same for human KNL-2 Myb (HsMyb) domain. Those Myb domains bind DNA *in vitro*, however they do not bind any specific DNA sequence. Surprisingly, specific binding of the Myb domains to human CENP-A-YFP chromatin, compared to H2B-GFP, is observed by using a modified version of

SIMPull assay under the TIRF microscope. Therefore, the KNL-2 Myb domain is sufficient to recognize and bind a specific feature generated by the presence of CENP-A nucleosomes at centromeres.

Finally, a centromeric chromatin feature recognized by the Myb domains under the TIRF microscope is proposed. HsMyb and CeMyb domains bind ssDNA, a RNA like structure, *in vitro*. Moreover, HsMyb and CeMyb domains do not co-localize with CENP-A in HeLa cells, whether with PML body staining, which are nuclear bodies known to contain RNAs. Thereby, by recognizing and binding centromeric transcripts, KNL-2 Myb domains might specify the newly synthesized CENP-A incorporation only at centromere loci.


## KEY WORDS

Centromere • CENP-A • MgcRacGAP • KNL-2 • Myb domain • Centromeric chromatin • RNA • NMR • High resolution microscopy • TIRF

# TABLE OF CONTENT

## FIGURES LIST

## TABLES LIST

## ABBREVIATIONS LIST

| | |
|---|---|
| A | Adenine |
| ACA | Anti-Centromere Antibody |
| Ada2 | ADAptor 2 |
| AFM | Atomic Force Microscopy |
| Ago1 | Argonaute 1 |
| *A. thaliana* | *Arabidopsis thaliana* |
| ATP | Adenosine TriPhosphate |
| A.U. | Arbitrary Units |
| AUC | Analytical Ultracentrifugation |
| bp | Base pair |
| BG | Benzylguanine |
| BUB1 | Budding Uninhibited by Benzimidazoles 1 |
| CAF-1 | Chromatin Assembly Factor-1 |
| CAL1 | Chromosome ALignement 1 |
| CATD | CENP-A Targeting Domain |
| CCAN | Constitutive Centromere-Associated Network |
| Cdc42 | Cell Division Cycle 42 |
| CDEI-III | Centromere DNA Element I to III |
| CDK | Cyclin-Dependent Kinase |
| CeKNL-2 | *Caenorhabditis elegans* KNL-2 |
| CenH3 | Centromeric Histone H3, also called CENP-A |
| *C. elegans* | *Caenorhabditis elegans* |
| CeMyb | *Caenorhabditis elegans* KNL-2 Myb domain |
| CENP-A | CENtromere Protein A |
| CENP-B | CENtromere Protein B |
| CENP-C | CENtromere Protein C |
| CENP-S | CENtromere Protein S |
| CENP-T | CENtromere Protein T |
| CENP-W | CENtromere Protein W |

| | |
|---|---|
| CENP-X | CENtromere Protein X |
| ChIP | Chromatin ImmunoPrecipitation |
| Chp1 | Chromodomain protein 1 |
| CID | Centromere IDentifier |
| CIN | Chromosome Instability Number |
| CLC | Centromere Licensing Complex |
| Clr4 | Cryptic Loci Regulator |
| Cnp1 | CeNtromere Protein 1 |
| CpG | Cytosine Phosphodiester bond Guanine |
| CREST | Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, and Telangiectasia |
| Cse4 | Chromosome SEgregation 4 |
| Ct | Cycle threshold |
| CTCF | Insulator protein CCCTC-binding factor |
| Cug2 | Cancer-Upregulated Gene 2 |
| CytoD | Cytochalasin D |
| DAPI | 4', 6-diamidino-2-phenylindole |
| *D. melanogaster* | *Drosophila melanogaster* |
| DMEM | Dulbecco’s modified Eagle’s medium |
| DMSO | DiMethylSulfOxide |
| DNA | DeoxyriboNucleic Acid |
| DNMT3A/3B | DNA MethylTransferase 3A and 3B |
| dsDNA | double stranded DNA |
| ECT2 | Epithelial Cell Transforming sequence 2 |
| EMSA | ElectroMobility Shift Assay |
| FACT | FAcilitates Chromatin Transcription |
| FAM | 6-carboxy- fluorescein |
| FCS | Fluorescence Correlation Spectroscopy |
| FRAP | Fluorescence Recovery After Photobleaching |
| GAP | GTPase Activating Protein |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |

| | |
|---|---|
| GDP | Guanine DiPhosphate |
| GEF | Guanine Exchange Factor |
| GFP | Green Fluorescent Protein |
| GO | Gene Ontology |
| GST | Glutathione S-Transferase |
| GTP | Guanine TriPhosphate |
| G1 | Growth phase 1 |
| G2 | Growth phase 2 |
| HA | Hemagglutinin |
| HAT | Histone Acetyl Transferase |
| HCP-3 | HoloCentric chromosome binding Protein-3 |
| HCP-4 | HoloCentric chromosome binding Protein-4 |
| H/DX-MS | Hydrogen/Deuterium eXchange technique coupled to Mass Spectrometry |
| HeLa | Henrietta Lacks |
| HJURP | Holliday Junction Recognition Protein |
| HP1 | Heterochromatin Protein 1 |
| HRPT | Hypoxanthine-guanine phosphoribosyltransferase |
| *H. sapiens* | *Homo sapiens* |
| HsKNL-2 | *Homo sapiens* KNL-2 |
| HsMis18 | *Homo sapiens* Mis18 |
| HsMyb | *Homo sapiens* KNL-2 Myb domain |
| HSQC | $^{1}H$-$^{15}N$ heteronuclear single quantum coherence |
| H2A | Histone H2A |
| H2B | Histone H2B |
| H3 | Histone H3 |
| H3K4me2 | Histone H3 Lysine 4 DiMethylation |
| H3K9 | Histone H3 Lysine 9 |
| H3K9me2 | Histone H3 Lysine 9 DiMethylation |
| H3K9m3 | Histone H3 Lysine 9 TriMethlylation |
| H3.1 | Histone H3.1, variant |

| | |
|---|---|
| H3.3 | Histone H3.3, variant |
| H4 | Histone H4 |
| KNL-2 | Kinetochore NulL-2 |
| IF | ImmunoFluorescence |
| IP | ImmunoPrecipitation |
| ITC | Isothermal Titration Calorimetry |
| LacI | Lactose Inhibitor |
| LacO | Lactose Operon |
| LatA | Latrunculin A |
| LC-MS/MS | Liquid Chromatography followed by MS/MS analysis |
| *L. nivea* | *Luzula nivea* |
| LSD1 | Lysine-Specific Demethylase 1 |
| M | Mitosis phase |
| Mbp | Mega base pair |
| MgcRacGAP | Male Germ Cell Rac GTPase Activating Protein |
| Mis16 | Mis-segregation 16 |
| Mis18 | Mis-segregation 18 |
| MKLP1 | Mitotic-Kinesin Like Protein 1 |
| MNase | Micrococcal Nuclease |
| mRNA | messenger RNA |
| MS | Mass Spectrometry |
| MTOC | MicroTubule Organization Center |
| M18BP1 | Mis18 Binding Protein 1 |
| N-CoR | Nuclear receptor Co-Repressor |
| ncRNA | NonCoding RNA |
| NDC80 | Nuclear Division Cycle 80 |
| NLP | Nucleophosmin-Like Protein |
| NMP1 | Nucleophosmin 1 |
| NMR | Nuclear Magnetic Resonance |
| PARP | Poly (ADP-Ribose) Polymerase |
| PBS | Phosphate Buffer Solution |

| | |
|---|---|
| PCR | Polymerase Chain Reaction |
| pdb | Protein databank |
| pI | Isoelectric Point |
| PML | Promyelocytic Leukemia |
| PSF | point-spread function |
| qPCR | quantitative PCR |
| ppm | Parts per million |
| Rac1 | RAs-related C3 botulinum toxin substrate 1 |
| RbAp46/48 | Rb-Associated Protein 46/48 |
| rDNA | ribosomal DNA |
| RFP | Red Fluorescent Protein |
| RhoA | Ras HOmolog gene family, member A |
| RITS | RNA-induced Initiation of Transcriptional gene Silencing |
| RNA | RiboNucleic Acid |
| RNAi | RNA interference |
| RNAse | RiboNuclease |
| RNAPII | RNA polymerase II |
| RSF1 | Remodelling and Spacing Factor 1 |
| RT-PCR | Reverse Transcription PCR |
| S | S-phase |
| SAC | Spindle Assembly Checkpoint |
| SANT | (Swi3, Ada2, N-CoR, TFIIIB) |
| SANTA | SANT Associated |
| *S. cerevisiae* | *Saccharomyces cerevisiae* |
| Scm3 | Suppressor of Chromosome Mis-segregation 3 |
| *S. frugiperda* | *Spodoptera frugiperda* |
| shRNA | Short Hairpin RNA |
| siRNA | Small Interfering RNA |
| SMC | Structural Maintenance of Chromosomes |

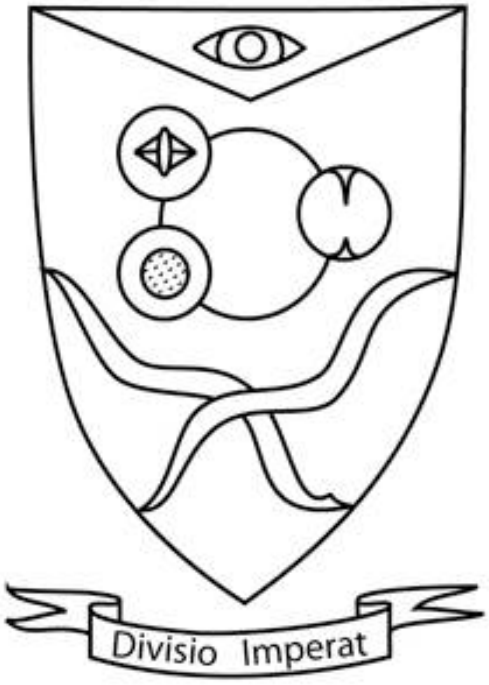
| | |
|---|---|
| SNAP | Modified variant of the suicide enzyme $O^6$-alkylguanine-DNA alkyltransferase |
| *S. pombe* | *Schizosaccharomyces pombe* |
| ssDNA | Single stranded DNA |
| Stc1 | Stanniocalcin 1 |
| Suv39 | SUppressor of Variegation 3-9 |
| Swi3 | SWItching-defective protein 3 |
| Swi6 | SWItching-defective protein 6 |
| T | Thymine |
| TFIIIB | Transcription Factor IIIB |
| TIRFm | Total Internal Reflection Fluorescence microscopy |
| TMR | TetraMethylRhodamine |
| *X. laevis* | *Xenopus laevis* |
| YFP | Yellow Fluorescent Protein |

*À Grand-papa Omer et Maman Gigi*

*Les études, c'est important!*

## ACKNOWLEDGEMENTS, REMERCIEMENTS

My PhD adventure was incredible! I learn a lot of cutting edge techniques, which I applied to my project, and I was aware of a lot of cool scientific ideas through annual meeting of ASCB (some speakers that I think will become Nobel Laureate one day). Thank you ***Paul*** for letting me come and grow in your scientific Family! The project was exciting; most of the time it was a lot of fun, and sometimes it was very hard and deceiving (none of the projects I know about are easy...). But at the end, my PhD project was very instructive! Thank you for your support, your advices, your cool ideas, and most importantly, for your trust and for believing in me! You were the perfect match to my expectations and you are an extraordinary boss!




Also, I want to thank my colleagues that helped me at the beginning of my graduate studies. ***Anaïck***, merci pour tes judicieux conseils techniques et expérimentales. Tu m'as appris la grande majorité des techniques que j'ai appliquées en laboratoire et j'ai eu beaucoup de plaisir à visiter quelques villes américaines en ta présence lors des congrès d'ASCB. J'en garde de très bons souvenirs! ***Jonas***, tu m'as montré que les formules mathématiques, la physique et l'informatique ne sont pas si complexes lorsque c'est bien expliqué. J'ai grandement apprécié tes conseils professionnels et personnels, ainsi que la bonne raclette suisse! ***Lily***, thank you for your patience, and for being calm and peaceful under certain circumstances. I enjoyed the time we spent visiting some cities in the USA, especially San Diego! It was a pleasure to see you around the lab every day! To my colleagues that helped me with my PhD project: ***Mike, Jack,***

***Abbas, Joël, Corentin and Émilie***, thank you for bringing your motivation and your expertise to this incredibly tough biological question. As I always said, KNL-2 is a BI***! Without you, it would have been impossible to publish in a high impact journal and for me to graduate in 5 years! Thanks to ***Rajesh, Abbas, Karine, Anne-Marie, Carlos, Ben, Vincent, Maxime, Patrick and Abby*** for our time complaining about a person in particular (of course Paul it's you!), for nice discussions during lunchtime, and for unforgettable time spent at ASCB meetings. A special thanks to ***Amy***, for scientific and personal advices, and to ***Jean-Claude***, merci d'avoir cru en moi et de m'avoir poussé à appliquer à la maîtrise à l'IRIC. Et finalement, merci aux membres du laboratoire de Jean-Claude avec qui j'ai partagé l'espace de bureau: ***Alexia, Rana, Laura, Marianne, Catherine, Eugénie, Nicolas et Vincent***. Nous avons eu beaucoup de plaisir durant ces dernières années!

Grâce à mes charges d'enseignement, mes années au doctorat ont passées très vite. Merci à ***Daniel, Nathalie, André, Shona, Nicole, Jacques, Philipe et Audrey*** du Département de biochimie, sans qui je n'aurais pas eu autant de plaisir avec les mitochondries de coeur de porc et *E. coli* (BiP BiP BiP)! Et merci aux ***centaines d'étudiants*** à qui j'ai eu la chance d'enseigner! J'ai eu énormément de plaisir à vous côtoyer et à vous expliquer pourquoi je vous ai enlevé des points!

Je n'aurais pas pu passer au travers de mon doctorat sans le support inconditionnel de ma famille. Merci à ma maman ***Suzanne***, mon papa ***Alain,*** mon beau-papa ***Gaétan***, ma belle-maman ***Louise***, mes frères ***Nicolas, Maxime, Tommy*** et ma soeur ***Cindie***. Même si vous ne comprenez pas ce que je fais en laboratoire et que vous n'avez pas de réponse claire à votre question: que feras-tu à la fin de ton doctorat? vous avez toujours été là pour moi. Merci à la famille de mon mari, ma belle-mère ***Johanne***, mon beau-père ***Richard***, ma belle-soeur ***Mélissa***, mon beau-frère ***Mathieu*** ainsi que leur fille ***Mila***. Votre support à chacune des étapes importantes de mon doctorat a été grandement apprécié.

À mes amis qui comprennent exactement tout ce que j'ai vécu: ***Stéphanie et Nicolas***, je ne sais pas ce que j'aurais fait sans vous! J'ai eu énormément de plaisir à enseigner à vos côtés (je me souviendrai de plusieurs fous rires!) et sans le savoir, vous avez été mes mentors à ce niveau. Merci à ***Catherine*** (à tous ces midis où nous avons dîner ensemble), ***Mila*** (pour nos escapades de quelques jours à Boston question de décompresser), ***Charly*** (parce que tu es le meilleur!), ***Louis*** (Ben voyons donc! Je te le dis que tu as bien fait d'aller en médecine), ***Jasmine*** (encore tes bébittes! Non Jas, elles ne peuvent pas survivre seules plus que trois jours!), ***Amélie*** (je ne suis pas SI intelligente que ça, c'est plutôt une question de persévérance et de motivation!), ***Mélissa*** (tu es mon petit rayon de soleil orange et vert écolo!), ***Chantal*** (we don't see each other very often but when we do it is always a pleasure!), et mon homonyme ***Valérie*** (grâce à toi, j'ai réussi à surmonter des étapes que je croyais infranchissables!).

À mon meilleur ami, mon confident, mon amoureux et mon mari ***Guillaume***. Merci d'avoir été présent et d'avoir été à l'écoute de mes questionnements, de mes peurs, de mes colères, de mon désespoir, de mes joies et de mes moments de folie. Tu as toujours eu les bons mots pour me motiver à continuer et à ne jamais lâcher. Grâce à tes grandes exigences, je n'aurai pas perdu beaucoup de temps les soirs de semaine et les fins de semaine au labo!

Merci aux organismes subventionnaires: les IRSC pour l'octroie de ma bourse à la maîtrise ainsi que la Société Canadienne du Cancer pour les subventions de recherches. Merci à aux Programmes de biologie moléculaire pour les généreuses bourses de l'excellence ainsi qu'à la Faculté des études supérieures et postdoctorales pour la bourse de fin d'études doctorales

Finalement, un merci tout spécial à Marieke Rozendaal sans qui cette thèse aurait été une vraie cacophonie!

Merci!

Thank you!

# CHAPTER 1

# INTRODUCTION

## 1.1 General organization of chapters

The **introduction** chapter describes a chromosomal locus called the centromere and how this region is specified and regulated through cell division. The different proteins involved in this process are described as well as the potential role for RNA in centromere identity and propagation. Experimental techniques used in centromere biology studies are described as well as the one used in the papers presented in this thesis.

**Chapter 2** presents a review that I wrote and it was published in Chromosoma in 2012. This review describes the latest findings in centromere biology, and focuses on specific structural details of CENP-A, showing how distinct this protein is compared to Histone H3. Centromere identity specification and CENP-A incorporation regulation by essential proteins and a variety of chromatin epigenetic marks are discussed.

**Chapter 3** presents my first co-author paper published in 2010 in Nature Cell Biology. This paper demonstrates the role of the small GTPase Cdc42 in CENP-A incorporation at the centromere. Through KNL-2 immunoprecipitation experiments, we found new binding partners that might have a role in CENP-A localization at the centromere. MgcRacGAP caught our attention and depleting this protein showed a CENP-A fluorescence intensity decrease at the centromere. Upon depletion of MgcRacGAP in the CENP-A SNAP-tag cell line, we were able to establish that this protein is required for the stability of newly incorporated CENP-A at the centromere. In fact, after depletion of MgcRacGAP, only old CENP-A was detected at the centromere. Since MgcRacGAP is the GAP of the Rho GTPase family, depletion of RhoA, Rac1, and Cdc42 were also performed. We found that Cdc42 is required for CENP-A localization at the centromere. Thus, we discovered a new maintenance step for CENP-A incorporation at the centromere in G1 phase.

**Chapter 4** presents my co-first author paper under revision in Developmental Cell. This paper demonstrates the role of KNL-2 as a true licensing factor for CENP-A incorporation at the centromere. We depicted the molecular function of the predicted Myb domain by using biochemical and cell biological approaches to show a Myb binding preference of CENP-A-GFP chromatin over H2B-GFP chromatin. Though not conserved at the sequence level, we demonstrated that both HsMyb and CeMyb domains preferentially bind human CENP-A-GFP chromatin over H2B-GFP chromatin. Moreover, we solved the 3D NMR structure of the CeMyb domain (Protein DataBank accession number: 2m3a). Thus, the Myb domain is sufficient to recognize and bind the centromeric chromatin, and this preference appears to be conserved through metazoans.

**Chapter 5** presents **preliminary data**, which demonstrate the capability of HsMyb and CeMyb domain to bind single stranded DNA *in vitro*. When transfected in HeLa cells, those domains form nuclear puncta without colocalizing with CENP-A. Surprisingly, immunofluorescence experiments showed a colocalization of those punctae with the PML bodies, nuclear structures known to contain RNA. This observation suggests a potential, but highly hypothetic, role of the KNL-2 Myb domain binding to RNA.

**Chapter 6** will discuss the different breakthroughs that the papers, presented in this thesis, are bringing in the centromere biology field. A perspective section describing experiments that can be performed following the presented publications is also discussed in this chapter.

## 1.2 Cell division

Development of various living organisms resides in the competence of cells to divide. The goal of cell division is to replicate the genetic material called DNA (DeoxyriboNucleic Acid) and to segregate it equally into two daughter cells, giving rise to an increase in cell population. Intrinsic and extrinsic stimuli are required for the activation or inhibition of molecular mechanisms controlling cell division, and they are well tuned, since a defect in those mechanisms can lead to cell death or persistent proliferation, the cause of cancer in complex eukaryotes.

In eukaryotic cells, several ordered molecular processes have to occur to initiate and complete cell division. These highly regulated processes are part of a cellular mechanism called the cell cycle (Morgan, 2007), and it is divided in four major phases: G1, S, G2, and M (Figure 1.1). In G1, some cells may arrest their cell cycle progression, and stop dividing for a long period of time. In this case, the cells exit the cell cycle, and enter a quiescent phase called G0. This cellular process is reversible; if intrinsic and extrinsic stimuli are favourable for cell division, cells can re-enter the cell cycle in G1 and undergo cell division.

The different phases of the cell cycle were first observed and described by Anton Schneider and Walther Flemming in the nineteen century (Schneider, 1873; Flemming, 1882). The latter is famous for his stunning schematic of cells in different cell cycle stages, and for demonstrating the presence of some cellular features (e.g. centrioles, microtubules, and chromatin), and the dynamics of DNA «threads», later called chromosomes. Those drawings were made using a simplified light microscope, and since then, a lot of research has been performed to better understand the different stages of the cell cycle by using modern fluorescence microscopy and advanced biochemical approaches. The different molecular components and their

dynamics have been extensively studied in the last decades and specific cell cycle mechanisms have been described.

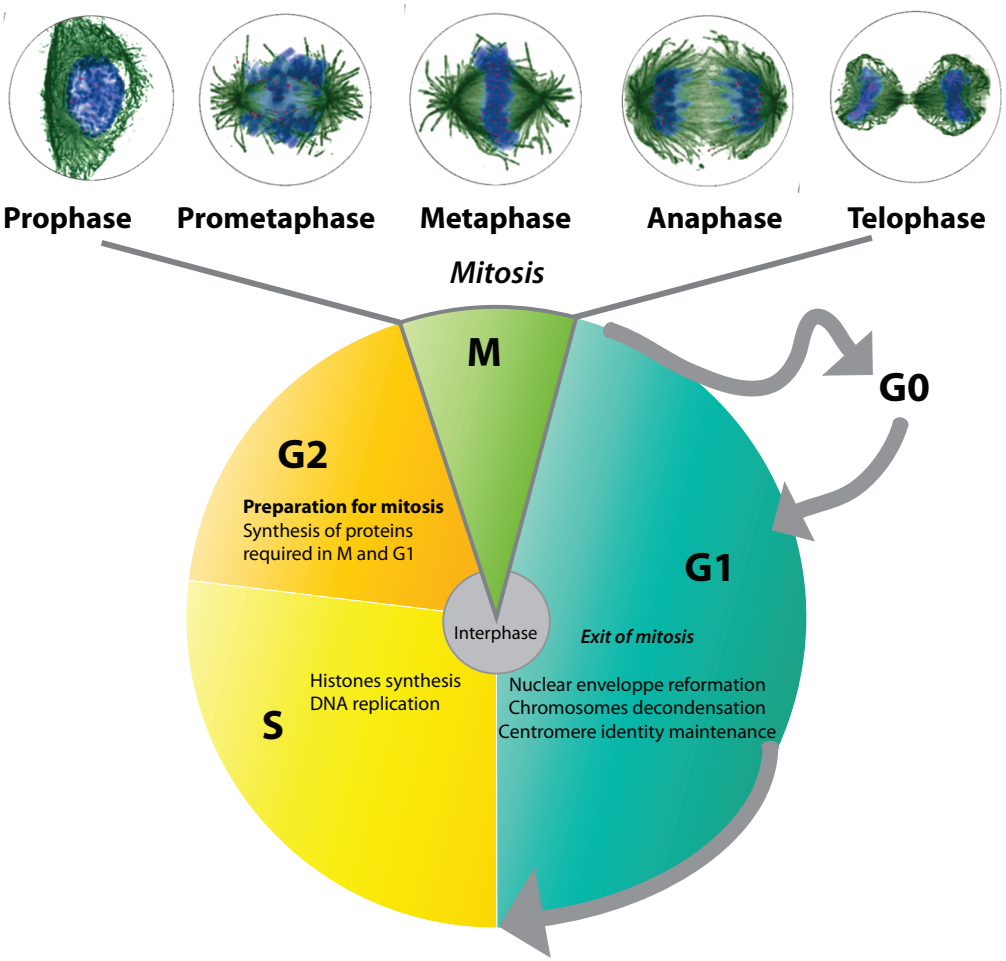



**Figure 1.1 Schematic of the different phases of the cell cycle.** The pie chart was divided based on the time human cells spend in each phase (% time of one phase/total time of the cell cycle). *The Mitosis images were adapted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology (Cheeseman and Desai, 2008), copyright 2008.*

### 1.2.1 S phase

S phase takes around 6 hours to be completed in human cells. It is during this time window that cells replicate their genome and re-arrange their chromatin structure. When origins of replication are fired all along the chromosomes, the replication fork components double the amount of DNA by copying each strand of DNA (5'-3' and 3'-5'), giving rise to a duplicated chromosome pair called sister chromatids. The pre-existing DNA is wrapped around a protein complex called the histone octamer, which is composed of two of each canonical histone proteins: H2A, H2B, H3, and H4. Many

octamers are disrupted when the replication fork passes, and re-assembled after the replication fork on the nascent double stranded DNAs, a reaction called the parental nucleosome segregation (Krude, 1999). The pre-existing histone complexes (H3/H4 tetramers) are diluted between the two double stranded DNAs and gaps have to be filled by newly synthesized histones, expressed at the beginning of S phase, a reaction called de novo nucleosome assembly (Verreault, 2000). At the end of this phase, the amount of genetic material has doubled, and the cell is ready to further compact its chromatin into chromosomes in order to segregate it in M phase.

### 1.2.2 G2 phase

G2 phase is a cell cycle gap between S phase and M phase whose completion requires around 4 hours in human cells. During this short time window of interphase, cells prepare to enter mitosis by a period of rapid growth. The proteins required for M and G1 phases are expressed and cells decide to enter mitosis or not through the action of a cell cycle control system called the G2/M checkpoint (Morgan, 2007). Cells having DNA damage (single strand breaks, double strand breaks, mismatch mutations or covalent modification of base pairs) that occurred throughout replication can pause their cell cycle progression. This delay will give them time to repair the damage and prevent further complications throughout the cell cycle.

### 1.2.3 M phase

M phase is the most described and studied phase of the cell cycle, historically because of the striking cell morphology changes that occur during this time window (around 2 hours). It is composed of two major events: mitosis, which is the nuclear division (separation and segregation of chromosomes) and cytokinesis, which is the cell division (partition of the cytoplasm into daughter cells by the ingression of a physical barrier called the cytokinetic ring).

#### 1.2.3.1 Prophase and Prometaphase

For chromosomes to be equally segregated into two daughter cells, they have to be condensed into their characteristic «X» shape. Many factors are involved in the condensation process: Histones (H1, H2A, H2B, H3, and H4), SMC (Structural Maintenance of Chromosomes) complexes composed of Condensin and Cohesin proteins, and Topoisomerase II. These protein complexes function together in order to condense approximately four meters of DNA in a ~5µm diameter nucleus to make fully condensed chromosomes (Uhlmann, 2001). Throughout Prophase, centrioles are duplicated and the cytoskeletal microtubules are re-organized to form the mitotic spindle. At Prometaphase, the nuclear envelope breaks down, thereby allowing the microtubules of the mitotic spindle to interact with chromosomes through the intermediate of the megadalton protein complex called the kinetochore (the trilaminar structure observed at the centromere under electron microscopy), that assembles only at centromeres in mitosis (Robbins and Gonatas, 1964; Rieder, 1979). Subsequently, those attached chromosomes will congress to the metaphase plate and get ready to be segregated into daughter cells.

#### 1.2.3.2 Metaphase

For chromosomes to be segregated properly, they have to be attached through a amphitelic-anchoring configuration (bi-orientation) (Maiato, 2004), meaning that duplicated chromosomes have to be attached to microtubules coming from opposite poles of the mitotic spindle (Figure 1.2). Then, the tension between sister kinetochores will be equal, and this will inactivate the second cell cycle control system called the SAC (Spindle Assembly Checkpoint) (Nicklas and Koch, 1969; Nicklas et al., 2001). This checkpoint inactivation involves a myriad of molecular processes that will permit the sister chromatid separation at anaphase.

#### 1.2.3.3 Anaphase

After the SAC is satisfied, a series of molecular processes occur and sister chromatids are separated from one another into distinct daughter cells. First, the

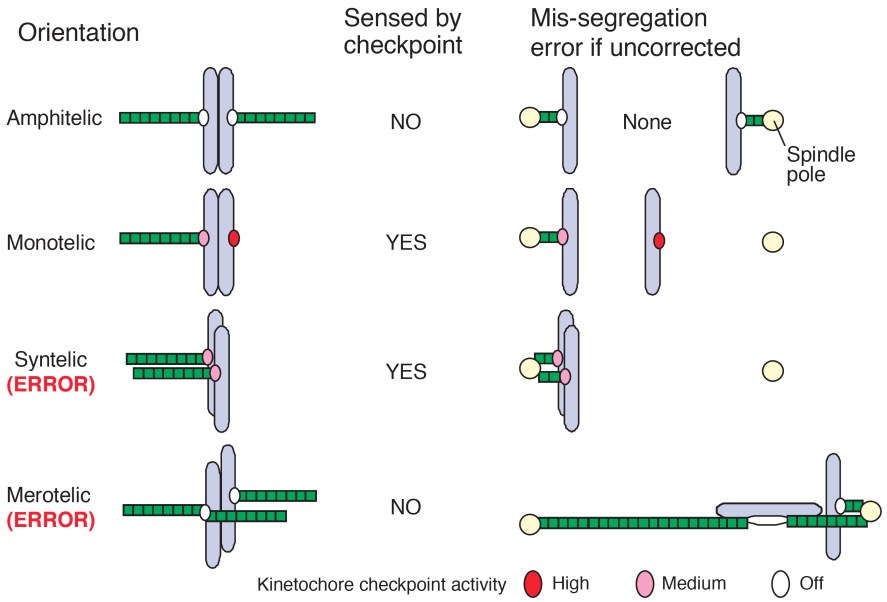



**Figure 1.2 Common Microtubule-Kinetochore attachments.** Amphitelic orientation is the preferred attachment for proper segregation of chromosomes. If not corrected, all other orientations will have mis-segregation of the chromosomes. *Adapted permission from Company of Biologists Ltd: Journal of Cell Science (Maiato et al. 2004), copyright (2004).*

cohesin complexes holding sister chromatids together will be degraded allowing the individual movements of chromatids (Morgan, 2007). Then, the microtubules attached to the kinetochores (kMT) will pull on the chromosomes, bringing them closer to the MTOC (MicroTubule Organization Center) (Cheeseman and Desai, 2008). During this period, the antiparallel microtubules emanating from both MTOCs meeting at the mitotic midzone interact resulting in microtubule based pushing to promote the spindle elongation (Roostalu et al., 2010). Finally, the cytokinetic ring will be assembled at the equatorial cortex between the segregated masses of chromosomes and it will constrict to separate the cytoplasm of the mother cell into two daughter cells.(Glotzer, 2005)

#### 1.2.3.4 Telophase

At Telophase, the cytokinetic ring completes its constriction, the nuclear envelope reforms around the chromosomes, and finally they decondense (Morgan, 2007). The final step of M phase is abscission, which is the physical separation of the daughter cells by the fusion of membranes, resulting in two independent daughter cells (Lafaurie-Janvore et al., 2013). After telophase, the cells will either re-cycle (by re-

entering the cell cycle in G1 phase) or exit the cell cycle and go into quiescence (by entering G0 phase) (Morgan, 2007).

### 1.2.4 G1 phase or Exit of Mitosis

G1 phase is often referred to as the cell cycle gap between M phase and S phase, where the cells prepare for DNA replication. Although this statement is not false, it is not complete, since many molecular mechanisms occur during this 10 hours window of the cell cycle. Early G1 phase, also called exit of mitosis, is the time where the cell propagates and maintains its centromere identity (Jansen et al., 2007; Schuh et al., 2007). Centromeres are the chromosomal loci where the kinetochore assembles in mitosis, and its identity maintenance at these regions is essential for the success of the subsequent cell divisions (Black et al., 2011). In fact, a centromere identity loss will prevent assembly of the kinetochore on chromosomes, causing a mis-segregation of chromosomes (Maddox et al., 2007). This will trigger the appearance of aneuploid cells, which is a hallmark of cancer cells. Regulation of centromere identity and maintenance is the main subject of this thesis and it will be extensively described in the next sections.

### 1.2.5 Consequences of problematic events taking place in the cell cycle

The cell cycle control system, the cell cycle checkpoints, involves many players that are highly regulated. This control system reduces the probability that cells escape those checkpoints and proceed in cell division with aberrant problems. Even if those mechanisms are highly regulated, some cells are still able to bypass the system. One interesting example involves microtubule-kinetochore attachments. As mentioned in section 1.2.3.2, the ideal configuration is the amphitelic attachment, where sister chromatids are attached at opposite spindle pole. However, merotelic attachments, where a single kinetochore is attached to both spindle poles, can occur and they are highly problematic, since the SAC is not able to detect this problem (Cimini et al., 2001). With sister chromatids attached to the opposite poles, and having a single

kinetochore attached at both poles, this attachment will satisfy the SAC. However, it will result in unexpected forces of both poles on a single kinetochore, resulting in either chromosome mis-segregation (having an extra chromosome in one cell) or chromosome breakage. Those situations are deleterious to cells, possibly causing the appearance of aneuploid cells in a mis-segregation situation (Cimini et al., 2001), and broken chromosomes could rearrange by translocation and cause various genetic problems, e.g., creating an oncogenic fusion protein like NUP98-HOXA9, which is thought to play a causative role in myeloid leukemogenesis (Kuwata et al., 2005). Aneuploid cells may be able to propagate over many generations, and some populations can be characterized as CIN (Chromosome Instability Number), having too few or too many chromosomes in their genome. This situation reveals also many problems linked to gene expression, cell growth, and cell division (Compton, 2011).

## 1.3 Chromosomes structure

The human genome is approximately four meters in length and it has to be packaged into chromatin to fit a nucleus of five to eight microns in diameter. This amazing compaction of DNA is performed through the fundamental packaging unit called the nucleosome; composed of canonical histones H2A, H2B, H3, and H4, forming an octamer, around which the DNA is wrapped (Kornberg, 1974; Oudet et al., 1975). When swelled chromatin is observed under the electron microscope, nucleosomes are spaced by ~200bp of DNA, showing a bead-on-the-string pattern (11nm in diameter) (Olins and Olins, 1974; Olins et al., 1976; Woodcock et al., 1976). A higher order structured chromatin is achieved when adjacent nucleosomes interact through their histone tails, forming the 30-nm fiber (Robinson et al., 2006). This structure is stabilized through the binding of the linker histone H1 protein, which neutralizes the negative charge on DNA (Schwarz and Hansen, 1994). A single histone H1 molecule localizes at the entry/exit of a nucleosome, and binds two DNA duplexes through two

independent DNA binding sites (Belikov and Karpov, 1998; Carruthers et al., 1998). Further levels of compaction are required for the formation of a 1,400nm mitotic chromosome; however, the molecular players involved in this process, the possible intermediate structures, and the components dynamics are not well understood, and are still highly debated (Daban, 2000; Woodcock and Dimitrov, 2001).

All along the chromosome, the proteanous composition and the chromatin structures are not uniform, creating distinct chromosomal structures, e.g. the centromere and the telomeres. These differences are crucial for many cellular processes including proper segregation of chromosomes in anaphase. Chromosome segregation fidelity depends on the kinetochore assembly only at the centromere regions and nowhere else. The composition and the structure of centromeres are described in this section.

### 1.3.1 Histone proteins: canonical vs variants

The X-ray crystal structure of a nucleosome core particle shows 146 bp of DNA wrapped in 1.65 left-handed superhelical turns around the histone octamer. Moreover, the canonical histone proteins H2A, H2B, H3 and H4 have a conserved C-terminal histone fold domain and a unique highly basic N-terminal tail (Luger et al., 1997). The histone fold domain is composed of three helices (α1 to α3) connected by 2 loops of 6 to 8 amino acids (L1 and L2). Having no functional role in the nucleosome structure, the unique N-terminal sequence on each histone is prone to post-translational modifications, which is the addition of chemical functional group on Lysine and arginine residues, e.g., Methylation, Acetylation, Sumoylation, and Phosphorylation (Taverna et al., 2007; Shiio and Eisenman, 2011). Those modifications were shown to have a great impact on chromatin structure and genes expression (Verreault, 2000).

The canonical histones, forming the majority of the genome nucleosomes, have fraternal twins called histone variants. Those non-allelic variants differ from the canonical histone by their sequences and their biophysical characteristics. Histone

variants are also components of nucleosomes and they localize to specific regions of the genome, and, in some cases, they are expressed in specific tissues. Their expression and incorporation in the genome are regulated by a replication independent mechanism, which means these processes occur throughout the cell cycle and are not restricted to S-phase. Furthermore, histone variants have additional functions besides genome compaction, by facilitating several cellular processes (Kamakaka and Biggins, 2005).

In the centromere field, a Histone H3 variant is critical for the centromere identity and is called CENP-A (CENtromere Protein A). Serendipitously discovered in 1985, CENP-A is a 17kDa protein localizing to the centromere, and it is essential for the centromere identity, propagation, and for the kinetochore assembly (Valdivia and Brinkley, 1985; Earnshaw and Rothfield, 1985; Earnshaw et al., 1986; Warburton et al., 1997; Heun et al., 2006; Jansen et al., 2007; Schuh et al., 2007). CENP-A is a distinctive histone protein that replaces both copies of canonical histone H3 in the centromeric nucleosome (Palmer et al., 1991; 1987; Yoda et al., 2000; Padeganeh et al., 2013b). It has only 60% similarity to histone H3 sequence, with major differences being at the N-terminus (Sullivan et al., 1994). CENP-A containing nucleosomes also differ from histone H3 nucleosomes at the structural level: CENP-A octamers are more compact consequently leading to a lower height and a wider width, with histone H2A and H2B being further from the dyad axis (Sekulic et al., 2010). Moreover, the CENP-A/CENP-A interface of the nucleosome is rotated relative to the H3/H3 interface, the N-terminal part of L1 loop protrudes giving the opportunity for interactions with non-histone proteins, and hydrophobic interactions at the CENP-A/H4 interface increases the rigidity of CENP-A nucleosome (Sekulic et al., 2010). Therefore, DNA is more compacted and less flexible around CENP-A nucleosomes compare to histone H3 nucleosomes. See section 2.3 for more structural details on CENP-A nucleosome and DNA topology.

### 1.3.2 Centromeres

First described by Flemming as the primary constriction of chromosomes, the centromere is the region of the chromosome where the kinetochore assembles (Figure 1.3) (Flemming, 1882; Robbins and Gonatas, 1964; Brinkley and Stubblefield, 1966; Warburton et al., 1997; Vafa and Sullivan, 1997; Van Hooser et al., 2001). Centromere nomenclature is based on its chromosomal position and it is divided in four categories: metacentric, submetacentric, acrocentric, and telocentric (Figure 1.4). Metacentric chromosomes have their centromere in their middle and it roughly has equal length of chromosome arms. Submetacentric chromosomes have unequal length of chromosome arms, relative to the centromere position. Acrocentric chromosomes have a very short arm, and the other being longer than the

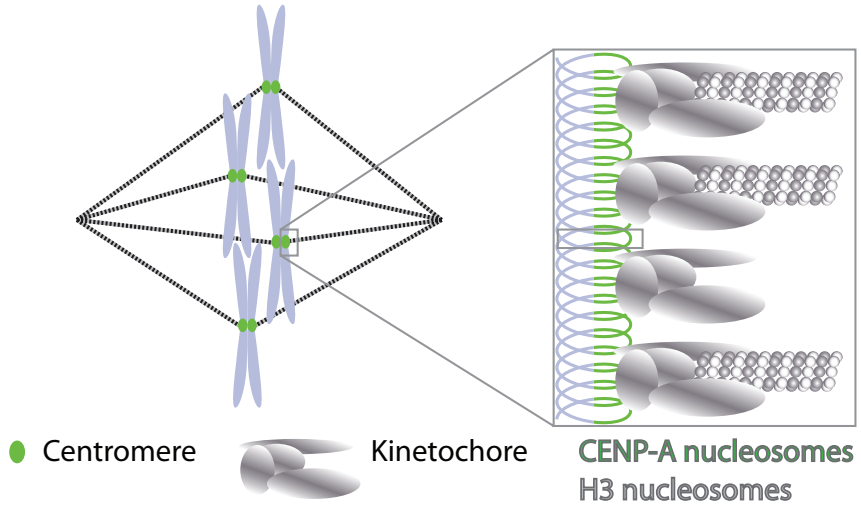



**Figure 1.3 Human mitotic spindle schematic showing a magnification of the centromeric region.**
Human centromeres are monocentric, and are composed of α-satellite DNA and CENP-A. This locus is where the megadalton complex called the kinetochore assembles and this drives the mitotic chromosome attachment to the mitotic spindle (dashed lines). Elongated centromeric DNA shows interspersed localization of CENP-A together with histone H3, thus centromeres are not solely composed of CENP-A nucleosomes (Blower et al., 2002; Dunleavy et al., 2011).

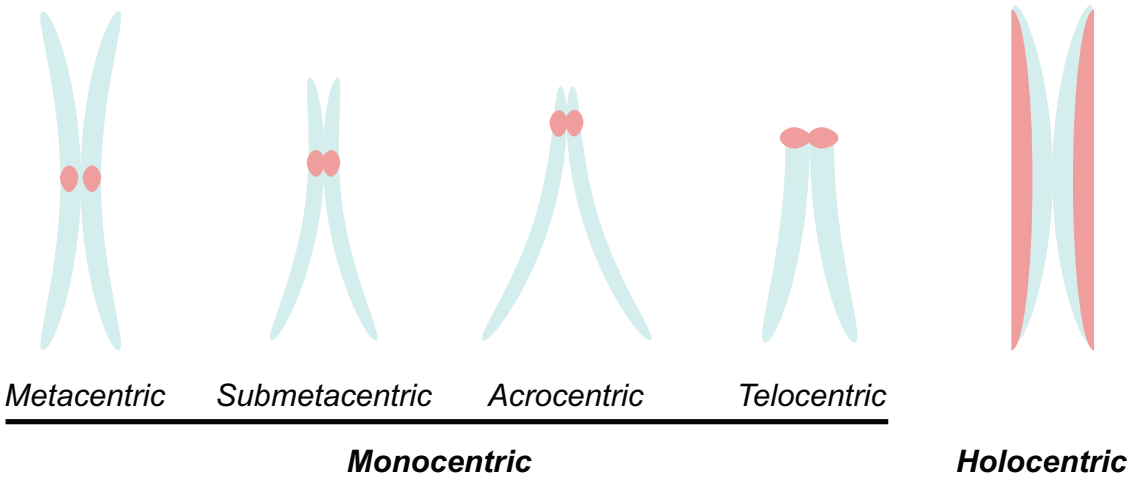



**Figure 1.4 Centromere nomenclatures based on chromosomal position.** Metacentric chromosomes have their centromere in the middle, submetacentric chromosomes have a chromosome arm shorter than the other one by referring to centromere location, acrocentric chromosomes have their centromere just before the end, and telocentric chromosomes have their centromere completely at the end of one chromosome arm. Holocentric chromosomes have their centromere on the entire length of chromosome, and they are present in some species only.

submetacentric centromere, and it is localized just before the end of chromosomes. Telocentric chromosomes have their centromeres at end of the chromosome arms (Levan et al., 1964). Some species have a mixture of these categories within their karyotype, whereas some categories are absent, for example the human karyotype does not have telocentric chromosomes. Furthermore, centromeres can be restricted to a single region of the chromosome, called monocentric chromosome, or they can be located on the entire length of chromosome, called holocentric chromosome (Maddox et al., 2004). Holocentric chromosomes are distinct under the light microscope since they do not have a primary constriction, and this type of chromosome architecture is present in some species, for example *L. nivea* (plant), *S. frugiperda* (insect), *and C. elegans* (nematode). This part of the introduction will mainly focus on human centromeres, and significant information regarding *C. elegans* centromere will be further discussed in chapter 4.

#### 1.3.2.1 α-satellite DNA

Centromeres, a term coined by Waldeyer in 1903, occupy 3 to 5% of the human genome, and they are the only region of chromosomes not to be sequenced; largely because of their repetitive nature at the sequence level (Manuelidis, 1976; 1978; Singer, 1982; Battaglia, 2003). In fact, centromeres are composed of a subclass of repetitive DNA called α-satellite DNA, comprising a 171 bp A (adenine) and T (thymine) rich DNA sequence arranged in a large head to tail array that can reach up to 1 to 5 Mbp in size (Manuelidis, 1976; 1978; Darling et al., 1982; Willard et al., 1983; Mitchell et al., 1985; Willard and Waye, 1987a; Ikeno et al., 1998;). Within a single 171bp monomer, there is a high sequence heterogeneity, meaning that there is no reliable consensus sequence for a single repeat (Vissel and Choo, 1987; Waye and Willard, 1987; Willard and Waye, 1987b). Moreover, there is an extensive size variation between chromosomes ranging from 200kb to 4Mb, and thus a variation in the number of repeats (Willard, 1985; Lo et al., 1999). Even if there are differences at the sequence level for each repeat, there are two motifs within a single α -satellite repeat that are specifically recognized by two proteins: the 17-bp CENP-B box bound by CENP-B (CENtromere Protein B), and a 9bp-*pJα* sequence bound by PARP (Poly (ADP-Ribose) Polymerase) (Muro et al., 1992; Earle et al., 2000). Until now, those two proteins are the only ones identified so far to recognize a specific motif within the human centromeric region.

#### 1.3.2.2 Centromeric sequences are not conserved through species

As previously described, centromere sequences are not constant between chromosomes, suggesting that centromeric DNA might not have a role in the centromere identity. This conclusion is supported by the fact that centromeric sequences are not conserved through species (Figure 2.1) (De Rop et al., 2012). With the exception of *S. cerevisiae* (*Saccharomyces cerevisiae*), mutations within the centromeric sequence of most organisms have no effect on the kinetochore assembly and subsequent chromosomes segregation (Spencer and Hieter, 1992). Therefore, with the exception of *S. cerevisiae*, centromeric DNA of all known species is not sufficient to drive centromere identity through many cell divisions.

#### 1.3.2.3 CENP-A is the epigenetic marker of centromeres

As previously described, CENP-A is a histone H3 variant, replacing histone H3 in centromeric nucleosomes (Padeganeh et al., 2013b). This protein is essential for the kinetochore assembly, and thus for chromosomes segregation in mitosis (Warburton et al., 1997; Heun et al., 2006; Gascoigne et al., 2011). Even if CENP-A mostly localizes to α-satellite DNA, it was discussed in the late 90's that α-satellite DNA, though capable of building functional centromeres when transfected in cultured cells, is not sufficient for centromere identity propagation and sustainable cell division. Indeed, α-satellite DNA is present on both active and inactive centromeres (α-satellite DNA that can not build a functional kinetochore) (Haaf et al., 1992; Sullivan and Schwartz, 1995; Ando et al., 2002;). However, CENP-A is present only on active centromeres (Warburton et al., 1997; Vafa and Sullivan, 1997; Van Hooser et al., 2001). Thus, this subclass of repetitive DNA has no function in centromere identity and CENP-A is the epigenetic mark defining centromeres.

#### 1.3.2.3 De novo centromere formation

There are some situations in nature where de novo centromeres have to be established. One interesting example is the nematode *C. elegans* sperm genome. Although this haploid genome does not contain any HCP-3 protein (HoloCentric chromosome binding Protein-3; CENP-A homolog in worm) on the centromere loci, after fertilization, the chromosomes segregate properly at the first embryonic division having incorporated HCP-3 before mitotic entry (Gassmann et al., 2012). Another example is the proper segregation of plasmids injected in the *C. elegans* gonad. When injected, plasmids do not have any histone proteins bound, and by a not well understood mechanism, CENP-A is incorporated to this genetic material and can be propagated for many cell divisions (Yuen et al., 2011).

Some experimental observations led researchers to propose a role for a protein named CENP-B in de novo centromere formation. CENP-B is a 80kDa protein that was discovered in conjunction with CENP-A and CENP-C (an inner kinetochore protein) (Earnshaw and Rothfield, 1985; Earnshaw et al., 1987; Saitoh et al., 1992).

CENP-B binds directly to the centromeric α-satellite DNA through the 17bp-CENP-B box**,** and this protein is required for heterochromatin formation, as well as translational positioning of the CENP-A octamer onto centromeric DNA (Masumoto et al., 1989; Muro et al., 1992; Yoda et al., 1998; 1992; Ikeno et al., 1994; Ohzeki et al., 2002; Tanaka et al., 2005; Okada et al., 2007; Hasson et al., 2013). It was demonstrated that the CENP-B box together with alphoid DNA are sufficient for de novo centromere formation, and the kinetochore assembly in cultured human cells (Masumoto et al., 1989; Ohzeki et al., 2002). This model integrates a possible role for α-satellite DNA in the centromere identity. However, some experimental observations bring additional information to challenge this attractive model. In fact, CENP-B knockout mice are viable, with cells undergoing mitosis and meiosis with no obvious problems (Hudson et al., 1998). Also, CENP-B is absent from human and mouse Y chromosome as well as neocentromeres (Earnshaw et al., 1987; Voullaire et al., 1993; Sart et al., 1997; Depinet et al., 1997; Choo, 1997; Saffery et al., 2000). And as for α-satellite DNA, CENP-B is present on active and inactive centromeres of dicentric chromosomes, where as CENP-A is only present on active centromeres (region where the functional kinetochore is assembled) (Earnshaw et al., 1989; Sullivan and Schwartz, 1995). Thus, de novo centromere formation and the role of CENP-B in this mechanism are not completely clear.

Recently, it was proposed that CENP-B has a redundant role with CENP-C for kinetochore assembly (Fachinetti et al., 2013). CENP-C is an inner kinetochore protein that binds α-satellite DNA, and is important for kinetochore assembly (Tomkiel et al., 1994; Politi et al., 2002). It was previously shown that CENP-A, CENP-B and CENP-C cooperate for the formation of the kinetochore, but how this cooperation functions is not clear (Ando et al., 2002; Suzuki et al., 2003). By a new approach using CENP-A rescue in a depleted cell, it was demonstrated that different parts of CENP-A bind either CENP-B or CENP-C, and either the N- or the C-terminus of CENP-A alone can recruit kinetochore proteins at the centromere. However, either CENP-B or CENP-C alone are not sufficient for functional centromere formation (Fukagawa et al., 1999; Fachinetti et al., 2013).

Many efforts were performed to better understand the mechanism of de novo centromere formation. However, it is still unclear what are the first molecular steps leading to de novo formation. Particularly, de novo centromere propagation mechanism over many cell divisions is still unidentified.

### 1.3.3 Neocentromeres

Neocentromeres are active centromeres forming away from the original, identified chromosomal locus. This chromatin feature was discovered by a cytologist who noted patients that possessed an abnormal karyotype (Depinet et al., 1997; Sart et al., 1997). Localized in euchromatin, this "new centromeric" region contains CENP-A nucleosomes (Sart et al., 1997). Although there is no α-satellite sequence, neocentromeres are able to build a functional kinetochore and to propagate the CENP-A mark through many cell divisions, consequently having a normal mitosis and meiosis (Voullaire et al., 1993; Sart et al., 1997; Depinet et al., 1997; Choo, 1997). This observation is the major argument in favour of the hypothesis that epigenetic inheritance of centromeres requires the CENP-A mark.

## 1.4 Regulation of centromere identity and propagation

It is well accepted in the literature that CENP-A is the epigenetic mark of the centromeres. This mark has to be propagated at each cell division in order to maintain centromere identity. If not, CENP-A will be diluted at each cell division (CENP-A equally distributed to each daughter strand (Mellone et al., 2011)), until the cell reaches a threshold where it will not be able to build a functional kinetochore. Then, chromosomes will not segregate properly, which will lead to cell death. Moreover, if too much CENP-A gets incorporated to the centromeres, or if it is mislocalized elsewhere than the centromeric locus, this extra CENP-A can cause serious kinetochore-microtubule attachment problems, e.g., merotelic attachment. Again, problems in chromosome segregation will occur, though to a lesser extent than CENP-A absence at the centromeres. Mis-segregation will lead to aneuploidy, a hallmark of cancer cells. Thus, tight regulation of CENP-A incorporation at the centromeres is necessary to avoid subsequent cellular problems.

Expressed in G2 and incorporated in G1, CENP-A incorporation is performed through a three-step mechanism: (1.4.1) Licensing, (1.4.2) Loading, and (1.4.3) Maintenance (Figure 2.3) (Shelby et al., 1997; Jansen et al., 2007; Schuh et al., 2007; Lagana et al., 2010).

### 1.4.1 Licensing

Important cellular processes are in control of restricting the centromere location and size. If those cellular processes are disturbed, CENP-A might be incorporated anywhere in the genome, consequently having major negative impacts in cells. Therefore, the incorporation of another level of control of the newly synthesized CENP-A, e.g. through a licensing mechanism, is crucial for cell survival.

#### 1.4.1.1 Mis18α and Mis18β

The first proteins identified as essential players for CENP-A localization to the centromere were a subset of the Mis proteins discovered in a genetic screen for chromosome mis-segregation in *S. pombe* (*Schizosaccharomyces pombe*): Mis6, Mis12, and Mis14-18. Depletion of those proteins leads to a Mis-segregation of chromosomes phenotype (Saitoh et al., 1997; Goshima et al., 1999; Hayashi et al., 2004). The CENP-A homolog Cnp1 (CeNtromere Protein 1) incorporation and localization highly depends on Mis18, and this protein was the most upstream factor identified for Cnp1 localization to the centromeres and for the kinetochore assembly (Takahashi et al., 2000; Hayashi et al., 2004). The Mis18 ortholog was later discovered as two closely related human genes and they were named Mis18α and Mis18β (Fujita et al., 2007). These proteins form a complex, which only localizes to the centromeres in telophase/early G1, and which is required for CENP-A localization to the centromeres. However, the exact role and their functions are not well understood. It was proposed that this complex regulates the epigenetic code of centromeric chromatin either by histone acetylation or methylation (Fujita et al., 2007; Kim et al., 2012). However, the implications and the timing of these modifications are also not well understood. Additionally, the Mis18 complex is regulated by the proteasome degradation through the action of Ubiquitin E3 ligase βTrCP, and Mis18β degradation disrupts the Mis18 complex (Kim et al., 2013). Once more, more investigations have to be performed to better understand the consequences of this degradation in the CENP-A incorporation pathway.

#### 1.4.1.2 KNL-2/M18BP1

To find other essential proteins required for CENP-A localization to the centromere, different genetic screens and mass spectrometry studies were performed (Obuse et al., 2004; Izuta et al., 2006; Maddox et al., 2007; Dunleavy et al., 2009; Foltz et al., 2009; Lagana et al., 2010). In *C. elegans*, depletion of all genes involved in chromosome segregation (previously identified in an embryonic lethal screen) was performed by RNAi (RNA interference), and Kinetochore NulL (KNL) phenotypes were scored by a fluorescence microscopy-based assay (Maddox et al., 2007). This

KNL phenotype is characterized by a clustering of chromosomes from each pronucleus (forming two DNA balls), premature spindle pole separation, defective chromosome alignment, and failed chromosomes segregation, since the kinetochore is not able to assemble and properly segregate chromosomes (Oegema and Hyman, 2006; Maddox et al., 2007). KNL-2 (Kinetochore NulL 2) was identified as an essential protein for HCP-3 (CENP-A) localization to the centromere. When KNL-2 is depleted, HCP-3 no longer localize to the centromere in a manner that is independent of HCP-3 protein level. The same phenotype is observed when the human ortholog HsKNL-2 is depleted in HeLa cells. Therefore, KNL-2 is essential for CENP-A localization to the centromere by stabilizing its binding to the centromeres (Maddox et al., 2007).

At the same time, another group identified the human ortholog of KNL-2 and named it M18BP1 (Mis18 Binding Protein 1) (Fujita et al., 2007).They showed that this protein localizes to centromeres and is required for CENP-A, Mis18α, and Mis18β localization to the centromeres (by directly binding to Mis18 α). Together, those findings suggest that KNL-2/M18BP1 is one of the most upstream components for CENP-A localization to the centromeres and forms a complex with Mis18α, and Mis18β. How this complex directly recognizes the centromeric chromatin in order to perform its licensing function is still not understood. Some evidence shows that KNL-2 is not able to bind soluble and mononucleosomal particles of CENP-A (Foltz et al., 2006; Maddox et al., 2007; Carroll et al., 2009), and other evidence suggests a role for CENP-C in the recruitment of the Mis18 complex to the centromeres (Moree et al., 2011; Dambacher et al., 2012).

#### 1.4.1.3 CENP-C

CENP-C is an inner kinetochore protein that can bind alphoid DNA containing CENP-A (Saitoh et al., 1992; Politi et al., 2002; Ando et al., 2002; Carroll et al., 2010; Erhardt et al., 2008). It is required for building a functional kinetochore as well as maintaining a proper kinetochore size (Tomkiel et al., 1994; Ando et al., 2002; Oegema et al., 2001). Since CENP-C directly binds the CENP-A nucleosome, it was suggested that

CENP-C is required for KNL-2 recruitment to the centromere (Dambacher et al., 2012; Moree et al., 2011). Two orthologs of KNL-2 were identified in *X. laevis* (*Xenopus laevis*), and the authors observed that depletion of CENP-C drastically reduces KNL-2 ortholog 1 localization to CENP-A chromatin, whereas the ortholog 2 localization does not change. However, a closer look at the results shows a significant amount of KNL-2 remaining at the centromeres after an immunodepletion of CENP-C in *X. laevis* egg extract. Also, it was previously shown in *C. elegans* that HCP-4 (HoloCentric chromosome binding Protein 4, CENP-C homolog in worms) depletion does not affect KNL-2 localization to the centromere (Maddox et al., 2007). Thus, the exact role of CENP-C in centromere licensing is not clear and needs more investigation.

### 1.4.2 Loading

Soluble histones are not incorporated into chromatin by a random diffusion event. They are negatively charged just like DNA, and they would repulse each other and therefore have less binding affinity. To mask histone charges and to actively incorporate it in chromatin, histone has to bind a chaperone, and each histone protein/complex has its designated chaperone.

#### 1.4.2.1 HJURP

To protect newly synthesized histone proteins from degradation and to actively incorporate them into chromatin, they must be bound by chaperones. HJURP (Holliday Junction Recognition Protein) was identified as the CENP-A chaperone, and this protein binds the CENP-A/H4 tetramer by recognizing the CATD (CENP-A Targeting Domain) on CENP-A and the CENP-A/H4 dimer interface through its N-terminal domain (Black et al., 2007; Foltz et al., 2009; Dunleavy et al., 2009; Shuaib et al., 2010; Barnhart et al., 2011; Hu et al., 2011; Bassett et al., 2012; Fachinetti et al., 2013). HJURP has a nucleosome assembly activity and incorporates newly synthesized CENP-A at the centromeres in a KNL-2 dependent manner (Mizuguchi et al., 2007; Camahort et al., 2007; Shivaraju et al., 2011; Dechassa et al., 2011;

Barnhart et al., 2011). HJURP orthologs were identified in many species, all having the same role in protecting CENP-A from degradation, and actively incorporate it in the centromere. The HJURP ortholog in *S. cerevisiae* is Scm3 (Suppressor of Chromosome Mis-segregation 3), and it binds the CENP-A/H4 tetramer as a dimer (Stoler et al., 2007; Zhou et al., 2011). This CENP-A/H4/HJURP complex forms a hexasome complex and it does not contain any histone H2A and H2B (Padeganeh et al., 2013a). Recently, the *D. melanogaster* HJURP ortholog was identified as CAL1 (Chromosome ALignment 1)(Chen et al., 2014). This protein was previously identified to be required for newly synthesized CID (Centromere IDentifier, CENP-A homolog in fly) localization and incorporation to the centromere together with CENP-C in metaphase (Erhardt et al., 2008; Mellone et al., 2011). HJURP is also conserved in the tetraploid organism *X. laevis*, and was shown to require the presence of condensin II for CENP-A incorporation to the centromere (Bernad et al., 2011). Although the HJURP orthologs have similar function, the primary sequence is highly divergent between species. Thereby, HJURP absence from some eukaryotes such as fish, nematodes (*C. elegans*) and plants can be explained by a sequence assembly limitation (Sanchez-Pulido et al., 2009). Also, some evidences suggest the lost of this protein over evolution, giving the possibility of CENP-A nucleosome assembly activity to another protein involved in the CENP-A incorporation pathway (Chen et al., 2014). Thus, HJURP is essential for CENP-A loading at the centromeres and it is conserved in many species.

#### 1.4.2.2 RbAp46/48

As described in section 1.4.1.1, Mis-segregation proteins (Mis6, Mis12, and Mis14-18) were the first identified players for the localization of CENP-A to the centromere. One of them being Mis16 is required for Cnp1 localization to the centromere, and the human orthologs are RbAp46/48 (Hayashi et al., 2004). Those proteins are part of the CAF-1 complex, and they were shown to be required for CENP-A recruitment at the centromere. Since those proteins are part of the CAF-1 complex, a complex known to incorporate the H3/H4 tetramer in S phase, they are not specific to an association

with the HJURP chaperone. Thus, not much has been discovered so far about the exact function of those proteins in the CENP-A incorporation pathway.

#### 1.4.2.3 The missing link between Licensing and Loading: Mis18 proteins

The sequential events and the dynamic of the different players involved in CENP-A incorporation and localization to the centromere is a hot topic in the centromere field. At the American Society for Cell Biology Annual meeting in 2013, the group of Daniel Foltz presented its latest work and they demonstrated that Mis18α and Mis18β form a dimer and they directly bind to HJURP (Nardi et al., 2013). Through this interaction, those proteins bring the new CENP-A/H4 tetramer in close proximity to the centromeric chromatin. This observation is an important piece of data showing a direct link between the licensing and the loading steps.

#### 1.4.2.4 Stable CENP-A incorporation by FACT and RSF1 complex

Identified in independent IP (immunoprecipitation) experiments, FACT (FAcilitates Chromatin Transcription) and RSF1 (Remodelling and Spacing Factor 1) were shown to interact directly with CENP-A nucleosomes (Obuse et al., 2004; Foltz et al., 2006; Izuta et al., 2006). FACT is a known nucleosome destabilizing protein that promotes transcription activity, and RSF1 an ATP-dependent nucleosome remodelling complex, which favours transcription initiation *in vitro* (LeRoy et al., 1998; Belotserkovskaya et al., 2003). Relative to the centromere, FACT contributes to the heterochromatin integrity, which supports the formation of functional centromeres (Lejeune et al., 2007). The RSF1 complex, composed of Rsf1 and SNF2h, localizes to the centromere in mid-G1 and is required for CENP-A nucleosome assembly at the centromeres (Perpelescu et al., 2009). It was suggested that FACT and RSF1 are involved in the stabilization of CENP-A incorporation to the chromatin (Bernad et al., 2009). However, this has not been clearly established.

### 1.4.3 Maintenance

In order to maintain centromere size throughout cell divisions, cells have to regulate the amount of CENP-A that has to be incorporated into the centromere. If there is an excess of CENP-A incorporation, it has to be removed and degraded in order to maintain genome stability. Then, the newly incorporated CENP-A has to become a mature epigenetic mark of the centromere, changing its "new" CENP-A mark to an "old" one for the next cell division.

#### 1.4.3.1 MgcRacGAP

*The role of this protein for CENP-A maintenance at the centromeres is the subject of a publication in Nature Cell Biology, see chapter 3.*

MgcRacGAP has a well-known function in cytokinesis as being part of the centralspindlin complex, together with the kinesin-6 family member MLKP1 (Jantsch-Plunger et al., 2000). This complex is required for the central spindle assembly, an array of antiparallel microtubules between segregated chromosomes at anaphase that recruits and activates key regulators for the contractile ring assembly and constriction, e.g., the small GTPase Rho family proteins RhoA, Rac1, and Cdc42 (Hirose et al., 2000; Etienne-Manneville and Hall, 2002; Glotzer, 2005; Canman et al., 2008). Those small GTPases cycle between two states: the active GTP bound state and the inactive GDP bound state, and they hydrolyse the GTP to GDP for various cell functions through effector proteins. The switch between the inactive form and the active forms is achieved by a protein having a Guanine Exchange Factor (GEF) activity, and the switch between the active form and the inactive form is performed by a GTPase Activating Protein (GAP); for the Rho GTPases family, ECT2 (Epithelial Cell Transforming sequence 2) is the GEF and MgcRacGAP is the GAP (Toure et al., 1998; Yüce et al., 2005; Glotzer, 2005).

MgcRacGAP was previously shown to interact with CENP-A (Izuta et al., 2006; Perpelescu et al., 2009). However, this observation only caught the field's attention in 2010, when MgcRacGAP was shown to have an important function in the centromere

identity maintenance (Lagana et al., 2010). After depletion of this protein, only old CENP-A is detected at the centromere by an advanced pulse-chase ImmunoFluorescence (IF) technique called SNAP-Tag. MgcRacGAP localizes to the centromeres in late G1, whereas KNL-2 localizes to the centromere in telophase/early G1 (Jansen et al., 2007). With these observations, it was proposed that MgcRacGAP has a role in the maintenance of centromere identity by changing a "mark" of the newly synthesized CENP-A for an old one, thereby stabilizing the CENP-A nucleosome, and maintaining the epigenetic mark of the centromere. However, this mark has not been identified so far.

## 1.5 KNL-2 regulation and function

*The role of this protein for newly synthesized CENP-A incorporation at the centromeres is the subject of a publication in revision in Developmental Cell, see chapter 4.*

As described previously, KNL-2 is a major player in the centromere identity. Through sequence analysis, KNL-2 has a predicted Myb like DNA binding domain located in the C-terminus of the protein (Maddox et al., 2007). This type of domain is known to bind DNA, therefore how this Myb domain contributes to the centromeric chromatin recognition by KNL-2 is an interesting avenue to better understand the centromere identity and maintenance.

### 1.5.1 Predicted domains

KNL-2 protein was identified in *C. elegans* and humans, however the two proteins are highly divergent at the sequence level (Maddox et al., 2007). A predicted Myb domain was identified in the C-terminus of the worm protein, and it is through alignment of this sequence with closely related nematode species that the human KNL-2 protein

was identified (the Myb domain is also present in the C-terminus of human KNL-2). Also, a SANTA (SANT Associated) domain was identified in the N-terminus of both the *C. elegans* and human proteins. These two predicted domains are known to bind chromatin and are of high interest in order to better understand the molecular function of KNL-2.

#### 1.5.1.1 Myb domain

Named after the discovery of this domain in the proto-oncogene c-Myb, a known transcriptional regulator of the hematopoietic system, the Myb binding domain is a repeat of three α helices arranged in a homeo-domain like helix-turn-helix motif (Westin et al., 1982; Otting et al., 1988; Ogata et al., 1992). Each helix contains a conserved aromatic residue forming a hydrophobic core important for the tertiary structure of the domain (Anton and Frampton, 1988; Kanei-Ishii et al., 1990; Saikumar et al., 1990; Ogata et al., 1992). Usually, Myb domains bind the major groove of DNA through residues located in proximity of the third helix, and it has an overall basic isoelectric point (pI) and a positive electrostatic surface (Peters et al., 1987; Biedenkapp et al., 1988; Saikumar et al., 1990; Ogata et al., 1994). Myb DNA binding domains are different from other helix-turn-helix DNA binding motifs since they are not found in proteins having enzymatic activity, e.g. chromatin remodelling activity (Boyer et al., 2004). For the KNL-2 protein, the predicted Myb domain is localized in the C-terminus whereas most Myb domains in other well-known proteins are all localized in the N-terminus. This might be an important feature of KNL-2 for its specific function in the centromere recognition.

The Myb domain is often confused with the SANT domain, mostly because the latter is also called Myb-like domain. Even if their structures are similar (the SANT domain also has three α helices and a hydrophobic core), their binding properties to chromatin are different. The SANT (Swi3, Ada2, N-CoR, TFIIIB) domain is a motif found in many chromatin remodelling enzymes where it binds to unacetylated histone tails and mediates protein-protein interactions (Aasland et al., 1996). This domain

does not bind directly to DNA, caused by a tryptophan hindrance, and it has an acidic pI, and a negative electrostatic surface (Boyer et al., 2004). With those different biochemical characteristics, SANT and Myb domains clearly have different functions in the cell nucleus.

#### 1.5.1.2 SANTA domain

The SANTA (SANT Associated) domain is another predicted domain within KNL-2. This protein motif was identified in many proteins already having a SANT or a Myb domain; therefore it was named SANT Associated domain (Zhang et al., 2006). This domain shows clear biochemical properties that are different than its related SANT domain. The SANTA domain is composed of four central β sheets flanked by three α helices, and many highly conserved tryptophan residues predicted to have a protein-protein interaction function. The SANTA domain is predicted to be present in many proteins having chromatin remodelling activities. However, its function has not been demonstrated.

### 1.5.2 KNL-2 conservation in metazoans

KNL-2 was identified in *C. elegans* and humans, and since the two orthologs are highly divergent at the sequence level, it has been challenging to find orthologs in other model organisms. Recently, the KNL-2 orthologs in *X. laevis*, named M18BP1, and in *A. thaliana,* named KNL2, were identified (Moree et al., 2011; Lermontova et al., 2013). Those proteins were shown to be upstream in the CENP-A centromere incorporation pathway, and are essential for CENP-A localization to the centromere (Moree et al., 2011; Lermontova et al., 2013). More interestingly, the C-terminus of *A. thaliana* KNL2, containing no predicted DNA binding domain, was shown to be sufficient for KNL2 localization to the centromeres (Lermontova et al., 2013). This observation is of interest since the predicted Myb domains in *C. elegans* and human are also located in the C-terminus of KNL-2 proteins. However, in our hands, sequence analysis of the C-terminal sequence of *A. thaliana* KNL2 did not predict any

Myb domain. It will be interesting to investigate whether this C-terminal sequence is able to bind any centromeric chromatin *in vitro*.

### 1.5.3 Cell cycle regulation of KNL-2

KNL-2 is essential for CENP-A localization to the centromeres and it is regulated by the cell cycle master regulator CDK (Cyclin-Dependent Kinase) (Silva et al., 2012). G2 synchronized HeLa cells treated with Roscovitine, a CDK1 and 2 inhibitor (Meijer et al. 1997), show newly synthesized CENP-A incorporation as well as KNL-2 localization to the centromeres. This observation suggests that CDK1 and 2 are the cell cycle regulators required to restrain KNL-2 localization to the centromeres and for CENP-A incorporation timing to be only in G1. It also suggests that KNL-2 is present throughout the cell cycle without being localized at the centromeres. This observation brings plenty of new biological questions regarding which phosphatases are regulating KNL-2 localization, and thereby CENP-A incorporation, where KNL-2 is sequestered outside of G1, and how it is regulated.

## 1.6 RNA function in CENP-A localization to centromeres

In the last decades, strong evidences show RNA as a component required for proper centromere function. First observed in the 70's, RNA localizes to the kinetochore and it is essential for its integrity (Rieder, 1979). This observation was interesting and led to many hypotheses on the actual role of RNA in the centromere biology.

### 1.6.1 Transcription, RNAi pathway, and heterochromatin formation for CENP-A incorporation to the centromeres

RNA function within the kinetochore has been extensively studied in *S. pombe*, and many observations have shown a role for transcription, the RNAi pathway and heterochromatin formation in Cnp1 (CENP-A homolog in *S. pombe*) localization to the centromere. First, heterochromatin in the pericentromeric region (region around the centromere locus) is essential for Cnp1 localization to the centromere. This region has high H3K9 methylation, a known heterochromatin marker, and this epigenetic mark is recognized by Swi6 (SWItching-defective protein 6, HP1 (Heterochromatin Protein 1) homolog in *S. pombe*) (Lachner et al., 2001; Bannister et al., 2001; Folco et al., 2008; Kagansky et al., 2009). Furthermore, the Clr4 (Cryptic Loci Regulator, Suv39 (SUppressor of Variegation 3-9) homolog in *S. pombe*) methyltransferase responsible for H3K9 methylation is also required for Cnp1 localization to the centromere (Folco et al., 2008). This enzyme is recruited to the centromere through the RNAi pathway by directly binding to Stc1 (Stanniocalcin 1), which localizes to the centromere through an association with Ago1 (Argonaute 1), a known RNAi component. Moreover, the ribonuclease Dicer, which cleaves double-stranded RNA to siRNA (small interfering RNA), and Chp1 (Chromodomain protein 1), part of the RITS complex (RNA-induced initiation of transcriptional gene silencing), are required for Cnp1 localization to the centromere (Volpe et al., 2002; Folco et al., 2008). Therefore, the RNAi pathway together with Swi6 are involved in the heterochromatin

formation at the pericentromere, and this cellular process is essential for Cnp1 localization to the centromere (Lejeune et al., 2011; Gent and Dawe, 2012).

Transcription is involved in the RNAi pathway and in the heterochromatin formation; therefore, it is also involved in Cnp1 localization to the centromere. RNAPII (RNA polymerase II) localizes to the centromere and transcribes ncRNA (noncoding-RNA) involved in the RNAi pathway (Chan et al., 2011; Choi et al., 2011). Double stranded ncRNA is processed into siRNA by Dicer, and those siRNAs are bound by Ago1, part of the effector complex RITS, that will hybridize the single stranded RNA of nascent mRNA, triggering its degradation (Lejeune et al., 2011). Then, the transcribed centromeric core transcript will be bound by the siRNA-Ago1 complex, which will recruit the Stc1 protein and the methyltransferase Clr4, which methylates H3K9 and forms heterochromatin that is required for Cnp1 localization to the centromere. Finally, when CENP-A is depleted, there is an increase of centromeric transcripts in *S. pombe* (Choi et al., 2011). Knowing that centromeres only have half of CENP-A at anaphase onset in human cells, it suggests that centromeric RNA might be transcribed and involved in newly synthesized CENP-A incorporation in G1.

### 1.6.2 RNA function in CENP-A incorporation to centromere in G1

As described in section 1.4.1.3, CENP-C has a role in CENP-A incorporation to the centromeres. CENP-C has a DNA binding domain, which is sufficient to bind DNA *in vitro*. Moreover, through the same domain, CENP-C is able to bind RNA. It was suggested that CENP-C binding to DNA is stabilized by RNA binding (Wong et al., 2007; Du et al., 2010). It might be possible that other components involved in CENP-A incorporation to the centromere may be regulated through RNA binding.

### 1.6.3 Centromere clustering at nucleoli

In interphase, CENP-A localizes to the nucleolus, which is a nuclear body where the

rDNA (ribosomal DNA) is transcribed and the ribosomes are assembled (Ochs and Press, 1992). CENP-A is anchored to the nucleolus through interactions with NLP (Nucleophosmin-Like Protein), and CTCF (Insulator protein CCCTC-binding factor) proteins (Padeken et al., 2013). NLP is a Nucleophosmin-like protein in Drosophila, and the human NMP1 (Nucleophosmin 1) localizes to the nucleolus, enhances rDNA transcription, and interacts directly with CENP-A pre-nucleosomal particle (CENP-A/H4 tetramer) (Murano et al., 2008; Foltz et al., 2009). NPM-1 interacts with CTCF, which is a chromatin insulator that mediates DNA interactions regulating gene expression (Yusufzai et al., 2004). These two proteins anchor CENP-A chromatin to the nucleolus by interacting with Modulo (Nucleolin homolog in Drosophila), and this protein is important for the ribosome synthesis and maturation (Lee et al., 1992). CENP-A tethering to the nucleolus seems to be linked to RNA biosynthesis; however, the reason why centromeric chromatin is localized to this nuclear body is not known.

Other centromeric components also localize to the nucleolus. CENP-C localizes through a nucleolus targeting signal, and this localization is sensitive to RNAse digestion in human cells (Pluta and Earnshaw, 1996; Wong et al., 2007). Also, CENP-W, identified as Cug2 (Cancer-Upregulated Gene 2), is part of the CCAN complex (Constitutive Centromere-Associated Network), and forms a complex with CENP-T, CENP-S, and CENP-X (having a histone-like fold domain). This protein also localizes to the nucleolus and interacts with NPM1 (Hori et al., 2008; Kim et al., 2009; Chun et al., 2011; Nishino et al., 2012). These observations are interesting and promising for a better understanding of centromere regulation in interphase.

## 1.7 Experimental techniques used for centromere studies

Most experimental techniques used to study centromere biology rely on fluorescence microscopy, mass spectrometry, and some biochemistry. Since most of the components required for the centromere identity propagation, the kinetochore assembly, and the microtubule binding to mitotic chromosomes have been identified, the centromere field is trying to better understand the cellular dynamics, and the regulation of those components *in vivo*. Therefore, microscopy based experiments are ideal to answer those questions. However, some components are essential for cell viability, and depletion of those components leads to rapid and drastic lethal phenotypes, like KNL-2 depletion. Biochemical approaches and biochemistry coupled to microscopy using the TIRF (Total Internal Reflection Fluorescence) microscope have been successful tools to understand the KNL-2 binding specificity.

### 1.7.1 Cell lines

Widely used in the centromere biology field, HeLa cells were obtained from a female patient (Henrietta Lacks) with a cervical cancer. These cells were immortalized and became the most widely used cell line in laboratories (Shah et al., 2014). Using these cells is advantageous since they propagate under minimal conditions, and they are relatively easy to manipulate. However, they are chromosomally instable, or CIN, meaning that at each cell division chromosomes are gained or lost, and they can make chromosomal rearrangements. Consequently, it creates a heterogeneous population with different numbers of chromosomes that can increase expression of different components important for the centromere identity propagation. This information must at all time be taken into consideration when results are generated using this cell line.

### 1.7.2 Live cell imaging and advanced microscopy techniques

By using specific antibodies recognizing proteins of interest, immunofluorescence is useful to establish the localization of a protein in the cell. Additionally, by having a cell cycle marker or by looking at DNA morphology, it is possible to know in which cell cycle phase our protein localizes to a specific locus. However, this microscopy technique does not give any dynamic parameters. Live cell imaging coupled to fluorescence microscopy resolves this fixed cell restriction. By looking at fluorescence over time, residency time can be established as well as quantities of protein by measuring the fluorescence intensity. Moreover, FRAP (Fluorescence Recovery After Photobleaching) is used to measure the rate of diffusion of a protein in cells. Also, by looking at the recovery rate, this information can be interpreted as tight (low recovery rate) or weak binding (high recovery rate) of a protein to a cell locus (Sprague et al., 2004). Live cell imaging is useful for some biochemical reactions within cells; however, the specific binding partners and the affinity constant cannot be obtained with this microscopic technique. With TIRF microscopy, it is possible to obtain association and dissociation rates between two purified cellular components fused to a fluorophore by measuring the fluorescence intensity fluctuations. A cellular component can be trapped on a coverslip using a nanobody binding to GFP, and its binding with another cellular component fused to RFP can be observed and quantified: this technique is called SIMPull assay (Jain et al., 2011). Background elimination due to evanescence illumination will allow the high-resolution observation of the trapped proteins by fluorescence microscopy, and permit calculation of an affinity constant.

Another microscopic technique was developed to study CENP-A dynamics. The SNAP-tag is a modified quench-pulse technique that enables to distinguish the newly synthesized CENP-A from the old chromatin-incorporated CENP-A (Jansen et al., 2007). CENP-A is coupled to a SNAP tag (a modified variant of the suicide enzyme $O^6$-alkylguanine-DNA alkyltransferase), and it can be bound by two molecules: TMR (TetraMethylRhodamine), a fluorescent tag, and BG (Benzylguanine), a quench

binding molecule. Also, the SNAP contains an HA (Hemagglutinin) epitope, and an antibody can recognize this tag. The technique consists of synchronizing SNAP-tag CENP-A cells in S-phase using double thymidine assay and treating with TMR (already incorporated CENP-A-SNAP will be fluorescent). Then cells are released in G2 and are treated with BG, quenching all newly synthesized CENP-A with the fluorophore. Finally, cells are fixed 24 hours after thymidine block and IF is performed to label total CENP-A in the cell. If a problem occurred in the incorporation process, only the old CENP-A will be present at the centromere, therefore a total intensity signal diminution will be observed. This technique is extremely useful for CENP-A incorporation studies.

### 1.7.3 Molecular biology and biochemistry

Many molecular biology and biochemical techniques can be used to answer a variety of questions. Mainly, qPCR (quantitative Polymerase Chain Reaction) and western blot are used to determine gene expression, and protein stability within a cell, or even to know if a depletion using shRNA (short hairpin RNA) is total or partial. Protein expression and purification are a must in biochemistry, mostly to determine its 3D structure through in solution NMR, to calculate the protein binding affinity to another protein or DNA, to see if the protein is a monomer or not, etc. This biochemistry can be coupled to the TIRF microscope to determine the binding preference of a protein, e.g., KNL-2 Myb domain. This will be later discussed in chapter 4.

## 1.8 Outstanding questions and research objectives

Even if the major players involved in the centromere identity propagation and maintenance have been identified, the functions and the regulation of those proteins are still not clear. KNL-2 was identified as an essential protein required for CENP-A localization to the centromere and for its maintenance to this chromosomal locus. However, by which mechanisms this protein performs those tasks remains unclear. Also, KNL-2 has a predicted Myb DNA binding domain, but how this protein binds and recognizes centromeric chromatin via this domain is still under investigation. Through the next chapters, KNL-2 functions in CENP-A incorporation to the centromere will be described and discussed.

## 1.9 References


Aasland, R., Stewart, A.F., and Gibson, T. (1996). The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIB. Trends Biochem Sci *21*, 87–88.

Ando, S., Yang, H., Nozaki, N., Okazaki, T., and Yoda, K. (2002). CENP-A, -B, and -C Chromatin Complex That Contains the I-Type α-Satellite Array Constitutes the Prekinetochore in HeLa Cells. Mol Cell Biol *22*, 2229–2241.

Anton, I.A., and Frampton, J. (1988). Tryptophans in myb proteins. Nature *336*, 719–719.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229–243.

Bassett, E.A., DeNizio, J., Barnhart-Dailey, M.C., Panchenko, T., Sekulic, N., Rogers, D.J., Foltz, D.R., and Black, B.E. (2012). HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. Dev Cell *22*, 749–762.

Battaglia, E. (2003). Centromere, kinetochore, kinochore, kinetosome, kinosome, kinetomere, kinomere, kinetocentre, kinocentre: history, etymology and intepretation. Caryologia *56*, 1–21.

Belikov, S., and Karpov, V. (1998). Localization of Histone H1 Binding Sites Within the Nucleosome by UV-induced H1-DNA Crosslinking *In Vivo*. J Biomol Struct Dyn *16*, 35–39.

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT Facilitates Transcription-Dependent Nucleosome Alteration. Science *301*, 1090–1093.

Bernad, R., Sanchez, P., and Losada, A. (2009). Epigenetic specification of centromeres by CENP-A. Exp Cell Res *315*, 3233–3241.

Bernad, R., Sanchez, P., Rivera, T., Rodriguez-Corsino, M., Boyarchuk, E., Vassias, I., Ray-Gallet, D., Arnaoutov, A., Dasso, M., Almouzni, G., et al. (2011). Xenopus HJURP and condensin II are required for CENP-A assembly. J Cell Biol *192*, 569–582.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E., and Klempnauer, K.H. (1988). Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature *335*, 835–837.

Black, B.E., Jansen, L.E., Foltz, D.R., and Cleveland, D.W. (2011). Centromere Identity, Function, and Epigenetic Propagation across Cell Divisions. Cold Spring Harb Symp Quant Biol *75*, 403–418.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell *25*, 309–322.

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

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

Brinkley, B.R., and Stubblefield, E. (1966). The fine structure of the kinetochore of a mammalian cell *in vitro*. Chromosoma *19*, 28–43.

Camahort, R., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., and Gerton, J.L. (2007). Scm3 is essential to recruit the histone h3 variant cse4 to centromeres and to maintain a functional kinetochore. Mol Cell *26*, 853–865.

Canman, J.C., Lewellyn, L., Laband, K., Smerdon, S.J., Desai, A., Bowerman, B., and Oegema, K. (2008). Inhibition of Rac by the GAP activity of centralspindlin is essential for cytokinesis. Science *322*, 1543–1546.

Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. J Cell Biol *189*, 1143–1155.

Carroll, C.W., Silva, M.C., Godek, K.M., Jansen, L.E., and Straight, A.F. (2009). Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. Nat. Cell Biol. *11*, 896–902.

Carruthers, L.M., Bednar, J., Woodcock, C.L., and Hansen, J.C. (1998). Linker Histones Stabilize the Intrinsic Salt-Dependent Folding of Nucleosomal Arrays: Mechanistic Ramifications for Higher-Order Chromatin Folding. Biochemistry *37*, 14776–14787.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H., and Wong, L.H. (2011). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore–microtubule interface. Nat Rev Mol Cell Biol *9*, 33–46.

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

Choi, E.S., Stralfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of Noncoding Transcripts from within CENP-A Chromatin at Fission Yeast Centromeres. J Biol Chem *286*, 23600–23607.

Choo, K.H. (1997). Centromere DNA Dynamics: Latent Centromeres and Neocentromere Formation. Am J Hum Genet *61*, 1225–1233.

Chun, Y., Park, B., Koh, W., Lee, S., Cheon, Y., Kim, R., Che, L., and Lee, S. (2011). New Centromeric Component CENP-W Is an RNA-associated Nuclear Matrix Protein That Interacts with Nucleophosmin/B23 Protein. J Biol Chem *286*, 42758–42769.

Cimini, D., Howell, B., Maddox, P.S., Khodjakov, A., Degrassi, F., and Salmon, E.D. (2001). Merotelic Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue Cells. J Cell Biol *153*, 517–527.

Compton, D.A. (2011). Mechanisms of aneuploidy. Curr Opin Cell Biol *23*, 109–113.

Daban, J.-R. (2000). Physical Constraints in the Condensation of Eukaryotic Chromosomes. Local Concentration of DNA versus Linear Packing Ratio in Higher Order Chromatin Structures. Biochemistry *39*, 3861–3866.

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

Darling, S.M., Crampton, J.M., and Williamson, R. (1982). Organization of a family of highly repetitive sequences within the human genome. J Mol Biol *154*, 51–63.

De Rop, V., Padeganeh, A., and Maddox, P.S. (2012). CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. Chromosoma *121*, 527–538.

Dechassa, M.L., Wyns, K., Li, M., Hall, M.A., Wang, M.D., and Luger, K. (2011). Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. Nat Commun *2*, 1–10.

Depinet, T., Zackowski, J.L., Earnshaw, W., Kaffe, S., Sekhon, G.S., Stallard, R., Sullivan, B.A., Vance, G.H., Van Dyke, D.L., Willard, H.F., et al. (1997). Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. Hum Mol Genet *6*, 1195–1204.

Du, Y., Topp, C.N., and Dawe, R.K. (2010). DNA Binding of Centromere Protein C (CENPC) Is Stabilized by Single-Stranded RNA. PLoS Genet *6*, e1000835.

Dunleavy, E.M., Almouzni, G., and Karpen, G.H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase. Nucleus *2*, 146–157.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

Earle, E., Saxena, A., MacDonald, A., Hudson, D.F., Shaffer, L.G., Saffery, R., Cancilla, M.R., Cutts, S.M., Howman, E., and Choo, K.H.A. (2000). Poly(ADP-ribose) polymerase at active centromeres and neocentromeres at metaphase. Hum Mol Genet *9*, 187–194.

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

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

Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F., and Cleveland, D.W. (1987). Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. J Cell Biol *104*, 817–829.

Earnshaw, W., Bordwell, B., Marino, C., and Rothfield, N. (1986). Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. J Clin Invest *77*, 426–430.

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

Etienne-Manneville, S., and Hall, A. (2002). Rho GTPases in cell biology. Nature *420*, 629–635.

Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., et al. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. Nat. Cell Biol. *15*, 1–20.

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

Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science *319*, 94–97.

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

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

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

Fukagawa, T., Pendon, C., Morris, J., and Brown, W. (1999). CENP-C is necessary but not sufficient to induce formation of a functional centromere. Embo J *18*, 4196–4209.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. Cell *145*, 410–422.

Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., et al. (2012). An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. Nature *484*, 534–537.

Gent, J.I., and Dawe, R.K. (2012). RNA as a Structural and Regulatory Component of the Centromere. Annu. Rev. Genet. *46*, 443–453.

Glotzer, M. (2005). The Molecular Requirements for Cytokinesis. Science *307*, 1735–1739.

Goshima, G., Saitoh, S., and Yanagida, M. (1999). Proper metaphase spindle length is determined by centromere proteins Mis12 and Mis6 required for faithful chromosome segregation. Genes Dev *13*, 1664–1677.

Haaf, T., Warburton, P.E., and Willard, H.F. (1992). Integration of human α-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell *70*, 681–696.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Ben E Black (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat Struct Mol Biol *20*, 1–10.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell *118*, 715–729.

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

Hirose, K., Kawashima, T., Iwamoto, I., Nosaka, T., and Kitamura, T. (2000). MgcRacGAP Is Involved in Cytokinesis through Associating with Mitotic Spindle and Midbody. J Biol Chem *276*, 5821–5828.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., McEwen, B.F., Shang, W.H., Suzuki, E., Okawa, K., et al. (2008). CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. Cell *135*, 1039–1052.

Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y., et al. (2011). Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. Genes Dev *25*, 901–906.

Hudson, D.F., Fowler, K.J., Earle, E., Saffery, R., Kalitsis, P., Trowell, H., Hill, J., Wreford, N.G., de Kretser, D.M., Cancilla, M.R., et al. (1998). Centromere Protein B Null Mice are Mitotically and Meiotically Normal but Have Lower Body and Testis Weights. J Cell Biol *141*, 309–319.

Ikeno, M., Masumoto, H., and Okazaki, T. (1994). Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. Hum Mol Genet *3*, 1245–1257.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. (1998). Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol *16*, 431–439.

Izuta, H., Ikeno, M., Suzuki, N., Tomonaga, T., Nozaki, N., Obuse, C., Kisu, Y., Goshima, N., Nomura, F., Nomura, N., et al. (2006). Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. Genes Cells *11*, 673–684.

Jain, A., Liu, R., Ramani, B., Arauz, E., Ishitsuka, Y., Ragunathan, K., Park, J., Chen, J., Xiang, Y.K., and Ha, T. (2011). Probing cellular protein complexes using single-molecule pull-down. Nature *473*, 484–488.

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

Jantsch-Plunger, V., Gönczy, P., Romano, A., Schnabel, H., Hamill, D., Schnabel, R., Hyman, A.A., and Glotzer, M. (2000). Cyk-4: a Rho family GTPase activating protein (GAP) required for central spindle formation and cytokinesis. J Cell Biol *149*, 1391–1404.

Kagansky, A., Folco, H.D., Almeida, R., Pidoux, A.L., Boukaba, A., Simmer, F., Urano, T., Hamilton, G.L., and Allshire, R.C. (2009). Synthetic Heterochromatin Bypasses RNAi and Centromeric Repeats to Establish Functional Centromeres. Science *324*, 1716–1719.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev *19*, 295–316.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D.N., Ogata, K., Nishimura, Y., and Ishii, S. (1990). The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product. J Biol Chem *265*, 19990–19995.

Kim, H., Lee, M., Lee, S., Park, B., Koh, W., Lee, D.J., and Lim, D.S. (2009). Cancer-upregulated gene 2 (CUG2), a new component of centromere complex, is required for kinetochore function. Mol Cells *27*, 697–701.

Kim, I.S., Lee, M., Park, J.H., Jeon, R., Baek, S.H., and Kim, K.I. (2013). βTrCP-mediated ubiquitylation regulates protein stability of Mis18β in a cell cycle-dependent manner. Biochem Bioph Res Co *443*, 62–67.

Kim, I.S., Lee, M., Park, K.C., Jeon, Y., Park, J.H., Hwang, E.J., Jeon, T.I., Ko, S., Lee, H., Baek, S.H., et al. (2012). Roles of Mis18alpha in epigenetic regulation of centromeric chromatin and CENP-A loading. Mol Cell *46*, 260–273.

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

Krude, T. (1999). Chromatin assembly during DNA replication in somatic cells. Eur J Biochem *263*, 1–5.

Kuwata, T., Yamazaki, Y., Jenkins, N.A., Copeland, N.G., Osato, M., Ito, Y., Kroon, E., Sauvageau, G., and Nakamura, T. (2005). Identification of cooperative genes for NUP98-HOXA9 in myeloid leukemogenesis using a mouse model. Blood *105*, 784–793.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lafaurie-Janvore, J., Maiuri, P., Wang, I., Pinot, M., Manneville, J.B., Betz, T., Balland, M., and Piel, M. (2013). ESCRT-III Assembly and Cytokinetic Abscission Are Induced by Tension Release in the Intercellular Bridge. Science *339*, 1625–1629.

Lagana, A., Dorn, J.F., De Rop, V., Ladouceur, A.M., Maddox, A.S., and Maddox, P.S. (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat. Cell Biol. *12*, 1186–

1193.

Lee, W.C., Zabetakis, D., and Mélèse, T. (1992). NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. Mol Cell Biol *12*, 3865–3871.

Lejeune, E., Bayne, E.H., and Allshire, R.C. (2011). On the Connection between RNAi and Heterochromatin at Centromeres. Cold Spring Harb Symp Quant Biol *75*, 275–283.

Lejeune, E., Bortfeld, M., White, S.A., Pidoux, A.L., Ekwall, K., Allshire, R.C., and Ladurner, A.G. (2007). The Chromatin-Remodeling Factor FACT Contributes to Centromeric Heterochromatin Independently of RNAi. Curr Biol *17*, 1219–1224.

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., Demidov, D., Schubert, V., and Schubert, I. (2013). Arabidopsis KINETOCHORE NULL2 Is an Upstream Component for Centromeric Histone H3 Variant cenH3 Deposition at Centromeres. Plant Cell *25*, 3389–3404.

LeRoy, G., Orphanides, G., Lane, W.S., and Reinberg, D. (1998). Requirement of RSF and FACT for Transcription of Chromatin Templates in Vitro. Science *282*, 1900–1904.

Levan, A., Fredga, K., and Sandberg, A.A. (1964). Nomenclature for centromeric position on chromosomes. Hereditas *52*, 201–220.

Lo, A.W.I., Liao, G.C.C., Rocchi, M., and Choo, K.H.A. (1999). Extreme Reduction of Chromosome-Specific α-Satellite Array Is Unusually Common in Human Chromosome 21. Genome Res *9*, 895–908.

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

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol *176*, 757–763.

Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). "'Holo'"er than thou: Chromosome segregation and kinetochore function in *C. elegans*. Chromosome Res. *12*, 641–653.

Maiato, H. (2004). The dynamic kinetochore-microtubule interface. J Cell Sci *117*, 5461–5477.

Manuelidis, L. (1976). Repeating restriction fragments of human DNA. Nucleic Acids Res *3*, 3063–3076.

Manuelidis, L. (1978). Chromosomal localization of complex and simple repeated human DNAs. Chromosoma *66*, 23–32.

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

Meijer, L., Borgne, A., Mulner, O., Chong, J.P.J., Blow, J.J., Inagaki, N., Inagaki, M., Delcros, J.-G., Moulinoux, J.-P. (1997) Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of cyclin-dependent kinases cdc2, cdk2 and cdk5. Eur J Biochem *243*, 527-536.

Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I., and Karpen, G.H. (2011). Assembly of Drosophila Centromeric Chromatin Proteins during Mitosis. PLoS Genet *7*, e1002068.

Mitchell, A.R., Gosden, J.R., and Miller, D.A. (1985). A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma *92*, 369–377.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M., and Wu, C. (2007). Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. Cell *129*, 1153–1164.

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

Morgan, D.O. (2007). The Cell Cycle, Principles of Control (Primers in Biology).

Murano, K., Okuwaki, M., Hisaoka, M., and Nagata, K. (2008). Transcription Regulation of the rRNA Gene by a Multifunctional Nucleolar Protein, B23/Nucleophosmin, through Its Histone Chaperone Activity. Mol Cell Biol *28*, 3114–3126.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J Cell Biol *116*, 585–596.

Nardi, I.K., Stellfox, M.E., Knipper, C.M., Zasadzińska, E., and Foltz, D.R. (2013). Mis18 forms a conserved tetrameric complex required for centromere recognition in humans. Annual Meeting of American Society for Cell Biology.

Nicklas, R.B., and Koch, C.A. (1969). Chromosome micromanipulation, III. Spindle Fiber Tension and the Reorientation of Mal-Oriented Chromosomes. Jcb *43*, 40–50.

Nicklas, R.B., Waters, J.C., Salmon, E.D., and Ward, S.C. (2001). Checkpoint signals

in grasshopper meiosis are sensitive to microtubule attachment, but tension is still essential. J Cell Sci *114*, 4173–4183.

Nishino, T., Takeuchi, K., Gascoigne, K.E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I.M., and Fukagawa, T. (2012). CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. Cell *148*, 487–501.

Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (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*, 105–120.

Ochs, R.L., and Press, R.I. (1992). Centromere autoantigens are associated with the nucleolus. Exp Cell Res *200*, 339–350.

Oegema, K., and Hyman, A.A. (2006). Cell Division, (WormBook).

Oegema, K., Desai, A., Rybina, S., Kirkham, M., and Hyman, A.A. (2001). Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. J Cell Biol *153*, 1209–1226.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc Natl Acad Sci U S A *89*, 6428–6432.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, H., and Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell *79*, 639–648.

Ohzeki, J., Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol *159*, 765–775.

Okada, T., Ohzeki, J.I., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B Controls Centromere Formation Depending on the Chromatin Context. Cell *131*, 1287–1300.

Olins, A.L., and Olins, D.E. (1974). Speroid chromatin units (N bodies). Science *183*, 330–332.

Olins, A.L., Senior, M.B., and Olins, D.E. (1976). Ultrastructural features of chromatin nu bodies. J Cell Biol *68*, 787–792.

Otting, G., Qian, Y., Müller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988). Secondary structure determination for the Antennapedia homeodomain by nuclear

magnetic resonance and evidence for a helix-turn-helix motif. Embo J *7*, 4305–4309.

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

Padeganeh, A., De Rop, V., and Maddox, P.S. (2013a). Nucleosomal composition at the centromere: a numbers game. Chromosome Res. *21*, 27–36.

Padeganeh, A., Ryan, J., Boisvert, J., Ladouceur, A.-M., Dorn, J.F., and Maddox, P.S. (2013b). Octameric CENP-A Nucleosomes Are Present at Human Centromeres throughout the Cell Cycle. Curr Biol *23*, 1–6.

Padeken, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.-J., Kremmer, E., and Heun, P. (2013). The Nucleoplasmin Homolog NLP Mediates Centromere Clustering and Anchoring to the Nucleolus. Mol Cell *50*, 236–249.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A *88*, 3734–3738.

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

Perpelescu, M., Nozaki, N., Obuse, C., Yang, H., and Yoda, K. (2009). Active establishment of centromeric CENP-A chromatin by RSF complex. J Cell Biol *185*, 397–407.

Peters, C.W., Sippel, A.E., Vingron, M., and Klempnauer, K.H. (1987). Drosophila and vertebrate myb proteins share two conserved regions, one of which functions as a DNA-binding domain. Embo J *6*, 3085–3090.

Pluta, A.F., and Earnshaw, W.C. (1996). Specific Interaction between Human Kinetochore Protein CENP-C and a Nucleolar Transcriptional Regulator. J Biol Chem *271*, 18767–18774.

Politi, V., Perini, G., Trazzi, S., Pliss, A., Raska, I., Earnshaw, W.C., and Valle, Della, G. (2002). CENP-C binds the alpha-satellite DNA *in vivo* at specific centromere domains. J Cell Sci *115*, 2317–2327.

Rieder, C.L. (1979). Ribonucleoprotein staining of centrioles and kinetochores in newt lung cell spindles. J Cell Biol *80*, 1–9.

Robbins, E., and Gonatas, N.K. (1964). The Ultrastructure of a Mammalian Cell during the Mitotic Cycle. J Cell Biol *21*, 429–463.

Robinson, P.J.J., Fairall, L., Huynh, V.A.T., and Rhodes, D. (2006). EM

measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A *103*, 6506–6511.

Roostalu, J., Schiebel, E., and Khmelinskii, A. (2010). Cell cycle control of spindle elongation. Cell Cycle *9*, 1084–1090.

Saffery, R., Irvine, D.V., griffith, B., Kalitsis, P., Wordeman, L., and Choo, K.H. (2000). Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. Hum Mol Genet *9*, 175–185.

Saikumar, P., Murali, R., and Reddy, E.P. (1990). Role of tryptophan repeats and flanking amino acids in Myb-DNA interactions. Proc Natl Acad Sci U S A *87*, 8452–8456.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., III, Maurer, M., Rothfield, N.F., and Earnshaw, W.C. (1992). CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. Cell *70*, 115–125.

Saitoh, S., Takahashi, K., and Yanagida, M. (1997). Mis6, a Fission Yeast Inner Centromere Protein, Acts during G1/S and Forms Specialized Chromatin Required for Equal Segregation. Cell *90*, 131–143.

Sanchez-Pulido, L., Pidoux, A.L., Pontig, C.P., Allshire, R.C. (2009). Common Ancestry of the CENP-A Chaperones Scm3 and HJURP. Cell *137*, 1173-1174.

Sart, D.D., Cancilla, M.R., Earle, E., Mao, J.-I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E., and Choo, K.H.A. (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. Nat Genet *16*, 144–153.

Schneider, A. (1873). Untersuchungen über Plathelminthen. Jahrb. Oberhess. Ges. Naturwiss. Heilk. 69–81.

Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol *17*, 237–243.

Schwarz, P.M., and Hansen, J.C. (1994). Formation and stability of higher order chromatin structures. Contributions of the histone octamer. J Biol Chem *269*, 16284–16289.

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

Shah, K., McCormack, C.E., and Bradbury, N.A. (2014). Do you know the sex of your cells? Am J Physiol Cell Physiol *306*, C3–C18.

Shelby, R.D., Vafa, O., and Sullivan, K.F. (1997). Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. J Cell Biol *136*, 501–513.

Shiio, Y., and Eisenman, R.N. (2011). Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A *100*, 13225–13230.

Shivaraju, M., Camahort, R., Mattingly, M., and Gerton, J.L. (2011). Scm3 is a centromeric nucleosome assembly factor. J Biol Chem *286*, 12106–12023.

Shuaib, M., Ouararhni, K., Dimitrov, S., and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc Natl Acad Sci U S A *107*, 1349–1354.

Silva, M.C., Bodor, D.L., Stellfox, M.E., Martins, N.M., Hochegger, H., Foltz, D.R., and Jansen, L.E. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev Cell *22*, 52–63.

Singer, M.F. (1982). Highly repeated sequences in mammalian genomes. Int Rev Cytol *76*, 67–112.

Spencer, F., and Hieter, P. (1992). Centromere DNA mutations induce a mitotic delay in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A *89*, 8908–8912.

Sprague, B.L., Pego, R.L., Stavreva, D.A., and McNally, J.G. (2004). Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching. Biophys J *86*, 3473–3495.

Stoler, S., Rogers, K., Weitze, S., Morey, L., Fitzgerald-Hayes, M., and Baker, R.E. (2007). Scm3, an essential *Saccharomyces cerevisiae* centromere protein required for G2/M progression and Cse4 localization. Proc Natl Acad Sci U S A *104*, 10571–10576.

Sullivan, B.A., and Schwartz, S. (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum Mol Genet *4*, 2189–2197.

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581–592.

Suzuki, N., Nakano, M., Nozaki, N., Egashira, S., Okazaki, T., and Masumoto, H. (2003). CENP-B Interacts with CENP-C Domains Containing Mif2 Regions Responsible for Centromere Localization. J Biol Chem *279*, 5934–5946.

Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 Centromere Connector for Localizing a CENP-A-Like Protein in Fission Yeast.

Science *288*, 2215–2219.

Tanaka, Y., Tachiwana, H., Yoda, K., Masumoto, H., Okazaki, T., Kurumizaka, H., and Yokoyama, S. (2005). Human Centromere Protein B Induces Translational Positioning of Nucleosomes on Alpha-Satellite Sequences. J Biol Chem *280*, 41609–41618.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025–1040.

Tomkiel, J., Cooke, C.A., Saitoh, H., Bernat, R.L., and Earnshaw, W.C. (1994). CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. J Cell Biol *125*, 531–545.

Toure, A., Dorseuil, O., Morin, L., Timmons, P., Jégou, B., Reibel, L., and Gacon, G. (1998). MgcRacGAP, A New Human GTPase-activating Protein for Rac and Cdc42 Similar to Drosophila rotundRacGAP Gene Product, Is Expressed in Male Germ Cells. J Biol Chem *273*, 6019–6023.

Uhlmann, F. (2001). Chromosome condensation: Packaging the genome. Curr Biol *11*, R384–R387.

Vafa, O., and Sullivan, K.F. (1997). Chromatin containing CENP-A and alpha-satellite DNA is a major component of the inner kinetochore plate. Curr Biol *7*, 897–900.

Valdivia, M.M., and Brinkley, B.R. (1985). Fractionation and initial characterization of the kinetochore from mammalian metaphase chromosomes. J Cell Biol *101*, 1124–1134.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci *114*, 3529–3542.

Verreault, A. (2000). De novo nucleosome assembly: new pieces in an old puzzle. Genes Dev *14*, 1430–1438.

Vissel, B., and Choo, K.H. (1987). Human alpha satellite DNA--consensus sequence and conserved regions. Nucleic Acids Res *15*, 6751–6752.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science *297*, 1833–1837.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein:

activation of a latent centromere? Am J Hum Genet *52*, 1153–1163.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., D, W., Tyler-Smith, C., Sullivan, K.F., et al. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol *7*, 901–904.

Waye, J.S., and Willard, H.F. (1987). Nucleotide sequence heterogeneity of alpha satellite repetitive DNA: a survey of alphoid sequences from different human chromosomes. Nucleic Acids Res *15*, 7549–7569.

Westin, E.H., Gallo, R.C., Arya, S.K., Eva, A., Souza, L.M., Baluda, M.A., Aaronson, S.A., and Wong-Staal, F. (1982). Differential expression of the amv gene in human hematopoietic cells. Proc Natl Acad Sci U S A *79*, 2194–2198.

Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. Am J Hum Genet *37*, 524–532.

Willard, H.F., and Waye, J.S. (1987a). Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet *3*, 192–198.

Willard, H.F., and Waye, J.S. (1987b). Chromosome-specific subsets of human alpha-satellite DNA: analysis of sequence divergence within and between chromosomal subset and evidence for an ancestral pentameric repeat. J Mol Evol *25*, 207–214.

Willard, H.F., Smith, K.D., and Sutherland, J. (1983). Isolation and characterization of a major tandem repeat family from the human X chromosome. Nucleic Acids Res *11*, 2017–2033.

Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., et al. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160.

Woodcock, C.L., and Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. Curr Opin Genet Dev *11*, 130–135.

Woodcock, C.L., Safer, J.P., and Stanchfield, J.E. (1976). Structural repeating units in chromatin: I. Evidence for their general occurence. Exp Cell Res *97*, 101–110.

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. Proc Natl Acad Sci U S A *97*, 7266–7271.

Yoda, K., Ando, S., Okuda, A., Kikuchi, A., and Okazaki, T. (1998). *In vitro* assembly of the CENP-B/α-satellite DNA/core histone complex: CENP-B causes nucleosome

positioning. Genes Cells *3*, 533–548.

Yoda, K., Kitagawa, K., Masumoto, H., Muro, Y., and Okazaki, T. (1992). A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH2 terminus, which is separable from dimerizing activity. J Cell Biol *119*, 1413–1427.

Yuen, K.W., Nabeshima, K., Oegema, K., and Desai, A. (2011). Rapid de novo centromere formation occurs independently of heterochromatin protein 1 in *C. elegans* embryos. Curr Biol *21*, 1800–1807.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. Mol Cell *13*, 291–298.

Yüce, Ö., Piekny, A., and Glotzer, M. (2005). An ECT2–centralspindlin complex regulates the localization and function of RhoA. J Cell Biol *170*, 571–582.

Zhang, D., Martyniuk, C.J., and Trudeau, V.L. (2006). SANTA domain: a novel conserved protein module in *Eukaryota* with potential involvement in chromatin regulation. Bioinformatics *22*, 2459–2462.

Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., Miller Jenkins, L.M., Xiao, H., Tjandra, N., Wu, C., et al. (2011). Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. Nature *472*, 234–238.

# CHAPTER 2

Publication

# CENP-A: THE KEY PLAYER BEHIND CENTROMERE IDENTITY, PROPAGATION, AND KINETOCHORE ASSEMBLY

Review published in Chromosoma (2012) Dec;121(6):527-38. PMID 23095988

## 2.1 Preface

This published review discusses the latest findings in centromere biology field until 2012. I wrote the main text with suggestions and corrections brought by Abbas Padeganeh and Dr. Paul S. Maddox. Dr. Paul S. Maddox and I designed the figures and I drew all of them using Illustrator program.

# CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly

Valérie De Rop, Abbas Padeganeh, Paul S. Maddox

Institute for Research in Immunology and Cancer (IRIC),
Department of Pathology and Cell Biology
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7  Canada

## 2.2 Abstract

Chromosome segregation is the one of the great problems in biology with complexities spanning from biophysics and polymer dynamics to epigenetics. Here, we summarize the current knowledge and highlight gaps in understanding of the mechanisms controlling epigenetic regulation of chromosome segregation.

*Key words*
Centromere, CENP-A, epigenetic, cell cycle

## 2.3 Introduction

Centromeres were first described by Walther Flemming as the primary constriction on condensed chromosomes, where cellular fibers, now known to be microtubules, attached during mitosis (Flemming 1882). Electron microscopy images later led to the realization that centromeres are chromosomal loci where the megadalton protein complex named the kinetochore assembles (Robbins and Gonatas 1964; Brinkley and Stubblefield 1966). The definition of centromeres, while still following these early rules, has become more complex with the discovery of molecular players involved in centromere identity. Counter intuitively, extensive studies on the expression and localization of different centromere components throughout the cell cycle have not led to a consensus mechanism that defines centromere specification. In this review, we focus on the latest discoveries in the field and specifically on epigenetic markers for centromere identity.

## 2.4 CENP-A as an epigenetic marker for centromere identity

Centromeres are epigenetically defined by a variant of histone H3, centromere protein-A (CENP-A). CENP-A was serendipitously discovered in 1985 by William Earnshaw in the course of immunoblotting and immunostaining experiments. Blotting serum isolated from CREST syndrome patients identified three recurrent bands common among many patients. Immunostaining using the same sera in tissue culture cells showed an enrichment at centromeres (Earnshaw and Rothfield 1985; Earnshaw et al. 1986; Valdivia and Brinkley 1985). These proteins were accordingly named centromere proteins A, B, and C. Later, biochemical approaches demonstrated that CENP-A copurified with histones and is a bona fide component of nucleosome particles (Palmer et al. 1987; Palmer et al. 1991). CENP-A contains a histone fold domain and is able to replace histone H3 in centromeric nucleosomes (Sullivan et al. 1994; Yoda et al. 2000). Sequence analysis of CENP-A and histone H3 reveals a 60% similarity within the histone fold domains with major differences

concentrated in the N-terminus (Sullivan et al. 1994). Through the following years, CENP-A homologues were identified in all eukaryotic model systems investigated: for example HCP-3 in *Caenorhabditis elegans*, CID in *Drosophila melanogaster*, Cse4 in *Saccharomyces cerevisiae*, Cnp1 in *Schizosaccharomyces pombe*, and CenH3 in *Arabidopsis thaliana* (Buchwitz et al. 1999; Blower and Karpen 2001; Stoler et al. 1995; Takahashi et al. 2000; Talbert et al. 2002). Interestingly, CENP-A is poorly conserved compared to other histone proteins, which are almost invariant through evolution at the amino acid level. Divergence, while extreme in the N-terminal tail of CENP-A, is prevalent even within the C-terminal histone fold of closely related species. Even if CENP-A is poorly conserved at the sequence level, its structure or active chemical tags (i.e., acetylation and methylation) may be specific features that are keys to epigenetic mechanisms controlling chromosome segregation.

Understanding differences between CENP-A and histone H3 has long been thought to hold the answer to epigenetic regulation of centromeres. In an attempt to understand the differences in molecular dynamics between the two histone proteins, the Cleveland lab identified a specific domain of CENP-A which they called the CENP-A targeting domain (CATD) using a hydrogen/deuterium exchange technique coupled to mass spectrometry (H/DX-MS). With this technique, they showed less deuterium exchange in the CATD domain of CENP-A and thus concluded that the CATD is less flexible and more compacted (Black et al. 2004; Black et al. 2007a). Remarkably, the CATD domain was shown to be sufficient for CENP-A localization to centromeres as demonstrated by swapping the CATD domain from CENP-A in histone H3 chimeric proteins (Black et al. 2007b). This exciting result was later shown to be due to recognition by a centromere-specific chaperone protein, Holiday junction-recognizing protein (HJURP, see below) (Bassett et al. 2012). Consistent with protein structure playing a critical role, HJURP binding to CENP-A-H4 induces more compaction and less flexibility compared to a control condition. Thus, it is clear that structural distinction in CENP-A is essential for centromere identity and function.

Subsequent structural studies have a yielded a more precise understanding of the atomic differences between H3 and CENP-A nucleosomes. For instance, the CENP-A-H4 tetramer crystal obtained by the Black lab showed unique biophysical properties of CENP-A nucleosomes (Sekulic et al. 2010). Overlay of H3-H4 and CENP-A-H4 tetramers revealed that the centromere-specific tetramer is rotated between dimer pairs compared to H3-H4 tetramer. This rotation is caused by two specific residues, His104 and Leu122, located at the CENP-A/CENP-A interface. Also, these residues were shown to be essential for CENP-A localization to the centromeres. Moreover, this rotation makes the CENP-A-H4 tetramer more compacted and less flexible as demonstrated by H/DX-MS. Overall, these features make the CENP-A-H4 tetramer narrower and shorter in 2D and wider in 3D compared to H3-H4 tetramers. In addition, the same group demonstrated a bulged structure in the L1 loop conferred by the Arg80 and Gly81 residues. Concomitantly, Kurumizaka's group also identified these residues and showed their importance in the stability of CENP-A localization (Tachiwana et al. 2012). This group solved the CENP-A nucleosome crystal structure and discovered that the CENP-A αN is shorter than that of histone H3, although they did not demonstrate the importance of this structural feature. All together, these results clearly demonstrate significant structural differences between CENP-A and H3 derived octamers.

The octamer structure is not the only source of distinction between CENP-A and H3. Recent studies of DNA wrapping topology by H/DX-MS showed that residues causing structural deformation are found in the αN part of CENP-A sequence adjacent to the DNA entry-exit site of the nucleosome (Panchenko et al. 2011). In canonical chromatin, this site is known to be recognized by diverse functional proteins, e.g., for transcriptional control by stabilization of nucleosomes, inhibition of nucleosome sliding, and compaction of chromatin in mitosis (Zlatanova et al. 2008). Even more radical differences in DNA wrapping have been reported. Henikoff's group provided evidence suggesting that centromeric DNA, instead of having a left-handed wrapping around nucleosomes with negative supercoiling, as for canonical H3 nucleosomes, is in fact wrapped in a right-handed manner causing positive supercoils or less negative

supercoils (Furuyama and Henikoff 2009). This was shown biochemically using an extrachromosomal plasmid and determining the state of DNA by inducing structural deformation by chloroquine infused gel electrophoresis. In summary, on one hand, we now have a good understanding of the structural differences between CENP-A and H3 nucleosomes, while on the other hand, new questions are arising regarding the precise topology of centromeric chromatin. Some progress has been made in the last year to understand if the observed structures are species and/or cell cycle specific (Bui et al. 2012; Shivaraju et al. 2012). Nonetheless, it is still controversial whether those structures are critical for CENP-A loading, incorporation, and stabilization at centromeres.

## 2.5 Does DNA sequence have any role in centromere propagation and identity?

For decades, centromere identity has been thought to rely on an epigenetic mechanism due to a myriad of experimental evidences, with the exception of *S. cerevisiae*. Centromere sequences are highly divergent throughout species (Figure 2.1). In *S. cerevisiae*, centromeres are 125bp and composed of three Centromere DNA elements (CDEI-III). A single mutation in the CDEII element abrogates Cse4 incorporation and leads to a loss of centromere function and cell death (Cottarel et al. 1989; Spencer and Hieter 1992). In *S. pombe*, centromeres consist of 40 to 100 Kbp of repeated and inverted sequences (Clarke et al. 1986; Fishel et al. 1988). *D. melanogaster* centromeres are repeats of transposon and satellite sequences (AATAT and CTCTT) that can measure up to 420 Kbp (Murphy and Karpen 1995; Sun et al. 1997). In humans, centromeres are composed of A/T rich α-satellite repeat sequences of 171bp arranged in a head to tail orientation, ranging in total size from 1 up to 5 Mbp (Tyler-Smith et al. 1993; Manuelidis 1978; Mitchell et al. 1985; Willard 1985). Interestingly, human centromeres vary in genetic length on different chromosomes. The discovery of neocentromeres in humans is the key experimental evidence (provided by the stochastic nature of biology) for epigenetic regulation of

centromeres. Found in exceedingly low frequency, neocentromeres form on a unique chromosomal locus distinct, and typically displaced by mega bases, from the “normal” centromere genomic position. This genomic displacement is not due to translocation or rearrangements and represents a novel epigenetic event. Neocentromeres contain CENP-A nucleosomes and are able to direct assembly of functional kinetochores supporting normal development, all in the absence of α-satellite DNA (Warburton et al. 1997). In addition, considering that α-satellite sequences can be found in non-centromeric loci (called inactive centromere), it is assumed that DNA sequence is not sufficient to incorporate CENP-A and build a functional kinetochore (Van Hooser et al. 2001; Earnshaw et al. 1989; Haaf et al. 1992; Warburton et al. 1997). To demonstrate this point, experiments performed by Earnshaw’s group on dicentric chromosomes showed that only the active centromere, and not an inactive one, can build a functional kinetochore (by recruiting CENP-C and nucleating kinetochore assembly) even if the two centromeres are composed of α-satellite DNA (Figure 2). In sum, all observations lead to the conclusion that centromeric DNA is neither necessary nor sufficient for centromere identity (with the noteworthy exception of *S. cerevisiae*).

There is a clear distinction to make between centromere propagation and de novo formation. Recently, evidence has emerged to indicate a role for centromeric DNA sequences in *de novo* formation of centromeres. Human alphoid DNA repeats found at centromeres can be classified into two types: α21-I and α21-II, with α21-I containing a 17-bp sequence called the CENP-B box (Ikeno et al. 1994). The CENP-B box is recognized and bound specifically by CENP-B, the only centromeric protein recognizing the centromeric DNA (Masumoto et al. 1989; Muro et al. 1992). Masumoto’s group demonstrated that the CENP-B box together with alphoid DNA sequence is sufficient for de novo centromere formation in human tissue culture cells. These artificial centromeres contain CENP-A and are competent to recruit kinetochore components such as CENP-C and CENP-E (Ohzeki et al. 2002). *C. elegans* sperm DNA is known to be deficient of CENP-A protein before fertilization and gains it after entering the oocyte. Despite the lack of CENP-B protein in worms, the sperm DNA is able to form a new centromere via an unknown molecular

mechanism (Gassmann et al. 2012). Another study showed that the Aurora B mitotic kinase, an inner centromere protein known to be required for correcting abnormal kinetochore-microtubule attachments, decreases at neocentromeres, concluding that this chromosomal environment, for an unknown reason, is less favorable for inner centromere maturation (Bassett et al. 2010). Hypothetically, decreased inner centromere assembly could be due to the DNA sequence itself, or it could also be that alphoid DNA sequences which are A/T rich could promote a more compacted chromatin structure that is favorable for inner centromere assembly

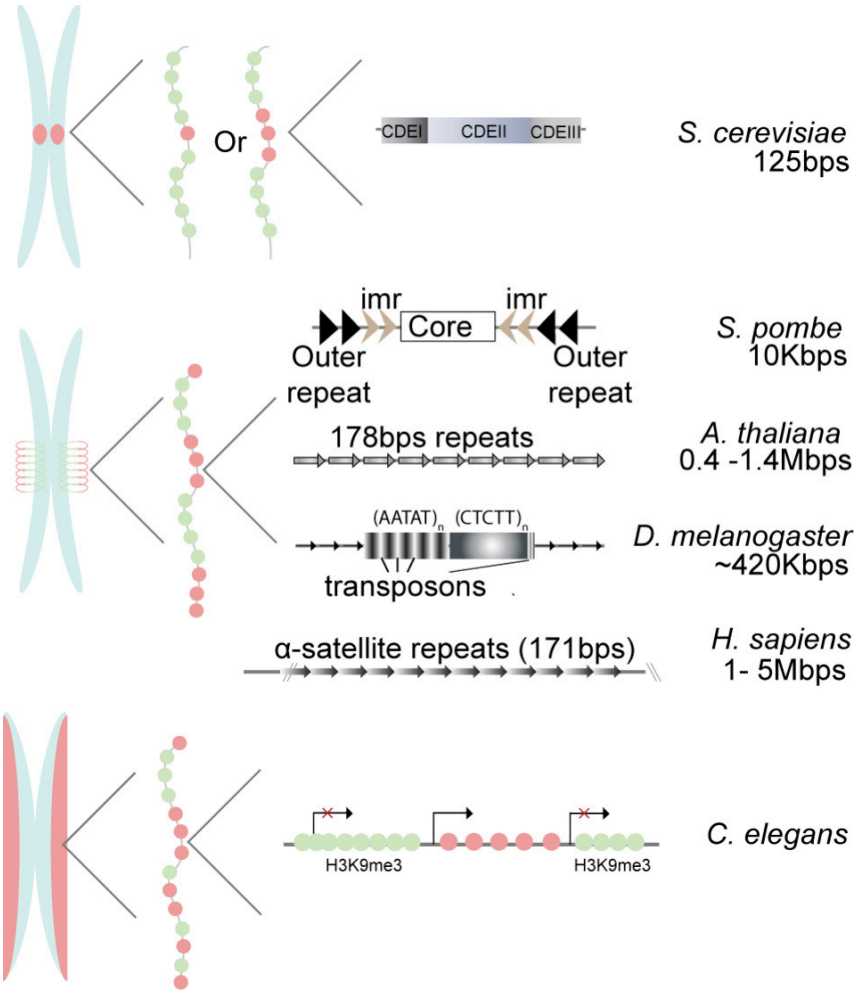



**Figure 2.1 Centromeric DNA sequences are not conserved through species.**

With the exception of *S. cerevisiae*, most evidence shows that DNA sequence has no role in centromere identity. Species comparison reveals vast differences in nucleotide composition as well as the centromere length. *S. cerevisiae* has a 125bp centromere which is divided in three centromere DNA elements (CDE), with CDEII being the sequence important for Cse4 incorporation. It is still somewhat controversial whether this centromere is composed of a single or two to three CENP-A nucleosomes. (Lawrimore et al. 2011; Furuyama and Biggins 2007) *S. pombe* 10Kbp centromere consists of inner and outer repeats located outside the core region and in a head-to-head orientation. *D. melanogaster* has relatively large centromeres made up of DNA repeats and transposon elements for a genomic size of ~420Kbp. *A. thaliana* and

*Homo sapiens* centromeres are made of head-to-tail 171-178bp repeats that can go up to 1.4Mbp for the plant and 5Mbp for human. General centromere structures display two general types of centromere: monocentric for *S. cerevisiae* to *H. sapiens*, and holocentric (whole length of chromosome) for *C. elegans* (generally nematodes and several other species) (Maddox et al. 2004; Melters et al. 2012) Elongated centromeric chromatin may have different arrangements of CENP-A and H3 nucleosomes arrays, which are repetitive and exclusive from one another (Blower et al. 2002). When compacted, the CENP-A arrays form a hypothetical centromeric plate required for kinetochore formation in mitosis.

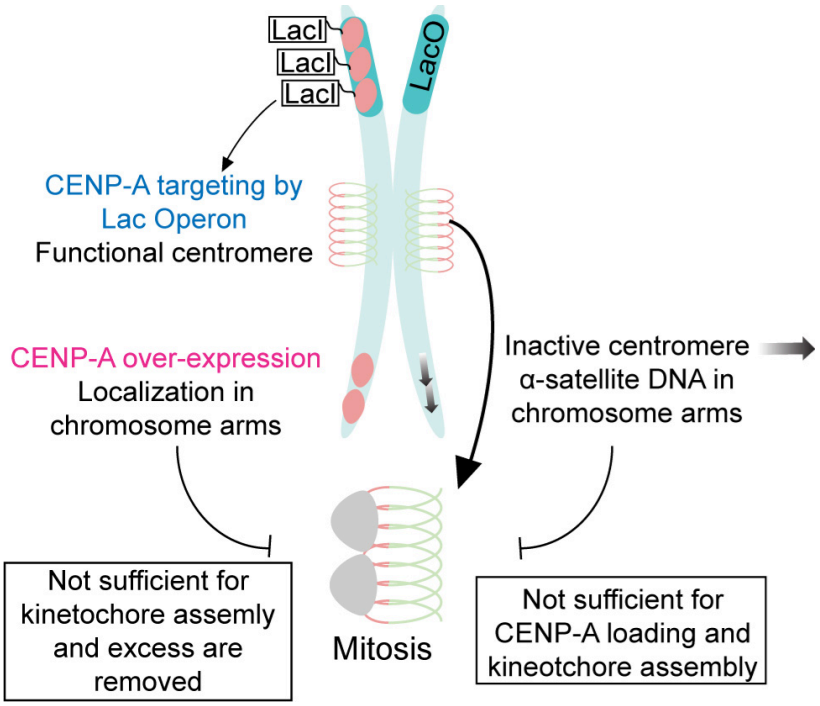



**Figure 2.2 Compacted CENP-A chromatin promotes kinetochore assembly and centromere propagation.** An inactive centromere, which is a chromosomal locus having alpha-satellite sequence without CENP-A protein, demonstrated that centromeric sequence is not sufficient for kinetochore assembly. Kinetochore proteins are recruited to this chromosomal locus only when CENP-A is present (active centromere) (Warburton et al. 1997). Conversely, when CENP-A is overexpressed in cells, it incorporates on chromosome arms. However, this mislocalized CENP-A does not recruit kinetochore proteins in mitosis. Ultimately, by targeting CENP-A on chromosome arms using the lac operon technique, this creates a high-density array of CENP-A protein and forms a neocentromere: it recruits kinetochore components and is able to propagate for a limited time. Those observations suggest that CENP-A density is important for centromere identity and its essential role of kinetochore assembly.

(Ganter et al. 1999; Dlakic and Harrington 1996). In summary, the precise role of centromeric DNA is still unclear; however, the fact that there is centromeric DNA signatures indicates that these genomic regions are important for centromere function.

## 2.6 A role for CENP-A nucleosome composition in centromere propagation

CENP-A nucleosome composition has become a fascinating debate. Some pieces of evidence demonstrate that CENP-A is an octamer while another scenario proposes it is half an octamer, a hemisome, or a tetrasome (Camahort et al. 2009; Palmer et al. 1987; Sekulic et al. 2010; Dalal et al. 2007; Williams et al. 2009; Mizuguchi et al. 2007) and reviewed in (Black and Cleveland 2011; Henikoff and Furuyama 2012). Here, we specifically describe how CENP-A nucleosome composition could be dynamic through the cell cycle and how this feature can manifest in conserved centromere identity over repetitive cell divisions.

In 2007, two key papers showed that CENP-A incorporation occurs after mitotic exit and takes place from late anaphase to early G1 (Schuh et al. 2007; Jansen et al. 2007). Using live cell imaging coupled to quantitative super-resolution analysis, we showed that CENP-A loading in human cells takes the entire length of G1 (about 8 h) (Lagana et al. 2009). Interestingly, in plants, CENP-A seems to be loaded in G2 as shown in two examples: *A. thaliana* and *Hordeum vulgare* (Lermontova et al. 2007; Lermontova et al. 2011). Regardless of the precise timing of CENP-A incorporation, it universally occurs outside of S-phase leading to a centromere identity problem in the next cell cycle. After the passage of the replication fork, canonical nucleosomes are randomly distributed onto replicated daughter strands, leaving gaps that are filled with newly synthesized histone proteins (H3, H4, H2A, and H2B) (Smith and Stillman 1991). CENP-A nucleosomes should also be randomly distributed, but gaps will be left after the replication fork passage since newly synthesized CENP-A is not loaded

in S-phase, thus diluting the epigenetic mark. Karpen's group tackled this problem and discovered that histone H3.3, another non-replication coupled histone H3 variant, fills in the gaps until the next G1 phase where it is specifically replaced by CENP-A (Dunleavy et al. 2011). Thus, centromeric chromatin is thought to be stabilized and reinforced by a combination of H3 variants in S-phase. An alternate model proposed by Henikoff's group relies on evidence of CENP-A hemisomes (tetramers of CENP-A/H4/H2A/H2B) at centromeres, supported by atomic force microscopy (AFM) analysis showing centromeric nucleosomes having half the height of a canonical octameric nucleosome (Dalal et al. 2007). The model suggests that CENP-A nucleosomes are split in half on each daughter strands in S-phase, keeping the centromeric epigenetic mark and the size of the centromere locus for each cell division. This is an attractive model for centromere identity preservation, which does not include an intermediate centromeric composition. However, a hemisome model does not take into account the packing of DNA as hemisomes will wrap approximately half the length of DNA compared to octamers.

Even if the hemisome model is controversial, it raises the possibility that CENP-A nucleosome composition is dynamic through the cell cycle. Recently, two papers concluded that CENP-A chromatin dynamically switches between octamer and tetramer ($(CENP\text{-}A\text{-}H4)_2$) compositions in different phases of the cell cycle (Bui et al. 2012; Shivaraju et al. 2012). Dalal's group measured nucleosome height and volume using AFM and they demonstrated that CENP-A nucleosomes are tetramers in G1, convert into octamers in early S, and revert back into tetramers at the end of S-phase. Whereas the Gerton lab, using fluorescence correlation spectroscopy (FCS) coupled to calibrated imaging, observed that yeast centromeres have one copy of Cse4 during the majority of the cell cycle and two copies at anaphase B. There are caveats with these results however. AFM data are based on affinity purification of centromeric nucleosomes whose precise biochemical makeup is not clear. Thus, differences in height could be attributed to the presence of additional proteins or other artefacts generated in the purification scheme. FCS is a powerful technique that uses peak fluorescence intensity from a diffraction limited spot (less than 300 nm in this case) to

determine protein concentration and diffusion rates. However, yeast centromeres are often dispersed (especially in metaphase) and do not all occupy a diffraction limited spot, which would result in an underestimation of Cse4 number. During anaphase, yeast centromeres are highly compacted fitting nicely into a diffraction limited spot. This topographical difference would nicely explain the FCS results. Regardless, centromere identity and composition are necessarily dynamic because of genome replication. It will be of great interest to determine if these observations are born out by the test of time.

## 2.7 CENP-A loading onto centromeric chromatin is a three-step mechanism

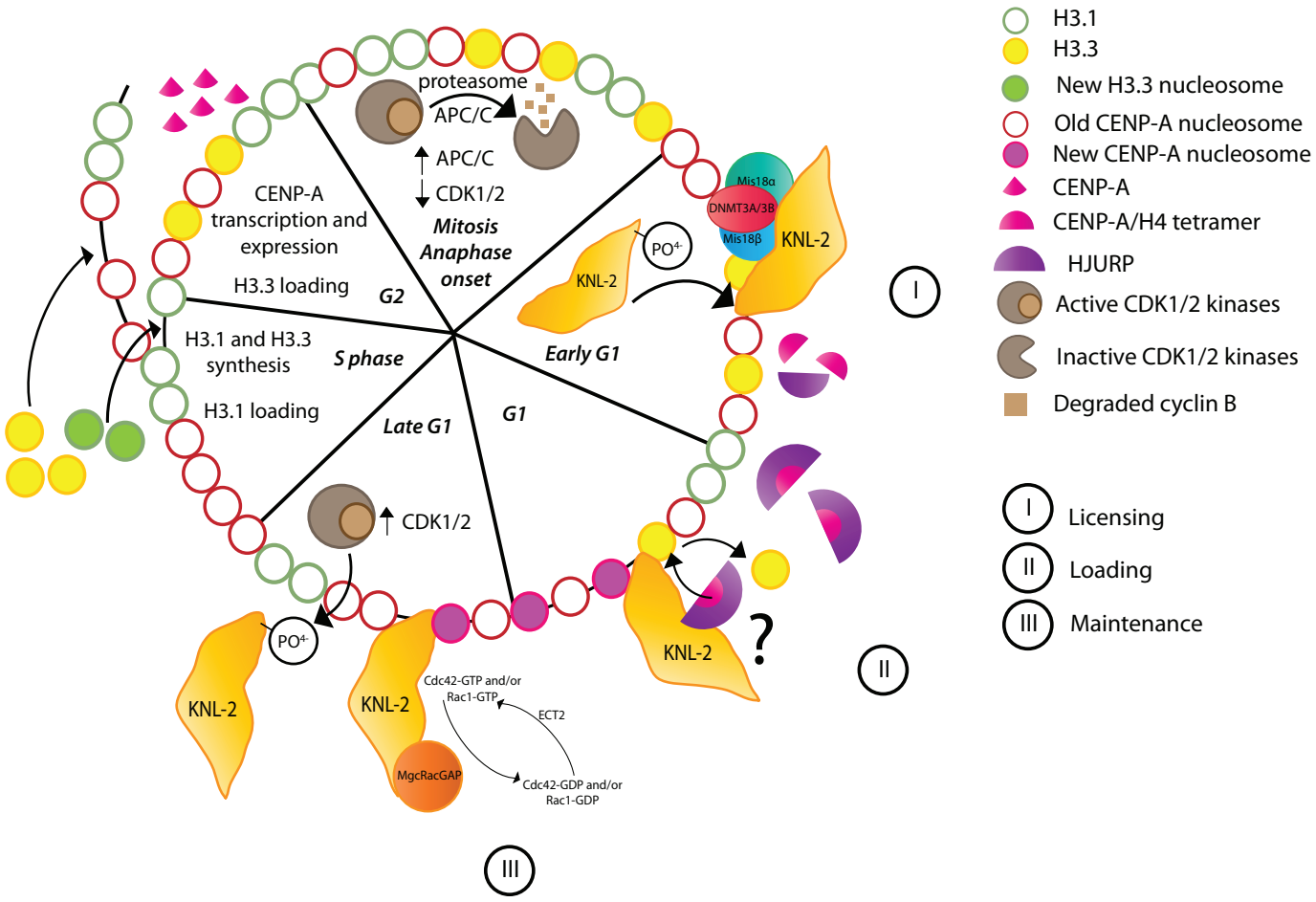
It has been clearly demonstrated in virtually every model system that nucleosome components (H3, H4, H2A, and H2B) are expressed in S-phase (Prescott 1966; Borun et al. 1975). Nucleosome assembly occurs in an ordered manner with the help of chaperone proteins, such as CAF-1 and Asf1, and the resultant octamers are formed after the replication fork (Smith and Stillman 1991, 1989; Hayashi et al. 2004). However, CENP-A, like other histone variants, has a distinct expression pattern. Human CENP-A mRNA peaks in G2 prior to mitosis, and its incorporation is restricted to G1 phase as discussed above (Shelby et al. 2000; Jansen et al. 2007; Schuh et al. 2007; Hemmerich et al. 2008). CENP-A incorporation in this window of the cell cycle is very interesting and raises many questions. One model wherein mitotic forces transmitted through kinetochore microtubule attachments somehow modify centromere structure was disproven by bypassing the spindle assembly checkpoint (O'Connell et al. 2008). Centromeres that never experienced forces (due to absence of microtubules) incorporated new CENP-A with normal G1 timing (Jansen et al. 2007; Schuh et al. 2007). Thus, the cellular mechanism propagating centromere identity in G1 is still largely obscure on a cell biological level.

On the molecular level, several studies have addressed the mechanism of CENP-A loading onto centromeric chromatin. Currently, a three-step mechanism loosely describes the process: (1) recognition and licensing of centromeres, (2) loading of newly synthesized CENP-A with the help of chaperone proteins, and (3) maintenance of newly incorporated CENP-A (Figure 2.3 A).

### 2.7.1 Recognition and licensing of centromeres

1. Recognition of centromeric chromatin for CENP-A loading only at centromeres is a complex question. In 2007, a licensing complex, also called the Mis18 complex, consisting of KNL-2 (M18BP1, hereafter referred to as KNL-2) and its partners Mis18α/β was shown to be required for CENP-A localization to centromeres in diverse model systems (Fujita et al. 2007; Maddox et al. 2007). Our understanding of the mechanism of the licensing complex is limited; however, it is clear that its recruitment to centromeres following anaphase is the most upstream event known for CENP-A deposition. It has been proposed that CENP-C recruits the licensing complex (Moree et al. 2011; Dambacher et al. 2012). This model is based on the observation that KNL-2 had reduced localization at centromeres when CENP-C was depleted in *Xenopus* egg extracts (Moree et al. 2011). However, KNL-2 localization is not fully lost and CENP-C is clearly downstream of KNL-2 in other model systems (Maddox et al. 2007; Fujita et al. 2007). Thus, the question of how the licensing factors recognize and bind centromeric chromatin with high specificity and subsequently recruit downstream components required for CENP-A loading needs more investigation.

A


proteasome
APC/C
APC/C
CDK1/2
Mitosis
Anaphase
onset
CENP-A
transcription and
expression
H3.3 loading
G2
H3.1 and H3.3
synthesis
S phase
H3.1 loading
Late G1
G1
Early G1
Mis18α
DNMT3A/3B
Mis18β
KNL-2
PO4-
KNL-2
CDK1/2
PO4-
KNL-2
KNL-2
Cdc42-GTP and/or
Rac1-GTP
ECT2
MgcRacGAP
Cdc42-GDP and/or
Rac1-GDP
KNL-2
?
I
II
III
H3.1
H3.3
New H3.3 nucleosome
Old CENP-A nucleosome
New CENP-A nucleosome
CENP-A
CENP-A/H4 tetramer
HJURP
Active CDK1/2 kinases
Inactive CDK1/2 kinases
Degraded cyclin B
I Licensing
II Loading
III Maintenance

B

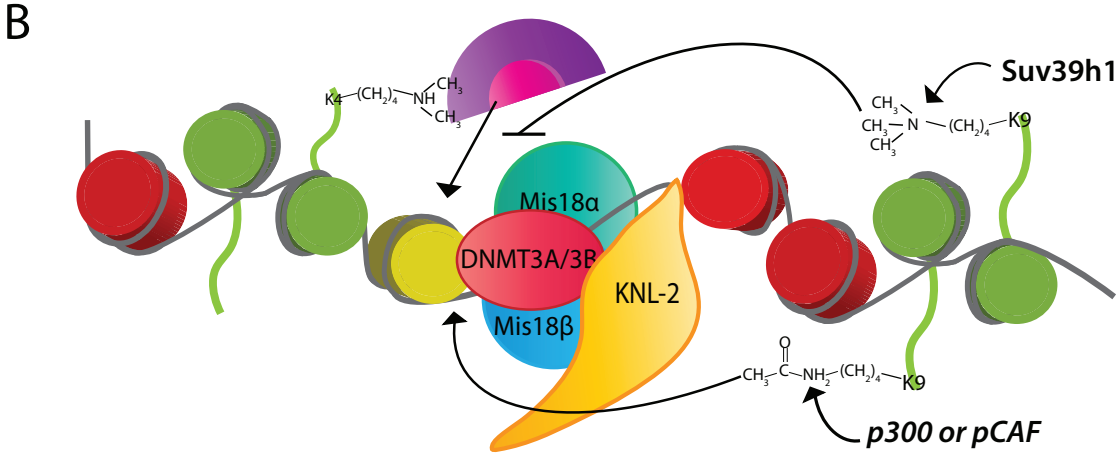

Suv39h1
K4
K9
Mis18α
DNMT3A/3B
Mis18β
KNL-2
K9
p300 or pCAF

**Figure 2.3 CENP-A incorporation is cell cycle regulated and depends on epigenetic marks.**
CENP-A incorporation is a replication independent mechanism; CENP-A-H4 tetramers are loaded in G1 instead of S-phase as for H3-H4 tetramers. The incorporation is a three-step mechanism: I- Licensing, II-Loading, and III-Maintenance. First, the licensing complex (KNL-2, Mis18α and β) recognizes and binds the centromeric chromatin. This will license the centromere for loading of newly synthesized CENP-A. Next, the licensing complex recruits, by an unknown mechanism, the CENP-A chaperone HJURP, which directly binds and stabilizes CENP-A/H4 complexes. Finally, when newly synthesized CENP-A is incorporated, a mark is removed or CENP-A conformation is changed in order to change the newly synthesized identity to that of an old one. Cdc-42 or Rac1 might be part of this cellular process through the action of the GAP MgcRacGAP and the GEF ECT-2. CENP-A loading to centromeres is regulated through the cell cycle by the CDK1/2 kinases, which phosphorylate KNL-2 in mitosis and block its localization to centromeres. At anaphase onset, CDK activity diminishes, KNL-2 is dephosphorylated and is able to localize to centromeres. a CENP-A incorporation to centromeres also depends on a post-translational modification of histone H3. H3K4 dimethylation is important to eventually HJURP, which localizes to centromeres and loads newly synthesized CENP-A. H3K9 trimethylation by the methyltransferase Suv39h1 inhibits CENP-A loading to centromeres, whereas H3K9 acetylation by the histone acetyl transferase (HAT) p300 or pCAF triggers CENP-A loading at centromeres. We hypothesize that KNL-2 to binds the linker DNA in the centromeric chromatin and, together with its partners Mis18α and Misβ, this protein complex acts as a licensing mark of centromeres for CENP-A loading. Also, a DNA methlyltransferase DNMT3A and DNMT3B interacts with Mis18α, and depletion of the protein leads to a decrease of other epigenetic marks such as H3K9me2, H3K9m3, and H3K4me2 at centromeric loci.

### 2.7.2 Loading of newly synthesized CENP-A with the help of chaperone proteins

2. There have been two factors identified that stabilize the CENP-A/H4 complex, RbAp46/48 (Mis16, a member of the CAF-1 complex) and HJURP (or SCM3 in yeast). These essential proteins were shown to be required to prevent the degradation of soluble CENP-A molecules and thus considered chaperones (Foltz et al. 2009; Dunleavy et al. 2009; Hayashi et al. 2004). As RbAp46/48 seems to be a nonspecific histone chaperone, research has focused on HJURP showing it to have nucleosome assembly activity specifically for newly synthesized CENP-A

(Dunleavy et al. 2009; Foltz et al. 2009; Barnhart et al. 2011). The CENP-A binding domain (CBP) in the N-terminus of HJURP recognizes the CATD, localized in the histone fold domain within the L1 and α2, of CENP-A (Shuaib et al. 2010). Further details of the nature of this interaction were uncovered by co-structural studies of HJURP and CENP-A/H4, identifying specific residues mediating direct binding (Hu et al. 2011b). Although it is well accepted that HJURP recognizes CENP-A through CATD binding, in 2012, the Black lab discovered another binding interface located in the N-terminus of CENP-A (Bassett et al. 2012). This distinct interface is not required for the specificity of the binding but rather for stabilization, a mechanism that results in changing the structure of the histone fold domains of both CENP-A and H4. H3 nucleosome assembly order is well described in the literature; a histone H3/H4 tetramer first sits on the DNA and then two H2A/H2B dimers complete the assembly (Smith and Stillman 1991). However, some biochemical and structural studies show that HJURP binds a dimer of CENP-A/H4 (Hu et al. 2011a). Thus, it is unclear which CENP-A/H4 complex (dimer or tetramer) gets incorporated in the centromeric chromatin. Regardless of the stoichiometry, HJURP localization to centromeres in G1 is dependent on the licensing complex; however, direct binding of the licensing complex components and HJURP has not been observed (Barnhart et al. 2011; Lagana et al. 2009; Fujita et al. 2007).

### 2.7.3 Maintenance of newly incorporated CENP-A

3. Immunoprecipitation of KNL-2 allowed our group to identify a new protein involved in centromere identity, MgcRacGAP (Lagana et al. 2009). This GTPase activating protein was previously described as part of the centralspindlin complex with a role in cytokinesis via regulation of Rho-type small G-proteins (Mishima et al. 2002; Canman et al. 2008). Nevertheless, two independent labs had revealed a possible role for MgcRacGAP in centromere function (Izuta et al. 2006; Perpelescu et al. 2009). In our study, we showed that MgcRacGAP localizes to centromeres in late G1 for a brief (1 to 2 h) window following CENP-A loading. Depletion of

MgcRacGAP resulted in the loss of newly incorporated CENP-A nucleosomes. Interestingly, a GAP dead mutant expressed in HeLa cells showed persistent localization to centromeres, and depletion of ECT-2 (the partner GTP exchange factor protein) recapitulated MgcRacGAP depletion. Thus a small GTPase cycle is likely required for maintaining CENP-A at centromeres. By depleting known small GTPases, we identified Cdc42 and Rac1 as possible targets of MgcRacGAP and ECT-2, and localization studies led us to favor Cdc42 as the relevant small GTPase. Since the localization of MgcRacGAP to centromeres is very late in G1 and depletion resulted in loss of centromere stability, we hypothesized that MgcRacGAP maintains the newly incorporated CENP-A and prevents overincorporation of CENP-A. This maintenance mechanism undoubtedly requires downstream effectors that are as yet unidentified.

## 2.8 Post-translational modifications of newly synthesized CENP-A

The importance of histone post-translational modifications has become clear in genome regulation (Figure 2.3B). Some modifications affect gene expression through activation or repression of transcription by changing chromatin compaction and state. In centromeres, histone H3 lysine 4 dimethylation (H3K4me2), a marker of transcriptional activation, was shown to be intersperse with CENP-A nucleosomes on elongated Drosophila and human chromatin (Sullivan and Karpen 2004). This post-translational modification was shown to be important for centromere regulation, as increased activity of the demethylase LSD1 at human artificial chromosome (HAC) centromere decreases HJURP localization (Bergmann et al. 2011). Furthermore, loss of H3K4me2 prevents loading of newly synthesized CENP-A to this alphoid DNA, revealing an important role of this post-translational modification for CENP-A localization to centromere. In the same study, a loss of H3K4me2 decreased the transcription of alphoid DNA; however, this was not clearly demonstrated to have a direct role with HJURP localization and CENP-A loading. As post-translational modifications such as H3K4me2 affect chromatin compaction state, the physical

topology of centromeric chromatin is likely to be critical for CENP-A localization. Additionally, Masumoto's group demonstrated that histone H3 lysine 9 trimethylation (H3K9me3) prevents de novo CENP-A assembly on HAC alphoid DNA by tethering Suv39h1 (a methyltransferase) at this specific locus in mammalian cells (Ohzeki et al. 2012). Also, after depleting Suv39h1, an increase in CENP-A at HAC alphoid DNA was observed. On the other hand, tethering of histone acetyltransferases p300 or pCAF increased the acetylation state of H3K9 and also increased CENP-A level at alphoid DNA. This modification is important only for de novo CENP-A assembly, as removal of acetyltransferases from cells did not affect preexisting centromeres over several cell divisions.

Interestingly, the balance between methylation and acetylation of H3K9 is linked generally to transcriptional activity indicating a possible link between with CENP-A localization to centromeres. In support of this hypothesis, the Desai lab demonstrated by genomic studies in *C. elegans* that CENP-A is incorporated in regions of low germline transcriptional activity (Figure 2.1) (Gassmann et al. 2012). One possible mechanism linking post-translational modification, transcription, and CENP-A localization to centromeres could be explained by the observations that RNA polymerase II together with its associated transcription factors are localized to centromeres and these active proteins increase α-satellite transcripts (Chan et al. 2011). Interestingly, inhibition of RNA polymerase II decreased CENP-C localization to centromeres; however, it is unclear if this is cause or effect.

DNA, as well as histone proteins, can be modified by methylation on cytosines of CpG islands. This state of the chromatin is usually linked to transcription repression due to chromatin compaction (Gros et al. 2012). Interestingly, a DNA methyltransferase enzyme called DNMT3A/3B has been shown to co-localize with Mis18α in mouse embryonic fibroblast (MEF) cells (Kim et al. 2012). Depletion of Mis18α protein leads to a decrease in centromere DNA methylation as well as a decrease of DNMT3A localization to centromeres. Moreover, this depletion leads to a decrease of some post-translational marks on histone H3 such as H3K9me2,

H3K9m3, and H3K4me2 at centromeric locus of chromosomes. However, this study does not demonstrate a direct link between those modifications, transcription of α-satellite DNA, and CENP-A localization to centromeres. All together, there is clear evidence that post-translational modifications of centromeric chromatin affect CENP-A loading. This field of research is blooming and we expect great advances in the near future providing a better understanding of those modifications in centromere identity.

## 2.9 CENP-A loading to centromere is regulated by cell cycle kinase

The timing of CENP-A loading invokes clear hypotheses of a cell cycle-dependent mechanism. Recently, the Jansen lab nicely demonstrated that cyclin-dependent kinases (CDKs) temporally regulate CENP-A loading to centromeres (Silva et al. 2012). Synchronized human cells treated with roscovitine, a CDK1 and 2 inhibitor, showed apparently normal, however mis-timed, CENP-A assembly in G2. More precisely, they demonstrated a role for CDK1 as the kinase regulating CENP-A loading. Using DT40 avian cell line, which is genetically null for CDK2, and inhibiting CDK1 function chemically, the authors observed an inappropriate CENP-A loading in G2, compared to control DT40 cells. Also, this regulation seems to be at the level of KNL-2, as expression of a mutant form with 24 potential phosphorylation sites changed to alanine also resulted in precocious localization of KNL-2 and downstream CENP-A loading in otherwise untreated G2 cells. This regulation through CDK activity is only true for KNL-2 and did not have any influences on the other licensing complex components. Thus, CENP-A loading is inversely timed relative to mitotic CDK activity and KNL-2 seems to be the most upstream component of the centromere epigenetic regulation pathway (Figure 2.3A). It is not clear if these rules will hold true for plants where CENP-A loading is in G2, however some cell cycle timing mechanism must exist in these models also.

## *2.*10 Is the ultimate mark for centromere identity CENP-A?

The epigenetic mechanism for centromere identity has been well defined over the years. However, only recently was it shown that CENP-A is sufficient to drive centromeric identity over multiple cell divisions (Figure 2.4). This was accomplished by generating a high-density region of CENP-A chromatin; it is needed because previous overexpression experiments resulted in CENP-A incorporation in the chromosome arms, however no kinetochore assembly. Briefly, many repeats of the lac operon were integrated in series at a non-centromeric (region on a chromosome arm) locus in flies (Mendiburo et al. 2011). Overexpression of LacI fused to CID (CENP-A) resulted in CID incorporation ectopically, and this ectopic centromere recruited kinetochore components. Remarkably, these ectopic centromeres can, with extremely low fidelity, functionally replace the endogenous centromeres for a short period of time. Similar studies fusing HJURP to LacI in human cells also generated localized high-density arrays of CENP-A with similar results. Interestingly, the lac operon system bypassed the need of the licensing complex for HJURP, since this technique forces direct interactions of the protein with a specific DNA sequence (Barnhart et al. 2011) (Figure 2.2). This is a strong argument in favor of the licensing complex being upstream of HJURP for CENP-A loading to centromeres given that HJURP, in absence of KNL-2 and Mis18, does not localize to normal centromeres. Therefore, high-density CENP-A chromatin is not only necessary but is sufficient for centromere identity and function.

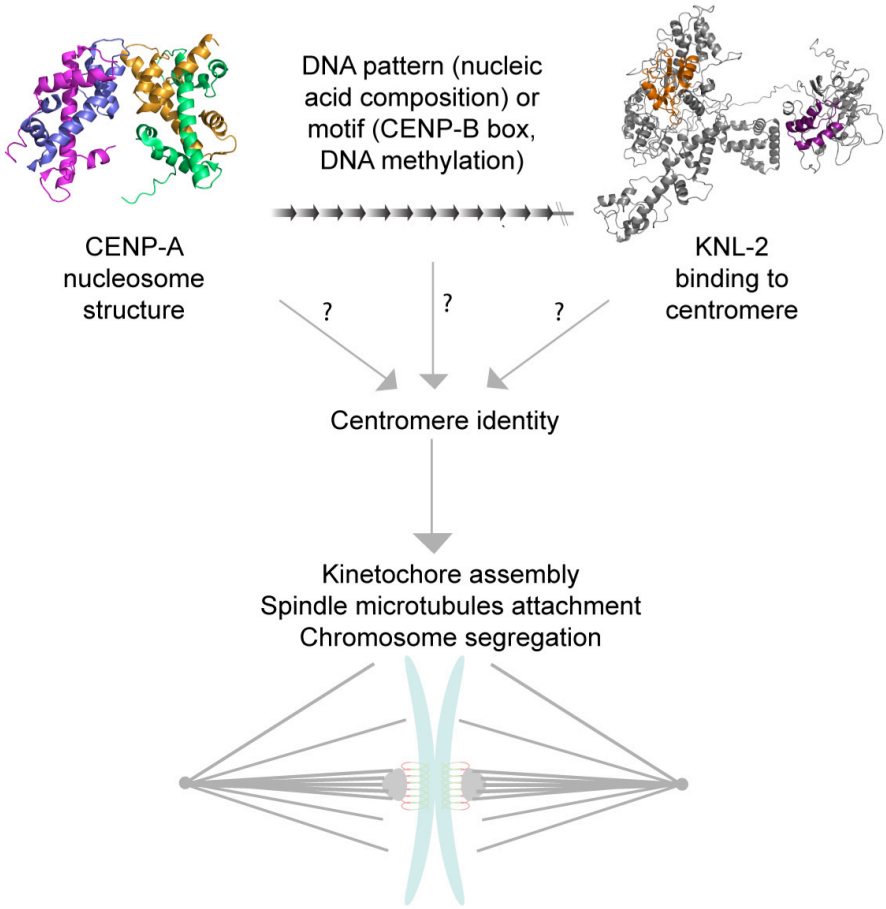



**Figure 2.4 What is the first epigenetic mark of centromeres?** Information provided by the literature brings the question of what is the first epigenetic mark of centromeres. Some evidence has shown that CENP-A nucleosome structure might be the feature for centromere identity. However, CENP-A incorporation and localization to centromeres are not passive cellular processes; therefore, CENP-A might not be the first epigenetic mark. Centromere identity could be through genetic features of DNA sequence (nucleic acid composition) or motif (CENP-B box, DNA methylation), but strong experimental evidence disputes this model. One marker that could link the two other hypotheses is KNL-2. This protein was shown to be essential for CENP-A localization to centromeres and genetically, it is the most upstream protein for this important cellular process. Thus, KNL-2 could be the first epigenetic mark of centromeres. For KNL-2 predicted structure: *Orange* is the SANTA domain and in *purple* is the Myb domain. (Published CENP-A nucleosome structure, Protein Data Bank reference# 3AN2 (Tachiwana et al. 2012); KNL-2 predicted structure (Zhang 2008))

In normal cells, the licensing complex is the first known step in recognizing centromeric. Interestingly, in *C. elegans* it has been observed for quite some time that exogenous DNA injected into oocytes forms stably transmitted arrays that are properly segregated over many generations. Accordingly, the Desai lab reasoned that if this DNA is segregated, then it should form a functional kinetochore (Yuen et al. 2011). To test this hypothesis, they showed that arrays segregate passively until early

embryogenesis where at the five to eight cell stages, lacO extrachromosomal arrays recruit centromeric proteins such as CENP-A, kinetochore proteins BUB-1 and NDC-80, and the licensing factor KNL-2, actively segregating the array. This interesting approach demonstrated that random DNA sequences are competent for de novo centromere formation; however, the timing and the sequence of events are not defined. It should be noted that this could be specific to *C. elegans* given the fact that alphoid DNA has been demonstrated to be critical for de novo centromere formation in human cells (Ohzeki et al. 2002).

Emerging evidence is expanding our knowledge on the epigenetic mechanism of centromere identity but still raises the question: what is the first mark for centromere identity? We propose two broad possibilities: (1) CENP-A structure and the surrounding post-translational modifications confers to centromere a specific docking site for the licensing complex to bind, loading newly synthesized CENP-A via the chaperone HJURP, (2) centromeric DNA, composed of motifs like the CENP-B box is directly recognized by the licensing complex, or centromeric DNA composition in nucleic acids confers a centromere-specific chromatin state. In either case, we propose that licensing protein KNL-2 through its Myb domain binds the centromeric DNA and recruits the loading of new CENP-A during G1 in human cells (Figure 2.4).

## 2.11 The role of CENP-A in mitosis and its relation with cancer therapy

Many labs working in a myriad of model systems have concluded that centromere identity propagation and maintenance through CENP-A is essential for cell division. CENP-A localization to centromeres creates a platform that is essential for kinetochore assembly; a loss in centromere identity results in chromosome segregation defects caused by a misattachment of chromosomes to the mitotic spindle. To date, centromere proteins such as CENP-A and KNL-2 have no known role outside of mitosis, thus making them appealing targets for chemotherapeutic

development. Inhibiting centromere function leads to chromosome segregation defects and ultimately to cell death, thus possibly a potent mechanism for slowing down cancer progression. Importantly, inhibiting mitosis-specific mechanisms will diminish the possibility of undesirable side effects as seen in taxol treatment. This broadly used chemotherapeutic treatment stabilizes microtubules leading to mitotic defects and cell death. However, microtubules also have essential roles as cytoskeleton components for most cell types, including neurons. Therefore, taxol also affects microtubules in the nervous system resulting in severe undesired side effects usually accompanied by neurodegenerative pathology. Thus, investigations with the aim of elucidating how the cell manages to preserve its centromeres well defined during cell division will likely provide us with new drug targets with higher specificity for mitotic cells.

Centromere fields of research are expanding, trying to better understand the complex cellular process of centromere identity and maintenance. In the next years, our efforts will bring more highlights to the spatial and temporal regulation of CENP-A loading to centromere, thus better define the first marks of centromere identity.

## 2.12 Acknowledgements

We would like to thank our lab members, especially Joël Ryan for his precious help, Dr. Amy S. Maddox and her lab members for discussions and innovative ideas. Also, we are grateful to the centromere/kinetochore community at large for sharing ideas and exciting results. P.S.M. is the Canada Research Chair in Cell Division and Chromosomal Organization and supported by research grants from the Canadian Institutes of Health Research (MOP-106548) and Canadian Cancer Society Research Institute (700824).

## 2.13 References


Barnhart MC, Kuich PH, Stellfox ME, Ward JA, Bassett EA, Black BE, Foltz DR (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol 194 (2):229-243. doi:jcb.201012017 [pii]10.1083/jcb.201012017

Bassett EA, DeNizio J, Barnhart-Dailey MC, Panchenko T, Sekulic N, Rogers DJ, Foltz DR, Black BE (2012) HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. Dev Cell 22 (4):749-762. doi:S1534-5807(12)00059-7 [pii]10.1016/j.devcel.2012.02.001

Bassett EA, Wood S, Salimian KJ, Ajith S, Foltz DR, Black BE (2010) Epigenetic centromere specification directs aurora B accumulation but is insufficient to efficiently correct mitotic errors. J Cell Biol 190 (2):177-185. doi:10.1083/jcb.201001035

Bergmann JH, Rodriguez MG, Martins NM, Kimura H, Kelly DA, Masumoto H, Larionov V, Jansen LE, Earnshaw WC (2011) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J 30 (2):328-340. doi:emboj2010329 [pii]10.1038/emboj.2010.329

Black BE, Brock MA, Bedard S, Woods VL, Jr., Cleveland DW (2007a) An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Proc Natl Acad Sci U S A 104 (12):5008-5013. doi:0700390104 [pii]10.1073/pnas.0700390104

Black BE, Cleveland DW (2011) Epigenetic centromere propagation and the nature of CENP-A nucleosomes. Cell 144 (4):471-479. doi:S0092-8674(11)00116-4 [pii]10.1016/j.cell.2011.02.002

Black BE, Foltz DR, Chakravarthy S, Luger K, Woods VL, Jr., Cleveland DW (2004) Structural determinants for generating centromeric chromatin. Nature 430 (6999):578-582. doi:10.1038/nature02766nature02766 [pii]

Black BE, Jansen LE, Maddox PS, Foltz DR, Desai AB, Shah JV, Cleveland DW (2007b) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell 25 (2):309-322. doi:S1097-2765(06)00886-0 [pii]10.1016/j.molcel.2006.12.018

Blower MD, Karpen GH (2001) The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat Cell Biol 3 (8):730-739. doi:10.1038/3508704535087045 [pii]

Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. Dev Cell 2 (3):319-330

Borun TW, Gabrielli F, Ajiro K, Zweidler A, Baglioni C (1975) Further evidence of transcriptional and translational control of histone messenger RNA during the HeLa S3 cycle. Cell 4 (1):59-67. doi:0092-8674(75)90134-8 [pii]

Brinkley BR, Stubblefield E (1966) The fine structure of the kinetochore of a mammalian cell in vitro. Chromosoma 19 (1):28-43

Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S (1999) A histone-H3-like protein in *C. elegans*. Nature 401 (6753):547-548. doi:10.1038/44062

Bui M, Dimitriadis EK, Hoischen C, An E, Quenet D, Giebe S, Nita-Lazar A, Diekmann S, Dalal Y (2012) Cell-cycle-dependent structural transitions in the human CENP-A nucleosome In Vivo. Cell 150 (2):317-326. doi:10.1016/j.cell.2012.05.035

Camahort R, Shivaraju M, Mattingly M, Li B, Nakanishi S, Zhu D, Shilatifard A, Workman JL, Gerton JL (2009) Cse4 is part of an octameric nucleosome in budding yeast. Mol Cell 35 (6):794-805. doi:10.1016/j.molcel.2009.07.022

Canman JC, Lewellyn L, Laband K, Smerdon SJ, Desai A, Bowerman B, Oegema K (2008) Inhibition of Rac by the GAP activity of centralspindlin is essential for cytokinesis. Science 322 (5907):1543-1546. doi:322/5907/1543 [pii]10.1126/science.1163086

Chan FL, Marshall OJ, Saffery R, Kim BW, Earle E, Choo KH, Wong LH (2011) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A 109 (6):1979-1984. doi:1108705109 [pii]10.1073/pnas.1108705109

Clarke L, Amstutz H, Fishel B, Carbon J (1986) Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci U S A 83 (21):8253-8257

Cottarel G, Shero JH, Hieter P, Hegemann JH (1989) A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in Saccharomyces cerevisiae. Mol Cell Biol 9 (8):3342-3349

Dalal Y, Wang H, Lindsay S, Henikoff S (2007) Tetrameric structure of centromeric nucleosomes in interphase *Drosophila* cells. PLoS Biol 5 (8):e218. doi:07-PLBI-RA-1092 [pii]10.1371/journal.pbio.0050218

Dambacher S, Deng W, Hahn M, Sadic D, Frohlich J, Nuber A, Hoischen C, Diekmann S, Leonhardt H, Schotta G (2012) CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. Nucleus 3 (1). doi:18955 [pii]

Dlakic M, Harrington RE (1996) The effects of sequence context on DNA curvature. Proc Natl Acad Sci U S A 93 (9):3847-3852

Dunleavy EM, Almouzni G, Karpen GH (2011) H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase. Nucleus 2 (2):146-157. doi:10.4161/nucl.2.2.152111949-1034-2-2-10 [pii]

Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, Daigo Y, Nakatani Y, Almouzni-Pettinotti G (2009) HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 137 (3):485-497. doi:S0092-8674(09)00254-2 [pii]10.1016/j.cell.2009.02.040

Earnshaw W, Bordwell B, Marino C, Rothfield N (1986) Three human chromosomal autoantigens are recognized by sera from patients with anti-centromere antibodies. J Clin Invest 77 (2):426-430. doi:10.1172/JCI112320

Earnshaw WC, Ratrie H, 3rd, Stetten G (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma 98 (1):1-12

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

Fishel B, Amstutz H, Baum M, Carbon J, Clarke L (1988) Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Mol Cell Biol 8 (2):754-763

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

Foltz DR, Jansen LE, Bailey AO, Yates JR, 3rd, Bassett EA, Wood S, Black BE, Cleveland DW (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 137 (3):472-484. doi:S0092-8674(09)00253-0 [pii]10.1016/j.cell.2009.02.039

Fujita Y, Hayashi T, Kiyomitsu T, Toyoda Y, Kokubu A, Obuse C, Yanagida M (2007) Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell 12 (1):17-30. doi:S1534-5807(06)00507-7 [pii]10.1016/j.devcel.2006.11.002

Furuyama S, Biggins S (2007) Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc Natl Acad Sci U S A 104 (37):14706-14711. doi:0706985104 [pii]10.1073/pnas.0706985104

Furuyama T, Henikoff S (2009) Centromeric nucleosomes induce positive DNA supercoils. Cell 138 (1):104-113. doi:S0092-8674(09)00510-8 [pii]10.1016/j.cell.2009.04.049

Ganter B, Chao ST, Lipsick JS (1999) Transcriptional activation by the myb proteins requires a specific local promoter structure. FEBS letters 460 (3):401-410

Gassmann R, Rechtsteiner A, Yuen KW, Muroyama A, Egelhofer T, Gaydos L, Barron F, Maddox P, Essex A, Monen J, Ercan S, Lieb JD, Oegema K, Strome S, Desai A (2012) An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. Nature 484 (7395):534-537. doi:nature10973 [pii]10.1038/nature10973

Gros C, Fahy J, Halby L, Dufau I, Erdmann A, Gregoire JM, Ausseil F, Vispe S, Arimondo PB (2012) DNA methylation inhibitors in cancer: recent and future approaches. Biochimie. doi:10.1016/j.biochi.2012.07.025

Haaf T, Warburton PE, Willard HF (1992) Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell 70 (4):681-696. doi:0092-8674(92)90436-G [pii]

Hayashi T, Fujita Y, Iwasaki O, Adachi Y, Takahashi K, Yanagida M (2004) Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell 118 (6):715-729. doi:10.1016/j.cell.2004.09.002S0092867404008323 [pii]

Hemmerich P, Weidtkamp-Peters S, Hoischen C, Schmiedeberg L, Erliandri I, Diekmann S (2008) Dynamics of inner kinetochore assembly and maintenance in living cells. J Cell Biol 180 (6):1101-1114. doi:10.1083/jcb.200710052

Henikoff S, Furuyama T (2012) The unconventional structure of centromeric nucleosomes. Chromosoma 121 (4):341-352. doi:10.1007/s00412-012-0372-y

Hu H, Liu Y, Wang M, Fang J, Huang H, Yang N, Li Y, Wang J, Yao X, Shi Y, Li G, Xu RM (2011a) Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. Genes Dev 25 (9):901-906. doi:gad.2045111 [pii]10.1101/gad.2045111

Hu H, Liu Y, Wang M, Fang J, Huang H, Yang N, Li Y, Wang J, Yao X, Shi Y, Li G, Xu RM (2011b) Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. Genes Dev. doi:gad.2045111 [pii]10.1101/gad.2045111

Ikeno M, Masumoto H, Okazaki T (1994) Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. Hum Mol Genet 3 (8):1245-1257

Izuta H, Ikeno M, Suzuki N, Tomonaga T, Nozaki N, Obuse C, Kisu Y, Goshima N, Nomura F, Nomura N, Yoda K (2006) Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. Genes Cells 11 (6):673-684. doi:GTC969 [pii]10.1111/j.1365-2443.2006.00969.x

Jansen LE, Black BE, Foltz DR, Cleveland DW (2007) Propagation of centromeric chromatin requires exit from mitosis. J Cell Biol 176 (6):795-805. doi:jcb.200701066 [pii]10.1083/jcb.200701066

Kim IS, Lee M, Park KC, Jeon Y, Park JH, Hwang EJ, Jeon TI, Ko S, Lee H, Baek SH, Kim KI (2012) Roles of Mis18alpha in epigenetic regulation of centromeric chromatin and CENP-A loading. Mol Cell 46 (3):260-273. doi:S1097-2765(12)00228-6 [pii]10.1016/j.molcel.2012.03.021

Lagana A, Dorn JF, De Rop V, Ladouceur AM, Maddox AS, Maddox PS (2009) A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat Cell Biol 12 (12):1186-1193. doi:ncb2129 [pii]10.1038/ncb2129

Lawrimore J, Bloom KS, Salmon ED (2011) Point centromeres contain more than a single centromere-specific Cse4 (CENP-A) nucleosome. J Cell Biol 195 (4):573-582. doi:10.1083/jcb.201106036

Lermontova I, Fuchs J, Schubert V, Schubert I (2007) Loading time of the centromeric histone H3 variant differs between plants and animals. Chromosoma 116 (6):507-510. doi:10.1007/s00412-007-0122-8

Lermontova I, Rutten T, Schubert I (2011) Deposition, turnover, and release of CENH3 at *Arabidopsis* centromeres. Chromosoma 120 (6):633-640. doi:10.1007/s00412-011-0338-5

Maddox PS, Hyndman F, Monen J, Oegema K, Desai A (2007) Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol 176 (6):757-763. doi:jcb.200701065 [pii]10.1083/jcb.200701065

Maddox PS, Oegema K, Desai A, Cheeseman IM (2004) "Holo" er than thou: chromosome segregation and kinetochore function in *C. elegans.* Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology 12 (6):641-653. doi:10.1023/B:CHRO.0000036588.42225.2f

Manuelidis L (1978) Chromosomal localization of complex and simple repeated human DNAs. Chromosoma 66 (1):23-32

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol 109 (5):1963-1973

Melters DP, Paliulis LV, Korf IF, Chan SW (2012) Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome research : an international journal on the molecular,

supramolecular and evolutionary aspects of chromosome biology 20 (5):579-593. doi:10.1007/s10577-012-9292-1

Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P (2011) *Drosophila* CENH3 is sufficient for centromere formation. Science 334 (6056):686-690. doi:334/6056/686 [pii]10.1126/science.1206880

Mishima M, Kaitna S, Glotzer M (2002) Central spindle assembly and cytokinesis require a kinesin-like protein/RhoGAP complex with microtubule bundling activity. Dev Cell 2 (1):41-54. doi:S1534580701001101 [pii]

Mitchell AR, Gosden JR, Miller DA (1985) A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma 92 (5):369-377

Mizuguchi G, Xiao H, Wisniewski J, Smith MM, Wu C (2007) Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. Cell 129 (6):1153-1164. doi:10.1016/j.cell.2007.04.026

Moree B, Meyer CB, Fuller CJ, Straight AF (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J Cell Biol 194 (6):855-871. doi:jcb.201106079 [pii]10.1083/jcb.201106079

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J Cell Biol 116 (3):585-596

Murphy TD, Karpen GH (1995) Localization of centromere function in a *Drosophila* minichromosome. Cell 82 (4):599-609. doi:0092-8674(95)90032-2 [pii]

O'Connell CB, Loncarek J, Hergert P, Kourtidis A, Conklin DS, Khodjakov A (2008) The spindle assembly checkpoint is satisfied in the absence of interkinetochore tension during mitosis with unreplicated genomes. J Cell Biol 183 (1):29-36. doi:jcb.200801038 [pii]10.1083/jcb.200801038

Ohzeki J, Bergmann JH, Kouprina N, Noskov VN, Nakano M, Kimura H, Earnshaw WC, Larionov V, Masumoto H (2012) Breaking the HAC Barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly. EMBO J 31 (10):2391-2402. doi:emboj201282 [pii]10.1038/emboj.2012.82

Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol 159 (5):765-775. doi:10.1083/jcb.200207112

Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A 88 (9):3734-3738

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kDa centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104 (4):805-815

Panchenko T, Sorensen TC, Woodcock CL, Kan ZY, Wood S, Resch MG, Luger K, Englander SW, Hansen JC, Black BE (2011) Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. Proc Natl Acad Sci U S A 108 (40):16588-16593. doi:1113621108 [pii]10.1073/pnas.1113621108

Perpelescu M, Nozaki N, Obuse C, Yang H, Yoda K (2009) Active establishment of centromeric CENP-A chromatin by RSF complex. J Cell Biol 185 (3):397-407. doi:jcb.200903088 [pii]10.1083/jcb.200903088

Prescott DM (1966) The syntheses of total macronuclear protein, histone, and DNA during the cell cycle in *Euplotes eurystomus*. J Cell Biol 31 (1):1-9

Robbins E, Gonatas NK (1964) The ultrastructure of a mammalian cell during the mitotic cycle. J Cell Biol 21:429-463

Schuh M, Lehner CF, Heidmann S (2007) Incorporation of *Drosophila* CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol 17 (3):237-243. doi:S0960-9822(06)02569-3 [pii]10.1016/j.cub.2006.11.051

Sekulic N, Bassett EA, Rogers DJ, Black BE (2010) The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. Nature 467 (7313):347-351. doi:nature09323 [pii]10.1038/nature09323

Shelby RD, Monier K, Sullivan KF (2000) Chromatin assembly at kinetochores is uncoupled from DNA replication. J Cell Biol 151 (5):1113-1118

Shivaraju M, Unruh JR, Slaughter BD, Mattingly M, Berman J, Gerton JL (2012) Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast. Cell 150 (2):304-316. doi:10.1016/j.cell.2012.05.034

Shuaib M, Ouararhni K, Dimitrov S, Hamiche A (2010) HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc Natl Acad Sci U S A 107 (4):1349-1354. doi:0913709107 [pii]10.1073/pnas.0913709107

Silva MC, Bodor DL, Stellfox ME, Martins NM, Hochegger H, Foltz DR, Jansen LE (2012) Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev Cell 22 (1):52-63. doi:S1534-5807(11)00466-7 [pii]10.1016/j.devcel.2011.10.014

Smith S, Stillman B (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. doi:0092-8674(89)90398-X [pii]

Smith S, Stillman B (1991) Stepwise assembly of chromatin during DNA replication in vitro. EMBO J 10 (4):971-980

Spencer F, Hieter P (1992) Centromere DNA mutations induce a mitotic delay in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 89 (19):8908-8912

Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M (1995) A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev 9 (5):573-586

Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nature structural & molecular biology 11 (11):1076-1083. doi:10.1038/nsmb845

Sullivan KF, Hechenberger M, Masri K (1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol 127 (3):581-592

Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional *Drosophila* centromere. Cell 91 (7):1007-1019. doi:S0092-8674(00)80491-2 [pii]

Tachiwana H, Kagawa W, Shiga T, Osakabe A, Miya Y, Saito K, Hayashi-Takanaka Y, Oda T, Sato M, Park SY, Kimura H, Kurumizaka H (2012) Crystal structure of the human centromeric nucleosome containing CENP-A. Nature 476 (7359):232-235. doi:nature10258 [pii]10.1038/nature10258

Takahashi K, Chen ES, Yanagida M (2000) Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science 288 (5474):2215-2219. doi:8597 [pii]

Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell 14 (5):1053-1066

Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, Muenke M, Zuffardi O, Jobling MA (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. Nat Genet 5 (4):368-375. doi:10.1038/ng1293-368

Valdivia MM, Brinkley BR (1985) Fractionation and initial characterization of the kinetochore from mammalian metaphase chromosomes. J Cell Biol 101 (3):1124-1134

Van Hooser AA, Ouspenski, II, Gregson HC, Starr DA, Yen TJ, Goldberg ML, Yokomori K, Earnshaw WC, Sullivan KF, Brinkley BR (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci 114 (Pt 19):3529-3542

Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG, Earnshaw WC (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol 7 (11):901-904. doi:S0960-9822(06)00382-4 [pii]

Willard HF (1985) Chromosome-specific organization of human alpha satellite DNA. Am J Hum Genet 37 (3):524-532

Williams JS, Hayashi T, Yanagida M, Russell P (2009) Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. Mol Cell 33 (3):287-298. doi:10.1016/j.molcel.2009.01.017

Yoda K, Ando S, Morishita S, Houmura K, Hashimoto K, Takeyasu K, Okazaki T (2000) Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc Natl Acad Sci U S A 97 (13):7266-7271. doi:10.1073/pnas.130189697130189697 [pii]

Yuen KW, Nabeshima K, Oegema K, Desai A (2011) Rapid de novo centromere formation occurs independently of heterochromatin protein 1 in C. elegans embryos. Curr Biol 21 (21):1800-1807. doi:S0960-9822(11)01018-9 [pii]10.1016/j.cub.2011.09.016

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40. doi:1471-2105-9-40 [pii]10.1186/1471-2105-9-40

Zlatanova J, Seebart C, Tomschik M (2008) The linker-protein network: control of nucleosomal DNA accessibility. Trends Biochem Sci 33 (6):247-253. doi:S0968-0004(08)00069-8 [pii]10.1016/j.tibs.2008.04.001

# CHAPTER 3

Publication

# A SMALL GTPASE MOLECULAR SWITCH REGULATES EPIGENETIC CENTROMERE MAINTENANCE BY STABILIZING NEWLY INCORPORATED CENP-A

Article published in Nature Cell Biology (2010) Dec; 12(12): 1186-93. PMID 21102442

## 3.1 Preface

Following the publication in The Journal of Cell Biology in 2007, demonstrating the essential role of KNL-2 for CENP-A localization to centromere, Dr. Paul S. Maddox wanted to find new KNL-2 binding partners. Through IP experiments followed by Mass Spectrometry analysis, our group identified over fifty potential candidates. We chose to investigate the role of MgcRacGAP since this protein did not have any identified function in CENP-A incorporation to centromere. The two co-first authors made the majority of experiments: Dr. Anaïck Lagana performed the shRNA depletion experiments, the qPCR analysis and the SNAP-tag experiments. Dr. Jonas F. Dorn performed the live cell experiments and fluorescence quantification using a program he wrote using Matlab. I contributed to this work by doing immunofluorescence experiments of small GTPases (Figure 3.6 A), the site directed mutagenesis of MgcRacGAP to make it a GAP-inactivated mutant, and the expression of this mutant in CENP-A-YFP cells (Figure 3.4 A, and Figure 3.S5). This contribution confirmed the role of Cdc42 as the small GTPase involved in CENP-A maintenance to centromere, and GAP inactivated mutant shows persistent localization to centromere in late G1 phase. Dr. Anaïck Lagana, Dr. Jonas F. Dorn, Anne-Marie, Dr. Paul S. Maddox and I conceived experiments. Dr. Paul S. Maddox, Dr. Amy S. Maddox and Dr. Jonas F. Dorn wrote the manuscript and they got feedbacks from Dr. Anaïck Lagana and I.

**A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A**

Anaïck Lagana*, Jonas F. Dorn*, Valérie De Rop[#], Anne-Marie Ladouceur, Amy S. Maddox[#&] and Paul S. Maddox[#%&]

Institute for Research in Immunology and Cancer (IRIC)
Dept of Pathology and Cell Biology
Université de Montréal
P.O. Box 6128, Station Centre-Ville
Montréal QC, H3C 3J7
CANADA

*These authors contributed equally to this work
[#]Systems biology option in the graduate program in Molecular Biology, Université de Montréal
[&]Department of Pathology and Cell Biology, Université de Montréal
[%]To whom correspondence should be addressed

Running title: Cdc42 is required for centromere maintenance

Word count:
1st paragraph: 158
Total: 2,995

## 3.2 Abstract

Epigenetic mechanisms regulate genome activation in diverse events including normal development and cancerous transformation. Centromeres are epigenetically designated chromosomal regions that maintain genomic stability by directing chromosome segregation during cell division. The histone H3 variant CENP-A resides specifically at centromeres, is fundamental to centromere function, and is thought to act as the epigenetic mark defining centromere loci. Mechanisms directing assembly of CENP-A nucleosomes have recently emerged, but how CENP-A is maintained after assembly is unknown. Here we show that a small GTPase switch functions to maintain newly assembled CENP-A nucleosomes. Using functional proteomics, we found that MgcRacGAP (a Rho family GTPase Activating Protein) interacts with the CENP-A licensing factor HsKNL2. High-resolution live-cell imaging assays, designed in this study, demonstrated that MgcRacGAP, the Rho family Guanine nucleotide Exchange Factor (GEF) Ect2, and the small GTPases Cdc42 and Rac1 are required for stability of newly incorporated CENP-A at centromeres. Thus, a small GTPase switch ensures epigenetic centromere maintenance after loading of new CENP-A.

## 3.3 Introduction

Epigenetic patterning is a dynamic process requiring maintenance of regulatory marks on chromatin across both cellular and organismal generations regardless of DNA sequence [1,2]. Centromere regions of chromosomes are highly diverse throughout phylogeny and are marked in most metazoan species by large stretches of imperfectly repeating DNA elements. To maintain genomic position, size and activity, centromeres are epigenetically marked by the conserved and essential histone H3 variant CENP-A [3-6]. Centromeres direct kinetochore assembly during mitosis and are thus required for chromosome segregation, making epigenetic maintenance of centromeric chromatin critical for genomic stability (reviewed in [7]).

During the normal cell cycle, CENP-A must be distributed to each daughter strand during DNA replication in S-phase, generating two equivalent sister centromeres, each possessing one half the number of CENP-A molecules present at the end of G1. In preparation for S-phase of the following cell cycle, new CENP-A is added during G1, doubling the number of CENP-A molecules at centromeres [8,9]. Thus, at the end of G1, but before S-phase, centromeres contain approximately 50% CENP-A from the previous cell cycle and 50% newly incorporated molecules. Incorporation of new CENP-A has been proposed to occur via a two-step process. First, the Myb domain protein HsKNL2/M18BP1 and its binding partner Mis18 'license', or mark, centromere regions, after which CENP-A is loaded through a chaperone protein (HJURP) coupled to the chromatin assembly factor (CAF1/RbAp46/48/Mis16) nucleosome assembly pathway (Fig 3.1A, [10-15]). Additionally, the chromatin remodelling complex protein RSF1 has an important, although as yet unidentified, role during G1, as loss of RSF1 function leads to decreased CENP-A levels at centromeres [16]. Thus, an understanding of centromere CENP-A replenishment is emerging [10-16]. However, whether specific mechanisms maintain centromere identity after CENP-A chromatin replenishment in G1 is currently unknown.

Here we employed functional proteomics and high-resolution imaging to identify a novel mechanism in centromere epigenetic regulation. This mechanism, controlled by a small G-protein molecular switch, is required to specifically retain newly incorporated CENP-A at the end of G1 [9,17]. Thus, newly incorporated CENP-A nucleosomes marking the centromere are molecularly distinct from those of the previous cell cycle.

## 3.4 Results

### 3.4.1 MgcRacGAP co-purifies with HsKNL2 and is required for CENP-A localization at centromeres

To identify proteins controlling CENP-A maintenance at centromeres, we used a functional proteomics approach. To restrict the analysis to G1 regulation of centromere chromatin, we purified interactors of HsKNL2, the G1-specific CENP-A licensing factor. Immunoprecipitations were conducted from micrococcal-nuclease-treated HeLa cell extracts using two different affinity-purified antibodies against HsKNL2 (targeted to distinct regions; Supplementary Information, Fig 3.S1, [11,12]). Mass-spectrometry analysis identified over 500 proteins in two or more HsKNL2 immunoprecipitations. Analysis of all samples (12 immunoprecipitations in sum) identified approximately 60 proteins unique to the HsKNL2 samples (absent from two independent control immunoprecipitations). HsKNL2, Mis18 and RbAp46/48 (CAF1) were found exclusively in anti-HsKNL2 purifications, confirming the specificity of the immunoprecipitations [11]. Nine of the remaining proteins were predicted to have nuclear localization, and thus be novel HsKNL2 interactors (Supplementary Information, Fig 3.S1).

One HsKNL2-interacting protein, MgcRacGAP, has been previously reported to co-purify with CENP-A chromatin [16,18] and was therefore chosen for further study. MgcRacGAP, a GAP for the Rho family of small GTPases, is part of the

centralspindlin complex together with MKLP1 (a kinesin-6 microtubule motor protein) that localizes to the spindle midzone after chromosome segregation [19,20,21]. Centralspindlin is thought to modulate actin dynamics required for cytokinesis through small GTPase regulation [22,23,24]. MgcRacGAP has been shown to exhibit GAP activity for the three Rho family small GTPases, RhoA, Rac1 and Cdc42 [20,25]. However, centralspindlin and Rho family small GTPases have not been implicated in regulation of centromere chromatin.

To test whether MgcRacGAP is required for centromere regulation, we used short hairpin RNA (shRNA)-mediated protein depletions in HeLa cells (48 hours treatment; Fig 3.1B-D and Supplementary Information, Table 3.S1). MgcRacGAP depletion resulted in increased binucleation as expected, owing to defects in cytokinesis [23,25,26] (Fig 3.1B, and Supplementary Information 3.S2). Visual inspection of CENP-A levels at centromeres revealed a clear reduction in, but not elimination of, CENP-A (as determined by either antibody staining or a yellow fluorescent protein (YFP)-tagged CENP-A fusion protein). Blocking cytokinesis independently of MgcRacGAP knockdown (by depleting RhoA or MKLP1, or by treating with actin depolymerizing drugs [27,28]) did not affect the level of CENP-A at centromeres. Thus, we conclude that the function of MgcRacGAP in centromere epigenetic regulation is independent from its role in cytokinesis.

### 3.4.2 MgcRacGAP function is independent of both the licensing and loading steps in centromere epigenetic regulation

Given that MgcRacGAP co-purified with HsKNL2, we hypothesized that MgcRacGAP functions in the licensing step of CENP-A replenishment in G1. To test this idea, we designed custom image-analysis software to specifically quantify CENP-A levels without human bias (Supplementary Information Fig 3.S3; see Methods). Using this software, we measured the level of CENP-A-YFP (confirmed by antibody staining) at

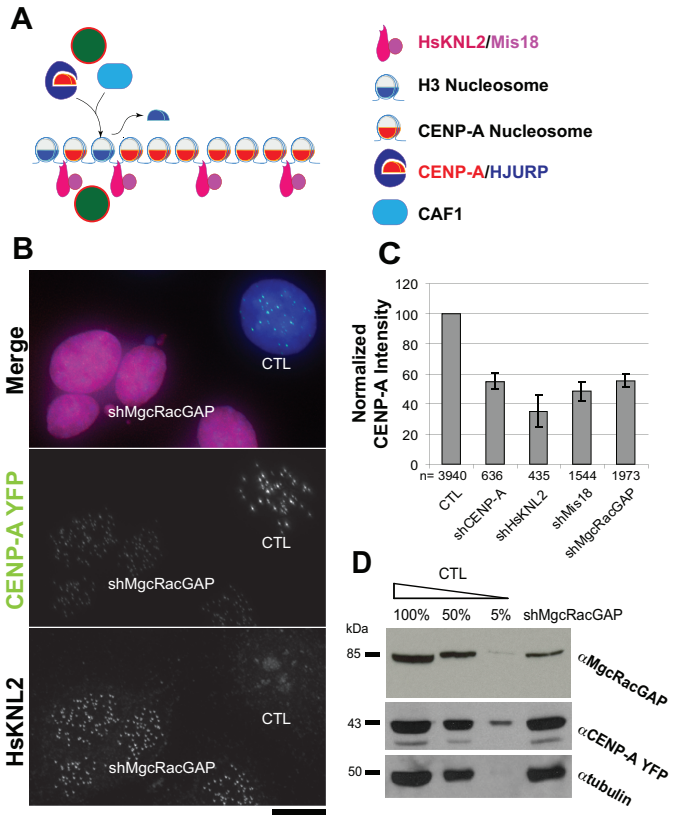



**Figure 3.1. MgcRacGAP is required for CENP-A protein localization to centromeres.**
(A) Model of the licensing (1) and loading (2) steps of CENP-A into centromere chromatin shows the hypothetical replacement of H3 nucleosomes with CENP-A (see text for details). (B) Representative immunofluorescence images of HeLa cells transfected with shRNA specific to MgcRacGAP and expressing CENP-A-YFP (middle) and treated with antibodies specific to HsKNL-2. Top image is merged from bottom two images. Depleted cells were identified by co-transfection with RFP-Histone H2B (red in merge); an untransfected control cell is also indicated. HsKNL-2 localization (bottom panel) is missing from the control because HsKNL-2 is normally lost after the end of G1. Thus, HsKNL-2 localization is normal in both cases. Scale bar, 10 µm. (C) Intensity of CENP-A-YFP at centromeres in cells expressing CENP-A-YFP and transfected with the indicated shRNA oligonucleotides, as assessed by high-resolution imaging of cells. Data are means +/- s.e.m., n; number of cells analysed. (D) Western blot of lysates from control cells transfected with MgcRacGAP shRNA (right). The indicated antibodies were used for blotting. All depletions were confirmed by qPCR (Supplementary Information, Table 3.S6).

centromeres in high-resolution images of large numbers of cells chosen at random from shRNA-treated (48 hours, see Methods) or control samples. MgcRacGAP depletion resulted in reduction of CENP-A levels at centromeres to a mean value of 45% of control (Fig 3.1C). The partial reduction in CENP-A levels in MgcRacGAP depleted cells is phenotypically distinct from the virtually complete loss of CENP-A localization after depletion of licensing factors (HsKNL2, Mis18; [11,12]). Interestingly,

HsKNL2 localization was unaltered by MgcRacGAP depletion (Fig 3.1B). These results indicate that MgcRacGAP does not function in CENP-A licensing through HsKNL2/Mis18 for centromere epigenetic regulation and possibly acts in the loading step (step 2, Fig 3.1A) of CENP-A replenishment in G1.

A hallmark of defective CENP-A loading (for example, following depletion of HJURP or CAF1; as opposed to licensing, Fig 3.1A) is CENP-A protein destabilization. Specifically, CENP-A protein levels have been shown to decrease to 50% of control (measured by western blot) probably because of the degradation of unincorporated, non-chromatin CENP-A protein [10,13,14,29]. If MgcRacGAP functioned in the CENP-A loading pathway, we would expect that CENP-A protein levels should be destabilized following MgcRacGAP depletion. Western blot analysis demonstrated that CENP-A protein levels (either YFP-transgene or endogenous CENP-A, Fig 3.1D and Supplementary Information, Figs 3.S4, respectively) were not altered by MgcRacGAP depletion. We therefore hypothesized that MgcRacGAP functions in a mechanism distinct from CENP-A replenishment to maintain CENP-A levels at centromeres.

Antibody staining revealed that MgcRacGAP localized to centromeres in interphase. However, early telophase cells (when licensing and loading factors both begin to localize to centromeres) did not show MgcRacGAP centromere localization (confirmed by expression of MgcRacGAP-mCherry fusion, Fig 3.2A, B). To define the step in which MgcRacGAP functions to maintain centromere identity, we developed a live-cell assay using high-resolution time-lapse imaging of cells expressing fluorescently tagged CENP-A (tagged with YFP [30]) and MgcRacGAP (tagged with mCherry; Fig 3.2B). CENP-A loading began immediately (within 10 min) after anaphase and continued for approximately 10hr, a period that coincides with the temporal localization pattern of both licensing and loading factors involved with replenishing CENP-A at individual centromeres. After the level of CENP-A doubled (approximately 10 h after anaphase, late G1), MgcRacGAP was detected in nuclear foci that co-localized with CENP-A. Localization of MgcRacGAP to individual

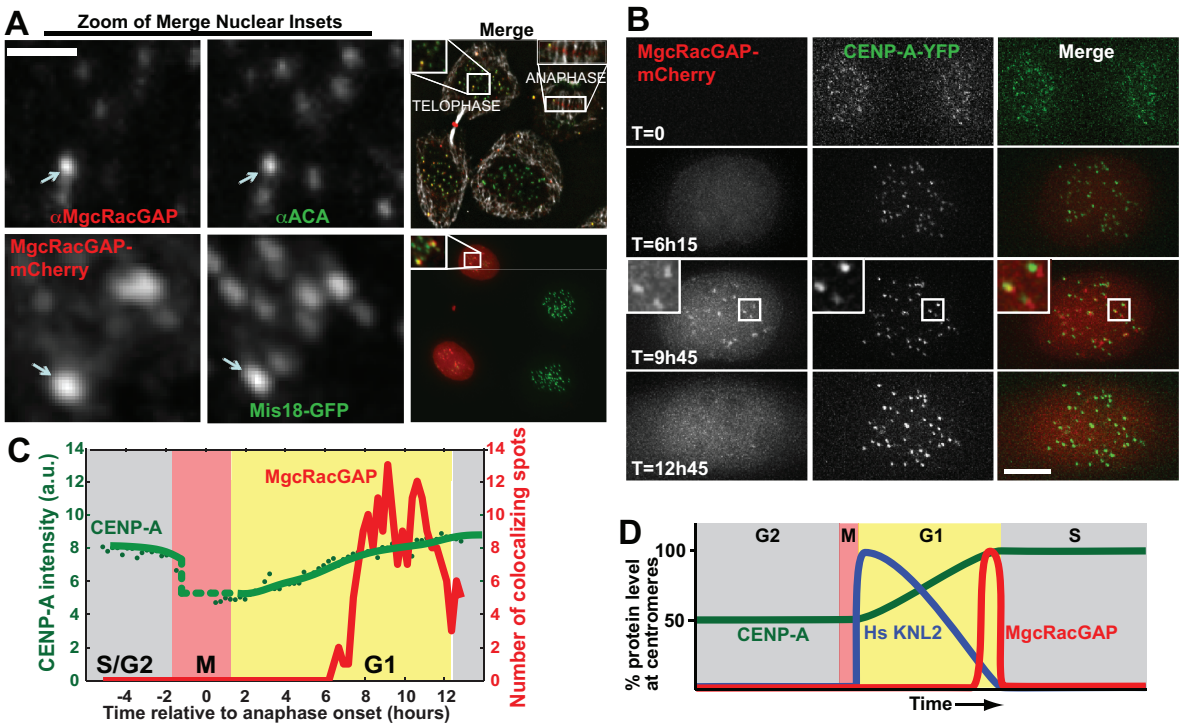



**Figure 3.2. MgcRacGAP localizes to centromeres transiently at the end of CENP-A loading.**
(A) Representative fluorescence microscopy images of endogenous MgcRacGAP and ACA (top; treated with indicates antibodies) and exogenous MgcRacGAP and Mis18 (MgcRacGAP-mCherry and Mis18-GFP; bottom panels) in HeLa cells in G1 phase. Arrow indicates co-localization of MgcRacGAP and centromeres. Images on the right are merged from images on the left, and insets are zooms of indicated nuclear regions; tubulin staining is overlaid in the top right panel in grayscale. Scale bar in nuclear insets, 2 µm and merged images, 10 µm. (B) Cells expressing MgcRacGAP-mCherry and CENP-A-YFP were imaged by time-lapse microscopy. Representative images are shown of MgcRacGAP-mCherry (left) and CENP-A-YFP localization (middle) at indicated nuclear regions. Scale bar=5 µm. (C) Quantification of CENP-A-YFP intensity (A.U.; arbitrary units) at centromeres and number of MgcRacGAP m-Cherry spots that co-localize with centromeres, from an experiment performed as in **B.** (D) Schematic representation of CENP-A (green), HsKNL2 (blue) and MgcRacGAP (red) localization to centromeres with respect to the cell cycle. M is mitosis and S is S-phase.

centromeres was brief (20-60 min) and only approximately 40% of centromeres showed localization at any given time. Our automated analysis was unable to detect

MgcRacGAP at all centromeres owing to high levels of nuclear background, but manual analysis suggests that MgcRacGAP localizes to all centromeres at some point in late G1 (Fig 3.2). Importantly, localization of MgcRacGAP to centromeres was restricted in time: it occurred within a 3-h period following detection at the first centromere (Fig 3.2C and Supplementary Information, Movie S1). Thus, MgcRacGAP is recruited to centromeres after loading of new CENP-A is completed in late G1. This is a cell-cycle pattern different from that of both licensing and loading factors that localize to centromeres from late anaphase until the end of G1, [11,12,13,14]. Together with phenotypic distinctions between MgcRacGAP and licensing or loading factors, localization analysis in living cells suggests that MgcRacGAP functions to maintain CENP-A at the centromere after loading (Fig 3.2D).

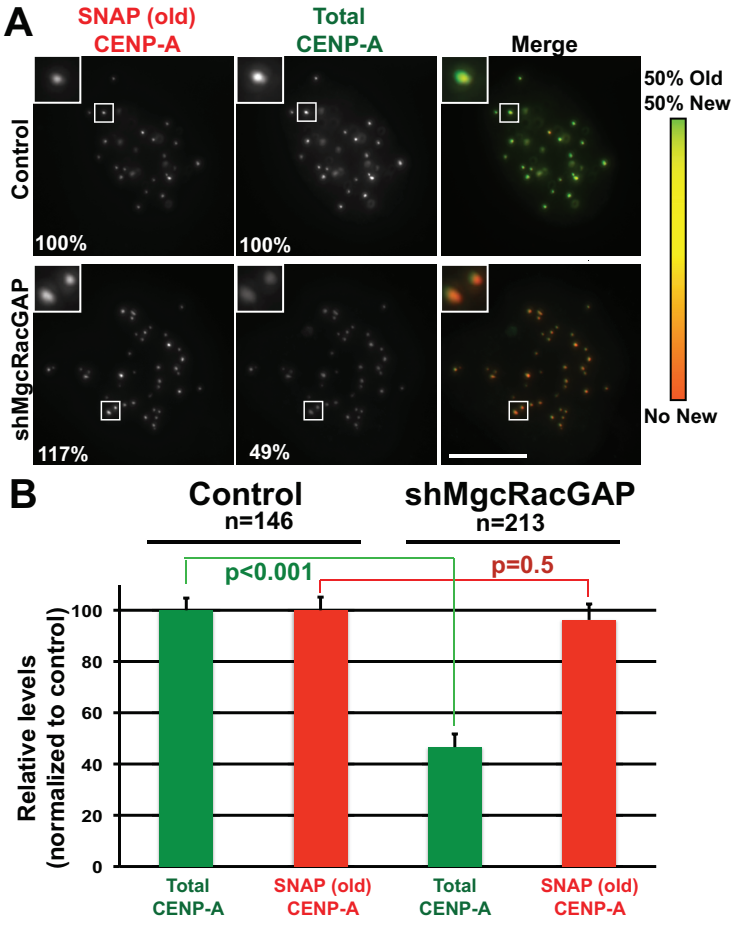



**Figure 3.3. MgcRacGAP is required specifically to stabilize newly incorporated CENP-A.**
(A) Newly deposited CENP-A is specifically lost in the absence of MgcRacGAP. Cells stably expressing SNAP-labelled CENP-A were pulse-labelled with TMR. After pulse-labelling, total CENP-A was labelled by immunostaining (green; middle). Images indicate cells after treatment with (bottom) or without (control; top) MgcRacGAP shRNA. The average (±s.e.m.) intensity of signal at centromeres in cell, normalized with respect to the control cells, is reported in the lower left corner. Colour overlay and a colour scale of the relative ratio (right) indicate that MgcRacGAP is required for maintenance of newly incorporated centromere CENP-A. Scale bar, 5 μm. (B) Quantification of SNAP-tag-labelled CENP-A and total

CENP-A in cells treated with MgcRacGAP shRNA, compared with control cells from experiment performed as in **A**. Levels of SNAP-labelled CENP-A (old, red) were relatively unchanged (P=0.5), but total CENP-A (green) was reduced to less than 50% (P<0.001) after MgcRacGAP depletion, compared with controls. This indicates that the CENP-A lost was not at centromeres when the SNAP pulse label was administered in the previous cell cycle and was therefore newly incorporated CENP-A protein. Data are means ± s.e.m.

### 3.4.3 MgcRacGAP specifically maintains newly incorporated centromere CENP-A

At the end of CENP-A replenishment, before S-phase, there are two distinct populations of CENP-A at centromeres, 50% 'old' and 50% 'new'. Given that CENP-A levels at centromeres are reduced to 50% following MgcRacGAP depletion, one hypothesis is that a distinct CENP-A population is lost. To test this hypothesis, we used the SNAP-tag pulse-chase method to determine whether newly incorporated CENP-A or both old and new CENP-A are destabilized by MgcRacGAP depletion [9]. Briefly, old CENP-A already incorporated into centromeres was labelled with a fluorescent mark using a short (15 min) pulse. All additional CENP-A was chased using a non-fluorescent label. After pulse-labelling, cells were transfected with shRNA vectors to deplete MgcRacGAP. The ratio of marked and total CENP-A at centromeres, compared with control cells in the following G2 (24h later), was then measured. If old CENP-A were specifically lost, the ratio of pulse-labelled CENP-A in depleted cells when compared with controls would be 0:1. Conversely, if new CENP-A were specifically lost, the ratio of pulse-labelled CENP-A would be 1:1. If both populations of CENP-A were lost from centromeres equally, the labelled ratio would be 1:2. Our analysis demonstrated that approximately 50% of the total CENP-A remained, as was observed in previous experiments (see above), following MgcRacGAP depletion. However, the labelled old CENP-A levels were unchanged in depleted cells relative to controls resulting in a ratio of 1.1:1. Reciprocal analysis, specifically labelling new as opposed to old CENP-A confirmed that old CENP-A levels are unchanged in the absence of MgcRacGAP (data not shown). Thus, pulse-chase analysis indicated that cells depleted of MgcRacGAP had a reduced total level

of centromere CENP-A (as shown above) and that centromeres specifically lacked newly incorporated CENP-A (Fig 3.3). Combined with the observations above, these data indicate that MgcRacGAP acts to specifically maintain new CENP-A after it is assembled in G1.

### 3.4.4 Canonical small G-protein GTPase cycling is required for CENP-A maintenance

GAP proteins function in conjunction with a GEF to regulate G-protein GTPase cycles [31]. If the GAP activity of MgcRacGAP is required for centromere maintenance, depletion of its partner GEF, ECT2, should result a phenotype in centromere maintenance. Indeed, depletion of ECT2 reproduced the CENP-A-loss phenotype seen in MgcRacGAP depletion, whereas depletion of MKLP1 (which is required for localization of the GAP to the spindle midzone in anaphase) did not (see below). In addition, an MgcRacGAP mutant lacking GAP activity persisted for an extended duration at centromeres compared with wild-type protein (Fig 3.4 and Supplementary Information, Fig 3.S5 and Movie 2). Thus, MgcRacGAP and ECT2 regulate a small Rho family GTPase cycle, which has a function in centromere maintenance, independently of their roles in cytokinesis. In sum, these results demonstrate that the canonical function of MgcRacGAP in promoting GTPase cycling, but not its spindle midzone localization, is required for ensuring maintenance of CENP-A at centromeres. To determine which small GTPase is the target of MgcRacGAP-ECT2, we depleted RhoA, Rac, or Cdc42 in cells and measured CENP-A levels. Depletion of Cdc42 or Rac, but not RhoA, led to a statistically significant decrease in centromere-localized CENP-A (Fig 3.5). Cdc42 has been implicated in mitotic chromosome segregation. Two studies have shown that loss of MgcRacGAP, ECT2, or Cdc42 function lead to defects in mitotic chromosome congression, a phenotype that could be explained by defective centromere chromatin [32,33]. Additionally, the CENP-A-maintenance phenotype following Cdc42 depletion was slightly more severe than that

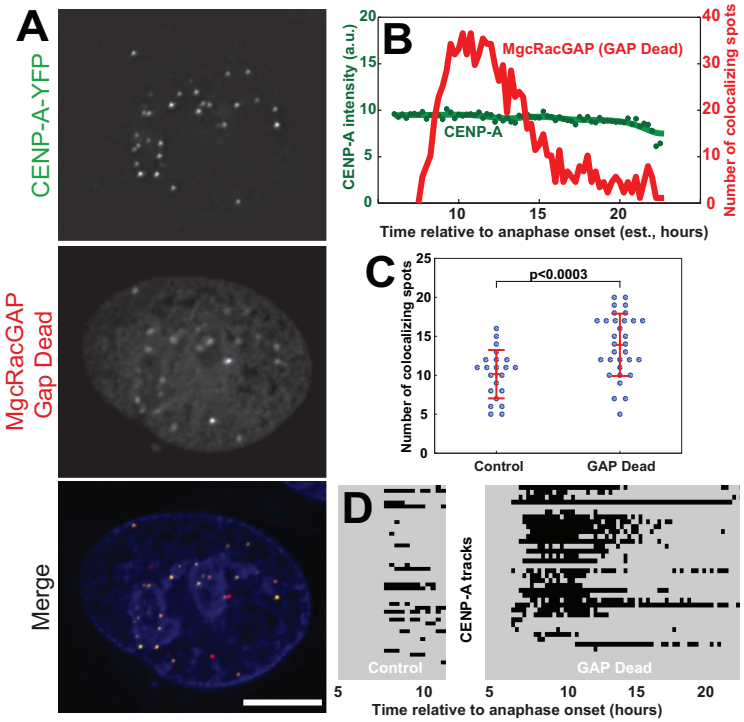



**Figure 3.4. GAP-inactive MgcRagGAP mutant localizes persistently to centromeres.** (A) Fluorescence microscopy images of a HeLa cell expressing GAP-inactive MgcRacGAP-mCherry and CENP-A-YFP. Scale bar, 10 µm. (B) Quantification of CENP-A-YFP intensity at centromeres and number of GAP-inactive MgcRacGAP-mCherry spots that co-localize with centromeres, as assessed from an automated time-lapse live-cell analysis. (C) The number of MgcRacGAP spots that co-localize with centromeres in cells expressing GAP-inactive MgcRacGAP-mCherry or wild-type MgcRacGAP-mCherry was compared by live-cell imaging. Co-localized spots were identified by visual inspection of dual-colour images. Blue, number of co-localizing spots per analysed cell; red, mean ± s.d. (D) Live-cell imaging was used to quantify time of co-localization in cells expressing CENP-A-YFP and either GAP-inactive MgcRacGAP-mCherry or wild-type MgcRacGAP-mCherry. Each row corresponds to a centromere that was tracked through time (horizontal axis). Black dots indicate co-localization of MgcRacGAP-mCherry with the centromere track.

of Rac depletion (Fig 3.5). Therefore, although we cannot eliminate the possibility that Rac also acts in centromere specification, we favour the hypothesis that Cdc42 is the relevant downstream target. Cross-regulation of small GTPases is well documented and may account for the reduction in CENP-A levels following Rac depletion [34]. In support of the hypothesis that Cdc42 is the primary target of MgcRacGAP-ECT2, Cdc42 (but not Rac or RhoA) localized to interphase centromeres (Fig 3.6A). Cdc42 centromere localization is probably transient, as only a few centromeres in a given cell were positive for anti-Cdc42 antibodies. As was also the case for MgcRacGAP, very few cells showed Cdc42 centromere localization. Importantly, co-localization studies revealed that centromeres positive for Cdc42 also had MgcRacGAP signal

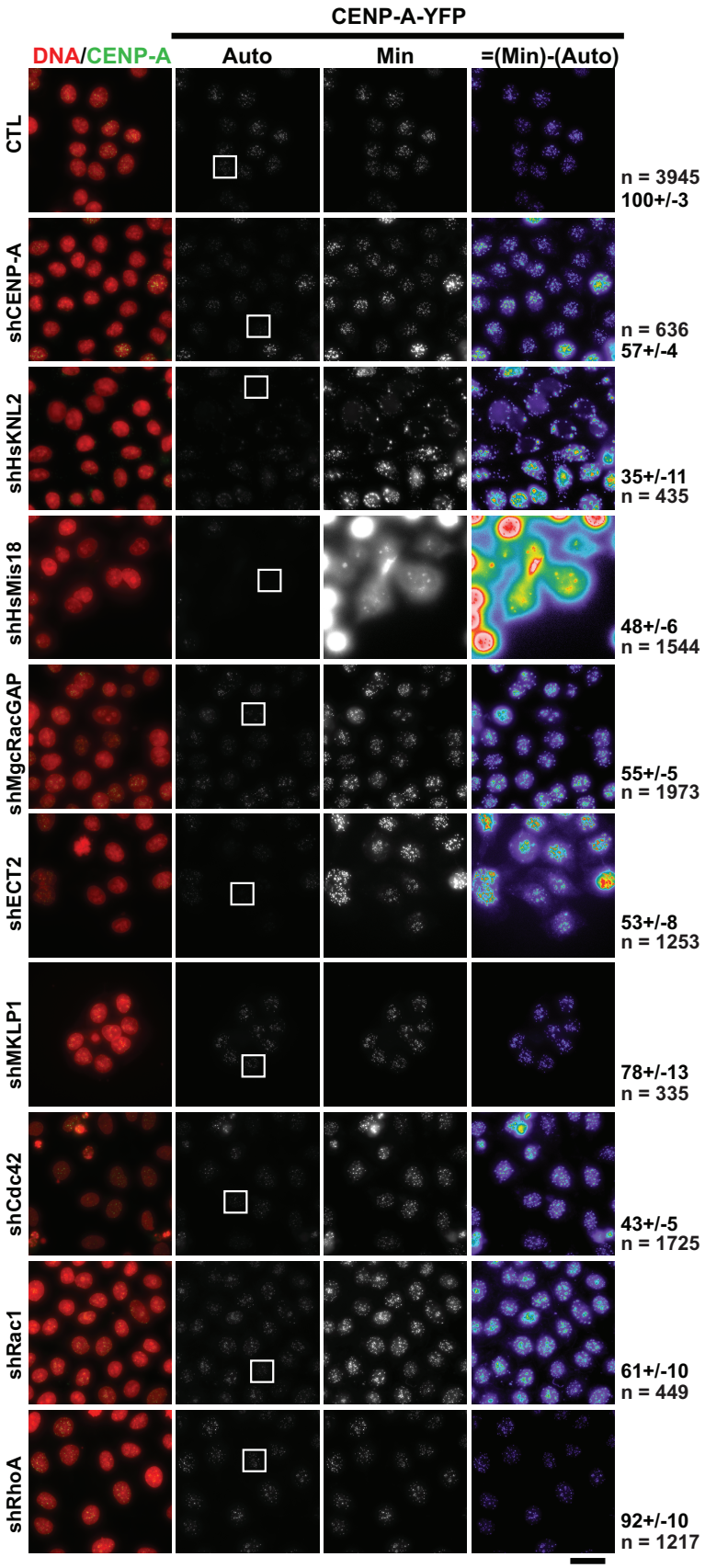



**Figure 3.5. Automated analysis of CENP-A levels following shRNA depletion of various target proteins reveals differential defects in epigenetic regulation of centromeres.**
Representative images of cells at 60x magnification from cells treated with indicated shRNA. Cells were stained with DAPI and the nuclei were segmented (Supplementary Information, Fig 3.S3). Images of DAPI (red) and CENP-A-YFP (green) localization are overlaid in the left column. The column labelled 'Auto' shows the CENP-A-YFP channel rescaled from 12bits to 8bits (required for display purposes) using an auto-scale where the highest pixel value is assigned 255 and the lowest 0 on the 8bit scale. The column labelled 'Min' shows the same images as the 'Auto' column, but it is scaled from 12 bits to 8 bits using the dynamic range of the dimmest cell in the field of view (indicated by the white box in the Auto image). The resulting image reveals cells that have reduced CENP-A signal (affected cells) and saturates cells with 'normal' signal (unaffected cells). Comparing the Auto and Min columns gives a quick view of the level of CENP-A lost in various treatments; conditions resulting in greater loss of CENP-A have more bright cells in the Min column. In the last column, the Auto column has been subtracted from the Min column and pseudo-coloured to facilitate visualization. Redder colours indicate a greater difference between the Min and Auto columns. Therefore, fields with more red colours represent a

greater effect on CENP-A levels. Numbers on the right of each image give the mean ±s.e.m. of CENP-A intensity per cell. *n*; number of cells measured. Scale bar, 20 µm.

(although the reverse was not the case, Supplementary Information, Fig 3.S5A). This localization pattern supports the hypothesis that MgcRacGAP at centromeres transiently activates a Cdc42-mediated switch to maintain centromere identity.

### 3.4.5 Polymeric actin is not required for CENP-A maintenance

Rho family GTPases regulate dynamics of the actin cytoskeleton. Actin has been reported to localize to the nucleus, where it is involved in various metabolic functions [35]. Therefore, it is possible that Cdc42 is required for centromere maintenance through a role in modulating nuclear actin dynamics. To test whether actin filament dynamics are required for centromere specification, we treated cells with Latrunculin A or Cytochalasin D and measured CENP-A levels at centromeres. As expected, most cells were binucleated, indicating penetrant cytokinesis failure (Fig. 3.6B, [36,37,38]). However, quantitative analysis revealed no difference in CENP-A-YFP levels at centromeres after 12h in Latrunculin A or Cytochalasin D when compared with controls, indicating that the function of Cdc42 in this context is independent of actin dynamics and regulates an as yet unknown chromatin remodelling activity (Fig 3.6C). Actin-related-proteins are functional components of a subset of chromatin remodelling complexes; these as well as other chromatin remodelling enzymes or unknown nuclear factors are attractive candidates for downstream Cdc42 effector proteins in regulation of centromere chromatin [39].

## 3.5 Discussion

Epigenetic regulation of genome activity is critical during development and stem cell maintenance, and increasing amounts of evidence highlight its importance in cancers.

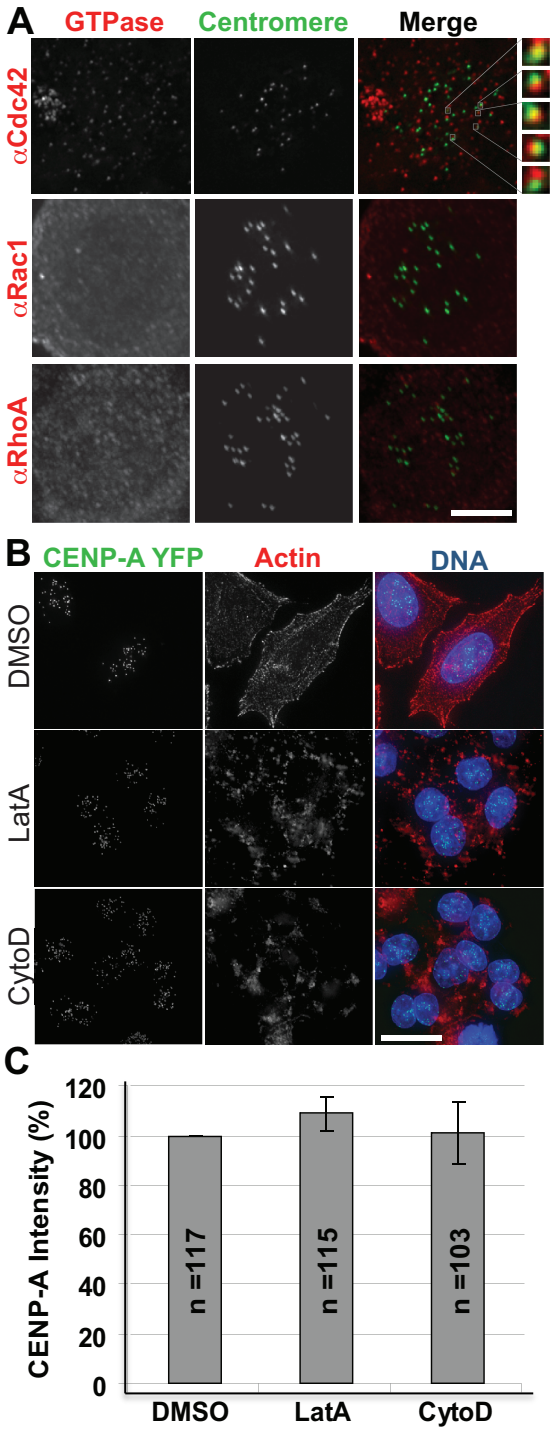



**Figure 3.6. Cdc42 localizes to centromeres and functions to maintain CENP-A levels independent of polymeric actin.**
(A) Representative fluorescence microscopy images of cells stained with antibodies against the indicated GTPases (left). Cells were expressing CENP-A-YFP (as shown in the middle images). Right: merge of images on the left. Co-localization of endogenous CDC42 and CENP-A-YFP is indicated in the insets. All images are projections of 4-8 optical sections from deconvolved image stacks. Scale bar, 5 µm. (B) Representative fluorescence microscopy images of cells expressing CENP-A-YFP and stained with antibodies against actin. Cells were treated as indicated. Right: merge of images on the left, with DAPI used to stain nuclei. Scale bar, 20 µm. (C) Quantification of CENP-A-YFP intensity at centromeres in cells treated as indicated in **B**, normalized to cells treated with DMSO. Data are means ±s.e.m.; *n*, number of cells analysed.

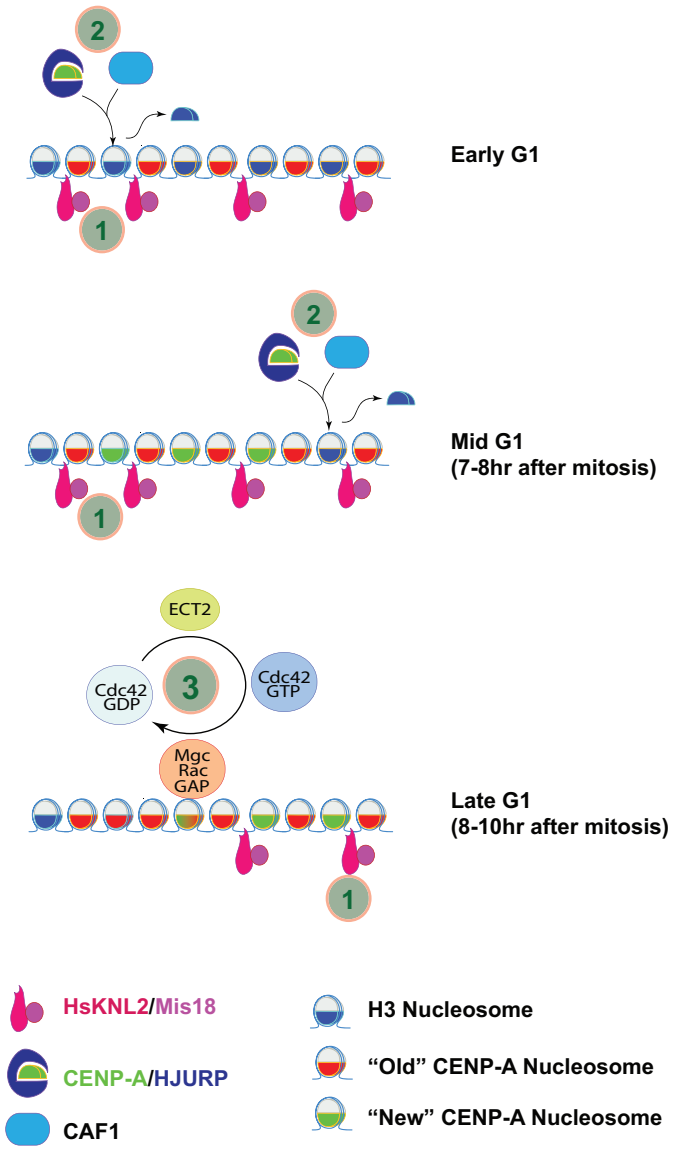



**Figure 3.7. Updated model of centromere specification with the addition of step 3,** Cdc42-mediated maintenance of CENP-A after licensing (step 1) and loading (step 2). MgcRacGAP and ECT2 cycle Cdc42 GTPase activity to modify newly incorporated CENP-A to make it molecularly identical to pre-existing CENP-A.

However, mechanisms controlling epigenetic regulation during a single cell cycle are generally less well understood, compared with those involved in transcriptional programmes. Centromere specification is an epigenetic regulatory event that controls genome activity at singular chromosomal loci and occurs at each cell cycle. Nucleosomes that contain CENP-A are thought to epigenetically define centromeres. During DNA replication, centromere identity is maintained by segregating CENP-A equally to the two daughter chromosomes. Before the subsequent S-phase, additional CENP-A must be incorporated at centromeres, thus propagating the centromere epigenetic mark. Critical to this cycle is maintenance of the proper amount of CENP-A; too little or too much CENP-A incorporation could result in either loss of centromere identity or errors in chromosome segregation [40]. Here, we

describe a mechanism to ensure maintenance of the proper CENP-A levels during the cell cycle regulated by a Rho family small GTPase molecular switch.

We used proteomics and quantitative imaging assays to identify a previously unknown step in centromere maintenance. MgcRacGAP, together with the GEF ECT2, and their cognate small GTPase Cdc42 (or possibly Rac) specifically maintain CENP-A at centromeres. MgcRacGAP localization to centromeres at the end of G1 is incongruous with a role in CENP-A loading and strongly suggests that MgcRacGAP acts in maintenance and not licensing or loading of CENP-A. Pulse-chase analysis revealed that MgcRacGAP is required specifically for maintenance of newly incorporated CENP-A as old CENP-A from the previous cell cycle was present at normal levels at centromeres. Reciprocal immunoprecipitation of MgcRacGAP did not isolate HsKNL2, probably because of a large excess of MgcRacGAP bound to other known interacting proteins in the cytoplasm (data not shown). These results support the conclusion that a minor subset of MgcRacGAP is bound to HsKNL2 for a brief period each cell cycle and imply that non-overlapping MgcRacGAP containing protein complexes function in cells. Overall, our work defines a new event in epigenetic centromere regulation and reveals its control by an unexpected small GTPase molecular switch.

We propose a model wherein the HsKNL2-Mis18 complex licenses centromeres for loading of new CENP-A by the combined activities of HJURP and CAF1 (model, Fig 3.7). After loading (approximately 8-12h after anaphase onset), HsKNL2-Mis18 recruits Cdc42. The activity of Cdc42 is required for preservation of newly incorporated CENP-A and thus finalizes centromere repopulation. Cdc42 activity requires GTPase cycling facilitated by MgcRacGAP and the GEF ECT2. Our results predict that newly incorporated CENP-A is distinct from CENP-A remaining from the previous cell cycle and can be recognized and removed. We propose that Cdc42 activity modifies (by either adding or removing a mark on) newly incorporated CENP-A, rendering it identical to old CENP-A. The manifestation of this mark could be any distinguishing modification, including but not limited to, recruitment of an additional

protein, conformational change of the CENP-A nucleosome, or any of a range of post-translational modifications. New CENP-A that is not modified would be recognized as erroneously incorporated and removed from chromatin during a late-G1 surveillance step, or during DNA replication.

In budding yeast, excess CENP-A (CSE-4) mislocalized to the chromosome arms is removed and selectively degraded through a proteasome-based mechanism [41]. If this mechanism is conserved in human cells, we expect it to be less stringent, since over-expressed CENP-A localizes diffusely to chromosome arms without causing obvious defects in cell division. Alternatively or additionally, centromere maintenance could involve the chromatin remodelling protein RSF-1, which is required for CENP-A nucleosome stability. However, because RSF-1 is proposed to act in mid-G1 before MgcRacGAP and Cdc42 localize to centromeres, it is unlikely to be the downstream target of small GTPase activity at centromeres [16]. Regardless of the removal mechanism, we propose that a GTPase switch is spatially and temporally restricted through regulated localization to centromeres precisely after CENP-A doubling to promote the removal of spurious CENP-A (either excess at centromeres, or outside true centromere loci). By restricting centromere size, this 'quality control' mechanism helps to ensure proper centromere function and kinetochore assembly, thus preventing aneuploidy. Furthermore, it is possible that this mechanistic theme will apply to other epigenetic events that contribute to genomic regulation.

## 3.6 Materials and Methods

All antibodies, plasmids, oligonucleotides, qRT-PCR results and shRNA constructs are listed with detailed information on sequence and suppliers in supplemental tables 1-6. All analysis with the exception of the Snap-Tag experiments was performed on asynchronous cultures 48 hours after transfection with shRNA plasmids.

### 3.6.1 Cell culture

The HeLa cell line and its derivative, which expresses CENP-A-YFP, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and penicillin-streptomycin (Invitrogen, Burlington, ON, Canada) at 37°C in a humidified 5% CO2–95% air incubator [12]. For treatment with actin depolymerizing drugs, cells were incubated for 12 hours in either Latrunculin A (500nM), Cytochalasin D (10 mM) or DMSO (control) [42]. Cells were fixed and stained as described below with the addition of rhodamine Phalloidin to label filamentous actin [43]. Immunofluorescence and Western blot analysis was performed as described [12].

### 3.6.2 Fixed Imaging

All fixed imaging was performed at room temperature on a DeltaVision microscope using Softworx software (Applied Precision, Issaquah, WA, USA) equipped with a CoolSnap HQ2 camera (Photometrics, Tucson, AZ, USA) at 1 × 1 binning and a 100x or 60x planApo objective or on a Nikon Swept Field Confocal (SFC, Nikon Canada, Mississauga, ON, Canada; and Prairie Technologies, Madison, WI, USA) using the 45mm pinhole setting, 60x planApo objective, and 1x1 binning on a CoolSnap HQ2 camera. All SFC acquisitions and additional components including laser exposure setting were controlled by Elements software (Nikon). For the purpose of visualization, some images were deconvolved using Softworx software (Applied Precision). All image analysis was performed on non-deconvolved data.

### 3.6.3 Image and data analysis

All image analysis was carried out using software developed with Matlab (The MathWorks, Natick, MA, USA; release 2008a or newer). Data analysis was carried out either using Matlab or Microsoft Excel (Microsoft, Redmond, WA, USA; version 2003 or newer).

All fluorescence intensity analysis on fixed cells was performed on asynchronous cultures 48 hours after transfection with shRNA plasmids. Fluorescence intensity measurements on fixed cells were carried out in a 2-step process: (1) nuclear segmentation, (2) intensity calculation. Nuclei were segmented using the DAPI channel. The average intensity projection of the z-stack was filtered with a Gaussian mask whose size corresponded to 7 times the width of the point-spread function [44]. The projection image was segmented into foreground (i.e. nuclei) and background by splitting the intensity histogram using Otsu's algorithm [45]. Closely adjacent nuclei were separated by applying a watershed algorithm [46]. This provided a series of masks defined by the DAPI-stained nuclei, which were used to measure nuclear protein intensity in the other fluorescent channels. To avoid measuring erroneously low intensity totals, all nuclear masks touching the border of the image were removed. To further ensure that the measured intensities are guaranteed to originate from nuclear protein, the nuclear masks were shrunk by ~1.6 μm.

Fluorescent intensities were calculated in each channel for each cell by first subtracting the background intensity of the channel from the image, and then by calculating the average fluorescent intensity within the nuclear mask. The background intensity was estimated as the robust mean of the intensities of the pixels classified as background after Otsu thresholding. Qualitatively similar fluorescence measurements were obtained by an alternative, more time-consuming method, which involved detecting each individual centromere, and fitting a point-spread function to each of the 3D intensity distributions to finally determine the intensity as the sum of all the individual intensity integrals (see below).

This analysis returned an intensity value for each cell. In order to compare the measured intensities between experiments, we normalized them in the following way: first, each imaging run was initiated by imaging a control slide prepared in parallel with the experimental samples. Imaging parameters were therefore determined empirically for each experiment each time they were imaged. Importantly, we analysed all cells and did not rely on the use of a co-transfection marker (although individual results were also confirmed by co-transfection with RFP-histone H2B as shown in Figure 3.1). Normalization began by calculating the mean value for the control data and setting this to 1. All experimental values were then normalized to this and the mean value of the normalized data calculated and is the final value listed in the publication.

To measure live-cell fluorescent CENP-A-YFP intensities, we used a modified version of the algorithm presented in [47,48,49] that we will describe in detail elsewhere. Here it was necessary to detect and fit individual CENP-A-YFP spots since in live cells, signal-to-noise ratio was lower than in fixed cells and since there could be non-negligible nuclear background fluorescence. Briefly, 3D images were filtered with what is in essence a spatial band-pass filter, i.e. a filter that emphasizes features of a specific size range, thus reducing high-frequency noise as well as large-scale background variation. We chose to emphasize the size range of 1-4 point-spread functions. Candidate spots for fitting intensities were chosen by unimodal thresholding of the signal-to-noise estimate for each local intensity maximum. The candidate spots were fitted with a Gaussian approximation of the 3D point spread function [44].

For measuring colocalization between CENP-A-YFP spots and spot-shaped signals of other fluorescent proteins, we first performed spot detection in both channels. Since there may be chromatic aberration between the two fluorescent channels, we defined two spots to be co-localized if they are less than 0.75 µm apart. We identified

co-localizing spots by assigning, if possible, each spot in the YFP channel to exactly one spot in the other channel.

### 3.6.4 SNAP-tag pulse chase analysis

SNAP tag analysis was modelled on protocols used in previous studies [9,50] with the following modifications. Briefly, SNAP labelled CENP-A was expressed constitutively from a viral integration site in a clonally selected cell culture; yielding a homogeneous population. Pulse labelling with TMR (tetra-methyl-rhodamine) SNAP substrate was incubated with cells (diluted 1:1000 from 0.6mM stock) for 15 minutes and then the cells were blocked (using unlabelled substrate, New England Biolabs, Ipswich, MA, USA). Cells were then depleted of MgcRacGAP by transfection of shRNA plasmid, as above, then fixed and stained for endogenous CENP-A. The levels of TMR-SNAP labelled ("old") and total CENP-A were measured as described above.

### 3.6.5 Immunoprecipitations and Mass Spectrometry Analysis

For all biochemical isolations, standard protocols were adapted [12,51,52]. For immunoprecipitation experiments, affinity purified antibodies against HsKNL2 (408 and 409), GFP (used as a negative control) and HCP-3 (the C. elegans CENP-A homolog which shows virtually no homology to any human proteins in the n-terminus where this antibody is directed, used as a negative control) were crosslinked to Affi-Prep Protein A beads (Biorad, Mississauga, ON, Canada), using dimethylpimelimidate (DMP). After incubation with DMP, beads were washed 2X with 1 ml 0.2M ethanolamine, 0.2M NaCl pH 8.5 to inactivate the residual crosslinker and then Pre-eluted with 0.1 M Triethlyamine pH 11.5 and washed with washing buffer containing 300 mM KCl, 0.05% NP-40 and DTT 0.5M and with Lysis Buffer without detergent to neutralize Triethlyamine. Cell lysates were treated with Micrococcal Nuclease (MN) (Fermentas, Burlington, ON, Canada; EN 0181) for 15 min. at RT. Beads were incubated with cellular extract for 1 hr at 4° C washed with washing buffer and associated proteins were eluted with 0.1 M Triethlyamine pH 11.3

neutralized with 150 µl 1 M Hepes pH 7.2. Proteins were then precipitated with TCA (Sigma-Aldrich, St. Louis, MO, USA) 100% (200 µl). Protein pellet was washed with Acetone and resuspended into Ammonium bicarbonate 100mM for in solution tryptic digestion and the entire mixture was analysed by Mass spectrometry.

### 3.6.6 Quantitative real-time PCR (qRT-PCR)

Transfected CENP-A-YFP HeLa cells were disrupted and homogenized using RNeasy Kit and QIAshredder (QIAGEN, Mississauga, ON, Canada) and total cellular RNA was isolated according to the manufacturer's instructions. All qRT-PCR analysis was performed on asynchronous cultures 48 hours after transfection with shRNA plasmids. Two micrograms of RNA were reverse transcribed using High-Capacity cDNA Reverse Transcription Kits (Applied Biosystems, Foster City, CA, USA). Gene expression was assessed by qRT-PCR using 7900HT with Roche Universal ProbeLibrary assays. For each gene, reactions were performed in triplicate from 3 different experiments, in 384 well plates for 50 amplification cycles (95°C 10 s; 95°C 15 s; 60°C 1 min). Gene expression assays were designed using the Roche Universal ProbeLibrary assay design software and were tested for maximum efficiency by standard curve analysis (slope = 3.1–3.6). Reference gene assays (GAPDH and HRPT) were purchased from ABI (20× primer-probe mix, VIC labeled). For each cDNA sample, a Cycle threshold (Ct) value was determined for every gene of interest ($Ct_{target}$) and endogeneous control (GAPDH and HRPT) ($Ct_{reference}$). Gene depletion was determined relative to cDNA control samples ($Ct_{calibrateur}$) for every experiment and normalized to endogenous reference genes (e.g. $\Delta Ct = Ct_{target} - Ct_{reference}$). Standard deviation (SD) was calculated for each triplicate and only samples with SD under 0.25 were considered. Relative quantification ($RQ = 2^{-\Delta\Delta Ct}$) of the gene expression values was determined relative to a calibrator value of 1. Data were analysed via SDS 2.2.2 software.

### 3.6.7 Statistical evaluations

All pair wise statistical comparisons were performed using Excel's T-test (two sided).

## 3.7 Acknowledgements

The authors would like to thank Jean-Claude Labbé, Arshad Desai, Kerry Bloom, and Sylvain Meloche for helpful comments and discussions, the IRIC proteomics facility (specifically Éric Bonneil and Pierre Thibault) for LC-MS/MS analysis, the IRIC genomics facility for sequencing and qRT-PCR analyses, and Aaron Straight and Ben Moree for the generous gift of the CENP-A SNAP-tag cell line. JFD was supported by a postdoctoral fellowship from the Swiss National Science Foundation; VDR was supported by a student fellowship from the CIHR. ASM receives salary support from the FRSQ. PSM holds the Canada Research Chair in Chromosome Organization and Mitotic Mechanisms. This work was funded by grants #018450 awarded to PSM and #019162 awarded to ASM from the Terry Fox Foundation and the CCSRI.

Author Contributions: AL conducted all IP/MS experiments. AL and JD performed all shRNA experiments and analysis respectively. JD performed live cell experiments. AL evaluated all shRNA experiments (qRT-PCR, western blot). VDR performed the small GTPases localization experiments. All authors performed essential tasks in generating figures. PSM, AL, JD, AML, and ASM conceived the experiments. PSM wrote the manuscript, assisted by all authors, in particular ASM and JD.

## 3.8 References


1 Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635-638, (2007).

2 Karpen, G. H. & Allshire, R. C. The case for epigenetic effects on centromere identity and function. *Trends Genet* **13**, 489-496 (1997).

3 Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H. & Margolis, R. L. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc Natl Acad Sci U S A* **88**, 3734-3738 (1991).

4 Choo, K. H. Centromerization. *Trends Cell Biol* **10**, 182-188 (2000).

5 Bernad, R., Sanchez, P. & Losada, A. Epigenetic specification of centromeres by CENP-A. *Exp Cell Res* **315**, 3233-3241, (2009).

6 Sullivan, K. F., Hechenberger, M. & Masri, K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J Cell Biol* **127**, 581-592 (1994).

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

8 Schuh, M., Lehner, C. F. & Heidmann, S. Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. *Curr Biol* **17**, 237-243, (2007).

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

10 Hayashi, T. *et al.* Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* **118**, 715-729 (2004).

11 Fujita, Y. *et al.* Priming of Centromere for CENP-A Recruitment by Human hMis18alpha, hMis18beta, and M18BP1. *Dev Cell* **12**, 17-30 (2007).

12 Maddox, P. S., Hyndman, F., Monen, J., Oegema, K. & Desai, A. Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. *J Cell Biol* **176**, 757-763 (2007).

13 Foltz, D. R. *et al.* Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* **137**, 472-484, (2009).

14 Dunleavy, E. M. *et al.* HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485-497, (2009).

15 Silva, M. C. & Jansen, L. E. At the right place at the right time: novel CENP-A binding proteins shed light on centromere assembly. *Chromosoma* **118**, 567-574, (2009).

16 Perpelescu, M., Nozaki, N., Obuse, C., Yang, H. & Yoda, K. Active establishment of centromeric CENP-A chromatin by RSF complex. *J Cell Biol* **185**, 397-407, (2009).

17 Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629-635, (2002).

18 Izuta, H. *et al.* Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. *Genes Cells* **11**, 673-684, (2006).

19 Kawashima, T. *et al.* MgcRacGAP is involved in the control of growth and differentiation of hematopoietic cells. *Blood* **96**, 2116-2124 (2000).

20 Minoshima, Y. *et al.* Phosphorylation by aurora B converts MgcRacGAP to a RhoGAP during cytokinesis. *Dev Cell* **4**, 549-560, (2003).

21 Toure, A. *et al.* MgcRacGAP, a new human GTPase-activating protein for Rac and Cdc42 similar to Drosophila rotundRacGAP gene product, is expressed in male germ cells. *J Biol Chem* **273**, 6019-6023 (1998).

22 Yuce, O., Piekny, A. & Glotzer, M. An ECT2-centralspindlin complex regulates the localization and function of RhoA. *J Cell Biol* **170**, 571-582, (2005).

23 Canman, J. C. *et al.* Inhibition of Rac by the GAP activity of centralspindlin is essential for cytokinesis. *Science* **322**, 1543-1546, (2008).

24 Miller, A. L. & Bement, W. M. Regulation of cytokinesis by Rho GTPase flux. *Nat Cell Biol* **11**, 71-77, (2009).

25 Hirose, K., Kawashima, T., Iwamoto, I., Nosaka, T. & Kitamura, T. MgcRacGAP is involved in cytokinesis through associating with mitotic spindle and midbody. *J Biol Chem* **276**, 5821-5828, (2001).

26 Jantsch-Plunger, V. *et al.* CYK-4: A Rho family gtpase activating protein (GAP) required for central spindle formation and cytokinesis. *J Cell Biol* **149**, 1391-1404 (2000).

27 Pavicic-Kaltenbrunner, V., Mishima, M. & Glotzer, M. Cooperative assembly of CYK-4/MgcRacGAP and ZEN-4/MKLP1 to form the centralspindlin complex. *Mol Biol Cell* **18**, 4992-5003, (2007).

28 Maddox, A. S. & Oegema, K. Closing the GAP: a role for a RhoA GAP in cytokinesis. *Mol Cell* **11**, 846-848 (2003).

29 Black, B. E. *et al.* Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell* **25**, 309-322 (2007).

30 Foltz, D. R. *et al.* The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* **8**, 458-469 (2006).

31 von Dassow, G., Verbrugghe, K. J., Miller, A. L., Sider, J. R. & Bement, W. M. Action at a distance during cytokinesis. *J Cell Biol* **187**, 831-845, (2009).

32 Yasuda, S. *et al.* Cdc42 and mDia3 regulate microtubule attachment to kinetochores. *Nature* **428**, 767-771 (2004).

33 Oceguera-Yanez, F. *et al.* Ect2 and MgcRacGAP regulate the activation and function of Cdc42 in mitosis. *J Cell Biol* **168**, 221-232, (2005).

34 Sanders, L. C., Matsumura, F., Bokoch, G. M. & de Lanerolle, P. Inhibition of myosin light chain kinase by p21-activated kinase. *Science* **283**, 2083-2085 (1999).

35 Rando, O. J., Zhao, K. & Crabtree, G. R. Searching for a function for nuclear actin. *Trends Cell Biol* **10**, 92-97, (2000).

36 Glotzer, M. The molecular requirements for cytokinesis. *Science* **307**, 1735-1739, (2005).

37 Maddox, A. S. & Oegema, K. Deconstructing cytokinesis. *Nat Cell Biol* **5**, 773-776 (2003).

38 Field, C., Li, R. & Oegema, K. Cytokinesis in eukaryotes: a mechanistic comparison. *Curr Opin Cell Biol* **11**, 68-80, (1999).

39 Chen, M. & Shen, X. Nuclear actin and actin-related proteins in chromatin dynamics. *Curr Opin Cell Biol* **19**, 326-330, (2007).

40 Tomonaga, T. *et al.* Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* **63**, 3511-3516 (2003).

41 Collins, K. A., Furuyama, S. & Biggins, S. Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr Biol* **14**, 1968-1972 (2004).

42 Ye, J., Zhao, J., Hoffmann-Rohrer, U. & Grummt, I. Nuclear myosin I acts in concert with polymeric actin to drive RNA polymerase I transcription. *Genes Dev* **22**, 322-330, (2008).

43 Theriot, J. A., Mitchison, T. J., Tilney, L. G. & Portnoy, D. A. The rate of actin-based motility of intracellular Listeria monocytogenes equals the rate of actin polymerization. *Nature* **357**, 257-260 (1992).

44 Stelzer, E. H., Wacker, I. & De Mey, J. R. Confocal fluorescence microscopy in modern cell biology. *Semin Cell Biol* **2**, 145-152 (1991).

45 Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on* **9**, 62-66 (1979).

46 Lindeberg, T. & Eklundh, J. O. in *Computer Vision, 1990. Proceedings, Third International Conference on.* 416-426.

47 Dorn, J. F., Danuser, G. & Yang, G. Computational processing and analysis of dynamic fluorescence image data. *Methods Cell Biol* **85**, 497-538, (2008).

48 Jaqaman, K. *et al.* Kinetochore alignment within the metaphase plate is regulated by centromere stiffness and microtubule depolymerases. *J Cell Biol* **188**, 665-679, (2010).

49 Dorn, J. F. *et al.* Yeast kinetochore microtubule dynamics analyzed by high-resolution three-dimensional microscopy. *Biophys J* **89**, 2835-2854, (2005).

50 Carroll, C. W., Silva, M. C., Godek, K. M., Jansen, L. E. & Straight, A. F. Centromere assembly requires the direct recognition of CENP-A nucleosomes by CENP-N. *Nat Cell Biol* **11**, 896-902, (2009).

51 Cheeseman, I. M. & Desai, A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci STKE* **2005**, pl1 (2005).

52 Cheeseman, I. M. *et al.* A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev* **18**, 2255-2268 (2004).

# CHAPTER 4

Publication

# CENTROMERE EPIGENOME STABILITY IS MEDIATED BY KNL-2 STRUCTURAL RECOGNITION OF CENP-A CHROMATIN

Article in revision in Developmental Cell (2014)

## 4.1 Preface

Following the identification of KNL-2 as an essential protein for CENP-A localization to centromere, work published in The Journal of Cell Biology in 2007, Dr Paul S. Maddox wanted to investigate the biochemical characterization of the predicted Myb DNA binding domain. This characterization lead to a solved 3D structure of *C. elegans* KNL-2 Myb domain in collaboration with Dr. Michael J. Osborne from the Biophysics platform at IRIC, who performed the NMR experiments and analysis. Dr. Jonas F. Dorn conceived and performed the live cell experiments and the image quantification. Corentin Moevus performed the Myb domain deletion cloning, which I conceived. Also, I conceived and performed the all the protein expressions and purifications experiments, together with all DNA binding assays *in vitro*, as well as the quantification of gene expression by qPCR (Supplementary figures). I conceived and performed the TIRF experiments with the help of Joël Ryan and Abbas Padeganeh. Quantifications of pull-down efficiency of the TIRF images were performed using a program wrote by Jacques Boisvert in Matlab. Joël Ryan performed the actual quantification analysis. Dr. Jonas F. Dorn, Dr. Paul S. Maddox, Dr. Amy S. Maddox, Dr. Katherine L.B. Borden and I wrote the manuscript.

**Centromere epigenome stability is mediated by KNL-2 structural recognition of CENP-A chromatin**

Authors: Valérie De Rop[1*], Jonas F. Dorn[1*], Michael J. Osborne[1], Jacques Boisvert[1], Joël Ryan[1], Abbas Padeganeh[1], Corentin Moevus[2], Katherine L.B. Borden[1], Amy S. Maddox[1,3], Paul S. Maddox[1,3]

Affiliations:
[1] Institut de Recherche en Immunologie et en Cancérologie (IRIC), Département de Pathologie et de Biologie Cellulaire, Université de Montréal, Montréal.
[2] Current address: Department of Pathology and Cell Biology, Columbia University, New-York.
[3] Current address: Department of Biology, University of North Carolina, Chapel Hill.
[*]These authors contributed equally to this work

Correspondence to
Institute for Research in Immunology and Cancer (IRIC),
Department of Pathology and Cell Biology
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7 CANADA

Running title: Specific binding of KNL-2 to CENP-A chromatin

## 4.2 Abstract

In metazoans, centromere specification is epigenetically controlled by the histone H3 variant CENP-A. Each cell cycle, CENP-A is replenished in a manner requiring the conserved centromere licensing complex (CLC) that includes the Myb-domain protein KNL-2. Here we show that KNL-2 is critical for the retention and segregation of the centromere epigenome: depletion of KNL-2 blocked loading of new CENP-A and caused loss of pre-existing CENP-A. Structural dynamics analysis revealed that centromere targeting by the CLC requires recognition of a CENP-A-generated DNA feature by the KNL-2 C-terminal Myb type DNA binding domain. This interaction is atypical in two key mechanisms. Firstly, unlike all known Myb domains, the KNL-2 Myb domain does not recognize a sequence motif. Secondly, both CeKNL-2 and HsKNL-2 interact with DNA in distinct, novel fashion. Thus, we have uncovered novel diversity within the Myb family, and have identified the molecular mechanism of interpretation of centromeric epigenetic identity.

Highlights

CLC is dynamic in G1 and is required for CENP-A loading and retention.

*C. elegans* KNL-2 Myb domain was solved by NMR and displays a Myb like structure.

KNL-2 Myb domain is sufficient to recognize and bind specifically CENP-A chromatin.

## 4.3 Introduction

Centromeres are genomic loci that direct the assembly of kinetochores in mitosis (Brinkley and Stubblefield, 1966; Robbins and Gonatas, 1964). This megadalton protein complex contains centromeric proteins responsible for direct attachment of condensed chromosomes to the mitotic spindle, and outer kinetochore components essential for mitotic checkpoint responses. Consequently, loss of centromere function results in dramatic failure of chromosome segregation (Blower and Karpen, 2001; Buchwitz et al., 1999; Stoler et al., 1995; Takahashi et al., 2000). The histone H3 variant CENP-A (CENtromere Protein A; CenH3) (Palmer et al., 1991; 1987; Sullivan et al., 1994; Yoda et al., 2000), found exclusively at centromeres, is both necessary and sufficient for centromere identity (Gascoigne et al., 2011; Mendiburo et al., 2011). Given that centromeric DNA sequence is ineffective for centromere maintenance, it is broadly accepted that CENP-A is the epigenetic mark for centromere identity and is essential for centromere epigenome stability (Bodor et al., 2013; Mendiburo et al., 2011; Vafa and Sullivan, 1997; Van Hooser et al., 2001; Warburton and Cooke, 1997).

CENP-A incorporation to centromeric loci is replication-independent, occurring in G1, and requires a cell cycle regulated mechanism comprising three steps: Licensing, Loading, and Maintenance (De Rop et al., 2012; Jansen et al., 2007; Schuh et al., 2007). Licensing is the first step for newly synthesized CENP-A incorporation and it is accomplished by the Centromere Licensing Complex (CLC, also known as the Mis18 complex or Licensing complex) made up of Mis18α, Mis18β, and KNL-2 (also termed M18BP1) (Fujita et al., 2007; Maddox et al., 2007). Depletion of any single CLC complex member results in a diminution of CENP-A level at centromeres and a loss of centromere identity (Fujita et al., 2007; Maddox et al., 2007). It was demonstrated in the nematode *C. elegans* that KNL-2 depletion results in the loss of pre-existing CENP-A at centromeres, thus showing its role in centromere identity maintenance (Lagana et al., 2010; Maddox et al., 2007). Loading of new CENP-A is tightly

regulated by its chaperone HJURP (Dunleavy et al., 2009; Foltz et al., 2009). Prior to incorporation of CENP-A to the centromere, HJURP dimerises and binds a CENP-A/H4 tetramer (Zasadzińska et al., 2013). Depletion of HJURP reveals a decrease in CENP-A intensity without affecting the stability of old incorporated CENP-A at centromeres (Foltz et al., 2009), thus demonstrating its role for newly synthesized CENP-A incorporation by recognizing a mark on the old centromeric chromatin by an unknown mechanism.

KNL-2, having a predicted Myb like DNA binding domain, is a candidate for the most upstream component of the CLC since it is required for subsequent Loading and Maintenance steps in centromere epigenome inheritance (Barnhart et al., 2011; Lagana et al., 2010). Moreover, *Arabidopsis thaliana* homolog KNL2 is required for CenH3 deposition in G2 and its C-terminus part is sufficient to recognize and bind centromeres *in vivo* (Lermontova et al., 2013). However, the *A. thaliana* KNL2 C-terminus has no predicted Myb binding domain. Also, KNL-2 is negatively regulated by CDK phosphorylation, preventing CENP-A incorporation outside of G1 phase (Silva et al., 2012). Despite this wealth of information, the molecular mechanism of centromere licensing, explicitly how KNL-2 recognizes and binds to centromere is unknown.

Here we use super-resolution analysis of time-lapse microscopy to show that KNL-2 inhibition results in arrest of new CENP-A incorporation, as well as a loss of centromere epigenetic identity in a single cell cycle. Structural dynamics studies reveal that KNL-2 binds to centromeric chromatin via a Myb type DNA binding domain. Solving the structure of the nematode *C. elegans* KNL-2 Myb domain combined with Total Internal Reflection Fluorescence microscopy (TIRFm) based biochemical assays revealed that KNL-2 preferentially binds to CENP-A containing chromatin independent of DNA sequence. Structural recognition of CENP-A chromatin is conserved in metazoan, since *C. elegans* KNL-2 Myb domain recognizes human centromeric chromatin. These results define a unique mechanism regulating

centromere epigenetic identity through KNL-2 function, which may serve as a paradigm for general epigenome segregation during cell division.

## 4.4 Results

### 4.4.1 CLC dynamically localizes to centromeres to initiate CENP-A chromatin assembly

The CLC was identified via functional screens as required for CENP-A localization to centromeres (Hayashi et al., 2004; Maddox et al., 2007). To better define the cell cycle timing of CLC centromere activity, we used high-resolution time-lapse imaging followed by super-resolution based analysis of the CLC component Mis18-GFP at individual centromeres in single live cell approach. The CLC is recruited to centromeres within approximately 5 min of anaphase onset (Fig. 4.1a and b), rapidly increasing to a maximal level in about 2 hours, after which time it gradually leaves centromeres with a first order decay constant of 1.74±0.14/hour, translating to a $t_{½}$ of approximately 23 minutes. After more than 5 hours post anaphase onset however, the CLC is still detectable at centromeres, indicating the existence of a stable as well as a dynamic population (data not shown). In sum, recruitment of the CLC to centromeres is concomitant with the timing of CENP-A loading in G1 phase (Jansen et al., 2007; Schuh et al., 2007; Shelby et al., 2000), consistent with the hypothesis that the CLC interprets a centromere epigenetic mark and directs replenishment of CENP-A.

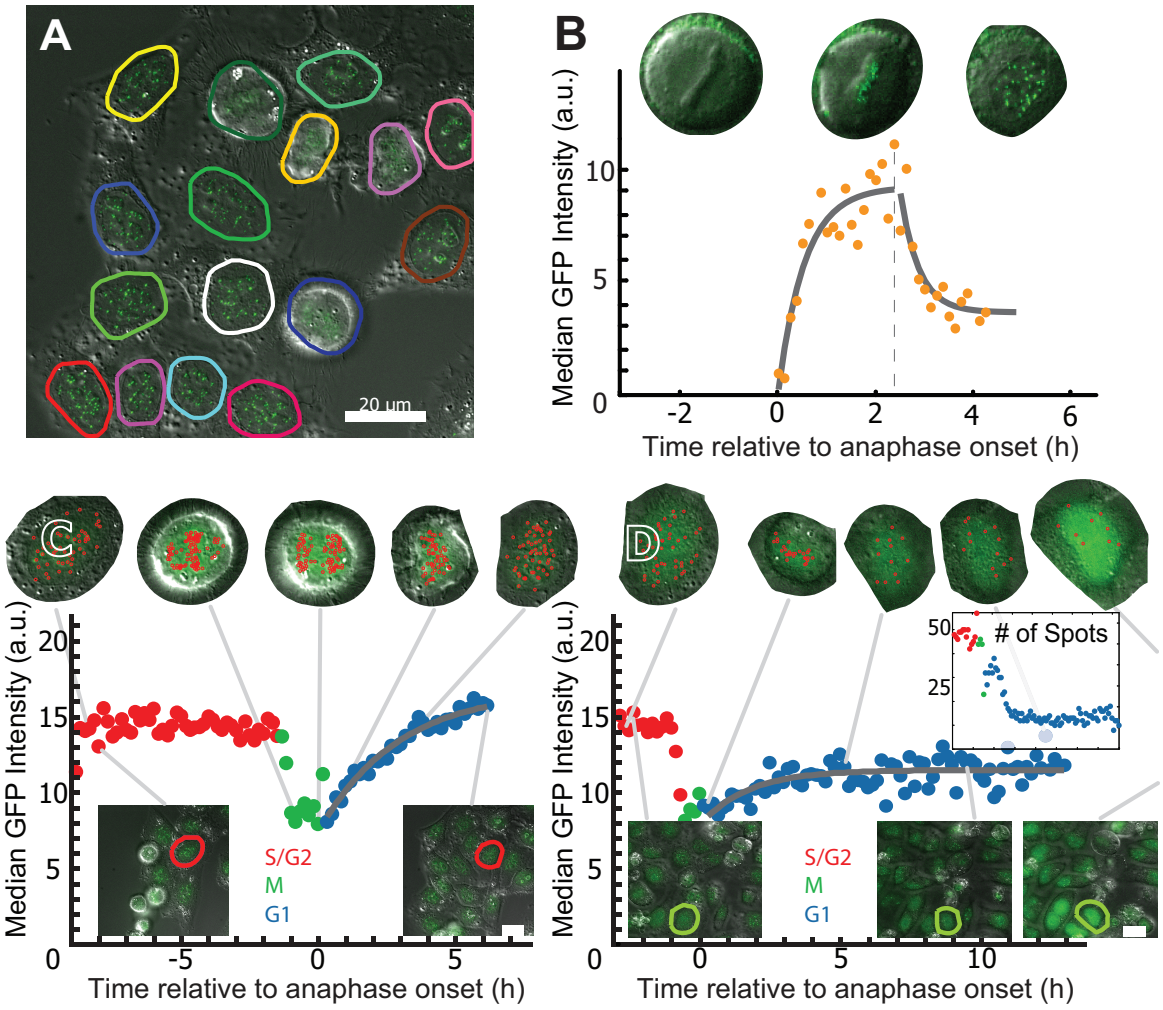



**Figure 4.1 Dynamics of CENP-A and Mis18 at the centromere.**
**a**, Quantification of CENP-A-GFP and Mis18-GFP levels by automated image analysis (see methods). **b**, Quantification of Mis18-GFP fluorescence intensity showing its dynamics loading after anaphase onset. Within approximately 2 hours after anaphase onset, centromeric Mis18 levels peak, after which part of Mis18 is lost from centromeres, while part of Mis18 persists. **c**, Quantification of CENP-A-GFP fluorescence intensity showing its dynamics level at the centromere. The dotted orange curve shows median centromeric CENP-A levels and the solid grey curve is a mono-exponential fit to the CENP-A loading dynamics. **d**, Quantification of CENP-A-GFP fluorescence intensity shows that KNL-2 depletion leads to a reduction in final CENP-A levels, while the loading dynamics are not perturbed.

CLC activity at centromeric chromatin is proposed to license the centromeric region (Fujita et al., 2007), and then recruit the loading machinery, composed of the chaperone HJURP, to incorporate new CENP-A (Barnhart et al., 2011; Dunleavy et al., 2009; Foltz et al., 2009). Therefore, we predicted that CLC inhibition would not alter the rate of new CENP-A chromatin assembly, but rather will result in decreased CENP-A incorporation. To determine the function of the CLC, we analyzed CENP-A-GFP at individual centromeres using our super-resolution single-cell time-lapse imaging assay. In control cells (Fig. 4.1c), CENP-A levels at individual centromeres doubled during G1 over the course of 6 to 8 hours. Doubling began within minutes of anaphase onset. The dynamics were best described by a first-order model with a rate constant *k* of approximately 0.39±0.04/hour ($t_{½}$ of approximately 110 minutes). Considering that there are at most 15,000 CENP-A nucleosomes per centromere in HeLa cells (Black et al., 2007), we estimated that the speed of CENP-A deposition is approximately 10 CENP-A nucleosomes per minute per centromere. This is comparable to the speed of nucleosome deposition at replication forks during S phase in human cells (1-2kb/s) (Conti et al., 2007). Thus, CENP-A loading dynamics in early G1 comprise a first-order reaction and may be mechanistically analogous to nucleosome deposition after replication.

Cells partially depleted of CENP-A or HJURP (Fig. 4.S1) loaded new CENP-A at a reduced rate and to a reduced final level. The extent of rate reduction averaged approximately 30% but was variable and likely proportional to the level of knockdown; in some cases, no new CENP-A was loaded. In these examples, each centromere maintained the amount of centromere-bound CENP-A from the previous cell cycle, as has been observed by fixation based pulse-chase assays (Lagana et al., 2010). Depletion of KNL-2 resulted in loss of CENP-A localization to centromeres. However, under partial depletion conditions, some foci were still present and tracked by using our super-resolution single-cell time-lapse imaging assay. Depletion of KNL-2 also resulted in reduced CENP-A incorporation (Fig. 4.1d), without altering cellular CENP-A levels (Lagana et al., 2010). However, the rate constant of CENP-A incorporation was not changed by KNL-2 depletion, indicating the speed of incorporation does not

noticeably depend on the quantity of KNL-2 present at the centromere, consistent with a proposed enzymatic activity of the chaperone HJURP which is recruited by the CLC (Wang et al., 2014). These results are consistent with the CLC acting as an intermediate, interpreting the centromere epigenetic mark to recruit the loading machinery, which in turn accomplishes CENP-A incorporation.

Depletion of KNL-2 not only blocked loading of new CENP-A, but also caused an overall loss of CENP-A from centromeres (Fig. 4.1d). This was previously reported in *C. elegans* where depletion of KNL-2 caused total loss of CENP-A at centromeres (Maddox et al., 2007). Thus the CLC is required for both recruiting the loading complex for new CENP-A incorporation and also for maintaining pre-existing CENP-A at centromeres. Previously, we identified a maintenance step during late G1 in which improperly incorporated CENP-A is removed (Lagana et al., 2010). Here our results suggest that the removal process can occur earlier in the cell cycle, and that the CLC antagonizes this activity in a manner independent of CENP-A loading. In sum, our super-resolution time-lapse data predict that CLC localization to centromeres is crucial for G1 centromere identity maintenance. Therefore, the mechanism of CLC binding to centromeres is key to understanding centromere epigenetic regulation.

### 4.4.2 The HsKNL-2 Myb domain is sufficient for CENP-A chromatin recognition

Our understanding of CLC function in accurate centromere epigenome segregation is based on phenotypic observations of cultured cells (Barnhart et al., 2011; Fujita et al., 2007; Maddox et al., 2007). In order to better define the biochemical mechanism of CLC action, we turned to *in vitro* assays. KNL-2 is weakly conserved and has a predicted Myb type DNA binding domain in its C-terminus (Fig. 4.2a). Myb domains can mediate protein-DNA interactions (Biedenkapp et al., 1988; Mizuguchi et al., 1990; Tanikawa et al., 1993), thus we hypothesized that this domain targets KNL-2 and therefore, the CLC to centromeric chromatin. The structure of the HsKNL-2

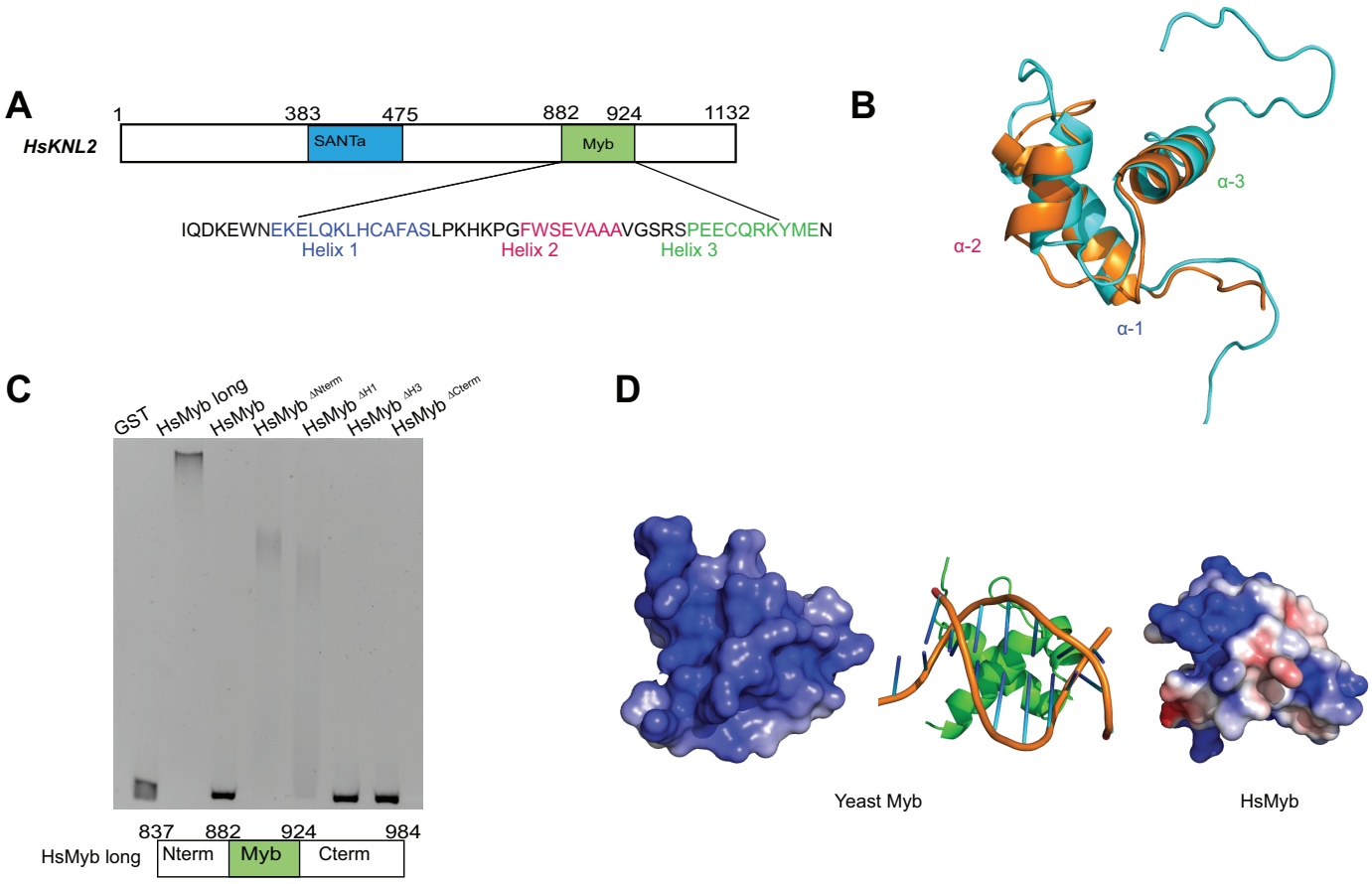



**Figure 4.2 HsKNL-2 Myb domain binds DNA *in vitro*.**
**a**, Schematic representation of HsKNL-2 sequence showing predicted SANT and Myb domains positions, at N-terminus and C-terminus respectively. Helix colors are associated with the NMR solved structure of HsKNL-2 Myb domain (pdb #1WGX) in **B** (cyan). **b**, Overlay of HsMyb domain (cyan) and a conventional Myb DNA binding domain from yeast (orange, pdb #1WOT), showing similar structures of both domains. Proteins are superimposed based on the helical elements (RMSD 0.5Å). **c**, EMSA of HsKNL-2 Myb domain truncations performed with 208bp human α-satellite DNA. C-terminus is required for stable interaction with DNA. **d**, Electrostatic comparison of Yeast Myb (pdb #1WOT) and HsMyb (pdb #1WGX). Proteins are oriented according to the central panel, which shows the importance of helix 3 for DNA binding of Yeast Myb.

(Human KNL-2) Myb domain (Fig. 4.2b) has been solved by Nuclear Magnetic Resonance (NMR): it comprises a helix-turn-helix motif as found in other Myb proteins (pdb #1WGX). To test if the KNL-2 Myb domain can bind DNA *in vitro*, we performed native Electro Mobility Shift Assays (EMSA) (Fig. 4.2c). Indeed, recombinant KNL-2 Myb recognized purified human centromeric DNA (180 bp of alpha satellite DNA) (Conde e Silva et al., 2007). By comparing the HsKNL-2 Myb domain structure to a conventional Myb DNA binding domain (Yeast Myb DNA binding domain, pdb #1WOT), it is electrostatically realistic that the HsKNL-2 Myb domain binds DNA in the same manner has other Myb domain families (Fig. 4.2b, 4.2d). Indeed, the HsKNL-2 Myb domain requires its third helix to bind DNA as previously shown by the c-Myb co-structure with DNA binding through its third helix of the third Myb domain repeat (Ogata et al., 1992). Myb domains in other protein families (e.g., transcription factors) bind to specific short DNA motifs (normally 6-8bp) (Biedenkapp et al., 1988; Mizuguchi et al., 1990), thus we next tested the idea that the KNL-2 Myb domain utilizes a sequence based mechanism for centromere recognition. Human centromeres are composed of vast regions of inverted, imperfect repeated of approximately 180bp elements termed alpha-satellite DNA. We generated an artificial array of 20mers (by oligonucleotide annealing) covering the entire 180bp of alpha-satellite DNA, and by EMSA and NMR, were unable to detect preferential binding to any of these fragments (data not shown). Thus, in accordance with the idea that the CLC recognizes the centromere via an epigenetic mechanism, the KNL-2 Myb domain does not bind to a specific DNA sequence found at centromeres.

Given the apparent absence of sequence specificity, we tested the idea that KNL-2 recognizes a structural element unique to CENP-A chromatin. To test this hypothesis, we developed a TIRFm based interaction assay (modified version of SIMpull (Jain et al., 2011; Padeganeh et al., 2013), Fig. 4.3a) to evaluate the specificity of KNL-2 for CENP-A chromatin. In this assay, the coverslip surface of a flow chamber was functionalized with a nanobody that recognizes YFP and GFP. Nuclear extract prepared from randomly cycling cells expressing either CENP-A-YFP or Histone H2B-GFP was first sonicated and then introduced into the flow chamber, thus

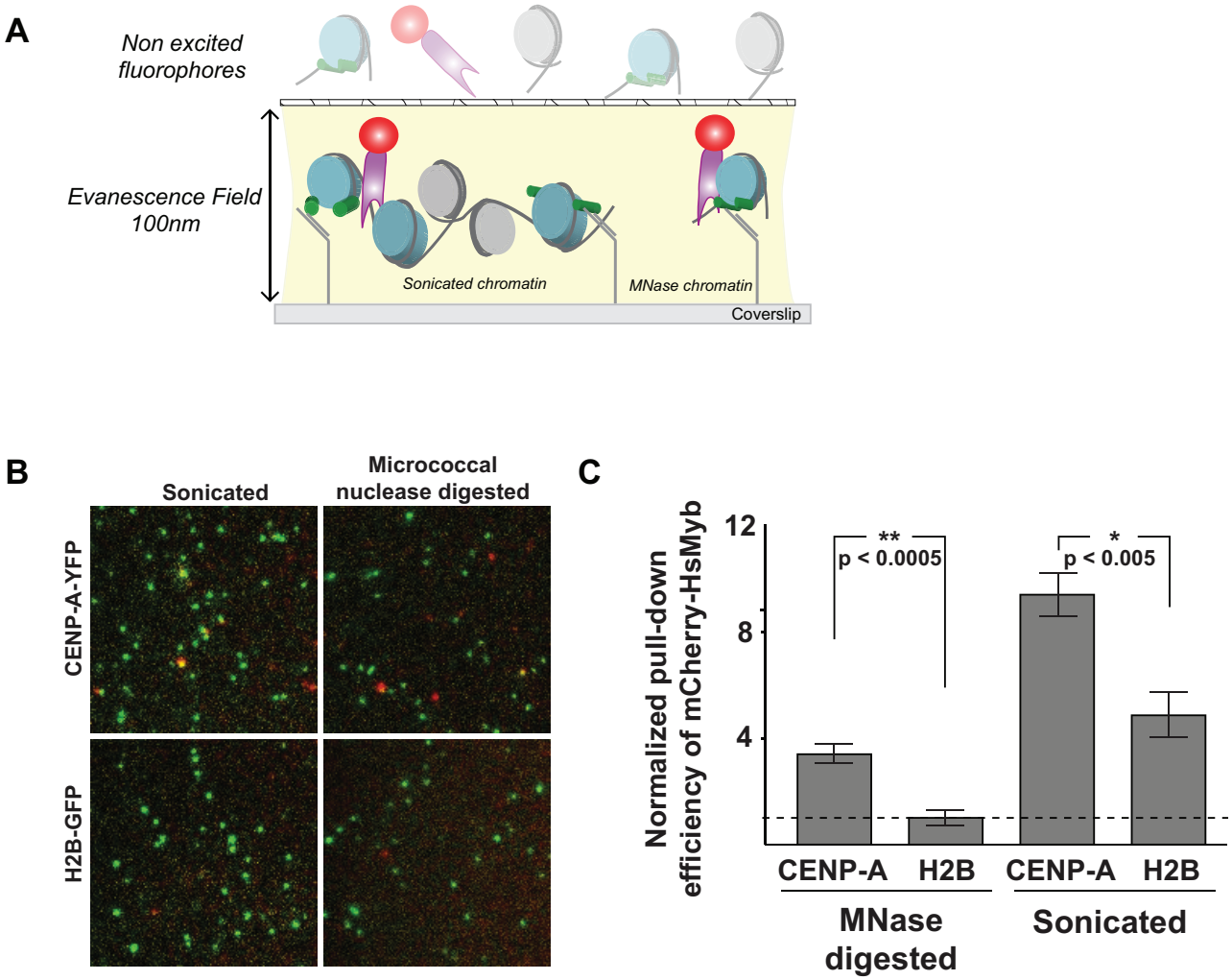



**Figure 4.3 HsKNL-2 Myb domain recognizes and binds specifically to CENP-A chromatin.**

**a**, Schematic representation of the TIRFM assay used to determine the Myb domain binding specificity to centromeric chromatin: a nanobody (single chain antibody; gray) is introduced in a flow chamber, where GFP or YFP (green cylinder), fused to H2B or CENP-A respectively, is trapped by the nanobody. Purified mCherry-Myb domain is introduced in the flow chamber for preferential binding analysis (mCherry: red circle, Myb domain: purple). **b**, Overlay examples of mCherry-HsMyb to either YFP-CENP-A or H2B-GFP chromatin (sonicated or MNase treated). Those microscopy images are then treated and quantified. **c**, Quantification of pull-down efficiency of the HsMyb domain to YFP-CENP-A or H2B-GFP chromatin. Quantification analysis is performed using probabilistic segmentation to measure fluorescence pull-down efficiency of HsMyb domain to CENP-A-YFP and H2B-GFP chromatins; quantification is normalized to histone H2B Mnase intensity. HsMyb domain preferentially binds CENP-A-YFP chromatin compared to H2B-GFP.

immunoprecipitate the chromatin onto the coverglass. After washing, purified recombinant mCherry-tagged HsKNL-2 Myb domain was introduced. Finally, HsKNL-2 Myb domain binding to the immunoprecipitated chromatin was analyzed by TIRF microscopy and images were analysed using probabilistic segmentation (Padeganeh et al., 2013). Interestingly, HsKNL-2 Myb domain showed a clear preference for CENP-A containing chromatin when compared to histone H2B chromatin (Fig. 4.3b and c). Given that KNL-2 has never been reported to interact directly with individual CENP-A nucleosomes (only chromatin), we hypothesized that the interaction was with linker DNA, or proteins binding the linker DNA, and not the histone core particle directly (Maddox et al., 2007). To test this hypothesis, CENP-A-YFP chromatin was pre-treated with Micrococcal Nuclease (MNase) to generate mono-nucleosomes and reduce the linking DNA. As predicted, KNL-2 Myb interaction with MNase treated chromatin was drastically reduced. These results show that the KNL-2 Myb domain binds preferentially to CENP-A chromatin and requires nucleosomal linking DNA stretches.

### 4.4.3 KNL-2 Myb domain recognition of CENP-A chromatin is conserved in metazoans

Several recent studies have suggested that CENP-A nucleosomes generate unconventional DNA topology (Furuyama and Henikoff, 2009; Hasson et al., 2013). Thus, it is possible that the targeting mechanism of KNL-2 is structure rather than DNA sequence based. A structure based epigenetic recognition mechanism might be conserved across phylogeny. *C. elegans* KNL-2 is poorly conserved while being a true functional ortholog (Maddox et al., 2007) (Fig. 4.4a). The DNA sequences that comprise holocentric *C. elegans* centromeres have no apparent sequence-based similarity to human centromeric DNA (Gassmann et al., 2012). Taking advantage of poor sequence-level but absolute functional conservation, we tested the idea of a structural mechanism for centromere recognition by the KNL-2 Myb domain. CeKNL-2

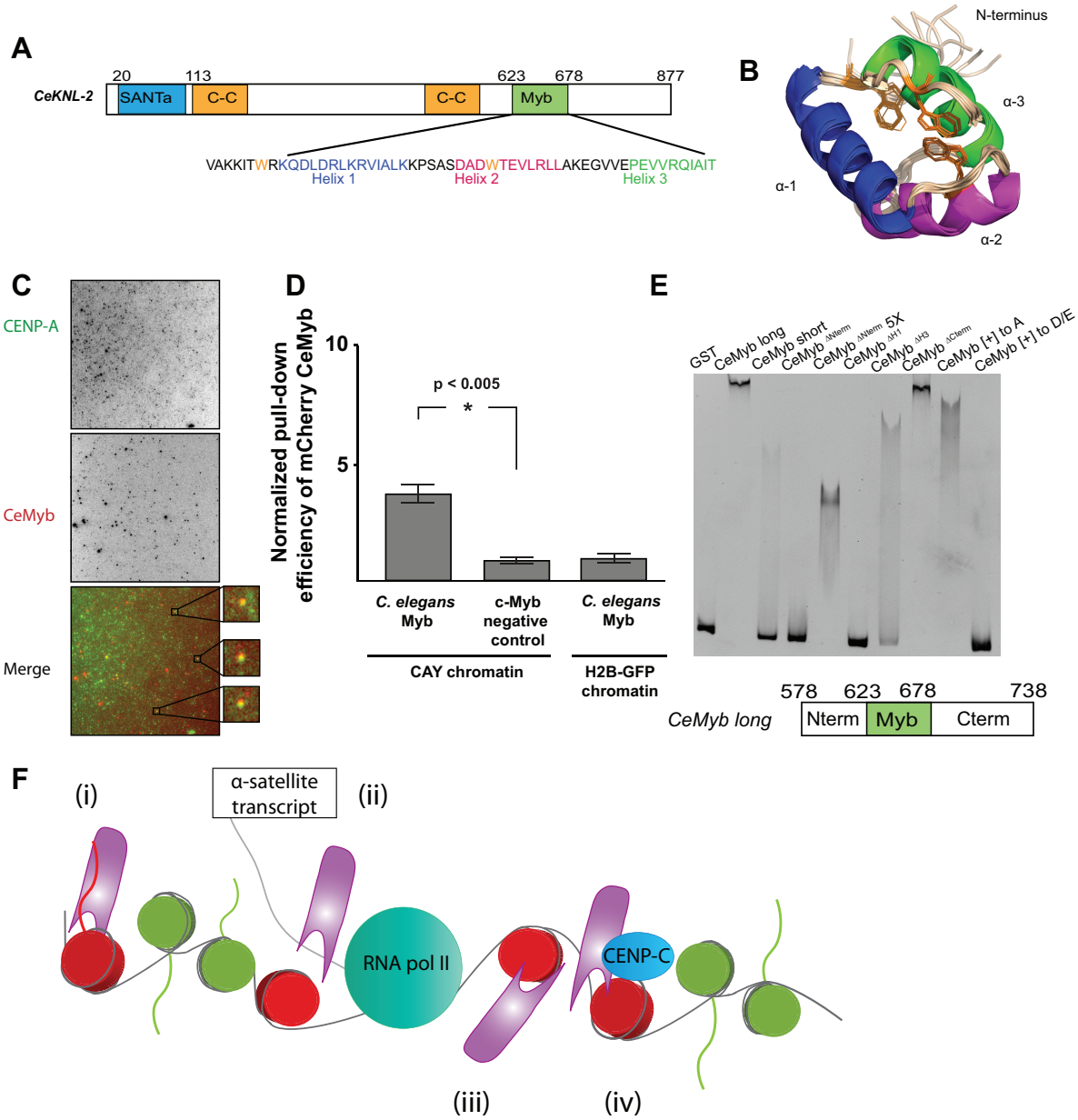



**Figure 4.4 CeKNL-2 Myb domain is binding specifically to human centromeric chromatin.**

**a**, Schematic representation of CeKNL-2 sequence showing predicted Myb domain position at C-terminus and SANTa domain at N-terminus. **b**, Overlay of the 10 lowest energy CeKNL-2 Myb domain structures, showing side chains of tryptophans of the hydrophobic core. Thus, CeKNL-2 Myb domain structure has a helix-turn-helix motif. **c**, Example of TIRF microscopy images overlay of mCherry-CeMyb domain with human CENP-A-YFP chromatin **d**, Quantification analysis performed using probabilistic segmentation and fluorescence pull-down efficiency of the CeMyb domain to CENP-A-YFP or H2B-GFP chromatin from human cells was measured.

CeMyb domain preferentially binds CENP-A-YFP chromatin compared to H2B-GFP. **e**, EMSA of CeKNL-2 Myb domain truncations and mutants performed with 208bp human α-satellite DNA. The N-terminus is required for stable interaction with DNA. **f,** Model of DNA binding preferences of KNL-2 Myb domain to CENP-A chromatin; (i) it may recognize the N-terminus tail of CENP-A or the CATD (CENP-A targeting domain) without the presence of a binding partner, (ii) it may bind to α-satellite transcripts generated by RNAPII passage to centromere, (iii) It may recognize and bind a specific nucleosomal structure given by the presence of CENP-A, (iv) or it can be recruited to centromere through a constitutive centromere protein like CENP-C.

(*C. elegans* KNL-2) contains a degenerate C-terminal Myb type domain (Fig. 4.4a) not predicted by standard bioinformatics (Maddox et al., 2007). To confirm that CeKNL-2 contains a *bone fide* Myb domain, we solved the atomic structure of recombinant CeKNL-2 Myb domain using NMR (Fig. 4.4b). As expected, the CeKNL-2 Myb domain comprises three helixes and a hydrophobic core of tryptophans. We next tested the centromere binding specificity of the CeKNL-2 Myb domain in our TIRFm based assay. Consistent with a conserved structure being the target, CeKNL-2 interacted preferentially to human CENP-A chromatin in a manner very similar to human KNL-2 (Fig. 4.4c and d). Thus, the KNL-2 family Myb domain recognizes a conserved linker-region feature of xenogeneic CENP-A chromatin.

Myb domains are well documented to bind target DNA motifs via insertion of the third helix into the DNA major groove (Ogata et al., 1992). Additionally, residues in the proximal N-terminal of the domain provide stability to the interaction. To determine the mechanism of KNL-2-DNA interaction, we used $^{1}H$-$^{15}N$ Heteronuclear Single Quantum Coherence (HSQC); an NMR-based method that identifies residues whose environment is altered upon ligand binding. After addition of double stranded DNA 15mers to the CeKNL-2 Myb domain, we observed spectral shifts in residues just N-terminal to as well as within the first helix, however there were only minor shifts in the third helix (Fig. 4.S2a and 4.S2b). These results suggested an unconventional binding mechanism of CeKNL-2 Myb domain. To better understand the binding mechanism, we generated a series of point mutants and truncations and tested these for DNA

binding by EMSA. In summary, single point mutants did not affect binding, nor did removing positively charged residues in the first helix (Fig. 4.4e). However, truncation of the N-terminus, removal of the first helix, or replacement of positive for negative charge in the first helix all dramatically inhibited CeKNL-2 Myb DNA binding (Fig. 4.4e). This data, along with the observation that truncation of the third helix did not alter binding, show that the CeKNL-2 Myb DNA interaction is unconventional in nature, and requires both the first helix as well as unstructured regions directly N-terminal to the defined Myb domain.

Comparison of the CeKNL-2 Myb structure with conventional a Myb-DNA binding domain revealed that the CeKNL-2 Myb domain could not bind DNA in the same manner as the second helix would generate steric hindrance (Fig. 4.5a, b, and c). However, as mentioned above, the human KNL-2 Myb domain is more similar in structure and charge to other Myb proteins. Interestingly, EMSA analysis of mutant and truncated forms of the HsKNL-2 Myb domain revealed DNA interactions via the third helix as expected for conventional Myb domains. Unconventionally, a small region C-terminal to and outside of the Myb domain is also required for stable interactions with DNA (Fig. 4.2c). Thus, KNL-2 from *C. elegans* and Human bind centromeric DNA in mechanistically distinct manners (Fig. 4.5b), via structural mechanisms distinct from those of other Myb-domain proteins.

## 4.5 Discussion

Understanding the molecular mechanism of the CLC, and more specifically the role of KNL-2 in the licensing process is crucial for the global understanding of centromere identity. We have shown that the CLC is critical for the retention and segregation of the centromere epigenome, as depletion of KNL-2 blocked loading of new CENP-A

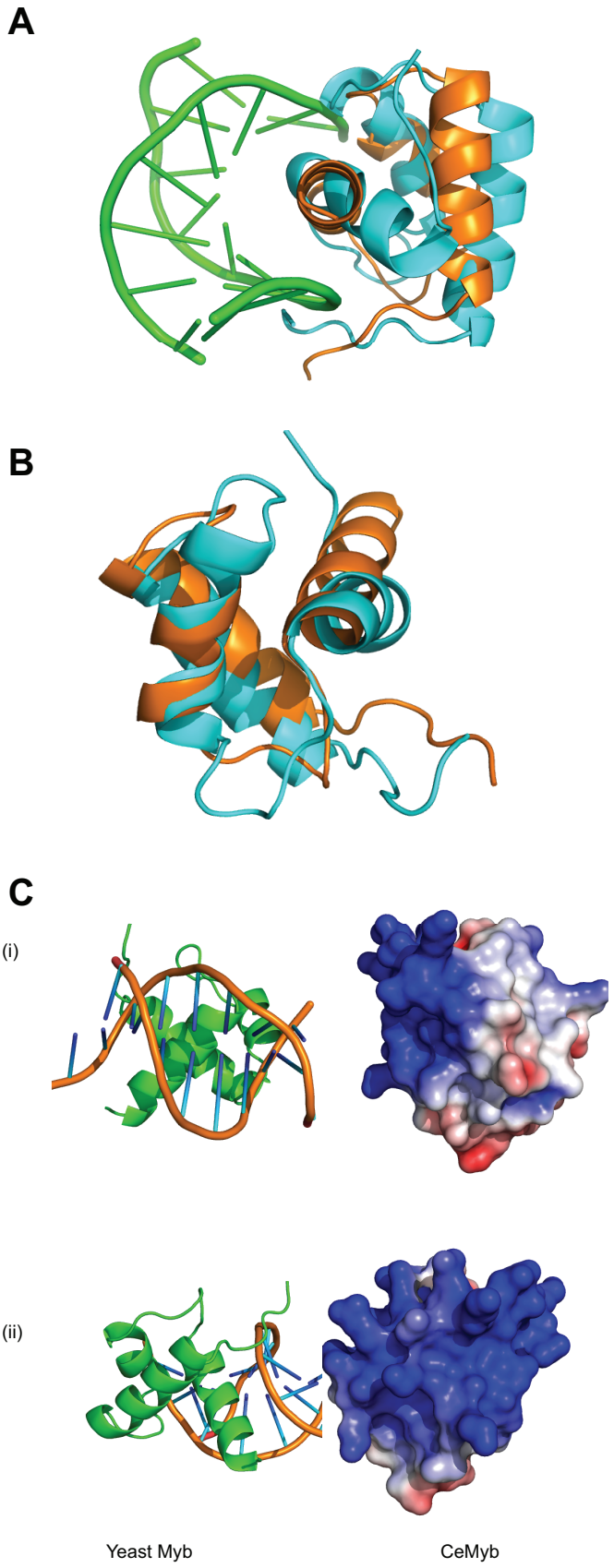



**Figure 4.5 CeKNL-2 Myb domain has a unique binding mechanism.** **a**, Overlay of CeKNL-2 Myb domain (cyan) with Yeast Myb domain (orange) as Myb-DNA complex (pdb 1WOT). The extended second helix in CeMyb structure clashes with DNA. **b**, Overlay of HsKNL-2 Myb domain (orange, pdb #1WGX) and CeKNL-2 Myb domain (cyan). Proteins are superimposed based on the helical elements (RMSD 1.5Å). Distinct differences can be observed between the structures: the second helix is extended and the orientations of the helices are different. **c**, Two different orientations (*i* and *ii*) of Yeast-Myb-DNA structure and CeMyb electrostatic potential maps.

and caused loss of pre-existing CENP-A. We propose that loss of pre-existing CENP-A is due to inappropriate activity of an as yet unidentified centromere maintenance complex. Structural dynamics analysis revealed that centromere targeting of the CLC requires recognition of a CENP-A-generated structural element by a C-terminal Myb type DNA binding domain. The structural recognition of CENP-A is conserved in

metazoan as CeKNL-2 Myb domain preferentially bound to human centromeric chromatin. While the centromere epigenetic mark is seemingly conserved across phylogeny, this interaction is atypical for the Myb domain family in two key mechanisms. First, unlike conventional Myb domains, the KNL-2 Myb domain does not recognize a sequence motif. Secondly, CeKNL-2 Myb binds DNA via its first and not its third helix as is the case for all known Myb domains and for HsKNL-2. The HsKNL-2 Myb domain requires C-terminal adjacent residues to stably interact with DNA, while in other Myb domains the N-terminal flanking residues play this role. Thus, we have uncovered novel diversity within the vast Myb family, and have identified a molecular mechanism of CLC centromeric interpretation of epigenetic identity. Future work on the identity of the higher-order DNA structure recognized by KNL-2 is necessary to reveal the identity of the epigenetic mark regulating centromeres.

## 4.6 Experimental Procedures

### 4.6.1 Cell Culture

The HeLa cell lines, which express EGFP-CENP-A, YFP-CENP-A, and Mis18-GFP, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin-streptomycin, and 6μg/ml puromycin (Invitrogen, Burlington, ON, Canada) at 37°C in a humidified 5% CO2–95% air incubator (Maddox et al., 2007).

### 4.6.2 Live-cell imaging

Cells were seeded in dishes (MatTek, MatTek Corporation, Ashland, MA, USA) and imaged in Gibco's CO2-independent media (Invitrogen) supplemented with 10% FCS, 4.5 g/l glucose (Invitrogen). To avoid drying out, dishes were filled to the rim with median (9.5ml) and sealed with vasoline. Time points comprised of 31 1024 by 1024 pixel z sections spaced 0.5 μm were acquired every 15 min for 24-72 hours with a 60× 1.42 NA planApo objective lens (Olympus) on a DeltaVision microscope fitted with a 37°C environmental chamber using Softworx software (Applied Precision, Issaquah, WA, USA) equipped with a CoolSnap HQ2 camera (Photometrics, Tucson, AZ, USA).

For presentation, 3D frames were protected along the z, taking, for every pixel, the difference between the maximum and the median along z (max-med projection) in order to emphasize spot features. Where possible, the frames were overlaid with a brightfield image of the cells.

### 4.6.3 Image and data analysis

All image analysis was carried out using software developed with Matlab (The MathWorks, Natick, MA, USA; release 2008a or newer).

To measure live-cell fluorescent EGFP-CENP-A intensities, we extended the algorithm presented in Lagana, Dorn et al., 2010 (Lagana et al., 2010). On each microscopy movie, the algorithm performed 5 tasks: (1) detection and tracking of individual nuclei, (2) initial spot detection, (3) cell cycle classification, (4) super-resolution spot fitting, (5) bleach correction and estimation of CENP-A loading dynamics.

*Task 1: Detection and tracking of individual nuclei*

To identify nuclei, we first averaged all 31 z-slices onto a single slice, since the average projection preserves background intensity. This single slice was band-pass filtered to emphasize objects of roughly the size of centromere clusters (Jaqaman et al., 2010): first, it was filtered with a Gaussian filter with radius 1.25 μm, then with a Gaussian filter with radius 2.5 μm, and the second image was subtracted from the first one, after which it was thresholded by setting all the positive differences to one, and all the negative differences to zero. Connected groups of pixels with values above zeros in the thresholded image were considered potential locations of nuclei. If two nuclei were positioned closely next to one another, it was possible that their signals in the thresholded image touched. Thus, we performed morphological erosion on the thresholded image with a disk of radius 0.8 μm, and if a connected group of pixels thus separated into two similarly sized groups, we considered them as two separate cells. Since the GFP intensity is not homogeneously distributed in the nucleus, the potential nuclei did not have a smooth exterior. Thus, we replaced the potential nuclei with their filled convex hull, and performed a morphological opening with a disk of radius 2.5 μm to eliminate corners from the convex hull. This approach reliably resulted in a mask that was ~1μm wider than the nucleus, thus including the signal of all centromeres.

After nuclei had been identified for each time point, we applied the uTrack algorithm (Jaqaman et al., 2008) to track the nuclei through time. The uTrack algorithm was able to identify dividing cells, and it was able to recover tracks even after cells had intermittently moved out of the frame. If the algorithm suggested that two nuclei had

merged temporarily, we considered this as a failure of segmentation, and removed the merged nuclei from our analysis. Furthermore, we removed all nuclei that touched the edge of the image, since this would lead to artifactually low numbers of detected centromeres.

This first analysis step thus resulted, for every nucleus, in a list of masks that indicated the location of the nucleus in every given frame.

*Task 2: Initial spot detection*

For every cell, I applied the algorithm described in Jaqaman et al., 2010 (Jaqaman et al., 2010) to estimate centromere positions. Briefly, the 3D frame was filtered with a band-pass filter to emphasize point-spread function (PSF) sized features. Candidate spots for subsequent fitting of intensities (task 4) were chosen by unimodal thresholding of the signal-to-noise estimate for each local intensity maximum.

Although the initial spot detection was unable to separate closely clustered centromeres, it provided, for each cell, a good estimate of centromere geometry.

*Task 3: Cell cycle classification*

To classify cells into cell cycle states, three measures were used: median z-position of the estimated centromere position, flatness of the nucleus, and nuclear volume.

Median z-position was low during interphase and high during metaphase, when cells round up. After mitotic exit, median z-position slowly decreased, as cells start adhering and flattening again. Using minimum z-positions of all cells on the coverslip, z-position was corrected for a possible tilt of the coverslip.

Flatness of the nucleus was measured as the orientation of the smallest eigenvector of the centromere position distribution. The interphase nucleus was rather flat, which means that the largest variance of centromere positions was in the x-y plane, whereas the direction of the smallest positional variance was parallel to the z-axis.

During metaphase, when the nucleus rounded up, the direction of the smallest positional variance became highly variable, until it stabilized perpendicular to the metaphase plate.

Nuclear volume was measured as the volume of the convex hull around all the centromeres. This measure was a good approximation to the true nuclear volume for most of interphase, when centromeres were positioned at the nuclear envelope (Solovei et al., 2004). During metaphase, the volume of the convex hull was very small, and right after mitotic exit, nuclear volume increased rapidly as the nucleus was being reorganized. Following this rapid growth, nuclear volume grew more slowly, but steadily in a linear fashion until the following mitosis.

Mitosis was thus identified as the period following a steep drop in nuclear volume when the direction of the smallest positional variance became unstable and the median z-position was high. Mitotic exit was characterized by a steep growth in nuclear volume, decreasing median z-position and stabilization of the direction of the smallest positional variance in z-direction. Recovery duration was set to 12 hours after anaphase onset, and S/G2 was set to start 15 hours after anaphase onset, or up to 20 hours before division.

This automatic classification made it possible to perform automatic fitting, and to run the spot fitting task (task 4) only where needed.

*Task 4: Super-resolution spot fitting*

Super-resolution spot fitting was used to separate, if possible, tightly clustered centromeres, and to improve the intensity estimation by measuring the spot intensity above local background. For super-resolution spot fitting the algorithm described in Dorn et al., 2005 (Dorn et al., 2005) was used. Briefly, this algorithm fits 3D Gaussian intensity profiles that closely match the PSF to the image at the positions estimated in Task 2, and then tests the fitted spots for statistically significant intensity above background and for statistically significant separation between spots. If the spot

passes both tests, the algorithm attempts to fit two intensity profiles into the estimated position, in case there were multiple tightly clustered centromeres at this position. If the jointly fitted spots all pass the statistical tests for significant intensity and separation, and if fitting an additional spot has significantly reduced the residual error in intensities between the fit and the raw data, the algorithm attempts to fit yet another spot, until the additional spot does not bring anymore significant improvements. Since we were not interested in super-resolution positions of metaphase centromeres, and to reduce computational cost, we only fitted single 3D Gaussian intensity profiles into the estimated spot positions in order to improve the intensity estimates.

Super-resolution spot fitting thus allowed separating many centromere clusters and measuring EGFP-CENP-A levels above local background at centromeres with high accuracy.

*Task 5: Bleach correction and estimation of CENP-A loading dynamics*

Exposing fluorophores to light leads to photobleaching. Since this study required accurate intensity measurement, the loss of fluorescence intensity due to photobleaching was a possible source of artifacts and had to be corrected for. An additional source of artificial intensity variation was the change in light intensity emitted by the xenon lamp to excite fluorescence. When the lamp was old (>1000h), intensity was changing by more than 50% over the course of a 60 h movie.

To correct for lamp intensity variation, measured intensities were divided by the total light intensity to which the sample was exposed per frame measured by the built-in photosensor on the microscope and recorded in the header of the microscope files.

To estimate bleaching as a function of light intensity, we first determined the dependence of bleaching rate on light intensity. Since we had learned from our experiments that CENP-A is dynamic throughout the cell cycle, and since we needed a lot (>200) of data points to be able to accurately estimate bleaching rates at light intensities commonly used in our experiments, we weren’t able to image live cells,

since the natural variation in CENP-A levels would otherwise confound our measurement. Therefore, we fixed cells in the MatTek dish with methanol (wash with PBS, fix in methanol for 25 minutes at -20°C, image in PBS) prior to imaging at different exposure times to simulate changing lamp intensities. From these data we estimated that for our experimental conditions, the bleach rate depended linearly on the total amount of light to which the sample was exposed. We finally estimated the bleaching rate for our experiment to scale with the light intensity by a factor of $3*10^{-11}$/timepoint (typical light intensity is on the order of $10^7$, so that after 150 timepoints, fluorescence has typically decayed by 20%.

These corrections allowed us to accurately estimate CENP-A loading dynamics following mitosis, as well as CENP-A maintenance dyanamics in S/G2 using separable nonlinear least squares fitting (based on the function *fminspleas* from the Matlab File Exchange) of a single exponential function with constant.

### 4.6.4 shRNA treatment of cells

Transfection with fusion with shRNA was performed using Lipofectamine 2000 reagent (Invitrogen, Grand Island, NY, USA) following manufacturer recommendations. Briefly, ~400000 cells were seeded in MatTek dishes 16-24 hours before transfection in culture media without puromycin. After two washes with PBS, cells were transfected for 8-10 hours in OptiMem. Imaging (in imaging media as described above) was started 1-2 hours later.

### 4.6.5 Total Internal Reflection Fluorescence Microscopy (TIRFM)

#### 4.6.5.1 Nuclear lysate preparation

Crude nuclear lysates were obtained by resuspending CENP-A-YFP or H2B-GFP expressing cells in a Digitonin lysis buffer supplemented with proteinase inhibitors, followed by dounce homogenization, centrifugation and resuspension, these latter steps repeated three times. Finally, nuclei were pelleted by centrifugation at 4000g, resuspended and kept at -80°C until use.

Isolated nuclear lysates were then diluted in PBST to 3.6 µg/ml and sonicated with a probe sonicator. Aliquots of sonicated nuclei were then treated with 100 units of micrococcal nuclease (Fermentas, Ottawa, Canada), in a digestion buffer consisting of PBS and 5 mM $CaCl_2$ and incubated at 37°C for 30 minutes. The reaction was stopped by adding 30 µl of a 0.5 M EDTA solution and chilling of the reaction tube on ice.

#### 4.6.5.2 Flow chamber

To build a flow chamber, 22 x 22 mm coverslips of 1.5 thickness were treated with previously described Piranha Solution (Labit et al., 2008) and stored in ultrapure water prior to usage. To make an imaging flow chamber, two narrow pieces of double-sided tape were applied to the coverslip, parallel to each other, over which a smaller, 18x18 mm cover slip was placed.

Flow chambers were then placed on microscope, prior to adding solutions, allowing imaging of coverslip at each step, to check for background, noise, etc. PBST was first perfused into the flow chamber. Then, 10 µl of a 1 mg/ml solution of YFP nanobodies was perfused, three times into the flow chamber, and incubated for 10 minutes, followed by 10 µl of 1X PBST to wash. The flow chamber was then blocked with 0.5% BSA in PBS, followed by several PBST washes. 10 µl of digested or sonicated nuclei was then perfused three times, incubated for 10 minutes and eventually washed with 1X PBS. Then, a 0.1 nM solution of HsMyb-mCherry was added to the flow chamber, and followed by PBS washes.

Images were then analysed offline. Pull-down efficiency analysis was performed in Matlab by applying probabilistic segmentation to images, and counting overlapping spots. Photobleach counting analysis was performed manually in ImageJ, by selecting regions of interest (ROI) surrounding CENP-A-YFP spots, and plotting

intensity fluctuation over time, which reveals the number of molecules that have undergone photobleaching within the ROI.

#### 4.6.5.3 Single-Molecule Microscopy

TIRF imaging was carried out on a Nikon TI Eclipse inverted microscope, equipped with a 100X 1.47NA APO-TIRF objective (Nikon, Melville, NY, USA), illuminated with a 488, 568, and 647nm laser launch (Nikon, Melville, NY, USA). Image acquisition was performed with a Cascade II EMCCD camera (Photometrics, Tucson, AZ, USA), controlled with NIS-Elements software (Nikon). The EMCCD was operated in full frame (unbinned) normal (non-EM) readout mode at the slowest speed to reduce noise.

### 4.6.6 Nuclear Magnetic Resonance Spectroscopy

#### 4.6.6.1 Protein purification

*C. elegans* KNL-2 Myb domain called CeMyb 11 was cloned in pGEX6P-1 vector. Isotopically enriched samples for NMR were prepared from E. coli BL-21 cells grown on minimal M9 media containing 1g of [$^{15}$N] ammonium chloride with or without 2g of [$^{13}C_6$] glucose, 2mM $MgSO_4$, 0.1mM $CaCl_2$, and 10mL of Vitamin Mix (Sigma, B6891) (Cambridge Isotopes Laboratory, Andover, MA). The protein was eluted using PreScission protease (GE Healthcare, Mississauga, ON, Canada).

CeMyb11:

**GPLGS**VAKKITWRKQDLDRLKRVIALKKPSASDADWTEVLRLLAKEGVVEPEVVRQI AITRLKWVEP

#### 4.6.6.2 NMR spectroscopy

NMR data were collected at 25º C on a 600 MHz Varian Inova spectrometer on uniformly $^{15}$N/$^{13}$C labeled **CeMyb11** samples in 20 mM sodium phosphate buffer, 100 mM NaCl, 5mM $MgSO_4$, 2mM DTT, 0.02% $NaN_3$ at pH 6.5 in 90% $H_2O$/10% $D_2O$.

Standard 3D triple resonance experiments for backbone and side chain assignments were acquired using Biopack. Distance restraints for structure calculations were obtained from 3D $^{15}$N and $^{13}$C-edited NOESY spectra. Dihedral restraints were obtained from TALOS+ (Shen et al., 2009). All NMR data were processed with NMRPipe (Delaglio et al., 1995) and analysed with NMRView (Johnson and Blevins, 1994). Structures were generated from distance and dihedral restraints using the autoassign module in CYANA2.1 (Herrmann et al., 2002). Structural statistics for the 10 lowest energy structures are given are given in Table S-I. Structural coordinates and NMR chemical shifts have been deposited to the PDB (**accession number 2m3a**).

Chemical shift mapping of **CeMyb11** was obtained by comparison of $^{1}$H-$^{15}$N HSQC spectra of 25 μM $^{15}$N-labelled **MYB11** in the absence and presence of 1mM **VDR59 dsDNA**. Both samples were in identical buffers (see buffer used for NMR) and care was taken to maintain identical pH.
VDR59 sequence:
5'- GTTGAACGATCCTTT-3'
3'- CAACTTGCTAGGAAA-5'

### 4.6.7 Proteins purification

Proteins of interest were cloned in pGEX6P-1, resulting in proteins possessing a GST tag as well as a PreScission cleavage site. Proteins fused to GST were expressed in *E. coli* BL-21 strain and were purified using Glutathione beads (GE Healthcare, Mississauga, ON, Canada). Standard protocol was used for purification and proteins were eluted either using excess of 10mM Glutathione or cleaved GST-tag with PreScission protease (GE Healthcare, Mississauga, ON, Canada).

HsMyb long:

*GPLGSP*ARPSVKETLQKSGVRKEFPITEAVGSDKTNRHPLECLPGLIQDKE**WNEKE**
**LQKLHCAFASLPKHKPGFWSEVAAAVGSRSPEECQRKYME**NPRGKGSQKHVTKK
KPANSKGQNGKRGDADQKQTIKITAKVGTLKRKQQMREFLEQLPKDDHD

(*Myb domain in bold, PreScission site in italic*)

HsMybΔNterm:

*GPLGSP***WNEKELQKLHCAFASLPKHKPGFWSEVAAAVGSRSPEECQRKYME**NP
RGKGSQKHVTKKKPANSKGQNGKRGDADQKQTIKITAKVGTLKRKQQMREFLEQL
PKDDHD

HsMybΔH1:

*GPLGSP*EFLPKHKPG**FWSEVAAAVGSRSPEECQRKYME**NPRGKGSQKHVTKKKP
ANSKGQNGKRGDADQKQTIKITAKVGTLKRKQQMREFLEQLPKDDHD

HsMybΔH3:

*GPLGSP*EFARPSVKETLQKSGVRKEFPITEAVGSDKTNRHPLECLPGLIQDKE**WNE**
**KELQKLHCAFAS**LPKHKPG**FWSEVAAA**VGSRS

HsMybΔCterm:

*GPLGSP*EFARPSVKETLQKSGVRKEFPITEAVGSDKTNRHPLECLPGLIQDKE**WNE**
**KELQKLHCAFAS**LPKHKPG**FWSEVAAA**VGSRS**PEECQRKYME**

CeMyb long:

*GPLGSP*NSTKSPAQARKKKRASLEDNRDLNDSIACNRPRRSCVTPVAKKIT**WRKQD**
**LDRLKRVIALKKPSA*SDAD*WTEVLRLLAKEGVVEPEVVRQIAITRLKWVE**PEQNEE
VLKQVEEVEQKRRRGAVARVKENVKMHEELREGGNHRAEDLQSGVESMEDYQPE
DVAA

(*Myb domain in bold, PreScission site in italic*)

**CeMyb**ΔNterm:

*GPLGSP***WRKQDLDRLKRVIALKKPSA*SDAD*WTEVLRLLAKEGVVEPEVVRQIAITR LKWVE**PEQNEEVLKQVEEVEQKRRRGAVARVKENVKMHEELREGGNHRAEDLQS GVESMEDYQPEDVAA

**CeMyb**ΔH1:

*GPLGSP***DADWTEVLRLLAKEGVVEPEVVRQIAITRLKWVE**PEQNEEVLKQVEEVEQ KRRRGAVARVKENVKMHEELREGGNHRAEDLQSGVESMEDYQPEDVAA

**CeMyb**ΔH3:

*GPLGSP*NSTKSPAQARKKKRASLEDNRDLNDSIACNRPRRSCVTPVAKKIT**WRKQD LDRLKRVIALKKPSA*SDAD*WTEVLRLLAKEGVVE**

**CeMyb**ΔCterm:

*GPLGSP*NSTKSPAQARKKKRASLEDNRDLNDSIACNRPRRSCVTPVAKKIT**WRKQD LDRLKRVIALKKPSA*SDAD*WTEVLRLLAKEGVVEPEVVRQIAITRLKWVE**

### 4.6.8 Electrophoretic Mobility Shift Assay (EMSA)

Protein and DNA were incubated in a binding buffer (40mM Tris, pH7.5, 1mM EDTA, 250mM NaCl, 5mM $MgCl_2$). Protein and DNA complexes were loaded using a loading buffer that does not contained SDS, and migrated on a native 5% acrylamide/bis-acrylamide gel 29:1 (Bio-Rad, Hercules, CA, USA), with running buffer (0.29M Glycine, 45mM Tris-base). The alpha-satellite DNA used was amplified from pUCαNCS (Conde e Silva et al., 2007) and purified on column. Affinity for different 20mers sequences (coming from alpha-satellite sequence) was determined using dsDNA. Those dsDNA were obtained by annealing complementary pairs of oligonucleotides using Thermo Scientific protocol (TR0045.1, TechTip #45).

Alpha-satellite DNA sequence:

5’ACCCCTTTGAGGCcttcgttggaaacgggaTTTCTTCATATTATGCTAGACAGAATAATT
CTCAGTAACTTCCCTGTGTTGTGTGTATTCAATTCACAGAGTTGAACGATCCTTTA
CAGAGAGCAGACTTGAAACACTCTTTTTGTGGAATTTGCAAGTGGAGATTTCAGC
CGCTTTGAGTTCAATGGTAGAATAGGAAATATCTTCC-3’

(*CENP-B box in lower case*)

### 4.6.9 qPCR

qPCR analysis were performed on asynchronous cells transfected for 8 jours with shRNA plasmids (Sigma Aldrich, St. Louis, MO, USA). Transfected CENP-A-YFP HeLa cells were disrupted and homogenized using RNeasy Kit and QIAshredder (QIAGEN, Mississauga, ON, Canada) and total RNA was isolated according to the manufacturer's instructions. Two µg of RNA were reverse transcribed using High-Capacity cDNA Reverse Transcription Kits (Applied Biosystems, Foster City, CA, USA). Gene expression was assessed by qPCR using BioRad system together with SYBR green JumpStart Taq ReadyMix (Sigma Aldrich, St. Louis, MO, USA). Reference gene assays used is GAPDH.

## 4.7 References


Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229–243.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E., and Klempnauer, K.H. (1988). Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature *335*, 835–837.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V., and Cleveland, D.W. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol Cell *25*, 309–322.

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

Bodor, D.L., Valente, L. P., Mata, J.F., Black, B.E., and Jansen, L.E. (2013). Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. Mol Biol Cell *24*, 923–932.

Brinkley, B.R., and Stubblefield, E. (1966). The fine structure of the kinetochore of a mammalian cell *in vitro*. Chromosoma *19*, 28–43.

Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., and Henikoff, S. (1999). A histone-H3-like protein in C. elegans. Nature *401*, 547–548.

Conde e Silva, N., Black, B.E., Sivolob, A., Filipski, J., Cleveland, D.W., and Prunell, A. (2007). CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. J Mol Biol *370*, 555–573.

Conti, C., Sacca, B., Herrick, J., Lalou, C., Pommier, Y., and Bensimon, A. (2007). Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. Mol Biol Cell *18*, 3059–3067.

De Rop, V., Padeganeh, A., and Maddox, P.S. (2012). CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. Chromosoma *121*, 527–538.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277–293.

Dorn, J.F., Jaqaman, K., Rines, D.R., Jelson, G.S., Sorger, P.K., and Danuser, G. (2005). Yeast Kinetochore Microtubule Dynamics Analyzed by High-Resolution

Three-Dimensional Microscopy. Biophys J *89*, 2835–2854.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

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

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

Furuyama, T., and Henikoff, S. (2009). Centromeric nucleosomes induce positive DNA supercoils. Cell *138*, 104–113.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. Cell *145*, 410–422.

Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., et al. (2012). An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. Nature *484*, 534–537.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Ben E Black (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat Struct Mol Biol *20*, 1–10.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell *118*, 715–729.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. J Mol Biol *319*, 209–227.

Jain, A., Liu, R., Ramani, B., Arauz, E., Ishitsuka, Y., Ragunathan, K., Park, J., Chen, J., Xiang, Y.K., and Ha, T. (2011). Probing cellular protein complexes using single-molecule pull-down. Nature *473*, 484–488.

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

Jaqaman, K., king, E.M., Amaro, A.C., JR, W., Dorn, J.F., Elliott, H.L., Mchedlishvili,

N., McClelland, S.E., Porter, I.M., Posch, M., et al. (2010). Kinetochore alignment within the metaphase plate is regulated by centromere stiffness and microtubule depolymerases. J Cell Biol *188*, 665.

Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L., and Danuser, G. (2008). Robust single-particle tracking in live-cell time-lapse sequences. Nat Methods *5*, 695–702.

Johnson, B.A., and Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR *4*, 603–614.

Labit, H., Goldar, A., Guilbaud, G., Douarche, C., Hyrien, O., and Martheineke, K. (2008). A simple and optimized method of producing silanized surfaces for FISH and replication mapping on combed DNA fibers. Bitechniques *45*, 649–652–654–656–658.

Lagana, A., Dorn, J.F., De Rop, V., Ladouceur, A.M., Maddox, A.S., and Maddox, P.S. (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat. Cell Biol. *12*, 1186–1193.

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., Demidov, D., Schubert, V., and Schubert, I. (2013). Arabidopsis KINETOCHORE NULL2 Is an Upstream Component for Centromeric Histone H3 Variant cenH3 Deposition at Centromeres. Plant Cell *25*, 3389–3404.

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol *176*, 757–763.

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

Mizuguchi, G., Nakagoshi, H., Nagase, T., Nomura, N., Date, T., Ueno, Y., and Ishii, S. (1990). DNA binding activity and transcriptional activator function of the human B-myb protein compared with c-MYB. J Biol Chem *265*, 9280–9284.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc Natl Acad Sci U S A *89*, 6428–6432.

Padeganeh, A., Ryan, J., Boisvert, J., Ladouceur, A.-M., Dorn, J.F., and Maddox, P.S. (2013). Octameric CENP-A Nucleosomes Are Present at Human Centromeres throughout the Cell Cycle. Curr Biol *23*, 1–6.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991).

Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A *88*, 3734–3738.

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

Robbins, E., and Gonatas, N.K. (1964). The Ultrastructure of a Mammalian Cell during the Mitotic Cycle. J Cell Biol *21*, 429–463.

Schuh, M., Lehner, C.F., and Heidmann, S. (2007). Incorporation of Drosophila CID/CENP-A and CENP-C into centromeres during early embryonic anaphase. Curr Biol *17*, 237–243.

Shelby, R.D., Monier, K., and Sullivan, K.F. (2000). Chromatin assembly at kinetochores is uncoupled from DNA replication. J Cell Biol *151*, 1113–1118.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR *44*, 213–223.

Silva, M.C., Bodor, D.L., Stellfox, M.E., Martins, N.M., Hochegger, H., Foltz, D.R., and Jansen, L.E. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev Cell *22*, 52–63.

Solovei, I., Schermelleh, L., D ring, K., Engelhardt, A., Stein, S., Cremer, C., and Cremer, T. (2004). Differences in centromere positioning of cycling and postmitotic human cell types. Chromosoma *112*.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev *9*, 573–586.

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581–592.

Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 Centromere Connector for Localizing a CENP-A-Like Protein in Fission Yeast. Science *288*, 2215–2219.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993). Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. Proc Natl Acad Sci U S A *90*, 9320–9324.

Vafa, O., and Sullivan, K.F. (1997). Chromatin containing CENP-A and alpha-satellite DNA is a major component of the inner kinetochore plate. Curr Biol *7*, 897–900.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci *114*, 3529–3542.

Wang, J., Liu, X., Dou, Z., Chen, L., Jiang, H., Fu, C., Fu, G., Liu, D., Zhang, J., Zhu, T., et al. (2014). Mitotic regulator Mis18β interacts with and specifies the centromeric assembly of molecular chaperone HJURP. J Biol Chem.

Warburton, P.E., and Cooke, H.J. (1997). Hamster chromosomes containing amplified human alpha-satellite DNA show delayed sister chromatid separation in the absence of de novo kinetochore formation. Chromosoma *106*, 149–159.

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. Proc Natl Acad Sci U S A *97*, 7266–7271.

Zasadzińska, E., Barnhart-Dailey, M.C., Kuich, P.H.J.L., and Foltz, D.R. (2013). Dimerization of the CENP-A assembly factor HJURP is required for centromeric nucleosome deposition. Embo J 1–12.

## 4.8 Acknowledgements

We would like to thank the Maddox labmembers for their support and their ideas, especially A. Lagana, and É. Rachiele-Tremblay. Also, we thank J. Filipski and A. Prunell for their gift for an α-satellite containing plasmid, and B.E. Black for is help with the α-satellite DNA purification. V. Villeneuve, G. Sylvain-Drolet, S. Gravel, and D. D'Amours for discussions. The authors declare neither conflict of interest nor competing financial interests in relation to the work described. VDR was supported by the Frederick Banting and Charles Best master scholarship from Canadian Institutes of Health Research, JFD was supported by a postdoctoral fellowship from the Swiss National Science Foundation, and PSM is the Canada Research Chair in Cell Division and Chromosomal Organization and supported by research grants from the Canadian Institutes of Health Research (MOP-106548) and Canadian Cancer Society Research Institute (700824).

Atomic coordinates and chemical shifts for the reported Nuclear Magnetic Resonance structure of CeKNL-2 Myb domain have been deposited with the **Protein Data Bank** under accession number **2m3a**.

# CHAPTER 5

Manuscript in preparation

# KNL-2 IS A BINDING PARTNER OF RNA PROCESSING PROTEINS, HAVING A MYB DOMAIN BINDING TO SSDNA

## 5.1 Preface

This paper is a follow up of the KNL-2 Myb domain biochemical and cell biology characterization. The exact recognition and binding mechanism of KNL-2 Myb domain to CENP-A chromatin remains elusive. Results presented in this paper suggest a potential binding preference of KNL-2 Myb domain to RNA. I conceived and performed all experiments. Corentin Moevus performed the Myb deletion cloning. Dr. Paul S. Maddox and I wrote the manuscript.

**KNL-2 is a binding partner of RNA processing proteins, having a Myb domain binding to ssDNA**

Authors: Valérie De Rop[1], Corentin Moevus[2], Paul S. Maddox[1,3]

Affiliations:
[1] Institut de Recherche en Immunologie et en Cancérologie (IRIC), Département de Pathologie et de Biologie Cellulaire, Université de Montréal, Montréal.
[2] Current address: Department of Pathology and Cell Biology, Columbia University, New-York.
[3] Current address: Department of Biology, University of North Carolina, Chapel Hill.

Correspondence to
Institute for Research in Immunology and Cancer (IRIC),
Department of Pathology and Cell Biology
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7 CANADA

## 5.2 Abstract

Centromere identity propagation in G1 is crucial for the subsequent cell division. The centromeric epigenetic mark, CENP-A, is diluted at each S phase since this protein is only expressed in G2. In order to maintain centromeric identity, newly synthesized CENP-A has to be incorporated through a replication independent mechanism in G1. KNL-2 was demonstrated to be the most upstream protein required for CENP-A incorporation. This protein has a Myb DNA binding domain, and it preferentially binds to CENP-A containing chromatin compared to histone H2B. However, the exact recognition and binding mechanism have not been shown. Here, KNL-2 Myb domain binding to both dsDNA and ssDNA is demonstrated as well as the co-localization of this domain with PML bodies, a known nuclear feature containing RNA. Thus, KNL-2 Myb domain might regulate CENP-A incorporation to the centromere by recognizing and binding to the centromeric transcripts.

## 5.3 Introduction

The primary constriction of mitotic chromosomes is the centromeric region where the megadalton protein complex called the kinetochore assembles. This forms a trilaminar plate, which can be observed under the electron microscope, and contains various functional proteins required for chromosome attachments to the mitotic spindle (Cheeseman and Desai, 2008; Maiato, 2004; Rieder, 1979; Robbins and Gonatas, 1964). Centromeres are epigenetically defined by a histone H3 variant called CENP-A (centromere protein A) that has been shown to be sufficient and essential for centromere function, whereas human centromeric DNA sequences have no known role in the centromere identity (Gascoigne et al., 2011; Mendiburo et al., 2011; Palmer et al., 1991; 1987; Sullivan et al., 1994; Yoda et al., 2000). Thus, centromeres are epigenetically defined by CENP-A (Bodor et al., 2013; Haaf et al., 1992; Mendiburo et al., 2011; Van Hooser et al., 2001; Warburton et al., 1997).

In human cells, CENP-A is expressed in G2 and is incorporated through a replication independent mechanism in early G1. The molecular mechanism of CENP-A incorporation to the centromere is divided in three specific steps: licensing, loading, and maintenance (De Rop et al., 2012). The CLC complex, containing KNL-2/M18BP1, Mis18α, and Mis18β proteins, performs the licensing step. The CLC complex components are essential for CENP-A localization to the centromeres, since depleted cells have reduced amounts of CENP-A at the centromeres (Fujita et al., 2007; Maddox et al., 2007). KNL-2 was shown to be the most upstream CLC component regulator for CENP-A incorporation to the centromere by binding the centromeric chromatin through its Myb DNA binding domain (Barnhart et al., 2011; De Rop et al., 2014; Fujita et al., 2007; Maddox et al., 2007). However, the recognition and binding mechanisms are not fully understood.

In the last decades, many lines of evidence have shown a possible role for RNA in the incorporation mechanism of newly synthesized CENP-A to the centromeres. In

fact, the RNAi pathway components and RNAPII are essential for Cnp1 localization to the centromere in *S. pombe* (Chan et al., 2011; Choi et al., 2011; Folco et al., 2008; Gent and Dawe, 2012; Lejeune et al., 2011; Volpe et al., 2002). Transcription of non-coding RNA by RNAPII will result in the recruitment of the RNAi machinery to the centromere, as well as other regulatory proteins all required for Cnp1 incorporation to the centromeres (Chan et al., 2011; Choi et al., 2011; Lejeune et al., 2011). In human cells, RNAPII was shown to co-localize with CENP-C, an inner kinetochore protein known to have a RNA binding domain (Du et al., 2010; Wong et al., 2007). However, no direct role was attributed to RNA for CENP-A localization to the centromeres in human cells.

Here, we show that the KNL-2 Myb domain binds both ssDNA and dsDNA. Immunofluorescence experiments show limited but significant HsKNL-2 co-localizing with RNAPII in G1, demonstrating the possible binding of KNL-2 to RNA *in vivo*. The KNL-2 Myb domain alone co-localizes with PML bodies when expressed in human cells. All together, these observations depict part of the mechanism of how KNL-2 recognizes and binds centromeric chromatin via its Myb binding domain.

## 5.4 Results

### 5.4.1 HsKNL-2 has RNA binding, processing and splicing factors as binding partners.

To better understand the mode of action of HsKNL-2 Myb domain, proteins, shown to interact with KNL-2 through immunoprecipitation experiments, were clustered by Gene Ontology (GO) terms using the STRAP program (Supplementary data, Table 5.S1) (Bhatia et al., 2009; Lagana et al., 2010). The results are illustrated in a pie chart (Figure 5.1). With the known role of KNL-2 being involved in centromere identity propagation, as expected, most of its binding partners are located in the nucleus. Indeed, the second major GO term indicates that KNL-2 binding partners are located in the cytoplasm. Interestingly, a small subset of binding partners is associated to the ribosome (6%). Within this group, there are proteins involved in the RNA binding, RNA processing and RNA splicing. RNAPII was also found in the HsKNL-2 immunoprecipitate, and this is of high interest since RNAPII was shown to co-localize with CENP-C, a known inner kinetochore protein that might be involved in the newly synthesized CENP-A incorporation to the centromere. To confirm the possibility that KNL-2 might bind RNAPII *in vivo*, immunofluorescence was performed on HeLa cells transfected with endogenous HsKNL-2-mCherry (Figure 5.2). Cells in anaphase or in telophase did not show any co-localization of RNAPII with HsKNL-2. In G1, some HsKNL-2 foci co-localized with RNAPII, having the majority of HsKNL-2 spots not co-localizing with RNAPII. Thus, RNAPII co-localization with KNL-2 may be transient and rapid in G1.

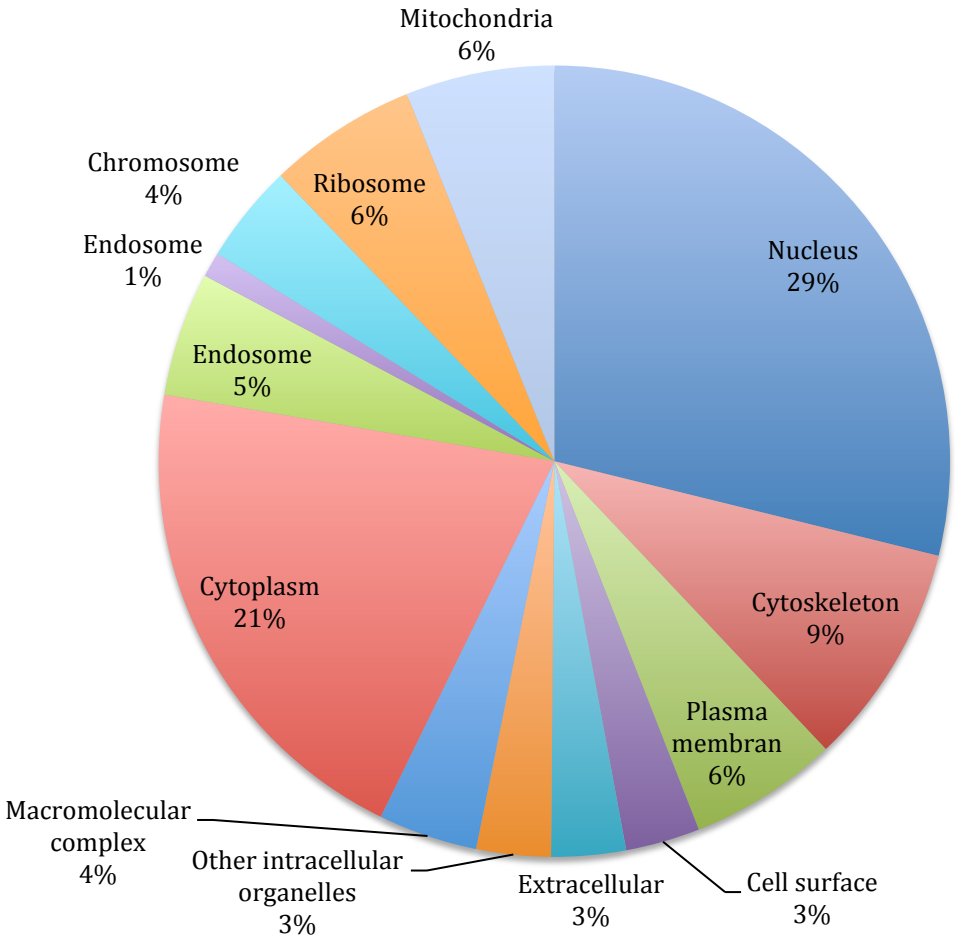



**Figure 5.1 KNL-2 binding partners clustered by GO term.**
Potential binding partners of KNL-2 identified by immunoprecipitation assay and mass spectrometry analysis are clustered by their Gene Ontology (GO) term using STRAP program (Bhatia et al., 2009). Interestingly, some binding partners have a function in ribosomal regulation including RNA binding, processing and splicing factors (Supplementary data, Table 5.S1).

### 5.4.2 HsKNL-2 Myb domain binds ssDNA

The previous observations lead to a specific question whether the HsKNL-2 Myb domain is able to bind RNA or ssDNA *in vitro*. Electro Mobility Shift Assay (EMSA) was performed using bacterially expressed and purified HsKNL-2 Myb domain together with different sizes of ssDNA coupled to a 488nm excitable fluorophore (Figure 5.3). By comparing the fluorescence intensities at the bottom of the gel (free DNA, unbound by proteins) of DNA with and without the HsMyb domain, fluorescence intensity decreases in presence of the protein. Moreover, increasing the ssDNA size increases the possibilities of the HsMyb domain to bind DNA, therefore showing less fluorescence intensity at the bottom of the gel as the mobility of the ssDNA in the gel decreases. Thus, the HsKNL-2 Myb domain binds ssDNA *in vitro*. Moreover, the CeKNL-2 Myb domain was used as control to compare fluorescence intensity at the bottom of the gel of the HsMyb domain binding to a ssDNA. The native gel experiment shows a higher affinity of the HsMyb domain to ssDNA compared to CeMyb domain (less fluorescence at the bottom for HsMyb domain compared to CeMyb domain for the same molar concentration of proteins). Therefore, both CeKNL-2 and HsKNL-2 Myb domains are able to bind ssDNA *in vitro*.

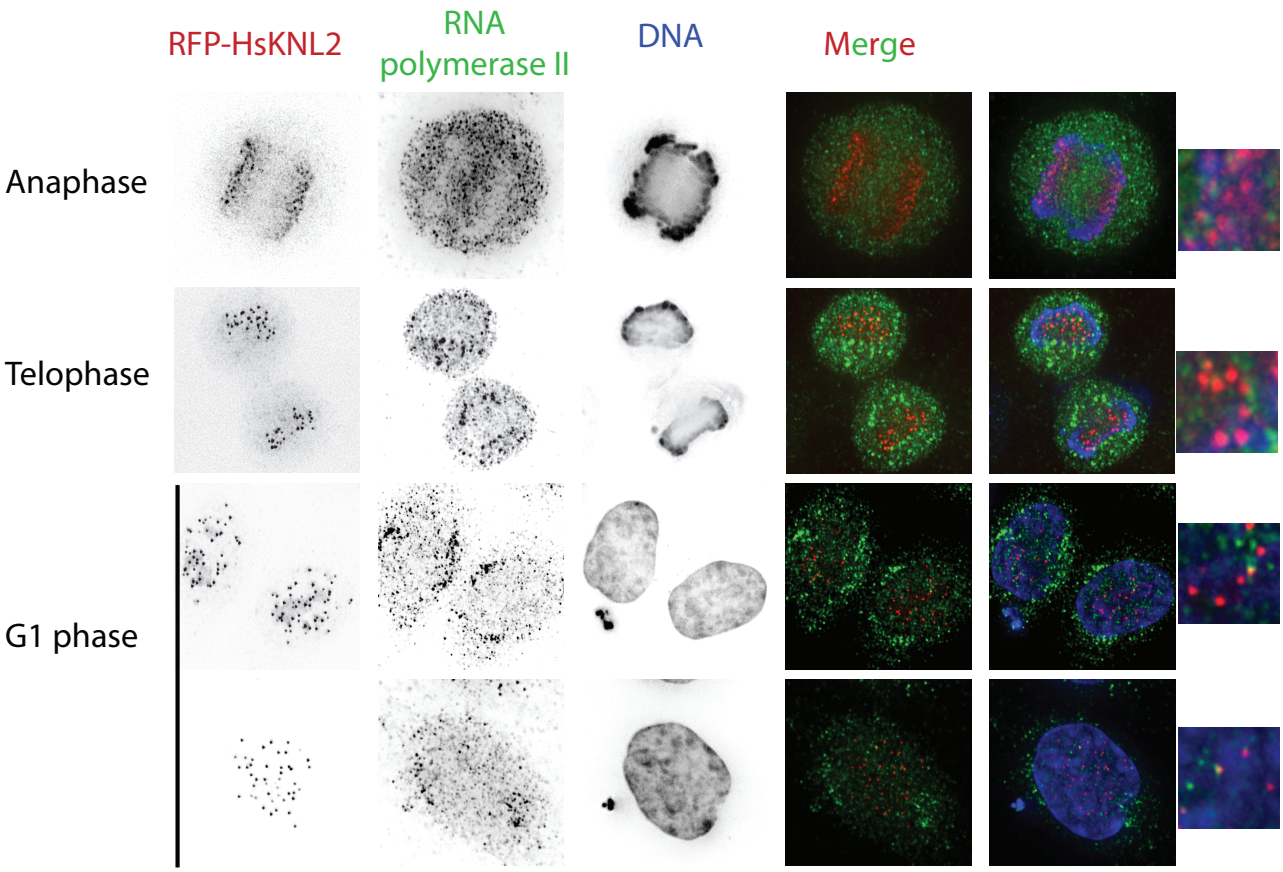



**Figure 5.2 HsKNL-2 slightly co-localized with RNAPII.**
Representative immunofluorescence images of HeLa cells transfected with exogenous HsKNL-2-mCherry and treated with antibodies specific to RNAPII. Right columns are merges from images on the left; first merge on the left represents HsKNL-2-mCherry and RNAPII, second merge represents HsKNL-2-mCherry, RNAPII and DNA in blue, and insets represent enlargement of spots within the nucleus. Three different cell cycle stages are represented in each lane, with two examples of G1 phase cells. Cells in G1 show some co-localization of HsKNL-2-mCherry with RNAPII.

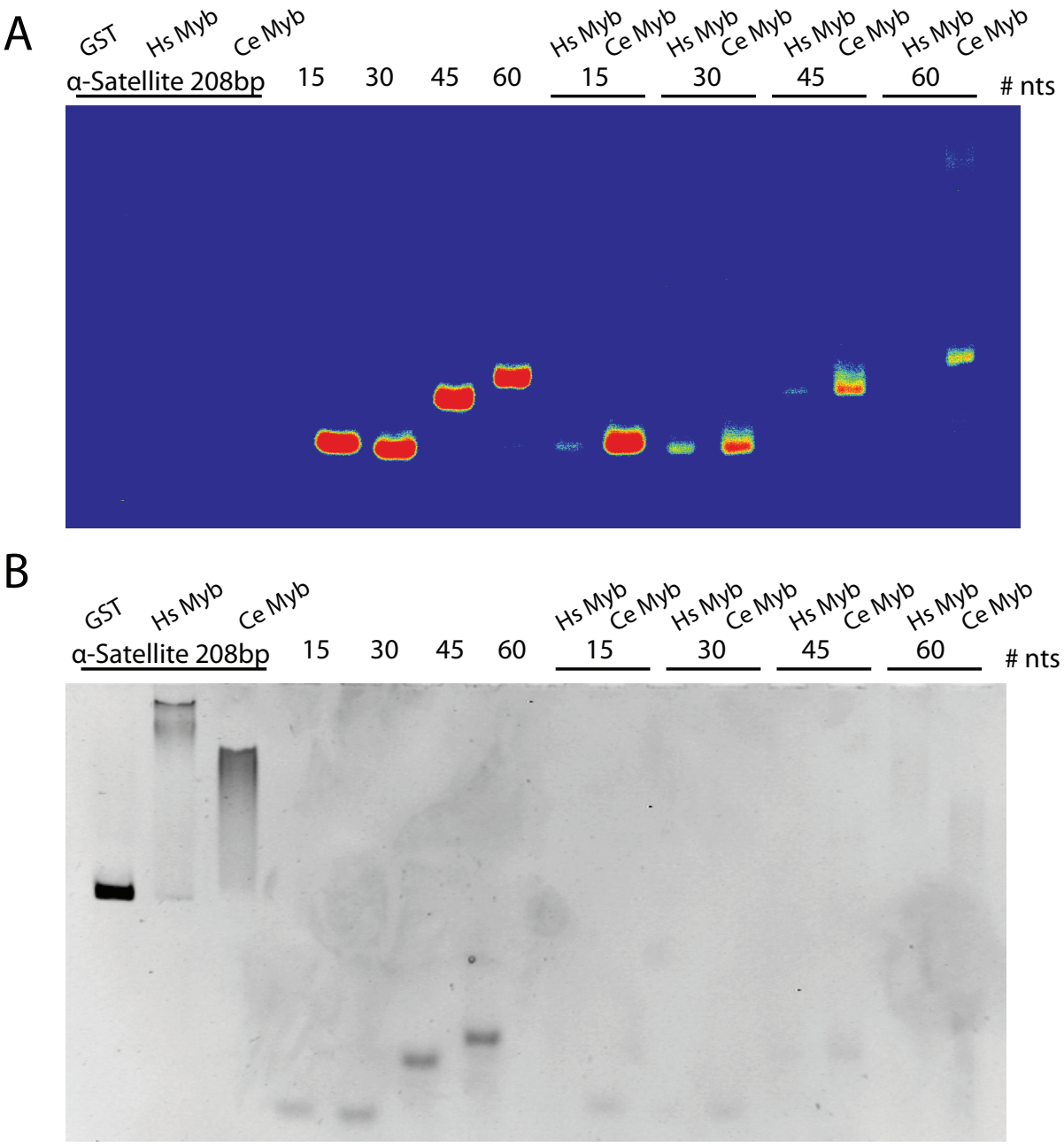



**Figure 5.3 KNL-2 Myb domains bind ssDNA *in vitro*.**

EMSA of HsKNL-2 and CeKNL-2 Myb domains performed with ssDNAs of different sizes, which are coupled to a 488nm excitable fluorophore (equimolar concentration of DNA and protein at 5μM). Both Myb domains are able to bind ssDNA *in vitro*, shown by a decrease in the fluorescence intensity at the bottom of the gel compared with control ssDNA without the protein. Control is performed with dsDNA 208bp human α-satellite DNA to show the protein binding capabilities. HsMyb and CeMyb domains sequences are the Myb long constructs published in De Rop, Dorn et al. (2014) (De Rop et al., 2014) **A** is a fluorescent detectable gel, and **B** is Ethidium bromide stained gel.

### 5.4.3 HsKNL-2 Myb domain form foci within the nucleus and co-localizes with PML bodies.

When HeLa cell line stably expressing CENP-A-YFP is transfected with the HsKNL-2 Myb domain, the transfected protein shows two distinct phenotypes under fluorescence microscope, (1) either it is diffusely localized (rarely) or (2) it forms foci within the nucleus (Figure 5.4 A and B). One interesting observation was that punctate formation of the HsMyb domain occurs independently of the centromeric CENP-A-YFP foci, thus showing no co-localization between those two proteins (Figure 5.4 A). However, the HsMyb domain co-localizes with PML bodies, forming huge nuclear foci where sometimes a single HsMyb domain spot may recruit many PML stained bodies (Figure 5.4 B). Interestingly, full length HsKNL-2 protein shows some PML body localization, but to a lesser extent than the HsKNL-2 Myb domain alone. It might be possible that the HsMyb domain itself is able to bind both DNA and RNA, and protein regions outside this DNA binding domain of KNL-2 are required for its regulation through the cell cycle. Outside of G1, KNL-2 might localize to a region of high RNA content like the nucleolus, and consequently be sequestered away from the centromere. Then in G1, KNL-2 is able to localize to the nucleus and to bind the centromeric chromatin. Thus, extra sequences outside the Myb domain drive its affinity and/or localization to either of those cellular regions. Observation of HsKNL-2 Myb domain co-localizing to PML bodies could explain how the *A. thaliana* KNL2 C-terminus, possibly having a Myb domain, localizes to the nucleolus, a known cellular feature having high RNA content. Moreover, different truncations of the HsMyb domain were also tested: deletions of the N-terminus ($\Delta$ N) and the N-terminus plus first helix ($\Delta$ H1). Those truncations were previously shown to have no effect on the DNA binding capabilities of the HsMyb domain *in vitro*. Here, they localize as foci within the nucleus, and co-localize with PML bodies just like the HsMyb domain. Therefore, the N-terminal and the first helix are not required for the Myb domain co-localization with PML bodies in HeLa cells.

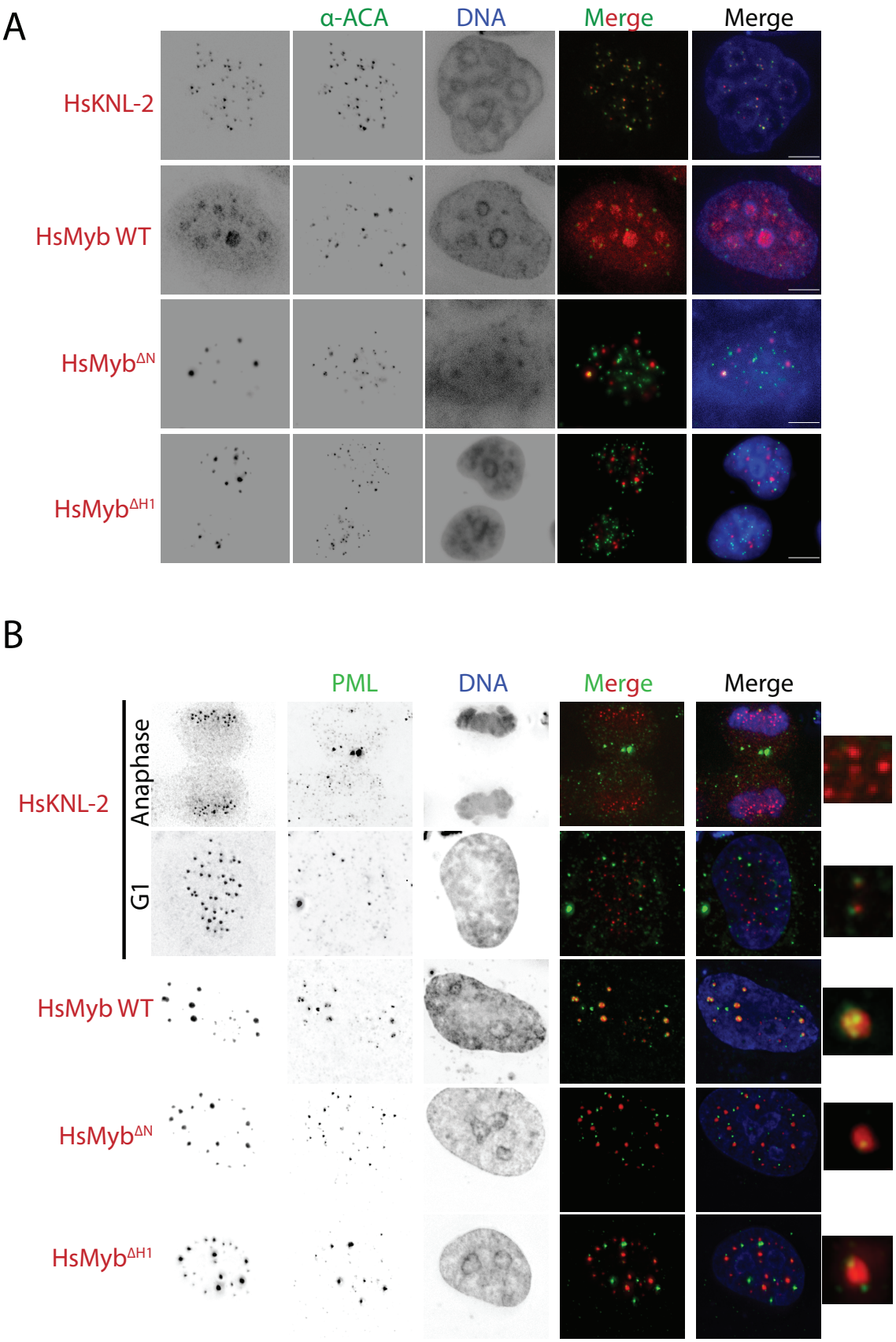

A
α-ACA
DNA
Merge
Merge
HsKNL-2
HsMyb WT
HsMyb^ΔN
HsMyb^ΔH1
B
PML
DNA
Merge
Merge
Anaphase
G1
HsKNL-2
HsMyb WT
HsMyb^ΔN
HsMyb^ΔH1

**Figure 5.4 HsKNL-2 Myb domain localizes as foci within the nucleus and co-localizes with PML bodies.**
**A** Representative immunofluorescence images of HeLa cells transfected with exogenous HsKNL-2-mCherry or HsKNL-2 Myb domain truncations and treated with antibodies specific to CENP-A, CENP-B, and CENP-C (ACA antibody) endogenous proteins. Right columns are merges from images on the left; first merge on the left represents HsKNL-2-mCherry and ACA, second merge represents HsKNL-2-mCherry, ACA and DNA in blue. HsKNL-2 co-localizes with ACA whereas the HsKNL-2 Myb domain truncations localizes as foci within nucleus but do not co-localize with ACA. Scale bar, 5 μm. **B** Representative immunofluorescence images of HeLa cells transfected with exogenous HsKNL-2-mCherry or HsKNL-2 Myb domain truncations and treated with antibodies specific to PML protein. Right columns are merges from images on the left; first merge on the left represents HsKNL-2-mCherry and ACA, second merge represents HsKNL-2-mCherry, ACA and DNA in blue, insets represent enlargement of co-localizing spots within nucleus. HsKNL-2 images show cells in anaphase (top) and in G1 (bottom). HsKNL-2 is in close proximity to PML bodies and HsKNL-2 Myb domain truncations does co-localize with PML bodies. HsMyb domains used are the same as the one published in De Rop, Dorn et al. (2014).

## 5.5 Discussion

Although the immunoprecipitation experiment of KNL-2 published in 2010 only depicted the role of a single partner MgcRacGAP, 60 new binding partners were identified. A cluster of these identified partners by their GO terms gives a schematic view of the major role of these proteins in the cell. As expected, many proteins fall in the nucleus and the cytoplasm fraction of the pie chart. Interestingly, a portion (6%) of these potential binding partners are linked to ribosome function. They are RNA binding, processing and splicing factors (Table 5.S1). KNL-2 binding to these functional proteins is of high interest since it becomes clearer in the literature that RNA has a direct role in CENP-A incorporation to the centromere.

It was demonstrated in *S. pombe* that centromeric transcripts have a role for CENP-A incorporation to the centromere. *S. pombe* Cnp1 (CENP-A homolog) incorporation to the centromere highly depends on the heterochromatin barrier formed by Swi6 (HP1 homolog) binding to H3K9 methylation at the pericentromere (Lejeune et al., 2011). Depletion of this protein or the methyltransferase Clr4 completely abolishes Cnp1 localization to the centromere. Moreover, Clr4 tethering to centromeres highly depends on the RNAi pathway featuring the centromeric transcripts. After depletion of any of the RNAi machinery components or RNAPII, Clr4 will not localize in proximity of the centromeres, therefore it will result in abolition of Cnp1 incorporation to centromere (Chan et al., 2011; Choi et al., 2011; Folco et al., 2008; Gent and Dawe, 2012; Lejeune et al., 2011; Volpe et al., 2002). Also, when Cnp1 is depleted, it shows an increase of the centromeric transcripts in *S. pombe* (Choi et al., 2011). Therefore, it might be possible that KNL-2 preferential binding to the centromeric chromatin is linked to the presence of the centromeric transcripts.

To test the possible binding of the KNL-2 Myb domain to a RNA-like structure, ssDNA binding capabilities were tested *in vitro* by EMSA. This technique coupled to fluorescent DNA has the advantage to discriminate rapidly between non-binding (free

DNA) and binding (bound DNA at the top of the gel). However, it might be challenging to determine the affinity constant by using this technique. Therefore, using an alternative technique like intrinsic fluorescence measurements (possible by the presence of two tryptophans in the Myb domain) or ITC (Isothermal Titration Calorimetry) will be more suitable to have precise binding measurements and find the precise affinity constant. Also, it will be interesting to test the different truncations known to affect dsDNA binding and to look for possible ssDNA binding. Thus, it is possible to conclude with the presented results that KNL-2 Myb domain is able to bind ssDNA *in vitro*.

When transfecting different KNL-2 Myb domain constructs in HeLa cells, they form foci in the nucleus but rarely co-localize with the centromere. However, a strong co-localization with PML bodies was observed. Those bodies are super-structures of the nucleus, and they anchor proteins involved in the regulation of many nuclear processes like replication, transcription, and epigenetic silencing (Lallemand-Breitenbach and de Thé, 2010). Therefore, the possible presence of RNA in these nuclear bodies could explain the co-localization of the HsKNL-2 Myb domain to these nuclear foci. However, there is no strong co-localization of HsKNL-2 full-length protein to PML bodies. One explanation could be that extra sequences outside the Myb domain are regulatory sequences that specify the Myb binding to RNA. They could drive the KNL-2 binding to the centromeric RNA transcripts in G1 and prohibit its localization to PML bodies and/or to the nucleoli. As such, the co-localization of the HsMyb domain with PML bodies could be an artefact created by the absence of those extra sequences of KNL-2.

In 2007, it was demonstrated that RNAPII co-localizes with CENP-C (Wong et al., 2007). Knowing that CENP-C might play a role for KNL-2 recruitment to the centromere, HsKNL-2 co-localization with RNAPII was tested. It was demonstrated that only some HsKNL-2 protein co-localize with RNAPII. However, exogenous HsKNL-2-mCherry was overexpressed in HeLa cells and it might be possible that endogenous KNL-2 (not detected in this experiment) could co-localize with RNAPII.

This possibility could be tested by immunofluorescence. Also, total RNAPII is detected with the antibody used. It might be possible that co-localization foci of HsKNL-2 and RNAPII are not active transcription sites. Therefore, this co-localization might not be significant. However, KNL-2 immunoprecipitation analysis detected RNAPII in one of the immunoprecipitation experiments, thus demonstrating the possible interaction with those two proteins. The published data of the immunoprecipitation together with the presented results in this paper suggest that where newly synthesized CENP-A has to be incorporated, centromeric transcripts may be produced by RNAPII and may be recognized by HsKNL-2 Myb domain to fulfill its licensing function. However, this hypothesis has to be further tested since this current model still has major gaps.

Another possibility for the HsKNL-2 Myb domain function binding to ssDNA or RNA is its regulation through the cell cycle. In *C. elegans*, it was demonstrated that KNL-2 is present throughout the cell cycle. However, this was never observed in human cells due to technical difficulties to detect endogenous KNL-2 (hard to perform suitable immunofluorescence in human cells). However, despite the lack of detection of KNL-2 outside of G1, it is still possible that some KNL-2 might still be present in the cell. A highly hypothetical mechanism will be that the HsKNL-2 Myb domain might have different binding affinity depending on the cell cycle stage. Outside of G1, the Myb domain might bind RNA featured organelles, e.g., the nucleolus and/or PML bodies, and when the extra sequences of KNL-2 (upstream and downstream of the Myb domain) are phosphorylated by CDK1/2, this might change the KNL-2 tertiary structure and affect the Myb domain binding affinity (Silva et al., 2012). Then, when KNL-2 is dephosphorylated at anaphase onset, the Myb domain conformation could change, allowing KNL-2 to bind more strongly to the centromeric dsDNA chromatin versus the RNA elsewhere in the cell (De Rop et al., 2012).

In this paper, it was demonstrated that KNL-2 might co-localize with RNAPII in G1, which could present a mechanism for KNL-2 Myb domain recognition of the centromeric chromatin by binding directly to the nascent centromeric RNA. It could

also be possible that the RNA binding properties of the KNL-2 Myb domain act as a regulatory mechanism by which KNL-2 is sequestered to a high RNA content feature within the cell away from the centromere outside of G1. This is the first evidence of KNL-2 Myb domain binding to both dsDNA and ssDNA *in vitro*.

## 5.6 Material and Methods

### 5.6.1 Cell culture

HeLa cells were grown in DMEM supplemented with 10% fetal bovine serum and penicillin-streptomycin (Invitrogen, Burlington, ON, Canada) at 37°C in a humidified 5% CO2–95% air incubator.

### 5.6.2 Fixed Imaging

All fixed imaging was performed at room temperature on a DeltaVision microscope using Softworx software (Applied Precision, Issaquah, WA, USA) equipped with a CoolSnap HQ2 camera (Photometrics, Tucson, AZ, USA) at 1 × 1 binning and a 100x planApo objective. For the purpose of visualization, some images were deconvoluted using Softworx software (Applied Precision, Issaquah, WA, USA).

### 5.6.3 Immunofluorescence

Immunofluorescence was performed as described (Maddox et al., 2007). PML immunostaining were performed using anti-PML ab423B (Santa Cruz Biothechnology Inc., Dallas, TX, USA). RNAPII immunostaining were performed using anti-PolII SC-899 (Santa Cruz Biothechnology Inc., Dallas, TX, USA). ACA immunostaining were performed using anti-ACA (Antibodies Inc, Davis, CA, USA). Transfections of all constructs were performed using X-tremeGENE 9 DNA transfection reagent (Roche Applied Science, Indianapolis, IN, USA).

### 5.6.4 Electrophoretic Mobility Shift Assay EMSA

Protein and DNA were incubated in a binding buffer (40mM Tris, pH7.5, 1mM EDTA, 250mM NaCl, 5mM $MgCl_2$). Protein and DNA complexes were loaded using a loading buffer that does not contained SDS, and were run on a native 5% acrylamide/bis-acrylamide gel 29:1 (Bio-Rad, Hercules, CA, USA), with running buffer (0.29M Glycine, 45mM Tris-base). The alpha-satellite DNA used was amplified from

pUCαNCS (Conde e Silva et al., 2007) and purified on column. The ssDNA is coupled to a FAM fluorophore (Integrated DNA Technologies, San Jose, CA, USA), sequences 5'- CCA GTG AAT TGT AAT-3' (15nts), 5'-CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG-3' (30nts), 5'- CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG GCG AAT TGG AGC TCC-3' (45nts), 5'- CCAGTG AAT TGT AAT ACG ACT CAC TAT AGG GCG AAT TGG AGC TCC ACC GCG GTG GCG GCC-3' (60nts) (Roy et al., 2011).

## 5.7 References


Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229–243.

Bhatia, V.N., Perlman, D.H., Costello, C.E., and McComb, M.E. (2009). Software Tool for Researching Annotations of Proteins: Open-Source Protein Annotation Software with Data Visualization. Anal. Chem. *81*, 9819–9823.

Bodor, D.L., Valente, L. P., Mata, J.F., Black, B.E., and Jansen, L.E. (2013). Assembly in G1 phase and long-term stability are unique intrinsic features of CENP-A nucleosomes. Mol Biol Cell *24*, 923–932.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H., and Wong, L.H. (2011). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore–microtubule interface. Nat Rev Mol Cell Biol *9*, 33–46.

Choi, E.S., Stralfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of Noncoding Transcripts from within CENP-A Chromatin at Fission Yeast Centromeres. J Biol Chem *286*, 23600–23607.

Conde e Silva, N., Black, B.E., Sivolob, A., Filipski, J., Cleveland, D.W., and Prunell, A. (2007). CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. J Mol Biol *370*, 555–573.

De Rop, V., Dorn, J.F., Osborne, M.J., Boisvert, J., Ryan, J., Padeganeh, A., Moevus, C., Borden, K., Maddox, A.S., and Maddox, P.S. (2014). Centromere epigenome stability is mediated by KNL-2 structural recognition of CENP-A chromatin. Dev Cell *In revision*.

De Rop, V., Padeganeh, A., and Maddox, P.S. (2012). CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly. Chromosoma *121*, 527–538.

Du, Y., Topp, C.N., and Dawe, R.K. (2010). DNA Binding of Centromere Protein C (CENPC) Is Stabilized by Single-Stranded RNA. PLoS Genet *6*, e1000835.

Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science *319*, 94–97.

Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and

Yanagida, M. (2007). Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev Cell *12*, 17–30.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. Cell *145*, 410–422.

Gent, J.I., and Dawe, R.K. (2012). RNA as a Structural and Regulatory Component of the Centromere. Annu. Rev. Genet. *46*, 443–453.

Haaf, T., Warburton, P.E., and Willard, H.F. (1992). Integration of human α-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. Cell *70*, 681–696.

Lagana, A., Dorn, J.F., De Rop, V., Ladouceur, A.M., Maddox, A.S., and Maddox, P.S. (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat. Cell Biol. *12*, 1186–1193.

Lallemand-Breitenbach, V., and de Thé, H. (2010). PML Nuclear Bodies. Cold Spring Harb Perspect Biol *2*, a000661.

Lejeune, E., Bayne, E.H., and Allshire, R.C. (2011). On the Connection between RNAi and Heterochromatin at Centromeres. Cold Spring Harb Symp Quant Biol *75*, 275–283.

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol *176*, 757–763.

Maiato, H. (2004). The dynamic kinetochore-microtubule interface. J Cell Sci *117*, 5461–5477.

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

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc Natl Acad Sci U S A *88*, 3734–3738.

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

Rieder, C.L. (1979). Ribonucleoprotein staining of centrioles and kinetochores in newt lung cell spindles. J Cell Biol *80*, 1–9.

Robbins, E., and Gonatas, N.K. (1964). The Ultrastructure of a Mammalian Cell during the Mitotic Cycle. J Cell Biol *21*, 429–463.

Roy, M.-A., Siddiqui, N., and D'amours, D. (2011). Dynamics and selective DNA-binding activity of Smc5, a core component of the Smc5-Smc6 complex. Cell Cycle *10*, 690–700.

Silva, M.C., Bodor, D.L., Stellfox, M.E., Martins, N.M., Hochegger, H., Foltz, D.R., and Jansen, L.E. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev Cell *22*, 52–63.

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol *127*, 581–592.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J Cell Sci *114*, 3529–3542.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science *297*, 1833–1837.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., D, W., Tyler-Smith, C., Sullivan, K.F., et al. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr Biol *7*, 901–904.

Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., et al. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160.

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. Proc Natl Acad Sci U S A *97*, 7266–7271.

# CHAPTER 6

# DISCUSSION, CONCLUSION AND PERSPECTIVES

## 6.1 Discussion

Cell division is essential for the development and survival of all cells, including single-cell organisms like bacteria, and multi-cellular organisms like humans. Successful cell division in eukaryotes depends on centromere identity maintenance, in order to build a functional kinetochore that will permit proper chromosome segregation. Thus, studying the molecular mechanisms involved in the centromere identity is crucial for a better understanding of cell division. Those findings will later be fundamental to design new drugs that will preferentially act on cancer cell division, without creating any undesirable side effects for healthy cells.

### 6.1.1 KNL-2 is a major regulator of centromere identity propagation

First identified in *C. elegans* and human cells, KNL-2 is essential for CENP-A incorporation to the centromere (Fujita et al., 2007; Maddox et al., 2007). Although not well conserved at the sequence level, the two orthologs show the same phenotypes when depleted, which are less CENP-A incorporation at the centromeres and defects in chromosome segregation due to problems related to the kinetochore assembly. Furthermore, *C. elegans* KNL-2 is required for pre-existing CENP-A stability at the centromere, since CeKNL-2 depletion leads to a total loss of CENP-A at the centromeres (Maddox et al., 2007). Therefore, a thorough understanding of KNL-2 function for CENP-A localization and maintenance at the centromeres is necessary for a better understanding of the centromere identity maintenance in cell division.

The CLC complex, composed of KNL-2, Mis18α, and Mis18β, is the first complex to be recruited to the centromere in early G1 for the licensing step of the CENP-A incorporation pathway (Maddox et al., 2007; Barnhart et al., 2011; De Rop et al., 2014). It was previously demonstrated that KNL-2 is the most upstream protein

required for the newly synthesized CENP-A incorporation to the centromere, since depletion of KNL-2 prevents Mis18 proteins and HJURP localization to the centromere (Barnhart et al., 2011). Thus, KNL-2 is the first protein to recognize and bind the centromere in early G1, and consequently recruits the Mis18 proteins and HJURP to the centromeres. KNL-2 has a predicted Myb DNA binding domain, and it was suggested that KNL-2 is a licensing factor recognizing and binding the centromeric chromatin, acting as a flag to indicate where to incorporate the newly synthesized CENP-A (Fujita et al., 2007; Maddox et al., 2007). This protein might be able to bind directly to the centromeric chromatin, but this was never tested before (see chapter 4). It should be mentioned that aside from its licensing function, KNL-2 possibly has an additional role, since depletion of this protein in *C. elegans* leads to destabilization and ejection of CENP-A protein from the centromere loci (Maddox et al., 2007). By finding new binding partners, like MgcRacGAP, it could give indications on the possible molecular mechanism involved in the maintenance of CENP-A at the centromeres (see chapter 3) (Lagana et al., 2010).

Although *C. elegans* KNL-2 localizes to the centromeres throughout cell division, human KNL-2 is only present at the centromeres in G1 (Fujita et al., 2007; Maddox et al., 2007). Moreover, it has not been possible to observe KNL-2 elsewhere during the cell cycle, due to low antibody affinity and a lower expression of KNL-2 outside of G1. An interesting observation made by Jansen's group is the localization of KNL-2 to the centromeres in G2 when cells are treated with the CDK 1/2 inhibitor Roscovitine (Silva et al., 2012). This phenotype demonstrates that KNL-2 is present throughout the cell cycle, and not only expressed in mitosis. A possible regulation of KNL-2 localization outside of G1 could rely on the capabilities of the Myb domain to bind ssDNA *in vitro* and to localize to PML bodies; a nuclear structure containing RNA (see chapter 5). KNL-2 localization to these nuclear structures might sequester KNL-2 away of the centromere outside of G1 and prevent over-incorporation of CENP-A to the centromeres.

Discovered in 2007, KNL-2 function and regulation are not well understood. The study of this protein is of great interest since it is in tight proximity with the centromere identity propagation regulation. Being the first protein to localize to the centromeres, this is an excellent target for CENP-A destabilization from the centromere, consequently inhibiting the kinetochore formation that will lead to chromosome mis-segregation in mitosis. Having only a role in cell division, it might be possible to design a drug that will target its functional Myb domain, thereby inhibiting its licensing function for CENP-A incorporation to the centromeres. Thus, inhibiting KNL-2 binding to the centromeric chromatin would stop cell division and cell propagation.

### 6.1.2 Challenges of studying KNL-2 function and regulation

Studying KNL-2 *in vivo* is a big challenge. Endogenous KNL-2 expression is very low and it is only concentrated to the centromeres in G1. Outside this cell cycle stage, KNL-2 is still present in the cell but its location is unknown. Moreover, its low expression makes it hard to immunoprecipitate and alternative cellular fractionation is almost impossible since its location is unknown. Only very sensitive detection like Mass Spectrometry can be used for immunoprecipitation experiments. Also, depletion of this protein in HeLa cells is lethal to the cells within 24h, making the phenotypes impossible to observe after two cell cycles. In addition, exogenous overexpression of KNL-2 fused to a fluorescent protein is lethal in human cells. With those two important limitations, it was not possible to perform rescue experiments with exogenous KNL-2 expression under depletion conditions of the endogenous protein.

Moreover, co-localization of exogenous KNL-2 is observed with CENP-A in a stably expressing CENP-A-YFP HeLa cell line. However, those images should be carefully analysed since the endogenous protein is still present in the cell. In those designed experiments, it is impossible to conclude that KNL-2 directly recognizes CENP-A; it might be possible that endogenous KNL-2 recruits exogenously expressed KNL-2 and forms dimers. This could be a possible cellular mechanism since GFP-KNL-2 immunoprecipitation experiment performed using specific GFP antibody results in

equivalent amounts of endogenous and exogenous KNL-2 protein levels shown by western blot (Maddox et al., 2007). Moreover, the available antibodies against KNL-2 in the laboratory are not specific, shown by the presence of many bands on a western blot. Consequently, immunofluorescence of endogenous KNL-2 is problematic, having difficulties to detect distinct foci with acceptable intensities. Thus, with the cell biology tools now available, *in vivo* study of KNL-2 is complex and requires careful analysis.

One way to study the possible roles of KNL-2 in the cell is to look at its binding partners. This can lead to interesting discoveries and give possible indications for its actual function in the cell without depleting or overexpressing it. Another way to study KNL-2 function is to express and purify KNL-2 protein, and undertake a biochemical approach to characterize its structure and its binding affinity.

### 6.1.3 Identification of new KNL-2 binding partners

By doing an immunoprecipitation assay, we found new KNL-2 binding partners, one of them being MgcRacGAP (Lagana et al., 2010). Depletion of this protein by shRNA shows only old CENP-A at the centromeres in G1. By time-lapse microscopy, we discovered that this protein, aside from its localization to the midbody in telophase, also localizes to the centromeres in late G1. Since MgcRacGAP is a GTPase activating protein of the small GTPase Rho family, we performed shRNA knockdown of Cdc42, Rac1 and RhoA in a cell line stably expressing CENP-A-YFP. We identified Cdc42 as the small GTPase required for CENP-A stabilization to the centromere. Since Rho family GTPases are involved in actin cytoskeleton regulation, the possible role of nuclear actin, previously identified and purified from Xenopus oocyte nuclei and observed in mouse leukemia L5178Y cells under fluorescence microscope, was tested for CENP-A maintenance to the centromeres (Rando et al., 2000). However, treatment of the CENP-A-YFP cell line with Latrunculin A and Cytochalasin D, agents blocking actin polymerisation, did not affect the total CENP-A level at the centromere.

Therefore, Cdc42 function in the centromere maintenance might be independent of its actin regulating function.

Because MgcRacGAP localizes to the centromeres at the end of G1 and is involved in the stability of newly incorporated CENP-A at the centromere, we suggested that MgcRacGAP, together with ECT-2, a Guanine exchange factor, and Cdc42, changes the CENP-A identity by adding or removing a mark on the newly incorporated CENP-A. However, we did not identify the actual mark on the CENP-A protein that would be responsible for its maintenance at the centromere. An easy assumption will be to assess the role of post-translational modifications of CENP-A N-terminal tail. But, enzymes involved in such modifications were not identified in the immunoprecipitation experiments (Lagana et al., 2010). Moreover, if a new mark is changed to an old one in the course of a couple of hours in G1 compared to the rest of cell cycle, the majority of CENP-A found at the centromere must be old marked CENP-A. Yet, very few post-translational modifications were found on the N-terminal tail of CENP-A. One possible mark could be tri-methylation of Glycine-1, a mark found only on pre-nucleosomal CENP-A: CENP-A bound to its chaperone HJURP (Bailey et al., 2013). If this mark is only found on pre-nucleosomal CENP-A, it has to be removed when CENP-A gets incorporated to the centromere in mid G1. However, this does not fit with the timing of MgcRacGAP localization to the centromere, which is in late G1. However, it could be possible that a very small population of MgcRacGAP is already present at the centromeres in mid-G1 and performs this job of removing the Glycine-1 tri-methylation mark.

Another hypothesis is MgcRacGAP, together with ECT-2 and Cdc42, destabilizes the Glycine-1 tri-methylated CENP-A incorporated to the centromere. This statement means that tri-methylation mark has to be removed from Glycine-1 before CENP-A gets incorporated to the centromere. One way to test this hypothesis would be to deplete MgcRacGAP from synchronized cells in S phase and quantify the Glycine-1 tri-methylation mark of the centromeric CENP-A and the pre-nucleosomal CENP-A in the next G1 phase using high-resolution mass spectrometry (Bailey et al., 2013). The

pre-nucleosomal CENP-A and the centromeric chromatin can be purified using already published protocols (Bailey et al., 2013; Foltz et al., 2009; Padeganeh et al., 2013). If MgcRacGAP is involved in the destabilization of Tri-methylated Glycine-1 CENP-A at the centromere, this post-translational mark should be more abundant at the centromere in the next G1 phase compared to control condition cells. This observation could explain the centromere size control mechanism and how it does not expand at each cell division.

### 6.1.4 Limitations of studying MgcRacGAP *in vivo*

One challenge to study MgcRacGAP *in vivo* is the fact that this protein has also a very important function in cytokinesis (Canman et al., 2008; Etienne-Manneville and Hall, 2002; Glotzer, 2005; Hirose et al., 2000). MgcRacGAP is the GTPase activating protein of Cdc42, Rac1 and RhoA, and they are essential proteins for the assembly and the constriction of the cytokinetic ring (Canman et al., 2008; Hirose et al., 2000). Those cellular processes are performed before the re-entry in G1 after mitosis, at the transition from anaphase to telophase. Inhibiting MgcRacGAP leads to cytokinetic failure, which could be problematic for the observation of interesting centromere phenotypes in G1. Thus, it is difficult to observe MgcRacGAP depletion phenotypes after more than two cell cycles. One way to bypass this limitation is to perform rescue experiments. We were able to deplete endogenous MgcRacGAP proteins and rescue its function by using a construct resistant to shRNA pathway (Lagana et al., 2010). With this strategy, it is possible to determine which domain within MgcRacGAP sequence is required for CENP-A maintenance at the centromere.

### 6.1.5 KNL-2 acts as a licensing factor and its Myb domain is sufficient to recognize and bind the centromeric chromatin.

It was proposed that KNL-2 acts as licensing factor for CENP-A incorporation to the centromere (Fujita et al., 2007). However, the exact timing of KNL-2 recruitment to the human centromere and the impact of its depletion on CENP-A incorporation to the

centromere were not characterized. By measuring the fluorescence intensity of Mis18-GFP stably expressed in HeLa cells, we demonstrated that Mis18 proteins are recruited to the centromere approximately 5 minutes after anaphase onset. Thus, KNL-2, an obligatory Mis18 partner, is also recruited to the centromere approximately 5 minutes after anaphase onset (Fujita et al., 2007; Lagana et al., 2010). The CLC localization to the centromere is rapid after mitosis exit and it is achieved before HJURP recruitment to the centromere (Foltz et al., 2009). Additionally, depletion of KNL-2, being partial as shown by qPCR, shows newly synthesized CENP-A incorporation at the same rate as the control condition, whereas HJURP depletion shows slower incorporation rates compared to the control cells. These observations are in favour of the licensing role of KNL-2 at the centromere, shown by its rapid recruitment to the centromeres. Thus, the CLC complex is one of the first complexes to be recruited to the centromere in G1 phase.

Using biochemical approaches, we characterized the function of the KNL-2 Myb domain. Since the *C. elegans* KNL-2 Myb domain was not predicted from primary sequence, we solved the in solution NMR 3D structure of the Myb domain. This showed an expected helix-turn-helix motif present in other Myb like DNA binding domains (Ogata et al., 1992; Otting et al., 1988). The human KNL-2 Myb domain structure was already solved by in solution NMR and is available on the protein databank website. After obtaining the expected structure, we expressed those domains in bacteria and purified them to perform DNA binding experiments. Indeed, those two domains are able to bind dsDNA *in vitro* shown by EMSA assay. Furthermore, we were able to detect CeKNL-2 Myb domain binding to DNA by NMR, but we have been unable to obtain similar results for the HsKNL-2 Myb domain. Its binding to dsDNA (a short 15 nucleotides sequence) is too tight for NMR binding experiments. We tried to find a specific motif recognized by the Myb domain by dissecting 208bp α-satellite DNA to short 20 nucleotides sequences, however the Myb domains were not specific to any of those sequences. Moreover, some years ago, CeKNL-2 Myb domain binding affinity was tested using all possible 7 nucleotides long sequences tethered to a microchip, and the CeKNL-2 Myb domain was binding

every 7 nucleotides sequences. Thus, KNL-2 Myb domain binding to DNA is not sequence specific.

We turned our hypothesis toward the possibility that the KNL-2 Myb domains recognizes and binds a specific structure generated by the presence of CENP-A at the centromere. To test this hypothesis, we optimized conditions to test preferential binding of the KNL-2 Myb domains binding to CENP-A-YFP chromatin compared to H2B-GFP chromatin. It should be mentioned that we used H2B-GFP chromatin over H3-GFP for our reference because we were unable to get a stably expressing H3-GFP cell line. Using the TIRF microscope and by doing co-localization analysis, we were able to observe a preferential binding of the KNL-2 Myb domains to CENP-A-YFP chromatin compared to H2B-GFP. It should be pointed out that CeKNL-2 Myb domain also preferentially binds human CENP-A-YFP chromatin. Even if those two domains are highly divergent at the sequence level, they have similar 3D structure and they both preferentially bind to the human CENP-A chromatin. Therefore, KNL-2 function might be conserved through metazoans. However, the exact recognized feature of CENP-A-YFP chromatin was not identified.

Many possibilities are likely to occur (Figure 4.4 F). First, it might be possible that the Myb domains recognize and bind the N-terminal tail of CENP-A. However, this is very unlikely, since KNL-2 was never found in pre-nucleosomal CENP-A immunoprecipitation assays (Obuse et al., 2004; Maddox et al., 2007; Dunleavy et al., 2009; Foltz et al., 2009; 2006). Secondly, Myb domains might bind centromeric transcripts. This assumption is based on the capabilities of the Myb domains to bind ssDNA *in vitro* (see chapter 5), and on the observation that *S. pombe* Cnp1 (CENP-A ortholog) localization to centromere highly depends on the RNAi machinery (Lejeune et al., 2011). This hypothesis could be tested under the TIRF microscope by treating CENP-A-YFP chromatin with RNAse and quantifying the preferential binding of Myb domains compared to the non-treated CENP-A-YFP chromatin. The third possibility is the recognition of a specific structure generated by the presence of CENP-A at the centromere. As mentioned in the introduction, CENP-A nucleosome structure is

different than H3 nucleosome structure (Sekulic et al., 2010). This hypothesis could be tested using *in vitro* reconstituted CENP-A chromatin having mutations that will disturb its initial structure to test, by either size exclusion chromatography or AUC (Analytical Ultracentrifugation) whether the Myb domains are still able to bind those CENP-A nucleosomes. The fourth possibility is that the Myb domains recognize a protein that is constitutively present at the centromere (present throughout the cell cycle), like CENP-C. It was demonstrated in Xenopus egg extract that CENP-C might be involved in KNL-2 recruitment (Moree et al., 2011). To determine if this protein is present where the Myb domains bind to the CENP-A chromatin, TIRF microscopy assays could be performed, and co-localization spots of Myb domains with CENP-A-YFP together with CENP-C, detected by immunofluorescence in the flow chamber, could be quantified (Padeganeh et al., 2013). Finally, a way to determine if the CENP-A nucleosome itself is sufficient for preferential binding of the Myb domain to CENP-A chromatin would be to test the binding of the Myb domains to reconstituted CENP-A nucleosomes *in vitro*. The binding could be assessed using size exclusion chromatography or AUC. Whichever possibility is revealed to reflect the observed and published phenotypes, it will bring deeper understanding of KNL-2 function for CENP-A incorporation to the centromeres.

Lastly, one interesting observation made is the differential mechanism of how those two Myb domains bind DNA. In fact, the HsKNL-2 Myb domain needs the C-terminal part of the Myb domain for DNA binding whereas CeKNL-2 Myb domain needs the N-terminal part to bind DNA. The CeKNL-2 Myb domain binding to DNA via its N-terminal is surprising since most of the Myb domains binding to DNA use their C-terminal part for stable interaction with DNA (Mohrmann et al., 2002; Boyer et al., 2004; Karamysheva, 2004). This differential mechanism could be due to sequence diversity between *C. elegans* and human, having less than 15% similarity between the two orthologs. Another explanation for this differential binding mechanism is relative to the centromere organisation, which is holocentric in *C. elegans* and monocentric in human (Maddox et al., 2004). As mentioned in the introduction, the *C. elegans* centromere is holocentric, meaning that CENP-A is localized all along the

chromosome. Comparatively, human centromeres are monocentric, meaning that the centromere is concentrated to a small region of the condensed chromosome. Those different centromere organisations could be linked to a different structure of the overall centromere region and could explain the differential binding mechanism between the two Myb domains.

### 6.1.6 Limitations of KNL-2 and its Myb domain study *in vitro*.

The major reason why we focused our research on the Myb domain characterization *in vitro* is the difficulty to express full length KNL-2, a 130kDa protein, in bacteria. Indeed, it will be worth the try to express this protein in HeLa cell extract kit or rabbit reticulocytes extract, which was shown to be very successful for the expression of large proteins. However, defining the actual role of the predicted Myb domain is a challenge on its own.

Using NMR techniques like HSQC is very useful to determine which residues on the protein of interest are involved in direct binding to DNA. We were able to get this information for the CeKNL-2 Myb domain (Figure 4.S2), but it was impossible to have a HSQC analysis for HsKNL-2 Myb domain. The reason for this is that the HsKNL-2 Myb domain binds more tightly to dsDNA and ssDNA, an inconvenient for NMR analysis. Also, the structure assignment information from Protein Data Bank is not perfectly fitting our HsMyb domain structure. Unfortunately, the HsKNL-2 Myb domain structure will have to be re-assigned.

TIRF microscopy was used to determine the preferential binding of the Myb domains to either the CENP-A-YFP of the H2B-GFP chromatin. But, TIRF microscopy has its own limitations as only proteins located in the range of the 100nm evanescent field are excited (Cox and Jones, 2013). For our experimental design this is fine, but it will be impossible to study KNL-2 localization and dynamics *in vivo*, since it is localized in the nucleus, which is too far from the cellular membrane.

### 6.1.7 Research impacts in the centromere biology field

A better understanding of the centromere identity propagation and maintenance is a field of high interest having repercussion in other field of research. For instance, knowing the number of CENP-A molecules at the *S. cerevisiae* centromere is crucial for molecule counting, since it is the reference number we use for microscopy-based molecule counting assays (Joglekar et al., 2008; 2006; Lawrimore et al., 2011). But more interestingly, identifying the recognition and binding mechanisms of the first event for CENP-A incorporation can direct the field towards the possibility to target and inhibit KNL-2 function for cell cycle arrest. Thus, KNL-2 is an interesting drug target for future cancer therapy.

MgcRacGAP is a new binding partner of KNL-2 and this finding lead to identify a third step for CENP-A maintenance, changing a new mark to an old one (Lagana et al., 2010). Stabilizing CENP-A at the centromere is a crucial event for the centromere identity maintenance over many cell cycles to keep a proper centromeric size for the kinetochore assembly and chromosomes segregation. On the other hand, this late G1 step is also important when there is mis-incorporated or over-incorporated CENP-A at the centromere. This step will destabilize the undesirable CENP-A, and prevent merotelic attachment, if CENP-A is incorporated in chromosome arms, or growing centromeric size that could be lethal to cells. Thus, the identification of this third step in CENP-A incorporation pathway has high impact in the centromere biology field.

KNL-2 localization to the centromere is the first event to occur, recognizing and binding preferentially CENP-A containing chromatin through its Myb DNA binding domain. This first event will lead to the centromere replenishment with CENP-A and cells will maintain their centromere identity. This study has high impact in the centromere biology field since it will bring researchers to think about what is unique about CENP-A nucleosomes compared to H3 nucleosomes that KNL-2 might recognize and bind. Further studies about the CENP-A incorporation mechanism and

regulation will have to consider the presented data in chapter 4 and 5 to deepen their understanding of CENP-A replenishment at the centromere.

### 6.1.8 Impacts for future cancer research

One of the biggest challenges of cancer therapy is the diversity of the molecular problems involved in cancerous cells leading to heterogeneous tumours (Visvader, 2011). However, one event that all cancer cells depend on is cell division. In the last decades, many drugs were designed to stop cellular processes involved in cell division, but lack of specificity and proteins having multiple functions are huge problems. One popular example of a problematic drug is taxol. Named Paclitaxel in the pharmacological field, this drug is largely used in chemotherapy to stop cell division. In fact, taxol inhibits microtubule dynamics, and thus the formation of the mitotic spindle. This leads to mitotic arrest and apoptosis. However, microtubules have another important role in neurons, providing tracks to vesicular transport (Alberts et al., 2002). Thus, treating a patient with paclitaxel increases the risk of developing axonal degeneration and important neuropathy diseases (Gornstein and Schwartz, 2013).

Having only one identified role in cell division, KNL-2 is an interesting target for chemotherapy. In fact, targeting KNL-2 should affect cells that are cycling and will not disturb differentiated healthy cells such as neurons and muscles. Thus, KNL-2 might be a more specific target for cancer cells. Small molecule screens will be performed to find a KNL-2 Myb domain inhibitor that will disturb its dsDNA binding. Intrinsic fluorescence will be used to determine which molecules could affect the Myb domain binding. This protocol was optimized and it was found that tryptophan, present within the Myb domains, fluorescence is sensitive to solvent upon DNA binding, decreasing its fluorescence intensity. Thus, the experimental set up will be fairly simple, and finding a KNL-2 inhibitor will be of high interest to the centromere biology field.

## 6.2 Conclusion

Centromere biology study is of high interest because of the importance of the centromere in cell division. The kinetochore exclusively assembles at the centromere, attaching the mitotic chromosomes to the mitotic spindle. Any deregulation of the centromere identity maintenance leads to the appearance of aneuploid cells, which are a hallmark in cancer cells. Centromeres are epigenetically defined by the histone H3 variant CENP-A, and the presence of this protein is essential for centromere function. On the other hand, the centromeric DNA sequence has no known functional role. CENP-A is incorporated to the centromeric chromatin in G1 phase following three defined steps: licensing, loading, and maintenance. In this thesis, we describe the discovery and characterisation of the maintenance step, showing the role of the small GTPase Cdc42, and its effector proteins MgcRacGAP and ECT-2, in the maintenance of newly incorporated CENP-A to the centromere. Moreover, we characterized the first step for newly synthesized CENP-A incorporation to the centromere. KNL-2 is the first protein to localize and bind to centromeric loci *in vivo*, and it preferentially binds to CENP-A containing chromatin *in vitro* through its Myb DNA binding domain. Those discoveries have a great impact in the centromere biology field, going toward a better understanding of the centromere identity maintenance. With those new and exciting results, a lot has to be achieved for a global understanding of the centromere identity maintenance. Altogether, the fixed objectives for this thesis were obtained and the future perspectives will be discussed in the next section.

## 6.3 Perspective

In the last decades, some studies have shown a possible role of transcription in the centromere identity. In *S. pombe*, it is well known that the RNAi pathway and the centromere transcription are essential for newly synthesized Cnp1 incorporation to the centromere (see section 1.6) (Lejeune et al., 2011). Recently, some publications have suggested a possible role for transcription in CENP-A incorporation at the human centromere (Chan et al., 2011; Choi et al., 2011). For this thesis, this hypothetical model is interesting since the KNL-2 Myb domain is able to bind ssDNA *in vitro* and co-localizes with PML bodies, a known nuclear structure containing RNA. Interestingly, it was demonstrated in human cells that RNAPII co-localizes with CENP-C (RNAPII is a KNL-2 binding partner shown by immunoprecipitation experiments, see chapter 5 and (Lagana et al., 2010)), and more centromeric transcripts were detected when Cnp1 is depleted in *S. pombe* (Choi et al., 2011). Another piece of interesting information is the clustering of CENP-A to the nucleolus in interphase cells (Ochs and Press, 1992). This nuclear feature is also known to contain RNA, where rDNA is transcribed and ribosomes assembled (Ochs and Press, 1992). Recently, it was demonstrated that CTCF, a chromatin insulator element involved in transcription regulation, together with Modulo and NLP, are required for CENP-A clustering to the nucleolus (Padeken et al., 2013).

All these observations put together lead to a highly hypothetical model for KNL-2 and CENP-A incorporation regulation (Figure 6.1). Centromeres might be transcribed by RNAPII and generate centromeric transcripts that could be recognized and bound by KNL-2 through its Myb domain (anaphase/early G1 phase). This recognition would license the centromere and consequently recruit the CENP-A chaperone HJURP through the CLC, more specifically through interactions with Mis18β (De Rop et al., 2014; Wang et al., 2014). When centromeres are replenished by the newly

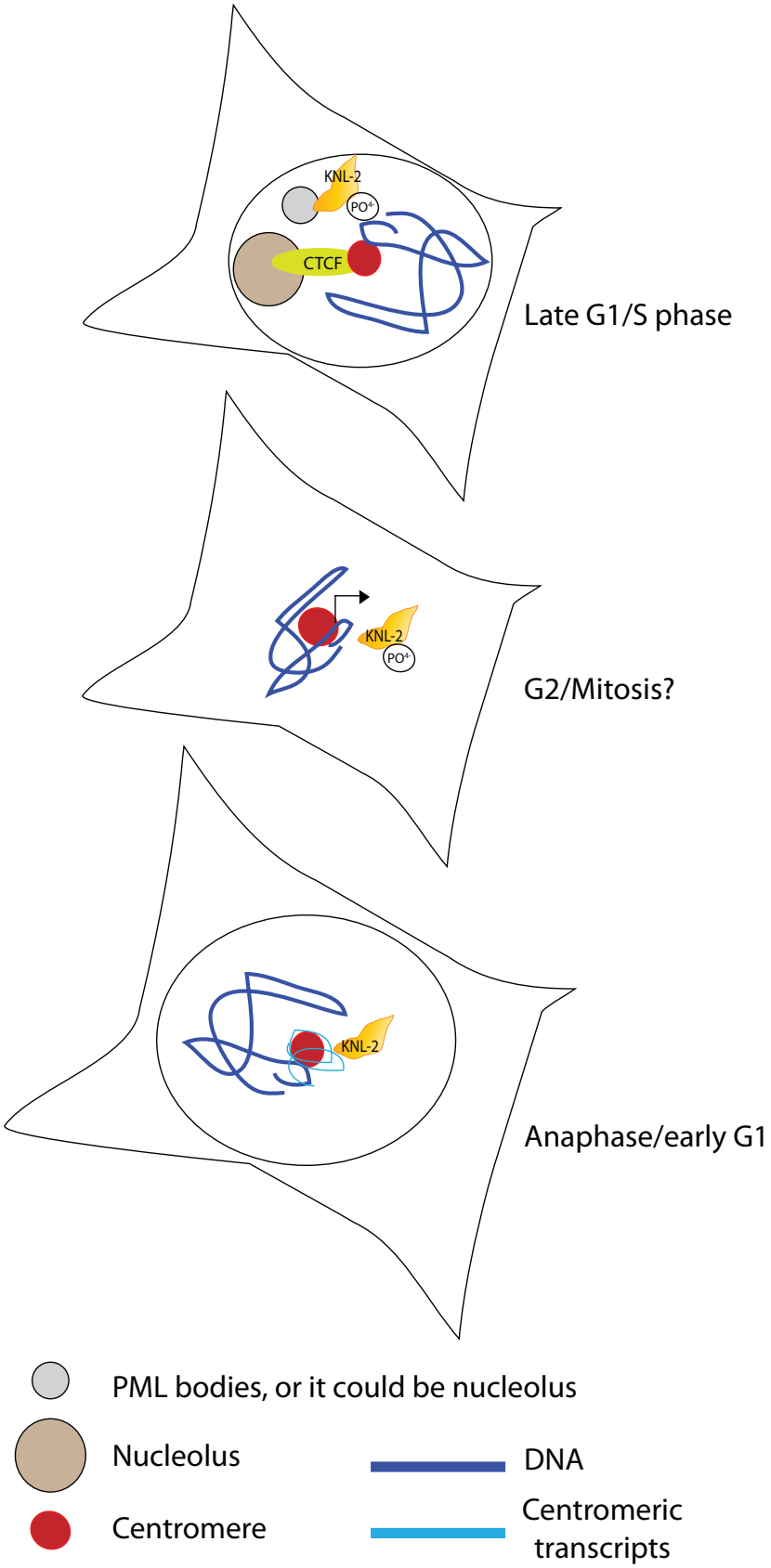



**Figure 6.1 Model of KNL-2 binding regulation to the centromere.**
In late G1 until mitosis, phosphorylated KNL-2 is localized to a cellular feature containing RNA, either the PML bodies or the nucleolus. When it is dephosphorylated by the active phosphatase at exit of mitosis, KNL-2 no longer localizes to the RNA containing feature and might be able to bind to centromeric transcripts, which are transcribed by RNAPII. At late G1, KNL-2 is back to its phosphorylated state and it is sequestered from centromere.

synthesized CENP-A, the centromeres might be clustered to the nucleolus in interphase. As mentioned, CENP-A localization to the nucleolus depends on the NLP, Modulo and CTCF proteins (Padeken et al., 2013). Since CTCF is an insulator for the transcription

regulation, it could be possible that CENP-A clustering to the nucleolus causes inhibition of the centromere transcription. Furthermore, at this stage of the cell cycle, CDK1/2 activities are restored and the kinases might be able to phosphorylate KNL-2. This could influence the KNL-2 affinity for RNA by decreasing its binding affinity to the centromeric transcripts and increasing its affinity to rDNA at the nucleolus or RNA localized to the PML bodies. Then, when cells reach anaphase onset, KNL-2 might be dephosphorylated and might re-establish its centromeric transcripts affinity. This kind of cell cycle regulation of KNL-2 could explain its presence at the centromeres when CDK1/2 are inhibited in G2 (Silva et al., 2012). However, each and every step described in the hypothetical model needs to be tested since it only relies on independent observations and publications. Still, this model or a related one would be interesting for a better understanding of the centromere biology outside of G1.

Another interesting avenue is what happens to the centromere replenishment when a cell enters G0 and subsequently re-enters the cell cycle. Is the mechanism the same as for cycling cells, involving the three defined steps for CENP-A incorporation to the centromere? If so, where does KNL-2 localize and how it is regulated? If not, do cells replenish their centromeres or wait for the next G1 phase? A highly probable mechanism would be the recognition and binding of the centromeric chromatin by KNL-2 as the first event to occur in order to incorporate the newly synthesized CENP-A, leading to the CENP-A replenishment at the centromere before entering S phase. Since no studies were performed on that specific question, every possibility is plausible. Finally, finding the mark of the newly synthesized CENP-A versus the mark of old incorporated CENP-A will be very interesting. If this mark is linked to an overall CENP-A stability at the centromere and a known and well-described regulator controls it, it could be possible to target this regulatory protein and completely abolish the centromere, therefore killing specifically all cells performing cell division without affecting non-dividing healthy cells.

Centromere biology is a young field of research. The presented results will for sure greatly influence the field and give ideas that will propel research toward very creative and amazing research!

## 6.4 References


Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) Biologie moléculaire de la cellule (Flammarion Médecine-Sciences).

Bailey, A.O., Panchenko, T., Sathyan, K.M., Petkowski, J.J., Pai, P.-J., Bai, D.L., Russell, D.H., Macara, I.G., Shabanowitz, J., Hunt, D.F., et al. (2013). Posttranslational modification of CENP-A influences the conformation of centromeric chromatin. Proc Natl Acad Sci U S A 1–6.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J Cell Biol *194*, 229–243.

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

Canman, J.C., Lewellyn, L., Laband, K., Smerdon, S.J., Desai, A., Bowerman, B., and Oegema, K. (2008). Inhibition of Rac by the GAP activity of centralspindlin is essential for cytokinesis. Science *322*, 1543–1546.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H., and Wong, L.H. (2011). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984.

Choi, E.S., Stralfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of Noncoding Transcripts from within CENP-A Chromatin at Fission Yeast Centromeres. J Biol Chem *286*, 23600–23607.

Cox, S., and Jones, G.E. (2013). Imaging cells at the nanoscale. Int J Biochem Cell Biol *45*, 1669–1678.

De Rop, V., Dorn, J.F., Osborne, M.J., Boisvert, J., Ryan, J., Padeganeh, A., Moevus, C., Borden, K., Maddox, A.S., and Maddox, P.S. (2014). Centromere epigenome stability is mediated by KNL-2 structural recognition of CENP-A chromatin. Dev Cell *In revision*.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

Etienne-Manneville, S., and Hall, A. (2002). Rho GTPases in cell biology. Nature *420*, 629–635.

Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black,

B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell *137*, 472–484.

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

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

Glotzer, M. (2005). The Molecular Requirements for Cytokinesis. Science *307*, 1735–1739.

Gornstein, E., and Schwartz, T.L. (2013). The paradox of paclitaxel neurotoxicity: Mechanisms and unanswered questions. Neuropharmacology *76*, 175–183.

Hirose, K., Kawashima, T., Iwamoto, I., Nosaka, T., and Kitamura, T. (2000). MgcRacGAP Is Involved in Cytokinesis through Associating with Mitotic Spindle and Midbody. J Biol Chem *276*, 5821–5828.

Joglekar, A.P., Bouck, D., Finley, K., Liu, X., Wan, Y., Berman, J., He, X., Salmon, E.D., and Bloom, K.S. (2008). Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. J Cell Biol *181*, 587–594.

Joglekar, A.P., Bouck, D.C., Molk, J.N., Bloom, K.S., and Salmon, E.D. (2006). Molecular architecture of a kinetochore–microtubule attachment site. Nat. Cell Biol. *8*, 581–585.

Karamysheva, Z.N. (2004). A C-terminal Myb Extension Domain Defines a Novel Family of Double-strand Telomeric DNA-binding Proteins in Arabidopsis. J Biol Chem *279*, 47799–47807.

Lagana, A., Dorn, J.F., De Rop, V., Ladouceur, A.M., Maddox, A.S., and Maddox, P.S. (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A. Nat. Cell Biol. *12*, 1186–1193.

Lawrimore, J., Bloom, K.S., and Salmon, E.D. (2011). Point centromeres contain more than a single centromere-specific Cse4 (CENP-A) nucleosome. J Cell Biol *195*, 573–582.

Lejeune, E., Bayne, E.H., and Allshire, R.C. (2011). On the Connection between RNAi and Heterochromatin at Centromeres. Cold Spring Harb Symp Quant Biol *75*, 275–283.

Maddox, P.S., Hyndman, F., Monen, J., Oegema, K., and Desai, A. (2007). Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol *176*, 757–763.

Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). "'Holo'"er than thou: Chromosome segregation and kinetochore function in *C. elegans*. Chromosome Res. *12*, 641–653.

Mohrmann, L., Kal, A.J., and Verrijzer, C.P. (2002). Characterization of the extended Myb-like DNA-binding domain of trithorax group protein Zeste. J Biol Chem *277*, 47385–47392.

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

Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (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*, 105–120.

Ochs, R.L., and Press, R.I. (1992). Centromere autoantigens are associated with the nucleolus. Exp Cell Res *200*, 339–350.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc Natl Acad Sci U S A *89*, 6428–6432.

Otting, G., Qian, Y., Müller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988). Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. Embo J *7*, 4305–4309.

Padeganeh, A., Ryan, J., Boisvert, J., Ladouceur, A.-M., Dorn, J.F., and Maddox, P.S. (2013). Octameric CENP-A Nucleosomes Are Present at Human Centromeres throughout the Cell Cycle. Curr Biol *23*, 1–6.

Padeken, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.-J., Kremmer, E., and Heun, P. (2013). The Nucleoplasmin Homolog NLP Mediates Centromere Clustering and Anchoring to the Nucleolus. Mol Cell *50*, 236–249.

Rando, O.J., Zhao, K., and Crabtree, G.R. (2000). Searching for a function for nuclear actin. Trends Cell Biol *10*, 92–97.

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

351.

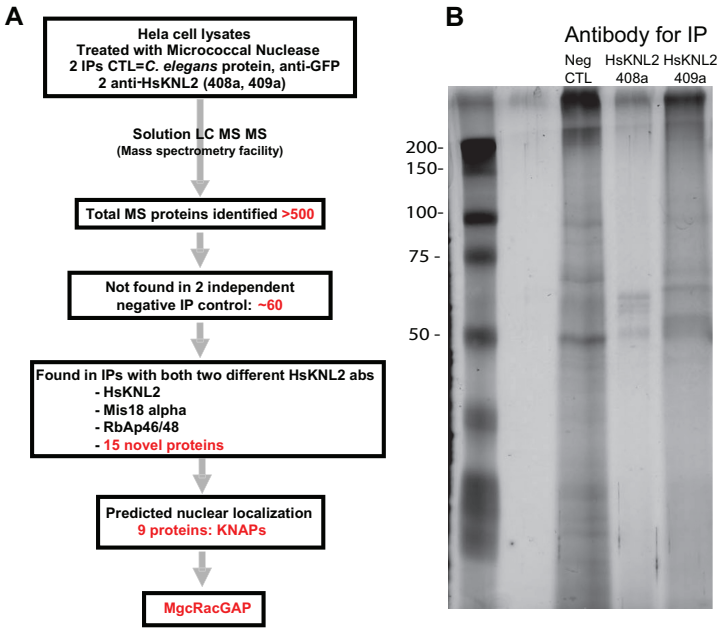
Silva, M.C., Bodor, D.L., Stellfox, M.E., Martins, N.M., Hochegger, H., Foltz, D.R., and Jansen, L.E. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev Cell *22*, 52–63.

Visvader, J.E. (2011). Cells of origin in cancer. Nature *469*, 314–322.

Wang, J., Liu, X., Dou, Z., Chen, L., Jiang, H., Fu, C., Fu, G., Liu, D., Zhang, J., Zhu, T., et al. (2014). Mitotic regulator Mis18β interacts with and specifies the centromeric assembly of molecular chaperone HJURP. J Biol Chem.

# ANNEXE 1

## Chapter 3 Supplementary figures and tables




C

Proteins identified by LC MS MS

| | Name | MW (kDa) | score 408a | score 409a |
|---|---|---|---|---|
| NM_018353 | HsKNL2 | 129 | 56 | 70 |
| NM_007280 | Mis18 alpha | 24 | 78 | 46 |
| NM_002893 | RbAp48 | 47 | 79 | 65 |
| NM_005610 | RbAp46 | 47 | 131 | 120 |
| NM_001126103 | MgcRacGAP | 70 | 72 | 65 |
| NM_016625 | RSRC1 | 31 | 56.2 | 68.45 |
| NM_023012 | RSRC2 | 45 | 185 | 119 |
| NM_023012 | LATS1 | 126 | 55 | 53 |
| NM_022457 | UBR4 | 573 | 217 | 1046 |
| NM_005870 | SAP18 | 17 | 72 | 53 |
| NM_04960 | FUS | 53 | 81 | 72 |
| NM_139168 | SP12 | 59 | 156 | 64 |
| NM_006185 | NUMA-1 | 236 | 105 | 197 |

**Figure 3.S1. Proteomic identification of MgcRacGAP as a HsKNL2 interacting protein.**

**(A)** Scheme of our IP/MS analysis. Two antibodies to distinct regions of HsKNL2 as well as two controls were used for IP/MS. After multiple replicates as well as removing probable background bands, 9 novel proteins were isolated. **(B)** Silver stained gel of eluates from IPs. Eluates were digested as a mixture for analysis by LC-MS/MS. **(C)** List of HsKNL2 interacting proteins isolated and defined as unique by the scheme in A.

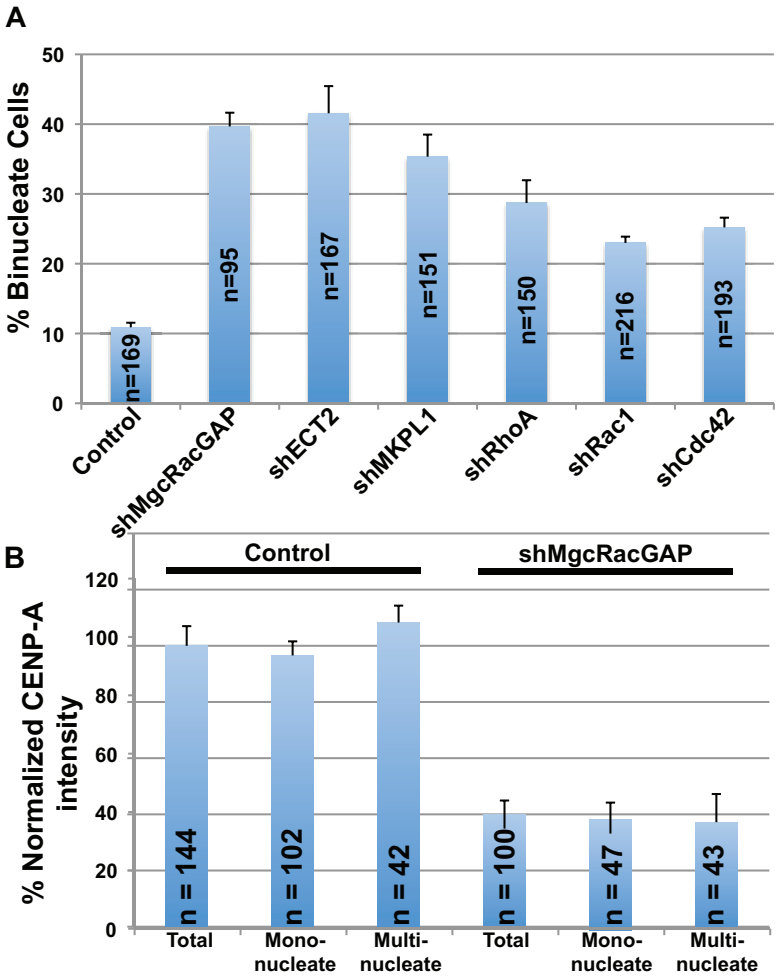



**Figure 3.S2. Depletion of MgcRacGAP results in loss of CENP-A localization to centromeres independent of successful cytokinesis.**
**(A)** Depletion of MgcRacGAP, ECT2, MKLP1, RhoA, Rac1, or Cdc42 caused an increase in multinucleate cells, as expected. **(B)** Multinucleate cells from either control or shMgcRacGAP treated cells displayed the same level of CENP-A as found in mono-nucleate cells. Results are average +/- SEM in both graphs. N= number of cells analyzed.

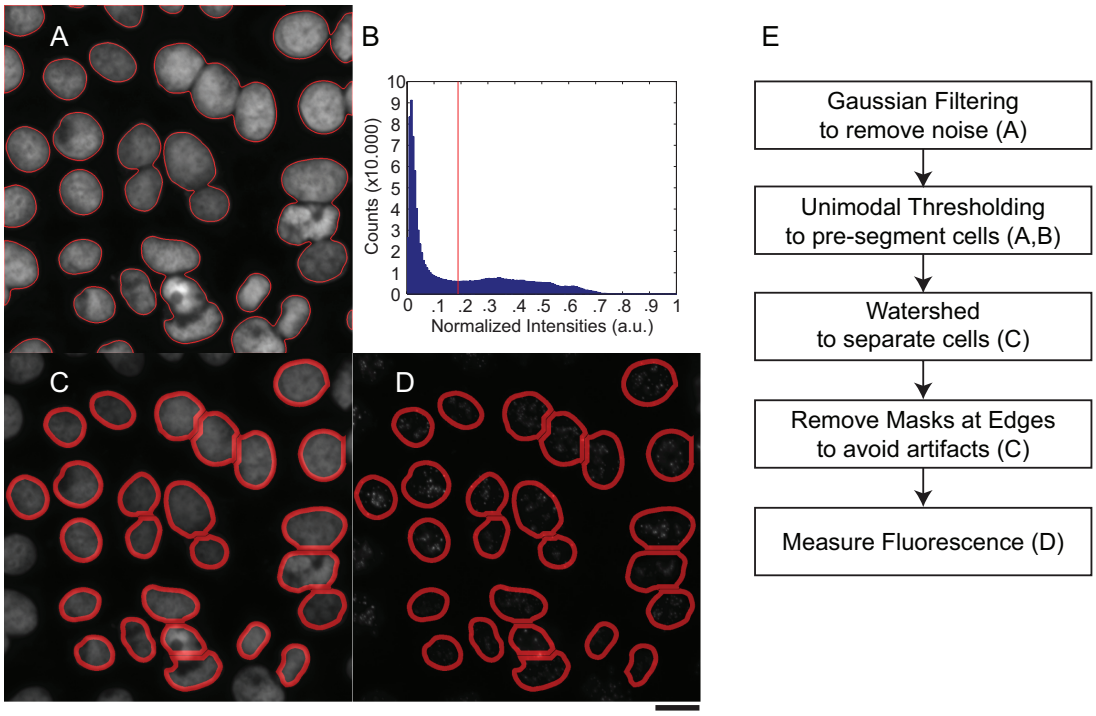



**Figure 3.S3. Scheme of our automated analysis regime for generating large, unbiased data sets of CENP-A levels in interphase cells.**
**(A)** DAPI images were used to generate a mask after noise filtering. **(B)** Masks were generated by automated thresholding, **(C)** then individual nuclei were separated by watershed. **(D)** Final masks generated on the DAPI image were transferred to the CENP-A image and intensities were measured for all cells. **(E)** A flow chart of our analysis regime. Scale Bar, 10 μm.

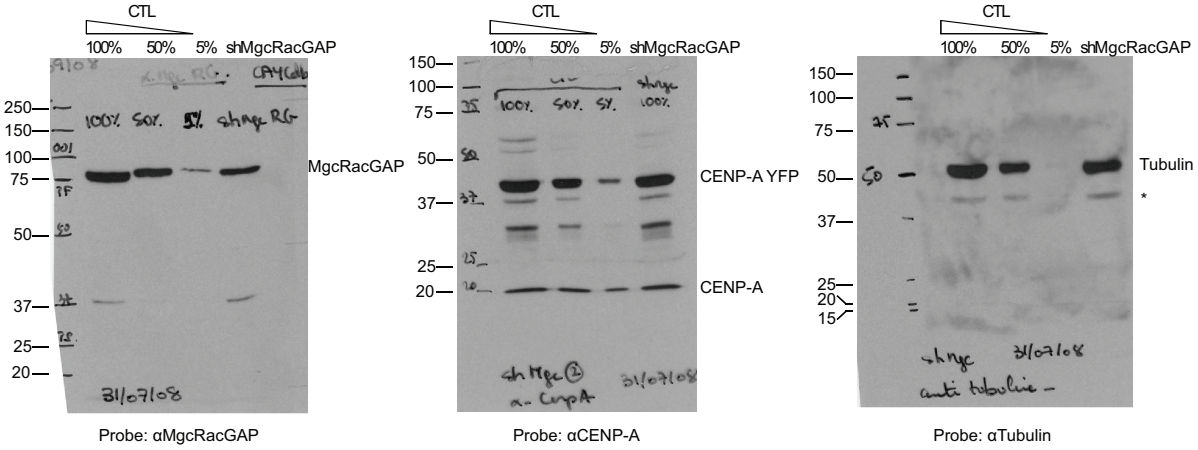



**Figure 3.S4. Western blot analysis shows that endogenous CENP-A protein levels are unaffected after depletion of MgcRacGAP.**
A dilution curve of control extract is compared to 100% load of shMgcRacGAP treated cell extract (as in Figure 3.1). The tubulin control is the same blot as shown in Figure 3.1. "*" marks a non-specific band.

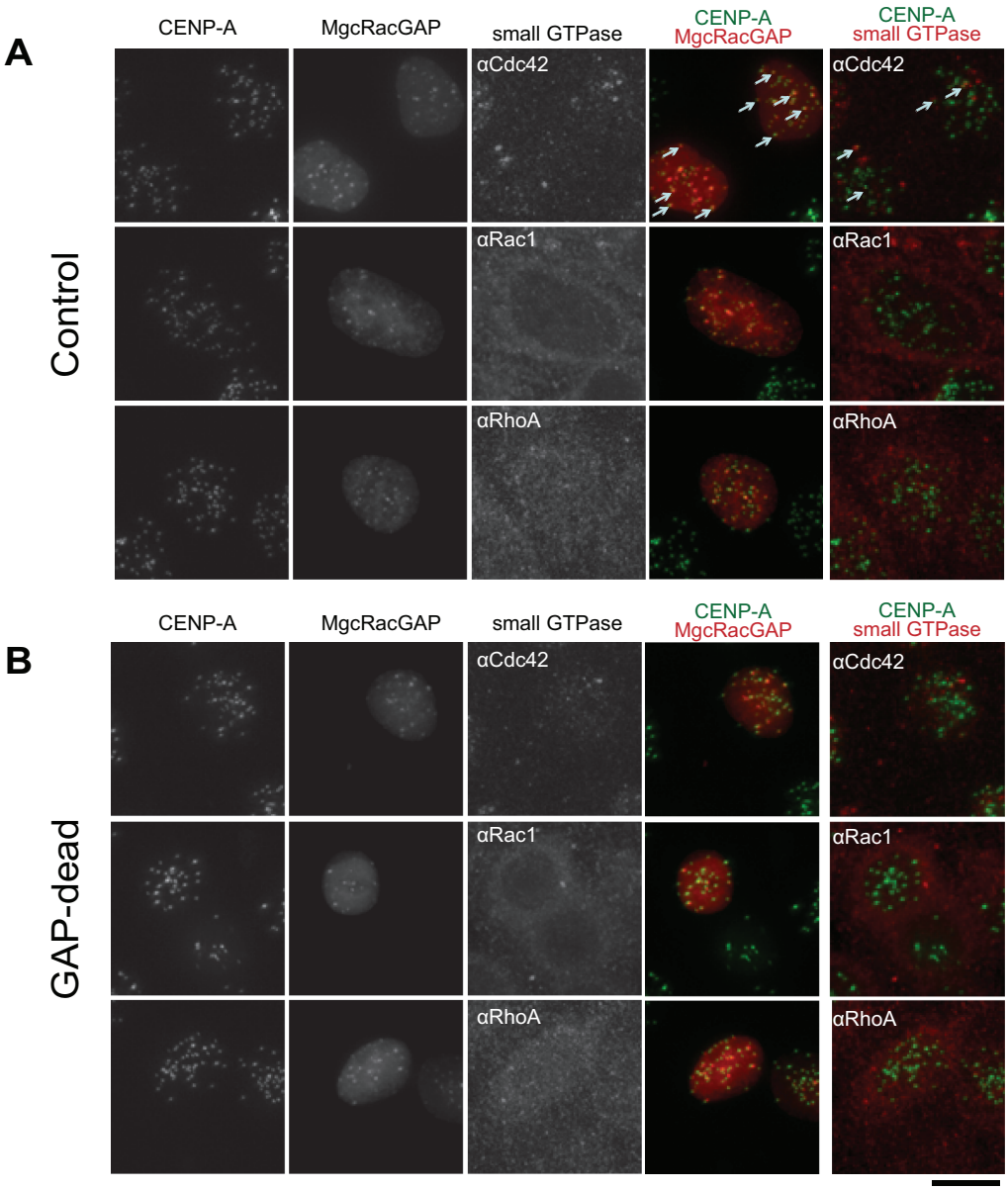



**Figure 3.S5. MgcRagGAP GAP-dead mutant localizes more persistently to centromeres.**
**(A)** MgcRacGAP-mCherry colocalizes (arrows, yellow in merged images) with CENP-A-YFP and Cdc42
**(B)** MgcRacGAP-dead mutant colocalizes with CENP-A and Cdc42 more persistently to centromeres in late G1

## Table 3.S1 Detail of shRNA constructs used in this study

| Protein Names | Accession numbers | Clone numbers | Sequences |
|---|---|---|---|
| Non-target | | SHC002 | |
| shLuciferase | | SHC007 | |
| CENP-A | NM_001809 | NM_001809.2-433s1c1 | CCGGGCAGCAGAAGCATTTCTAGTTCTCGAGAACTAGAAATGCTTCTGCTGCTTTTTG |
| HsKNL2 | NM_018353.4 | SH2456-E-11 | TGCTGTTGACAGTGAGCGCGCTGAGACTCTCAGTACTAATTAGTGAAGCCACAGATGTAATTAGTACTGAGAGTCTCAGCATGCCTACTGCCTCGG |
| | | SH2464-E-1 | TGCTGTTGACAGTGAGCGCGCTGATCAGAAACAAACTATTTAGTGAAGCCACAGATGTAAATAGTTTGTTTCTGATCAGCATGCCTACTGCCTCGGA |
| | | SH2760-E-4 | TGCTGTTGACAGTGAGCGCGGAGGAGATGACTTATCTAATTAGTGAAGCCACAGATGTAATTAGATAAGTCATCTCCTCCTTGCCTACTGCCTCGG |
| | | SH2771-A-1 | TGCTGTTGACAGTGAGCGACCAGATGACCAAGTAAATAATTAGTGAAGCCACAGATGTAATTATTTACTTGGTCATCTGGGTGCCTACTGCCTCGGA |
| Mis18 | NM_007280 | NM_007280.1-383s1c1 | CCGGGCATCAGAGATGGATATTCAACTCGAGTTGAATATCCATCTCTGATGCTTTTTG |
| MgcRacGAP | NM_001126103 | SH2781-B-11 | TGCTGTTGACAGTGAGCGAGCATGCTAGGTTCTATAAATATAGTGAAGCCACAGATGTATATTTATAGAACCTAGCATGCCTGCCTACTGCCTCGGA |
| | | SH2314-B-7 | TGCTGTTGACAGTGAGCGCAGCCAAATGCTGCTCATAGAATAGTGAAGCCACAGATGTATTCTATGAGCAGCATTTGGCTTTGCCTACTGCCTCGG |
| | | SH2508-B-2 | TGCTGTTGACAGTGAGCGAGCTCAGAGTCCACATACTAAATAGTGAAGCCACAGATGTATTTAGTATGTGGACTCTGAGCCTGCCTACTGCCTCGG |
| ECT2 | NM_018098 | SH2330-g-8 | TGCTGTTGACAGTGAGCGACCAGCTTCTCTTAAGCATATTTAGTGAAGCCACAGATGTAAATATGCTTAAGAGAAGCTGGGTGCCTACTGCCTCGGA |
| | | SH2229-e-1 | TGCTGTTGACAGTGAGCGCGCAGTTGATGACTTTAGAAATTAGTGAAGCCACAGATGTAATTTCTAAAGTCATCAACTGCTTGCCTACTGCCTCGGA |
| | | SH2335-d-8 | TGCTGTTGACAGTGAGCGCCCAGAGTCTAGCATTAACTATTAGTGAAGCCACAGATGTAATAGTTAATGCTAGACTCTGGTTGCCTACTGCCTCGGA |
| MKLP1 | NM_004856 | NM_004856.4-622s1c1 | CCGGGCCTTATTAGAACGTCAGAAACTCGAGTTTCTGACGTTCTAATAAGGCTTTTTG |
| RhoA | NM_001664 | NM_001664.1-382s1c1 | CCGGCGATGTTATACTGATGTGTTTCTCGAGAAACACATCAGTATAACATCGTTTTTG |
| | | NM_001664.1-199s1c1 | CCGGTGGAAAGACATGCTTGCTCATCTCGAGATGAGCAAGCATGTCTTTCCATTTTTG |
| | | NM_001664.1-300s1c1 | CCGGGAAAGCAGGTAGAGTTGGCTTCTCGAGAAGCCAACTCTACCTGCTTTCTTTTTG |
| Rac1 | NM_006908 | NM_006908.3-459s1c1 | CCGGCGCAAACAGATGTGTTCTTAACTCGAGTTAAGAACACATCTGTTTGCGTTTTT |
| Cdc42 | NM_1791.2 | NM_001791.2-328s1c1 | CCGGCAGATGTATTTCTAGTCTGTTCTCGAGAACAGACTAGAAATACATCTGTTTTTG |
| | | NM_001791.2-193s1c1 | CCGGCGGAATATGTACCGACTGTTTCTCGAGAAACAGTCGGTACATATTCCGTTTTTG |
| | | NM_001791.2-471s1c1 | CCGGCCCTCTACTATTGAGAAACTTCTCGAGAAGTTTCTCAATAGTAGAGGGTTTTTG |

**Table 3.S2 Oligos used for cloning**

| Proteins | Primers | Primer sequences |
|---|---|---|
| MgcRacGAP | oPM206 | GCGCGCTCGAGGATACTATGATGCTGAATGTG |
| | oPM207 | GCGCGGGATCCTCACTTGAGCATTGGAGAA |
| R386A_MgcRacGAP | oAL40 | CTGACTGAGACAGGCCTGTATgcGATCTCTGGCTGTGACCGC |
| | oAL41 | GCGGTCACAGCCAGAGATCGCATACAGGCCTGTCTCAGTCAG |
| HsKNL2-ML | oOD1872 | GCGCGCGAATTCTCATTTGTGCTTTGGAAGAGATC |
| | oOD1871 | GCGCGCGGATCCGAGCAAAGATGCATGAGGTATAA |
| HsKNL-2-NT-2 | oOD1875 | GCGCGCGGATCCAAAGAAAAAGTACTGCGTGACAA |
| | oOD1876 | GCGCGCGAATTCTCACGTGTCCTTTTTGGATTTAGTG |

**Table 3.S3 Plasmids used in this study**

| Plasmid Name | Template | Primers |
|---|---|---|
| MgcRacGAP-YFP | MgcRacGap Full Length | oPM 206/207 |
| MgcRacGAP-mcherry | MgcRacGap Full Length | oPM 206/207 |
| R386A-MgcRacGAP-mcherry | MgcRacGAP-RFP | oAL40/41 |
| GST-HsKNL2, 408 | HsKNL2-ML | oOD1872/oOD1871 |
| GST-HsKNL2, 409 | HsKNL-2-NT-2 | oOD1875 /oOD1876 |

## Table 3.S4 Antibodies used in this study

| Target | Antibody | Catalogue number | Company | Used | dilution |
|---|---|---|---|---|---|
| E-GFP | Rabbit, E-GFP anti-EGFP GST fusion | | Covance | IP | 7ug/ml |
| | | | | WB | 1ug/ml |
| WCES4 | Rabbit, HCP-3 F58A4.3 1GST fusion | | Covance | IP | 7ug/ml |
| KNL-2, 408a | Rabbit, anti-KNL-2 GST fusion | | Covance | IP | 7ug/ml |
| | | | | IF | 3ug/ml |
| | | | | WB | 0.5ug/ml |
| KNL-2, 409a | Rabbit, anti-KNL-2 GST fusion | | Covance | IP | 1ug/ml |
| | | | | IF | 1ug/ml |
| | | | | WB | 1ug/ml |
| Tubulin | Mouse, anti-a-tubulin | T6199 | Sigma | IF | 1ug/ml |
| | | | | WB | 0.25ug/ml |
| MgcRacGAP | Goat, anti-MgcRacGAP | ab2270 | abcam | IF | 1ug/ml |
| MgcRacGAP | Rabbit, anti-MgcRacGAP | ab61192 | abcam | WB | 1ug/ml |
| CENP-A | Mouse, anti-CenpA | ab13939 | abcam | WB | 1ug/ml |
| ACA | Human, anti-Centromere Protein | 15-235 | Antibodies inc. | IF | 2ug/ul |
| RhoA | Rabbit, anti-RhoA (kit) | 89854z | Pierce | IF | 2.5ug/ml |
| Rac1 | Mouse, anti-Rac1 | 05-389 | Upstate | IF | 1ug/ml |
| Cdc42 | Rabbit, anti-Cdc42 | 786627 | abcam | IF | 1ug/ml |
| Actin | Alexa Fluor® 546 Phalloidin | A-22283 | Invitrogen | IF | 5units/ml |
| Mouse | Alexa Fluor® 647 donkey anti-mouse IgG (H+L) | A-31571 | Invitrogen | IF | 1ug/ml |
| Rabbit | Alexa Fluor® 647 goat anti-rabbit IgG (H+L) | A-21244 | Invitrogen | IF | 1ug/ml |
| Rabbit | Alexa Fluor® 546 goat anti-rabbit IgG (H+L) | A-11010 | Invitrogen | IF | 1ug/ml |
| Human | Alexa Fluor® 488 goat anti-Human IgG (H+L) | A-11013 | Invitrogen | IF | 1ug/ml |
| Human | Alexa Fluor® 546 goat anti-Human IgG (H+L) | A-21089 | Invitrogen | IF | 1ug/ml |
| Goat | Alexa Fluor® 546 donkey anti-goat IgG (H+L) | A-11056 | Invitrogen | IF | 1ug/ml |
| Mouse | Anti-mouse HRP IgG (H+L) | 115-035-003 | Jackson | WB | 0.04ug/ml |
| Rabbit | Anti-rabbit HRP IgG (H+L) | 111-035-003 | Jackson | WB | 0.04ug/ml |

**Table 3.S5 Oligos used for RT**

| Protein Names | Probe number | Primer for RT-PCR |
|---|---|---|
| CENP-A | 41 | cacacacctcttgataaggaagc |
| | | acaccacgagtgaatttaacacata |
| HsKNL2 | 77 | tgcacgagaaaaccaaacag |
| | | cactgtgcttattactcttcagttcc |
| Mis18 | 29 | ctggctgccttgagaggt |
| | | tgatgcatttactatggcttttg |
| MgcRacGAP | 18 | ttcgccttaacagagccttt |
| | | acatggcagctatgctgttg |
| ECT2 | 82 | ctctaggtgagcacccctgt |
| | | tgtgccgttttcttgctatct |
| MKLP1 | 85 | ctgtggctgtggagatgaga |
| | | ctgggactgtcagttcatgg |
| RhoA | 8 | ggagctagccaagatgaagc |
| | | gccaatcctgtttgccata |
| Rac1 | 77 | gatgcaggccatcaagtgt |
| | | caggaaatgcattggttgtg |
| Cdc42 | 37 | tggagtgttctgcacttacaca |
| | | gctcttcttcggttctggag |

**Table 3.S6 Summary of RT-PCR results**

| Gene | Detector | RQ |
|---|---|---|
| Non Target | | 1 |
| CENP-A | CENP-A | 0.58 ± 0.05 |
| HsKNL-2 | HsKNL-2 | 0.68 |
| Mis18 | Mis18 | 0.60 ± 0.11 |
| MgcRacGAP | MgcRacGAP | 0.64 ± 0.03 |
| MgcRacGAP | CENP-A | 1.05 ± 0.04 |
| Cdc42 | Cdc42 | 0.62 ± 0.014 |
| Cdc42 | CENP-A | 1.04 ± 0.035 |
| ECT2 | ECT2 | 0.59 ± 0.055 |
| MKLP1 | MKLP1 | 0.63 ± 0.1 |
| RhoA | RhoA | 0.57 ± 0.05 |
| Rac1 | Rac1 | 0.65 ± 0.063 |

# ANNEXE 2

## Chapter 4 Supplementary figures and tables

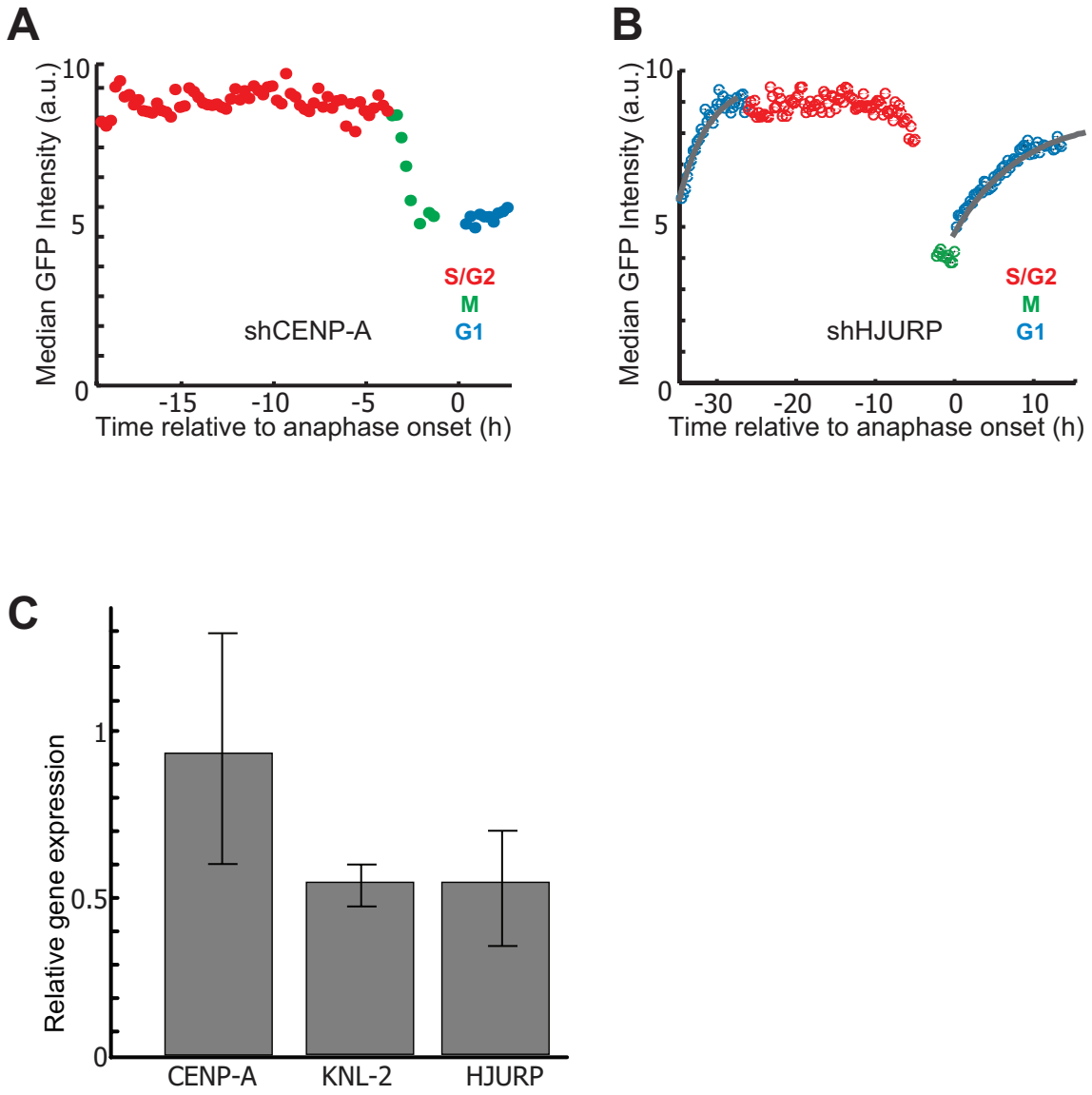



**Figure 4.S1: Depletion of CENP-A and HJURP affect loading of new CENP-A.** **(A)** Thorough depletion of CENP-A abolishes replenishment, but does not affect pre-existing CENP-A at the centromere. **(B)** HJURP depletion affects both speed and level of replenishment. After a replenishment phase with normal kinetics, the cell divides for a second time, after which both the rate constant and the level of replenishment are reduced by ~30%. This example of partial depletion of HJURP confirms its role as a CENP-A loading factor. **(C)** qPCR of cells depleted with shRNA for the presented target in HeLa cell line. KNL-2 and HJURP are depleted approximately 50% of total RNA content in the cell, showing a partial depletion phenotype.

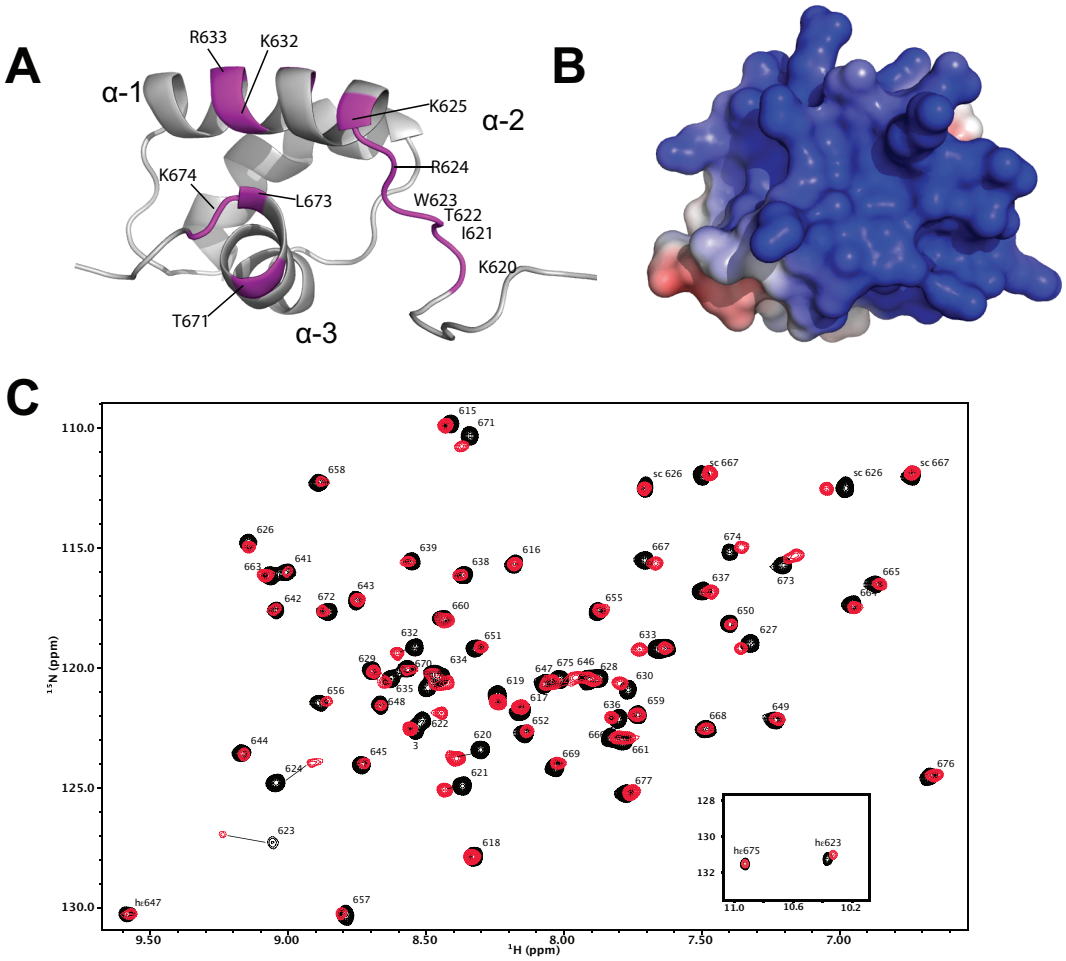



**Figure 4.S2 Biochemical characterizations of CeMyb domain.**
**(A)** Summary of chemical shift mapping of the CeMyb 11 domain upon addition of VDR59 DNA sequence (refer to Materials and Methods). Peaks most shifted are mapped onto the structure of CeKNL-2. **(B)** The electrostatic surface for the same orientation of CeKNL-2 in A showing a highly positive charge surface around the first helix. **(C)** HSQC spectra of CeKNL-2 Myb domain in presence of 15mers DNA sequence taken from alpha-satellite DNA. Inset represent residue that have proton that shift over 10ppm.

**Table 4.S1 NMR and refinement statistics for Ce-KNL2-myb**

| | |
|---|---|
| **NMR restraints** | |
| Distance restraints | |
| Total NOE | 1336 |
| Sequential ($\|i - j\| = 1$) | 649 |
| Medium-range ($\|i - j\| \leq 5$) | 332 |
| Long-range ($\|i - j\| > 5$) | 335 |
| Dihedral angle restraints[a] | |
| $\phi$ | 48 |
| $\psi$ | 48 |
| **Structure statistics** | |
| Deviations from idealized geometry[b] | |
| Bond lengths, r.m.s. (Å) | 0.001 |
| Bond angles, r.m.s. (º) | 0.2 |
| Close Contacts | 0 |
| Violations[c] | |
| Upper limits (number, max value (Å)) | 0, 0.20 |
| Dihedral angle restraints (number, max value (º)) | 0, 2.1 |
| VdW (number, max value (Å)) | 0, 0.19 |
| Average rmsd to mean (Å)[d] | |
| Backbone | 0.35±0.08 |
| Heavy | 0.91±0.10 |
| Ramachandran plot (PROCHECK)[d] | |
| Most favored regions | 84.6% |
| Allowed regions | 15.2% |
| Generously allowed regions | 0.2% |
| Disallowed regions | 0.0% |

Statistics are given for the 10 lowest-energy structures out of 100 calculated. [a]Derived from TALOS+. [b]Calculated from PSVS structure validation software. [c]Average number of violations larger than the cutoff (0.2 Å for distance constraints, 0.2 Å for VdW and 5 º for dihedral restraints). [d]Values are reported for structured residues 10-64.

# ANNEXE 3

## Chapter 5 Supplementary table

**Table 5.S1 KNL-2 potential binding partners identified by IP/MS analysis with their Gene Ontology based on cellular components**

| Proteins Name* | Gene Ontology based on cellular components | Proteins Name* | Gene Ontology based on cellular components |
|---|---|---|---|
| 60 kDa heat shock protein, mitochondrial | Cell surface<br>Plasma membrane<br>Other intracellular organelles<br>Cytoplasm<br>Endosome<br>Extracellular macromolecular complex<br>Mitochondria<br>Other | Rac GTPase-activating protein 1 | Other intracellular organelles<br>Cytoskeletion<br>Cytoplasm<br>Nucleus<br>Other |
| | | POTE ankyrin domain family member F | Cytoplasm |
| | | Serine/Arginine-related protein 53 | Cytoplasm<br>Nucleus |
| Stress-70 protein, mitochondrial | Cell surface<br>Mitochondria<br>Nucleus | 60S ribosomal protein L26 -like 1 | Cytoplasm<br>Ribosome |
| Arginine/serine -rich coiled coil protein 2 | - | | |
| | | Splicing regulatory glutamine/lysine-rich protein 1 | Mitochondria<br>Nucleus |
| Mis18-binding protein 1 | Chromosome<br>Nucleus | | |
| | | Tubulin alpha -1C chain | Cytoskeleton |
| Nuclear mitotic apparatus protein 1 | Cytoskeleton<br>Chromosome<br>Cytoplasm<br>Nucleus<br>Other | Tubulin beta -2B chain | Cytoplasm, cytoskeleton, plasma membrane<br>Cytoskeletion<br>Plasma membrane |

| | |
|---|---|
| Tubulin beta -6 chain | Cytoplasm<br>Cytoskeleton<br>Plasma membrane |
| Heat shock 70 kDa protein 1 -like | - |
| 60S ribosomal protein L23a | Ribosome<br>Nucleus |
| 60S ribosomal protein L26 | Ribosome |
| 40S ribosomal protein S19 | Ribosome<br>Nucleus |
| Heterogeneous nuclear ribonucleoproteins C1/C2 | Nucleus |
| RNA -binding protein 39 | Cytoskeleton<br>Nucleus |
| Tubulin beta -8 chain | Cytoplasm<br>Cytoskeleton<br>Plasma membrane |
| Putative RNA binding protein Luc 7 -like 1 | Nucleus |
| Serine/arginine -rich splicing factor 6 | Nucleus |
| RNA -binding protein with rich serine -rich domain 1 | Cytoplasm<br>Nucleus |
| Nucleolar protein 56 | Nucleus<br>Cytoplasm |
| Zinc finger Ran -binding domain -containing protein 2 | Nucleus |
| Elongation factor 1 -alpha 2 | Macromolecular complex<br>Nucleus<br>Other |
| Heterogeneous nuclear ribonucleoprotein K | Nucleus<br>Cytoplasm<br>Chromosome |
| Serine/arginine -rich factor 3 | Nucleus |
| Polyadenylate binding protein 3 | Cytoplasm |
| Serine/arginine -rich splicing factor 10 | Cytoplasm<br>Nucleus |
| U1 small nuclear ribonucleoprotein 70 kDa | Nucleus |

| | | | |
|---|---|---|---|
| Complement component 1 Q subcomponent -binding protein, mitochondrial | Cell surface | Protein Mis18 -beta | Nucleus |
| | Cytoplasm | | Chromosome |
| | Extracellular<br>Mitochondria<br>Nucleus<br>Plasma membrane | Thioredoxin | Cytoplasm<br>Extracellular<br>Mitochondria<br>Nucleus |
| 14-3-3 protein zeta/delta | Macromolecular complex | | |
| | Cytoplasm | Activated RNA polymerase II transcriptional coactivator p15 | Nucleus |
| | other intracellular organelles<br>Mitochondria<br>Nucleus<br>Cytoskeleton<br>Other | 60S ribosomal protein L9 | Ribosome<br>Nucleus |
| | | Arginine and glutamine -rich protein 1 | Nucleus |
| | | Heterogeneous nuclear ribonucleoprotein | Nucleus |
| Serine/arginine repetitive matrix protein 1 | Nucleus, cytoplasm | A2/B1 | Cytoplasm |
| | Cytoplasm | | |
| | | Splicing factor, proline- and glutamine -rich | Cytoplasm<br>Nucleus |

* Proteins identified by KNL-2 immunoprecipitation. The majority of those hits were not published in Lagana et al., 2010.

# CURRICULUM VITAE

## Publications

**De Rop V**, Dorn JF, Osborne MJ, Boisvert J, Ryan J, Padeganeh A, Moevus C, Borden KLB, Maddox AS, and Maddox PS (2014), Centromere epigenome stability is mediated by KNL-2 structural recognition of CENP-A chromatin**,** in revision at Developmental Cell.

Padeganeh A, **De Rop V**, Maddox PS (2013). Nucleosomal composition at centromere: a numbers game, *Chromosome Res*.Vol.21, Issue 1, p 27-36.

**De Rop V**, Padeganeh A, Maddox PS (2012). CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly, *Chromosoma*, Vol. 121, Issue 6, p 527-538.

Lagana A, Dorn JF, **De Rop V**, Ladouceur AM, Maddox AS, and Maddox PS (2010). A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A, *Nat Cell Biol*. Vol.12, p 1186-93.

## Scientific Presentations

### *Oral presentations*

April 2013: XIX Journée scientifique des programmes de biologie moléculaire de l'Université de Montréal.

November 2012 : 4ème Journée scientifique de l'IRIC, Université de Montréal

October 2011 : 28ème Journée du Département de pathologie et de biologie cellulaire.

May 2010 : XVIe Journée scientifique des programmes de biologie moléculaire de l'Université de Montréal.

January 2007 : Comité d'Organisation du Programme des Stagiaires d'Été (COPSÉ) de la Faculté de médecine de l'Université de Montréal.

### *Poster presentations*

#### International

December 2013: 53rd Annual meeting of the American Society for Cell Biology (ASCB), in New Orleans, USA.

December 2012: 52nd Annual meeting of the American Society for Cell Biology (ASCB), in San Francisco, USA

December 2011 : 51th Annual meeting of the American Society for Cell Biology (ASCB), in Denver, USA.

June 2011 : 2nd Dynamic Kinetochore Workshop, Bio-Center, Vienna, Austria.

December 2010 : 50th Annual meeting of the American Society for Cell Biology (ASCB), in Philadelphia, USA**.**

December 2009: 49th Annual meeting of the American Society for Cell Biology (ASCB), in San Diego, USA.

December 2008: 48th Annual meeting of the American Society for Cell Biology (ASCB), in San Francisco, USA**.**

**Regional**

February 2012 : The 2012 Annual McGill Biomedical Graduate Conference (AMBGC), Experimental medicine department, McGill University

February 2011 : The 2011 Annual McGill Biomedical Graduate Conference (AMBGC), Experimental medicine department, McGill University

February 2010: The 2010 Annual McGill Biomedical Graduate Conference (AMBGC), Experimental medicine department, McGill University.

**Academic**

October 2013: 30ème Journée du Département de pathologie et de biologie cellulaire de l'Université de Montréal.

May 2012 : XVIIIe Journée scientifique des programmes de biologie moléculaire de l'Université de Montréal.

June 2011: Symposium sur les développements technologiques des therapies par ARN, IRIC, Université de Montréal.

April 2011 : XVIIe Journée scientifique des programmes de biologie moléculaire de l'Université de Montréal.

March 2011 : 2ème compétition par affiche du Département de biochimie de l'Université de Montréal.

November 2010 : 3ème Journée scientifique de l'IRIC, Université de Montréal

March2010: 1ère compétition par affiche du Département de biochimie de l'Université de Montréal.

October 2009: 26ème Journée du Département de pathologie et de biologie cellulaire de l'Université de Montréal.

November 2008 : 2ème Journée de la recherche de l'IRIC, Université de Montréal