Decoding the molecular network between centromeric transcription, transcript and chromosomal instability

Pia Mihic

UNIVERSITÉ PARIS CITÉ

ÉCOLE DOCTORALE HÉMATOLOGIE, ONCOGENÈSE ET BIOTHÉRAPIES (HOB) – ED 561

Unité Epigénétique et Destin Cellulaire UMR 7216

Decoding the molecular network between centromeric transcription, transcript and chromosomal instability

Décodage du réseau moléculaire à l'intersection de la transcription centromérique, des transcripts et de l'instabilité chromosomique

Par **Pia MIHIC**

THÈSE DE DOCTORAT DE ONCOGENÈSE

Thèse dirigée par le Dr. Claire FRANCASTEL et le Dr. Ekaterina BOYARCHUK

Présentée et soutenue publiquement le 30 Septembre 2022

Membres du jury

Pr. Claire VOURC'H, Université Grenoble Alpes	Rapportrice
Dr. Clément CARRÉ, Sorbonne Université	Rapporteur
Dr. Florence LARMINAT, Université Toulouse III – Paul Sabatier	Examinatrice
Dr. Benoit PALANCADE, Université Paris Cité	Examinateur
Pr. Michaela FONTENAY, Université Paris Cité	Invité
Dr. Claire FRANCASTEL, Université Paris Cité	Directrice de thèse
Dr. Ekaterina BOYARCHUK, Université Paris Cité	Co-directrice de thèse

Acknowledgments

What an adventure! These four years of PhD have flown by, and I cannot believe I made it to this point. I always thought I would have the right words when the time comes... but now, I don't know how to put on paper all the immense gratitude I have for everyone who partook in this great voyage with me, whether for 4 years or 4 days. Maybe things will get soppy, but bear with me!

First of all, I want to thank all the members of my jury for accepting to evaluate my work. Many thanks to my rapporteurs Pr. Claire Vourc'h and Dr. Clément Carré, for taking interest in my work and for your constructive comments that enriched my work. Thank you to the examiteur Dr. Benoit Palancade, who also held the role of the president of the jury. A very big thank you to the examinatrice (and thesis committee member) Dr. Florence Larminat for all of our discussions and for reminding me that science is fun !

Thank you to my thesis committee members, Dr. Christian Jaulin, Dr. Christophe Escudé, and Dr. Paola Arimondo, that followed me during my PhD and had given me precious advice.

All of this would have not been possible without my host team:

Thank you to Dr. Claire Francastel, who took me in her team, despite my rough start, but based on my willingness to learn. I was a *tabula rasa,* and you though me so much.

Florent, thank you for the laughter and for not judging me (too much) for all the timers I've killed.... (that is pure coincidence, no causal link there!).

Guillaume! Thank you for all of your immense support over the years. I will never forget the text messages you would send me to check how I am doing and to remind me that I can do this. Your kindness helped me push and do my best, and get up when an experiment failed.

Baptiste, thank you for laughing with me every day (and sometimes listening to me complain). You are a big inspiration to me, your work ethics, your knowledge and determination thought me so much. I will miss asking you what do you have for lunch and then being appalled at your choices. I mean, how can you NOT like almonds? Or coconut? Okay, okay, I will make sure that at my pot de these there is some (vegan) sneackers =)

Marie-Cecile, you came into the lab (and in my life) towards the end, but also when I needed it the most. Our pottery sessions and lunches together were my safe space. Thank you for reminding me that there is life outside of the lab. You reminded me, every day, what I am worthy of =)

Helene, we shared a brief time together but it was a meaningful one. I had the pleasure of seeing you become a scientist. I hope you enjoy discovering R-loops =)

Katia ! I saved you for last because I don't know how to say thank you enough. You were a true mentor to me. You supported me, scientifically and otherwise, and your constant reminders of "don't panic" are what got me through to the end. Thank you for your optimism (both the real and not real ones). Thank you for reminding me that the length of my results is not the most important thing. Your knowledge of everything is truly fascinating. You have this particular way of seeing science and explaining it to me, that really shaped me in the last months and the way I think. No panic, we are almost at the finish line.

I am grateful to all the great UMR7216. It has been a real pleasure to work with you, to share with you weeks and weekends of work. Your smiling faces in the hallways really did make a difference. A big thank you to directrice Valerie Mezger who made such a positive difference in my PhD life ! Thank you for making sure I was in the right (head)space to finish my work.
Thank you Slimane for inviting me to your lab meetings and making me feel welcome, and for all the advice and support. Thank you to Jonathan for always taking the time to ask me how I am doing and for sharing poetry with me.
I am also very grateful for all of my fellow PhD collegues that helped me with protocols, advice (scientific or personal) and the time we spent together.
I am grateful to the Ministère de la recherche et de l'enseignement superieur as well as the Fondation pour la Recherche Médicale (FRM) for believing in me and my project and allow me to work for four intense years on this project.

Thank you to Chris Luno for the amazing electronic sets that made weekends in the lab (almost) fun. Thank you to Seannan McGuire for the amazing October Day series that kept me company in the tram, while waiting for cell transfections, or whenever I had 5min of spare time.

Now we get to the very personal thank yous, that have nothing to do with pipetting and everything to do with amazing people lending me their unquestionable support.
La famiglia! Dominik, Iva, Iris, Viktor. Hvala vam za vase navijanje, za to sto ste samnom slavili a neki puta i plakali. Uvijek ste mi bili velika ste mi potpora, kroz sve.

Roditelji: hvala sto ste vjerovali u mene. Znam da ste se brinuli za mene u “bijelom” svijetu, ali uspjela sam! Uvijek ste mi govorili da sam pametna, a sada sam to morala provjeriti za sebe.
Dajana i Miljenko, vi ste bili veliki dio ovog doktorata. Pratili ste me kroz transformaciju iz criceto incazzato u nekoga tko se zna boriti za sebe (na zen nacin). Nadam se da ce nam se putevi uskoro opet krizati =)
Matko, hvala ti.... za SVE! Pokazao si mi neumornu potporu, od prvog pa do zadnjeg dana, uvijek. Svaki puta kada bih pala, podsjetio bi me da se isto tako mogu i dici i nastaviti dalje. S tobom sam naucila o strpljenju, bez kojeg nebih uspjela ovaj pothvat. Ovaj doktorat je barem malim dijelom i tvoji.
Un grazie enorme ai miei Ritals. Antonio, Arianna, Flaminia, Manzo, Paolo, Stefano (ordine alfabetico). Mi avete accettat* senza esitare e mi avete sembre fatt* sentire un* dei vostri <3
E chi piu di voi poteva capirmi quando mi lamentavo dello studio matto e disperatissimo? Ma cosa ancora piu’ importante: abbiamo condiviso delle giornate/serate/pomeriggi bellissimi, pieni di gioia, riso, cibo, musica. Con voi, Parigi sembra anche come una citta’ carina =D

Anna, dal primo giorno all’ultimo, sempre insieme! Grazie per avermi fatto ridere tantissimo, per il tuo aiuto ogni volta quando ne avevo bisogno. Ancora un po, e ci sei anche tu! <3

Un grand merci a Josee, Camille, Johan, Theo, pour des vacances tres belles et merci de m'avoir accepte a bras ouverts.

Alix, my specialest goose. All our laughter together was a breath of fresh air that gave me all the energy to conquer the world (honk!). You made me discover so much of what I now love. It’s almost like we share one brain. Thank you for all the experiences we had and I cannot wait to discover what life has in store for us. Maybe it’s in a Berlin store ;) but for now: they are just (Paris) roommates! Back to serious business: thank you for never judging me, not for my (crazy) love for plants, not for my (crazier) love of pasta and pizza (hey, now you love them too =P), nor for my crazy butterflies that no one else understands. But you do, you always do. You are the thread that I follow to get back to myself.

Table of contents

ABBREVIATIONS

5caC – 5-carboxyl-cytosine

5fC – 5-formyl-cytosine

5hmC – 5-hydroxy-methyl-cytosine

5hmU – 5-hydroxy-methyl-uracil

5mC – 5-methyl-cytosine

5-AZA – 5-azacytidine

DNA – Deoxyribonucleic acid

RNA – Ribonucleic acid

ATM – Ataxia Telangectasia Mutated

ATR – Ataxia Telangectasia and Rad3 Related

CRISPR – clustered regularly interspaced short palindromic repeats

CCAN – Constitutive Centromere Associated Network

CDC – Cell Division Cycle

CDK – Cyclin Dependent Kinase

CENP – CENtromeric Protein

ChIP – Chromatin Immunoprecipitation

Chk – Checkpoint Kinase

CPC – Chromosomal Passenger Complex

DAPI – 4',6'-diamidino-2-phenylindole

DDR – DNA Damage Response

DNA-PK – DNA-dependent Protein Kinase

DNMT – DNA Methyltransferase

Dpc – days post coitum

DSB – Double-strand break

ENCODE – Encyclopedia of DNA Elements

FACT – Facilitates Chromatin Transcription

FISH – Fluorescence In Situ Hybridisation

HAC – Human Artificial Chromosome

HDAC – Histone Deacetylase

hnRNPs - Heterogeneous Nuclear Ribonucleoproteins

HP1 – Heterochromatin Protein 1

HURP – Holliday Junction Recognition Protein

IAP – Intracisternal A Particle

ICF – Immunodeficiency with Centromeric Instability and Facial Anomalies

Kb – Kilobase

KRAB – Kruppel-Associated box

LINE – Long Interspersed Element

Mb – Megabase

MEF – Mouse Embryonic Fibroblast

MEL – Murine Erythroleukemia cell

mESCs – murine Embryonic Stem Cell

mRNA – messenger RNA

MSRE – Methylation-Sensitive Restriction Enzymes

ncRNA– non-coding RNA

RdRp – RNA Dependent RNA Polymerase

RISC – RNA-induced Silencing Complex

RNAi – RNA Interference

RNA Pol – RNA Polymerase

rRNA – ribosomal RNA

Sat – Satellites

siRNA – small-interfering RNA

SINE – Short Interspersed Element

TE – Transposable Elements

TERRA – telomeric repeat-containing RNA

TET – Ten-Eleven Translocation enzymes

TKO – DNMT triple knock-out

tL1 - truncated L1 repeat

TLC – Telocentric repeat

tRNA – tranfer RNA

XIST – X-Inactive Specific Transcript

ZBTB – zinc finger and Broad complex, Tramtrack, and Bric-a`-brac

RÉSUMÉ COURT

Titre : Décodage du réseau moléculaire à l'intersection de la transcription centromérique, des transcripts et de l'instabilité chromosomique

Résumé :

Les centromères constituent un paradigme intéressant pour l'assemblage de domaines chromatiniens spécialisés. Ils servent de site de fixation pour les microtubules du fuseau et sont essentiels à la ségrégation fidèle du matériel génétique au cours des divisions cellulaires. Les centromères s'assemblent sur de grands réseaux de séquences répétées en tandem (répétitions satellites), peu conservés entre les espèces, mais sont définis épigénétiquement dans toutes les espèces par une variante d'histone H3 spécifique, CENP-A. Il est frappant de constater que les répétitions centromériques sont transcriptionnellement compétentes dans la plupart des organismes, bien qu'à de faibles niveaux dans des conditions physiologiques, les transcrits qui en résultent (cenRNAs) participant à l'architecture et à la fonction des centromères. En revanche, la transcription non programmée ou l'accumulation de cenRNAs est une signature pathologique des maladies d'instabilité chromosomique comme le cancer, en corrélation avec la perte de méthylation de l'ADN au niveau des répétitions satellites sous-jacentes. Pourtant, les liens fonctionnels n'ont jamais été formellement établis. J'ai d'abord entrepris de caractériser les cenRNA murins et leur transcription et j'ai constaté qu'ils sont préférentiellement transcrits dans l'orientation opposée aux télomères. De plus, le knock-down du brin le plus abondant semble plus délétère que la perte de l'autre, mettant en évidence leurs rôles potentiellement différents dans la fonction des centromères. Ensuite, pour disséquer les conséquences de l'augmentation de la transcription centromérique et de celle de l'accumulation de cenRNAs, sur la perte de fonction des centromères, nous avons mis en œuvre des outils dédiés à l'édition de l'épigénome pour manipuler la transcription au niveau des centromères. Mes travaux montrent que l'augmentation de la transcription centromérique conduit à l'apparition d'hybrides ADN-ARN aberrants qui sont corrélés avec des dommages à l'ADN au niveau des centromères, et aboutissent finalement à des défauts mitotiques et à une viabilité cellulaire altérée. Nous avons approfondi le lien entre la transcription aberrante des centromères et les erreurs mitotiques en montrant que les hybrides ADN-ARN non programmés sont la cause de la perte de l'architecture des centromères. En revanche, l'accumulation de cenRNAs ectopiques, globalement dans le noyau ou spécifiquement au

niveau du centromère, ne modifie pas l'architecture du centromère. Cependant, l'attachement des ARN aux centromères entraîne une augmentation des dommages à l'ADN du locus, probablement en altérant ses propriétés physiques et en déplaçant les protéines de leur site d'action. De plus, nous avons étudié la relation fonctionnelle entre l'hypométhylation de l'ADN et la transcription des centromères, en tirant parti du syndrome ICF (Immunodéficience, Instabilité Centromérique et Anomalies Faciales), une maladie génétique humaine rare, où l'intégrité compromise des centromères et l'instabilité chromosomique sont directement liées à l'hypométhylation de l'ADN au niveau des répétitions sous-jacentes. À cet égard, nos résultats montrent que si la méthylation de l'ADN n'est ni suffisante ni nécessaire pour une transcription accrue des centromères, elle la facilite.

Mots clefs : centromères ; ARN ; transcription ; boucles R ; syndrome ICF ;

ABSTRACT

Title : Decoding the molecular network between centromeric transcription, transcript and chromosomal instability

Abstract :
Centromeres provide an interesting paradigm for the assembly of specialized chromatin domains. They serve as the attachment site for spindle microtubules and are essential for faithful segregation of genetic material during cell divisions. Centromeres assemble on large arrays of tandemly repeated sequences (satellite repeats), poorly conserved across species, but are epigenetically defined in all species by a specific histone H3 variant, CENP-A. Strikingly, centromeric repeats are transcriptionally competent in most organisms, although at low levels in physiological conditions, with the resulting transcripts (cenRNAs) participating in centromere architecture and function. In contrast, unscheduled transcription or accumulation of cenRNAs is a pathological signature of chromosomal instability diseases like cancer, in correlation with DNA methylation loss at the underlying satellite repeats. Yet, functional links have never been formally established. I first set to characterize the murine cenRNAs and their transcription and found that they are preferentially transcribed in the orientation away from telomeres. Moreover, the knock-down of the more abundant strand seems more deleterious than loss of the other, highlighting their potentially different roles in centromere function. Next, to dissect the consequences of increased centromere transcription and that of the accumulation of cenRNAs, on centromere loss of function, we implemented dedicated epigenome editing tools to manipulate transcription at centromeres. My work shows that increased centromeric transcription leads to the appearance of aberrant DNA-RNA hybrids which correlate with DNA damage at centromeres, and finally culminates in mitotic defects and impaired cell viability. We have further dissected the link between aberrant centromere transcription and mitotic errors by showing that unscheduled DNA-RNA hybrids are causal to loss of centromere architecture. On the other hand, accumulation of ectopic cenRNAs, globally in the nucleus or specifically at the centromere, does not alter centromere architecture. However, tethering RNAs to centromeres results in increased DNA damage of the locus, possibly by altering its physical properties and displacing proteins from their site of action. Moreover, we have investigated the functional relationship between DNA hypomethylation and centromere transcription, by taking advantage of the ICF (Immunodeficiency, Centromeric Instability and Facial Abnormalities) syndrome, a rare

human genetic disease, where compromised centromere integrity and chromosomal instability are directly linked to DNA hypomethylation at underlying repeats. In that regard, our results show that while DNA methylation is not sufficient nor necessary for increased centromere transcription, it facilitates it.

Keywords : centromeres ; RNA ; transcription ; R-loops ; ICF syndrome ;

RÉSUMÉ SUBSTANTIEL

Pour fonctionner correctement et survivre, les cellules doivent relever le défi de maintenir l'intégrité de leur génome et d'assurer la transmission correcte de l'information génétique à chaque cellule fille après chaque division cellulaire. Cela dépend en partie du maintien de l'intégrité des centromères, qui sont les domaines chromosomiques servant de plateformes pour l'assemblage des kinétochores, des complexes multiprotéiques responsables de la fixation des chromosomes au fuseau mitotique et de la ségrégation fidèle du matériel génétique dans les cellules en division. Chez les mammifères, les centromères sont caractérisés par la présence de séquences satellites, qui s'organisent comme de grands domaines composés de séquences hautement répétées en tandem, et sont généralement intégrés dans des domaines d'hétérochromatine péricentrique. Leur séquence est peu conservée entre espèces, c'est pourquoi les centromères sont définis épigénétiquement dans toutes les espèces par une variante spécifique de l'histone H3, CENP-A (Cenp-A chez la souris). De manière frappante, les répétitions centromériques sont transcriptionnellement compétentes dans la plupart des organismes, bien qu'à des niveaux faibles dans des conditions physiologiques. Dans ce contexte, mon équipe d'accueil a montré que les niveaux d'ARN centromériques (cenARN) murins sont régulés au cours du cycle cellulaire, atteignant un pic au début de la mitose où ils sont nécessaires à l'identité et à la fonction des centromères. Les transcrits font partie intégrante de la chromatine CENP-A et participent au recrutement de complexes protéiques qui assurent la ségrégation correcte des chromosomes mitotiques. Comme d'autres séquences répétées du génome des mammifères, les centromères sont méthylés dans toutes les cellules somatiques. La méthylation de l'ADN est l'un des principaux mécanismes qui contribuent à leur répression transcriptionnelle et à la protection du génome contre la recombinaison mitotique qui peut se produire entre les éléments répétés. Ainsi, la méthylation de l'ADN et l'intégrité des centromères sont deux facettes majeures du maintien de l'intégrité génomique. Des arguments importants en ce sens viennent de l'existence de maladies d'instabilité chromosomique, comme les cancers, où les schéma de méthylation sont souvent altérés. Dans ces contextes, les altérations de la méthylation de l'ADN sont corrélées à des altérations structurelles des centromères, en association avec des aberrations chromosomiques numériques et structurelles. Ces altérations sont également corrélées à des niveaux anormalement élevés de transcrits provenant de répétitions centromériques. Dans ce contexte, mon équipe d'accueil a fourni un premier indice d'un lien de causalité par lequel

l'accumulation ectopique de cenARN favorise l'instabilité centromérique et perturbe l'organisation nucléaire. Il apparaît alors que le maintien de la stabilité du centromère est lié au contrôle de sa transcription. Cependant, le rôle de la méthylation de l'ADN dans la prévention de l'activation transcriptionnelle des répétitions centromériques, et de leur recombinaison illicite, n'a toujours pas été formellement déchiffré. L'hypométhylation pathologique des répétitions satellites est une caractéristique de maladies complexes et multifactorielles comme le cancer, mais aussi d'une maladie rare du développement, l'ICF (Immunodéficience, instabilité Centromérique, anomalies Faciales). Les études sur l'étiologie du syndrome ont identifié des facteurs nécessaires au maintien de la méthylation de l'ADN au niveau des centromères et à la stabilité chromosomique : DNMT3B, une ADN méthyltransférase de novo (DNMT) et trois autres facteurs dépourvus d'activité DNMT et de fonction inconnue. Le syndrome ICF offre ainsi un contexte monogénique remarquable où l'altération de l'intégrité des centromères est directement liée à l'altération du statut de méthylation de l'ADN des répétitions satellites sous-jacentes, fournissant de nouveaux facteurs à considérer et remettant en question les liens fonctionnels qui conduisent à la perte de stabilité du génome.

De plus, des perturbations de la fonction des centromères ont également été documentées lors de dommages à l'ADN. En effet, le stress génotoxique entraîne une activation transcriptionnelle des répétitions centromériques qui provoque une désorganisation structurelle drastique de la chromatine centromérique. De manière importante, mon équipe hôte a montré que cette activation transcriptionnelle au niveau des centromères a des conséquences fonctionnelles distinctes sur les phénotypes cellulaires en fonction du contexte cellulaire. Dans les cellules primaires, le stress génotoxique induit une activation transcriptionnelle rapide des centromères, conduisant à une délocalisation frappante de Cenp-A loin de son emplacement par défaut et à un arrêt prématuré du cycle cellulaire. En revanche, en l'absence d'un point de contrôle p53, la transcription centromérique n'est pas activée, Cenp-A reste aux centromères, les cellules ne s'arrêtent pas mais accumulent des micronoyaux indiquant une instabilité chromosomique. En somme, le désassemblage des centromères déclenché par l'activation transcriptionnelle rapide des centromères sert de mécanisme de sauvegarde, mais il reste à déterminer si les dommages à l'ADN ou la transcription en soi déclenchent la perte d'identité des centromères.

Ces événements pathologiques au niveau des centromères coexistent dans différents contextes pathologiques, tels que le cancer ou le syndrome ICF. Cependant, leurs relations de cause à

effet et la manière dont ils conduisent à l'instabilité chromosomique restent largement inconnues.

Les objectifs généraux de mon projet de doctorat étaient d'établir des liens fonctionnels entre les différents événements pathologiques qui peuvent causer le dysfonctionnement des centromères : perte de la méthylation d'ADN aux répétitions satellites, apparition de dommages à l'ADN, augmentation de la transcription à ce locus et accumulation de cenARN. Pour éviter de dépendre des contextes cellulaires et de leurs effets secondaires, ma stratégie principale est de découpler ces différents événements en utilisant des outils d'édition (épi)génomique dédiés, dérivés de la technologie Crispr/Cas9 pour : 1. l'activation transcriptionnelle sélective en utilisant un Cas9 catalytiquement inactif (dCas9) fusionné à des domaines d'activation transcriptionnelle (VP64). 2. le ciblage des cenARN directement sur les centromères, en l'absence d'une transcription accrue au niveau des centromères. J'ai d'abord réalisé une caractérisation approfondie des ARN provenant des centromères en utilisant la RT-qPCR spécifique au brin, et la RNA-FISH. J'ai montré que les cenARN murins peuvent provenir des deux brins, avec une prédominance des cenARN Forward par rapport au bras long du chromosome. De plus, j'ai découvert que les cenARN Forward et Reverse n'ont pas la même stabilité d'ARN, où les niveaux des cenARN Forward diminuent plus rapidement que les cenARN Reverse, dans les premières heures après l'inhibition de la transcription. Il est intéressant de noter que leurs niveaux s'équilibrent progressivement et diminuent à la même vitesse, ce qui soulève des questions intéressantes sur leur régulation transcriptionnelle et leur stabilité. Comme les cenARN Forward constituent la population la plus abondante du cenARN, j'ai activé la transcription de ce brin et évalué son impact sur la fonction des centromères et les phénotypes cellulaires. Mes travaux montrent que l'augmentation de la transcription centromérique conduit à l'apparition d'hybrides ADN-ARN aberrants qui sont en corrélation avec des dommages à l'ADN au niveau des centromères. Une augmentation soutenue de la transcription (48h) entraîne une perte de l'architecture appropriée des centromères, visualisée par une localisation perturbée de Cenp-A, et une décondensation de l'ADN centromérique, et aboutit à des défauts mitotiques et une viabilité cellulaire altérée. Nous avons disséqué davantage le lien entre la transcription aberrante du centromère et les erreurs mitotiques en montrant que les hybrides ADN-ARN non programmés sont responsables de la perte de l'architecture du centromère. En effet, leur élimination rétablit la localisation correcte de Cenp-A et corrige les erreurs mitotiques. Puisqu'il a été démontré que les boucles R centromériques sont essentielles pour une division mitotique correcte, j'ai

cherché à déterminer dans quelle phase du cycle cellulaire la transcription centromérique produirait des boucles R aberrantes. Mes résultats montrent qu'une forte accumulation de boucles R en G2, avant le début de la mitose, a un effet délétère sur la ségrégation fidèle des chromosomes. Ce résultat pourrait s'expliquer par la perte de rigidité du centromère, dont on sait qu'elle est importante pour contrer les forces de traction exercées par les microtubules sur les chromosomes en mitose.

La transcription du centromère produit des cenARN qui sont impliqués dans les boucles R, mais aussi certains qui ne sont pas liés à la chromatine. Par conséquent, je me suis également interrogé sur l'effet de l'accumulation de cenARN solubles, ceux qui ne sont pas impliqués dans les boucles R, sur le dysfonctionnement des centromères. Il est intéressant de noter que l'accumulation de cenARN ectopiques, globalement dans le noyau ou spécifiquement ciblés sur le centromère, ne modifie pas l'architecture du centromère. Cependant, l'attachement de tous les ARN testés (contrôle ou cenARN) aux centromères entraîne une augmentation des dommages à l'ADN du locus, probablement en altérant ses propriétés physiques et en déplaçant les protéines de leur site d'action. Ces résultats suggèrent que l'augmentation de la transcription des centromères et l'accumulation de cenARN conduisent toutes deux à une altération de la fonction des centromères, mais peut-être par des mécanismes moléculaires distincts. À l'inverse, l'élimination des cenARN a eu un impact important sur la viabilité des cellules. La perte de cenARN Forward a eu un effet phénotypique plus drastique, les cellules apparaissant plus plates. et ont commencé à mourir, alors que les cellules ont toléré plus facilement la perte des cenARN Reverse. L'effet différent de la perte de l'une ou l'autre population de cenRNAs ouvre de nouvelles et intéressantes questions sur les partenaires protéiques potentiellement distincts des cenARN Forward ou Reverse.

Étant donné le rôle essentiel de la méthylation de l'ADN au niveau des ADN répétés dans la régulation de leur statut transcriptionnel, j'ai évalué si l'hypométhylation de l'ADN au niveau des centromères serait plus propice à des niveaux plus élevés de cenRNAs. Pour répondre à cette question, j'ai utilisé des cellules immortalisées provenant de modèles murins du syndrome ICF, qui sont hypométhylées au niveau des centromères. Bien que ces cellules présentent des niveaux de cenARN similaires à ceux de leurs homologues de type sauvage, le fait de cibler une transcription centromérique accrue dans ce contexte cellulaire a entraîné des niveaux plus élevés de cenARN. Ces résultats impliquent donc que l'hypométhylation de l'ADN n'est pas suffisante ou nécessaire à l'augmentation de la transcription centromérique, mais qu'elle semble la faciliter.

L'accumulation de cenARN est une caractéristique courante des conditions pathologiques, mais on ne sait toujours pas si elle représente une force motrice dans l'oncogenèse. J'ai montré que la surexpression de cenARN murins ectopiques Forward et Reverse pouvait être nécessaire et/ou suffisante pour le début de la sénescence, rappelant la sénescence induite par les oncogènes. Dans l'ensemble, l'établissement des liens de causalité entre l'accumulation des cenARN et l'état cancéreux mérite attention, car ils pourraient représenter de futurs biomarqueurs intéressants pour le pronostic de la maladie. Une exploration plus approfondie des liens de causalité entre les différents événements pathologiques au niveau des centromères et la (dys)fonction des centromères promet d'aboutir à des découvertes intéressantes, tant pour la recherche clinique que pour la science fondamentale.

INTRODUCTION

1. General overview of the (epi)genome

Genome composition

Every living organism possesses a genome that stores all the genetic information necessary to carry out cellular functions and sustain their life. In eukaryotes, this genomic DNA is packaged into the cell nucleus, in the form of a double helix of deoxyribonucleic acid (DNA) consisting of four nucleotides: adenine, thymine, guanine and cytosine (A, T, G, C) (Franklin and Gosling, 1953; Watson and Crick, 1953; Wilkins and Randall, 1953). The genome of eukaryotes is composed of both unique DNA sequences, that contain the genetic coding material along with a variety of diverse regulatory elements, and repetitive DNA sequences which may be present at high copy numbers. The latter were largely considered for many decades as non-functional DNA. Indeed, the first draft of the human genome has revealed that less than 2% of the human genome corresponds to protein coding genes, while 98% of the genome was composed of "junk DNA" without a known function (Lander *et al.*, 2001; Palazzo and Lee, 2015). A new road map was provided in 2012 by the Encyclopedia of DNA Elements (ENCODE) project, which estimated that 80% of the human genome is transcribed or may hold regulatory functions (ENCODE Project Consortium, 2012), which overrode the idea that "junk DNA" has no functional relevance. This non-coding DNA is mainly represented by repeated sequences (or DNA repeats) of different sizes, scattered between (or within) the genes (http://www.repeatmasker.org). Based on their distribution in the genome, DNA repeats can be divided into two categories: interspersed repeats, where individual repeat units are distributed around the genome (Choo, 2001; Ewing *et al.*, 2020; Sullivan *et al.*, 1996), and tandemly repeated DNA, where repeat units are adjacent to each other, forming megabase-scale arrays (Padeken *et al.*, 2015). Moreover, the first gapless, complete genome assembly, created by the Telomere-to-Telomere (T2T) consortium, has identified that approximately 54% of the human genome is composed of repetitive DNA (Hoyt *et al.*, 2022) (**Figure 1.a**), and it has enriched the repeat annotation of the human genome, especially for the tandem repeats. Tandem repeats underlie essential chromosomal landmarks, such as chromosome ends called telomeres (Padeken *et al.*, 2015), and at centromeres which are essential for the faithful chromosome segregation in dividing cells (McKinley and Cheeseman, 2016; Pluta *et al.*, 1995; Thakur *et al.*, 2021). In the mouse, model organism I used for my thesis work, pericentromeric Major satellites cluster together in the interphase nucleus,

forming cytologically detectable structures called chromocenters (Hsu *et al.*, 1971), whereas the centromeric Minor satellites form discrete foci around them (Guenatri, 2004) (**Figure 1.b**). Both centromeres and pericentromeres are key elements for the correct segregation in cell division (Pidoux and Allshire, 2000), and their roles will be described in more detail in Chapter 2.

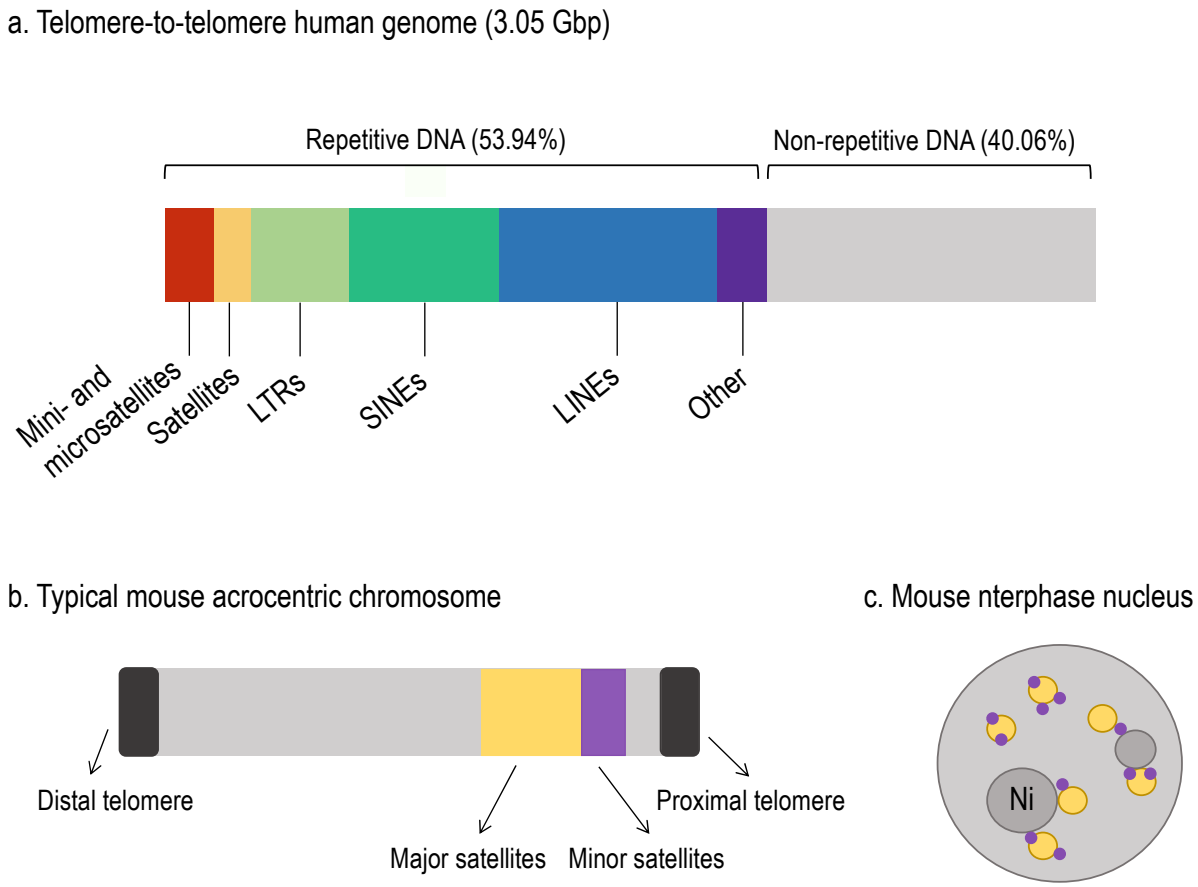


Figure 1. Types and distribution of repetitive DNA elements.
a. Bar representing the composition of the human genome: repetitive DNA sequences occupy around 54%, which includes mini and microsatellites, satellites, LTRs, SINEs and LINEs. Modified from Haws *et al.*, 2022.
b. Organization of murine Major and Minor satellite repeats on mouse acrocentric chromosomes. A typical acrocentric chromosome (gray) features telomeres at both ends (black). Major (yellow) and Minor (purple) satellites are located in proximity of the proximal telomere.
c. In murine somatic interphase nuclei, pericentric (Major satellite) heterochromatin domains cluster together to form chromocenters (yellow). Minor satellites form distinct domains (purple) adjacent to chromocenters. In the interphase nucleus, Minor satellites are often located in proximity of nucleoli (Ni). **b**. and **c**. are modified from Probst and Almouzni, 2011.

Both the unique DNA sequences and the repetitive ones are subjected to multiple nuclear functions: replication, repair, transcription and splicing of the resulting transcripts. These events occur in the confines of the nucleus, where approximately 6 billion nucleotides, in the case of the human diploid genome, are tightly packaged in a hierarchical manner. It starts with the flexible DNA polymer of 2 nm, which in eukaryotic genomes is not naked, but is instead associated with a family of basic proteins called histones. Histones are all rich in positively charged residues (lysine, K and arginine, R), allowing the interaction with the negatively charged backbone of the DNA. The DNA and histones are packaged into nucleosomes, the basic building blocks of chromatin, forming the 11 nm "beads on a string" structure. In each nucleosome, approximately two turns of DNA (146 bp) wrap around an octamer of core histone proteins consisting of a central $(H3/H4)_2$ tetramer flanked on either side by two H2A/H2B dimers. Core histones contain N-terminal and C-terminal unfolded tails that extend away from the nucleosome core and are subjected to many different post-translational modifications (Luger *et al.*, 1997). The H1 family of histones, called linker histones because they associate with linker DNA between two neighbouring nucleosomes, promotes the organization of nucleosomes into a 30 nm fiber (Hergeth and Schneider, 2015). The 30 nm structure folds further to form metaphase chromosomes, the most condensed form of DNA (Moser and Swedlow, 2011) (**Figure 2**).

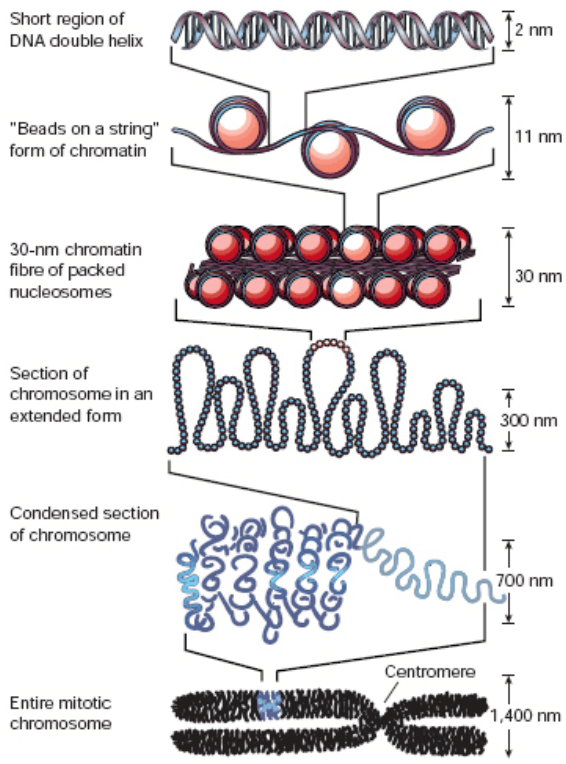


Figure 2. Chromatin folding.
The first layer of compaction corresponds to the "beads on a string" form, which can be further compacted to form a 30 nm fiber. Subsequent compaction of DNA fibers fives rise to fully compacted mitotic chromosomes (taken from Felsenfeld and Groudine, 2003).

While the folding of metaphase chromosomes is locus-independent and is homogeneous between different chromosomes (Naumova *et al.*, 2013), this is not the case for the interphase chromatin, which is not homogeneously organized in the nucleus. Instead, it is organized into distinct domains which are defined by their level of compaction, which are further associated with distinct regulatory functions. The typically transcription-permissive form, called euchromatin, is characterized by relatively loose compaction. It is generally gene-rich and replicates early in S phase (Goldman *et al.*, 1984). This form allows access of the genome to the machineries for replication, transcription or repair. Conversely, heterochromatin is highly condensed and dense to electrons in electron microscopy, generally poor in genes and replicates late in S phase (Goldman *et al.*, 1984) (**Figure 3**). Additionally, heterochromatin can be subdivided into two different classes: constitutive and facultative heterochromatin. Constitutive heterochromatin forms at the same genomic regions in every cell type, regardless of the development or cell cycle stages. It is mostly composed of tandem repetitive elements, such as those found at pericentromeres and at telomeres (Saksouk *et al.*, 2015). In contrast, facultative heterochromatin regions can change their chromatin state depending on the developmental stage and cellular stimuli (Penagos-Puig and Furlan-Magaril, 2020).

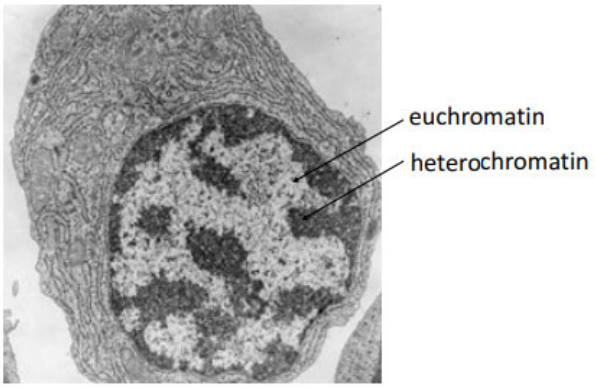


Figure 3. Euchromatin and heterochromatin in the interphase nucleus.
Electron microscopy image of a cell nucleus, where euchromatin and heterochromatin correspond to light and dark regions, respectively (Euchromatin Dr. Jastrow's electron microscopy atlas).

Not surprisingly, the condensation of the genome represents an obstacle to DNA accessibility for processes such as transcription, replication, and DNA repair (Clapier and Cairns, 2009). To deal with this problem, the genome needs to be dynamic in terms of chromatin structure, in order to expose underlying DNA sequences to genome maintenance factors and transcription and replication machineries. This implies the existence of mechanisms that control the chromatin state and, by consequence, its functions and transcriptional output.

A field guide to epigenetics

In multicellular organisms, all the cells carry the same genetic information, with the exception of cells that undergo programmed DNA elimination (Wang and Davis, 2014). To form the different cell types of an organism, this genetic information needs to be differentially exploited. This is largely achieved by the contribution of epigenetic mechanisms which control the proper organization and dynamics of the chromatin, to ensure accurate genome function. Epigenetics is an umbrella term for molecular mechanisms that, in response to internal and external cues, ensure the differential exploitation of genetic information to determine cell identity and fate. This term was introduced by Conrad Waddington in 1942 and defined as "the study of the mechanisms of development through which genes bring about phenotypic effects" (Waddington, 2012). Since then, this term been subjected to different iterations. The modern and generally accepted definition of epigenetics is the study of changes in gene function that are stably inherited and that do not lead to a change in the DNA sequence (Berger *et al.*, 2009).

Several molecular mechanisms mediate epigenetic phenomena including: histone variants and their post transcriptional modifications (PTMs), covalent DNA modifications such as DNA methylation, and non-coding RNAs (ncRNAs) (Felsenfeld, 2014; Feng and Riddle, 2020) (**Figure 4**). The interplay of different epigenetic mechanisms is essential to ensure that genome stability and function is maintained (Carmona and Michan, 2016; Feng and Riddle, 2020). Conversely, loss of genome stability is often linked to the breakdown of these epigenetic mechanisms (Aunan *et al.*, 2016; Jeffries, 2020; Pal and Tyler, 2016).

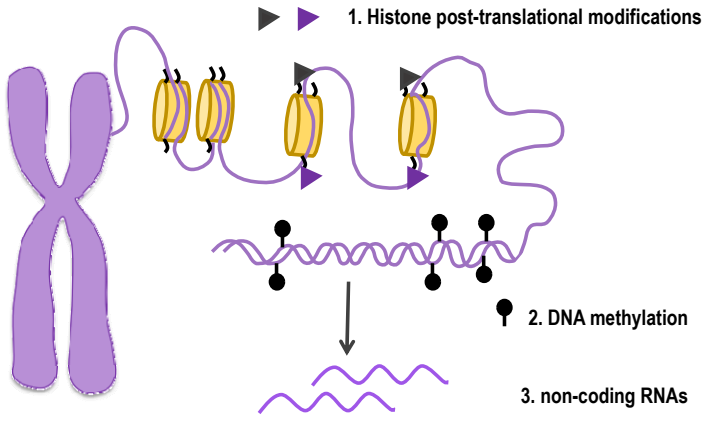


Figure 4. An overview of the most known epigenetic mechanisms regulating gene expression.
1. The regulation of chromatin structure and function through PTMs (arrowheads) to histone proteins (yellow circles).
2. DNA modifications such as DNA methylation, which results from the addition of a methyl group to cytosines (black lollipop).
3. non-coding RNAs also participate in the regulation of cellular processes.

Histones, their variants and their post-transcriptional modifications

Since histones are closely associated with DNA, they play an important role in the local chromatin structure and environment, which in turn facilitates access to transcription, replication or repair factors. This largely occurs through a variety of PTMs that are placed on the histone N-terminal tails exposed on the nucleosome surface. Moreover, the presence of histone variants adds an additional level of complexity and regu1ation of the chromatin.

Histone variants

Except for the histone H4, all core histones have variant counterparts (Talbert and Henikoff, 2021). Histone variants participate in the dynamic organization of chromatin during transcription and DNA replication, and can function as temporary place-holders for other histones (Buschbeck and Hake, 2017; Gurard-Levin and Almouzni, 2014).

Some histone variants occupy specific chromosomal regions, like the inactive X chromosome enriched in macroH2A (Costanzi and Pehrson, 1998), or centromeres that are defined by the histone H3 variant CENP-A (Earnshaw and Rothfield, 1985; Palmer *et al.*, 1991). CENP-A is the most specialized as well as the most divergent histone H3 variant (Allshire and Karpen, 2008; Black and Bassett, 2008; Guldner *et al.*, 1984). Interestingly, this rapid evolution of CENP-A is linked to the co-evolution of the underlying centromeric satellite DNA (Cooper and Henikoff, 2004; Talbert *et al.*, 2004). CENP-A epigenetically defines centromeres, and it acts as the platform for the assembly of the kinetochore, a multiprotein complex that connects mitotic chromosomes to microtubules of the mitotic and meiotic spindles (Musacchio and Desai, 2017). Perhaps not surprisingly, CENP-A depletion prior to mitosis results in the mislocalization of most kinetochore proteins (Fachinetti *et al.*, 2013). However, CENP-A is not required for the mitotic function of centromeres once the kinetochore is already assembled (Hoffmann *et al.*, 2016). Outside of its mitotic role, CENP-A seems to protect centromeric DNA integrity, since its abrogation results in chromosome breakage, recombination, and translocations (Giunta *et al.*, 2021a). More on the role of CENP-A at centromeres is discussed in Chapter 2.

Along with its centromeric variant, histone H3 comprises at least eight different proteins in human (Maze *et al.*, 2014; Sitbon *et al.*, 2017) (schematized in **Figure 5**). The two most known are the H3.1 and H3.3 (Malik and Henikoff, 2009). H3.1 is known as the replicative variant because it is incorporated during the S-phase, whereas H3.3 is the non-replicative variant, deposited in a cell cycle-independent manner (Borg *et al.*, 2021). In addition, these

two forms vary also in the genomic positions they occupy: while H3.1 is present throughout the genome, H3.3 is more specifically associated with promoters, where it is subjected to high turnover, associated with high transcriptional activity of these genes (Goldberg, 2010).

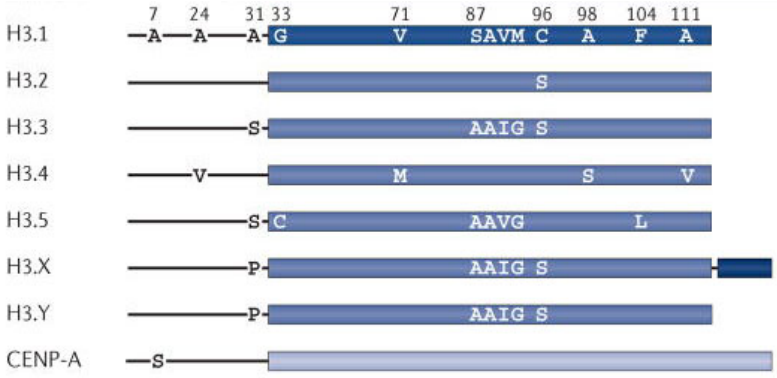


Figure 5.
Schematic representation of the diversity of histone variants. Different shades illustrate protein sequence similarity. Letters illustrate differences in specific residues. Adapted from (Maze *et al.* 2014).

Histone variants also play a crucial role in sensing DNA damage (Ferrand *et al.*, 2021). This is the case of H2AX, a key player in the DNA damage response (discussed in Chapter 3). Indeed, upon DNA damage, H2AX is phosphorylated on serine 139 (Li and Fang, 2015). This phosphorylated form of H2AX is also called γH2AX, since it was discovered in cells irradiated with gamma rays (Rogakou *et al.*, 1998). γH2AX marks the damaged DNA and recruits DNA damage response proteins on this site (Stucki *et al.*, 2005).
Moreover, in the last few years, several mutated histone forms have been reported and associated with cancers, and lead to the concept of oncohistones (Amatori *et al.*, 2021).

Histone PTMs
Histone PTMs are reversible modifications which depend on the fine balance between two distinct enzymatic activities, the deposition or removal performed by histone Writers or Erasers, respectively (Kouzarides, 2007). Histone PTMs are then recognized and bound by other proteins, commonly known as Readers, in order to give way to specific outcomes in terms of chromatin compaction and biological function (Nicholson *et al.* 2015). The histone PTMs include acetylation, phosphorylation, methylation, ubiquitylation, ADP-ribosylation and SUMOylation, or some lesser known modifications such as crotonylation and citrullination (Kleinsmith *et al.*, 1966; Ord and Stocken, 1967; Shiio and Eisenman, 2003; Stanley *et al.*, 2001). Table 1. briefly summarizes some of the best-known mammalian histone H3 marks and their biological function.

Histone, residue, modification	Function	Genomic element	Reference
H3K4me1/2/3 H3K9me1 H3K27me1/2 H3K36me1/3	Active transcription	Promoters /enhancers Gene bodies	(Beacon *et al.*, 2021; Ferrari *et al.*, 2014; Saha and Muntean, 2021)
H3K27me3 H3K9me2 H3K9me3	Repressed transcription	Promoters/enhancers Repeated sequences (pericentromeric repeats, transposons)	(Becker *et al.*, 2016; Guo *et al.*, 2013; Padeken *et al.*, 2015)
CENP-A-S7ph	Mitotic progression	Human centromeres	(Goutte-Gattat *et al.*, 2013; Kunitoku *et al.*, 2003)
CENP-A-K105ac	CENP-A deposition	Centromeres	(Bui *et al.*, 2012)

Table 1. Summary of post-transcriptional modifications on histone H3 in mammals, highlighting their chromatin function and the DNA elements on which they are found.

Importantly, one residue can be modified in multiple ways. However, not all PTMs can coexist on the same histone at the same time (Voigt *et al.*, 2012). Still, the resulting large number of possible combinations gave rise to the hypothesis of a "histone code", where histone modifications would work, singularly or in combination, to affect gene regulation (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Histone PTMs can directly influence chromatin structure. For example, acetylation and phosphorylation can reduce the positive charge of histones and will likely weaken histone–DNA interaction, resulting in increased nucleosome mobility (Cosgrove *et al.*, 2004). Another modality of how histone PTMs influence chromatin is by recruiting readers (Taverna *et al.*, 2007).

Certain PTM readers function as architectural proteins, with the potential to induce chromatin compaction or serve as a physical block to the underlying DNA. For instance, the heterochromatin protein 1 (HP1) can spread across a large region through self-propagation and oligomerization (Bannister *et al.*, 2001). Different isoforms of HP1 have been shown to perform a multitude of functions, ranging from transcriptional regulation (Kwon and

Workman, 2011), RNA splicing (Yearim *et al.*, 2015), sister chromatid cohesion (Dinant and Luijsterburg, 2009), chromosomal segregation (Azzaz *et al.*, 2014), or telomere maintenance (Chow *et al.* 2018) to name a few.

Other readers can determine the functional outcome of the histone PTMs by recruiting additional factors, such as the transcriptional or DNA damage repair machineries (Musselman *et al.*, 2012). Moreover, PTM readers can also crosstalk with, and guide, the DNA methylation machinery (discussed in detail in the next section) (Ben-Porath and Cedar, 2001; Fuks, 2005).

DNA methylation

DNA methylation is one of the most studied and best characterized epigenetic marks. In mammals, DNA methylation is essential for embryonic development (Li *et al.*, 1992; Okano *et al.*, 1999), thus strongly indicating its key role in molecular mechanisms involved during embryogenesis. Moreover, DNA methylation is key for maintenance of genome stability, where 70% – 80% of the somatic mammalian genome is methylated (Ehrlich *et al.*, 1982; Lister *et al.*, 2009). Historically associated with gene repression, DNA methylation is now considered to have a critical role in transcriptional regulation, genomic imprinting, X-chromosome inactivation and genomic stability by silencing and neutralizing repeated DNA elements (Greenberg and Bourc'his, 2019).

The how: establishment and erasure of DNA methylation patterns

In mammals, DNA methylation occurs predominantly on cytosines within a CpG dinucleotide context. It is catalyzed by DNA methyltransferases enzymes (DNMTs) that transfer a methyl group from the S-adenyl methionine (SAM) donor, to the fifth carbon of a cytosine residue to produce 5-methylcytosine (5mC) (Suzuki and Bird, 2008). DNA methylation patterns are established by *de novo* DNMTs, DNMT3A and DNMT3B, during early embryonic development (Goll and Bestor, 2005). During this process, another DNMT3 enzyme, devoid of catalytic activity, called DNTM3-Like (DNMT3L), is known to positively stimulate DNMT3A/DNMT3B activities (Growher 2005). Once established, DNA methylation profiles are maintained through cell divisions by the maintenance DNMT called DNMT1 (Ren *et al.*, 2018).

DNA methylation is highly dynamic during mammalian development. Indeed, in germ cells and pre-implantation embryos, the 5mC patterns are first erased and then established *de novo*, a process referred to the epigenetic reprogramming (Du *et al.*, 2022). Shortly after fertilization, parental genomes are demethylated with different kinetics. The paternal genome is demethylated by an active process, independent of cell division, hence termed "active demethylation" (Cimmino and Aifantis, 2017). It relies on demethylating enzymes from the Ten-eleven translocation (TET) methylcytosine dioxygenase family. TET proteins can cause the sequential oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Ito *et al.*, 2011). In contrast, the maternal genome is demethylated by a passive, replication-dependent mechanism (Eckersley-Maslin *et al.*, 2018; Mayer *et al.*, 2000; Oswald *et al.*, 2000). Only a few loci, such as imprinted genes, are not affected by this wave of demethylation. 5mC levels reach their lowest point during the blastocyst stage, after which the methylation status is gradually re-established during gastrulation, on almost the entire genome, except CpG islands which remain protected (Deaton and Bird, 2011). During primordial germ cell (PGC) formation, 5mC is globally erased and re-established once again, this event being essential for correct establishment of genomic imprinting (Guibert *et al.*, 2012) (**Figure 6**).

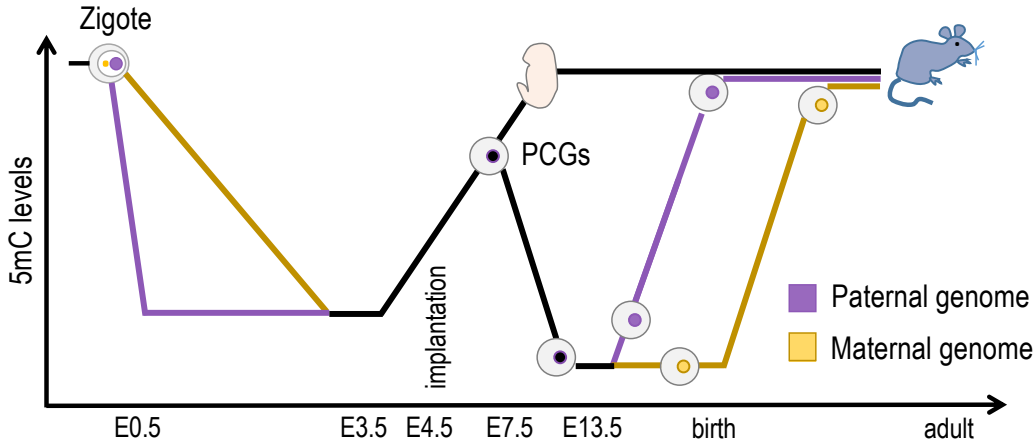


Figure 6. Waves of DNA methylation reprogramming during development. Immediately following fertilisation, global levels of DNA methylation decrease to reach a low point in pre-implantation balstocysts. In the paternal genome (purple), it involves a process of active demethylation and is accompanied by high levels of 5-hydroxymethylation. Conversely, DNA methylation in the maternal genome (yellow) is lost by a passive, replication dependent process. Note that many sequences (such as IAP retroelements) partially escape methylation erasure after fertilization. New patterns of DNA methylation are established following implantation by *de novo* DNMT3 enzymes. Primordial germ cells (PGCs) in the developing embryo undergo a second wave of global demethylation that is complete by embryonic day 13, after which female germ cells enter meiosis at male germ cells mitotic arrest. Modified from Auclair and Weber 2012.

The where: genomic elements enriched in DNA methylation

The genome is characterized by the presence of DNA stretches with a high CpG density, called CpG islands (CGI). These CGIs are non-randomly distributed and are present at the promoter regions of 60% of all human genes (Illingworth and Bird, 2009). Several studies showed that only a small fraction of CGIs seems to be methylated, even in terminally differentiated cells (Rollins, 2005; Weber *et al.*, 2007). In contrast, the bulk of the genome is methylated in somatic cells (Weber and Schübeler, 2007). The methylated cytosines concentrate in active gene bodies (Hackett and Surani, 2013). Indeed, intragenic 5mC has been observed in highly transcribed genes in murine embryonic stem cells, where it is suggested to prevent spurious transcription initiation by cryptic entry of RNA polymerase II (Neri *et al.*, 2017). DNA methylation is also a prominent feature at repetitive DNA, among which satellite DNA repeats are more specifically of interest in the frame of my thesis work (Francastel and Magdinier, 2019; Scelfo and Fachinetti, 2019). At DNA repeats, DNA methylation is linked to maintenance of genome integrity, by inhibiting their transposition and mitotic recombination, possibly by maintaining them in a silent state (Jones, 2012).

The why: the roles of DNA methylation in regulating transcription

DNA methylation has been traditionally associated with transcriptional silencing. Yet, DNA methylation is far from being a simple "ON/OFF" switch. Its function and influence on transcriptional activity depend on the genomic location and CpG density. At promoters, DNA methylation is negatively correlated with transcription only when CpG density ranges from intermediate to high. For instance, in mammals, at early embryonic stages, germline CGI promoters become methylated, leading to permanent somatic silencing of germline genes (Borgel *et al.*, 2010). Moreover, some CGI promoters, like those of lineage-specific genes and germline genes, become methylated and silenced during cellular differentiation (Mohn *et al.*, 2008; Payer and Lee, 2008). On the other hand, methylation at low CpG density promoters does not preclude transcription (Weber *et al.*, 2007). In addition, 5mC can coexist with active histone marks at enhancers, such as H3K27 (Charlet *et al.*, 2016). These last examples highlight that the relationship between 5mC levels and transcriptional activity is far from being completely deciphered. Furthermore, how some regions can be protected from DNA methylation or how DNA methylation is targeted to specific sites is still unclear.

DNA methylation has also been implicated in silencing of retrotransposable elements and satellite repeats (Jones, 2012; Weber *et al.*, 2007; Yoder *et al.*, 1997). Transposable elements are heavily interspersed in eukaryotic genomes and their silencing is important to maintain

genomic integrity. In germ and mouse ES cells, methylation levels at transposable elements are low compared to somatic cells, and correlate with their transcriptional activity (Efroni *et al.*, 2008; Fadloun *et al.*, 2013). Evidence that DNA methylation plays a critical role in silencing of transposable elements came from the study showing that hypomethylation of the retroelements LINE1 in DNMT3L knock-out germ cells correlated with aberrant transcription of these elements, leading to meiotic catastrophes (Bourc'his and Bestor, 2004). More on transposable elements and their implication in genome (in)stability is detailed Chapter 2.

Tandem satellite repeats are constitutively methylated in somatic cells. In the mouse, Minor (centromeric) and Major (pericentromeric) satellite repeats contain two and eight CpG dinucleotides per repeat unit, respectively. Despite the prevalence of methylation at centromeric regions, the link between DNA methylation and the transcription of these regions is still unclear, and it will be further analyzed in Chapter 2.

DNA methylation and disease

In light of the fundamental role of DNA methylation in influencing transcription of the genome and its stability, it is not surprising that aberrant DNA methylation profiles are pathological signatures reported in complex diseases and genetic disorders. Indeed, epigenetic defects are also hallmarks of cancer. Aberrant DNA methylation profiles in cancer include global genomic hypomethylation, and locus-specific hypermethylation of CpG islands at promoters of tumor suppressor genes (Francastel and Magdinier, 2019) (**Figure 7**). More specifically, cancer is accompanied by DNA hypermethylation of CpG islands near the promoter regions of tumor suppressors, genes that regulate cellular motility, apoptosis, angiogenesis and genes of the DNA repair system (Kisseljova *et al.*, 1998). In contrast, the targets for aberrant demethylation in tumors are unique DNA sequences encoding for imprinted genes, proto-oncogenes, cancer-testis genes and genes involved in invasion and metastases (Esteller, 2008; Kisseljova *et al.*, 1998). Moreover, this hypomethylation also affects DNA repeats that are normally methylated in physiological contexts, and often correlates with increased aberrant transcription of satellite sequences (Ting *et al.*, 2011). DNA hypomethylation in cancer cells increases with tumor progression and causes chromatin decondensation and chromosomal rearrangements that may result in chromosomal instability (Jaco *et al.*, 2008). However, no clear causality relationship between DNA methylation and deregulated satellite RNA accumulation has been clearly established. This will be discussed in more detail in Chapter 2 describing the mechanisms of transcriptional regulation of satellite repeats.

Perturbed DNA methylation is not only associated with cancer, but also with a broad range of genetic diseases (Francastel and Magdinier, 2019), like imprinting gene disorders such as Prader-Willi and Angelman syndromes (Madaan and Mendez, 2022; Mendiola and LaSalle, 2021) or repeat-instability diseases such as fragile X-syndrome and facioscapulohumeral muscular dystrophy (FSHD) (Gaillard *et al.*, 2014) and the Immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome (Ehrlich, 2003) on which I will focus more in the following section.

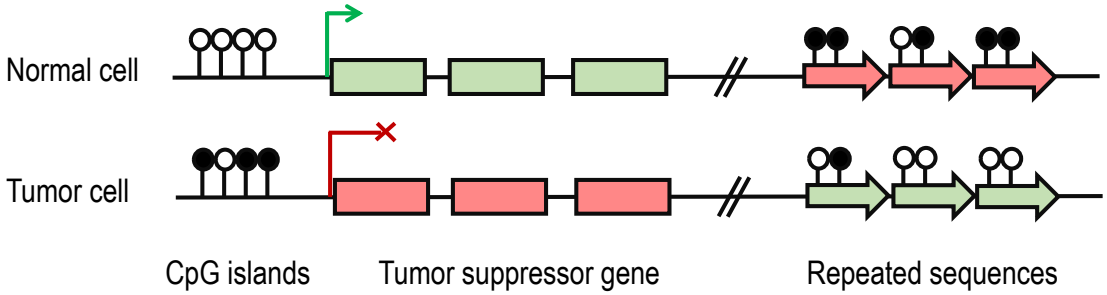


Figure 7. DNA methylation patterns in normal and cancer cells. In normal cells, repeated sequences are methylated, while CpGs present at the promoters of tumor suppressor genes (TSG) are normally hypomethylated and the genes are expressed. In cancer cells, a decrease in the level of methylation of repeated sequences lead to genomic instability as a consequence of an increase in mitotic recombination events. On the contrary, *de novo* methylation in tumor cells can affect CpG islands of tumor suppressor genes leading to their inactivation. Modified from Lopez-Serra and Esteller, 2011.

The ICF syndrome: a model to link DNA hypomethylation to aberrant transcription and chromosomal instability

Hypomethylation of satellite repeats around and within centromeres is a hallmark of pathological cases with chromosomal instabilities including cancer but also of a rare human recessive disease, the ICF syndrome (Immunodeficiency, Centromeric instability, Facial anomalies syndrome) (Ehrlich, 2003; Wijmenga *et al.*, 2000). ICF is an incredibly rare disease, and a little more than 70 patients have been reported worldwide (Kiaee *et al.*, 2021). The ICF is a clinically heterogeneous disease, although the common feature is severe impairment to humoral immunity (Blanco-Betancourt *et al.*, 2004). This leads to premature death of the patients in early childhood from repeated infections. Despite a few reported cases where patients developed cancer, it is not clear whether hypomethylation of (peri)centromeric repeats would favor later complications and further emergence of cancer. However, the study of this rare recessive genetic disease has already proven to be instrumental in the

identification of new players and mechanisms involved in DNA methylation and maintenance of genome stability (Velasco and Francastel, 2019). DNMT3B was the first gene to be incriminated, however only in about 60% of patients, causing the ICF syndrome type 1 (ICF1, OMIM 242860) (Xu *et al.*, 1999). Later, mutations in a gene encoding Zinc finger and BTB domain factor with unknown functions, ZBTB24, were shown to cause ICF2 (OMIM 614069) (de Greef *et al.*, 2011; Nitta *et al.*, 2013). More recently, two additional "ICF factors", CDCA7 and HELLS, were identified as the cause of ICF3 (OMIM 609937) and ICF4 (OMIM 603946), respectively (Thijssen *et al.*, 2015). Work in my host team showed that these factors, with little known functions and devoid of methyltransferase activity, were required for maintenance of DNA methylation at (peri)centromeric repeats in the mouse (Nitta *et al.*, 2013; Thijssen *et al.*, 2015). Data in the team has confirmed that ZBTB24 functions as a transcriptional activator in the case of the promoter of CDCA7 (Velasco *et al.*, 2018; Wu *et al.*, 2016). In turn, CDCA7 and HELLS form a bipartite nucleosome remodelling complex (Jenness *et al.*, 2018), where CDCA7 binds to nucleosomes and is required for HELLS localization onto chromatin and to activate its remodelling activity. Lastly, ZBTB24, CDCA7 and HELLS could function as a recruiting platform for DNMT3B at certain genomic loci (Velasco *et al.*, 2018).

Loss of ICF factor function in ICF syndrome results in significant changes in DNA methylation patterns, not only at certain gene promoters but also at the level of non-genic regions, and in particular at the level of constitutive heterochromatin. In fact one of the disease's most defining features is loss of DNA methylation in centromeric and pericentromeric region and marked loss of chromosomal condensation during mitosis (Xu *et al.*, 1999). In ICF patients, pericentromere decondensation in mitogen stimulated B lymphocytes leads to abundant chromosome breaks and rearrangements and often multibranched chromosomes containing three or more arms of chromosome 1 and 16 joined in the vicinity of the centromere (Carpenter *et al.*, 1988; Fryns *et al.*, 1981; Maraschio *et al.*, 1988; Tiepolo *et al.*, 1979) (**Figure 8 top**).

In ICF patients, centromeric and pericentromeric repeats are strongly hypomethylated (Jiang *et al.*, 2005). Mutations in DNMT3B (ICF1) result in hypomethylation at pericentromeric satellite 2 and satellite 3 repeats (Xu *et al.*, 1999), while mutations in ZBTB24, CDCA7 and HELLS (ICF2, 3, 4 respectively) share additional hypomethylation at centromeric alpha satellites (de Greef *et al.*, 2011; Thijssen *et al.*, 2015). Mouse models of the ICF syndrome have highlighted somewhat different methylation profiles compared to patients (**Figure 8**

bottom). This is notably the case of the ICF1 (Dnmt3b mutant) mouse embryos, where the loss of Dnam3b enzymatic activity is associated with a specific demethylation of centromeric Minor satellite repeats (Velasco *et al.*, 2010). On the other hand, ICF2, 3 and 4 mouse models share the hypomethylated signature at both centromeric (Minor) and pericentromeric (Major) satellite repeats (Dennis *et al.*, 2001; Velasco *et al.*, 2018).

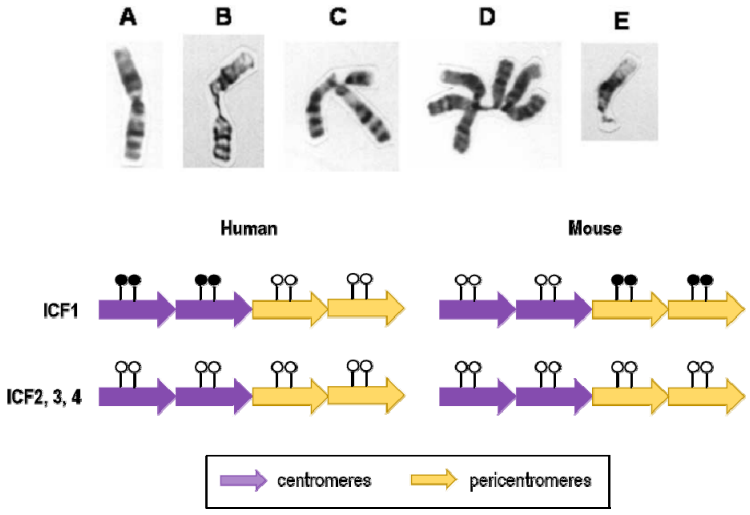


Figure 8. Chromosomal anomalies in the ICF syndrome. Top: examples of chromosomal anomalies in metaphases from ICF LCLs (**A**) normal Chr1, (**B**) Chr1 with decondensation in the qh region, (**C**) tri-radial (1;16)(q,q,p), (**D**) hexa-radial (1) (p,p,q,q,q,q), (E) deletion(1)(qh). Adapted from Ehrlich, 2003. **Bottom**: schematic of molecular defects affecting satellite repeats in ICF syndrome. Arrows represent centromeric (purple) and pericentromeric (yellow) tandem repeats. Black circles represent methylated CpGs, while white circles represent unmethylated ones. Black and white circles represent the methylation status, not the number of CpGs *per* repeat unit. In human, ICF1 displays hypomethylation on pericentromeric but not centromeric repeats, while it is the opposite for the ICF1 model. ICF2, 3 and 4 patients and mouse models harbor hypomethylation on both types of repeats.

Besides satellite sequences, ICF1/2/3/4 patients share hypomethylation at some regions of the inactive X chromosome (Hansen *et al.*, 2000; Kondo *et al.*, 2000), while only ICF1 patients harbor subtelomeric hypomethylation, accompanied by high levels or TERRA expression (Toubiana *et al.*, 2018).

In sum, the ICF factors represent new players in DNA methylation of the genome, especially at satellite repeats, and thus play a crucial a role in chromosomal stability. The discovery of these new "guardians of the genome integrity" raises the question of what are the molecular and cellular consequences of their loss of function on the hypomethylation-driven aberrant accumulation of satellite RNAs, which will be described further in Chapter 2.

Non coding RNAs

In recent years, ncRNAs have also emerged as key mediators of epigenetic mechanisms, as RNA pathways have been shown to be involved in targeting specific sequences for DNA modifications and/or heterochromatin formation (Dumesic *et al.* 2013). However, for a long time, RNAs have been regarded as just intermediates between the DNA sequence and proteins. This idea underlies the central dogma of biology, as theorized by Francis Crick, which explains how genetic information is expressed in our cells (Crick, 1958). This idea was contrasted by the discovery of untranslated RNAs known as non-coding RNAs (ncRNAs), such as ribosomal RNAs (rRNAs) (Aronson and McCarthy, 1961; Sloan *et al.*, 2017) and transfer RNAs (tRNAs) (Hoagland *et al.*, 1958), both of which are involved in protein synthesis. In more recent years, genome sequencing and high-throughput transcriptomic studies have shown that the mammalian transcriptome is constituted in large part of ncRNAs (Uszczynska-Ratajczak *et al.*, 2018). Indeed, approximately 90% of cellular RNA consists of rRNA (if evaluated by mass) and by tRNA (if evaluated by molar number) (**Figure 9**).

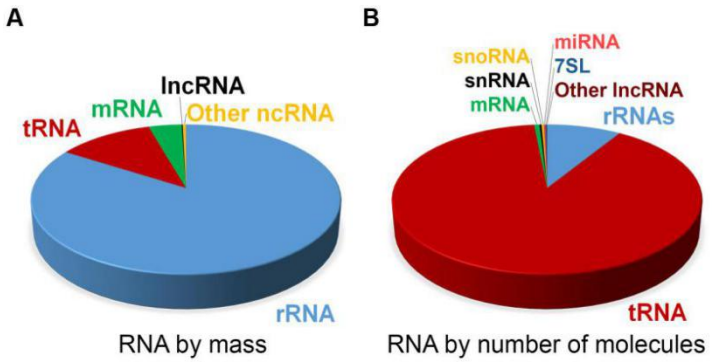


Figure 9. Estimate of RNA levels in a typical mammalian cell.
A. Proportion of the different classes of RNA in mammalian somatic cells by total mass and **B**. by number of molecules.
From Palazzo and Lee, 2015.

ncRNAs are gaining prominence as key players for proper cellular function. They largely contribute to nuclear organization and function (Clark and Mattick, 2011), regulation of gene transcription (Rinn, 2014) and they physically associate with protein complexes and direct them to genomic loci crucial for proper cellular functions (Kawaji and Hayashizaki, 2008; Mercer *et al.*, 2009). Importantly, ncRNAs are structural components of membraneless nuclear bodies (Clemson *et al.*, 2009; Sasaki *et al.*, 2009). Indeed, different studies have shown that RNAs can act as 'seeds' to recruit soluble RNA-binding nuclear body components from the nucleoplasm, and induce the formation of distinct types of nuclear bodies, such as paraspeckles, nuclear stress bodies or the Barr body (see bellow) (Mao *et al.*, 2011; Shevtsov and Dundr, 2011). ncRNAs play also roles for chromatin architecture (Mondal *et al.*, 2010;

Rinn, 2014), which can involve the interaction of the RNA with the DNA, *via* a DNA–RNA hybrid duplex (DNA-RNA hybrids are detailed in Chapter 5) (Aguilera and García-Muse, 2012).

ncRNAs are generally classified depending on their size in small or long ncRNAs (sncRNAs and lncRNAs, respectively). Satellite RNAs, which are of particular interest for my PhD work, are so diverse in length that they fit both definitions. Chapter 2 is dedicated to the analysis of satellite RNA characteristics and functions, in health and disease.

Small non coding RNAs

The category of sncRNAs is composed of RNAs with size between 19 and 30nt. The main role of these small ncRNAs is to keep the transcriptome under extensive surveillance by repressing gene expression (Michaux *et al.*, 2014). Various small RNAs of distinctive characteristics have been found and can be classified into three classes: microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-associated RNAs (piRNAs) (Kim *et al.*, 2009). These different classes of sncRNAs target different types of RNAs, associate with distinct subsets of effector proteins and are different in terms of biogenesis (**Figure 10**).

Briefly, miRNAs and siRNAs derive from double-stranded precursors (Carthew and Sontheimer, 2009) while piwiRNA precursors seem to be single-stranded (Malone and Hannon, 2009). These precursors are subsequently cleaved by enzymes of the RNase III family (Dicer for miRNAs and siRNAs). miRNAs and siRNAs are incorporated then into two ribonucleoprotein complexes: the cytoplasmic RNA-induced Silencing Complex (RISC) and the nuclear RNA-induced transcriptional silencing complex (RITS). On the other hand, piRNAs are incorporated in the Piwi complex. The main role of these small RNAs is the post-transcriptional regulation of genes and more particularly their repression. In particular, each complex can direct site-specific cleavage of target RNAs mediated by the ‘‘slicer’’ activity of an Argonaute protein (Kim *et al.*, 2009). Notably, these different classes of small RNAs target different types of RNAs. miRNAs are mainly directed against mRNAs. In fact, more than a third of human genes are predicted to be a direct target of miRNAs (Agarwal *et al.*, 2015; Lu and Rothenberg, 2018). While siRNAs also associate target certain mRNAs, they are mainly required for protection against viral infections. Finally, piRNAs are mainly expressed in germ cells, and they repress the expression and transposition of mobile elements (Kawaji and Hayashizaki, 2008; Kim *et al.*, 2009).

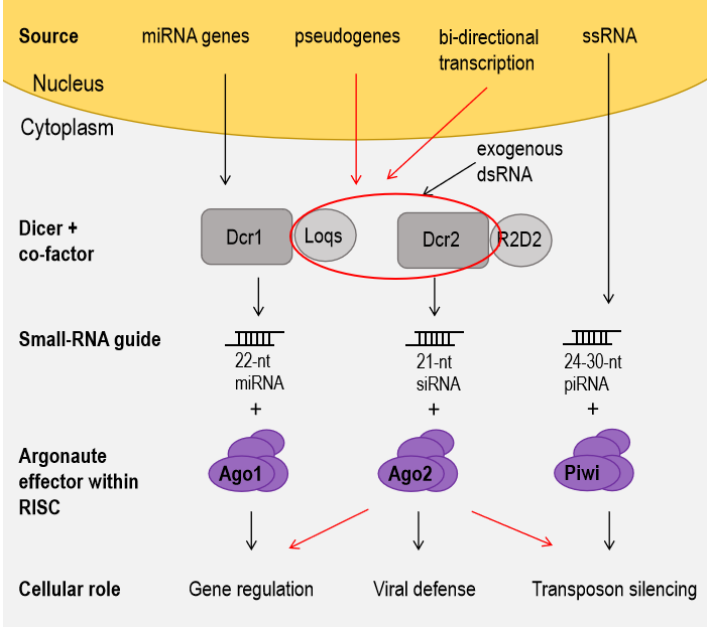


Figure 10.
Biogenesis and function of small non-coding RNAs. The three main classes of sncRNAs: 1. miRNAs, 2. siRNAs, and 3. piRNAs, differ in their biogenesis and target type.
Modified from Sasidharan and Gerstein, 2008.

Long non coding RNAs

lncRNAs are arbitrarily considered to be longer than 200nt, on the basis of a convenient practical cut-off in RNA purification protocols (Kapranov *et al.*, 2012). Most lncRNAs have expression profiles that are restricted to certain developmental stages and/or specific to certain cell types (Dinger *et al.*, 2008; Mercer *et al.*, 2009). They are involved in a wide range of cellular processes, such as dosage compensation of the mammalian X chromosome, parental imprinting, regulation of transcription and nuclear architecture (Mattick and Makunin, 2006). lncRNAs also associate with protein complexes and regulate their localization and/or activity. For example, in the context of transcriptional regulation, they can recruit chromatin activator or repressor complexes to specific genomic loci. This is the case of the lncRNA XIST, which is expressed from one of the two female X chromosome. XIST coats the X chromosome *in cis*, and recruits Polycomb group complexes (Engreitz *et al.*, 2013; Vaskova *et al.*, 2014). This leads to silencing across the majority of the X chromosome (Peeters *et al.*, 2014), which can be identified as the so-called Barr body, located in the periphery of the nucleus or nucleolus (**Figure 11.a**). Another function of lncRNAs is to act as scaffolds for nuclear compartments. This is the case, for example, of paraspeckles, formed by the assembly of the lncRNA NEAT1 and paraspeckle proteins (PSPs) (Naganuma *et al.*, 2012) (**Figure 11.b**). These nuclear domains are involved in splicing and editing of RNA (Wang and Chen, 2020). In addition, these compartments allow the sequestration of spliced, polyadenylated and capped RNAs (mature RNAs) encoding stress response proteins. This in turn allows a rapid response to stress, when the sequestrated RNAs are exported to the cytoplasm and translated (Clemson *et*

al., 2009; Naganuma *et al.*, 2012; Sasaki *et al.*, 2009). Another type of distinct nuclear structures, called nuclear stress bodies (nSBs), forms in response to heat shock in human cells (**Figure 11.c**). nSBs form at the pericentromeric regions where Sat III RNAs are transcribed and accumulate as lncRNAs, ranging from less than 2kb to more than 5kb in size (Jolly *et al.*, 2004; Rizzi *et al.*, 2004). nSBs seem to play a role in the fine control of splicing in response to stress by transiently sequestrating certain splicing factors, such as the SRp30c and SF2/ASF proteins, modulating gene expression programs required for the cells to recover from stress (Chiodi *et al.*, 2004; Denegri *et al.*, 2001; Metz *et al.*, 2004). More on satellite RNA roles in stress response is detailed in Chapter 2.

Several studies point to an important role of ncRNAs in the formation and maintenance of higher order chromatin regions that are repressive for transcription, such as heterochromatin domains (Maison *et al.*, 2002; Schoeftner and Blasco, 2008) (**Figure 11.d**). Indeed, the architectural protein HP1, involved in establishing and maintaining the repressive state of pericentromeric heterochromatin, is targeted to Major satellites by the Major satellite ncRNA (Maison *et al.*, 2002), upon the post-translational modification of HP1 with the small ubiquitin-like modifier (SUMO) (Maison *et al.*, 2011).

Besides being essential structural components of membraneless nuclear bodies, some lncRNAs exert their function by regulating the activity of proteins. This is the case of the lncRNA TERRA, which is transcribed from the subtelomeres (Luke and Lingner, 2009). TERRA interacts with the holoenzyme Telomerase and inhibits its activity outside of the S/G2 phase when it is no longer needed (Redon *et al.*, 2010; Schoeftner and Blasco, 2008).

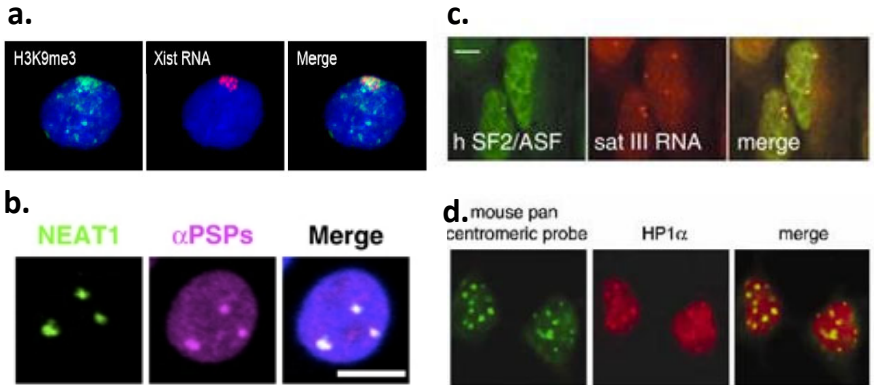


Figure 11. ncRNAs are integral parts of nuclear subcompartments. a. Barr body visualized by H3K9me3 and Xist RNA (from Vascova *et al.*, 2014); **b**. Paraspeckles visualized by NEAT1 RNA and pan-paraspeckle protein staining (from Naganuma *et al.*, 2012); **c**. nuclear stress bodies visualized by human SF2/ASF protein and Sat III RNA (from Metz *et al.*, 2004); **d**. pericentromeric heterochromatin visualized with pan-centromeric probe that hybridizes with Major satellites and HP1 protein staining (from Maison *et al.,* 2002).

The diverse functions of small and long ncRNAs firmly support the idea that ncRNAs are not just the result of passive read-through transcription of the genome. Evidence pointing in the same direction is provided by the knockdown of some ncRNAs in mice which is lethal or leads to developmental defects (Sauvageau *et al.*, 2013).

Unsurprisingly, the deregulation of ncRNAs seems to be one of the main features of many human diseases, including cancers, neuronal and muscle disorders (Esteller, 2011; Mattick and Makunin, 2006). Particularly in cancer, it has been shown that dysregulation of miRNAs and their processing machinery are a common hallmark (Calin *et al.*, 2002; Davalos *et al.*, 2012). A well-known example of the involvement of lncRNAs in cancer is that of the lncRNA ANRIL. ANRIL is found on the antisense strand of the INK4 locus, which encodes three tumor suppressor proteins. Hence, the overexpression of ANRIL leads to the repression of the INK4 locus and promotes cell proliferation (Aguilo *et al.*, 2011). ncRNAs are emerging as crucial players in both health and disease. Hence, better understanding of their transcriptional regulation, expression patterns and functional roles will lead the way to new approaches for disease diagnosis and therapy.

2. Diversifying the genome: contribution of repetitive DNA

To keep their genome intact, organisms have to manage diverse challenges, such as spontaneous mutations and DNA damage (see Chapter 3), while tolerating low level of mutation which represents the basis of evolutionary change. Repetitive DNA sequences have extensively shaped the mammalian genome (Frost *et al.*, 2005; Kuo *et al.*, 2021). In fact, repetitive DNA sequences are considered as fast-evolving, while genes that are essential to life, such as housekeeping genes, remain conserved across species (Eisenberg and Levanon, 2013). Moreover, differences in the amount of repetitive DNA sequences differentiate the genomes across species in both organization and size (de Koning *et al.*, 2011). The discrepancy in genome size across species stems mainly from repetitive sequences, which are hypothesized to have participated in genome size increase during speciation and evolution (Frazer *et al.*, 2003; Locke *et al.*, 2003). These repetitive DNA sequences include transposable elements (TEs), which constitute most of the interspersed DNA repeats, and tandem repetitive DNA (Sutherland and Richards, 1995). Bellow I will briefly describe the interspersed DNA repeats such as transposable elements, while the tandem repeats, like those

found at centromeres, are described in more detail as part of a book chapter found below (Mihic *et al.*, 2021).

Interspersed DNA repeats: the case of the transposable elements

Transposable elements (TEs) constitute approximately 45% and 37% of the human and mouse genomes, respectively (Deininger and Batzer, 2002), hence they are the most abundant family of repeated DNA. These genetic elements are capable of "moving" or transposing in the germ line, or in pathological conditions. They do so by either using a copy-and-paste mechanism that involves an RNA intermediate in the case of retrotransposons (type 1 TEs) or by a cut-and-paste mechanism, which is employed by DNA transposons (type 2 TEs) (**Figure 12**).

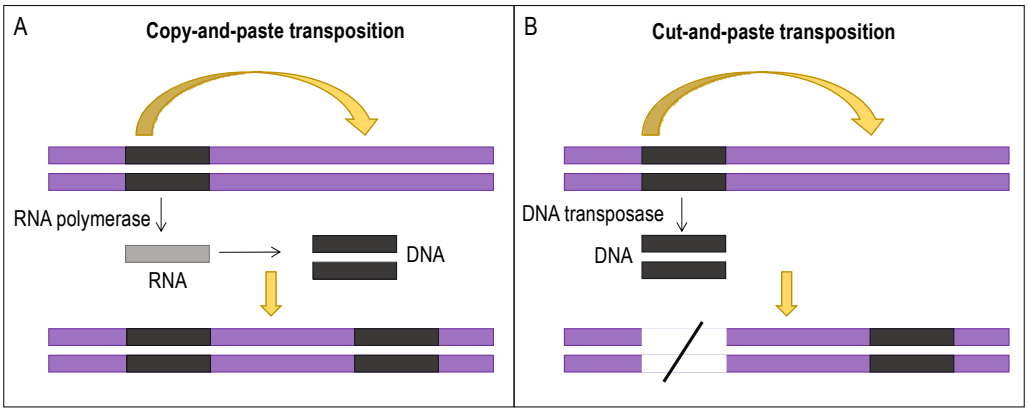


Figure 12. Mechanisms of transposition of TEs. A. the copy-and-paste mechanism of retrotransposons mobilization; **B**. the cut-and-paste mechanism of DNA transposons mobilization. Modified from Bonchev and Parisod 2013.

Type 1 transposons: Retrotransposons

Retrotransposons replicate via an RNA intermediate, which is then reverse transcribed back into DNA and reintegrated into the genome. They are classified according to their structure and putative viral origin in: i) Long Terminal Repeat (LTR) retrotransposons, that are retroviral-like in structure and ii) non-LTR retrotransposons, which include Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs). During the process of retrotransposition, LINE and LTR elements are transcribed by RNA polymerase II while SINEs are transcribed by RNA polymerase III. Moreover, LINE1 is the only retoelement capable of autonomous retrotransposition, and it's the most widespread family of retrotransposons representing approximately 20% of mammalian genomes (Martens *et al.*,

2005). Retroelements are particularly enriched at centromeric, pericentromeric and telomeric domains (Bartolomé *et al.*, 2002; Wright *et al.*, 2003). Interestingly, a family of truncated LINE1 (tL1) is highly conserved in chromosomal regions that separate telomeres and centromeres on mouse chromosomes (Kalitsis *et al.*, 2006a). The functions of the tL1 sequences remain poorly studied, and whether they simply serve to separate telomeric and centromeric domains, or they play a role in the transcriptional regulation of these domains is not completely understood. A suggested role for retroelements at centromeres is maintaining centromere size (Presting, 2018).

Type 2 transposons: DNA transposons

DNA transposons are mobile DNA that move using a single or double-stranded DNA intermediate (Cowan *et al.*, 2005). Some TEs have been "domesticated" by the host to perform a specific function in the cell (Sinzelle *et al.*, 2009). This is the case of RAG proteins, which have been derived from DNA transposases, and participate in V(D)J recombination during antibody class switch (van Gent *et al.*, 1996; Hiom *et al.*, 1998). Another example of the propensity of transposases to give rise to new host proteins is the centromeric protein CENP-B, which seems to have originated from the pogo-like transposon (Casola *et al.*, 2008; Kipling and Warburton, 1997).

Among many organisms, TEs are mostly active in the germline. Yet, there is evidence that TEs can be active also in somatic cells (Kazazian, 2011). Indeed, in humans, L1 expression and transposition has been detected also in early embryos and stem cells (Garcia-Perez *et al.*, 2007; Klawitter *et al.*, 2016). To persist in evolution, TEs need to strike a balancing act between expression and repression, in order not to lead to disadvantages for the host. TEs have contributed to shape mammalian genomes throughout evolution *via* various events such as formation of new genes or by creating new transcriptional regulatory elements, like promoters, thus acting as "functional elements" that can modulate transcriptional states (Faulkner and Carninci, 2009; Schmitz and Brosius, 2011). Although mutations and truncations have made most mammalian TEs incompetent of transposition, this does not preclude transcription initiation from promoters present within immobile elements. Indeed, TEs are hypothesized to act as promoters for centromeric transcription (Topp *et al.*, 2004) given their presence at or around centromeres in different species (Chang *et al.*, 2019; Kalitsis *et al.*, 2006; Miga *et al.*, 2014; Sun *et al.*, 2003). Yet, TEs represent a substantial burden to genome stability. In pathological conditions such as cancer, TE activity often arises as a

consequence of an already disregulated genome in disease situations. However, TE activity in somatic cells can drive cancer formation (Kaer and Speek, 2013). For instance, *de novo* L1 insertions can disrupt critical tumor suppressor genes (Scott *et al.*, 2016). Thus, transcription at these elements must be tightly controlled in order to avoid potentially aberrant events of transposition. Host organisms have developed a slew of different mechanisms to defend themselves against high rates of TE activity, which include DNA methylation (Bourc'his and Bestor, 2004; Greenberg and Bourc'his, 2019), RNA interference mediated mechanisms (Obbard *et al.*, 2009) or the action of specific proteins like the apolipoprotein B mRNA editing enzyme catalytic polypeptide 3 (APOBEC3) protein family that inhibit TE mobilization (Schumann, 2007).

Chapter 7
Centromeres Transcription and Transcripts for Better and for Worse

Pia Mihìc, Sabrine Hédouin, and Claire Francastel

Abstract Centromeres are chromosomal regions that are essential for the faithful transmission of genetic material through each cell division. They represent the chromosomal platform on which assembles a protein complex, the kinetochore, which mediates attachment to the mitotic spindle. In most organisms, centromeres assemble on large arrays of tandem satellite repeats, although their DNA sequences and organization are highly divergent among species. It has become evident that centromeres are not defined by underlying DNA sequences, but are instead epigenetically defined by the deposition of the centromere-specific histone H3 variant, CENP-A. In addition, and although long regarded as silent chromosomal loci, centromeres are in fact transcriptionally competent in most species, yet at low levels in normal somatic cells, but where the resulting transcripts participate in centromere architecture, identity, and function. In this chapter, we discuss the various roles proposed for centromere transcription and their transcripts, and the potential molecular mechanisms involved. We also discuss pathological cases in which unscheduled transcription of centromeric repeats or aberrant accumulation of their transcripts are pathological signatures of chromosomal instability diseases. In sum, tight regulation of centromeric satellite repeats transcription is critical for healthy development and tissue homeostasis, and thus prevents the emergence of disease states.

Keywords Centromere · Transcription · Satellite DNA · Cancer · ICF syndrome

Pia Mihìc and Sabrine Hédouin contributed equally with all other contributors.

P. Mihìc · C. Francastel (✉)
Université De Paris, Epigenetics and Cell Fate, CNRS UMR7216, Paris, France
e-mail: claire.francastel@univ-paris-diderot.fr; http://parisepigenetics.com/dmrhd/

S. Hédouin
Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Đ. Ugarković (ed.), *Satellite DNAs in Physiology and Evolution*, Progress in Molecular and Subcellular Biology 60,
https://doi.org/10.1007/978-3-030-74889-0_7

7.1 Satellite DNA Underlies (Peri)centromeric Chromosomal Regions

Centromeres are chromosomal domains specialized in the faithful segregation of the genetic material between daughter cells at each cell division. In most higher eukaryotes, centromeres are made up of satellite DNA, i.e., large arrays of short, mostly A/T-rich, DNA sequences repeated in tandem over chromosomal domains that can range from hundreds of kilobases (Kb) to tens of megabases (Mb) on each centromere. Different families of satellite DNA define two distinct functional chromosomal domains (Choo 2001) with distinct chromatin landmarks (Karpen and Allshire 1997), i.e., the centromere per se and the juxta- or pericentromeric domains. The centromere is defined by the presence of unique nucleosomes, in which histone H3 is replaced by its variant called CENP-A in humans (Cse4 in budding yeast, Cnp1 in fission yeast, and CID/CenH3 in fruit flies) (Earnshaw and Rothfield 1985), interspersed with canonical nucleosomes. The CENP-A nucleosomal domain creates a platform for the assembly of a proteinaceous structure known as the kinetochore, that links chromosomes to mitotic spindle microtubules (Palmer et al. 1991; Fukagawa and Earnshaw 2014; Gambogi et al. 2020). CENP-A deposition is tightly regulated and mediated by a specific histone chaperone Holliday Junction Recognition Protein (HJURP) (Dunleavy et al. 2009; Foltz et al. 2009), although it remains unclear how HJURP directs CENP-A to its default location (Hoffmann and Fachinetti 2017). In juxta- or pericentromeric position, satellite repeats are enriched in repressive epigenetic marks which makes up the bulk of constitutive heterochromatin in mammals (Schueler and Sullivan 2006; Eymery et al. 2009b; Saksouk et al. 2015), and to which several functions have been assigned, including in sister chromatid cohesion at centromeres (Pidoux and Allshire 2005), maintenance of genome stability (Peters et al. 2001), and functional organization of the interphase nucleus (Wijchers et al. 2015; Muller et al. 2019; Francastel et al. 2000).

In contrast to the presence of CENP-A nucleosomes being the main and evolutionary conserved determinant of centromere identity, with very few exceptions in some insect lineages and kinetoplastids (Akiyoshi and Gull 2014; Drinnenberg et al. 2014; Navarro-Mendoza et al. 2019), the evolution of the underlying DNA sequences has been quite dynamic and gave rise to highly divergent centromere organization (Malik and Henikoff 2009). Phylogenetic analysis also found little evidence for satellite sequence conservation, which contrasts with the ancestral conserved structure of tandem repeats at telomeres (Meyne et al. 1989). Centromeres in different species display a wide variety of sequences, repeats organization, and chromosomal positions (Melters et al. 2013; Plohl et al. 2014), that can even diverge between chromosomes of the same species like in human and *Drosophila* (Bracewell et al. 2019; Sullivan and Sullivan 2020). Tandem repeats are highly prevalent at (peri)centromeres of most animal and plant genomes, and monocentric centromeres are the fundamental unit for chromosome inheritance in most species. However, chromosomes in certain insects lack a primary constriction and have adopted holocentric centromeres, i.e., in which the activity of the kinetochore extends over

the whole chromosome arms, and therefore lacks a satellite repeats signature (Drinnenberg et al. 2014). There are also examples of atypical, yet functional, centromeres that spontaneously form on unique non-satellite sequences. Originally described in humans, the so-called neocentromeres are functionally and structurally similar to endogenous centromeres but lack the underlying repetitive sequences (Scott and Sullivan 2014). In the domestic horse, the discovery of satellite-less centromeres with variable positions among individuals (Giulotto et al. 2017) reinforced the idea that centromeres are defined, at least in part, by a centromere-specific histone variant, and not by the underlying DNA sequences.

Besides the abundance of repeats and association of centromeres with pericentromeric heterochromatin domains in most eukaryotes, a 17 base-pair (bp) motif is also highly conserved throughout species and called the CENP-B-box (B-Box). This sequence is the consensus binding site for the CENP-B protein, the only centromeric protein with sequence-specific DNA-binding activity (Masumoto et al. 1989). CENP-B protein is itself highly conserved, although its essential nature is debatable since certain centromeres lack a B-box in their satellite repeats, like on the human Y chromosome (Jain et al. 2018), and because CENP-B seems dispensable in the mouse (Kapoor et al. 1998). The development of human artificial chromosomes (HAC) has been instrumental in the determination of the minimal requirements for a functional centromere in terms of protein factors and DNA sequences (Bergmann et al. 2012). Many reports highlighted that both satellite DNA and CENP-B were necessary for de novo assembly of centromeres and HAC formation (Ohzeki et al. 2002; Masumoto et al. 2004), although alternative ways to establish a centromere during HAC formation have been reported (Logsdon et al. 2019), still questioning CENP-B requirement for establishment or maintenance of centromeres.

7.1.1 Examples of Centromere Organization Across Species

The budding yeast *Saccharomyces cerevisiae* and some relatives represent some sort of exceptions in the centromere world. In these organisms, a so-called "point centromere" is defined by short and unique sequences on which the centromere-specific nucleosome is positioned (Lechner and Ortiz 1996; Furuyama and Biggins 2007). This is in contrast with other eukaryotic organisms that feature regional centromeres assembled on kilo- to megabase-scale arrays of tandem repeats at the primary constriction site of the chromosome, and at which CENP-A nucleosomes are interspersed with canonical ones (Blower et al. 2002). The yeast *Schizosaccharomyces pombe* centromeres are composed of a 4–7 Kb-long central core element (ctr) flanked by centromere-specific innermost repeats (imr) sequences and pericentric outer repeats (otr), with an overall size range of 30–110 Kbs depending on the chromosome (Polizzi and Clarke 1991). In maize, centromeres are composed of CentC repeat of 156 bp, which form tandem arrays that span 180 Kb, separated by one or more copies of centromeric retrotransposable

(CR) elements. Similar organization at centromeres is shared by the fruit fly *Drosophila melanogaster*, where the centromere is primarily composed of AATAT and TTCTC satellites, interspersed with complex A/T-rich repeats and mobile elements (Sun et al. 2003; Chang et al. 2019).

A model of choice for the study of centromere organization is the laboratory mouse *Mus musculus* due to its fairly homogeneous centromeres across all chromosomes (Kalitsis et al. 2006). The basic repeat unit in murine centromeres, called a minor satellite, is 120 bp-long and repeated in tandem over about 600 Kbs, which represents around 0.45% of the mouse genome. Murine chromosomes are telocentric, meaning that the centromere is nearly adjacent to the telomere of the short chromosome arm. On this short arm, minor satellite repeats are flanked by a retrotransposable DNA element, the truncated Long Interspersed Nucleotide Element 1 (tL1) and clusters of telocentric tandem (TLC) repeats, which share between 74% and 77% of homology with minor satellites, but lay in the opposite orientation (Kalitsis et al. 2006). On the long chromosome arm, the flanking pericentromeric domains are made up of tandem repeats of 234 bp-long major satellite repeat units over around 6 Mb and representing up to 3% of the mouse genome (Choo 1997; Kalitsis et al. 2006).

As opposed to homogeneous murine satellite repeats, each human centromere shows distinct polymorphisms in the number and sequence of α-satellite repeats (Aldrup-Macdonald and Sullivan 2014). Centromeric regions contain 171 bp-long α-satellite repeat units arranged in a tandem head-to-tail fashion, into higher-order repeat (HORs) units, themselves repeated in a largely uninterrupted fashion up to 5 Mb. The re-iteration of the HOR forms the centromeric α-satellite array, with occasional interruptions by transposable elements (She et al. 2004). In a given HOR, individual α-satellite repeats may only share 50–70% sequence similarity, whereas different HORs from the same chromosome share up to 98% of homology (Miga 2019). Neighboring pericentromeric regions account for about 4% to 5% of the human genome. They can be made of three types of satellite repeats: type I, which are formed by an alternation of 17 and 25 bp monomers and are restricted to chromosomes 2, 3, and acrocentric chromosomes. Type II and type III satellites are made of a 5 bp-long GGAAT repeat unit, found on all chromosomes, although unevenly distributed over several Mb. Notably, large blocks of heterochromatin in juxtacentromeric position on the long arm of chromosomes 1, and 16, or the long arm of chromosome 9, are composed of satellites type II or III, respectively (Vourc'h and Biamonti 2011).

All in all, the few unifying features in centromere organization across eukaryote kingdoms pose somewhat of a paradox given the essential nature of centromeres for the maintenance of genome integrity and the conserved functions and dynamics of kinetochores. Perhaps the most common characteristic of DNA sequences underlying centromeres is that they are transcriptionally competent in most of the species studied.

7.2 Transcription of Centromeric Repeats and Their Transcripts

At a time when recognition of the relatively pervasive aspect of genome transcription was in its infancy, the findings that (peri)centromeric satellite repeats could be transcribed have almost gone unnoticed. However, transcription at centromeric repeats was hinted at by the existence of centromeric transcripts in murine cells (Harel et al. 1968; Cohen et al. 1973) and in lung cells of the newt *Taricha granulosa* (Rieder 1978).

Nowadays, both the transcription of centromeric repeats and its products, the centromeric RNAs (cenRNA), are viewed as a conserved feature of centromeres in a broad range of organisms (Table 7.1), including yeast (Volpe et al. 2002; Choi et al. 2011; Ohkuni and Kitagawa 2011), plants (Topp et al. 2004; Du et al. 2010), beetles (Pezer and Ugarković 2008), *Drosophila* (Grewal and Elgin 2007; Rošić et al. 2014), amphibians (Varley et al. 1980; Diaz et al. 1981; Blower 2016), mouse (Rudert et al. 1995; Bouzinba-Segard et al. 2006; Ferri et al. 2009), and humans (Chan et al. 2012; Quénet and Dalal 2014; McNulty et al. 2017). Interestingly, work using a structurally dicentric chromosome that contains two α-satellite arrays demonstrated that RNA Polymerase II (RNA Pol II) localizes at active centromeres, i.e., at which the kinetochore assembles, but not at the inactive one (Chan et al. 2012), although another study reported that inactive arrays can also produce cenRNAs but just less stable than those originating from active arrays (McNulty et al. 2017).

In essence, in the absence of conserved centromeric DNA sequences across species, it is tempting to speculate that transcription through centromeric repeats or their derived transcripts may be functionally relevant to centromeres identity or function.

7.2.1 Regulation of Centromere Transcription/Transcripts Levels

Most of our knowledge of centromeric repeats transcription has been inferred from the existence of transcripts with centromeric sequences. Yet, the consequence of their highly repetitive and near-identical nature in some species is that they are mostly absent from reference genomes and specifically excluded from high-throughput sequencing analysis. Yet, transcripts with sequences of the identified centromeric repeat units can be found in various databases, including Expressed Sequence Tags (EST) databases. In addition, dedicated experimental testing of the levels of cenRNAs in a given tissue or at a specific developmental stage argued against the idea of simple transcription noise and even provided evidence for some level of transcriptional regulation. The challenge is rather to determine whether the observed differences in transcript abundance are the result of transcriptional or posttranscriptional control mechanisms.

Table 7.1 Satellite transcripts observed in normal and pathological conditions from various model organisms

Organism	Origin	Condition	Size	Pol II	Characteristics	References
S. pombe	Centromeric **Cnt** and **imr**	Exosome mutants	~0.5 Kb	Yes	Not detected in WT cells, transcribed from both strands	Choi et al. (2011)
	Pericentromeric **dg** and **dh**	WT cells	21–25 nt	Yes	RNAi machinery-dependent siRNAs	Volpe et al. (2002)
		RNAi depletion	1.4–2.4 Kb	Yes	Accumulation of cenRNAs longer than siRNAs	
S. cerevisiae	Centromere	Exosome mutants	1.3 Kb	Yes	Not detected in WT cells	Ohkuni and Kitagawa (2012)
Plants	Centromeric **retroelement** Satellite **CentC** repeat	WT cells	40–900 bp	N.D.	Single-stranded, polyadenylated RNA	Du et al. (2010), Topp et al. (2004)
Beetles	Centromeric repeats	WT cells	0.5–5 Kb	Yes	Transcribed from both strands	Pezer and Ugarković (2008)
Drosophila	Centromeric repeats	WT cells	1.3 Kb	Yes	Transcribed from the X chromosome	Rošić et al. (2014)
Amphibians	Pericentromeric **Sat I**	WT cells	N.D.	Yes	Transcribed from both strands	Diaz et al. (1981)
Mouse	Centromeric **Minor satellite**	WT somatic cells	2–4 Kb	Yes	Transcribed from both strands Peak in G2/M	Bouzinba-Segard et al. (2006), Ferri et al. (2009), Hédouin et al. (2017)
		5AZA-treated or differentiated cells	120 nt	N.D.	Transcribed from both strands	
	Pericentromeric **Major satellite**	WT mESCs	25–30 nt	Yes	Transcribed from both strands	Kanellopoulou et al. (2005)
		Dicer-KO mESCs	>1 Kb	Yes	Accumulation of cenRNAs longer than siRNAs	Kanellopoulou et al. (2005)
		Murine embryo, 2-cell stage	N.D.	N.D.	Transcribed first from the Forward then from the Reverse strand	Probst et al. (2010)

Human	Centromeric **α-satellite**	WT cells	1.3 Kb	Yes	Transcribed from both strands Peak in G1	Quénet and Dalal (2014), Bury et al. (2020)
	Pericentromeric **Satellite II**	Pancreatic adenocarcinomas	N.D.	Yes	Transcribed from both strands	Ting et al. (2011)
	Pericentromeric **Satellite III**	Heat shock treated cells	2–5 Kb	Yes	Transcribed from the reverse strand	Rizzi et al. (2004)

Pol II transcribed by RNA polymerase II; *5AZA* 5-aza-2′-deoxycytidine; *KO* knock-out; *mESCs* mouse embryonic stem cells; *N.D.* not determined; *RNAi* RNA interference; *WT* wild-type

In most species, the levels of cenRNAs appear to vary with particular developmental stages and with cell types, tissues, or organs. They have been detected in coleopteran insect species at all three developmental stages: larvae, pupae, and adults (Pezer and Ugarković 2008). In chicken and zebrafish, transcripts from an α-like satellite repeat are detected during early embryogenesis but are limited to the cardiac neural crest, the head, and the heart (Li and Kirby 2003). During mouse early development, pericentromeric major satellite RNAs (pericenRNAs) start being detected at the 2-cells stage and are required for the major reorganization of the nucleus that occurs at this stage, most notably characterized by the assembly of pericentromeric heterochromatin nuclear compartments, concomitantly with zygotic gene activation (Probst et al. 2010). In somatic cells, murine cen- and pericenRNAs accumulate with terminal differentiation, a process also accompanied by major spatial reorganization of constitutive heterochromatin compartments and changes in gene expression programs (Terranova et al. 2005; Bouzinba-Segard et al. 2006).

Transcription of centromeric repeats seems to be also regulated during the cell cycle. In cycling murine cells, cenRNAs begin to accumulate at the end of S phase and peak in the G2/M phase, just before the onset of mitosis (Ferri et al. 2009). This accumulation coincides with the late S phase when murine centromeres are being replicated (Müller and Almouzni 2017), although no formal demonstration that this would facilitate active transcription has been established. In human cells, the levels of cenRNAs do not change throughout the cell cycle (McNulty et al. 2017), although a recent study showed that their levels could fluctuate and peak in G2/M (Bury et al. 2020). In contrast, active RNA Pol II has been detected at human centromeric repeats in G1 (Quénet and Dalal 2014), when cenRNAs levels are low (Bury et al. 2020). More strikingly, elongating RNA Pol II was detected at mitotic centromeres in humans and mice (Chan et al. 2012). This is paradoxical since mitosis is regarded as a phase during which the bulk of the genome is transcriptionally silent (Christova and Oelgeschläger 2002), and this RNA Pol II localization could represent storage or bookmarking for further transcriptional activation when cells reenter the cell cycle. However, incorporation of fluorescent uridine-5′-triphosphate nucleotides (UTP) at the mitotic centromere suggested that human and murine centromeres are indeed actively transcribed during mitosis (Chan et al. 2012).

7.2.2 Mechanisms of Transcriptional Regulation

7.2.2.1 Transcriptional Machinery at Centromeric Repeats

The characteristic organization of most satellite DNA sequences is based on tandem repeats devoid of canonical promoter sequences, which led to the proposal that they could be transcribed by read-through from upstream genes or promoters of transposable elements, which is the case in maize (Topp et al. 2004). It should be noted that not all centromeres, like the human Y chromosome, may contain transposon sequences (Miga et al. 2014). In addition, a candidate TATA-box has been identified

within human α-satellite sequences, as well as an SV40 enhancer-core sequence with spacing and orientation characteristic of RNA Pol II-transcribed genes (Vissel et al. 1992). The hypothesis that cryptic promoter elements are present within repeat sequences would not be surprising since, like any genomic sequence, centromeric repeats are stuffed with consensus binding sites for regulatory proteins. Some of these binding sites have long been known to serve as entry sites for the basal transcriptional machinery, like GATA sites, in place of a canonical TATA-box (Aird et al. 1994). Of note, consensus binding sites for GATA factors (WGATAR in which W indicates A/T and R indicates A/G) are frequently occurring in mammalian genomes, including at murine and human centromeric repeats, although occupancy by GATA-family members has not been described.

Whether cells have adapted to the fortuitous binding of various transcription factors to genomic loci essential for their survival is not known. Nevertheless, it may explain why centromere transcription appears to be regulated depending on cellular contexts, and hence, may rely on context-specific transcription factors. In addition, accumulation of cenRNAs of the size of a repeat unit in the mouse suggested that each repeat unit might contain a transcription start site (Bouzinba-Segard et al. 2006), although we cannot exclude that posttranscriptional processing of longer transcripts occurs, which is discussed below.

The question of the RNA polymerase(s) involved, and the means employed to regulate transcription of centromeric repeats, is also important. Centromeres of both budding and fission yeast are transcribed by RNA Pol II (Ohkuni and Kitagawa 2011; Sadeghi et al. 2014). In beetles, the presence of a cap structure and poly (A) tails in a subset of cenRNAs, termed PRAT, is also indicative of RNA Pol II-dependent transcription (Pezer and Ugarković 2008). Differential inhibition of RNA Pol I, II, and III and detection of active RNA Pol II at humans (Quénet and Dalal 2014; McNulty et al. 2017) and murine (Chan et al. 2012) centromeres indicated that RNA Pol II orchestrates the transcription of their centromeres.

In sum, RNA Pol II seems to be responsible for most part of the transcription of both point and regional centromeres, suggesting that transcription of this essential chromosomal domain has been conserved throughout evolution.

7.2.2.2 Transcription Factors

Compatible with the plethora of putative consensus binding sites for transcription factors in centromeric sequences, activators and repressors have been identified to control transcription of satellite repeats in various systems in a similar mode to the regulation of gene promoters.

In *S. cerevisiae*, the transcription factor Centromere-binding protein 1 (Cbf1) has been implicated as an activator of centromere transcription in an RNA Pol II-dependent manner (Ohkuni and Kitagawa 2011), although this is still debated. Indeed, other studies reported the upregulation of cenRNAs in cells lacking Cbf1, in association with chromosomal instability through the downregulation of the protein levels and mislocalization of CENP-A, HJURP, and components of the

Chromosome Passenger Complex (CPC) (Ling and Yuen 2019; Chen et al. 2019). Along the same line, Htz1 (human homolog of H2A.Z) was identified as a transcriptional repressor since its deletion resulted in an upregulation of cenRNAs levels (Ling and Yuen 2019). Interestingly, the double invalidation of Cbf1 and Htz1 resulted in an additive effect on the upregulation of cenRNAs, suggesting that these two proteins operate in distinct pathways to repress centromere transcription (Ling and Yuen 2019). Noteworthy, Cbf1 is conserved among species with point centromeres, but not in eukaryotic species that have regional centromeres.

The Daxx-like motif-containing GATA factor Ams2 was actually one of the first transcription factors shown to be required for centromere function in *S. pombe* (Chen et al. 2003). Ams2 is a cell cycle-regulated factor that occupies centromeric chromatin in S phase where it is required for SpCENP-A deposition, although if this occurs through promoting transcription of cenRNAs has not been established. Again, whether a role for GATA factors is conserved among species has not yet been tested.

The only reported transcriptional regulator for transcriptional activation at murine and human centromeric repeats is the Zinc Finger and AT-Hook Domain Containing (ZFAT) protein, through binding to the ZFAT box, a short sequence present at centromeres of all chromosomes in mouse and human (Ishikura et al. 2020). In mammals, more is known about the transcription of the neighboring pericentromeric satellite repeats. In the mouse, transcriptional repressors YY1 (Shestakova et al. 2004), C/EBPα (Liu et al. 2007), and Ikaros (Brown et al. 1997; Cobb et al. 2000) appear to bind directly to major satellite sequences, although their link with transcriptional repression of these repeats has not been tested. In contrast, heat-shock transcription factor 1 (HSF1) has been shown to promote transcriptional activation of Sat III repeats in response to cellular stress in human cells (Jolly et al. 2004; Rizzi et al. 2004).

7.2.2.3 Histone Marks and DNA Methylation

In addition to occupancy by transcription factors, epigenetic modifications are likely to participate in the control of the transcription of centromeric repeats. In contrast to nearby pericentromeric heterochromatin, and in addition to CENP-A-containing nucleosomes, centromeres exhibit marks of euchromatin such as dimethylation of lysine 4 of histone H3 (H3K4me2) and lack of heterochromatin marks such as di- and tri-methylation of lysine 9 at histone H3 (H3K9me2 or me3) (Sullivan and Karpen 2004). In addition, a hypoacetylated state at centromeres seems to be conserved across eukaryotes (Wako et al. 2003; Sullivan and Karpen 2004; Choy et al. 2011). More specifically, hypoacetylated lysine 16 of histone H4 (H4K16) was shown to be required for kinetochore function and accurate chromosome segregation, although the link with transcriptional competence of centromeric repeats was not assessed (Choy et al. 2011). More recently, acetylated lysine 4 of histone H3 (H3K4ac) was described at centromeres, at which it is required for the centromeric localization of the chromatin reader Bromodomain-containing protein 4 (BRD4),

which in turn recruits the RNA Pol II (Ishikura et al. 2020). At pericentromeric satellite repeats, SIRT6, a member of the Sirtuin family of deacetylases, has been implicated in the maintenance of their silent state through deacetylation of histone H3 at lysine 18 (H3K18ac) (Tasselli et al. 2016). Knockdown of SIRT6 caused an aberrant accumulation of pericenRNAs associated with mitotic errors through desilencing of pericentromeric repeats, which also lent support to the importance of heterochromatin maintenance in centromere function.

Additional important insights into the causal roles of histone modifications for transcription at centromeres came from the study of human artificial chromosomes (HACs) in which a CENP-B box was replaced by a tetracycline operator (tetO). Demethylation of H3K4me2, through targeting of the Lysine-Specific histone Demethylase 1A (LSD1) to the tetO sequence, induced a strong decrease in centromere transcription associated with impaired recruitment of HJURP, the CENP-A chaperone (Bergmann et al. 2011). This study provided a nice demonstration of a causal link between a specific activating histone mark, transcription of centromeric repeats, maintenance of CENP-A, and kinetochore functions.

Ubiquitination of histones is also important for the transcription of satellite repeats, with activating or repressive roles for ubiquitination of histone H2B or H2A (H2A-Ub; H2B-Ub), respectively (Zhu et al. 2011; Sadeghi et al. 2014). Loss of function of the Breast Cancer type 1 susceptibility protein (BRCA1) in cancer cells led to reduced H2A-Ub at pericentromeric satellite repeats, accompanied by their transcriptional derepression and loss of heterochromatin integrity (Zhu et al. 2011). Knockdown of Ring Finger Protein 20 (RNF20), the ubiquitin ligase responsible for H2B-Ub in *S. pombe,* led to reduced levels of transcription and nucleosome turnover at centromeres, associated with impaired centromere function. These data suggested that H2B-ub is essential for the maintenance of active centromeric chromatin (Sadeghi et al. 2014).

In contrast to their divergent sequence and structure, a common feature of mammalian satellite sequences is their methylated state at CpG dinucleotides, the main context for DNA methylation at least in mammals, which is also a major epigenetic mechanism to consider. The density of methylatable CpGs is higher at pericentromeres, which is consistent with their heterochromatin status, whereas centromeric repeat units in mice and humans contain only 2 to 3 CpGs *per* repeat unit. At repetitive elements, DNA methylation has been implicated in the inhibition of transposition and mitotic recombination between repeats, in part through maintaining these elements in a silent state (reviewed in Saksouk et al. 2014; Francastel and Magdinier 2019; Scelfo and Fachinetti 2019). However, our vision of the direct relationship between DNA methylation and transcriptional states of these regions may only be partial. For example, treatment of cells with the DNA demethylating agent 5-aza-2′-deoxycytidine (5AZA) led to increased levels of pericenRNAs in human cells (Eymery et al. 2009a) and of cenRNAs in murine cells (Bouzinba-Segard et al. 2006). However, it remains to be determined whether is it directly through demethylation of satellite repeats or through more indirect effects caused by 5AZA treatment, e.g., increased DNA damage also known to cause accumulation of Sat III pericenRNAs in humans (Valgardsdottir et al. 2008) or

cenRNAs in murine cells (Hédouin et al. 2017). The identification of the DNA methyltransferases (DNMTs) responsible for de novo methylation (DNMT3A and DNMT3B) (Okano et al. 1999) or maintenance of methylation (DNMT1) (Bestor et al. 1988) was decisive in the dissection of more direct links. Notably, mouse minor satellites and human pericentromeric Sat II and Sat III repeats were identified as specific targets of the de novo methyltransferase DNMT3B (Okano et al. 1999; Xu et al. 1999). Human cells deficient for DNMT1 and DNMT3B are hypomethylated on Sat III repeats but do not accumulate pericenRNAs compared to wild-type cells (Eymery et al. 2009a). Similarly, hypomethylation of (peri)centromeric regions in murine embryonic stem cells (mESCs) deficient for Dnmt1 and Dnmt3a/b is not sufficient to promote their transcriptional activation (Lehnertz et al. 2003; Martens et al. 2005). However, in physiopathological cellular contexts further discussed below, loss of DNA methylation at (peri)centromeres has been associated with their transcriptional derepression, although it remains to be determined whether this is a direct consequence or a mere byproduct of the disease states.

It is possible that the methylated state at satellite repeats could influence the binding of transcriptional activators or repressors. It is important to note that two of the methylatable CpGs in human and murine centromeric repeat units reside within the CENP-B-box. Yet, the impact of CpG methylation on CENP-B binding and centromere architecture is still debated (Scelfo and Fachinetti 2019). DNA methylation was shown to prevent CENP-B binding to the CENP-B-box in vitro (Tanaka et al. 2005). Conversely, global demethylation using 5AZA treatment in cultured cells led to CENP-B spreading over demethylated repeats (Mitchell et al. 1996), although nearby pericentromeric repeats without a CENP-B-box are also demethylated in these conditions, making it difficult to conclude. Nevertheless, DNA methylation could be directly linked to the correct assembly of centromere architecture independently of the transcription of the repeats.

In sum, hypomethylation may create a favorable environment for transcriptional activation of satellite repeats, although it might not be sufficient. Hence, one has to consider that specific cellular contexts and their associated tissue- or context-specific transcription factors would be a prerequisite.

7.3 Functional Relevance of Centromeric Transcripts/ Transcription

7.3.1 Centromeres Transcription and Chromatin Remodeling Processes

It appeared that low transcriptional activity is a characteristic of centromeric repeats in normal somatic cells. Several studies suggested that active transcription at centromeres would not only be compatible with but even required for centromere function. Indirect evidence came from the global inhibition of transcription by

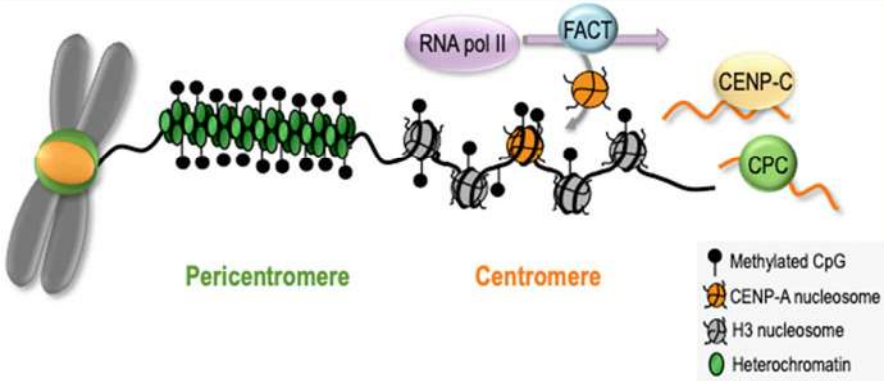


Fig. 7.1 Role of centromere transcription and transcripts in physiological conditions. Centromeric chromosomal domains are marked by a combination of nucleosomes containing histone H3 (gray) or CENP-A (orange) and are flanked by pericentromeric heterochromatin domains (green). RNA pol II-mediated transcription of centromeric repeats, together with chromatin remodelers such as the FACT complex, contribute to the dynamics of chromatin at centromeres and to the deposition of CENP-A. The derived centromeric transcripts serve as scaffolds or guides for the correct localization of centromeric proteins (e.g., CENP-C) and their associated complexes such as the chromosomal passenger complex (CPC), which includes Survivin, INCENP, and Aurora B kinase

RNA Pol II inhibitors which led to compromised centromere function (Chan et al. 2012; Quénet and Dalal 2014). Transcription or nucleosome turnover at centromeres may be important for the dynamics of nucleosomes at centromeric repeats and deposition of CENP-A (Fig. 7.1). Notably, the complex Facilitates Chromatin Transcription (FACT) is localized to centromeres (Foltz et al. 2006) and is involved in CENP-A deposition via the recruitment of the chromatin remodeler Chromodomain Helicase DNA Binding Protein 1 (CHD1) in chickens (Okada et al. 2009). In mammals, a subunit of FACT, the Structure Specific Recognition Protein 1 (SSRP1), co-localized with RNA Pol II at centromeres during mitosis and was necessary for the efficient deposition of CENP-A in early G1 (Chan et al. 2012). Similarly, FACT-mediated transcription was also shown to be required for the de novo incorporation of CENP-A in *Drosophila* S cells (Chen et al. 2015). In this study, knockdown of FACT led to the loss of transcription at centromeres and reduced CENP-A loading. In recent years, a two-step process for *Drosophila* CENP-A loading was proposed (Bobkov et al. 2018). The first step was transcription-independent, during which CENP-A localized to centromeres through the *Drosophila*-specific chaperone Chromosome Alignment defect 1 (CAL1) (Chen et al. 2014), but a second step of active transcription was necessary for its stable incorporation into chromatin. In *S. pombe* mutants that are unable to restart stalled RNA Pol II at centromeres, CENP-A was still efficiently deposited, suggesting that halting RNA Pol II at centromeres may promote local chromatin remodeling events sufficient for CENP-A deposition (Catania et al. 2015). In a context where centromeric transcription-dependent chromatin remodeling is required for stable incorporation of its epigenetic determinant CENP-A, which also relies on the eviction of

previously deposited H3/H3.3-placeholder nucleosomes (Dunleavy et al. 2011), the replication-independent histone chaperone and transcription elongation factor Spt6 (SUPT6H in human) was identified as a conserved CENP-A maintenance factor (Bobkov et al. 2020). Spt6 was shown to prevent loss of centromere identity in a transcription-dependent manner, through promoting the recycling of CENP-A and maintenance of parental CENP-A nucleosomes in both *Drosophila* and human cells (Bobkov et al. 2020).

Importantly, targeting of a strong trans-activation domain from Herpes simplex virus (VP16) to centromeres of HACs impaired the incorporation of newly synthesized CENP-A and led to the eviction of the parental one (Bergmann et al. 2011). Hence, even though transcriptional activation of centromeric repeats is necessary for centromere identity, increased transcription at this locus is incompatible with centromere function since it leads to loss of CENP-A at HAC centromeres (Bergmann et al. 2011).

To date, our knowledge of the immediate contribution of centromere transcription on centromere identity in different species still remains incomplete. Since the direct output from transcription at centromeres is the production of cenRNAs, whether long or short-lived, a question that arises is whether cenRNAs themselves could be implicated in the maintenance of centromere identity and function.

7.3.2 Functional Relevance of Centromeric Transcripts themselves

A growing body of evidence suggests that cenRNAs themselves may contribute to proper kinetochore assembly (Chen et al. 2003; Nakano et al. 2003; Topp et al. 2004; Ferri et al. 2009). Notably, cenRNAs are an integral part of the centromeric fraction (Ferri et al. 2009; McNulty et al. 2017; Kabeche et al. 2018), and coprecipitate with CENP-A in maize (Topp et al. 2004), mouse (Ferri et al. 2009), and humans (Chueh et al. 2009). Importantly, knockdown of human cenRNAs, without affecting transcription of the locus per se, led to impaired CENP-A deposition (Quénet and Dalal 2014). This finding suggested that correct loading of CENP-A does not only depend on active transcription at centromeres (see above), but also requires the transcripts themselves. This was further evidenced by the knockdown of cenRNAs in extracts of *Xenopus* oocytes, which led to decreased occupancy of CENP-A at centromeres (Grenfell et al. 2016).

Besides CENP-A deposition, several studies showed a direct association between cenRNAs and CENP-C in gel shift or immunoprecipitation experiments, probably through the CENP-C RNA binding domain. This was the case for example in maize (Du et al. 2010) and *Drosophila* (Rošić et al. 2014). RNase treatment of human cells induced the delocalization of CENP-C but not that of CENP-A (Wong et al. 2007). Similarly, inhibition of transcription in mitosis, a stage at which centromeres are

transcribed by RNA Pol II, impaired the localization of CENP-C at centromeres (Chan et al. 2012).

Beyond their association with the constitutive components of the centromere, cenRNAs also interact with components of the CPC: Aurora B, INCENP, and Survivin in the G2/M phase in murine cells (Ferri et al. 2009). More specifically, the association of cenRNAs with the mitotic kinase Aurora B was necessary for Aurora B interaction with its partners Survivin and CENP-A and potentiated its kinase activity (Fig. 7.1). Since cenRNAs peak in G2/M, these data therefore suggested a role for cenRNAs in the timely recruitment or stabilization of the CPC components specifically at centromeres before the onset of mitosis. In turn, unscheduled accumulation of murine cenRNAs throughout the cell cycle led to ectopic localization of CPC proteins, mitotic abnormalities, and loss of cohesion between sister chromatids (Bouzinba-Segard et al. 2006). This interaction between cenRNAs and CPC proteins is conserved in *Xenopus* and humans, as seen with the knockdown of cenRNAs which impaired the recruitment of CPC at centromeres (Ideue et al. 2014; Blower 2016). Strikingly, the association of cenRNAs with Aurora B was also shown to be required for both telomerase activity and maintenance of telomere length in mESCs (Mallm and Rippe 2015). In the mouse, centromeres and telomeres being in close proximity, it is possible that cenRNAs may favor a local concentration of the kinase for shared functions on two essential chromosomal domains.

Together, these data emphasize that the fine-tuning of centromeric transcription/transcript levels is absolutely required, since too low or too high transcription or cenRNAs levels have deleterious consequences for centromere identity and function, and hence for normal cell growth and survival, and are emerging as new kinds of players in the development of disease as discussed in the following chapters.

7.3.3 Regulation of the Levels of (Peri)centromeric Transcripts Themselves

As mentioned above, the dynamic levels of cenRNAs according to cellular contexts may not only result from regulatory processes at the level of transcription but also at the level of the transcripts themselves, through fine-tuning of their stability or their processing. In *S. cerevisiae*, low levels of cenRNAs might be insured by their degradation by the exosome (Houseley et al. 2007; Ling and Yuen 2019). In many species, centromeric repeats are transcribed in both sense and antisense orientations (Topp et al. 2004; Li et al. 2008; Carone et al. 2009; Ideue et al. 2014), which would favor the formation of double-stranded RNAs (dsRNAs) that are substrates for further processing by the RNA interference (RNAi) machinery. The first example of the possibility that cenRNAs could be processed into smaller species was provided by the discovery of endogenous small interfering RNA (siRNAs) of centromeric origin in *S. pombe* (Reinhart and Bartel 2002). In addition, deletion of factors

of the *S. pombe* RNAi machinery such as Dicer, the RNA-binding protein Argonaute (AGO), or the RNA-dependent RNA polymerase (RdRP), led to aberrant accumulation of cenRNAs and chromosome missegregation due to defective pericentromeric heterochromatin formation (Volpe et al. 2002). Similarly, in human–chicken hybrid cells (chicken DT40 cells carrying a human chromosome), ablation of Dicer led to mitotic defects and premature sister chromatid separation that was attributed to the loss of HP1 at pericentromeric heterochromatin and mislocalization of the cohesin complex (Fukagawa et al. 2004). In mouse embryonic stem cells (mESCs), Dicer deficiency also caused an accumulation of pericenRNAs, ranging from 40 nt to over 200 nt in size, i.e., not in the size range of siRNAs (Kanellopoulou et al. 2005; Murchison et al. 2005). Dicer has been involved in the repression of pericentromeric repeats in many species, but its depletion did not lead to mitotic defects in mESCs, although they exhibited differentiation and proliferation defects. In human cells, knockdown of Dicer or AGO2 resulted in chromosome lagging and increased levels of cenRNAs (Huang et al. 2015).

In fact, outside of the well-characterized *S. pombe* model, the literature is punctuated by opposing views of the role of Dicer/RNAi pathway in the regulation of the levels of satellite transcripts, with consequences for heterochromatin assembly and chromosomal stability. This is probably related to the failure to detect 25–30 nt RNA species, at least in the mouse (Kanellopoulou et al. 2005; Bouzinba-Segard et al. 2006). However, in mESCs, 150 nt-long cenRNAs, and smaller species but longer than siRNAs, have been detected and shown to rely on Dicer for their biogenesis (Kanellopoulou et al. 2005). Whether these transcripts play similar roles as (peri)centromeric siRNAs found in *S. pombe* is an interesting possibility. Of note, whereas exponentially growing somatic murine cells exhibit low levels of 2–4 Kb cenRNAs, the accumulation of 120 nt-long cenRNAs in physiopathological conditions recapitulated the same phenotypic defects observed in Dicer-deficient *S. pombe* that failed to produce centromeric siRNAs, including loss of sister chromatid cohesion, impaired centromere architecture and heterochromatin organization, associated with mitotic defects (Bouzinba-Segard et al. 2006). Although it is still not known whether these shorter cenRNAs result from cleavage through unknown mechanisms or are produced by multiple transcription initiation events, these data suggested that the absence of mature siRNAs or the accumulation of unprocessed longer RNA species have the same impact on the integrity of centromeric regions (Bouzinba-Segard et al. 2006).

The low levels of cenRNAs in normal conditions could indicate a short half-life caused by the rapid degradation of the transcripts through mechanisms exposed above or through posttranscriptional modifications shown to regulate the stability of transcripts (Nachtergaele and He 2017; Boo and Kim 2020). Among the myriad of known posttranscriptional RNA modifications, the adenosine to inosine (A to I) modification mediated by the adenosine deaminase acting on RNA (ADAR) machinery was shown to edit structured dsRNAs, with selectivity for certain internal loops and bulges rather than for a consensus sequence (Levanon et al. 2005). Just like the excitement of a connection between the RNAi machinery in keeping (peri)-centromeric transcription in check in a broad range of eukaryotes in the early

years of 2000 (Lippman and Martienssen 2004), the involvement of ADAR in the silencing of repetitive sequences in heterochromatin also attracted much interest a few years later (Fernandez et al. 2005). The A/T-rich and potentially dsRNAs produced from murine pericentromeric satellite repeats (Kanellopoulou et al. 2005) would make excellent substrates for the deamination of adenosine to inosine residues by ADAR. However, the search for such editing and for an immunolocalization of factors of the ADAR complex at (peri)centromeric domains remained unsuccessful (Lu and Gilbert 2008). Yet, it is interesting to note that RNA editing by ADAR is incompatible with the RNAi machinery (Scadden and Smith 2001). This is exemplified by the A to I conversion on microRNAs (miRNAs) derived from repetitive Long interspersed nuclear element 2 (LINE2), which blocks their cleavage by Dicer in human and mouse (Kawahara et al. 2007).

Thus, there is an exciting possibility that the dynamic size range of cenRNAs, with potentially distinct roles, could result from the fine-tuning of a balance between RNA modifications and RNA processing depending on phases of the cell cycle or cellular contexts.

7.4 Centromeric Transcripts: Cause or Consequence of Disease?

7.4.1 Accumulation of Satellite Transcripts in Various Types of Cellular Stress

Cells are constantly exposed to various environmental or endogenous stresses that may jeopardize their identity and viability. Exogenous sources of cellular stress include high temperatures [heat shock (HS)], DNA damaging chemicals, UV and ionizing radiation, hyperosmotic and oxidative stresses, whereas endogenous stress can originate from cellular metabolism or replication defects. In this context, a major challenge for the cells is therefore to safeguard their genome integrity. This is fundamental for their survival but also for the normal functioning of the whole organism and protection from the emergence of disease states. In that respect, cells have evolved sophisticated mechanisms that can trigger a rapid and adapted cellular response, including cell-cycle arrest, to allow time to repair DNA lesions or activate sets of genes to recover from stress, and therefore maintain their genomic integrity. There is now a large body of evidence showing that transcription of (peri)-centromeric satellite repeats is rapidly induced in cells under various stress conditions, through the activity of transcription factors belonging to different stress-response pathways, thereby increasing the levels of resulting satellite transcripts as integral components of the stress response.

The accumulation of human pericenRNAs from Sat III of chromosome 9 was the first and best-characterized example of the accumulation of satellite transcripts in stress conditions. Originally described in response to HS (Jolly et al. 2004; Rizzi

et al. 2004), it occurs in all of the above-cited types of stress conditions (Valgardsdottir et al. 2008). In response to HS, transient subnuclear organelles, called nuclear stress bodies (nSB), assemble on blocks of Sat III DNA repeats at which the Heat Shock Factor 1 (HSF1) binds, which in turn recruits the RNA Pol II and promotes transcriptional activation of the repeats (Biamonti and Vourc'h 2010). In stressed cells, nSBs are thought to contribute to the rapid and transient shutdown or reprogramming of gene expression programs required for the cells to recover from stress, through a Sat III transcripts-mediated trapping of a subset of splicing and transcription factors away from their site of action (Eymery et al. 2010). Pericentromeric heterochromatin is also central to the functional organization of the cell nucleus and to the maintenance of gene silencing through their positioning in the vicinity of subnuclear heterochromatin compartments (Francastel et al. 2000; Fisher and Merkenschlager 2002). Hence, transcriptional activation of pericentromeric Sat III sequences may also have a long-range impact on gene expression programs through the disorganization of these repressive nuclear compartments. It is interesting to note that activated transcription of satellite repeats in response to thermal perturbations appears to be a shared feature as it also occurs in beetles (Pezer and Ugarkovic 2012), *Arabidopsis thaliana* (Pecinka et al. 2010; Tittel-Elmer et al. 2010), and mouse (Hédouin et al. 2017). However, whether it triggers similar mechanisms in these organisms is not known. In fact, in the mouse, HS stimulates only a modest increase in cenRNAs levels but not that from pericentromeric repeats, whereas nSBs described in human cells do not form in murine cells. Hence, cellular responses to stress may vary between organisms although satellite transcripts appear to be central to the stress response.

In contrast to HS in murine cells, genotoxic stress led to a strong and rapid transcriptional activation of centromeric repeats, followed by local accumulation of cenRNAs at their site of transcription (Hédouin et al. 2017). Transcriptional activation, and not the transcripts themselves, has been causally linked to the loss of centromere identity characterized by the delocalization of CENP-A away from its default location. CENP-A delocalization, or eviction of CENP-A nucleosomes, was dependent on the chromatin remodeler FACT, pinpointing another function of FACT at centromeres in nucleosome destabilization at centromeres, and on the DNA Damage Response (DDR) effector ATM (Hédouin et al. 2017). Importantly, genotoxic stress-induced transcriptional activation of centromeric repeats had distinct functional consequences for cellular phenotypes depending on the integrity of the p53 checkpoint. Whereas immortalized cells continued to cycle, while accumulating micronuclei indicative of mitotic errors and centromere dysfunction, primary cells with normal p53 entered premature cell-cycle arrest and senescence (Hédouin et al. 2017). Hence, in the mouse, activated transcription at centromeric repeats provides a safeguard mechanism to prevent genomic instability in the context of persistent DNA damage signaling, through the disassembly of the core components of centromere identity and function. Whether this mechanisms is related to the acrocentric structure of murine chromosomes, i.e., with telomeres close to centromeric repeats, or to the mouse having very long telomeres so they enter senescence through mechanisms distinct than telomere shortening, is not known. However,

together with the example of increased levels of Sat III pericenRNAs in heat-shocked cells, this illustrated the functional relevance for satellite transcripts in the stress response and protection of genome integrity.

7.4.2 *Deregulation of Satellite Transcription/Transcripts in Cancer*

Consistent with the above-mentioned links between transcription of satellite repeats and the triggering of safeguard mechanisms, it is not surprising that aberrant accumulation of transcripts from satellite repeats characterizes disease states with chromosomal instability like cancer, and that they actually represent good biomarkers of cancerous lesions (Eymery et al. 2009a; Ting et al. 2011; Zhu et al. 2011; Bersani et al. 2015; Tasselli et al. 2016; Hall et al. 2017). Yet, whether they are mere byproducts of disease phenotypes or act as drivers of disease, and through which mechanisms, are still open questions.

Macroarray-based approaches, designed to assess levels of transcripts from various repeated elements including satellite sequences, showed that the levels of satellite transcripts are higher in a variety of cancer cells compared to their normal healthy counterparts (Eymery et al. 2009a). Aberrant accumulation of satellite transcripts has also been reported in a wide range of primary epithelial tumors, both in humans and mice (Eymery et al. 2009a; Ting et al. 2011; Zhu et al. 2011). High-throughput sequencing has further highlighted that satellite transcripts actually represent up to 50% of transcriptional output in these tumors, which is in stark contrast to the low levels of these RNAs in normal cells (Ting et al. 2011).

An important open question remains as to the mechanisms that lead to the pathological transcriptional derepression of satellite sequences. Abnormal levels of satellite transcripts are often associated with the global hypomethylation that characterizes cancer cells, which in fact reflects reduced DNA methylation at repeated sequences owing to the large fraction of the genome they represent and their heavily methylated state in normal cells (Ross et al. 2010). The derepression of satellite transcripts was indeed shown to correlate with reduced DNA methylation of the underlying repeats (Ting et al. 2011; Unoki et al. 2020).

More hints into causal links between the aberrant accumulation of satellite transcripts and chromosomal instability came from gain-of-function experiments where cenRNAs were ectopically transcribed from expression vectors. In murine cells, ectopic expression of minor satellites from one repeat unit was sufficient to promote mitotic defects and alterations of nuclear organization typical of cancer cells (Bouzinba-Segard et al. 2006). Unscheduled accumulation of cenRNAs led to mitotic errors and disorganized centromere architecture through the trapping of centromeric protein complexes away from their default location (Bouzinba-Segard et al. 2006), providing a direct link between high levels of cenRNAs and centromere dysfunction. Likewise, ectopic expression or injection of satellite transcripts in

cultured human or murine cells also led to mitotic errors (Zhu et al. 2011, 2018; Kishikawa et al. 2016, 2018), in correlation with the accumulation of foci of phosphorylated histone H2A.X (γ-H2A) that marks DNA double-strand breaks (DSBs) (Zhu et al. 2011). These data suggested that increased levels of satellite transcripts, and not transcriptional activation per se, is deleterious to the cells and leads to increased DNA mutation rates. They also put forward an interesting interplay between pathological high levels of satellite transcripts and DNA damage. In support of this hypothesis, there is the finding that these satellite transcripts accumulate strongly in breast cancer cells deficient for the *BRCA1* gene (Zhu et al. 2011). Likewise, BRCA1 depletion has been causally linked to elevated levels of cenRNAs associated with impaired centromere architecture and chromosome missegregation (Di Paolo et al. 2014). BRCA1 is an important repair factor which, in normal conditions, occupies centromeric chromatin in interphase and throughout mitosis in normal cells (Pageau and Lawrence 2006; Di Paolo et al. 2014; Gupta et al. 2018), and may serve as a guardian of centromere integrity. Aberrant levels of cenRNAs, just like it has been described for kinetochore proteins (Bouzinba-Segard et al. 2006), were proposed to lead to BRCA1 delocalization away from centromeres and further exposure of this locus to the accumulation of unrepaired genotoxic insults (Zhu et al. 2018). Yet, the molecular mechanisms may not be that straightforward since BRCA1, besides its role in DNA damage repair, operates pleiotropic functions linked to the maintenance of chromosomal stability including at the replication fork, control of the cell cycle, and many other regulatory functions (Savage and Harkin 2015). BRCA1 was also recently shown to be an important determinant of the epigenetic states of centromeric and pericentromeric chromatin, through its ubiquitin ligase activity (Zhu et al. 2011). H2A ubiquitination at Lys 19 by BRCA1 provides a repressive mark at centromeric repeats important for their transcriptional repression. Hence, pathogenic variants of BRCA1 would promote DNA damage through the derepression of centromeric repeats in addition to, or instead of, promoting the accumulation of satellite transcripts. Along the same lines, centromeric targeting of VP16 for transcriptional activation of the underlying repeats in murine cells also promoted chromosomal instability (Zhu et al. 2018). Conversely, genotoxic stress using DSB inducers triggered the rapid transcriptional activation of murine centromeric repeats in a p53- and ATM-dependent dependent manner (Hédouin et al. 2017). In that case, transcriptional activation preceded, and was required for, eviction of CENP-A from centromeric chromatin, suggesting a direct link between activated transcription at centromeres and loss of centromere function and identity.

All these data suggested that our vision of the functional consequences of activated transcription of satellite repeats or accumulation of their related transcripts, and the actors at play, on centromere function and chromosomal stability is still only partial and may depend on organisms and cellular contexts. In the mouse, activated transcription of centromeric repeats and delocalization of CENP-A is associated with premature senescence in primary cells, whereas immortalized cells with impaired p53 checkpoint continue to cycle while accumulating mitotic errors and micronuclei, indicative of chromosomal instability (Hédouin et al. 2017). Thus, at least in the

mouse, a functional p53 pathway is an important surveillance mechanism for centromere integrity, although the mechanisms remain unknown. Interestingly, p53-deficient mice ectopically expressing either human cenRNAs or murine pericenRNAs, were susceptible to tumor formation in mammary glands (Zhu et al. 2018). Hence, alterations to centromeric transcription may cooperate with oncogenic events or loss of tumor-suppressor function to promote oncogenesis.

Many questions remain unanswered as to the direct and reciprocal links between pathological hypomethylation of satellite repeats, their transcriptional derepression, the accumulation of DNA damage, and chromosomal instability. As cancer is a complex multifactorial disease, the current challenge is to dissect further and order these events.

7.4.3 Deregulation of Satellite Transcripts in the ICF Syndrome

A major breakthrough in the medical field came from the identification of inherited disorders of the epigenetic machinery, which provided interesting monogenic contexts and unsuspected players in a number of biological processes (Velasco and Francastel 2019). The first example of such developmental rare diseases was the Immunodeficiency with Centromeric instability and Facial anomalies (ICF) syndrome, a rare autosomal recessive immunological/neurological disorder with typical centromeric instability, including the presence of unusual multiradial chromosomal figures, decondensation, and rearrangement of (peri)centromeric regions (Ehrlich et al. 2006; Francastel and Magdinier 2019). At the molecular level, it is a remarkable case where these chromosomal alterations are caused by constitutive defects in DNA methylation, especially visible at heterochromatin blocks in juxtacentromeric position regions of chromosomes 1, 9, and 16 (Satellites type II and III) in all patients (Ehrlich et al. 2006). In a subset of patients, additional hypomethylation of centromeric α-satellite repeats suggested the genetic heterogeneity of the disease (Jiang et al. 2005; Toubiana et al. 2018).

Studies of the etiology of this rare disease have been instrumental in the identification of essential factors for the methylated state of (peri)centromeric repeats and the maintenance of their integrity. Hypomorphic mutations in the DNMT3B gene were the first identified genetic cause in about half of the patients (Xu et al. 1999), concomitantly implicated in de novo DNA methylation at centromeres in the mouse (Okano et al. 1999). The disease gained renewed interest when, in the reminder of patients, under the same diagnosis but with additional DNA methylation loss at centromeric repeats, exome sequencing identified mutations in factors with very few known functions and strikingly devoid of DNA methyltransferase (DNMT) activity (de Greef et al. 2011; Thijssen et al. 2015). These factors are transcription factors (ZBTB24, CDCA7) or chromatin remodeler of the SWI/SNF2 family (HELLS), the latter having already been shown to play a role in DNA methylation at murine

centromeric repeats (Zhu et al. 2006). RNA interference performed in somatic cells, where DNA methylation profiles are already established, further demonstrated the requirement for ZBTB24, CDCA7, and HELLS in DNA methylation maintenance at murine centromeric repeats (Thijssen et al. 2015). These findings represented a major breakthrough in the knowledge of the determinants of DNA methylation at centromeric repeats. Yet, they raised again the question of the mechanisms that link hypomethylation of centromeric repeats to centromere loss of integrity, and the question of the contribution of non-DNMT ICF factors in DNA methylation and integrity of centromeres.

Independently of a putative role in DNA methylation pathways, a role for CDCA7 and HELLS in DNA repair pathways has been recently reported, reinforcing the idea of a link between DNA damage and centromere integrity (Burrage et al. 2012; Unoki et al. 2019). Notably, human embryonic kidney HEK-293T cells engineered to reproduce CDCA7 and HELLS mutations found in ICF patients exhibited a compromised nonhomologous end joining (NHEJ) DNA repair pathway (Unoki et al. 2019). Consistent with an aberrant accumulation of defects in DNA repair at centromeres, these cells accumulated micronuclei and suffered from abnormal chromosome segregation, while satellite repeats retained their methylated status. These engineered ICF cells, as well as cells from ICF patients, also exhibited increased transcription of (peri)centromeric repeats (Unoki et al. 2020). Given that genotoxic stress promotes a rapid transcriptional activation at centromeres (Hédouin et al. 2017), and along with the findings that satellite transcripts accumulate in breast cancer cells deficient for the DNA repair factor BRCA1 (Zhu et al. 2011), these data, therefore, suggested that DNA damage may trigger transcriptional activation at satellite repeats. An alternative, or concomitant, the mechanism could be that factors like CDCA7, HELLS, or BRCA1 may protect transcribed centromeric repeats from the accumulation of deleterious DNA:RNA hybrids (R-loops), just like BRCA1 does at transcriptional termination pause sites of actively transcribed genes (Hatchi et al. 2015). R-loops are dynamic and abundant structures that have been involved in a variety of physiological processes including chromosome segregation (Kabeche et al. 2018), whereas their unscheduled accumulation is also a source of DNA damage and genome instability (Costantino and Koshland 2018; Mishra et al. 2021). R-loops have been observed to accumulate at (peri)centromeres in engineered ICF cells and cells from ICF patients (Unoki et al. 2020). However, whether the transcriptional derepression and subsequent R-loop formation arise directly through DNA methylation loss or loss of function of ICF factors acting as “guardians” or transcriptional repressors of (peri)centromeric repeats, remains to be determined.

Like in cancer cells, transcriptional derepression or accumulation of the related satellite transcripts may represent intermediate steps between pathological hypomethylation of satellite repeats and chromosomal instability (Fig. 7.2). Importantly, the ICF syndrome leads to the premature death of the patients in early childhood from repeated infections, and despite a few reported cases where patients developed cancer, it is not clear whether pathological hypomethylation of centromeric repeats would favor later complications and further emergence of cancer. It

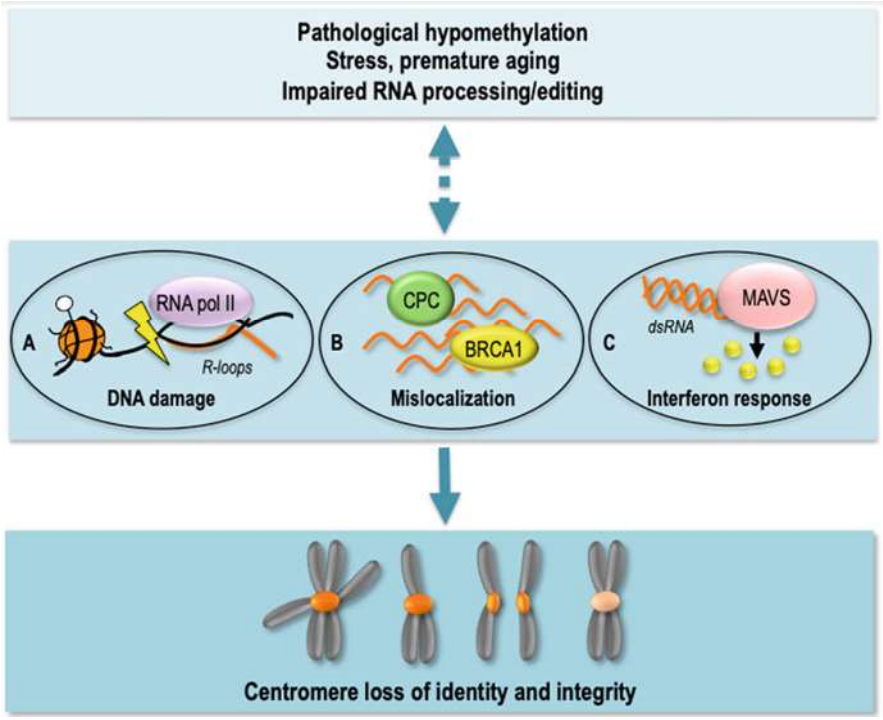


Fig. 7.2 Increased centromere transcription/transcripts levels: missing links between physiopathological DNA hypomethylation/DNA damage at centromeres and loss of centromere identity and function. Pathological hypomethylation, DNA damage at centromeres, or potentially impaired RNA processing or editing could promote: (**a**) unscheduled activated transcription of centromeric repeats, which in turn would lead to the formation of genotoxic R-loops; (**b**) aberrant accumulation of cenRNAs and trapping of kinetochore and DNA repair proteins away from centromeres; (**c**) the formation of double-stranded cenRNAs, known to trigger inflammatory responses. Although direct links between all these events remain to be formally dissected, the deregulation of centromeric transcription/transcripts ultimately lead to loss of centromere identity and integrity, as exemplified by multiradial chromosome figures, loss of sister chromatid cohesion, or recombination events between satellite repeats

also pointed out again that the impact of higher levels of satellite repeats transcripts on cellular phenotypes depends on the context in which it occurs.

A possibility that has not been evoked yet is that loss of DNA methylation at (peri)centromeric repeats and its associated abnormal levels of satellite RNAs may trigger surveillance mechanisms, which *in fine* would activate an interferon inflammatory response (Rajshekar et al. 2018). This has been nicely shown in a Zebrafish ICF model where one of the earliest in vivo consequences of ZBTB24 loss of function is a progressive loss of DNA methylation at pericentromeric regions associated with the derepression of sense and antisense pericenRNAs. This in turn triggered an interferon-dependent immune response mediated by the Melanoma Differentiation-Associated gene 5 (MDA5) and Mitochondrial AntiViral Signaling

(MAVS) machinery, an antiviral surveillance mechanism that senses dsRNAs (Berke et al. 2013). Injection of sense and antisense pericenRNAs in Zebrafish embryos was also sufficient to stimulate the innate immunity (Rajshekar et al. 2018), implicating the accumulation of pericenRNAs as an important trigger of autoimmunity in a variety of diseases.

7.5 Conclusion

All of the data exposed in this chapter lend support to the essential nature of temporal control of the act of transcription through centromeric satellite repeats for the determination and correct functioning of this chromosomal region in most of the species studied so far. Transcription per se would facilitate the dynamic exchange of nucleosomes for the deposition of the key determinant of centromere identity, CENP-A, but would also favor a local concentration of the transcripts themselves for the timely recruitment of other centromere components. Yet, it is still unclear which molecular mechanisms and regulatory pathways are involved for a timely control in normal conditions, although we mentioned transcription factors acting in defined chromatin environments and RNA-based mechanisms for the regulation of the transcripts levels.

In turn, unscheduled transcription or aberrant levels of the transcripts have profound consequences for both centromere function and cell fate. We have seen that activated transcription of (peri)centromeric repeats or ectopic accumulation of the transcripts, i.e. elsewhere than at centromeres, under stress conditions is a mechanism adopted by many organisms to trigger rapid cellular responses for cells to recover from stress. This type of response is possible through the trapping of various regulatory factors away from their site of action and impairment of their associated functions, to favor genome repair or remodeling of gene expression programs. In turn, unscheduled transcription or accumulation of the transcripts coincides with disease states. In that case, they are not seen as safeguard mechanisms, which implicitly infer collaborative effects with disease conditions like defective checkpoints, oncogenic events, or even an inflammatory environment that ultimately alter cellular phenotypes. Pathological hypomethylation of satellite repeats like in cancer or ICF syndrome is a good candidate for uncontrolled transcription of satellite repeats, although we have seen that it is not necessarily sufficient and that opportunistic tissue- or context-specific factors may come into play. This might explain why all cancer cells do not necessarily exhibit increased transcription of satellite repeats, and why ICF patients do not have widespread alterations in all their tissues. Alternatively, or in addition to, defective RNA-based surveillance mechanisms might also contribute to the abnormal elevation of the levels of satellite transcripts.

In sum, the use of a wide range of model organisms and artificial centromeres allowed to identify a large number of centromere and kinetochore proteins, to address the relevance of DNA sequences for centromere identity, and to tackle the

functional relevance of centromeres transcription/transcripts for centromere identity and function. Studies of the etiology of complex or monogenic human diseases further identified key determinants for centromere integrity and function, among which we can cite factors with DNA repair or chromatin remodeling activities, many of which could not be suspected before their implication in centromeric instability diseases. Yet, our vision of the intricate contribution of all the actors and mechanisms mentioned throughout this chapter still remains fragmentary and will require the development of targeted approaches, many of which are still missing in mammalian systems.

Acknowledgments We are grateful to Ekaterina Boyarchuk for the critical reading of the manuscript and Florent Hubé for the help with figures.

We thank our tutelles University de Paris, Centre National de la Recherche Scientifique (CNRS) and Institut National de la Recherche Médicale (INSERM). Work in C.F.'s lab is supported by research grants from Agence Nationale pour la Recherche (grant ANR-19-CE12-0022), Ligue Nationale Contre le Cancer, Fondation ARC and Fondation Maladies Rares. P.M. is a recipient of a fellowship from the French Ministry of Research through the Ecole Universitaire Génétique et Epigénétique Nouvelle Ecole (EUR GENE).

References

Aird WC, Parvin JD, Sharp PA, Rosenberg RD (1994) The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters A model of tissue-specific gene expression. J Biol Chem 269:883–889

Akiyoshi B, Gull K (2014) Discovery of unconventional kinetochores in kinetoplastids. Cell 156:1247–1258

Aldrup-Macdonald ME, Sullivan BA (2014) The past, present, and future of human centromere genomics. Genes (Basel) 5:33–50

Bergmann JH, Rodríguez MG, Martins NMC, Kimura H, Kelly DA, Masumoto H, Larionov V, Jansen LET, Earnshaw WC (2011) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J 30:328–340

Bergmann JH, Martins NMC, Larionov V, Masumoto H, Earnshaw WC (2012) HACking the centromere chromatin code: insights from human artificial chromosomes. Chromosom Res 20:505–519

Berke IC, Li Y, Modis Y (2013) Structural basis of innate immune recognition of viral RNA. Cell Microbiol 15:386–394

Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, MacKenzie OC, Brannigan BW, Wittner BS, Jung H et al (2015) Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. Proc Natl Acad Sci USA 112:15148–15153

Bestor T, Laudano A, Mattaliano R, Ingram V (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203:971–983

Biamonti G, Vourc'h C (2010) Nuclear stress bodies. Cold Spring Harb Perspect Biol 2:a000695

Blower MD (2016) Centromeric transcription regulates Aurora-B localization and activation. Cell Rep 15:1624–1633

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

Bobkov GOM, Gilbert N, Heun P (2018) Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol 217:1957–1972
Bobkov GOM, Huang A, van den Berg SJW, Mitra S, Anselm E, Lazou V, Schunter S, Feederle R, Imhof A, Lusser A et al (2020) Spt6 is a maintenance factor for centromeric CENP-A. Nat Commun 11:2919
Boo SH, Kim YK (2020) The emerging role of RNA modifications in the regulation of mRNA stability. Exp Mol Med 52:400–408
Bouzinba-Segard H, Guais A, Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci 103:8709–8714
Bracewell R, Chatla K, Nalley MJ, Bachtrog D (2019) Dynamic turnover of centromeres drives karyotype evolution in Drosophila. eLife 8
Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell 91:845–854
Burrage J, Termanis A, Geissner A, Myant K, Gordon K, Stancheva I (2012) The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. J Cell Sci 125:5524–5534
Bury L, Moodie B, Ly J, McKay LS, Miga KH, Cheeseman IM (2020) Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. eLife 9:e59770
Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ et al (2009) A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma 118:113–125
Catania S, Pidoux AL, Allshire RC (2015) Sequence features and transcriptional stalling within centromere DNA promote establishment of CENP-A chromatin. PLoS Genet 11:e1004986
Chan FL, Marshall OJ, Saffery R, Kim BW, Earle E, Choo KHA, Wong LH (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci USA 109:1979–1984
Chang C-H, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, Chen C-C, Erceg J, Beliveau BJ, Wu C-T et al (2019) Islands of retroelements are major components of Drosophila centromeres. PLoS Biol 17:e3000241
Chen ES, Saitoh S, Yanagida M, Takahashi K (2003) A cell cycle-regulated GATA factor promotes centromeric localization of CENP-A in fission yeast. Mol Cell 11:175–187
Chen C-C, Dechassa ML, Bettini E, Ledoux MB, Belisario C, Heun P, Luger K, Mellone BG (2014) CAL1 is the Drosophila CENP-A assembly factor. J Cell Biol 204:313–329
Chen C-C, Bowers S, Lipinszki Z, Palladino J, Trusiak S, Bettini E, Rosin L, Przewloka MR, Glover DM, O'Neill RJ et al (2015) Establishment of Centromeric chromatin by the CENP-A assembly factor CAL1 requires FACT-mediated transcription. Dev Cell 34:73–84
Chen C-F, Pohl TJ, Chan A, Slocum JS, Zakian VA (2019) *Saccharomyces cerevisiae* centromere RNA is negatively regulated by Cbf1 and its unscheduled synthesis impacts CenH3 binding. Genetics 213:465–479
Choi ES, Strålfors A, Castillo AG, Durand-Dubief M, Ekwall K, Allshire RC (2011) Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. J Biol Chem 286:23600–23607
Choo KH (1997) Centromere DNA dynamics: latent centromeres and neocentromere formation. Am J Hum Genet 61:1225–1233
Choo KHA (2001) Domain organization at the centromere and neocentromere. Dev Cell 1:165–177
Choy JS, Acuña R, Au W-C, Basrai MA (2011) A role for histone H4K16 hypoacetylation in *Saccharomyces cerevisiae* kinetochore function. Genetics 1:11–21
Christova R, Oelgeschläger T (2002) Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. Nat Cell Biol 4:79–82

3. DNA damage and the DNA damage response

Human cells are estimated to suffer from a staggering 100,000 damaging events per day (Saul and Ames, 1986). To survive and function properly, cells have evolved mechanisms to control the integrity of their genome, such as cell cycle checkpoints and signaling pathways (Poon, 2016). These control mechanisms operate before, during and after cell division, to ensure that the DNA has been faithfully replicated, that there is no DNA damage and that the correct segregation of chromosomes occurred (Takatsuka *et al.*, 2021).

Alterations to the DNA could potentially cause adverse effects on the fitness and survival of an organism. If left unrepaired, these lesions can accumulate and lead to genome instability, which includes mutations in the DNA sequence, chromosomal rearrangements and impaired chromosomal segregation that can lead to aneuploidy (Gollin, 2004), all phenomena that can be harmful to life. Indeed, genomic instability is a common feature of aging as well as diseases such as cancer (Hanahan and Weinberg, 2011; Vijg and Suh, 2013). Therefore, proper control and maintenance of the genome is essential to ensure a healthy development and life.

DNA damage and the DNA damage response

DNA lesions represent a serious threat for genome integrity (reviewed in Ferrand *et al.*, 2021). DNA breaks can be caused by both internal and external cues. Internal forces affecting genome integrity include, among others, the error rate of the DNA replication machinery, the efficiency of DNA repair pathways, reactive oxidant byproducts of cellular metabolism, active transposable elements (TEs) and aberrant DNA structures like R-loops (three-stranded structures which will be described in detail in Chapter 5) (Tubbs and Nussenzweig, 2017). On the other hand, external stressors can come from UV and ionizing radiations or carcinogenic chemicals (Ghosh and Ghosh, 2021; Mullenders, 2018).

To overcome the constant insults on the genome, cells employ efficient DNA damage response (DDR) pathways that aim to restore the DNA and to counteract the various genotoxic stresses, in order to survive or to activate cell death pathways if the damage is too great. These pathways are essential for the maintenance of genomic integrity, since unrepaired or unsuccessfully repaired damage can be a source of genetic mutations or chromosomal rearrangements, and can promote tumorigenesis (García-de-Teresa *et al.*, 2020; Helleday *et al.*, 2008; Ma *et al.*, 2018).

Briefly, in the DDR cascade, sensor proteins are first to act, by detecting DNA damage and/or chromatin alterations. Next, transducer proteins, which are usually kinases, amplify the damage signal. Last in the chain are the effector proteins that determine the outcome for the cell, which may be the repair of damaged DNA, cell cycle arrest, transcription of stress response genes, or even cell death if the damage is too vast (Ciccia and Elledge, 2010). In addition, specific proteins act as mediators by promoting the interaction between the other DDR proteins.

The nature of DNA damage can be very diverse, ranging from chemical modification of bases such as oxidation, pyrimidine dimers, intra- or inter-strand bridging and finally single or double strand breaks (Chatterjee and Walker, 2017). Different types of DNA damage can be repaired by multiple DNA repair pathways, which are highly conserved during evolution. In the following section, I will focus only on the double-stranded DNA damage and its repair pathways.

Repair of double-strand breaks (DSBs)

DSBs represent a substantial threat to the cell since, if the two free ends of DNA at the break site dissociate, this can lead to chromosomal alterations such as translocations or loss of genetic material. Two main pathways are involved in DSB repair: Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR). These two repair pathways involve different proteins and they differ also in their activation timing. Indeed, HR is mainly activated during the S and G2 phases of the cell cycle (except for centromeric repeats, which will be discussed bellow) since it requires the presence of a homologous strand to repair the lesion, whereas NHEJ is predominant in the G1 phase but is also active throughout the cell cycle (Li *et al.*, 2019; Lieber, 2010).

DNA repair by homologous recombination

Homologous recombination is generally an error-free pathway, since it relies on the homology present on the sister chromatid synthesized during replication (**Figure 13**). In short, the break is detected by the Mre11-Rad50-Nbs1 (MRN) complex, which possesses endonuclease and 3'-5' exonuclease activities that are responsible for the resection step (Qiu and Huang, 2021). The generated single-stranded region of DNA initially becomes coated with the single strand binding protein called Replication protein A (RPA) (Iyama and Wilson, 2013). Then, the BReast CAncer gene 1 (BRCA1) protein activates the recombination protein Rad51 (Zhao *et al.*, 2017). Next, Rad51 seeks sequence homology and promotes invasion of the template

strand. The 3' end of the cut is elongated by DNA polymerase, using the two strands of the sister chromatid as a template. The Holliday junctions, formed by the crossing over of the strands, are resolved by specific nucleases called Resolvases, as well as by helicases. Finally DNA Ligase 1 restores the continuity of the two DNA strands (Ciccia and Elledge, 2010).

Generally, HR is limited to the late S to G2 phase of the cell cycle, when sister chromatids are available. However, an interesting exception is the case of centromeric satellite repeats (Yilmaz *et al.*, 2021). At this locus, even in the absence of the sister chromatid, DNA breaks in G1 activate the HR pathway (Yilmaz *et al.*, 2021). This process relies on centromeric transcription and DNA–RNA hybrid formation (discussed in Chapter 5), which in turn promote DNA-end resection.

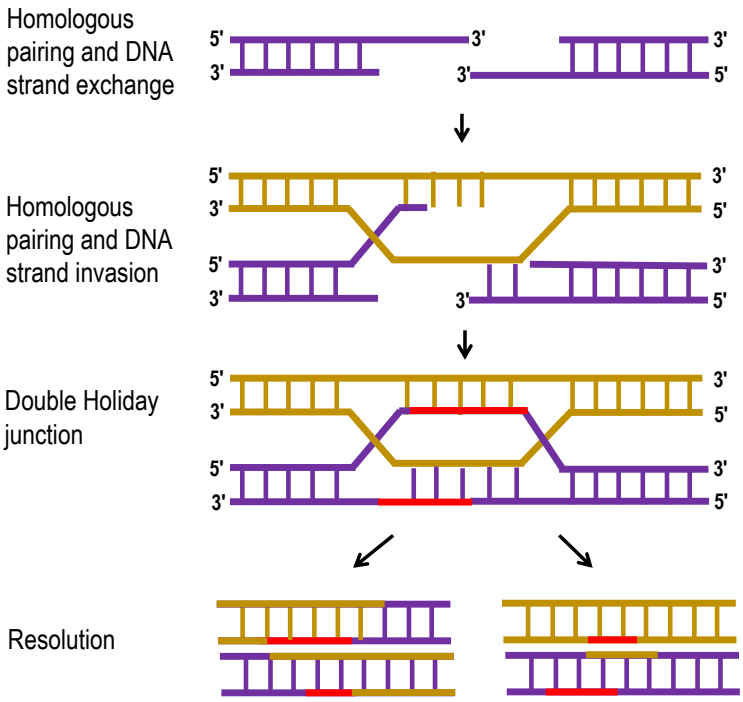


Figure 13. DNA damage repair by homologous recombination. HR involves the excision of damaged or incorrect bases from the damaged strand (purple) and requires the resynthesis of the DNA (red), using the intact strand as a template (yellow). Modified from Li and Heyer, 2008.

Non-homologous end joining

When no homologous strand is available for recombination, *i.e.* outside of the S/G2 phase, repair is done directly *via* ligation of the ends produced by DSBs, independently of the sequence. Since there is no sequence homology, NHEJ can cause errors such as the ligation of ends of two different chromosomes or the appearance of deletions caused by the loss of nucleotides around the break. Globally, NHEJ repair starts with the Ku heterodimer composed of Ku70 and Ku80, which recognize and bind the DSB ends of the damaged site (Mahaney *et al.*, 2009). This complex then recruits and activates DNA-PKcs, the catalytic subunit of DNA-PK (DNA-dependent protein kinase). The activated DNA-PKc phosphorylates itself as

well as several substrates, including p53 and replication protein A (RPA) (Wang *et al.*, 2003). Finally, DNA-PK recruits the ligation complex of DNA Ligase 4, X-Ray Repair Cross Complementing 4 (XRCC4) and XRCC4-like factor (XLF) (Drouet *et al.*, 2005) to restore DNA strand integrity (Roberts *et al.*, 2010) (**Figure 14**). If no further processing of the ends is required, DNA Ligase 4, XRCC4 and XLF can promote the rejoining reaction. Otherwise, the extremities might require processing by the nuclease Artemis and re-synthesis by DNA polymerase Terminal deoxynucleotidyl Transferase (TdT), polymerase lambda and polymerase mu (Pannunzio *et al.*, 2018).

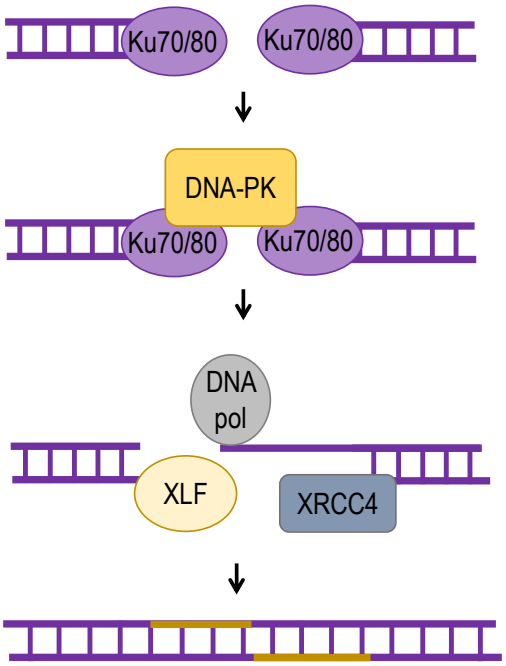


Figure 14. DNA repair by non-homologous end joining.
Broken DNA extremities are bound by the heterodimer Ku70/80, which activates the cascade that leads to synthesis and ligation of the broken ends. If the two DNA extremities require end processing (as depicted), then such processing may require the activities of DNA pol. Modified from Gonzalez-Marin et al, 2012.

Chromatin response to DNA damage

DNA lesions appear on a highly organized chromatin substrate. To relieve the chromatin barrier for subsequent DNA repair, chromatin is subjected to extensive remodeling and post-translational modifications (Reviewed in Ferrand *et al.*, 2021). Notably, at the break site, chromatin undergoes rapid local decondensation, characterized in particular by nucleosome destabilization (Price and D'Andrea, 2013). Furthermore, it appears that the ability of repair factors to detect and to be efficiently retained at DNA lesions is dependent on modifications of histones located around the lesion (Polo and Jackson, 2011). In addition to phosphorylation of the H2AX variant, other modifications, such as acetylation, methylation, phosphorylation, and ubiquitination of histones are also linked to various aspects of DNA repair by facilitating

DNA accessibility or by recruiting repair or remodeling proteins (Polo and Jackson, 2011). Histone variants are also highly dynamic during the DNA damage response and are often deposited at sites of damage (Adam *et al.*, 2015).

The case of centromeric DNA damage

Large arrays of tandem DNA repeats are considered to be chromosomal fragile sites since they are prone to DNA breaks and rearrangements (Barra and Fachinetti, 2018). Interestingly, DNA damage at centromeric and pericentromeric tandem repeats can activate different DNA repair pathways. More specifically, centromeric repeats seem to recruit both NHEJ and HR proteins throughout the cell cycle, while pericentromeric repeats recruit NHEJ factors in G1 and HR factors in S/G2 (Tsouroula *et al.*, 2016) (**Figure 15**). The discrepancy in DNA repair pathway is not completely understood, and it might be influenced by the DNA sequence and histone variant composition at these different families of tandem repeats.

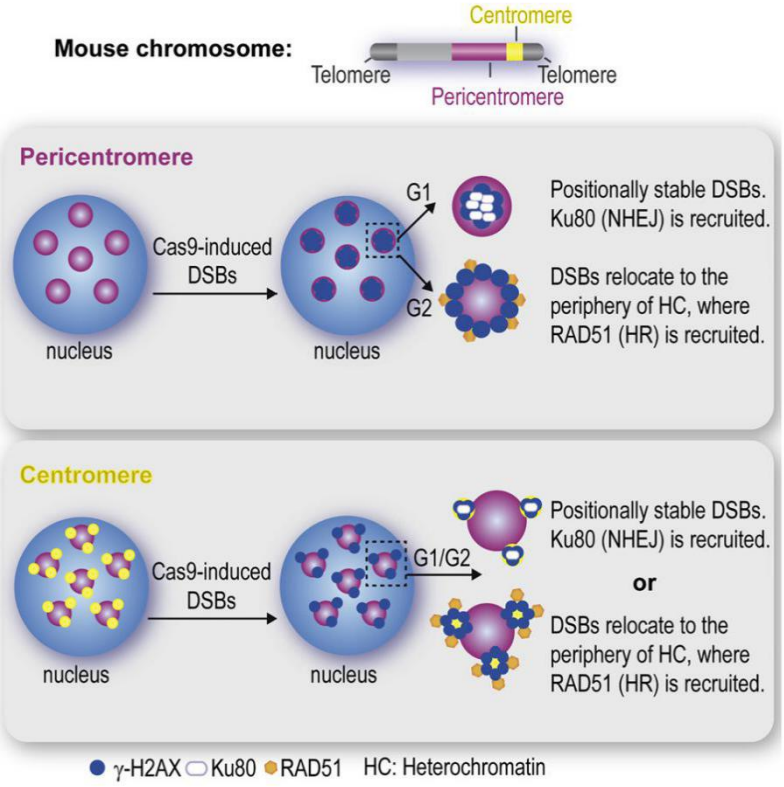


Figure 15. Double strand breaks on mouse (peri)centromeres. DSBs that occure at pericentric heterochromatin in G1 can activate NHEJ. In S/G2, DSBs activate NHEJ or HR. Centromeric DSBs activate NHEJ and HR, independently of the cell cycle.
From Tsouroula *et al.*. 2016

Centromeres have unique chromatin features that are essential for their maintenance (McKinley and Cheeseman, 2016), which include the histone H3 variant, Cenp-A. Depletion of this histone variant in S phase increases the fragility of centromeres and makes them more

susceptible to DNA damage and rearrangements (Giunta *et al.*, 2021). However, there are conflicting reports about Cenp-A localization in response to DNA damage. A study in *Drosophila* has shown that CID (*Drosophila* Cenp-A homolog) does not accumulate at damage sites, possibly to prevent the formation of an ectopic kinetochore and the induction of genomic instability (Mathew *et al.*, 2014). In murine cells, global genotoxic stress leads to Cenp-A eviction from centromeric repeats in a DDR-dependent manner (Hédouin *et al.*, 2017). Conversely, CENP-A has been observed to be rapidly recruited to laser irradiation-induced break sites in human cells (Zeitlin *et al.*, 2009). Along these lines, a recent study has shown that DSBs targeted at centromeres in murine cells promote the *de novo* deposition of Cenp-A at centromeric repeats (Yilmaz *et al.*, 2021). These discrepancies might be due to the nature of the DNA damage or the phase of the cell-cycle during which the lesion occurs. For example, in the study cited above, Cenp-A is deposited at centromeric repeats upon targeted DSBs in G1 (Yilmaz *et al.*, 2021), phase of the cell cycle when Cenp-A is deposited in physiological conditions (Hemmerich *et al.*, 2008; Jansen *et al.*, 2007). While many unknowns still remain on how centromere integrity is preserved in response to DNA damage, a better understanding of these mechanisms is of particular interest, given the essential role of centromeres in maintaining genome stability.

4. Cell cycle control

In physiological conditions, cells go through a series of events known as the cell cycle, during which cells grow and potentially divide in two daughter cells. In typical eukaryotic cells, the cell cycle is divided into four phases: the gap 1 (G1) phase when cells grow, after which cells may enter the synthesis (S) phase, where the nuclear DNA is replicated. In the succeeding gap 2 (G2) phase, cells continue to grow. Finally, in the mitosis (M) phase, sister chromatids are separated and distributed to the two daughter cells (Harashima *et al.*, 2013) (**Figure 16**). The succession of the different phases of the cell cycle is regulated by Cyclin Dependent Kinase (CDK)/Cyclin complexes (Malumbres, 2014). CDK activity is regulated at several levels: by cyclins whose levels vary during the cell cycle, by protein kinases such as Wee1 or CAK (Cdk Activating Kinase), by phosphatases such as Cdc25 and by inhibitory proteins such as $p21^{Cip1}$, $p16^{Ink4a}$ and $p27^{Kip1}$. The correct progression through the cell cycle is controlled at distinct checkpoints. The major checkpoints are at entry into S phase (G1–S checkpoint), mitosis (G2–M checkpoint) and the spindle checkpoint that controls progression into

anaphase. The activation of these checkpoints mainly leads to the inhibition of CDK/Cyclin complexes (Malumbres, 2014). This blocks the cell cycle progression as long as the genome is not repaired, thus avoiding jeopardizing genome integrity.

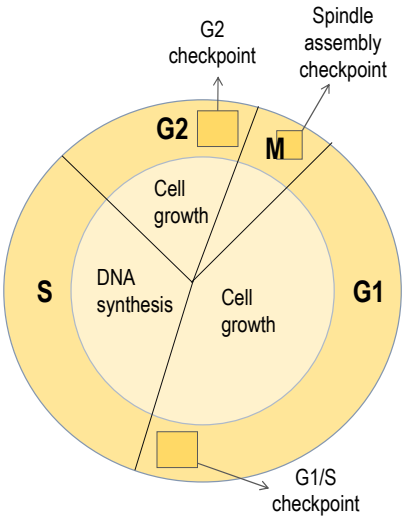


Figure 16. The mammalian cell cycle. The cell cycle progresses orderly through G1 (growth), S (DNA replication), G2 (growth) and M (cell division) phases. Three major checkpoint survey the health of the cell and its genome: the G1/S restriction point, the G2 checkpoint and the spindle assembly checkpoint (SAC) in mitosis. Modified from Harper and Brooks, 2005.

Spindle assembly checkpoint

Correct segregation of chromosomes is essential for the correct distribution of genetic material to both daughter cells during cell division. In methaphase, each sister chromatid needs to be correctly attached to the spindle microtubules (*via* the kinetochore, described in the next section) and positioned on the metaphase plate. A crucial surveillance mechanism controls this event: the Spindle Assembly Checkpoint (SAC), which allows for the continuation of anaphase and the separation of sister chromatids by contraction of mitotic microtubules (Przewloka and Glover, 2009). Correct segregation of chromosomes is therefore intimately linked to the maintenance of genomic integrity and chromosomal stability (Lara-Gonzalez *et al.*, 2021). Indeed, defects in the spindle checkpoint seem to promote aneuploidy and tumorigenesis (Wassmann and Benezra, 2001) or human genetic syndromes such as trisomies (Mikwar *et al.*, 2020).

Following DNA replication in S phase, the resulting duplicated DNA strands are held together by Cohesins, multimeric protein ring structures that encircle the replicated sister chromatids (Nasmyth, 2002). Cohesion of sister chromatids is maintained through the rest of S phase, G2 and into early mitosis. At this point the chromosomes align at the cell equator, through the

attachment to the microtubule spindle apparatus. This occurs in a stochastic manner, where elongating and contracting microtubules, emanating from the two spindle poles, capture the chromosomes (Cleveland *et al.*, 2003). Due to the random nature of this process, it is crucial that chromosome segregation occurs only after all of the sister chromatids are correctly attached to the spindle microtubules. The SAC has a crucial role in the surveillance of this process, by sensing if there is proper attachment and/ or tension at kinetochores of each sister chromatid (Millband *et al.*, 2002; Musacchio and Hardwick, 2002; Yu, 2002). In the presence of unattached kinetochores, the SAC is "ON" and anaphase is inhibited. Only when all kinetochores are stably bound to microtubules, the cell cycle can continue.

During normal mitoses, the sister chromatids are separated at the onset of anaphase. This is achieved by proteolysis of one of the Cohesin subunits, Scc1, by a protease called Separase (Nasmyth, 2002). Before this step, the Separase is kept inactive by its inhibitor, Securin (Nasmyth, 2002; Nasmyth *et al.*, 2000; Zou *et al.*, 1999). After proper microtubule attachment of sister chromatids, Securin is ubiquitinated by the anaphase-promoting complex or cyclosome (APC/C). The APC/C is a multi-subunit E3 ubiquitin ligase that requires activation by its co-factor, cell division cycle protein 20 (Cdc20), to target several proteins for proteolytic degradation, including mitotic cyclins (Au *et al.*, 2002; Peters, 2002; Yu, 2002). The ubiquitination of Securin results in its degradation, leading to Separase activation and Scc1 cleavage. Once chromatid cohesion is lost, sister chromatids separate, allowing spindle forces to pull them to opposite sides of the cell (**Figure 17, left**).

In a scenario where chromosomes are not properly attached to the spindle microtubules, the APC/C is inhibited and progression to anaphase is halted. This is achieved by the mitotic checkpoint complex (MCC) composed of BubR1, Bub3 and Mad2 (Manic *et al.*, 2017). The inhibition of APC/C results in the stabilization of Securin and preservation of sister chromatid cohesion. This in turn precludes the separation of sister chromatids (**Figure 17, right**).

The SAC is instrumental to avoid the formation of daughter cells with an abnormal number of chromosomes, an event called aneuploidy. What is then the trigger for the formation and activity of the MCC, and subsequent safeguard mechanism of the SAC? Research over the last decade has uncovered that the SAC signal that blocks anaphase onset is generated at kinetochores.

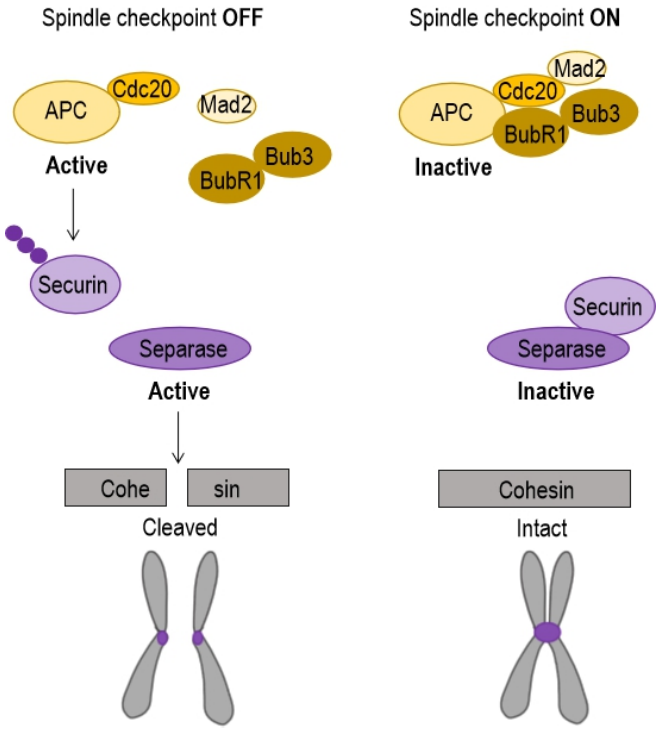


Figure 17. Overview of the spindle assembly (SAC) checkpoint pathway. At the metaphase–anaphase transition, APC/C-Cdc20 ubiquitinates Securin. Degradation of Securin activates Separase, which in turn cleaves the cohesin complex, allowing chromosome segregation (left).
If sister chromatids are not properly attached to the mitotic spindle, the SAC promotes the assembly of checkpoint protein complexes that inhibit the activity of APC/C. This leads to the stabilization of Securin, maintainance of sister chromatid cohesion, and finally results in the delay of anaphase onset. Modified from Bharadwaj and Yu, 2004.

The kinetochore:

The kinetochore is a multiprotein complex, comprising at least 80 different proteins (Kixmoeller *et al.*, 2020), which assembles only once on the centromere of each sister chromatid, and directs and controls the attachment of each chromosome to the mitotic spindle. The multiprotein complex of kinetochore assembles at centromeres in layers (**Figure 18**). First, the centromeric determinant CENP-A recruits the constitutive centromere-associated network (CCAN), composed of 17 proteins which occupy this region throughout the cell cycle (Westhorpe and Straight, 2013). As cells enter mitosis, the CCAN recruits the KMN network, consisting of the KNL1 (Kinetochore Scaffold 1), Mis12 and Ndc80 sub-complexes, which form the outer kinetochore (Varma and Salmon, 2012). Of note, the Ndc80 complex is essential for microtubule attachment (Ciferri *et al.*, 2008). Other kinetochore proteins regulating microtubule attachment include the RZZ complex, comprising Rod, Zw10 and Zwilch (named after the *Drosophila melanogaster* genes *Rough Deal, Zwilch, and Zeste White 10*), which recruits the dynein/dynactin minus-end motor complex (Barisic and Geley, 2011). Although not part of the kinetochore *per se*, the chromosome passenger complex (CPC) is also essential for the regulation of the SAC and the bi-orientation of chromosomes. Notably, the CPC is recruited to centromeres at the onset of mitosis and it is composed of the kinase

Aurora B, and regulatory subunits INCENP, Borealin and Survivin (Ruchaud *et al.*, 2007). During prophase, Aurora B is located on the chromosome arms, while during prometaphase and metaphase it migrates to (peri)centromeric regions (Crosio *et al.*, 2002), where it is responsible for phosphorylation of histone H3 on Serine 10 during mitosis, which in turn is important for mitotic chromosome condensation (Li *et al.*, 2017).

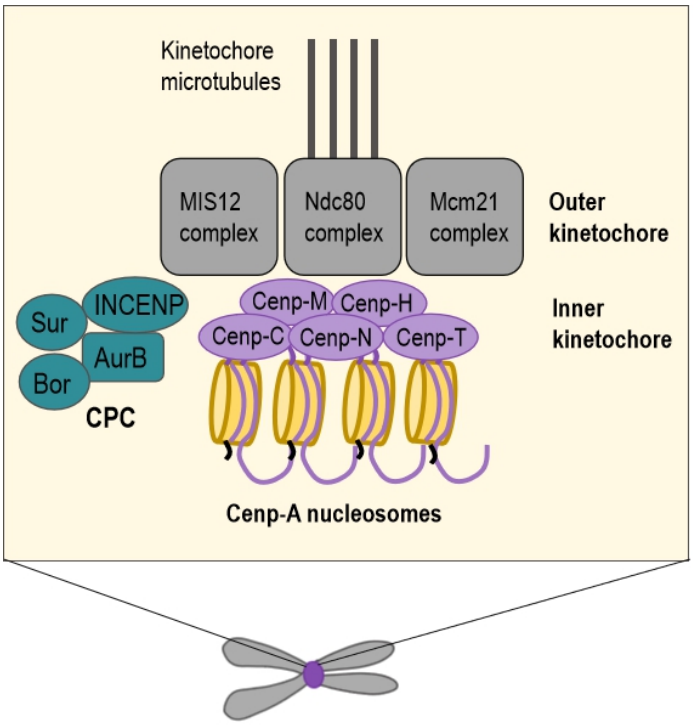


Figure 18. The kinetochore. Diagram showing the organization of the centromere region and the kinetochore components in mitosis. Yellow circles: Cenp-A nucleosomes. Purple circles: proteins of the inner kinetochore. Teal figures: CPC components. Gray boxes: complexes of the outer kinetochore. Gray lines: kinetochore microtubules. Modified from Musacchio and Salmon, 2007.

So how do unattached kinetochores generate the "ON" signal that activates the SAC? When unattached, kinetochores catalyze the formation of the MCC. This event goes through the RZZ and the Ndc80 complexes, which recruit the MCC components to unattached kinetochores, favoring the formation and activation of the MCC (Lara-Gonzalez *et al.*, 2021). One fascinating aspect is that the strength of the diffusible SAC signal is proportional to the number of unattached kinetochores present, indicating that the SAC signaling is modulated (Lara-Gonzalez *et al.*, 2021; Rieder *et al.*, 1995). Once kinetochores are securely attached to the microtubule, the SAC signal is rapidly silenced to promote the onset of anaphase and mitotic exit. In this scenario, the microtubule attachments trigger the poleward transport of the MCC components. In addition, kinetochore-based phosphatases oppose mitotic kinases that activate the SAC.

Given the essential role of the kinetochore for correct segregation of genetic material in daughter cells, it is not surprising that dysregulation of the kinetochore components is a signature of cancer cells. Analyses of numerous cancer databases has shown that overexpression of kinetochore genes correlates with increased levels of genomic instability (Zhang *et al.*, 2016), by generating dicentric or lagging chromosomes and other segregation errors (Thiru *et al.*, 2014).

DNA damage checkpoint

Cell cycle checkpoints are activated in response to different threats to cell viability and genome integrity, one of which is DNA damage. In response to DNA damage, the ATM (ataxia telangiectasia mutated) and ATR (Ataxia telangiectasia and Rad3-related) kinases phosphorylate and activate Checkpoint kinases (Chks) 1 and 2 (Chk1 and Chk2), which in turn phosphorylate and inactivate the Cell division control (Cdc) phosphatases Cdc25a and Cdc25c (Mailand *et al.*, 2000). As a result, CDK1 and CDK2 remain inactive and the cell cycle is arrested in G1/S and G2/M. Another key component of this response is the phosphorylation of p53 protein by ATM/ATR and Chk1/Chk2. Phosphorylation of p53 is the key event that leads to p53 stabilization and accumulation in the nucleus since it inhibits its degradation by the E3 ubiquitin ligase MDM2 (Mouse double minute 2 homolog). p53 regulates the expression of many genes involved in DNA repair, cell cycle arrest, apoptosis and senescence (Zilfou and Lowe, 2009). Moreover, it represses the transcription of several key factors for cell cycle progression, such as CDK1, Cdc25c and Cyclin B1 (Innocente *et al.*, 1999; St Clair and Manfredi, 2006). One of the main targets of p53 is the gene encoding $p21^{Cip1}$, a CDK inhibitor. The interaction of $p21^{Cip1}$ with different CDKs such as CDK1 and CDK2 leads to their inhibition and allows cycle arrest at any checkpoint (Gartel and Radhakrishnan, 2005). In addition, p53 induces the expression of proteins of the Growth Arrest And DNA Damage Inducible (GADD45) family that are involved in many functions in response to DNA damage, including cell cycle arrest. For example, GADD45 enables cycle arrest in G2/M by inhibiting the interaction of CDK1 with cyclin B1 (Wang *et al.*, 1999) (**Figure 19**).

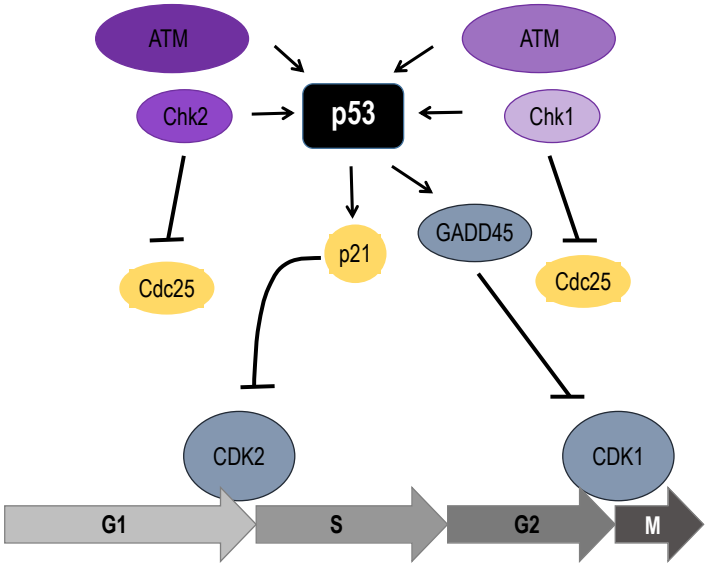


Figure 19. DNA damage checkpoint. ATM and ATR orchestrate transient cell cycle arrest in response to DNA damage.
Rapid phosphorylation of cdc25a and Wee1 enables rapid activation of the G1/S and G2/M checkpoint. Activation of p53 and its associated pathways contribute to checkpoint maintenance at later times. Adapted from Joaquin,Fernandez-Capetillo, 2012.

If transient arrest is insufficient to repair all the damages present, the DNA damage response may be chronically activated and cells may engage in irreversible cell-cycle arrest called senescence (Hayflick, 1965; Hayflick and Moorhead, 1961). Senescent cells continue to be metabolically active but they no longer respond to mitogenic signals (Schmeer *et al.*, 2019). Senescence can be induced by many stimuli including persistent DNA damage, telomere shortening, oxidative stress and oncogene activation. Induction of the $p16^{Ink4a}$/pRB pathway is crucial for entry into senescence. Like $p21^{Cip1}$, $p16^{Ink4a}$ is also a CDK inhibitor. It inhibits the CDK4/6 complex which becomes unable to phosphorylate the Retinoblastoma protein (RB). Hypophosphorylated RB remains associated with the transcription factor E2F1 and is sequestered in the cytoplasm where it cannot activate the transcription of its target genes that are essential for the G1/S transition (Rayess *et al.*, 2012). The expression of $p16^{Ink4a}$ is induced in response to various stresses including DNA damage (Ohtani *et al.*, 2001; Potapova *et al.*, 2001; Rayess *et al.*, 2012).

If the DNA damage is too great to be repaired, the cell may engage in programmed cell death known as apoptosis. The p53 protein also plays a key role in the activation of this process. It activates the transcription of pro-apoptotic genes such as Bax, Bid and PUMA and represses that of anti-apoptotic genes such as Survivin (Miyashita *et al.*, 1995; Nakano and Vousden, 2001).

5. R-loop: friend or foe? R-loops in physiopathology

An endogenous threat to genome integrity is represented by R-loops (Aguilera and García-Muse, 2012; Wahba *et al.*, 2013). First described in 1976, R-loops are non-B DNA structures that form when a nascent RNA molecule base-pairs with the template DNA strand, forming a DNA-RNA hybrid. This structure leaves a single-stranded DNA (ssDNA) exposed (Thomas *et al.*, 1976a). The term DNA-RNA hybrid refers only to the double-stranded structure, while R-loops refer to the three-stranded structure formed by the DNA-RNA hybrid and the ssDNA (Kim and Wang, 2021) (**Figure 20**).

There are two types of R-loops: regulated R-loops that serve important physiological roles, and unscheduled or excessive R-loops which interfere with crucial cellular functions (García-Muse and Aguilera, 2019a; Uruci *et al.*, 2021).

To better understand their implications in genome instability, I will first describe how R-loops are formed and what are their physiological roles, followed by what happens when R-loops go rogue.

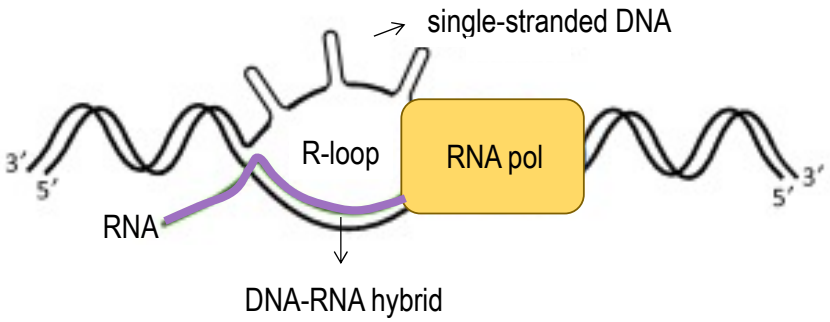


Figure 20. R-loops are three stranded structures composed of a DNA-RNA hybrid and a single-stranded DNA. Generally they form co-transcriptionally, when the nascent RNA (purple) hybridizes with the template DNA. Adapted from Sollier and Cimprich, 2015.

How and where: an overview of R-loop formation

R-loops predominantly form co-transcriptionally behind elongating RNA polymerases (RNA pol) (García-Pichardo *et al.*, 2017; Li and Manley, 2005; Yu *et al.*, 2003). The current model for R-loop formation is the "thread-back", where the negative supercoiling that occurs behind the RNA pol provides an ideal opportunity for the nascent RNA to anneal with the DNA as soon as it exits the RNA Pol (Westover *et al.*, 2004). This suggests that they generally form *in*

cis (Aguilera and García-Muse, 2012), however they have also been observed *in trans* (Wahba *et al.*, 2013; Zaitsev and Kowalczykowski, 2000). The precise mechanism of R-loop formation *in trans* is not fully understood, but it would seem that, at least in *S. cerevisiae,* it depends on strand exchange mediated by Rad51, a protein involved in DSB repair by HR (Wahba *et al.*, 2013).

R-loops have been observed predominantly at highly transcribed genes such as rRNA and tRNA loci (El Hage *et al.*, 2010), but also in the mitochondrial genome (Xu and Clayton, 1996), genomic regions where an antisense transcript exists, centromeres and telomeres (Kabeche *et al.*, 2018a; Toubiana *et al.*, 2018; Wahba *et al.*, 2016). R-loop levels correlate with increased transcription in physiological conditions (Chan *et al.*, 2012a; Stork *et al.*, 2016; Wahba *et al.*, 2016). However, it is important to note that high transcription levels do not necessarily promote the formation of DNA-RNA hybrids (Hraiky *et al.*, 2000; Huertas and Aguilera, 2003), since the nascent RNA is usually coated by RNA binding proteins and further processed ((Li and Manley, 2005).

The average size of R-loops ranges from less than 100 bp up to 2 kb (García-Pichardo *et al.*, 2017; Richard and Manley, 2017; Yu *et al.*, 2003). Several factors favor the formation of R-loops, such as i) the DNA sequence, ii) DNA topology and iii) protein factors. In terms of DNA sequence, R-loop hotspot regions are generally characterized by high GC content or G-quadruplex (G4)-containing sequences (Ginno *et al.*, 2012; Sanz *et al.*, 2016; Stork *et al.*, 2016). Of note, this observation might reflect a technical bias in the detection of more thermo-stable GC-rich DNA-RNA hybrids. However, genome topology can increase the propensity of R-loop formation independently of GC content. This is achieved by the negative supercoiling that occurs behind an advancing RNA Pol (Roy *et al.*, 2010). Moreover, the presence of ssDNA nicks can favor R-loop formation, whereby the free DNA end facilitates the intertwining of DNA and RNA molecules (Roy *et al.*, 2010). Lastly, DNA-binding proteins can also stabilize R-loops by binding the displaced ssDNA. A known example is the case of the homeodomain protein AtNDX which regulates COOLAIR, a set of antisense RNAs originating from the *Arabidopsis* FLOWERING LOCUS C (FLC). AtNDX binds the ssDNA of the R-loop in the COOLAIR promoter, which in turn regulates FLC expression (Sun *et al.*, 2013).

Rules of the R-loop game: physiological R-loop functions

R-loops can occur genome-wide (Xu *et al.*, 2017), and decorate up to 5% of the human genome (Sanz *et al.*, 2016). They have been detected in all four kingdoms of life (Zeng *et al.*, 2021), which suggests that these structures are critically involved in important cellular functions. Both stable R-loops (the three-stranded structures that feature displaced ssDNA) and transient DNA-RNA hybrids have been documented to occur in well-known physiological cellular processes. For instance, DNA-RNA hybrids are formed during a very known nuclear process: the synthesis of the lagging DNA strand during DNA replication. In this process, DNA polymerase α synthesizes RNA primers that are then used by DNA polymerase δ to produce mature Okazaki fragments (Stodola and Burgers, 2017; Sugino *et al.*, 1972). DNA-RNA hybrids form also during transcription, where an 8 bp DNA-RNA hybrid forms in the transcription bubble (Westover *et al.*, 2004). Telomere homeostasis is another example where a nuclear function relies on DNA-RNA hybrids. The reverse transcriptase Telomerase specifically recognizes chromosome ends through DNA-RNA hybrid formation. The Telomerase utilizes the telomere repeat-containing RNA (TERRA) that hybridizes to the 3' ssDNA at chromosome ends and enables telomere extension (Greider and Blackburn, 1989). Even the CRISPR-Cas9 activity (which I will elaborate more on in Chapter 6) relies on the formation of DNA-RNA hybrids formed when the guide RNA transiently binds the target DNA, in order to identify the correct sequence for Cas9-mediated cleavage (Jinek *et al.*, 2012).

In contrast to transient DNA-RNA hybrids, stable R-loops appear to be rarer event. Nonetheless, they have been reported to play a role in several controlled cellular processes: from the replication of *E.coli* plasmid (Itoh and Tomizawa, 1980), to the initiation of mitochondrial DNA replication (Aguilera and García-Muse, 2012; Crossley *et al.*, 2019a), and Immunoglobulin class switch recombination in vertebrate B lymphocytes (Yu *et al.*, 2003).

Recently, R-loops have been gaining more recognition as intermediates in regulating gene expression, suggested by the fact that these structures are enriched at gene promoters and termination sites (Aguilera and García-Muse, 2012). DNA-RNA hybrids have been shown to display an intermediate conformation between that of dsRNA (A form) and that of dsDNA (B form) (Shaw and Arya, 2008). Hence, these distinct structures might potentially be recognized by different epigenetic or transcription factors, that then influence gene expression. Interestingly, there seems to be a functional link between R-loop formation at gene promoters, and DNA hypomethylation of these loci. Indeed, R-loops have been reported at the 5' UTR

regions immediately downstream of certain non-methylated promoters in humans (Ginno *et al.*, 2012). For instance, this is the case of the R-loop at the promoter of the gene TCF21 (Transcription Factor 21), which is formed by the antisense lncRNA TARID (TCF21 Antisense RNA Inducing Promoter Demethylation). This R-loop is then recognized and bound by the stress response protein GADD45A, which in turn recruits TET1 that mediates promoter demethylation, inducing TCF21 transcription (Arab *et al.*, 2019). Otherwise, the presence of R-loops at gene promoters might protect these regions against DNA methylation, since DNMT1 has a lower affinity for DNA-RNA hybrids compared to dsDNA (Grunseich *et al.*, 2018).

Besides being enriched at promoter of certain genes, R-loops are found also at several G-rich terminator elements (Sanz *et al.*, 2016), which was suggested to promote RNA Pol II pausing before efficient termination (Skourti-Stathaki *et al.*, 2011). Another example of R-loops termination-promoting activity is the newly unveiled crosstalk between post-transcriptional RNA modifications and R-loops (Kim and Wang, 2021). The methylation of adenosine in RNA molecules (N^6-Methyladenosine, m6A) has been suggested to stabilize R-loops, facilitating transcriptional termination (Abakir *et al.*, 2020; Yang *et al.*, 2019). The altered DNA conformation created by the R-loop at the 3' end of a gene might be unfavorable for RNA Pol II progression, leading to transcriptional termination (Belotserkovskii *et al.*, 2013; Kireeva *et al.*, 2000) (**Figure 21**).

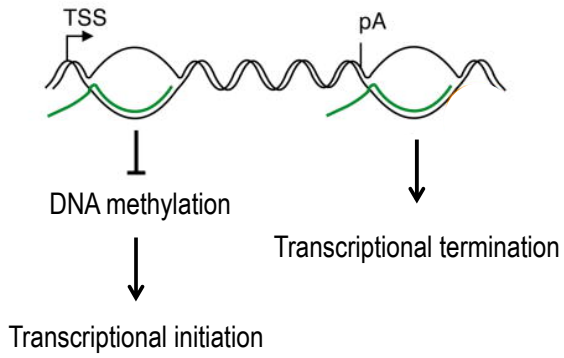


Figure 21. R-loops influence both transcriptional initiation and termination. R-loops located at promoters prevent DNA methylation, which correlates with transcriptional initiation. Conversely, R-loops located at termination regions promote transcriptional termination.
Adapted from Sollier and Cimprich, 2015

Apart from being involved in gene expression programs, R-loops also seem to be indispensable for maintaining genome integrity during mitosis. They do so by participating in the correct function of one of the most crucial domains of the mitotic chromosome - the centromere.

R-loops and their physiological roles at centromeres

A recent intriguing discovery has shown that mitosis-specific R-loops act at centromeres to promote faithful chromosome segregation (Kabeche *et al.*, 2018a; Moran *et al.*, 2021a). Mitotic R-loops at centromeres serve as a signaling hub for a non-canonical ATR pathway that is independent of DNA damage. The ssDNA present in the R-loop is coated by RPA, which is required for ATR activation at centromeres during mitosis. Centromere-localized ATR in turn promotes full activation of Aurora B, ensuring accurate chromosome segregation (Kabeche *et al.*, 2018a). Similarly, the DHX9 RNA helicase has been implicated in the formation of physiological centromeric R-loops, probably through unwinding RNA secondary structures (Chakraborty *et al.*, 2018).

Outside of mitosis, the appearance or physiological role of centromeric R-loops is less clear. In G1, centromeric R-loops safeguard the stability of the locus upon targeted double-strand breaks. In this context, low levels of centromeric transcription lead to R-loop formation, which in turn promotes DNA-end resection and the licensing of HR, preventing the activation of mutagenic DNA repair pathways (Yilmaz *et al.*, 2021). The different studies therefore point to the unique beneficial roles of centromeric R-loops in different stages of the cell cycle, in physiological conditions or in response to DNA damage.

When R-loops don’t play nice: R-loops as a source of genome instability

R-loops are constantly formed and resolved all along the transcribed genome, resulting in about 27,000 R-loops per day, out of which, only 300 are stable and maintained (Crossley *et al.*, 2019). Given the potentially deleterious consequences of unscheduled R-loops (R-loops that form at the wrong time or at the wrong genomic locus), cells developed different factors and mechanisms to resolve or prevent their formation. In eukaryotic cells, DNA-RNA hybrids are prevented by coating the nascent RNA molecule with proteins involved in processing and export (Li and Manley, 2005). When cells fail to prevent R-loop formation, or when physiological R-loops need to be resolved, cells employ factors to remove them. The most known and best characterized such factors are RNase H enzymes, which specifically cleave the RNA moiety of the DNA-RNA hybrid (Hyjek *et al.*, 2019a). The second type of factors that remove DNA-RNA hybrids are helicases with RNA-dependent DNA-RNA unwinding activity. These factors include Senataxin (SETX), Fanconi anemia complementation group M

(FANCM), DEAD-Box Helicase 5 (DDX5) and Bloom syndrome protein (BLM) among others (Chang and Stirling, 2017; Cristini *et al.*, 2018; Mischo *et al.*, 2011; Ribeiro de Almeida *et al.*, 2018; Skourti-Stathaki *et al.*, 2011). In addition, BRCA1 has been shown to antagonize R-loop accumulation at centromeres, most likely by recruiting its known partner SETX (Racca *et al.*, 2021a).

When cells are unable to efficiently prevent or remove unscheduled R-loops, this threatens genome integrity (Aguilera and García-Muse, 2012; Giunta *et al.*, 2021; Huertas and Aguilera, 2003). R-loops are more susceptible to the mutagenic action of DNA-modifying enzymes such as the activation-induced cytidine deaminase (AID) that converts cytidine to uracil (Chaudhuri *et al.*, 2004; Gómez-González and Aguilera, 2007; Petersen-Mahrt *et al.*, 2002), therefore introducing mutations in the genome (**Figure 22.A**). Moreover, ssDNA might be more vulnerable to the action of nucleases (Aguilera and García-Muse, 2012) and to DNA damage (Costantino and Koshland, 2018) **(Figure 22.B).** In agreement, centromeric R-loops accumulation correlates with increased DNA damage in BRCA1 deficient cells, which in turn favors crossover recombination between α-SAT tandem repeats and thereby centromeric instability (Racca *et al.*, 2021).

However, genome stability is mostly threatened by R-loops formation in S-G2 due to transcription-replication conflicts, which lead to replication fork collapse and DNA breaks (Gan *et al.*, 2011; Wellinger *et al.*, 2006) **(Figure 22.C).** In this scenario, the R-loop is a collateral event, which might fragilize further the locus. This is the case of centromeric R-loops that form in S phase (Giunta *et al.*, 2021). As opposed to centromeric R-loops in mitosis or G1 mentioned in the previous paragraph, those formed in S phase are toxic to the cell (Giunta *et al.*, 2021). Upon depletion of Cenp-A, centromeres are transcribed during DNA replication, which in turn leads to collision of the transcription and replication forks, and ultimately poses a threat to centromere integrity (Giunta *et al.*, 2021).

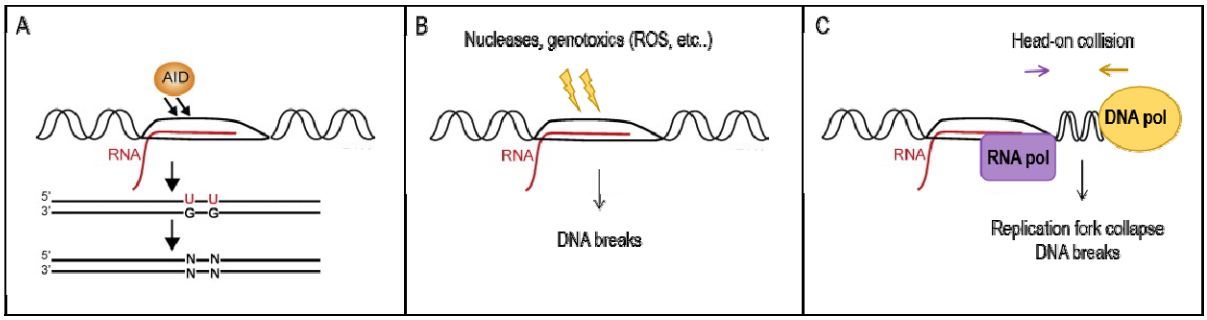


Figure 22. R-loops as threats to genome integrity. R-loops can lead to mutagenesis by exposing the displaced ssDNA to **A**. AID, or **B**. to endogenous genotoxic metabolites. **C.** In S phase, the most likely source of DNA damage are transcription-replication conflicts. This can lead to replication fork collapse, leaving an R-loop upstream. Adapted from Aguilera and Garcia-Muse, 2012.

Expectedly, aberrant R-loops have been documented in a number of syndromes, cancer cells and autoimmune diseases (Perego *et al.*, 2019; Richard and Manley, 2017; Sarkar *et al.*, 2018). This is not surprising since aberrant R-loops might lead to genome instability and translocations, hallmarks of cancer cells (Halazonetis *et al.*, 2008). Indeed, human breast cancer cell lines accumulate R-loops at estrogen-induced genes, the latter being also subjected to translocations (Stork *et al.*, 2016). As mentioned above, BRCA1 tumor-suppressors and Fanconi anemia factors are known players in R-loop prevention. Carriers of BRCA1 mutation showcase excessive R-loops formation in breast luminal epithelial cells (Zhang *et al.*, 2017). It has been proposed that mutations in BRCA1 lead to R-loop accumulation at the promoter of the ESR1 locus, which encodes estrogen receptor α (ERα). This in turn leads to silencing of neighbouring tumor supressor genes, and these events seem to be directly implicated in tumorigenesis (Chiang *et al.*, 2019; Zhang *et al.*, 2017).

Another pathological example is the ICF syndrome, which highlights the interplay between DNA hypomethylation at tandem repeats, increased transcription and R-loop formation. Indeed, cells from ICF2/3/4 patients show increased transcription of the hypomethylated (peri)centromeric repeats, resulting in R-loop formation (Unoki *et al.*, 2020). Moreover, ICF1 patients with DNA hypomethylation in subtelomeric regions, feature high levels of TERRA and excessive telomeric R-loop formation. This is accompanied by elevated DNA damage at this locus that mediates the telomere shortening observed in the ICF syndrome (Sagie *et al.*, 2017).

Finally, mutations in RNase H2 are associated with the rare Aicardi-Goutières Syndrome (AGS) (Crow *et al.*, 2006), characterized by an abnormal innate immune response. Cells of AGS patients feature R-loops that accumulate along the genome, which correlates with a

genome-wide loss of DNA methylation (Lim *et al.*, 2015). However, whether unscheduled or excessive R-loops are causal to disease, or are just a collateral consequence of it, has not fully been explored.

Now you see me, now you don't: the problem of R-loop detection

R-loops were first visualized in 1977 by electron microscopy, on rDNA in *Drosophila* (White and Hogness, 1977). Since then, many techniques for R-loop profiling have been developed, which include the detection of: i) ssDNA using bisulfite-based footprinting; ii) DNA-RNA hybrids using the monoclonal S9.6 antibody; iii) DNA-RNA hybrids using the catalytically dead RNase H (dRNAse H).

The sodium bisulfite treatment induces the deamination of unmethylated cytosines to uracils in ssDNA. This method has been used to detect DNA-RNA hybrid-dependent ssDNA at unmethylated promoters (Ginno *et al.*, 2012). However, this method might miss the detection of R-loops on methylated stretches of DNA, or AT-rich R-loops such as those found on centromeric repeats.

The most widely used method of R-loop detection is their capture using the monoclonal S9.6 antibody (Boguslawski *et al.*, 1986). This antibody recognizes 10bp long DNA-RNA hybrids, regardless of their sequence. The S9.6 antibody is widely used for the DNA-RNA immunoprecipitation (DRIP) assay, which is followed by a qPCR for specific DNA regions or by high-throughput sequencing (Sanz and Chédin, 2019). In the case of DRIP assays, samples treated with recombinant RNase H1 that serve as a negative control are essential to verify the specificity of the S9.6 immunoprecipitation (Smolka *et al.*, 2021).

The S9.6 antibody also allows for the visualization of R-loops in immunofluorescence, where it has been shown that it predominantly stains nucleolar R-loops signal on the rDNA (Lindström *et al.*, 2018; Shen and Skibbens, 2017). However, the detection of R-loops using this antibody has important limitations. It has been reported that S9.6 binds specifically to dsRNAs in fission yeast (Hartono *et al.*, 2018), as well as both ssRNAs and dsRNAs in human cells (Smolka *et al.*, 2021). Another major issue is the different S9.6 staining patterns in different studies, which depends on technical differences in the pre-extraction, fixation and post-fixation treatment of cells (Nguyen *et al.*, 2017; Silva *et al.*, 2018; Skourti-Stathaki *et al.*, 2011; Sollier *et al.*, 2014). The S9.6 antibody detection of R-loops by immunofluorescence often features an intense cytoplasmic signal, that has been attributed to R-loops in the

mitochondrial genome (El Hage *et al.*, 2010; Shen and Skibbens, 2017). However, the majority of cytoplasmic S9.6 signal localizes outside of mitochondria (Smolka *et al.*, 2021), possibly due to recognition of cytoplasmic RNAs.

Recently, a new immunofluorescence protocol has been published that advises the pre-treatment of samples with recombinant RNase T1 and RNase III, that degrade ssRNA and dsRNA respectively, to control for the specificity of the signal, since the pre-treatment of samples with recombinant RNaseT1 abolishes cytoplasmic and nucleolar staining (Smolka *et al.*, 2021).

An alternative approach to visualize R-loops is the use of dRNase H fused to a fluorescent protein. However, the drawback of this system is that dRNase H has been shown to stabilize R-loops, potentially leading to deleterious effects for the cell (Yilmaz *et al.*, 2021). Moreover, dRNase H-eGFP staining of R-loops has been reported to exhibit high background (Crossley *et al.*, 2019)

Since there are several limitations and technical challenges for the visualization and detection of R-loops, the best strategy seems to be the use of orthogonal approaches and opportune negative controls. Developing of new tools to detect or to perturb R-loops will open doors for a better understanding of their implications in health and disease.

6. Customizing the (epi)genome: (epi)genetic engineering and its tools

The ability to manipulate the (epi)genome in a specific and efficient manner represents a powerful tool for both basic research and development of therapeutic approaches for various diseases. To determine gene function, scientists often rely on gain of function (GOF) experiments which allow to ectopically express the product of the gene of interest using cDNA-containing constructs, or on loss of function (LOF) approaches through targeted knockout or using knockdown by RNA interference (RNAi). While these two complementary strategies have proven useful for their relative ease of use, they both have their limitations. GOF experiments don't represent the full complexity of transcript isoforms and don't allow to recreate the transcriptional regulation of this locus in its chromatin context. Likewise, knockdown by RNAi has its limitations: it is incomplete, varies between experiments and different studies, and often provides only temporary inhibition of gene function.

The advent of (epi)genetic engineering revolutionized biological research and refined the means of manipulating either the DNA sequence per se, or its epigenetic state. Genetic engineering involves gene disruption or replacement, by taking advantage of the DNA repair machinery of the cells This allows scientists to stably introduce exogenous DNA or make small deletions/insertion or Single nucleotide polymorphisms (SNPs), thereby modifying the genome of the cells. In contrast, epigenome engineering doesn't affect the genetic sequence. Rather, it consists of adding, removing or modifying preexisting epigenetic marks, with the goal of changing the chromatin organization and/or transcriptional state of the locus of interest. These tools consist of a DNA binding domain that confers sequence-specificity and an effector domain (nucleases, epigenetic enzymes, fluorescent proteins, etc). Several DNA binding domains have been developed, such as zinc finger (ZF) proteins, transcription activator-like domains (TALEs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) (Gaj *et al.*, 2013).

Zinc finger (ZF) proteins

ZF proteins were first discovered in *X. laevis* (Miller *et al.*, 1985), and subsequently found to be frequent transcription factors in the mammalian genome (Klug, 2010; Pavletich and Pabo, 1991; Urnov *et al.*, 2010). The Cys_2-His_2 zinc-finger domain, the most common ZF type in eukaryotes, consists of 30 amino acids that fold to interact with 3 bp in the major groove of DNA (Beerli and Barbas, 2002). The development of synthetic ZF proteins that recognized 18 bp DNA sequences of choice allowed to target specific sequences in the human genome for the first time (Beerli *et al.*, 1998, 2000). While in theory ZN proteins can be designed to target any sequence (**Figure 23**), their construction and validation are a laborious process.

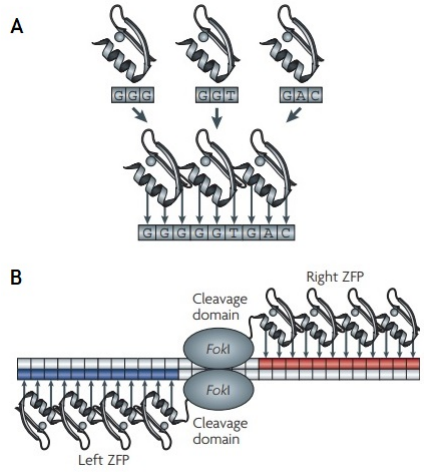


Figure 23. Structure and design of zinc finger proteins. A. Modular assembly of a three-finger protein from individual ZFs to generate a ZF protein with sequence specificity. **B**. Sketch of a zinc finger nuclease dimer bound to its target. Each ZF nuclease contains the cleavage domain of FokI linked to an array of ZFs, designed to recognize specific sequences that flank the cleavage site. Adapted from Urnov *et al.*, 2010.

TALEs

TALEs represent a simpler alternative for engineering programmable DNA-binding proteins. They were first discovered as naturally occurring proteins in the rice pathogenic bacteria Xanthomonas (Boch *et al.*, 2009). TALEs consist of repeated domains, each of which contains a highly conserved sequence of 33-35 amino acids that recognize a single DNA nucleotide within the target site (Moscou *et al.*, 2009) (**Figure 24**). The two amino acids that define the DNA specificity of a TALE are located in position 12 and 13 of each repeat, called Repeat-Variable Di-residues (RVD). Although the design of TALEs is less challenging than that of ZF proteins, both ZFPs and TALEs have their own intrinsic DNA sequence specificity, which makes high-throughput screening long and costly, and potentially requires the construction of several sets of ZF proteins or TALEs.

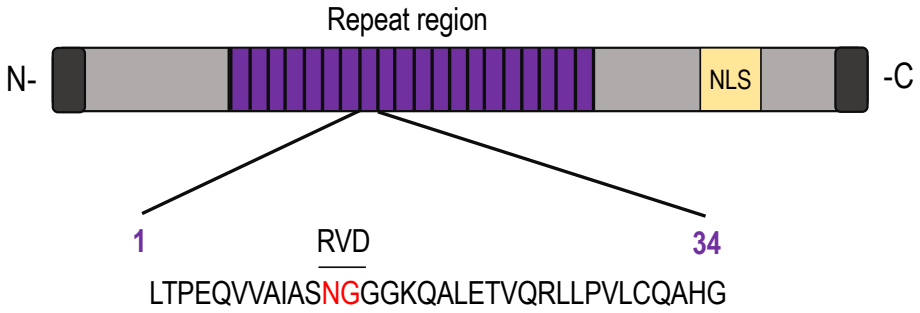


Figure 24. TALE architecture. TALEs harbor a repeat region with a varying number of highly conserved repeats, each of which is composed of 33–35 amino acids placed in tandem. The amino acid sequence of a consensus TALE repeat with 34-aa is shown and the amino acids responsible for the DNA specificity of a TALE, the repeat variable di-residues (RVD), are highlighted.
Modified from Becker and Boch, 2021.

CRISPR/Cas system

In recent years, CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats) / Cas (CRISPR-associated) systems have taken center stage in (epi)genome editing applications (reviewed in detail in Gaj *et al.*, 2013). The CRISPR/Cas system has evolved as an adaptive immune system in bacteria and archea (Grissa *et al.*, 2007; Kunin *et al.*, 2007) and it has first been observed in *E.coli* in 1987 (Ishino *et al.*, 1987). There are three types of CRISPR/Cas systems and they vary in their mode of action and recognized sequences. The type I and III systems employ multiple Cas proteins that form an effector complex (Santiago-Frangos *et al.*, 2021; Spilman *et al.*, 2013), while type II use a single endonuclease, such as Cas9, to locate

and cleave the target DNA (Garneau *et al.*, 2010). In the type II system, short DNA segments from the genome of an invading virus are integrated into a CRISPR locus (Cady *et al.*, 2012; Yosef *et al.*, 2012). They are then transcribed and processed into short CRISPR RNA (crRNAs), which anneal with trans-activating crRNAs (tracrRNAs). This pair of ncRNAs then directs the Cas-mediated sequence-specific cleavage of pathogenic DNA (**Figure 25**).

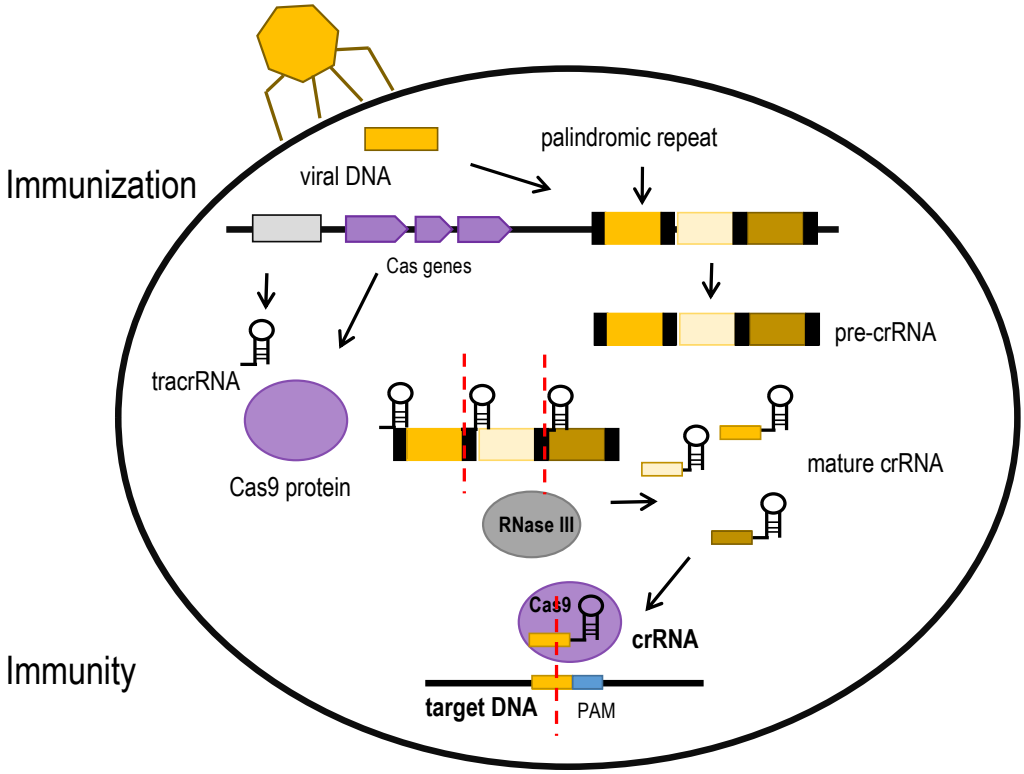


Figure 25. The CRISPR/Cas adaptive immune system. Upon viral infection, short DNA fragments of the virus are integrated into the bacterial genome. These sequences are transcribed to form the crRNA, which combines with the tracrRNA, and is loaded in the Cas9 protein to mediate sequence-specific cleavage. Modified from Hryhorowicz *et al.*, 2017.

Target recognition by the Cas9 protein requires a conserved dinucleotide-containing protospacer adjacent motif (PAM) sequence upstream of the crRNA binding region. The PAM sequence varies between the different CRISPR systems (Mojica *et al.*, 2009). The crRNA and tracrRNA can be combined in a chimeric single guide RNA molecules (sgRNA) (Hsu *et al.*, 2014) (Figure X, top), capable of directing Cas9 to its genomic target (Cong *et al.*, 2013; Mali *et al.*, 2013). More complex systems with added functionality are provided by engineered sgRNAs that include RNA aptamers, secondary RNA structures specifically recognized by RNA-binding proteins (Konermann *et al.*, 2015; Zalatan *et al.*, 2015). Using

aptamers such as MS2 and effectors fused to the corresponding RNA-binding proteins, allows for their recruitment to the dCas9–sgRNA complex (Mali *et al.*, 2013a; Zalatan *et al.*, 2015), forming a tripartite system with synergistic activity (Konermann *et al.*, 2015; Xu *et al.*, 2016) (**Figure 26**).

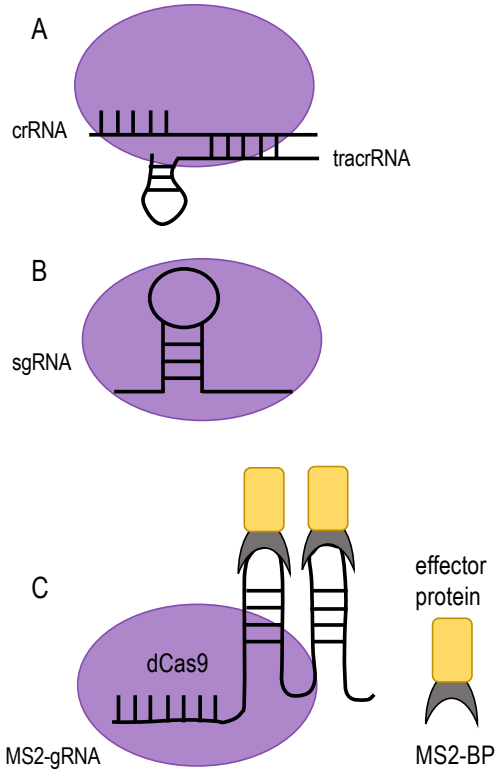


Figure 26. Guide RNA molecules used by Cas proteins.
A. Duplex of tracrRNA and crRNA that hybridize to generate a gRNA.
B. Chimeric single gRNA molecules (sgRNAs)
C. Engineered sgRNAs containing MS2 aptamers will recruit also effector proteins that are fused to the MS2 binding proteins (MS2-BP).

Cas9 is a multidomain DNA endonuclease (Jiang and Doudna, 2015), that contains two distinct nuclease domains responsible for dsDNA cleavage: the HNH domain that cuts the target DNA strand, and the RuvC-like domain which cuts the non-target DNA strand (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). In recent years, mutations in one or both nuclease domains have converted the Cas9 nuclease into a nickase that introduced single-strand cut (Cong *et al.*, 2013) or into a catalytically dead Cas9 (dCas9) (Jinek *et al.*, 2012), which has paved the way for epigenome editing and the manipulation of gene transcription. This versatile tool can be fused to different proteins that enable genome imaging, gene regulation, and epigenetic modification (**Figure 27**).

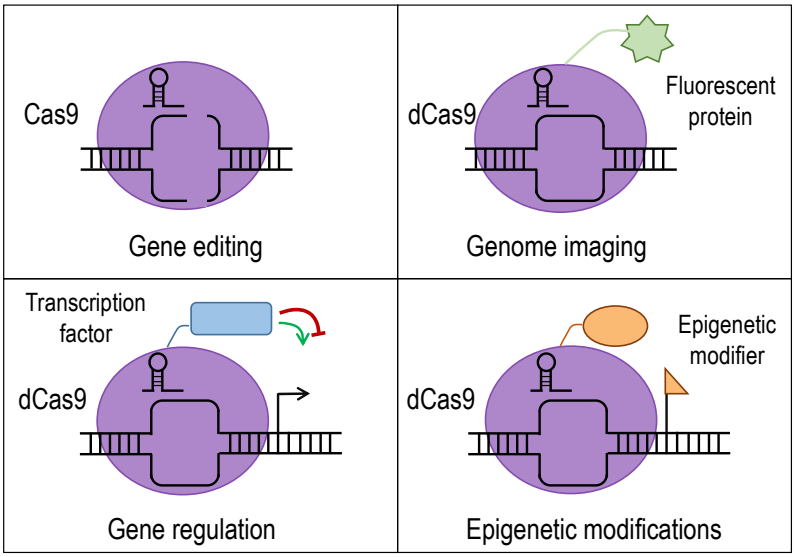


Figure 27.
Schematic representation of (epi)genome editing tools.
Cas9 is used to edit DNA by disrupting genes, creating deletions or to insert specific sequences. dCas9 fused to a fluorescent protein can be used for genome imaging. Fusing dCas9 to transcription factors or epigenetic modifiers allows to modulate transcription of targeted genes.

Transcriptional inhibition (CRISPRi) is achieved by fusing dCas9 with transcriptional repressors such as Krüppel-associated box KRAB, a common repression domain found in numerous mammalian transcription factors (Margolin *et al.*, 1994). KRAB blocks transcription by inducing heterochromatinization of the locus, by attracting a variety of epigenetic modifiers and chromatin remodeling proteins (Groner *et al.*, 2010).

Conversely, transcriptional activation (CRISPRa) is achieved by fusing dCas9 with transcriptional activators such the acidic domain of herpes simplex viral protein 16 (VP16) (Beerli *et al.*, 1998). Increasing the number of VP16 repeat domains, such as dCas9-VP96, VP64-dCas9-VP64, dCas9-VP160 and dCas9-VP192, have been used to efficiently upregulate the expression of endogenous genes (Balboa *et al.*, 2015; Chakraborty *et al.*, 2018; Kearns *et al.*, 2014).

For efficient gene regulation, dCas9 can be fused to additional motifs that recruit several copies of the effector proteins. This is the case of the SunTag, a protein scaffold with repeating peptide epitope arrays, that can recruit antibody-activator fusion proteins (Tanenbaum *et al.*, 2014). Moreover, efficient gene regulation can be achieved by using multiple sgRNAs tiling upstream or downstream of the promoter region, in the case of CRISPRa or CRISPRi respectively (Doench *et al.*, 2014; Simeonov *et al.*, 2017). The choice of the effector protein has to be experimentally tested and tailored for each locus of interest, along with validation of sgRNA targeting, to ensure proper transcriptional control (Brocken *et al.*, 2018).

The ease of engineering of the CRISPR/Cas system makes it a cheap and fast genome-editing platform. However, one of the major concerns is its off-target binding at a genome-wide scale

(Kuscu *et al.*, 2014; Slaymaker *et al.*, 2016). Predictive algorithms have been developed to help design sgRNAs, by searching target sequences that are distinct from any other sequence in the genome and include fewer off-targets. Nevertheless, CRISPR/Cas systems have already proven to be instrumental in basic and applied research. Further optimization of this system will increase its reliability and decrease the off-targets, with important perspectives for both basic and clinical research.

THESIS OBJECTIVES

Centromeres are essential for the correct segregation of sister chromatids during cell division, and thus, their integrity is essential for genomic stability. Outside of mitosis, they are also involved in the assembly of repressive nuclear compartments and long-range transcriptional silencing (reviewed in Francastel *et al.*, 2000). Consequently, given its functions cited above, it is not surprising that defects in centromere identity and function are associated with human pathologies (reviewed in Barra and Fachinetti, 2018).

Somewhat counter-intuitively given their localization in a heterochromatin-rich environment, centromeric repeats are transcribed in most organisms. The resulting cenRNAs have an essential role in centromere function (Nakano *et al.*, 2003; Topp *et al.*, 2004; Quénet and Dalal, 2014; Kabeche *et al.*, 2018; Burry *et al.*, 2020). Murine cenRNAs have been shown to accumulate in G2/M phase of the cell cycle, when they participate in the correct spatio-temporal assembly of the kinetochore (Ferri et al, 2009). In clear contrast, their unscheduled and increased levels are toxic to cells and promote chromosomal instability and perturbation of nuclear architecture (Bouzinba-Segard *et al.*, 2006). In addition, high levels of (peri)cenRNAs have been reported in pathological situations with chromosomal instability such as cancer (Ting *et al.*, 2011; Eymery *et al.*, 2009). However, the molecular mechanisms that link the accumulation of cenRNAs and chromosomal instability await to be formally established.

Perturbations of centromere function have been documented in several physiopathological contexts, as depicted in the graphical abstract bellow:

- **DNA damage** has been reported to activate centromere transcription and leads to a drastic structural disorganization of centromeric chromatin (Hedouin *et al.*, 2017) and appearance of DNA-RNA hybrids (Yilmaz *et al.*, 2021). Given the repetitive nature of centromeres, unrepaired DNA damage at centromeres could lead to aberrant recombination at this locus.
- **DNA hypomethylation** at centromeres leads to increased recombination events (Jaco *et al.*, 2008) and correlates with their illicit transcriptional activation (Unoki *et al.*, 2020). DNA hypomethylation at centromeres could also impact the organization of this locus, by affecting the ability of proteins to bind to centromeres or affecting centromeric DNA compaction.
- **Unscheduled and/or increased transcription** at centromeres could lead to the collision of the replication and transcription machineries in S phase (Giunta, Herve *et al.*, 2020). Moreover, unscheduled transcription could perturb centromeres by leading to unscheduled chromatin remodeling or by impacting the eviction or deposition of Cenp-A.

- **Accumulation of cenRNAs** alters the centromeric chromatin environment, by tethering and sequestering centromeric, kinetochore and CPC proteins away from centromeres (Bouzinba-Segard *et al.*, 2006) or hinders DDR proteins localization at centromeres (Zhu *et al.*, 2018). Accumulation of cenRNAs could affect the availability of the RNA modifying machinery by titrating these enzymes. Additionally, accumulation of cenRNA could activate RNA surveillance mechanisms and trigger autoimmune responses.

These pathological events at centromeres co-occur in different pathological context, such as cancer or the ICF syndrome. However, their causal relationships, and how they lead to chromosomal instability, remains largely unknown.

The objective of my thesis was therefore to explore the contribution of these different events on centromeric chromatin dysfunction (**Figure 21**). To uncouple these different events, I set up epigenome editing tools to target centromeres and I questioned:

- How does increased centromere transcription impact centromere architecture, and function?
- Does increased transcription at centromeres lead to DNA damage at this locus?
- What are the consequences of increased centromere transcription in the different phases of the cell cycle?
- Does accumulation of cenRNAs *per se* have an effect on centromere integrity and function and is it distinct from that of increased centromere transcription?
- Could DNA hypomethylation provide a permissive context in which centromeric repeats would be transcribed at higher levels?

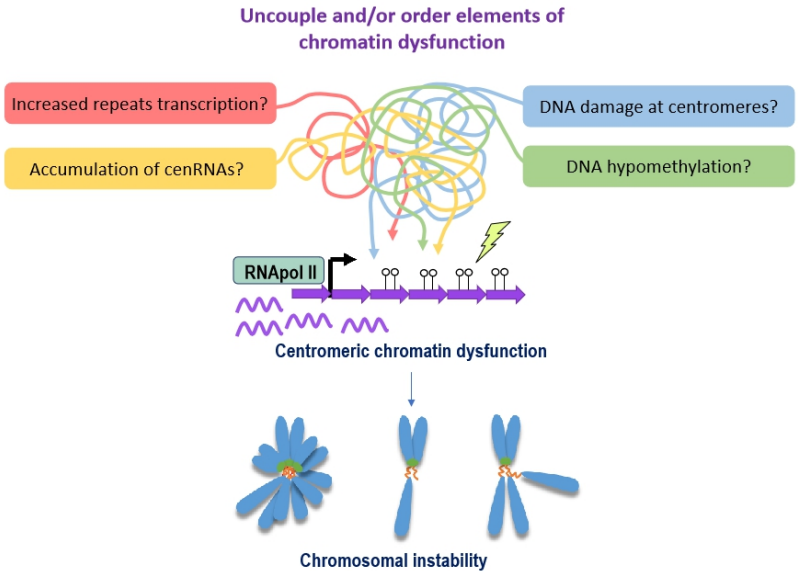


Figure 28. Graphical abstract of my PhD project

RESULTS

Part I: Mihic *et al.*, manuscript in preparation

Increased centromere transcription leads to R-loop-mediated loss of centromere architecture and function

Pia MIHIC[1], Guillaume VELASCO[1], Florence LARMINAT[2], Ekaterina BOYARCHUK[1]*, and Claire FRANCASTEL[1]*

Affiliations

[1] Université Paris Cité, CNRS, Epigenetics and Cell Fate, F-75013, Paris, France.

[2]Institut de Pharmacologie et Biologie Structurale, IPBS, Université de Toulouse, CNRS, UPS, Toulouse, France.

* Correspondence to: Ekaterina Boyarchuk, e-mail: ekaterina.boyarchuk@u-paris.fr, phone: +33 1 57 27 89 32 and Claire Francastel, e-mail: claire.francastel@gmail.com

Manuscript in preparation

ABSTRACT

Centromeres coordinate chromosome segregation through kinetochore assembly and their integrity is key for proper genome inheritance and stability. These loci have unique chromatin features and are assembled onto repetitive satellite sequences, which are transcribed at low levels in cycling cells. Although both the act of transcription and the resulting transcript are necessary for the proper centromere structure and function, their elevated levels are often linked to chromosomal instability that characterizes many human diseases. In this study, we deciphered a causal link between the dysregulation of centromeric transcription and centromere functions. Using CRISPR-Cas epigenome editing tools, we showed that moderate transcriptional activation of centromere repeats results in the accumulation of centromeric RNAs, R-loops, and DNA damage at these loci, followed by centromere decompaction and anaphase defects. We further demonstrate that the accumulation of R-loops at centromeres upon transcriptional activation is primarily responsible for these deleterious effects on centromeric structure and function. Our findings provide new insights into the molecular mechanisms that link aberrant centromere transcription to chromosomal instability.

INTRODUCTION

Centromeres are specialized chromosomal domains, essential for the proper segregation of the genetic material during each cell division. The highly specific chromatin organization of these domains governs the mitotic assembly of the kinetochore, a proteinaceous complex that binds spindle microtubules and ensures correct chromosomal attachment (reviewed in Kixmoeller *et al.* 2020). In higher eukaryotes, centromeres are composed of tandem satellite DNA repeats and are epigenetically defined by the presence of the centromere-specific histone H3 variant called CENP-A in humans and mice (reviewed in Sekulic & Black 2012). These CENP-A-enriched domains are embedded into constitutive heterochromatin regions, pericentromeres, that also participate in faithful chromosome segregation by contributing to the cohesion of the sister chromatids until the onset of anaphase.

Despite being embedded in neighboring heterochromatin, centromeres are transcribed at low levels, mainly by RNA polymerase II (RNAPol II) (Hall *et al.* 2012). Their transcription is regulated during the cell cycle (Bury *et al.* 2020; Ferri *et al.* 2009) and promotes the deposition of CENP-A (Bobkov *et al.* 2018; Chan & Wong 2012). However, when strongly elevated, centromeric transcription is not compatible with centromere function and leads to CENP-A eviction (Bergmann *et al.*, 2011; Hédouin *et al.*, 2017b). Importantly, a causal link and intermediate steps between the dysregulation of centromeric transcription and the loss of centromere integrity and function are still missing.

In line with transcription of their underlying repeats, centromeres yield centromeric RNAs (cenRNAs). The levels of cenRNAs peak at the onset of mitosis (Ferri *et al.*, 2009a), and to some extent hybridize with the accessible DNA template forming DNA-RNA hybrids, also known as R-loops (Kabeche *et al.*, 2018b; Moran *et al.*, 2021b; Thomas *et al.*, 1976b). In a normal physiological context, R-loops have been shown to be essential for faithful chromosome segregation (Kabeche *et al.* 2018; Moran *et al.* 2021). Generally, R-loops are known to have a regulatory role in transcription (García-Muse & Aguilera 2019). However, when encountering replication forks, they represent a potent source of replication stress (Gaillard & Aguilera 2016; Sollier & Cimprich 2015; Tuduri *et al.* 2009). When such collision occurs at centromeres, it results in the appearance of DNA breaks, recombination, and chromosome translocations at this locus (Giunta *et al.* 2021).

In addition to their potential in promoting the formation of R-loops co-transcriptionally, cenRNAs *per se* are an essential component of centromere chromatin, as they act as scaffolds to enable the binding of various centromeric proteins (Blower, 2016a; Ferri *et al.*, 2009a;

McNulty *et al.*, 2017a; Quénet and Dalal, 2014a; Rošić *et al.*, 2014a; Topp *et al.*, 2004b; Wong *et al.*, 2007a). However, it seems that cenRNA levels need to be precisely titrated. Indeed, their excessive accumulation is associated with loss of centromeric function (Bouzinba-Segard *et al.* 2006; Racca *et al.* 2021; Zhu *et al.* 2018) and is a common feature of chromosome instability diseases such as cancer (Hall *et al.* 2017; Ting *et al.* 2011; Zhu *et al.* 2011).

Here we have thus tackled the question of the still elusive direct consequences of excessive centromeric transcription on centromere function. We selectively induced transcription at murine centromeres, assembled on minor satellite repeats, using a CRISPR activation (CRIPSRa) system at levels that are similar to the pathological accumulation of centromeric transcripts observed *in vivo*. Using this approach, we showed that increased centromere transcription in interphase led to the formation of DNA-RNA hybrids, DNA damage at centromeres and culminated in the loss of centromere function as evidenced by lagging chromosomes and chromosome bridges in anaphase. Importantly, we demonstrated that the presence of the DNA-RNA hybrids, but not DNA damage, was a major cause of the perturbations in centromere architecture and mitotic function. We also highlighted that both the levels and the timing of centromeric transcription were crucial for securing the correct centromere functions.

RESULTS

CRISPRa targeting to minor satellites (cenCRISPRa) results in increased centromeric transcription and strong local accumulation of cenRNAs

To evaluate the impact of elevated centromeric transcription on centromere integrity and function, we forced transcription by using a CRISPR activation (CRISPRa) strategy (Chavez *et al.* 2015). To this end, we used the catalytically dead Cas9 (dCas9) fused to four minimal activation domains of VP16 protein (dCas9-VP64) (Konermann *et al.* 2015), which we targeted to the minor satellite repeats using the previously published gRNA (MiSgRNA) (Schmidtmann *et al.* 2016). As a control, we targeted a tagged version of dCas9 (dCas9-HA) (Braun *et al.* 2017) (Fig. 1a). We chose mouse cells as a model of study because of the fairly homogeneous centromere sequences across all chromosomes (Kalitsis *et al.* 2006), which allow us to achieve simultaneous centromeric targeting of CRISPRa (cenCRISPRa) at all centromeres, thus avoiding chromosome-specific effects.

First, we confirmed the correct targeting of dCas9 to centromeres by analyzing its co-localization by immunofluorescence (IF) with the centromere-specific protein Cenp-B, which directly binds DNA at a short centromeric sequence (Cenp-B box, (Masumoto *et al.* 1989)) and found a strong colocalization of the signals 24 h upon transduction with lentiviral particles that delivered either dCas9-HA or cenCRISPRa (Fig. S1a). We further confirmed correct targeting of dCas9 to the centromeres by chromatin immunoprecipitation (ChIP) and detected a strong enrichment of dCas9 at minor satellites (Fig. S1b).

Next, we tested by ChIP, followed by qPCR, the enrichment of transcriptionally active forms of RNAPII at centromeres upon cenCRISPRa transduction. We focused on two phosphorylated forms of RNAPII: on residue serine 5 on its C-terminal tail, mainly associated with transcriptional initiation (PolIIS5ph), or on serine 2, mainly associated with transcriptional elongation (PolIIS2ph) (Mayer *et al.* 2010) (Fig. 1b). PolIIS5ph ChIP showed a strong enrichment at 24 h post-transduction with cenCRIPSRa followed by relative loss of this form of PolII, whereas it was more modest for the elongating form (Fig 1b), strongly suggesting that the transduction with cenCRISPRa efficiently led to increased centromeric transcription.

We further assessed the levels of cenRNA levels after cenCRISPR-mediated transcriptional activation at centromeres by RT-qPCR analysis. We detected a 10-fold increase in cenRNA levels in a total cell population after 24 h of cenCRISPRa transduction, which increase was even more important at 48 h, reaching 25-fold (Fig. 1c). Additionally, we were not able to detect a drastic increase in the transcripts produced from major satellite repeats that constitute the adjacent pericentric domains (Fig S1c), confirming the specificity of the transcriptional activation of centromeric repeats with cenCRISPRa.

Finally, we analyzed the localization of centromeric transcripts upon cenCRISPRa by immuno-RNA-FISH and observed that cenRNAs accumulate in the vicinity of dCas9 in 36% of cells 24 h post-transduction with cenCRIPSRa (Fig. 1d), while they could hardly be detected upon targeting of dCas9-HA. Additionally, cenRNAs were still detected as foci at 48 h (Fig. S1d), suggesting that it is mostly the local concentration of cenRNA at centromeres that is elevated in our experimental system, and not only the total levels of cenRNAs detected by RT-qPCR.

We further investigated the strand-specificity of transcription promoted by cenCRISPRa. For this, we tested the presence of cenRNA that would originate from the other strand. Notably,

we did not detect any such transcript by RNA-FISH (Fig. S1e), suggesting that the cenCRISPRa specifically activates transcription from the targeted strand.

Together, these data indicate that transduction with cenCRISPRa efficiently drives increased transcription of centromeric minor satellite repeats by RNA Polymerase II followed by accumulation of cenRNAs at centromeres.

Increased transcription at minor satellites leads to perturbations of centromere architecture and function

Next, we investigated the effect of increased centromere transcription on centromere identity and architecture. Previous studies on human artificial chromosomes showed that when extremely elevated (more than 100 times) transcription levels cause decondensation of the centromeric region. In contrast, the transcriptional activation at lower levels, similar to those that we triggered here using cenCRISPRa, had no such effect (Bergmann *et al.*, 2011).

To test whether moderately increased transcription of endogenous centromeres affected centromere identity, we used Cenp-A as an epigenetic marker of this locus and assessed its localization by IF. We found that Cenp-A was still present at centromeres at 24 h and 48 h post-transduction with cenCRISPRa (Fig S2a and 2a), suggesting that, at those time points, the achieved levels of transcriptional activation did not affect centromere identity. At 24 h post-transduction with cenCRISPRa Cenp-A showed an expected sharp punctiform pattern, similar to the control condition (Fig. S2a). In contrast, at 48 h, Cenp-A exhibited stretched pattern in 40% of nuclei transduced with cenCRISPRa (Fig. 2a). This suggests that sustained moderately elevated centromere transcription either perturbs centromere architecture or that Cenp-A was spreading to the neighboring pericentromeric regions. To distinguish between these two possibilities, we tested the localization of Cenp-B, which is bound to centromeres in a sequence-specific manner (Masumoto *et al.* 1989). Similar to Cenp-A, at 24 h post-transduction, Cenp-B patterns were punctiform and comparable between all tested conditions, (Fig. S2b). Conversely, at 48 h Cenp-B pattern became more diffuse (Fig. 2a) in 35% of nuclei transduced with cenCRISPRa, suggesting that the underlying centromeric DNA was indeed decondensed.

To directly assess the integrity of DNA compaction, we performed DNA-FISH using probes specific for murine centromeric repeats. In agreement with our IF data, this analysis showed that at 24 h post-transduction with cenCRISPRa we could not detect any impact on centromeric DNA organization (Fig. S2c). However, at 48 h we observed diffuse foci in 41%

of cells transduced with cenCRISPRa compared to punctiform minor satellite signals of control conditions (Fig. 2b). Taken together, these observations suggested that sustained moderately increased transcription of minor satellites leads to decompaction and disorganization of centromeric DNA.

Given the essential role of centromeres in mitosis, we examined the effect of increased centromere transcription on cell division. While a 24 h increase of had no strong effect on mitosis in NIH-3T3 cells, at 48 h post-transduction with cenCRIPSRa, cells exhibited elevated levels of lagging chromosomes in 76% of observed anaphases (Fig. 2c).

Taken together, these data suggested that sustained elevated transcription of centromeric repeats leads to perturbations of centromere architecture, which is followed by the accumulation of mitotic error.

Increased transcription at Minsat leads to DNA damage and accumulation of DNA-RNA hybrids at centromeres

The progression of RNAPII requires chromatin remodeling and might be causal to changes in chromatin compaction (Studitsky *et al.* 2004). However, the emergence of centromeric disorganization at 48 h and not at 24 h suggests that the activation of transcription may not be the main causative event *per se*, but may initiate a cascade of events that results in the perturbation of centromeric architecture. We thus asked what were the consequences of the increased transcription at centromeres. Since chromatin reorganization and decondensation are known to coincide with DNA damage (Hauer & Gasser 2017; Jakob *et al.* 2011), we investigated the presence of γH2AX as a marker of DNA damage at centromeric chromatin. At 24 h post-transduction with cenCRIPSRa, before we can detect disorganization of centromeric DNA using DNA FISH (Fig. 2b), γH2AX co-localized with Cenp-B signal in 79% of nuclei (Fig 3a). Such colocalization between γH2AX and Cenp-B also persisted at 48 h (Fig S3a). Notably, there was no significant enrichment of γH2AX at centromeres in cells targeted with dCas9-HA (Fig 3a, S3a). We further confirmed the presence of γH2AX at centromeres after transduction with cenCRISPa by ChIP-qPCR at minor satellite repeats (Fig. 3b).

Minor satellite repeats are prone to form DNA-RNA hybrids known as R-loops, (Kabeche *et al.* 2018; Moran *et al.* 2021; Racca *et al.* 2021). Importantly, the presence of aberrant and/or unscheduled DNA-RNA hybrids is known to correlate with an elevated risk of DNA breakage (Aguilera & Gómez-González 2017; Hamperl *et al.* 2017). Therefore, we first analyzed the

presence of such hybrids upon transcriptional activation at centromeres by IF, using the monoclonal S9.6 antibody, which targets DNA-RNA heteroduplex structures in a non-sequence-specific manner (Sanz & Chédin 2019; Tan *et al.* 2020). In control conditions, we found that S9.6 antibody signal was only observed in nucleoli which are known to be enriched in DNA-RNA hybrids involving rRNA (Vydzhak *et al.* 2020). Upon induction of centromeric transcription with cenCRIPSRa, the signal of the S9.6 antibody co-localized with that of Cenp-A in 60% of nuclei (Fig. 3c), indicating the accumulation of DNA-RNA hybrids at centromeres. Consistently, using DNA-RNA immunoprecipitation (DRIP), we detected significant enrichment of DNA-RNA hybrids at minor satellites in cells transduced with cenCRISPRa compared to control cells or cells transduced with dCas9-HA (Fig. 3d). The enrichment was specific to DNA–RNA hybrids since they were eliminated by treatment of the samples with recombinant RNase H1 prior precipitation (Fig. S3b).

Taken together our findings suggest that increased centromere transcription leads to the formation of centromeric DNA-RNA hybrids in interphase nuclei which coincides with the appearance of a hallmark of DNA damage at these loci.

Removal of centromeric DNA-RNA hybrids restores centromere architecture and function but does not prevent DNA damage

We next sought to test the causal links between DNA damage at centromeres, the formation of DNA-RNA hybrids, and centromere dysfunction during mitosis by digesting the RNA moiety in the hybrids, using exogenously expressed RNaseH1 (Hyjek *et al.* 2019). DNA-RNA hybrids have essential physiological roles such as regulation of gene expression, chromosome segregation, and telomere stability (Crossley *et al.* 2019; Kabeche *et al.* 2018; Pérez-Martínez *et al.* 2020; Roy *et al.* 2008). Hence, to minimize indirect effects caused by the removal of DNA-RNA hybrids genome-wide and to ensure efficient removal of DNA-RNA hybrids specifically from centromeres, we targeted RNaseH1, or its catalytically dead counterpart dRNaseH1 as a control, directly to centromeres. We used a tripartite system composed of dCas9 (fused either to HA or VP64), a MiSgRNA modified with MS2 RNA aptamers, and (d)RNaseH1 fused to an MS2 binding protein (MS2BP, Fig. 4a) (Bertrand *et al.* 1998). With this system, we were able to simultaneously activate centromere transcription by cenCRISPRa and target RNaseH1 to centromeres (Fig S4a). As expected, expression of MS2-RNaseH1 together with cenCRIPSRa led to the decrease of DNA-RNA hybrids at centromeres as detected by S9.6 signal in IF (Fig 4b) and in DRIP (Fig 4c) experiments. Conversely, as

expected, co-expression of dRNaseH1 and cenCRIPSRa stabilized DNA-RNA hybrids at centromeres as evidenced by both, IF and DRIP approaches (Fig 4b and c).

Interestingly, neither the removal of DNA-RNA hybrids at centromeres by RNaseH1 nor their stabilization by dRNaseH1 affected the presence of centromeric DNA damage marker. Indeed, the γH2AX signal still co-localized with Cenp-B at 24 h in both conditions (Fig. 4d and Fig. S4b), suggesting that while the presence of DNA-RNA hybrids might exacerbate the DNA damage at centromeres, a fraction of γH2AX signal at these loci detected upon activation of centromeric transcription by cenCRISPRa seems independent of the formation of DNA-RNA hybrids.

Importantly, the removal of accumulated DNA-RNA hybrids with MS2-RNaseH1 restored the normal punctiform pattern of Cenp-A at 48 h post-transduction (Fig. 4e), otherwise perturbed by the cenCRIPSRa-mediated excessive transcription (Fig. 2a). On the other hand, consistent with previous observations (Racca *et al.* 2021), we observed that the Cenp-A signal was more dispersed in the nucleus in the presence of MS2-dRNaseH1 (Fig 4e). Together, these data strongly suggested that an accumulation and/or stabilization of DNA-RNA hybrids in interphase is a key driver of the disruption of centromere architecture upon activated transcription of centromeric repeats.

Strikingly, the co-targeting of cenCRISPRa and RNaseH1, which removes excessive DNA-RNA hybrids from centromeres but permits transcription (Fig S4a) of the repeats, was sufficient to prevent chromosomal aberrations in the majority (64%) of the observed anaphases (Fig 4f). Notably, the stabilization of excessive DNA-RNA hybrids at centromeres neither rescued nor aggravated the percentage of aberrant anaphases, compared to samples transduced with cenCRISPRa alone (Fig 1); in both cases, we observed 70% of aberrant anaphases. These data suggest that, at least in NIH3T3, certain levels of DNA damage can be present at centromeres and either do not lead to immediate negative effects on centromere function or are repaired before entering into mitosis, while the excess of DNA-RNA hybrids heavily affects mitosis fidelity.

Therefore, we concluded that, upon increased transcription at centromeres, the excess of DNA-RNA hybrids affects centromeric architecture and function. Additionally, the presence of DNA damage upon transcriptional activation at the centromeric repeats seems to be partly independent of DNA-RNA hybrids formation.

Increased transcription at centromeres in G1 or S/G2 has distinct consequences on anaphase fidelity

In human and mouse cells, centromeric DNA is transcribed in mitosis (Chan *et al.*, 2012b), producing DNA-RNA hybrids which are essential for correct centromere function and faithful chromosome segregation (Kabeche *et al.*, 2018b; Moran *et al.*, 2021b). In addition, cenRNA levels are cell-cycle regulated and peak in the G2/M phases in murine (Ferri *et al.* 2009) and human (Bury *et al.* 2020) cells. To assess the impact of the untimely centromere transcription on the loss of centromere function, we specifically triggered increased transcription either in S/G2 or in G1 and tested to which extent centromere transcription in S/G2 or G1 phases would lead to aberrant DNA-RNA hybrid formation and mitotic defects.

To activate centromere transcription in G1, we fused cenCRIPSRa to the N-terminal part of Cdt1, a pre-replication factor degraded in all phases of the cell cycle except in G1 phase (cenCRISPRa-Cdt1, Fig 5a) (Howden *et al.* 2016; Nishitani *et al.* 2004). Alternatively, to activate transcription specifically in S/G2, we modified cenCRISPRa by fusing it to the N-terminal part of human Geminin, a replication factor that is degraded in M and G1 (cenCRISPRa-Gem, Fig. 5a) (Charpentier *et al.* 2018; Nishitani *et al.* 2004).

Our above-described data (Fig 3) showed that the accumulation of DNA-RNA hybrids is one of the most deleterious outputs of activated transcription at centromeres. Therefore, we first questioned whether the activation of centromere transcription in either G1 or S/G2 phases leads to aberrant DNA-RNA hybrid formation to a similar extent. 24 h post-transduction with cenCRISPRa-Cdt1, DNA-RNA hybrids were detectable on centromeres in the majority of transduced cells (75%, Fig. 5c), reminiscent of the timing of appearance and severity of the defects trigged by centromeric transcriptional activation by cenCRISPRa (Fig. 3c). In contrast, we detected S9.6-positive signals at centromeres in only 20% of cells in interphase after triggering transcription with cenCRISPRa-Gem (Fig. 5b), which correlates with the percentage of the cells in G2 in NIH3T3. This also suggests that the DNA-RNA hybrids could be formed on the centromeres upon enhanced transcriptional activation regardless of the phase of the cell cycle. Importantly, we were not able to detect any significant increase in anaphase defects in cells 24h after the induction of centromeric transcription in G1, using cenCRISPRa-Cdt1 (Fig. 5c), similar to our results with cenCRISPRa (Fig. 2c). However, at 48h post-transduction, 61% of observed anaphases featured lagging chromosomes (Fig. 5c), reminiscent of the time of appearance and levels of the defects trigged by centromeric transcription activation by cenCRISPRa (Fig. 2).

Surprisingly, already 24 h after transduction with cenCRISPRa-Gem and triggering transcription in S/G2, we found that the majority (94%) of the observed anaphase displayed defects: lagging chromosomes in 75% and anaphase bridges in 18% (Fig. 5d). These observations show that even though endogenous cenRNAs peak in G2 (Bury *et al.* 2020; Ferri *et al.* 2009) and that the presence of DNA-RNA hybrids is necessary for faithful chromosome segregation (Kabeche *et al.* 2018; Moran *et al.* 2021), inducing increased transcription at centromeres during this time-window disrupts centromere function.

In conclusion, our study showed that the activation of centromeric transcription, which leads to the accumulation of cenRNA to levels similar to those observed in some pathological conditions characterized by chromosomal instability, results in the perturbation of centromere architecture, and is accompanied by chromosome segregation errors (Fig. 6). We provide evidence that the observed centromere disorganization arises mainly from the excessive accumulation of DNA-RNA hybrids on these loci, while the accumulation of resulting DNA damage has a less severe impact. Our data also brings forward the importance of fine-tuning the levels of centromeric transcription for controlling centromere architecture and function.

DISCUSSION

The data presented here revealed that the targeted transcriptional activation at centromeres, leading to increased cenRNA levels similar to those found in pathological situations of chromosomal instability (Ting *et al.*, 2011b; Unoki *et al.*, 2020b), promotes a rapid local accumulation of cenRNAs, R-loops, and DNA damage marker γH2A.X, at those loci. This further leads to centromeric DNA decondensation and anaphase defects. We also showed that the accumulation of R-loops at centromeres is mainly responsible for such deleterious effects on centromere structure and function (Fig. 6).

The role of R-loops at centromeres is still controversial. It has been previously reported that mitotic R-loops are necessary for the proper kinetochore function (Kabeche *et al.* 2018; Moran *et al.* 2021). Additionally, R-loops have been found to be beneficial for the integrity of centromeres in G1 in the presence of DNA breaks (Yilmaz *et al.* 2021). Conversely, excessive accumulation of R-loops at centromeres has been shown to affect kinetochore bi-orientation and chromosomal stability in budding yeast (Mishra *et al.*, 2021). Furthermore, R-loops accumulation has been linked to the reduction of Cenp-A levels at centromeres, accumulation of DNA damage, and centromere recombination, which are associated with chromosomal instability in human cells (Giunta *et al.* 2021; Racca *et al.* 2021). While the observed positive

effects of R-loops on centromeres are likely to be explained by the R-loops-specific recruitment of proteins involved in kinetochore function and Cenp-A deposition (Moran *et al.* 2021; Yilmaz *et al.* 2021), the mechanism(s) through which R-loops are deleterious for centromere function is intriguing. We found that the resolution of R-loops by RNAseH1 prevents centromeric decompaction that otherwise results from the enhanced transcriptional activation of the locus. This negative effect of R-loops on chromatin compaction partially contradicts the previous observation that the presence of these structures correlates with chromatin condensation genome-wide (Castellano-Pozo *et al.* 2013). Such discrepancy could be attributed to the repeated nature and chromatin states of the centromeres. Another possibility is that the presence of R-loops might directly affect the DNA-loop-mediated organization of centromeric chromatin that characterize functional centromeres (Chardon *et al.* 2022; Kasinathan & Henikoff 2018). Indeed, human and mouse centromeric satellite repeats have been shown to adopt non-B form DNA (Kasinathan & Henikoff 2018), forming secondary structures that have been implicated in the maintenance of centromere identity (Kasinathan & Henikoff 2018; Mellone & Fachinetti 2021) and centromeric stability (Black & Giunta 2018; Bloom & Costanzo 2017). Moreover, a recent study has shown that in human cells CENP-B compacts centromeric DNA by DNA-looping, which helps to maintain centromere DNA integrity (Chardon *et al.* 2022).

Importantly, a functional role for centromeric DNA loops has been proposed during mitosis, where they may act as molecular strings to absorb pulling forces of the microtubules exerted on centromeric chromatin (Bloom 2014). The loss of the centromere rigidity could be one of the explanations for our otherwise surprising observation that a strong accumulation of R-loops in G2, before the onset of mitosis when the presence of the transcripts and R-loops is expected to play a positive role (Kabeche *et al.*, 2018b; Moran *et al.*, 2021b), has a more deleterious effect on faithful chromosome segregation than continuous transcriptional activation during the whole cell cycle. Therefore, our findings open up the exciting question of whether the deregulated formation of R-loops could directly affect the DNA-folding at centromeres.

Apart from a potential impact on DNA folding, R-loops as well as the act of transcription *per se* might modify centromeric chromatin by affecting nucleosome density (Chédin 2016) and centromere-specific DNA- and Cenp-A nucleosome-binding components of the CCAN (constitutive centromere-associated network) complex (Hara & Fukagawa 2017). It would be particularly interesting to evaluate whether the activated transcription affects the presence or

amount of CENP-N, which has been shown to directly promote the compaction of centromeric chromatin (Zhou *et al.* 2022).

In addition to the accumulation of R-loops and DNA damage, we observed a strong local increase in cenRNA levels upon targeted centromeric transcriptional activation, which could have an impact on the observed centromere dysfunction. Although cenRNAs are necessary components of the centromere (Topp *et al.*, 2004; Quénet and Dalal, 2014; Kabeche *et al.*, 2018; Bouzinba-Segard *et al.* 2006; Ferri *et al.* 2009), they could function as sponges and sequester proteins away from their site of action when accumulated above a certain threshold (reviewed in (Mihìc *et al.* 2021)). Indeed, high levels of cenRNAs have been linked to centromere dysfunction by tethering kinetochore proteins away from centromeres (Bouzinba-Segard *et al.* 2006; Zhu *et al.* 2011). Furthermore, the local concentration of RNAs might influence the physical properties of chromatin subcompartments through the modulation of phase-separation abilities of certain proteins (reviewed in (Rippe 2022)). Indeed, the relatively high levels of the major satellite RNAs produced from the neighboring pericentromeric heterochromatin have been shown to modulate the physical properties and nuclear organization of heterochromatin in mouse embryonic stem cells (ESCs) and to promote heterochromatin clustering in interphase (Novo *et al.* 2022). The suggested mechanism relies on the ability of the locally accumulated major satellite RNAs to positively regulate the phase-separating properties of HP1 proteins (Larson *et al.* 2017; Novo *et al.* 2022; Strom *et al.* 2017). Interestingly, among the kinetochore constituents, the subunit of the essential component of the inner centromere chromosomal passenger complex (CPC), Borealin, has also been shown to phase-separate (Trivedi *et al.* 2019). This ability of the CPC to undergo phase separation has been demonstrated to be important for its inner-centromere localization and functions, which includes the surveillance of the correct chromosome segregation (Trivedi *et al.* 2019; Trivedi & Stukenberg 2016), the function that was severely affected upon increased centromere transcription in our experiments. Importantly, the assembly of the CPC to the centromeres is facilitated by its binding to cenRNAs (Ferri *et al.* 2009). Hence, it could conceivably be hypothesized that the changes in the local concentration of the cenRNA might provide means to alter the physical properties of the centromeres and influence centromere functions.

Notably, transcription of centromeric repeats under stress conditions is one of the mechanisms that trigger rapid cellular responses for cells to recover from stress (Hédouin *et al.* 2017;

reviewed in Mihìc *et al.* 2021). However, when coinciding with a chronic disease state, these safeguard mechanisms could be bypassed. Indeed, the levels of the transcripts induced in our experiments mimicked those observed in pathological conditions, suggesting that some adaptive mechanisms might partially compensate for the observed deleterious effects. Thus, it has been shown that cancer cells rely on Ino80 remodeler to promote R-loops resolution in S-phase to prevent the collapse of the replication fork (Prendergast *et al.* 2020). In line with this, the chromatin remodeler Ino80 was found at the centromeres and proved important for the CENP-A loading related to transcription (Choi *et al.* 2017; Singh *et al.* 2020). It suggests that chromatin remodelers could help to tip the balance between the necessity of centromere transcription and the deleterious effects of excessive R-loops accumulation on these loci. Similar adaptations could be envisaged for the other deleterious aspects of enhanced centromeric transcription and accumulation of the transcripts, described in current and other studies (Bergmann *et al.*, 2011; Bury *et al.*, 2020a; Giunta *et al.*, 2021b). The resolution of the local centromere proteome in various physiological contexts, including the pathologies that affect centromeric transcription, might provide the means to improve our understanding of the regulation of centromeric transcription and get mechanistic insights into the regulation of centromere chromatin structure and function.

FIGURE LEGENDS

Figure 1. CRISPRa targeting to minor satellites leads to increased centromeric transcription and accumulation of cenRNAs.

a. Schematic representation of the experimental approach. NIH-3T3 cells are transiently transduced with lentiviral particles containing a plasmid coding for dCas9 fused to either HA or VP64 (cenCRISPRa) and simultaneously transfected with a plasmid expressing minor satellite-targeting gRNA (MiSgRNA). 24h or 48h post-transduction, cells are fixed and/or processed for the following experiments.

b. Chromatin immunoprecipitation (ChIP) assays performed on chromatin prepared from control cells or cells transduced with cenCRISPRa/gRNA for 24 and 48 h, using antibodies against RNA PolII phophosphorylated on Ser5 or Ser2 on its C-terminal tail (PolIIS5ph and PolIIS2ph, respectively), followed by qPCR with sequence-specific primers for Minor satellites. The number of independent experiments (n)=1

c. Quantitative PCR after reverse transcription (RT–qPCR) showing fold change of cenRNA expression levels upon 24 and 48 h post-transduction with dCas9-VP64 and MiSgRNA

(cenCRISPRa) or dCas9-HA control (HA). n=3; the p-value was determined by the two-tailed t-test.

d. Immuno-RNA-FISH for cenRNA and dCas9 protein in cells transduced with dCas9-HA or cenCRISPRa/gRNA. Cells were analyzed 24 h post-transduction. n=3, 30 nuclei with dCas9 signal were quantified in each experiment, the p-value was determined using the two-tailed t-test. The results for 48 h post-transduction are shown in Fig. S1c

Figure 2. Increased centromeric transcription perturbs the centromere architecture and leads to mitotic errors.

a. Representative images (left & middle) and quantifications (right) of Cenp-A and Cenp-B patterns in cells transduced with either dCas9-HA/ or cenCRISPRaA/gRNA. Cells were fixed 24 h (shown Fig. S2a and S2b) and 48 h post-transduction and stained with antibodies specific for Cas9 in combination with Cenp-A (left) or Cenp- B (middle panel) antibodies. Nuclei were stained with DAPI. Insets are magnified an additional 2X. n=3, at least 100 nuclei with dCas9 signal were analyzed per experiment for Cenp-A and Cenp-B localization.

b. DNA-FISH analysis of the architecture of minor satellite repeats 48 h post-transduction with either dCas9-HA/ or cenCRISPRa/gRNA. Insets are magnified an additional 3X. n=3, at least 100 nuclei were analyzed per experiment.

c. Analysis of anaphase defects in cells transduced with either dCas9-HA or cenCRISPRa, stained with DAPI. Arrowhead on the right points to lagging chromosomes. Two independent experiments were analyzed, 20 anaphases for HA; 17 anaphases for cenCRISPRa.

All images represent one focal plane. Scale bar, 10 μm. The p-values were determined using the two-tailed t-test.

Figure 3. Increased transcription at centromeres leads to DNA damage and accumulation of DNA-RNA hybrids at these loci.

a. Representative images and percentages of the cells exhibiting the presented patterns. Cells were analyzed 24 h post-transduction with either dCas9-HA/ or cenCRISPRa/gRNA. Antibodies against γH2AX were used as a DNA damage marker and against Cenp-B as a marker of underlying centromeric repeats. n=3, 100 nuclei were quantified for each experiment.

b. ChIP-qPCR assays performed on chromatin prepared from cells transduced with cenCRISPRa/ or dCas9-HA/gRNA, using antibodies against γH2AX, followed by quantitative PCR with sequence-specific primers for minor satellites.

c. Representative images of IF analysis of DNA-RNA hybrids at 24 h post-transduction with either dCas9-HA/ or cenCRISPRa/gRNA. Nuclei were stained with DAPI and DNA-RNA hybrids were detected with the S9.6 antibody, Cenp-A antibody was used as a marker of centromere identity. n=3; 70 nuclei for Cenp-A and Cenp-B were analyzed for each experiment.

All images represent one focal plane. Scale bar, 10 μm. The p-values were determined using the two-tailed t-test.

d. DNA-RNA immunoprecipitation (DRIP) of DNA-RNA hybrids using S9.6 antibody followed by qPCR with sequence-specific primers for murine centromeric minor satellite repeats. Cells were collected 24 h post-transduction, n=3, the p-values were determined using the two-tailed t-test.

Figure 4. Resolution of centromeric DNA-RNA hybrids accumulation restores centromere architecture and function.

a. Schematic representation of the experimental setup. NIH-3T3 cells were transiently transduced with lentiviral particles containing dCas9-VP64 and lentiviral particles containing MS2-RNaseH1 or the catalytically dead RNaseH1 (dRNaseH1) and simultaneously transfected with MiSgRNA that contains MS2 RNA aptamers. 24 h after transfection, cells were fixed and the detection of S9.6 or γH2AX was performed in the following experiments.

b. Representative images of IF analysis of DNA-RNA hybrids 24 h post-transduction with cenCRISPRa/gRNA and either MS2-dRNaseH1 or MS2-RNaseH1. Nuclei were stained with DAPI and DNA-RNA hybrids were detected with the S9.6 antibody. Cenp-A antibody was used as a marker of centromere identity. n=3, 50 nuclei were quantified for each experiment. The p-value was determined using the two-tailed t-test.

c. DRIP assay followed by qPCR with primers for minor satellite repeats in cells transduced with cenCRISPRa with either MS2-dRNaseH1 or MS2-RNaseH1. n=2.

d. Representative images of cells co-stained with antibodies specific for Cenp-B as a marker of centromeric repeats, and γH2AX as a marker of DNA damage. Cells were transduced with either cenCRISPRa/gRNA together with MS2-dRNaseH1 or MS2-RNaseH1, and analyzed 24

h post-transduction. Nuclei were stained with DAPI. The quantification is shown in Fig S4b. n=2; 50 nuclei were quantified for each experiment.

e. Representative images of cells co-stained with antibodies specific for Cas9 and Cenp-A, 48 h post-transduction with cenCRISPRa/gRNA and either MS2-dRNaseH1 or MS2-RNaseH1. Nuclei were stained with DAPI. n=2, 20 nuclei were quantified for each experiment.

f. Analysis of anaphase defects in cells cenCRISPRa with either MS2-dRNaseH1or MS2-RNaseH1. Nuclei were stained with DAPI. n=2; 9 anaphases for MS2-dRNaseH1 and 17 anaphases for MS2-RNaseH1 were analyzed.

All images represent one focal plane. Scale bar, 10 μm.

Figure 5. Increased transcription at centromeres in G1 or in S/G2 has distinct consequences for the cells.

a. Schematic representation of dCas9-VP64–Gem or –Cdt1 stability during the cell cycle.

b. Representative images of IF analysis of DNA-RNA hybrids 24 h post-transduction with cenCRISPRa-Cdt1/ or cenCRISPRa-Gem/gRNA. DNA-RNA hybrids were detected with the S9.6 antibody, Cenp-A antibody was used as a marker of centromere identity. Nuclei were stained with DAPI. The percentages of the cells with colocalized signals for S9.6 and Cenp-A are shown. n=1, 40 nuclei were quantified.

c. Analysis of anaphase defects in cells 24 h or 48 h post-transduction with cenCRISPRa-Cdt1. DNA was stained with DAPI. The percentages of the cells with abnormal anaphases are shown. n=2, 23 anaphases were analyzed

d. Analysis of anaphase defects in cells 24 h post-transduction with cenCRISPRa-Gem. DNA was stained with DAPI. The percentages of the anaphases with lagging chromosomes (left) and chromosomal bridges (right) are shown. n=2, 13 anaphases were analyzed

All images represent one focal plane. Scale bar, 10 μm.

Figure 6. Model illustrating the impact of increased centromere transcription on centromere functions.

Transcription at centromeres, enhanced by cenCRISPRa, results in local accumulation of cenRNAs, DNA damage, and unscheduled R-loops at the underlying repeats, followed by anaphase defects. Resolution of the unscheduled R-loops by RNAse H1, while retaining accumulation of DNA damage and cenRNAs at centromeres, rescues the anaphase defects.

SUPPLEMENTAL FIGURES:

Figure S1, related to Figure 1.

a. Representative image (left) of the cells co-stained with antibodies specific for Cas9 and Cenp-B 24 h post-transduction with cenCRISPRa. Nuclei were stained with DAPI. The fluorescence intensity line scan (right) through each nucleus (red dotted line) shows the profiles for cenCRISPRa (green line) and Cenp-B (red line).

b. ChIP-qPCR assay performed on chromatin prepared from control cells or cells transduced with dCas9-HA or cenCRISPRa for 24 h, using antibodies against Cas9. The primers, specific for minor satellite sequences, were used for quantitative PCR. ChIP-qPCR results were normalized on the input signal (% input). n=1

c. Quantitative PCR with reverse transcription (RT–qPCR) showing fold change of major satellite RNA expression levels upon 24 and 48 h post-transduction with dCas9-HA and cenCRISPRa. n=3.

d. RNA-FISH for cenRNA in cells transduced with dCas9-HA or cenCRISPRa and transiently transfected with MiSgRNA. Cells were analyzed 48 h post-transduction. n=2, 100 nuclei were quantified in each experiment.

e. RNA-FISH for cenRNA originated from reverse DNA strain (cenRNA R) in cells transduced with cenCRISPRa. Cells were analyzed 48 h post-transduction. As a positive control, the cells were transiently transfected with a plasmid driving the expression of Reverse cenRNA under the control of the CMV promoter.

Figure S2, related to Figure 2.

a, b. Representative images (top) and quantifications (bottom) of cells with different Cenp-A (a) and Cenp-B (b) patterns in cells 24 h post-transduction with dCas9-HA or cenCRISPRa. Cells were stained with antibodies specific for Cas9 and Cenp-A or Cenp-B, respectively. Nuclei were co-stained with DAPI. n=3, at least 100 nuclei were analyzed for each experiment. The p-values were determined using the two-tailed T-test. All images represent one focal plan. Scale bar, 10 μm

c. Representative image of DNA-FISH analysis of the architecture of minor satellite repeats after 24 h post-transduction with cenCRISPRa.

Figure S3, related to Figure 3.

a. Representative images of cells stained with antibodies specific for dCas9 and γH2AX 48 h post-transduction with either dCas9-HA or cenCRISPRa. Nuclei stained with DAPI. The percentages of the cells with presented γH2AX nuclear distribution are shown. N=3, 50 nuclei per experiment were analyzed.

b. DRIP-qPCR analysis on the genomic DNA purified from the cells control cells or 24 h post-transduction with dCas9-HA or cenCRISPRa. RNase H1 treatment was carried out on half of each sample before the IP. The graph shows the DNA-RNA hybrids enrichment (as a percent of input) using qPCR with specific primers for murine centromeric minor satellite repeats**.** n=2

Figure S4, related to Figure 4.

a. RT–qPCR analysis showing fold change of cenRNA expression levels upon 24 and 48 h of transduction with dCas9-HA, cenCRISPRa alone, or together with RNaseH1 or dRNaseH1. The expression of cenRNA in dCas9-HA transduced cells was set as 1.

b. The percentages of the cells with colocalization of γH2AX and Cenp-B signals. Representative images are shown in Fig. 4d. n=1, 40 nuclei were analyzed.

MATERIALS AND METHODS

Cell culture and transfection

Immortalized murine fibroblasts NIH/3T3 (ATCC) were cultured in Dulbecco's Modified Eagle Medium/Glutamax (Gibco) supplemented with 10% fetal bovine serum (FBS, Deutscher), 100 U/ml penicillin and 100 µg/ml streptomycin (Gibco) at 37°C and 5% CO_2. Cells were transiently transduced with lentiviral particles produced in-house. Transient transfection was performed using JetPei (Polyplus-transfection) according to the manufacturer's instructions. The plasmids are listed in **Table 1**.

Immunofluorescence

Cells were fixed for 15 min at room temperature in 4% paraformaldehyde solution (Electron Microscopy Science) in PBS, followed by 3 washes in PBS. Then, cells were permeabilized for 15min in PBS supplemented with 0.1% Triton X-100. After 3 washes in PBS, samples were blocked with 2% donkey serum (Sigma, D9663) in PBS for a minimum of 30min before primary antibody staining overnight. After 3 PBS washes, samples were incubated in

secondary antibody for 1 h and counterstained with DAPI (1 mg/ml), and mounted in Vectashield (Vector labs). For visualization of mitoses, cells were permeabilized with PBS/0,1% Tween-20 for 15min and counterstained with DAPI. For immunostaining of S9.6 samples were prepared using a recently reported protocol (Tan *et al.* 2020). The antibodies are listed in the **Table 2**.

DNA & RNA FISH

DNA-FISH and RNA-FISH were performed as previously described (Bouzinba-Segard *et al.*, 2006a). Minor satellite oligonucleotide probes were labeled in 3' with fluorescent Cy3-dUTP using terminal transferase (New England Biolabs) according to the manufacturer's instructions. Probe sequences are provided in **Table 3**.

Immuno-RNA-FISH

Cells grown on 12 mm coverslips were fixed in 4% paraformaldehyde in PBS, for 15 min on ice. Cells were then permeabilized for 15min in ice-cold cytoskeletal buffer (NaCl 100 mM, sucrose 300 mM, MgCl2 3 mM, PIPES 10 mM) supplemented with 0.5% Triton, 1 mM EGTA, and 2 mM Vanadyl Ribonucleoside complex (VRC, New England Biolabs). Following that, samples were incubated in a blocking buffer composed of 2% donkey serum in PBS supplemented with 1X VRC, for 30 min at room temperature. First, an immunofluorescence assay was performed, with the incubation of the primary antibody for 2 h at room temperature and incubation of the secondary antibody for 30 min. After that, cells were post-fixed in 2% paraformaldehyde in PBS and hybridized with 100 ng of probes complementary to murine Minor satellite repeats (sequence provided in **Table 3**). Probes were labeled with cy3 on their 3' end using the terminal transferase (New England Biolabs), following the manufacturer's protocol. Coverslips were hybridized overnight with the probes at 37 °C in 50% formamide/50% hybridization buffer [4× SSC (saline-sodium citrate), 20% dextran sulfate, 2 mg/ml bovine serum albumin], and 20 U of RNase inhibitor (New England Biolabs). After 3 washes in 2xSSC/1xVRC at 37 °C for 5 min, coverslips were mounted in Vectashield containing DAPI.

Microscopy

Images were acquired on a Zeiss Axioplan 2 fluorescence microscope, with a Plan-Neofluar 100X/1.3 NA oil immersion objective (Zeiss) using a digital cooled camera (CoolSNAP fx;

Photometrics) and METAMORPH 7.04 software (Roper Scientific, Trenton, NJ). All images correspond to one focal plane. Analysis was performed by scoring at least 50 cells in each experiment.

RNA extraction and RT-qPCR analysis

Total RNA from cells was isolated using TRI Reagent® (Sigma) according to the manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNAfree kit (Ambion). Reverse transcription (RT) was carried out using 1µg DNA-free RNA and 60 µM random hexamers (Roche), 20U of RNase inhibitor (New England Biolabs), and 200U of ProtoScript II Reverse Transcriptase (New England Biolabs). Real-time quantitative PCR was performed using the LightCycler® DNA Master SYBR Green I mix (Roche) supplemented with 200 nM specific primer pairs (all listed in Table 3). Real-time quantitative PCR was run on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers, and analyzed by the comparative CT (ΔΔCT) method using U6 RNA as an invariant RNA. Used primers are listed in **Table 4**.

DRIP assay

DNA-RNA immunoprecipitation assay was adapted from Racca *et al.*, 2021. Briefly, genomic DNA was extracted from 3-5 million cells using the Monarch® Genomic DNA Purification Kit (New England Biolabs) according to the manufacturer's instructions. Genomic DNA was eluted in 100 µl of elution buffer (5 mM Tris-HCl pH 8) and then fragmented by sonication using the Bioraptor Pico (Diagenode), 6 cycles of 10 sec ON, 10 sec OFF, to yield an average fragment size of 800–200 bp. 2µg of each sample was treated with E. coli RNase HI (New England Biolabs) or mock-treated, overnight at 37 °C in 1x RNase H reaction buffer. Following that, DNA was diluted with 400 µl binding buffer (10 mM NaPO4, pH 7.0/140 mM NaCl/0.05% Triton X-100) and immunoprecipitated with 5 µg of the S9.6 antibody on a wheel at 4 °C for at least 4 h. At the end of the incubation, 25 µl of protein A/G magnetic beads (ThermoFisher Scientific), prewashed 3 times with binding buffer, were added to the DNA/antibody complex and incubated overnight at 4 °C on a rotating wheel. After four washes with 800 µl PBS/0.5% Triton X-100, at 4 °C for 10 min each, the beads were eluted using 100 µl of DNA isolation buffer (Diagenode) containing 1 µl proteinase K (20 mg/ml, New England Biolabs). They were incubated for 15 min in a thermomixer at 55 °C at a

mixing speed of 1400 rpm, and then for 15 min at 100 °C. The immunoprecipitated DNA, and 1% of input, were analyzed by qPCR using the comparative CT method. The DNA-RNA hybrid enrichment was calculated based on the IP/input ratio. Primers are listed in **Table 4**.

Chromatin Immunoprecipitation (ChIP) Cells were crosslinked with 1% formaldehyde for 10 min at room temperature (RT) and the reaction was quenched by adding glycine at a final concentration of 0.125M for 5min at RT. Fixed cells were washed and adherent cells were harvested with phosphate-buffered saline (PBS). Cell nuclei were isolated using a cell lysis buffer [5 mM Pipes pH 8.0, 85 mM KCl, 0.5% NP40, Complete protease inhibitor cocktail (Merck)] for 20min, at 4°C. Nuclei were then pelleted by centrifugation and lysed in a nuclei lysis buffer [20 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, Complete protease inhibitor cocktail (Merck)] for 20 min, 4°C. Chromatin was subjected to sonication with a Bioruptor (Diagenode) yielding genomic DNA fragments with a bulk size of 150- 300 bp. Immunoprecipitations were performed overnight at 4°C on 10 μg of chromatin with antibodies pre-bound to Dynabeads protein A/G (ThermoFisher Scientific). The antibodies used are listed in Table 4. The Beads were then washed once with Low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 150 mM NaCl), once with High salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8; 500 mM NaCl), once with LiCl wash buffer (10 mM Tris pH 8.0, 1% Na- deoxycholate, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and twice with TE (10mM Tris pH 8.0, 1mM EDTA pH 8.0). The elution of chromatin was performed in Elution buffer (50mM Tris pH 8.0, 1mM EDTA pH 8.0, 0.2 M NaCl, 1% SDS) at 65°C for 45min and crosslink was reversed O/N at 65°C after addition of NaCl to 0.2M final concentration. The eluted material was digested with 40 μg of Proteinase K (Thermo Fisher Scientific) at 65°C for 2h. DNA was purified using MinElute PCR Purification Kit (Qiagen, 28006), resuspended in H_2O. Quantitative PCRs were performed using the SensiFAST SYBR NoROX Kit (Bioline) and analyzed on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche). Sequences of primers used for the qPCR are listed in Table 3. ChIP-qPCR results were normalized on input signal (% input) and analyzed using the formula:

$2^{-[Ct\ IP-(Ct\ Input-\ log2DF)]}*100$, where DF=dilution factor

Primers are listed in **Table 4**.

Statistical analysis

A two-tailed t-test was used to determine the statistical significance between values presented as percentages.

Table 1. Plasmids used in this study

Indicated modifications were introduced using NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621S) Cloning primers were designed using NEBuilder® Assembly Tool (https://nebuilder.neb.com/)

Plasmid	Comments
dCas9-VP64	Addgene #61425
dCas9-HA	Addgene #102812
pMD2.G	Addgene #12259
psPAX2	Addgene #12260
pEX-A-U6-MiSgRNA_PuroR (MiSgRNA)	Addgene #84781
MS2_MiSgRNA	ACACTGAAAAACACATTCGT spacer were cloned into Addgene # 61424
MS2BP dRNAseH1-V5	Modified from Addgene #89308 and #139835
MS2BP-RNAseH1-V5	Modified from Addgene #89308 and #139836
dCas9-VP64-Cdt1	Modified from Addgene #61425 and #80426
dCas9-VP64-Geminin	Modified from Addgene #61425 and #71707
GOF plasmid	Modified from Invitrogene #V79520 and PCR amplified 4 units of Minor satellites

Table 2. Antibodies used in this study

Name	Supplier	Application and dilution
dCas9, Cas9 (7A9-3A3) Mouse mAb	CST, 14697S	IF (1:900), ChIP (2 µg)
Cenp-A, CENP-A (C51A7) Rabbit mAb (Mouse Specific)	CST, 2048S	IF (1:600)

Cenp-B	Santa Cruz Biotechnology (sc-376283)	IF (1:100)
γ H2A.X, Phospho-Histone H2A.X (Ser139) (D7T2V) Mouse mAb	CST,80312S	IF (1:700), ChIP (2 µg)
PolIIS5ph, Anti-RNA polymerase II subunit B1 (phospho-CTD Ser-5) Antibody, clone 3E8 from rat	Sigma, 04-1572	ChIP (2 µg)
PolIIS2ph, Anti-RNA polymerase II subunit B1 (phosphor-CTD Ser-2) Antibody, clone 3E10, from rat	Sigma, 04-1571	ChIP (2 µg)
S9.6, Anti-DNA-RNA Hybrid Antibody, clone S9.6, from mouse	Sigma, MABE1095	IF (1:200)
S9.6	In-house (Racca *et al.*, 2021)	DRIP (5 µg)
Mouse IgG	Sigma	DRIP (5 µg)

Table 3. DNA and RNA-FISH probes used in this study

Name	Application	Sequence
cenRNA Forward	DNA FISH RNA FISH Immuno-RNA FISH	ACATTCGTTGGAAAC GGGATTTGTAGAACA GTGTATATCAATGAG TTACAA
cenRNA Reverse	RNA FISH	ATGTTTCACATTGTA ACTCATTGATATACA CTGTTCTACAAATCC CGTTTC

Table 4. Primers used in this study

Name	Application	Sequence
U6 Forward	qPCR	CTCGCTTCGGCAGCACA
U6 Reverse	qPCR	AACGCTTCACGAATTTGCGT
Minor Satellite Forward	qPCR	GAACATATTAGATGAGTGAGTTAC
Minor Satellite Reverse	qPCR	GTTCTACAAATCCCGTTTCCAAC

ACKNOWLEDGEMENTS

The authors would like to thank Drs Florent Hubé and Baptiste Bogard for insightful discussions about this work, and Drs Slimane Ait-Si-Ali and Valérie Lallemand-Mezger for critical reading. We acknowledge the imaging platform "EPI2", ddPCR service, and PRG vectorology platform of the Epigenetics and Cell Fate center.

FUNDING

This study was funded by the Agence Nationale pour la Recherche (grant ANR-19-CE12-0022), Ligue Nationale Cnore le Cancer, Fondation ARC and Fondation Maladies Rares. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

P.M., G.V., F.L., and E.B. performed the experiments. P.M, E.B. and C.F. designed the study and supervised the experiments. P.M, and E.B. wrote the manuscript. C.F. acquired the funding.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

Abakir, A., Giles, T.C., Cristini, A., Foster, J.M., Dai, N., Starczak, M., Rubio-Roldan, A., Li, M., Eleftheriou, M., Crutchley, J., *et al.* (2020). N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. Nat Genet *52*, 48–55. https://doi.org/10.1038/s41588-019-0549-x.

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

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*. https://doi.org/10.7554/eLife.05005.

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115–124. https://doi.org/10.1016/j.molcel.2012.04.009.

Aguilera, A., and Gómez-González, B. (2017). DNA-RNA hybrids: the risks of DNA breakage during transcription. Nat Struct Mol Biol *24*, 439–443. https://doi.org/10.1038/nsmb.3395.

Aguilo, F., Zhou, M.-M., and Walsh, M.J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer Res *71*, 5365–5369. https://doi.org/10.1158/0008-5472.CAN-10-4379.

Alfeghaly, C., Aigueperse, C., Maenner, S., and Behm-Ansmant, I. (2021). Non-Coding RNA Silencing in Mammalian Cells by Antisense LNA GapmeRs Transfection. Methods Mol Biol *2300*, 31–37. https://doi.org/10.1007/978-1-0716-1386-3_4.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet *9*, 923–937. https://doi.org/10.1038/nrg2466.

Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., *et al.* (2022). Complete genomic and epigenetic maps of human centromeres. Science *376*, eabl4178. https://doi.org/10.1126/science.abl4178.

Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008). The eukaryotic genome as an RNA machine. Science *319*, 1787–1789. https://doi.org/10.1126/science.1155472.

Amatori, S., Tavolaro, S., Gambardella, S., and Fanelli, M. (2021). The dark side of histones: genomic organization and role of oncohistones in cancer. Clin Epigenetics *13*, 71. https://doi.org/10.1186/s13148-021-01057-x.

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. Nat Genet *51*, 217–223. https://doi.org/10.1038/s41588-018-0306-6.

Aronson, A.I., and McCARTHY, B.J. (1961). Studies of E. coli ribosomal RNA and its degradation products. Biophys J *1*, 215–226. https://doi.org/10.1016/s0006-3495(61)86885-9.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol *2*, e391. https://doi.org/10.1371/journal.pbio.0020391.

Au, S.W.N., Leng, X., Harper, J.W., and Barford, D. (2002). Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit. J Mol Biol *316*, 955–968. https://doi.org/10.1006/jmbi.2002.5399.

Aunan, J.R., Watson, M.M., Hagland, H.R., and Søreide, K. (2016). Molecular and biological hallmarks of ageing. Br J Surg *103*, e29-46. https://doi.org/10.1002/bjs.10053.

Azzaz, A.M., Vitalini, M.W., Thomas, A.S., Price, J.P., Blacketer, M.J., Cryderman, D.E., Zirbel, L.N., Woodcock, C.L., Elcock, A.H., Wallrath, L.L., *et al.* (2014). Human heterochromatin protein 1α promotes nucleosome

associations that drive chromatin condensation. J Biol Chem *289*, 6850–6861. https://doi.org/10.1074/jbc.M113.512137.

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., and Otonkoski, T. (2015). Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. Stem Cell Reports *5*, 448–459. https://doi.org/10.1016/j.stemcr.2015.08.001.

van Balkom, B.W.M., Eisele, A.S., Pegtel, D.M., Bervoets, S., and Verhaar, M.C. (2015). Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting. J Extracell Vesicles *4*, 26760. https://doi.org/10.3402/jev.v4.26760.

Banerjee, P.R., Milin, A.N., Moosa, M.M., Onuchic, P.L., and Deniz, A.A. (2017). Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. Angew Chem Int Ed Engl *56*, 11354–11359. https://doi.org/10.1002/anie.201703191.

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. https://doi.org/10.1038/35065138.

Barik, S. (2004). Development of gene-specific double-stranded RNA drugs. Ann Med *36*, 540–551. https://doi.org/10.1080/07853890410018817.

Barisic, M., and Geley, S. (2011). Spindly switch controls anaphase: spindly and RZZ functions in chromosome attachment and mitotic checkpoint control. Cell Cycle *10*, 449–456. https://doi.org/10.4161/cc.10.3.14759.

Barra, V., and Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat Commun *9*, 4340. https://doi.org/10.1038/s41467-018-06545-y.

Bartolomé, C., Maside, X., and Charlesworth, B. (2002). On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol *19*, 926–937. https://doi.org/10.1093/oxfordjournals.molbev.a004150.

Beacon, T.H., Delcuve, G.P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A.J., and Davie, J.R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clin Epigenetics *13*, 138. https://doi.org/10.1186/s13148-021-01126-1.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet *32*, 29–41. https://doi.org/10.1016/j.tig.2015.11.001.

Beerli, R.R., and Barbas, C.F. (2002). Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol *20*, 135–141. https://doi.org/10.1038/nbt0202-135.

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A *95*, 14628–14633. https://doi.org/10.1073/pnas.95.25.14628.

Beerli, R.R., Schopfer, U., Dreier, B., and Barbas, C.F. (2000). Chemically regulated zinc finger transcription factors. J Biol Chem *275*, 32617–32627. https://doi.org/10.1074/jbc.M005108200.

Belmont, A.S. (2022). Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. Cold Spring Harb Perspect Biol *14*, a041268. https://doi.org/10.1101/cshperspect.a041268.

Belotserkovskii, B.P., Neil, A.J., Saleh, S.S., Shin, J.H.S., Mirkin, S.M., and Hanawalt, P.C. (2013). Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks. Nucleic Acids Res *41*, 1817–1828. https://doi.org/10.1093/nar/gks1333.

Ben-Porath, I., and Cedar, H. (2001). Epigenetic crosstalk. Mol Cell *8*, 933–935. https://doi.org/10.1016/s1097-2765(01)00399-9.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev *23*, 781–783. https://doi.org/10.1101/gad.1787609.

Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., and Earnshaw, W.C. (2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J *30*, 328–340. https://doi.org/10.1038/emboj.2010.329.

Bergmann, J.H., Martins, N.M.C., Larionov, V., Masumoto, H., and Earnshaw, W.C. (2012). HACking the centromere chromatin code: insights from human artificial chromosomes. Chromosome Res *20*, 505–519. https://doi.org/10.1007/s10577-012-9293-0.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol Cell *2*, 437–445. https://doi.org/10.1016/s1097-2765(00)80143-4.

Bickmore, W.A., and Zaret, K.S. (2010). Altered states: how gene expression is changed during differentiation. Curr Opin Genet Dev *20*, 467–469. https://doi.org/10.1016/j.gde.2010.08.003.

Black, B.E., and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Curr Opin Cell Biol *20*, 91–100. https://doi.org/10.1016/j.ceb.2007.11.007.

Black, E.M., and Giunta, S. (2018). Repetitive Fragile Sites: Centromere Satellite DNA As a Source of Genome Instability in Human Diseases. Genes (Basel) *9*, E615. https://doi.org/10.3390/genes9120615.

Blanco-Betancourt, C.E., Moncla, A., Milili, M., Jiang, Y.L., Viegas-Péquignot, E.M., Roquelaure, B., Thuret, I., and Schiff, C. (2004). Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. Blood *103*, 2683–2690. https://doi.org/10.1182/blood-2003-08-2632.

Bloom, K.S. (2014). Centromeric heterochromatin: the primordial segregation machine. Annu Rev Genet *48*, 457–484. https://doi.org/10.1146/annurev-genet-120213-092033.

Bloom, K., and Costanzo, V. (2017). Centromere Structure and Function. Prog Mol Subcell Biol *56*, 515–539. https://doi.org/10.1007/978-3-319-58592-5_21.

Blower, M.D. (2016a). Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep *15*, 1624–1633. https://doi.org/10.1016/j.celrep.2016.04.054.

Blower, M.D. (2016b). Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep *15*, 1624–1633. https://doi.org/10.1016/j.celrep.2016.04.054.

Bobkov, G.O.M., Gilbert, N., and Heun, P. (2018a). Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol *217*, 1957–1972. https://doi.org/10.1083/jcb.201611087.

Bobkov, G.O.M., Gilbert, N., and Heun, P. (2018b). Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol *217*, 1957–1972. https://doi.org/10.1083/jcb.201611087.

Bodak, M., Cirera-Salinas, D., Yu, J., Ngondo, R.P., and Ciaudo, C. (2017). Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells. FEBS Open Bio *7*, 204–220. https://doi.org/10.1002/2211-5463.12174.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods *89*, 123–130. https://doi.org/10.1016/0022-1759(86)90040-2.

Boo, S.H., and Kim, Y.K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. Exp Mol Med *52*, 400–408. https://doi.org/10.1038/s12276-020-0407-z.

Borg, M., Jiang, D., and Berger, F. (2021). Histone variants take center stage in shaping the epigenome. Curr Opin Plant Biol *61*, 101991. https://doi.org/10.1016/j.pbi.2020.101991.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genetics *42*, 1093–1100. https://doi.org/10.1038/ng.708.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96–99. https://doi.org/10.1038/nature02886.

Bouzinba-Segard, H., Guais, A., and Francastel, C. (2006a). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci U S A *103*, 8709–8714. https://doi.org/10.1073/pnas.0508006103.

Bouzinba-Segard, H., Guais, A., and Francastel, C. (2006b). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proceedings of the National Academy of Sciences *103*, 8709–8714. .

Braun, S.M.G., Kirkland, J.G., Chory, E.J., Husmann, D., Calarco, J.P., and Crabtree, G.R. (2017). Rapid and reversible epigenome editing by endogenous chromatin regulators. Nat Commun *8*, 560. https://doi.org/10.1038/s41467-017-00644-y.

Brocken, D.J.W., Tark-Dame, M., and Dame, R.T. (2018). dCas9: A Versatile Tool for Epigenome Editing. Curr Issues Mol Biol *26*, 15–32. https://doi.org/10.21775/cimb.026.015.

Brogna, S., McLeod, T., and Petric, M. (2016). The Meaning of NMD: Translate or Perish. Trends Genet *32*, 395–407. https://doi.org/10.1016/j.tig.2016.04.007.

Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. Cell *150*, 317–326. https://doi.org/10.1016/j.cell.2012.05.035.

Bulut-Karslioglu, A., Perrera, V., Scaranaro, M., de la Rosa-Velazquez, I.A., van de Nobelen, S., Shukeir, N., Popow, J., Gerle, B., Opravil, S., Pagani, M., *et al.* (2012). A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol *19*, 1023–1030. https://doi.org/10.1038/nsmb.2382.

Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., and Cheeseman, I.M. (2020a). Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. Elife *9*, e59770. https://doi.org/10.7554/eLife.59770.

Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., and Cheeseman, I.M. (2020b). Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. Elife *9*. https://doi.org/10.7554/eLife.59770.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol *18*, 299–314. https://doi.org/10.1038/nrm.2016.166.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. J Bacteriol *194*, 5728–5738. https://doi.org/10.1128/JB.01184-12.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., *et al.* (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524–15529. https://doi.org/10.1073/pnas.242606799.

Carmona, J.J., and Michan, S. (2016). Biology of Healthy Aging and Longevity. Rev Invest Clin *68*, 7–16. .

Carpenter, N.J., Filipovich, A., Blaese, R.M., Carey, T.L., and Berkel, A.I. (1988). Variable immunodeficiency with abnormal condensation of the heterochromatin of chromosomes 1, 9, and 16. J Pediatr *112*, 757–760. https://doi.org/10.1016/s0022-3476(88)80698-x.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655. https://doi.org/10.1016/j.cell.2009.01.035.

Casanova, M., Pasternak, M., El Marjou, F., Le Baccon, P., Probst, A.V., and Almouzni, G. (2013). Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. Cell Rep *4*, 1156–1167. https://doi.org/10.1016/j.celrep.2013.08.015.

Casola, C., Hucks, D., and Feschotte, C. (2008). Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol Biol Evol *25*, 29–41. https://doi.org/10.1093/molbev/msm221.

Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell *52*, 583–590. https://doi.org/10.1016/j.molcel.2013.10.006.

Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. Nat Commun *9*, 4346. https://doi.org/10.1038/s41467-018-06677-1.

Chan, F.L., and Wong, L.H. (2012a). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res *40*, 11178–11188. https://doi.org/10.1093/nar/gks921.

Chan, F.L., and Wong, L.H. (2012b). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res *40*, 11178–11188. https://doi.org/10.1093/nar/gks921.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012a). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984. https://doi.org/10.1073/pnas.1108705109.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012b). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984. https://doi.org/10.1073/pnas.1108705109.

Chang, E.Y.-C., and Stirling, P.C. (2017). Replication Fork Protection Factors Controlling R-Loop Bypass and Suppression. Genes (Basel) *8*, E33. https://doi.org/10.3390/genes8010033.

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N.M.C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B.J., Wu, C.-T., *et al.* (2019). Islands of retroelements are major components of *Drosophila* centromeres. PLoS Biol *17*, e3000241. https://doi.org/10.1371/journal.pbio.3000241.

Chardon, F., Japaridze, A., Witt, H., Velikovsky, L., Chakraborty, C., Wilhelm, T., Dumont, M., Yang, W., Kikuti, C., Gangnard, S., *et al.* (2022). CENP-B-mediated DNA loops regulate activity and stability of human centromeres. Mol Cell *82*, 1751-1767.e8. https://doi.org/10.1016/j.molcel.2022.02.032.

Charlet, J., Duymich, C.E., Lay, F.D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., and Jones, P.A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. Mol Cell *62*, 422–431. https://doi.org/10.1016/j.molcel.2016.03.033.

Charpentier, M., Khedher, A.H.Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S., *et al.* (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. Nat Commun *9*, 1133. https://doi.org/10.1038/s41467-018-03475-7.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen *58*, 235–263. https://doi.org/10.1002/em.22087.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992–998. https://doi.org/10.1038/nature02821.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. Nat Methods *12*, 326–328. https://doi.org/10.1038/nmeth.3312.

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet *32*, 828–838. https://doi.org/10.1016/j.tig.2016.10.002.

Cheetham, S.W., Gruhl, F., Mattick, J.S., and Dinger, M.E. (2013). Long noncoding RNAs and the genetics of cancer. Br J Cancer *108*, 2419–2425. https://doi.org/10.1038/bjc.2013.233.

Chiang, H.-C., Zhang, X., Li, J., Zhao, X., Chen, J., Wang, H.T.-H., Jatoi, I., Brenner, A., Hu, Y., and Li, R. (2019). BRCA1-associated R-loop affects transcription and differentiation in breast luminal epithelial cells. Nucleic Acids Res *47*, 5086–5099. https://doi.org/10.1093/nar/gkz262.

Chintalaphani, S.R., Pineda, S.S., Deveson, I.W., and Kumar, K.R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. Acta Neuropathol Commun *9*, 98. https://doi.org/10.1186/s40478-021-01201-x.

Chiodi, I., Corioni, M., Giordano, M., Valgardsdottir, R., Ghigna, C., Cobianchi, F., Xu, R.-M., Riva, S., and Biamonti, G. (2004). RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res *32*, 4127–4136. https://doi.org/10.1093/nar/gkh759.

Choi, E.S., Strålfors, 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. https://doi.org/10.1074/jbc.M111.228510.

Choi, E.S., Cheon, Y., Kang, K., and Lee, D. (2017). The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3. Nat Commun *8*, 529. https://doi.org/10.1038/s41467-017-00704-3.

Chong, Z.-S., Ohnishi, S., Yusa, K., and Wright, G.J. (2018). Pooled extracellular receptor-ligand interaction screening using CRISPR activation. Genome Biol *19*, 205. https://doi.org/10.1186/s13059-018-1581-3.

Choo, K.H.A. (2001). Domain Organization at the Centromere and Neocentromere. Developmental Cell *1*, 165–177. https://doi.org/10.1016/S1534-5807(01)00028-4.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol Cell *40*, 179–204. https://doi.org/10.1016/j.molcel.2010.09.019.

Ciferri, C., Pasqualato, S., Screpanti, E., Varetti, G., Santaguida, S., Dos Reis, G., Maiolica, A., Polka, J., De Luca, J.G., De Wulf, P., *et al.* (2008). Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. Cell *133*, 427–439. https://doi.org/10.1016/j.cell.2008.03.020.

Cimmino, L., and Aifantis, I. (2017). Alternative roles for oxidized mCs and TETs. Curr Opin Genet Dev *42*, 1–7. https://doi.org/10.1016/j.gde.2016.11.003.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223.

Clark, M.B., and Mattick, J.S. (2011). Long noncoding RNAs in cell biology. Semin. Cell Dev. Biol. *22*, 366–376. https://doi.org/10.1016/j.semcdb.2011.01.001.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell *33*, 717–726. https://doi.org/10.1016/j.molcel.2009.01.026.

Cleveland, D.W., Mao, Y., and Sullivan, K.F. (2003). Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell *112*, 407–421. https://doi.org/10.1016/s0092-8674(03)00115-6.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823. https://doi.org/10.1126/science.1231143.

Cooper, J.L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol *21*, 1712–1718. https://doi.org/10.1093/molbev/msh179.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat Struct Mol Biol *11*, 1037–1043. https://doi.org/10.1038/nsmb851.

Costantino, L., and Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol Cell *71*, 487-497.e3. https://doi.org/10.1016/j.molcel.2018.06.037.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599–601. https://doi.org/10.1038/31275.

Cowan, R.K., Hoen, D.R., Schoen, D.J., and Bureau, T.E. (2005). MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol Biol Evol *22*, 2084–2089. https://doi.org/10.1093/molbev/msi202.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138–163. .

Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep *23*, 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025.

Crosio, C., Fimia, G.M., Loury, R., Kimura, M., Okano, Y., Zhou, H., Sen, S., Allis, C.D., and Sassone-Corsi, P. (2002). Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. Mol Cell Biol *22*, 874–885. https://doi.org/10.1128/MCB.22.3.874-885.2002.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019a). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019b). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., *et al.* (2006). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. Nat Genet *38*, 910–916. https://doi.org/10.1038/ng1842.

Cusanelli, E., and Chartrand, P. (2015). Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. Front Genet *6*, 143. https://doi.org/10.3389/fgene.2015.00143.

Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol *15*, 81–94. https://doi.org/10.1038/nrclinonc.2017.166.

Davalos, V., Moutinho, C., Villanueva, A., Boque, R., Silva, P., Carneiro, F., and Esteller, M. (2012). Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. Oncogene *31*, 2062–2074. https://doi.org/10.1038/onc.2011.383.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development *25*, 1010–1022. https://doi.org/10.1101/gad.2037511.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., *et al.* (2011). Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric

Instability, and Facial Anomalies Syndrome Type 2. The American Journal of Human Genetics *88*, 796–804. https://doi.org/10.1016/j.ajhg.2011.04.018.

Deininger, P.L., and Batzer, M.A. (2002). Mammalian retroelements. Genome Res *12*, 1455–1465. https://doi.org/10.1101/gr.282402.

Denegri, M., Chiodi, I., Corioni, M., Cobianchi, F., Riva, S., and Biamonti, G. (2001). Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors. Mol Biol Cell *12*, 3502–3514. https://doi.org/10.1091/mbc.12.11.3502.

Deng, Q., Li, Y., Tedesco, D., Liao, R., Fuhrmann, G., and Sun, P. (2005). The ability of E1A to rescue ras-induced premature senescence and confer transformation relies on inactivation of both p300/CBP and Rb family proteins. Cancer Res *65*, 8298–8307. https://doi.org/10.1158/0008-5472.CAN-05-0054.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P.M. (2009). TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol Cell *35*, 403–413. https://doi.org/10.1016/j.molcel.2009.06.025.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & Development *15*, 2940–2944. .

Di Paolo, A., Racca, C., Calsou, P., and Larminat, F. (2014). Loss of BRCA1 impairs centromeric cohesion and triggers chromosomal instability. FASEB J *28*, 5250–5261. https://doi.org/10.1096/fj.14-250266.

Dimauro, T., and David, G. (2010). Ras-induced senescence and its physiological relevance in cancer. Curr Cancer Drug Targets *10*, 869–876. https://doi.org/10.2174/156800910793357998.

Dinant, C., and Luijsterburg, M.S. (2009). The emerging role of HP1 in the DNA damage response. Mol Cell Biol *29*, 6335–6340. https://doi.org/10.1128/MCB.01048-09.

Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput Biol *4*, e1000176. https://doi.org/10.1371/journal.pcbi.1000176.

Dirac, A.M.G., and Bernards, R. (2003). Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. J Biol Chem *278*, 11731–11734. https://doi.org/10.1074/jbc.C300023200.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol *32*, 1262–1267. https://doi.org/10.1038/nbt.3026.

Drouet, J., Delteil, C., Lefrançois, J., Concannon, P., Salles, B., and Calsou, P. (2005). DNA-dependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks. J Biol Chem *280*, 7060–7069. https://doi.org/10.1074/jbc.M410746200.

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. https://doi.org/10.1371/journal.pgen.1000835.

Du, Z., Zhang, K., and Xie, W. (2022). Epigenetic Reprogramming in Early Animal Development. Cold Spring Harb Perspect Biol *14*, a039677. https://doi.org/10.1101/cshperspect.a039677.

Duda, K.J., Ching, R.W., Jerabek, L., Shukeir, N., Erikson, G., Engist, B., Onishi-Seebacher, M., Perrera, V., Richter, F., Mittler, G., *et al.* (2021). m6A RNA methylation of major satellite repeat transcripts facilitates chromatin association and RNA:DNA hybrid formation in mouse heterochromatin. Nucleic Acids Res *49*, 5568–5587. https://doi.org/10.1093/nar/gkab364.

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. https://doi.org/10.1007/BF00328227.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat Rev Mol Cell Biol *19*, 436–450. https://doi.org/10.1038/s41580-018-0008-z.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447. .

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin Immunol *109*, 17–28. https://doi.org/10.1016/s1521-6616(03)00201-8.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709–2721. https://doi.org/10.1093/nar/10.8.2709.

Ehrlich, M., Jackson, K., and Weemaes, C. (2006). Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet Journal of Rare Diseases *1*, 1. .

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet *29*, 569–574. https://doi.org/10.1016/j.tig.2013.05.010.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev *24*, 1546–1558. https://doi.org/10.1101/gad.573310.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., *et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973. https://doi.org/10.1126/science.1237973.

Enukashvily, N.I., Donev, R., Waisertreiger, I.S.-R., and Podgornaya, O.I. (2007). Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells. Cytogenet Genome Res *118*, 42–54. https://doi.org/10.1159/000106440.

Esteller, M. (2008). Epigenetics in cancer. N Engl J Med *358*, 1148–1159. https://doi.org/10.1056/NEJMra072067.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet *12*, 861–874. https://doi.org/10.1038/nrg3074.

Ewing, A.D., Smits, N., Sanchez-Luque, F.J., Faivre, J., Brennan, P.M., Richardson, S.R., Cheetham, S.W., and Faulkner, G.J. (2020). Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol Cell *80*, 915-928.e5. https://doi.org/10.1016/j.molcel.2020.10.024.

Eymery, Horard, B., El Atifi-Borel, M., Fourel, G., Berger, F., Vitte, A.-L., Van den Broeck, A., Brambilla, E., Fournier, A., Callanan, M., *et al.* (2009). A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Res *37*, 6340–6354. https://doi.org/10.1093/nar/gkp639.

Fachinetti, D., Folco, H.D., 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*, 1056–1066. https://doi.org/10.1038/ncb2805.

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol *20*, 332–338. https://doi.org/10.1038/nsmb.2495.

Faulkner, G.J., and Carninci, P. (2009). Altruistic functions for selfish DNA. Cell Cycle *8*, 2895–2900. https://doi.org/10.4161/cc.8.18.9536.

Felsenfeld, G. (2014). The evolution of epigenetics. Perspect Biol Med *57*, 132–148. https://doi.org/10.1353/pbm.2014.0004.

Feng, J.X., and Riddle, N.C. (2020). Epigenetics and genome stability. Mamm Genome *31*, 181–195. https://doi.org/10.1007/s00335-020-09836-2.

Ferrand, J., Plessier, A., and Polo, S.E. (2021). Control of the chromatin response to DNA damage: Histone proteins pull the strings. Semin Cell Dev Biol *113*, 75–87. https://doi.org/10.1016/j.semcdb.2020.07.002.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell *53*, 49–62. https://doi.org/10.1016/j.molcel.2013.10.030.

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009a). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res *37*, 5071–5080. https://doi.org/10.1093/nar/gkp529.

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009b). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res *37*, 5071–5080. https://doi.org/10.1093/nar/gkp529.

Francastel, C., and Magdinier, F. (2019). DNA methylation in satellite repeats disorders. Essays Biochem *63*, 757–771. https://doi.org/10.1042/EBC20190028.

Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740–741. https://doi.org/10.1038/171740a0.

Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V.K., Patil, N., and Cox, D.R. (2003). Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. Genome Res *13*, 341–346. https://doi.org/10.1101/gr.554603.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol *3*, 722–732. https://doi.org/10.1038/nrmicro1235.

Fryns, J.P., Azou, M., Jaeken, J., Eggermont, E., Pedersen, J.C., and Van den Berghe, H. (1981). Centromeric instability of chromosomes 1, 9, and 16 associated with combined immunodeficiency. Hum Genet *57*, 108–110. https://doi.org/10.1007/BF00271181.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev *15*, 490–495. https://doi.org/10.1016/j.gde.2005.08.002.

Fumagalli, M., Rossiello, F., Mondello, C., and d'Adda di Fagagna, F. (2014). Stable Cellular Senescence Is Associated with Persistent DDR Activation. PLoS ONE *9*, e110969. https://doi.org/10.1371/journal.pone.0110969.

Gaillard, H., and Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. Annu Rev Biochem *85*, 291–317. https://doi.org/10.1146/annurev-biochem-060815-014908.

Gaillard, M.-C., Roche, S., Dion, C., Tasmadjian, A., Bouget, G., Salort-Campana, E., Vovan, C., Chaix, C., Broucqsault, N., Morere, J., *et al.* (2014). Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology *83*, 733–742. https://doi.org/10.1212/WNL.0000000000000708.

Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev *25*, 2041–2056. https://doi.org/10.1101/gad.17010011.

Gao, X.D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S.A., Zhu, L.J., Wolfe, S.A., *et al.* (2018). C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods *15*, 433–436. https://doi.org/10.1038/s41592-018-0006-2.

García-de-Teresa, B., Rodríguez, A., and Frias, S. (2020). Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences. Genes (Basel) *11*, E1528. https://doi.org/10.3390/genes11121528.

García-Muse, T., and Aguilera, A. (2019a). R Loops: From Physiological to Pathological Roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.

García-Muse, T., and Aguilera, A. (2019b). R Loops: From Physiological to Pathological Roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.

Garcia-Perez, J.L., Marchetto, M.C.N., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet *16*, 1569–1577. https://doi.org/10.1093/hmg/ddm105.

García-Pichardo, D., Cañas, J.C., García-Rubio, M.L., Gómez-González, B., Rondón, A.G., and Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. Mol Cell *66*, 597-609.e5. https://doi.org/10.1016/j.molcel.2017.05.014.

Gartel, A.L., and Radhakrishnan, S.K. (2005). Lost in transcription: p21 repression, mechanisms, and consequences. Cancer Res *65*, 3980–3985. https://doi.org/10.1158/0008-5472.CAN-04-3995.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586. https://doi.org/10.1073/pnas.1208507109.

Gaubatz, J.W., and Cutler, R.G. (1990). Mouse satellite DNA is transcribed in senescent cardiac muscle. J Biol Chem *265*, 17753–17758. .

van Gent, D.C., Mizuuchi, K., and Gellert, M. (1996). Similarities between initiation of V(D)J recombination and retroviral integration. Science *271*, 1592–1594. https://doi.org/10.1126/science.271.5255.1592.

George, C.X., John, L., and Samuel, C.E. (2014). An RNA editor, adenosine deaminase acting on double-stranded RNA (ADAR1). J Interferon Cytokine Res *34*, 437–446. https://doi.org/10.1089/jir.2014.0001.

Ghosh, S., and Ghosh, A. (2021). Activation of DNA damage response signaling in mammalian cells by ionizing radiation. Free Radic Res *55*, 581–594. https://doi.org/10.1080/10715762.2021.1876853.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814–825. https://doi.org/10.1016/j.molcel.2012.01.017.

Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., *et al.* (2021a). CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc Natl Acad Sci U S A *118*, e2015634118. https://doi.org/10.1073/pnas.2015634118.

Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., *et al.* (2021b). CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc Natl Acad Sci U S A *118*, e2015634118. https://doi.org/10.1073/pnas.2015634118.

Goll, M.G., and Bestor, T.H. (2005). EUKARYOTIC CYTOSINE METHYLTRANSFERASES. Annual Review of Biochemistry *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721.

Gollin, S.M. (2004). Chromosomal instability. Curr Opin Oncol *16*, 25–31. https://doi.org/10.1097/00001622-200401000-00006.

Gómez-González, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc Natl Acad Sci U S A *104*, 8409–8414. https://doi.org/10.1073/pnas.0702836104.

Goutte-Gattat, D., Shuaib, M., Ouararhni, K., Gautier, T., Skoufias, D.A., Hamiche, A., and Dimitrov, S. (2013). Phosphorylation of the CENP-A amino-terminus in mitotic centromeric chromatin is required for kinetochore function. Proc Natl Acad Sci U S A *110*, 8579–8584. https://doi.org/10.1073/pnas.1302955110.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol *20*, 590–607. https://doi.org/10.1038/s41580-019-0159-6.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331–337. https://doi.org/10.1038/337331a0.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res *35*, W52-57. https://doi.org/10.1093/nar/gkm360.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet *6*, e1000869. https://doi.org/10.1371/journal.pgen.1000869.

Grunseich, C., Wang, I.X., Watts, J.A., Burdick, J.T., Guber, R.D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A.B., *et al.* (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. Mol Cell *69*, 426-437.e7. https://doi.org/10.1016/j.molcel.2017.12.030.

Guenatri, M. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. The Journal of Cell Biology *166*, 493–505. https://doi.org/10.1083/jcb.200403109.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res *22*, 633–641. https://doi.org/10.1101/gr.130997.111.

Guldner, H.H., Lakomek, H.J., and Bautz, F.A. (1984). Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. Clin Exp Immunol *58*, 13–20. .

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., *et al.* (2013). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Neuroscience *17*, 215–222. https://doi.org/10.1038/nn.3607.

Gurard-Levin, Z.A., and Almouzni, G. (2014). Histone modifications and a choice of variant: a language that helps the genome express itself. F1000Prime Rep *6*, 76. https://doi.org/10.12703/P6-76.

Gutbrod, M.J., Roche, B., Steinberg, J.I., Lakhani, A.A., Chang, K., Schorn, A.J., and Martienssen, R.A. (2022). Dicer promotes genome stability via the bromodomain transcriptional co-activator BRD4. Nat Commun *13*, 1001. https://doi.org/10.1038/s41467-022-28554-8.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., *et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369. https://doi.org/10.1016/j.stem.2013.06.002.

Hackett, J.A., and Surani, M.A. (2013). DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond B Biol Sci *368*, 20110328. https://doi.org/10.1098/rstb.2011.0328.

Hagan, C.R., and Rudin, C.M. (2007). DNA cleavage and Trp53 differentially affect SINE transcription. Genes Chromosomes Cancer *46*, 248–260. https://doi.org/10.1002/gcc.20406.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. Science *319*, 1352–1355. https://doi.org/10.1126/science.1140735.

Hall, L.E., Mitchell, S.E., and O'Neill, R.J. (2012). Pericentric and centromeric transcription: a perfect balance required. Chromosome Res *20*, 535–546. https://doi.org/10.1007/s10577-012-9297-9.

Hall, L.L., Byron, M., Carone, D.M., Whitfield, T.W., Pouliot, G.P., Fischer, A., Jones, P., and Lawrence, J.B. (2017). Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. Cell Rep *18*, 2943–2956. https://doi.org/10.1016/j.celrep.2017.02.072.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786.e19. https://doi.org/10.1016/j.cell.2017.07.043.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

Hansen, R.S., Stöger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D'Esposito, M., Feil, R., Gimelli, G., *et al.* (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol. Genet. *9*, 2575–2587. .

Hara, M., and Fukagawa, T. (2017). Critical Foundation of the Kinetochore: The Constitutive Centromere-Associated Network (CCAN). Prog Mol Subcell Biol *56*, 29–57. https://doi.org/10.1007/978-3-319-58592-5_2.

Harashima, H., Dissmeyer, N., and Schnittger, A. (2013). Cell cycle control across the eukaryotic kingdom. Trends Cell Biol *23*, 345–356. https://doi.org/10.1016/j.tcb.2013.03.002.

Harris, C.R., Dewan, A., Zupnick, A., Normart, R., Gabriel, A., Prives, C., Levine, A.J., and Hoh, J. (2009). p53 responsive elements in human retrotransposons. Oncogene *28*, 3857–3865. https://doi.org/10.1038/onc.2009.246.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chédin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol *430*, 272–284. https://doi.org/10.1016/j.jmb.2017.12.016.

Hauer, M.H., and Gasser, S.M. (2017). Chromatin and nucleosome dynamics in DNA damage and repair. Genes Dev *31*, 2204–2221. https://doi.org/10.1101/gad.307702.117.

Hayflick, L. (1965). THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. Exp Cell Res *37*, 614–636. https://doi.org/10.1016/0014-4827(65)90211-9.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp Cell Res *25*, 585–621. https://doi.org/10.1016/0014-4827(61)90192-6.

Héberlé, É., and Bardet, A.F. (2019). Sensitivity of transcription factors to DNA methylation. Essays Biochem *63*, 727–741. https://doi.org/10.1042/EBC20190033.

Hédouin, S., Grillo, G., Ivkovic, I., Velasco, G., and Francastel, C. (2017a). CENP-A chromatin disassembly in stressed and senescent murine cells. Sci Rep *7*, 42520. https://doi.org/10.1038/srep42520.

Hédouin, S., Grillo, G., Ivkovic, I., Velasco, G., and Francastel, C. (2017b). CENP-A chromatin disassembly in stressed and senescent murine cells. Sci Rep *7*, 42520. https://doi.org/10.1038/srep42520.

Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R.A. (2008). DNA repair pathways as targets for cancer therapy. Nat Rev Cancer *8*, 193–204. https://doi.org/10.1038/nrc2342.

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

Hergeth, S.P., and Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep *16*, 1439–1453. https://doi.org/10.15252/embr.201540749.

Hiom, K., Melek, M., and Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell *94*, 463–470. https://doi.org/10.1016/s0092-8674(00)81587-1.

Hoffmann, S., Dumont, M., Barra, V., Ly, P., Nechemia-Arbely, Y., McMahon, M.A., Hervé, S., Cleveland, D.W., and Fachinetti, D. (2016). CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. Cell Rep *17*, 2394–2404. https://doi.org/10.1016/j.celrep.2016.10.084.

Holland, A.J., Fachinetti, D., Han, J.S., and Cleveland, D.W. (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A *109*, E3350-3357. https://doi.org/10.1073/pnas.1216880109.

Howden, S.E., McColl, B., Glaser, A., Vadolas, J., Petrou, S., Little, M.H., Elefanty, A.G., and Stanley, E.G. (2016). A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. Stem Cell Reports *7*, 508–517. https://doi.org/10.1016/j.stemcr.2016.07.001.

Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., de Lima, L.G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., *et al.* (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science *376*, eabk3112. https://doi.org/10.1126/science.abk3112.

Hraiky, C., Raymond, M.A., and Drolet, M. (2000). RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in Escherichia coli. J Biol Chem *275*, 11257–11263. https://doi.org/10.1074/jbc.275.15.11257.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell *12*, 711–721. https://doi.org/10.1016/j.molcel.2003.08.010.

Hyjek, M., Figiel, M., and Nowotny, M. (2019a). RNases H: Structure and mechanism. DNA Repair (Amst) *84*, 102672. https://doi.org/10.1016/j.dnarep.2019.102672.

Hyjek, M., Figiel, M., and Nowotny, M. (2019b). RNases H: Structure and mechanism. DNA Repair (Amst) *84*, 102672. https://doi.org/10.1016/j.dnarep.2019.102672.

Ideue, T., Cho, Y., Nishimura, K., and Tani, T. (2014). Involvement of satellite I noncoding RNA in regulation of chromosome segregation. Genes Cells *19*, 528–538. https://doi.org/10.1111/gtc.12149.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands - 'A rough guide.' FEBS Letters *583*, 1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012.

Innocente, S.A., Abrahamson, J.L., Cogswell, J.P., and Lee, J.M. (1999). p53 regulates a G2 checkpoint through cyclin B1. Proc Natl Acad Sci U S A *96*, 2147–2152. https://doi.org/10.1073/pnas.96.5.2147.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303. https://doi.org/10.1126/science.1210597.

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc Natl Acad Sci U S A *77*, 2450–2454. https://doi.org/10.1073/pnas.77.5.2450.

Iyama, T., and Wilson, D.M. (2013). DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) *12*, 620–636. https://doi.org/10.1016/j.dnarep.2013.04.015.

Jaco, I., Canela, A., Vera, E., and Blasco, M.A. (2008). Centromere mitotic recombination in mammalian cells. J Cell Biol *181*, 885–892. https://doi.org/10.1083/jcb.200803042.

Jakob, B., Splinter, J., Conrad, S., Voss, K.-O., Zink, D., Durante, M., Löbrich, M., and Taucher-Scholz, G. (2011). DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. Nucleic Acids Res *39*, 6489–6499. https://doi.org/10.1093/nar/gkr230.

Jansen, L.E.T., 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. https://doi.org/10.1083/jcb.200701066.

Jeffries, M.A. (2020). The Development of Epigenetics in the Study of Disease Pathogenesis. Adv Exp Med Biol *1253*, 57–94. https://doi.org/10.1007/978-981-15-3449-2_2.

Jenness, C., Giunta, S., Müller, M.M., Kimura, H., Muir, T.W., and Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci U S A *115*, E876–E885. https://doi.org/10.1073/pnas.1717509115.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080. https://doi.org/10.1126/science.1063127.

Jiang, F., and Doudna, J.A. (2015). The structural biology of CRISPR-Cas systems. Curr Opin Struct Biol *30*, 100–111. https://doi.org/10.1016/j.sbi.2015.02.002.

Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hulton, M., Jonveaux, P., Maraschio, P., Mogarbano, A., *et al.* (2005). DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Human Mutation *25*, 56–63. https://doi.org/10.1002/humu.20113.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821. https://doi.org/10.1126/science.1225829.

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B.M., Khochbin, S., and Vourc'h, C. (2004). Stress-induced transcription of satellite III repeats. J Cell Biol *164*, 25–33. https://doi.org/10.1083/jcb.200306104.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484–492. https://doi.org/10.1038/nrg3230.

Jursch, T., Miskey, C., Izsvák, Z., and Ivics, Z. (2013). Regulation of DNA transposition by CpG methylation and chromatin structure in human cells. Mob DNA *4*, 15. https://doi.org/10.1186/1759-8753-4-15.

Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018a). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science *359*, 108–114. https://doi.org/10.1126/science.aan6490.

Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018b). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science *359*, 108–114. https://doi.org/10.1126/science.aan6490.

Kaer, K., and Speek, M. (2013). Retroelements in human disease. Gene *518*, 231–241. https://doi.org/10.1016/j.gene.2013.01.008.

Kalitsis, P., Griffiths, B., and Choo, K.H.A. (2006a). Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A *103*, 8786–8791. https://doi.org/10.1073/pnas.0600250103.

Kalitsis, P., Griffiths, B., and Choo, K.H.A. (2006b). Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A *103*, 8786–8791. https://doi.org/10.1073/pnas.0600250103.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489–501. https://doi.org/10.1101/gad.1248505.

Kang, H.J., Cheon, N.Y., Park, H., Jeong, G.W., Ye, B.J., Yoo, E.J., Lee, J.H., Hur, J.-H., Lee, E.-A., Kim, H., *et al.* (2021). TonEBP recognizes R-loops and initiates m6A RNA methylation for R-loop resolution. Nucleic Acids Res *49*, 269–284. https://doi.org/10.1093/nar/gkaa1162.

Kapranov, P., Ozsolak, F., and Milos, P.M. (2012). Profiling of short RNAs using Helicos single-molecule sequencing. Methods Mol Biol *822*, 219–232. https://doi.org/10.1007/978-1-61779-427-8_15.

Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. Mol Biol Evol *35*, 949–962. https://doi.org/10.1093/molbev/msy010.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469. https://doi.org/10.1126/science.1114955.

Kawaji, H., and Hayashizaki, Y. (2008). Exploration of small RNAs. PLoS Genet *4*, e22. https://doi.org/10.1371/journal.pgen.0040022.

Kazazian, H.H. (2011). Mobile DNA transposition in somatic cells. BMC Biol *9*, 62. https://doi.org/10.1186/1741-7007-9-62.

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223. https://doi.org/10.1242/dev.103341.

Keegan, L., Khan, A., Vukic, D., and O'Connell, M. (2017). ADAR RNA editing below the backbone. RNA *23*, 1317–1328. https://doi.org/10.1261/rna.060921.117.

Kiaee, F., Zaki-Dizaji, M., Hafezi, N., Almasi-Hashiani, A., Hamedifar, H., Sabzevari, A., Shirkani, A., Zian, Z., Jadidi-Niaragh, F., Aghamahdi, F., *et al.* (2021). Clinical, Immunologic and Molecular Spectrum of Patients with Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome: A Systematic Review. Endocr Metab Immune Disord Drug Targets *21*, 664–672. https://doi.org/10.2174/1871530320666200613204426.

Kim, A., and Wang, G.G. (2021). R-loop and its functions at the regulatory interfaces between transcription and (epi)genome. Biochim Biophys Acta Gene Regul Mech *1864*, 194750. https://doi.org/10.1016/j.bbagrm.2021.194750.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126–139. https://doi.org/10.1038/nrm2632.

Kipling, D., and Warburton, P.E. (1997). Centromeres, CENP-B and Tigger too. Trends Genet *13*, 141–145. https://doi.org/10.1016/s0168-9525(97)01098-6.

Kireeva, M.L., Komissarova, N., and Kashlev, M. (2000). Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. J Mol Biol *299*, 325–335. https://doi.org/10.1006/jmbi.2000.3755.

Kisseljova, N.P., Zueva, E.S., Pevzner, V.S., Grachev, A.N., and Kisseljov, F.L. (1998). De novo methylation of selective CpG dinucleotide clusters in transformed cells mediated by an activated N-ras. Int J Oncol *12*, 203–209. .

Kixmoeller, K., Allu, P.K., and Black, B.E. (2020a). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol *10*, 200051. https://doi.org/10.1098/rsob.200051.

Kixmoeller, K., Allu, P.K., and Black, B.E. (2020b). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol *10*, 200051. https://doi.org/10.1098/rsob.200051.

Klawitter, S., Fuchs, N.V., Upton, K.R., Muñoz-Lopez, M., Shukla, R., Wang, J., Garcia-Cañadas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A., *et al.* (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat Commun *7*, 10286. https://doi.org/10.1038/ncomms10286.

Kleinsmith, L.J., Allfrey, V.G., and Mirsky, A.E. (1966). Phosphorylation of nuclear protein early in the course of gene activation in lymphocytes. Science *154*, 780–781. https://doi.org/10.1126/science.154.3750.780.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol *18*, 490–495. https://doi.org/10.1016/j.cub.2008.03.016.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem *79*, 213–231. https://doi.org/10.1146/annurev-biochem-010909-095056.

Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. Nat Rev Cancer *1*, 157–162. https://doi.org/10.1038/35101031.

Kondo, T., Bobek, M.P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., Bourc'his, D., Viegas-Péquignot, E., Ehrlich, M., and Hanash, S.M. (2000). Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet *9*, 597–604. https://doi.org/10.1093/hmg/9.4.597.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015a). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015b). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. PLoS Genet *7*, e1002384. https://doi.org/10.1371/journal.pgen.1002384.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651–669. https://doi.org/10.1534/genetics.112.146704.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol *8*, R61. https://doi.org/10.1186/gb-2007-8-4-r61.

Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Honda, S., Kobayashi, O., Hatakeyama, K., Ushio, Y., Saya, H., and Hirota, T. (2003). CENP-A phosphorylation by Aurora-A in prophase is required for enrichment of Aurora-B at inner centromeres and for kinetochore function. Dev Cell *5*, 853–864. https://doi.org/10.1016/s1534-5807(03)00364-2.

Kuo, Y.-T., Ishii, T., Fuchs, J., Hsieh, W.-H., Houben, A., and Lin, Y.-R. (2021). The Evolutionary Dynamics of Repetitive DNA and Its Impact on the Genome Diversification in the Genus Sorghum. Front Plant Sci *12*, 729734. https://doi.org/10.3389/fpls.2021.729734.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol *32*, 677–683. https://doi.org/10.1038/nbt.2916.

Kwon, S.H., and Workman, J.L. (2011). The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. Bioessays *33*, 280–289. https://doi.org/10.1002/bies.201000138.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921. https://doi.org/10.1038/35057062.

Lara-Gonzalez, P., Pines, J., and Desai, A. (2021). Spindle assembly checkpoint activation and silencing at kinetochores. Semin Cell Dev Biol *117*, 86–98. https://doi.org/10.1016/j.semcdb.2021.06.009.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240. https://doi.org/10.1038/nature22822.

Lawrimore, J., and Bloom, K. (2019). The regulation of chromosome segregation via centromere loops. Crit Rev Biochem Mol Biol *54*, 352–370. https://doi.org/10.1080/10409238.2019.1670130.

Lee, Y.W., Arora, R., Wischnewski, H., and Azzalin, C.M. (2018). TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. Nat Struct Mol Biol *25*, 147–153. https://doi.org/10.1038/s41594-017-0021-5.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current Biology *13*, 1192–1200. .

Leonova, K.I., Brodsky, L., Lipchick, B., Pal, M., Novototskaya, L., Chenchik, A.A., Sen, G.C., Komarova, E.A., and Gudkov, A.V. (2013). p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. Proc Natl Acad Sci U S A *110*, E89-98. https://doi.org/10.1073/pnas.1216922110.

Li, M., and Fang, Y. (2015). Histone variants: the artists of eukaryotic chromatin. Sci China Life Sci *58*, 232–239. https://doi.org/10.1007/s11427-015-4817-4.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378. https://doi.org/10.1016/j.cell.2005.06.008.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926. https://doi.org/10.1016/0092-8674(92)90611-f.

Li, J., Sun, H., Huang, Y., Wang, Y., Liu, Y., and Chen, X. (2019). Pathways and assays for DNA double-strand break repair by homologous recombination. Acta Biochim Biophys Sin (Shanghai) *51*, 879–889. https://doi.org/10.1093/abbs/gmz076.

Li, M., Dong, Q., and Zhu, B. (2017). Aurora Kinase B Phosphorylates Histone H3.3 at Serine 31 during Mitosis in Mammalian Cells. J Mol Biol *429*, 2042–2045. https://doi.org/10.1016/j.jmb.2017.01.016.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131.

Lim, Y.W., Sanz, L.A., Xu, X., Hartono, S.R., and Chédin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. Elife *4*. https://doi.org/10.7554/eLife.08007.

Lindström, M.S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., and Volarevic, S. (2018). Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. Oncogene *37*, 2351–2366. https://doi.org/10.1038/s41388-017-0121-z.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322. https://doi.org/10.1038/nature08514.

Liu, S.J., Dang, H.X., Lim, D.A., Feng, F.Y., and Maher, C.A. (2021). Long noncoding RNAs in cancer metastasis. Nat Rev Cancer *21*, 446–460. https://doi.org/10.1038/s41568-021-00353-1.

Locke, D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D., and Eichler, E.E. (2003). Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res *13*, 347–357. https://doi.org/10.1101/gr.1003303.

Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., *et al.* (2021). The structure, function and evolution of a complete human chromosome 8. Nature *593*, 101–107. https://doi.org/10.1038/s41586-021-03420-7.

López de Silanes, I., Stagno d'Alcontres, M., and Blasco, M.A. (2010). TERRA transcripts are bound by a complex array of RNA-binding proteins. Nat Commun *1*, 33. https://doi.org/10.1038/ncomms1032.

Lopez-Bergami, P., Lau, E., and Ronai, Z. (2010). Emerging roles of ATF2 and the dynamic AP1 network in cancer. Nat Rev Cancer *10*, 65–76. https://doi.org/10.1038/nrc2681.

Lu, J., and Gilbert, D.M. (2008). Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence? Cell Cycle *7*, 1907–1910. https://doi.org/10.4161/cc.7.13.6206.

Lu, T.X., and Rothenberg, M.E. (2018). MicroRNA. J Allergy Clin Immunol *141*, 1202–1207. https://doi.org/10.1016/j.jaci.2017.08.034.

Luger, K., Mäder, 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. https://doi.org/10.1038/38444.

Luke, B., and Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. EMBO J *28*, 2503–2510. https://doi.org/10.1038/emboj.2009.166.

Ma, J., Setton, J., Lee, N.Y., Riaz, N., and Powell, S.N. (2018). The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nat Commun *9*, 3292. https://doi.org/10.1038/s41467-018-05228-y.

Madaan, M., and Mendez, M.D. (2022). Angelman Syndrome. In StatPearls, (Treasure Island (FL): StatPearls Publishing), p.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J *417*, 639–650. https://doi.org/10.1042/BJ20080413.

Mailand, N., Falck, J., Lukas, C., Syljuâsen, R.G., Welcker, M., Bartek, J., and Lukas, J. (2000). Rapid destruction of human Cdc25A in response to DNA damage. Science *288*, 1425–1429. https://doi.org/10.1126/science.288.5470.1425.

Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nature Genetics *30*, 329–334. https://doi.org/10.1038/ng843.

Maison, C., Quivy, J.-P., Probst, A.V., and Almouzni, G. (2010). Heterochromatin at mouse pericentromeres: a model for de novo heterochromatin formation and duplication during replication. Cold Spring Harb Symp Quant Biol *75*, 155–165. https://doi.org/10.1101/sqb.2010.75.013.

Maison, C., Bailly, D., Roche, D., Montes de Oca, R., Probst, A.V., Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., *et al.* (2011). SUMOylation promotes de novo targeting of HP1α to pericentric heterochromatin. Nat Genet *43*, 220–227. https://doi.org/10.1038/ng.765.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826. https://doi.org/10.1126/science.1232033.

Malik, H.S., and Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. Cell *138*, 1067–1082. https://doi.org/10.1016/j.cell.2009.08.036.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656–668. https://doi.org/10.1016/j.cell.2009.01.045.

Malumbres, M. (2014). Cyclin-dependent kinases. Genome Biol *15*, 122. https://doi.org/10.1186/gb4184.

Manic, G., Corradi, F., Sistigu, A., Siteni, S., and Vitale, I. (2017). Molecular Regulation of the Spindle Assembly Checkpoint by Kinases and Phosphatases. Int Rev Cell Mol Biol *328*, 105–161. https://doi.org/10.1016/bs.ircmb.2016.08.004.

Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol *13*, 95–101. https://doi.org/10.1038/ncb2140.

Maraschio, P., Zuffardi, O., Dalla Fior, T., and Tiepolo, L. (1988). Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. Journal of Medical Genetics *25*, 173–180. .

Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J. (1994). Krüppel-associated boxes are potent transcriptional repression domains. Proc Natl Acad Sci U S A *91*, 4509–4513. https://doi.org/10.1073/pnas.91.10.4509.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO Journal *24*, 800–812. .

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. https://doi.org/10.1083/jcb.109.5.1963.

Mathew, V., Pauleau, A.-L., Steffen, N., Bergner, A., Becker, P.B., and Erhardt, S. (2014). The histone-fold protein CHRAC14 influences chromatin composition in response to DNA damage. Cell Rep *7*, 321–330. https://doi.org/10.1016/j.celrep.2014.03.008.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum Mol Genet *15 Spec No 1*, R17-29. https://doi.org/10.1093/hmg/ddl046.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272–1278. https://doi.org/10.1038/nsmb.1903.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502. https://doi.org/10.1038/35000656.

Maze, I., Noh, K.-M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat Rev Genet *15*, 259–271. https://doi.org/10.1038/nrg3673.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol *17*, 16–29. https://doi.org/10.1038/nrm.2015.5.

McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017a). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Dev Cell *42*, 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001.

McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017b). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Developmental Cell *42*, 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001.

Mellone, B.G., and Fachinetti, D. (2021). Diverse mechanisms of centromere specification. Curr Biol *31*, R1491–R1504. https://doi.org/10.1016/j.cub.2021.09.083.

Mendiola, A.J.P., and LaSalle, J.M. (2021). Epigenetics in Prader-Willi Syndrome. Front Genet *12*, 624581. https://doi.org/10.3389/fgene.2021.624581.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet *10*, 155–159. https://doi.org/10.1038/nrg2521.

Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albà, M.M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics *18*, 333–334. https://doi.org/10.1093/bioinformatics/18.2.333.

Metz, A., Soret, J., Vourc'h, C., Tazi, J., and Jolly, C. (2004). A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci *117*, 4551–4558. https://doi.org/10.1242/jcs.01329.

Michaux, C., Verneuil, N., Hartke, A., and Giard, J.-C. (2014). Physiological roles of small RNA molecules. Microbiology (Reading) *160*, 1007–1019. https://doi.org/10.1099/mic.0.076208-0.

Miga, K.H., Newton, Y., Jain, M., Altemose, N., Willard, H.F., and Kent, W.J. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res *24*, 697–707. https://doi.org/10.1101/gr.159624.113.

Mihìc, P., Hédouin, S., and Francastel, C. (2021a). Centromeres Transcription and Transcripts for Better and for Worse. Prog Mol Subcell Biol *60*, 169–201. https://doi.org/10.1007/978-3-030-74889-0_7.

Mihìc, P., Hédouin, S., and Francastel, C. (2021b). Centromeres Transcription and Transcripts for Better and for Worse. Prog Mol Subcell Biol *60*, 169–201. https://doi.org/10.1007/978-3-030-74889-0_7.

Mikwar, M., MacFarlane, A.J., and Marchetti, F. (2020). Mechanisms of oocyte aneuploidy associated with advanced maternal age. Mutat Res Rev Mutat Res *785*, 108320. https://doi.org/10.1016/j.mrrev.2020.108320.

Millband, D.N., Campbell, L., and Hardwick, K.G. (2002). The awesome power of multiple model systems: interpreting the complex nature of spindle checkpoint signaling. Trends Cell Biol *12*, 205–209. https://doi.org/10.1016/s0962-8924(02)02276-6.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J *4*, 1609–1614. https://doi.org/10.1002/j.1460-2075.1985.tb03825.x.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol Cell *41*, 21–32. https://doi.org/10.1016/j.molcel.2010.12.007.

Mishra, P.K., Chakraborty, A., Yeh, E., Feng, W., Bloom, K.S., and Basrai, M.A. (2021). R-loops at centromeric chromatin contribute to defects in kinetochore integrity and chromosomal instability in budding yeast. Mol Biol Cell *32*, 74–89. https://doi.org/10.1091/mbc.E20-06-0379.

Misteli, T. (2001). The concept of self-organization in cellular architecture. J Cell Biol *155*, 181–185. https://doi.org/10.1083/jcb.200108110.

Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell *183*, 28–45. https://doi.org/10.1016/j.cell.2020.09.014.

Mitchell, A.R., Jeppesen, P., Nicol, L., Morrison, H., and Kipling, D. (1996). Epigenetic control of mammalian centromere protein binding: does DNA methylation have a role? Journal of Cell Science *109*, 2199–2206. .

Miyashita, T., Kitada, S., Krajewski, S., Horne, W.A., Delia, D., and Reed, J.C. (1995). Overexpression of the Bcl-2 protein increases the half-life of p21Bax. J Biol Chem *270*, 26049–26052. https://doi.org/10.1074/jbc.270.44.26049.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755–766. https://doi.org/10.1016/j.molcel.2008.05.007.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading) *155*, 733–740. https://doi.org/10.1099/mic.0.023960-0.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res *20*, 899–907. https://doi.org/10.1101/gr.103473.109.

Moran, E.C., Liu, L., Zasadzinska, E., Kestner, C.A., Sarkeshik, A., DeHoyos, H., Yates, J.R., Foltz, D., and Stukenberg, P.T. (2021a). Mitotic R-loops direct Aurora B kinase to maintain centromeric cohesion (Cell Biology).

Moran, E.C., Liu, L., Zasadzinska, E., Kestner, C.A., Sarkeshik, A., DeHoyos, H., Yates, J.R., Foltz, D., and Stukenberg, P.T. (2021b). Mitotic R-loops direct Aurora B kinase to maintain centromeric cohesion. BioRxiv 2021.01.14.426738. https://doi.org/10.1101/2021.01.14.426738.

Moser, S.C., and Swedlow, J.R. (2011). How to be a mitotic chromosome. Chromosome Res *19*, 307–319. https://doi.org/10.1007/s10577-011-9198-3.

Mullenders, L.H.F. (2018). Solar UV damage to cellular DNA: from mechanisms to biological effects. Photochem Photobiol Sci *17*, 1842–1852. https://doi.org/10.1039/c8pp00182k.

Musacchio, A., and Desai, A. (2017). A Molecular View of Kinetochore Assembly and Function. Biology (Basel) *6*, E5. https://doi.org/10.3390/biology6010005.

Musacchio, A., and Hardwick, K.G. (2002). The spindle checkpoint: structural insights into dynamic signalling. Nat Rev Mol Cell Biol *3*, 731–741. https://doi.org/10.1038/nrm929.

Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol *19*, 1218–1227. https://doi.org/10.1038/nsmb.2436.

Nachtergaele, S., and He, C. (2017). The emerging biology of RNA post-transcriptional modifications. RNA Biol *14*, 156–163. https://doi.org/10.1080/15476286.2016.1267096.

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J *31*, 4020–4034. https://doi.org/10.1038/emboj.2012.251.

Nakano, K., and Vousden, K.H. (2001). PUMA, a novel proapoptotic gene, is induced by p53. Mol Cell *7*, 683–694. https://doi.org/10.1016/s1097-2765(01)00214-3.

Nasmyth, K. (2002). Segregating sister genomes: the molecular biology of chromosome separation. Science *297*, 559–565. https://doi.org/10.1126/science.1074757.

Nasmyth, K., Peters, J.M., and Uhlmann, F. (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. Science *288*, 1379–1385. https://doi.org/10.1126/science.288.5470.1379.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953. https://doi.org/10.1126/science.1236083.

Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA *12*, 1802–1809. https://doi.org/10.1261/rna.165106.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77. https://doi.org/10.1038/nature21373.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Mol Cell *65*, 832-847.e4. https://doi.org/10.1016/j.molcel.2017.01.029.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods *6*, 917–922. https://doi.org/10.1038/nmeth.1401.

Nishitani, H., Lygerou, Z., and Nishimoto, T. (2004). Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of geminin through its N-terminal region. J Biol Chem *279*, 30807–30816. https://doi.org/10.1074/jbc.M312644200.

Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., *et al.* (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. Journal of Human Genetics *58*, 455–460. .

Novo, C.L., Wong, E.V., Hockings, C., Poudel, C., Sheekey, E., Wiese, M., Okkenhaug, H., Boulton, S.J., Basu, S., Walker, S., *et al.* (2022). Satellite repeat transcripts modulate heterochromatin condensates and safeguard chromosome stability in mouse embryonic stem cells. Nat Commun *13*, 3525. https://doi.org/10.1038/s41467-022-31198-3.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci *364*, 99–115. https://doi.org/10.1098/rstb.2008.0168.

Ohtani, N., Zebedee, Z., Huot, T.J., Stinson, J.A., Sugimoto, M., Ohashi, Y., Sharrocks, A.D., Peters, G., and Hara, E. (2001). Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. Nature *409*, 1067–1070. https://doi.org/10.1038/35059131.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257. https://doi.org/10.1016/s0092-8674(00)81656-6.

Ord, M.G., and Stocken, L.A. (1967). Phosphate and thiol groups in histone f3 from rat liver and thymus nuclei. Biochem J *102*, 631–636. https://doi.org/10.1042/bj1020631.

Oroz, J., and Laurents, D.V. (2019). RNA binding proteins: Diversity from microsurgeons to cowboys. Biochim Biophys Acta Gene Regul Mech *1862*, 194398. https://doi.org/10.1016/j.bbagrm.2019.06.009.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475–478. https://doi.org/10.1016/s0960-9822(00)00448-6.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Curr Opin Genet Dev *31*, 12–19. https://doi.org/10.1016/j.gde.2015.03.009.

Pal, S., and Tyler, J.K. (2016). Epigenetics and aging. Sci Adv *2*, e1600584. https://doi.org/10.1126/sciadv.1600584.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Front Genet *6*, 2. https://doi.org/10.3389/fgene.2015.00002.

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. https://doi.org/10.1073/pnas.88.9.3734.

Pannunzio, N.R., Watanabe, G., and Lieber, M.R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J Biol Chem *293*, 10512–10523. https://doi.org/10.1074/jbc.TM117.000374.

Papin, C., Ibrahim, A., Gras, S.L., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S., and Hamiche, A. (2017). Combinatorial DNA methylation codes at repetitive elements. Genome Res *27*, 934–946. https://doi.org/10.1101/gr.213983.116.

Pappalardo, X.G., and Barra, V. (2021). Losing DNA methylation at repetitive elements and breaking bad. Epigenetics Chromatin *14*, 25. https://doi.org/10.1186/s13072-021-00400-z.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809–817. https://doi.org/10.1126/science.2028256.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet *42*, 733–772. https://doi.org/10.1146/annurev.genet.42.110807.091711.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Mittelsten Scheid, O. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell *22*, 3118–3129. https://doi.org/10.1105/tpc.110.078493.

Peeters, S.B., Cotton, A.M., and Brown, C.J. (2014). Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. Bioessays *36*, 746–756. https://doi.org/10.1002/bies.201400032.

Penagos-Puig, A., and Furlan-Magaril, M. (2020). Heterochromatin as an Important Driver of Genome Organization. Front Cell Dev Biol *8*, 579137. https://doi.org/10.3389/fcell.2020.579137.

Perego, M.G.L., Taiana, M., Bresolin, N., Comi, G.P., and Corti, S. (2019). R-Loops in Motor Neuron Diseases. Mol Neurobiol *56*, 2579–2589. https://doi.org/10.1007/s12035-018-1246-y.

Pérez-Martínez, L., Öztürk, M., Butter, F., and Luke, B. (2020). Npl3 stabilizes R-loops at telomeres to prevent accelerated replicative senescence. EMBO Rep *21*, e49087. https://doi.org/10.15252/embr.201949087.

Perrin, A., Rousseau, J., and Tremblay, J.P. (2017). Increased Expression of Laminin Subunit Alpha 1 Chain by dCas9-VP160. Mol Ther Nucleic Acids *6*, 68–79. https://doi.org/10.1016/j.omtn.2016.11.004.

Peters, J.-M. (2002). The anaphase-promoting complex: proteolysis in mitosis and beyond. Mol Cell *9*, 931–943. https://doi.org/10.1016/s1097-2765(02)00540-3.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99–103. https://doi.org/10.1038/nature00862.

Pezer, Z., and Ugarković, D. (2008). RNA Pol II promotes transcription of centromeric satellite DNA in beetles. PLoS One *3*, e1594. https://doi.org/10.1371/journal.pone.0001594.

Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G., and Earnshaw, W.C. (1995). The centromere: hub of chromosomal activities. Science *270*, 1591–1594. https://doi.org/10.1126/science.270.5242.1591.

Polo, S.E., and Jackson, S.P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev *25*, 409–433. https://doi.org/10.1101/gad.2021311.

Poon, R.Y.C. (2016). Cell Cycle Control: A System of Interlinking Oscillators. Methods Mol Biol *1342*, 3–19. https://doi.org/10.1007/978-1-4939-2957-3_1.

Potapova, O., Basu, S., Mercola, D., and Holbrook, N.J. (2001). Protective role for c-Jun in the cellular response to DNA damage. J Biol Chem *276*, 28546–28553. https://doi.org/10.1074/jbc.M102075200.

Prendergast, L., McClurg, U.L., Hristova, R., Berlinguer-Palmini, R., Greener, S., Veitch, K., Hernandez, I., Pasero, P., Rico, D., Higgins, J.M.G., *et al.* (2020). Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. Nat Commun *11*, 4534. https://doi.org/10.1038/s41467-020-18306-x.

Presting, G.G. (2018). Centromeric retrotransposons and centromere function. Current Opinion in Genetics & Development *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004.

Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344–1354. https://doi.org/10.1016/j.cell.2013.02.011.

Probst, A.V., Okamoto, I., Casanova, M., El Marjou, F., Le Baccon, P., and Almouzni, G. (2010). A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev Cell *19*, 625–638. https://doi.org/10.1016/j.devcel.2010.09.002.

Przewloka, M.R., and Glover, D.M. (2009). The kinetochore and the centromere: a working long distance relationship. Annu Rev Genet *43*, 439–465. https://doi.org/10.1146/annurev-genet-102108-134310.

Qiu, S., and Huang, J. (2021). MRN complex is an essential effector of DNA damage repair. J Zhejiang Univ Sci B *22*, 31–37. https://doi.org/10.1631/jzus.B2000289.

Quénet, D., and Dalal, Y. (2014a). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. Elife *3*, e03254. https://doi.org/10.7554/eLife.03254.

Quénet, D., and Dalal, Y. (2014b). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. ELife *3*. https://doi.org/10.7554/eLife.03254.

Racca, C., Britton, S., Hédouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021a). BRCA1 prevents R-loop-associated centromeric instability. Cell Death Dis *12*, 896. https://doi.org/10.1038/s41419-021-04189-3.

Racca, C., Britton, S., Hédouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021b). BRCA1 prevents R-loop-associated centromeric instability. Cell Death Dis *12*, 896. https://doi.org/10.1038/s41419-021-04189-3.

Rayess, H., Wang, M.B., and Srivatsan, E.S. (2012). Cellular senescence and tumor suppressor gene p16. Int J Cancer *130*, 1715–1725. https://doi.org/10.1002/ijc.27316.

Redon, S., Reichenbach, P., and Lingner, J. (2010). The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic Acids Res *38*, 5797–5806. https://doi.org/10.1093/nar/gkq296.

Ren, W., Gao, L., and Song, J. (2018). Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. Genes (Basel) *9*, E620. https://doi.org/10.3390/genes9120620.

Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. Mol Cell *70*, 650-662.e8. https://doi.org/10.1016/j.molcel.2018.04.001.

Richard, P., and Manley, J.L. (2017). R Loops and Links to Human Disease. J Mol Biol *429*, 3168–3180. https://doi.org/10.1016/j.jmb.2016.08.031.

Rieder, C.L., Cole, R.W., Khodjakov, A., and Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. J Cell Biol *130*, 941–948. https://doi.org/10.1083/jcb.130.4.941.

Rinn, J.L. (2014). lncRNAs: linking RNA to chromatin. Cold Spring Harb Perspect Biol *6*, a018614. https://doi.org/10.1101/cshperspect.a018614.

Rippe, K. (2022a). Liquid-Liquid Phase Separation in Chromatin. Cold Spring Harb Perspect Biol *14*, a040683. https://doi.org/10.1101/cshperspect.a040683.

Rippe, K. (2022b). Liquid-Liquid Phase Separation in Chromatin. Cold Spring Harb Perspect Biol *14*, a040683. https://doi.org/10.1101/cshperspect.a040683.

Rizzi, N., Denegri, M., Chiodi, I., Corioni, M., Valgardsdottir, R., Cobianchi, F., Riva, S., and Biamonti, G. (2004). Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. Mol Biol Cell *15*, 543–551. https://doi.org/10.1091/mbc.e03-07-0487.

Roberts, S.A., Strande, N., Burkhalter, M.D., Strom, C., Havener, J.M., Hasty, P., and Ramsden, D.A. (2010). Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. Nature *464*, 1214–1217. https://doi.org/10.1038/nature08926.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858–5868. https://doi.org/10.1074/jbc.273.10.5858.

Rollins, R.A. (2005). Large-scale structure of genomic methylation patterns. Genome Research *16*, 157–163. https://doi.org/10.1101/gr.4362006.

Rošić, S., Köhler, F., and Erhardt, S. (2014a). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol *207*, 335–349. https://doi.org/10.1083/jcb.201404097.

Rošić, S., Köhler, F., and Erhardt, S. (2014b). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol *207*, 335–349. https://doi.org/10.1083/jcb.201404097.

Roy, D., Yu, K., and Lieber, M.R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol Cell Biol *28*, 50–60. https://doi.org/10.1128/MCB.01251-07.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.-L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol Cell Biol *30*, 146–159. https://doi.org/10.1128/MCB.00897-09.

Ruchaud, S., Carmena, M., and Earnshaw, W.C. (2007). Chromosomal passengers: conducting cell division. Nat Rev Mol Cell Biol *8*, 798–812. https://doi.org/10.1038/nrm2257.

Sage, J., Miller, A.L., Pérez-Mancera, P.A., Wysocki, J.M., and Jacks, T. (2003). Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. Nature *424*, 223–228. https://doi.org/10.1038/nature01764.

Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chédin, F., and Selig, S. (2017). Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. Nat Commun *8*, 14015. https://doi.org/10.1038/ncomms14015.

Saha, N., and Muntean, A.G. (2021). Insight into the multi-faceted role of the SUV family of H3K9 methyltransferases in carcinogenesis and cancer progression. Biochim Biophys Acta Rev Cancer *1875*, 188498. https://doi.org/10.1016/j.bbcan.2020.188498.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3. https://doi.org/10.1186/1756-8935-8-3.

Santiago-Frangos, A., Buyukyoruk, M., Wiegand, T., Krishna, P., and Wiedenheft, B. (2021). Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. Curr Biol *31*, 3515-3524.e6. https://doi.org/10.1016/j.cub.2021.05.068.

Sanz, L.A., and Chédin, F. (2019a). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734–1755. https://doi.org/10.1038/s41596-019-0159-1.

Sanz, L.A., and Chédin, F. (2019b). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734–1755. https://doi.org/10.1038/s41596-019-0159-1.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell *63*, 167–178. https://doi.org/10.1016/j.molcel.2016.05.032.

Sarkar, K., Han, S.-S., Wen, K.-K., Ochs, H.D., Dupré, L., Seidman, M.M., and Vyas, Y.M. (2018). R-loops cause genomic instability in T helper lymphocytes from patients with Wiskott-Aldrich syndrome. J Allergy Clin Immunol *142*, 219–234. https://doi.org/10.1016/j.jaci.2017.11.023.

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A *106*, 2525–2530. https://doi.org/10.1073/pnas.0807899106.

Saul, R.L., and Ames, B.N. (1986). Background levels of DNA damage in the population. Basic Life Sci *38*, 529–535. https://doi.org/10.1007/978-1-4615-9462-8_55.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., *et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife *2*, e01749. https://doi.org/10.7554/eLife.01749.

Scelfo, A., and Fachinetti, D. (2019). Keeping the Centromere under Control: A Promising Role for DNA Methylation. Cells *8*. https://doi.org/10.3390/cells8080912.

Schmeer, C., Kretz, A., Wengerodt, D., Stojiljkovic, M., and Witte, O.W. (2019). Dissecting Aging and Senescence-Current Concepts and Open Lessons. Cells *8*, E1446. https://doi.org/10.3390/cells8111446.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., and Leonhardt, H. (2016a). Determination of local chromatin composition by CasID. Nucleus *7*, 476–484. https://doi.org/10.1080/19491034.2016.1239000.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., and Leonhardt, H. (2016b). Determination of local chromatin composition by CasID. Nucleus *7*, 476–484. https://doi.org/10.1080/19491034.2016.1239000.

Schmitz, J., and Brosius, J. (2011). Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie *93*, 1928–1934. https://doi.org/10.1016/j.biochi.2011.07.014.

Schoeftner, S., and Blasco, M.A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol *10*, 228–236. https://doi.org/10.1038/ncb1685.

Schumann, G.G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans *35*, 637–642. https://doi.org/10.1042/BST0350637.

Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M., and Devine, S.E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res *26*, 745–755. https://doi.org/10.1101/gr.201814.115.

Sekulic, N., and Black, B.E. (2012). Molecular underpinnings of centromere identity and maintenance. Trends Biochem Sci *37*, 220–229. https://doi.org/10.1016/j.tibs.2012.01.003.

Senga, S.S., and Grose, R.P. (2021). Hallmarks of cancer-the new testament. Open Biol *11*, 200358. https://doi.org/10.1098/rsob.200358.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593–602. https://doi.org/10.1016/s0092-8674(00)81902-9.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026–1039. https://doi.org/10.1016/j.biochi.2008.04.011.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat Methods *12*, 664–670. https://doi.org/10.1038/nmeth.3433.

Shen, D., and Skibbens, R.V. (2017). Temperature-dependent regulation of rDNA condensation in Saccharomyces cerevisiae. Cell Cycle *16*, 1118–1127. https://doi.org/10.1080/15384101.2017.1317409.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat Cell Biol *13*, 167–173. https://doi.org/10.1038/ncb2157.

Shiio, Y., and Eisenman, R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A *100*, 13225–13230. https://doi.org/10.1073/pnas.1735528100.

Silva, B., Arora, R., Bione, S., and Azzalin, C.M. (2021). TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. Nat Commun *12*, 3760. https://doi.org/10.1038/s41467-021-24097-6.

Silva, S., Camino, L.P., and Aguilera, A. (2018). Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. Proc Natl Acad Sci U S A *115*, 11024–11029. https://doi.org/10.1073/pnas.1807258115.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., *et al.* (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115. https://doi.org/10.1038/nature23875.

Singh, P.P., Shukla, M., White, S.A., Lafos, M., Tong, P., Auchynnikava, T., Spanos, C., Rappsilber, J., Pidoux, A.L., and Allshire, R.C. (2020). Hap2-Ino80-facilitated transcription promotes de novo establishment of CENP-A chromatin. Genes Dev *34*, 226–238. https://doi.org/10.1101/gad.332536.119.

Sinzelle, L., Izsvák, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci *66*, 1073–1093. https://doi.org/10.1007/s00018-009-8376-3.

Sitbon, D., Podsypanina, K., Yadav, T., and Almouzni, G. (2017). Shaping Chromatin in the Nucleus: The Bricks and the Architects. Cold Spring Harb Symp Quant Biol *82*, 1–14. https://doi.org/10.1101/sqb.2017.82.033753.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794–805. https://doi.org/10.1016/j.molcel.2011.04.026.

Slack, F.J., and Chinnaiyan, A.M. (2019). The Role of Non-coding RNAs in Oncology. Cell *179*, 1033–1055. https://doi.org/10.1016/j.cell.2019.10.017.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88. https://doi.org/10.1126/science.aad5227.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol *14*, 1138–1152. https://doi.org/10.1080/15476286.2016.1259781.

Smolka, J.A., Sanz, L.A., Hartono, S.R., and Chédin, F. (2021). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. J Cell Biol *220*, e202004079. https://doi.org/10.1083/jcb.202004079.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol *25*, 514–522. https://doi.org/10.1016/j.tcb.2015.05.003.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell *56*, 777–785. https://doi.org/10.1016/j.molcel.2014.10.020.

Spector, D.L. (2003). The dynamics of chromosome organization and gene regulation. Annu Rev Biochem *72*, 573–608. https://doi.org/10.1146/annurev.biochem.72.121801.161724.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Mol Cell *52*, 146–152. https://doi.org/10.1016/j.molcel.2013.09.008.

St Clair, S., and Manfredi, J.J. (2006). The dual specificity phosphatase Cdc25C is a direct target for transcriptional repression by the tumor suppressor p53. Cell Cycle *5*, 709–713. https://doi.org/10.4161/cc.5.7.2628.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature https://doi.org/10.1038/nature10716.

Stanley, J.S., Griffin, J.B., and Zempleni, J. (2001). Biotinylation of histones in human cells. Effects of cell proliferation. Eur J Biochem *268*, 5424–5429. https://doi.org/10.1046/j.0014-2956.2001.02481.x.

Stein, G.H., Drullinger, L.F., Soulard, A., and Dulić, V. (1999). Differential roles for cyclin-dependent kinase inhibitors p21 and p16 in the mechanisms of senescence and differentiation in human fibroblasts. Mol Cell Biol *19*, 2109–2117. https://doi.org/10.1128/MCB.19.3.2109.

Stodola, J.L., and Burgers, P.M. (2017). Mechanism of Lagging-Strand DNA Replication in Eukaryotes. Adv Exp Med Biol *1042*, 117–133. https://doi.org/10.1007/978-981-10-6955-0_6.

Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chédin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife *5*, e17548. https://doi.org/10.7554/eLife.17548.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45. https://doi.org/10.1038/47412.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245. https://doi.org/10.1038/nature22989.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226. https://doi.org/10.1016/j.cell.2005.09.038.

Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., and Felsenfeld, G. (2004). Chromatin remodeling by RNA polymerases. Trends Biochem Sci *29*, 127–135. https://doi.org/10.1016/j.tibs.2004.01.003.

Sugino, A., Hirose, S., and Okazaki, R. (1972). RNA-linked nascent DNA fragments in Escherichia coli. Proc Natl Acad Sci U S A *69*, 1863–1867. https://doi.org/10.1073/pnas.69.7.1863.

Sullivan, B.A., Schwartz, S., and Willard, H.F. (1996). Centromeres of human chromosomes. Environ Mol Mutagen *28*, 182–191. https://doi.org/10.1002/(SICI)1098-2280(1996)28:3<182::AID-EM4>3.0.CO;2-G.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619–621. https://doi.org/10.1126/science.1234848.

Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional *Drosophila* centromere. Genome Res *13*, 182–194. https://doi.org/10.1101/gr.681703.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465–476. https://doi.org/10.1038/nrg2341.

Takatsuka, H., Shibata, A., and Umeda, M. (2021). Genome Maintenance Mechanisms at the Chromatin Level. Int J Mol Sci *22*, 10384. https://doi.org/10.3390/ijms221910384.

Talbert, P.B., and Henikoff, S. (2010). Centromeres convert but don't cross. PLoS Biol *8*, e1000326. https://doi.org/10.1371/journal.pbio.1000326.

Talbert, P.B., and Henikoff, S. (2021). Histone variants at a glance. J Cell Sci *134*, jcs244749. https://doi.org/10.1242/jcs.244749.

Talbert, P.B., Bryson, T.D., and Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. J Biol *3*, 18. https://doi.org/10.1186/jbiol11.

Tan, J., Duan, M., Yadav, T., Phoon, L., Wang, X., Zhang, J.-M., Zou, L., and Lan, L. (2020). An R-loop-initiated CSB-RAD52-POLD3 pathway suppresses ROS-induced telomeric DNA breaks. Nucleic Acids Res *48*, 1285–1300. https://doi.org/10.1093/nar/gkz1114.

Tanaka, Y., Kurumizaka, H., and Yokoyama, S. (2005). CpG methylation of the CENP-B box reduces human CENP-B binding. FEBS J *272*, 282–289. https://doi.org/10.1111/j.1432-1033.2004.04406.x.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646. https://doi.org/10.1016/j.cell.2014.09.039.

Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., Wei, Z., Su, J., Liu, G., Rong, R., *et al.* (2021). m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. Nucleic Acids Res *49*, D134–D143. https://doi.org/10.1093/nar/gkaa692.

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. https://doi.org/10.1038/nsmb1338.

Terranova, R., Sauer, S., Merkenschlager, M., and Fisher, A.G. (2005). The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. Exp Cell Res *310*, 344–356. https://doi.org/10.1016/j.yexcr.2005.07.031.

Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. Int J Mol Sci *22*, 4309. https://doi.org/10.3390/ijms22094309.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., *et al.* (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun *6*. https://doi.org/10.1038/ncomms8870.

Thiru, P., Kern, D.M., McKinley, K.L., Monda, J.K., Rago, F., Su, K.-C., Tsinman, T., Yarar, D., Bell, G.W., and Cheeseman, I.M. (2014). Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. Mol Biol Cell *25*, 1983–1994. https://doi.org/10.1091/mbc.E14-03-0837.

Thomas, M., White, R.L., and Davis, R.W. (1976a). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci U S A *73*, 2294–2298. https://doi.org/10.1073/pnas.73.7.2294.

Thomas, M., White, R.L., and Davis, R.W. (1976b). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci U S A *73*, 2294–2298. https://doi.org/10.1073/pnas.73.7.2294.

Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G.F., and Romano, C. (1979). Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. Hum Genet *51*, 127–137. https://doi.org/10.1007/BF00287166.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., *et al.* (2011a). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. Science *331*, 593–596. https://doi.org/10.1126/science.1200801.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., *et al.* (2011b). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science *331*, 593–596. https://doi.org/10.1126/science.1200801.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I. (2010). Stress-induced activation of heterochromatic transcription. PLoS Genet *6*, e1001175. https://doi.org/10.1371/journal.pgen.1001175.

Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004a). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci U S A *101*, 15986–15991. https://doi.org/10.1073/pnas.0407154101.

Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004b). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci U S A *101*, 15986–15991. https://doi.org/10.1073/pnas.0407154101.

Toubiana, S., Velasco, G., Chityat, A., Kaindl, A.M., Hershtig, N., Tzur-Gilat, A., Francastel, C., and Selig, S. (2018). Subtelomeric methylation distinguishes between subtypes of Immunodeficiency, Centromeric instability and Facial anomalies syndrome. Hum Mol Genet *27*, 3568–3581. https://doi.org/10.1093/hmg/ddy265.

Trivedi, P., and Stukenberg, P.T. (2016). A Centromere-Signaling Network Underlies the Coordination among Mitotic Events. Trends Biochem Sci *41*, 160–174. https://doi.org/10.1016/j.tibs.2015.11.002.

Trivedi, P., Palomba, F., Niedzialkowska, E., Digman, M.A., Gratton, E., and Stukenberg, P.T. (2019). The inner centromere is a biomolecular condensate scaffolded by the chromosomal passenger complex. Nat Cell Biol *21*, 1127–1137. https://doi.org/10.1038/s41556-019-0376-4.

Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. Mol Cell *63*, 293–305. https://doi.org/10.1016/j.molcel.2016.06.002.

Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell *168*, 644–656. https://doi.org/10.1016/j.cell.2017.01.002.

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., *et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol *11*, 1315–1324. https://doi.org/10.1038/ncb1984.

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., and Sasaki, H. (2020a). CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. Sci Rep *10*, 17865. https://doi.org/10.1038/s41598-020-74636-2.

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., and Sasaki, H. (2020b). CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. Sci Rep *10*, 17865. https://doi.org/10.1038/s41598-020-74636-2.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636–646. https://doi.org/10.1038/nrg2842.

Uruci, S., Lo, C.S.Y., Wheeler, D., and Taneja, N. (2021). R-Loops and Its Chro-Mates: The Strange Case of Dr. Jekyll and Mr. Hyde. Int J Mol Sci *22*, 8850. https://doi.org/10.3390/ijms22168850.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet *19*, 535–548. https://doi.org/10.1038/s41576-018-0017-y.

Valgardsdottir, R., Chiodi, I., Giordano, M., Rossi, A., Bazzini, S., Ghigna, C., Riva, S., and Biamonti, G. (2008). Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res *36*, 423–434. https://doi.org/10.1093/nar/gkm1056.

Varma, D., and Salmon, E.D. (2012). The KMN protein network--chief conductors of the kinetochore orchestra. J Cell Sci *125*, 5927–5936. https://doi.org/10.1242/jcs.093724.

Vaskova, E.A., Dementyeva, E.V., Shevchenko, A.I., Pavlova, S.V., Grigor'eva, E.V., Zhelezova, A.I., Vandeberg, J.L., and Zakian, S.M. (2014). Dynamics of the two heterochromatin types during imprinted X chromosome inactivation in vole Microtus levis. PLoS One *9*, e88256. https://doi.org/10.1371/journal.pone.0088256.

Velasco, G., and Francastel, C. (2019). Genetics meets DNA methylation in rare diseases. Clin Genet *95*, 210–220. https://doi.org/10.1111/cge.13480.

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc Natl Acad Sci U S A *107*, 9281–9286. https://doi.org/10.1073/pnas.1000473107.

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., and Francastel, C. (2018). Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Hum Mol Genet *27*, 2409–2424. https://doi.org/10.1093/hmg/ddy130.

Velazquez Camacho, O., Galan, C., Swist-Rosowska, K., Ching, R., Gamalinda, M., Karabiber, F., De La Rosa-Velazquez, I., Engist, B., Koschorz, B., Shukeir, N., *et al.* (2017). Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife *6*, e25293. https://doi.org/10.7554/eLife.25293.

Vijg, J., and Suh, Y. (2013). Genome instability and aging. Annu Rev Physiol *75*, 645–668. https://doi.org/10.1146/annurev-physiol-030212-183715.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193. https://doi.org/10.1016/j.cell.2012.09.002.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837. https://doi.org/10.1126/science.1074973.

Vourc'h, C., and Biamonti, G. (2011). Transcription of Satellite DNAs in Mammals. Prog Mol Subcell Biol *51*, 95–118. https://doi.org/10.1007/978-3-642-16502-3_5.

Vydzhak, O., Luke, B., and Schindler, N. (2020). Non-coding RNAs at the Eukaryotic rDNA Locus: RNA-DNA Hybrids and Beyond. J Mol Biol *432*, 4287–4304. https://doi.org/10.1016/j.jmb.2020.05.011.

Waddington, C.H. (2012). The Epigenotype. Int. J. Epidemiol. *41*, 10–13. https://doi.org/10.1093/ije/dyr184.

Wahba, L., Gore, S.K., and Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. Elife *2*, e00505. https://doi.org/10.7554/eLife.00505.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev *30*, 1327–1338. https://doi.org/10.1101/gad.280834.116.

Wang, J., and Davis, R.E. (2014). Programmed DNA elimination in multicellular organisms. Curr Opin Genet Dev *27*, 26–34. https://doi.org/10.1016/j.gde.2014.03.012.

Wang, Y., and Chen, L.-L. (2020). Organization and function of paraspeckles. Essays Biochem *64*, 875–882. https://doi.org/10.1042/EBC20200010.

Wang, H., Perrault, A.R., Takeda, Y., Qin, W., Wang, H., and Iliakis, G. (2003). Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Res *31*, 5377–5388. https://doi.org/10.1093/nar/gkg728.

Wang, K., Peng, J., and Yi, C. (2021). The m6A Consensus Motif Provides a Paradigm of Epitranscriptomic Studies. Biochemistry *60*, 3410–3412. https://doi.org/10.1021/acs.biochem.1c00254.

Wang, Y., Park, S., and Beal, P.A. (2018). Selective Recognition of RNA Substrates by ADAR Deaminase Domains. Biochemistry *57*, 1640–1651. https://doi.org/10.1021/acs.biochem.7b01100.

Wassmann, K., and Benezra, R. (2001). Mitotic checkpoints: from yeast to cancer. Curr Opin Genet Dev *11*, 83–90. https://doi.org/10.1016/s0959-437x(00)00161-1.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738. https://doi.org/10.1038/171737a0.

Weber, M., and Schübeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Current Opinion in Cell Biology *19*, 273–280. https://doi.org/10.1016/j.ceb.2007.04.011.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics *39*, 457–466. https://doi.org/10.1038/ng1990.

Wellinger, R.E., Prado, F., and Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol *26*, 3327–3334. https://doi.org/10.1128/MCB.26.8.3327-3334.2006.

Westhorpe, F.G., and Straight, A.F. (2013). Functions of the centromere and kinetochore in chromosome segregation. Curr Opin Cell Biol *25*, 334–340. https://doi.org/10.1016/j.ceb.2013.02.001.

White, R.L., and Hogness, D.S. (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. Cell *10*, 177–192. https://doi.org/10.1016/0092-8674(77)90213-6.

Wijmenga, C., Hansen, R.S., Gimelli, G., Björck, E.J., Davies, E.G., Valentine, D., Belohradsky, B.H., van Dongen, J.J., Smeets, D.F., van den Heuvel, L.P., *et al.* (2000). Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Human Mutation *16*, 509–517. .

Wilkins, M.H.F., and Randall, J.T. (1953). Crystallinity in sperm heads: molecular structure of nucleoprotein in vivo. Biochim Biophys Acta *10*, 192–193. https://doi.org/10.1016/0006-3002(53)90232-7.

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.* (2007a). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160. https://doi.org/10.1101/gr.6022807.

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.* (2007b). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160. https://doi.org/10.1101/gr.6022807.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem Soc Trans *38*, 1027–1032. https://doi.org/10.1042/BST0381027.

Wright, S.I., Agrawal, N., and Bureau, T.E. (2003). Effects of recombination rate and gene density on transposable element distributions in Arabidopsis thaliana. Genome Res *13*, 1897–1903. https://doi.org/10.1101/gr.1281503.

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Human Molecular Genetics ddw243. .

Xu, B., and Clayton, D.A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. EMBO J *15*, 3135–3143. .

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187–191. https://doi.org/10.1038/46052.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.-L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov *2*, 16009. https://doi.org/10.1038/celldisc.2016.9.

Yang, X., Liu, Q.-L., Xu, W., Zhang, Y.-C., Yang, Y., Ju, L.-F., Chen, J., Chen, Y.-S., Li, K., Ren, J., *et al.* (2019). m6A promotes R-loop formation to facilitate transcription termination. Cell Res *29*, 1035–1038. https://doi.org/10.1038/s41422-019-0235-7.

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H.S., Rippe, K., Meshorer, E., *et al.* (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. Cell Rep *10*, 1122–1134. https://doi.org/10.1016/j.celrep.2015.01.038.

Yilmaz, D., Furst, A., Meaburn, K., Lezaja, A., Wen, Y., Altmeyer, M., Reina-San-Martin, B., and Soutoglou, E. (2021a). Activation of homologous recombination in G1 preserves centromeric integrity. Nature *600*, 748–753. https://doi.org/10.1038/s41586-021-04200-z.

Yilmaz, D., Furst, A., Meaburn, K., Lezaja, A., Wen, Y., Altmeyer, M., Reina-San-Martin, B., and Soutoglou, E. (2021b). Activation of homologous recombination in G1 preserves centromeric integrity. Nature *600*, 748–753. https://doi.org/10.1038/s41586-021-04200-z.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523. https://doi.org/10.1038/nature06968.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics *13*, 335–340. https://doi.org/10.1016/S0168-9525(97)01181-5.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res *40*, 5569–5576. https://doi.org/10.1093/nar/gks216.

Yu, H. (2002). Regulation of APC-Cdc20 by the spindle checkpoint. Curr Opin Cell Biol *14*, 706–714. https://doi.org/10.1016/s0955-0674(02)00382-4.

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442–451. https://doi.org/10.1038/ni919.

Zagrovic, B., Bartonek, L., and Polyansky, A.A. (2018). RNA-protein interactions in an unstructured context. FEBS Lett *592*, 2901–2916. https://doi.org/10.1002/1873-3468.13116.

Zaitsev, E.N., and Kowalczykowski, S.C. (2000). A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange. Genes Dev *14*, 740–749. .

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., *et al.* (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350. https://doi.org/10.1016/j.cell.2014.11.052.

Zeitlin, S.G., Baker, N.M., Chapados, B.R., Soutoglou, E., Wang, J.Y.J., Berns, M.W., and Cleveland, D.W. (2009). Double-strand DNA breaks recruit the centromeric histone CENP-A. Proc Natl Acad Sci U S A *106*, 15762–15767. https://doi.org/10.1073/pnas.0908233106.

Zeng, C., Onoguchi, M., and Hamada, M. (2021). Association analysis of repetitive elements and R-loop formation across species. Mob DNA *12*, 3. https://doi.org/10.1186/s13100-021-00231-5.

Zhang, W., Lee, H.-R., Koo, D.-H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell *20*, 25–34. https://doi.org/10.1105/tpc.107.057083.

Zhang, W., Mao, J.-H., Zhu, W., Jain, A.K., Liu, K., Brown, J.B., and Karpen, G.H. (2016). Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. Nat Commun *7*, 12619. https://doi.org/10.1038/ncomms12619.

Zhang, X., Chiang, H.-C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S.J., Michalek, J., Jatoi, I., Lautner, M., *et al.* (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nat Commun *8*, 15908. https://doi.org/10.1038/ncomms15908.

Zhao, W., Steinfeld, J.B., Liang, F., Chen, X., Maranon, D.G., Jian Ma, C., Kwon, Y., Rao, T., Wang, W., Sheng, C., *et al.* (2017). BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. Nature *550*, 360–365. https://doi.org/10.1038/nature24060.

Zhou, K., Gebala, M., Woods, D., Sundararajan, K., Edwards, G., Krzizike, D., Wereszczynski, J., Straight, A.F., and Luger, K. (2022). CENP-N promotes the compaction of centromeric chromatin. Nat Struct Mol Biol *29*, 403–413. https://doi.org/10.1038/s41594-022-00758-y.

Zhu, J., Cheng, K.C.L., and Yuen, K.W.Y. (2018a). Histone H3K9 and H4 Acetylations and Transcription Facilitate the Initial CENP-AHCP-3 Deposition and De Novo Centromere Establishment in Caenorhabditis elegans Artificial Chromosomes. Epigenetics Chromatin *11*, 16. https://doi.org/10.1186/s13072-018-0185-1.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011a). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184. https://doi.org/10.1038/nature10371.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011b). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184. https://doi.org/10.1038/nature10371.

Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K.H., Ke, E., Verma, S., Soroczynski, J., *et al.* (2018b). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol Cell *70*, 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023.

Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K.H., Ke, E., Verma, S., Soroczynski, J., *et al.* (2018c). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol Cell *70*, 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023.

Zilfou, J.T., and Lowe, S.W. (2009). Tumor suppressive functions of p53. Cold Spring Harb Perspect Biol *1*, a001883. https://doi.org/10.1101/cshperspect.a001883.

Zou, H., McGarry, T.J., Bernal, T., and Kirschner, M.W. (1999). Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science *285*, 418–422. https://doi.org/10.1126/science.285.5426.418.

(2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. The EMBO Journal *30*, 328–340. https://doi.org/10.1038/emboj.2010.329.

Figure 1

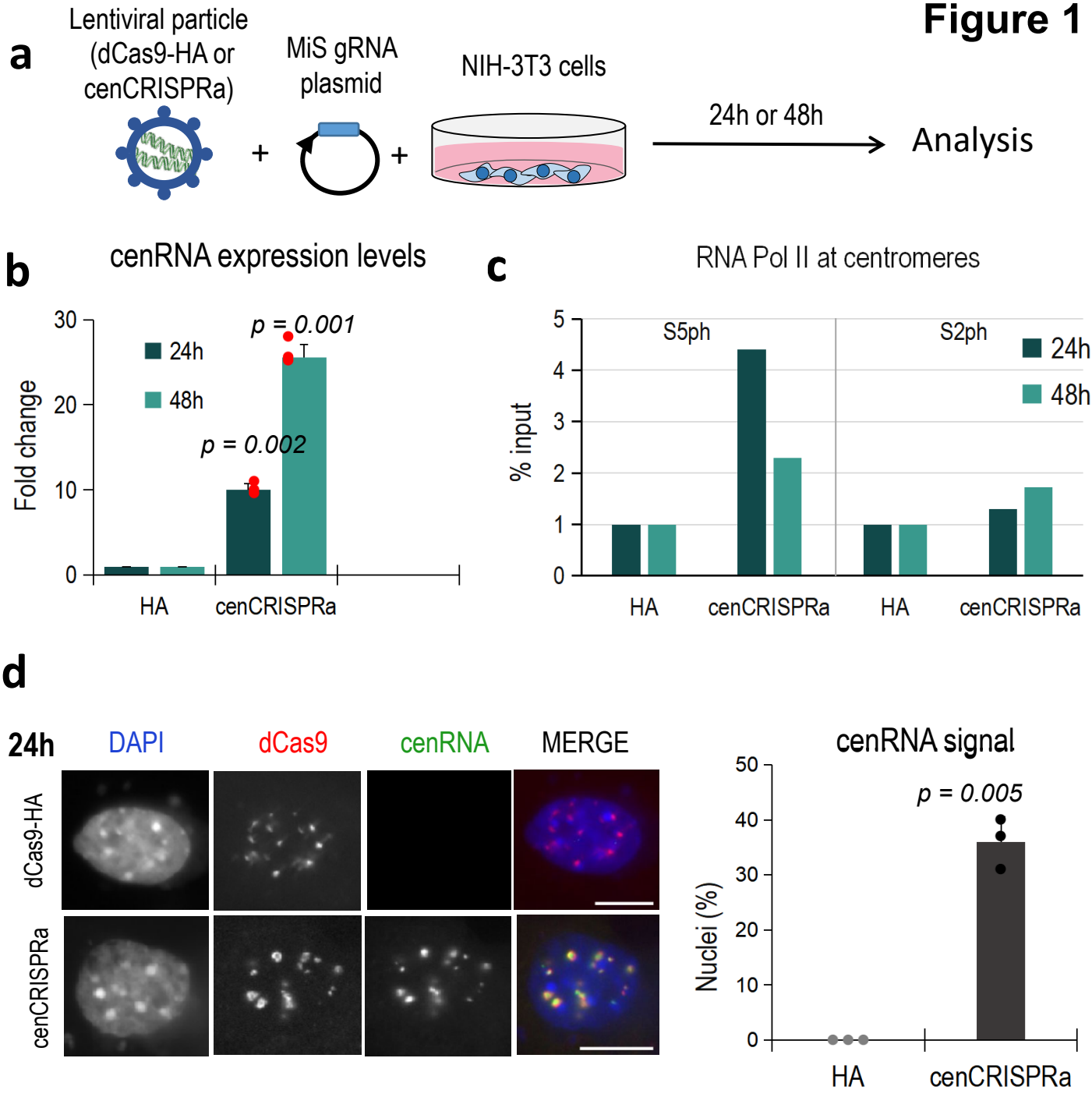

a
Lentiviral particle (dCas9-HA or cenCRISPRa)
MiS gRNA plasmid
NIH-3T3 cells
24h or 48h
Analysis
b
cenRNA expression levels
Fold change
24h
48h
p = 0.001
p = 0.002
HA
cenCRISPRa
c
RNA Pol II at centromeres
% input
S5ph
S2ph
24h
48h
HA
cenCRISPRa
HA
cenCRISPRa
d
24h
DAPI
dCas9
cenRNA
MERGE
dCas9-HA
cenCRISPRa
cenRNA signal
Nuclei (%)
p = 0.005
HA
cenCRISPRa

Figure 2

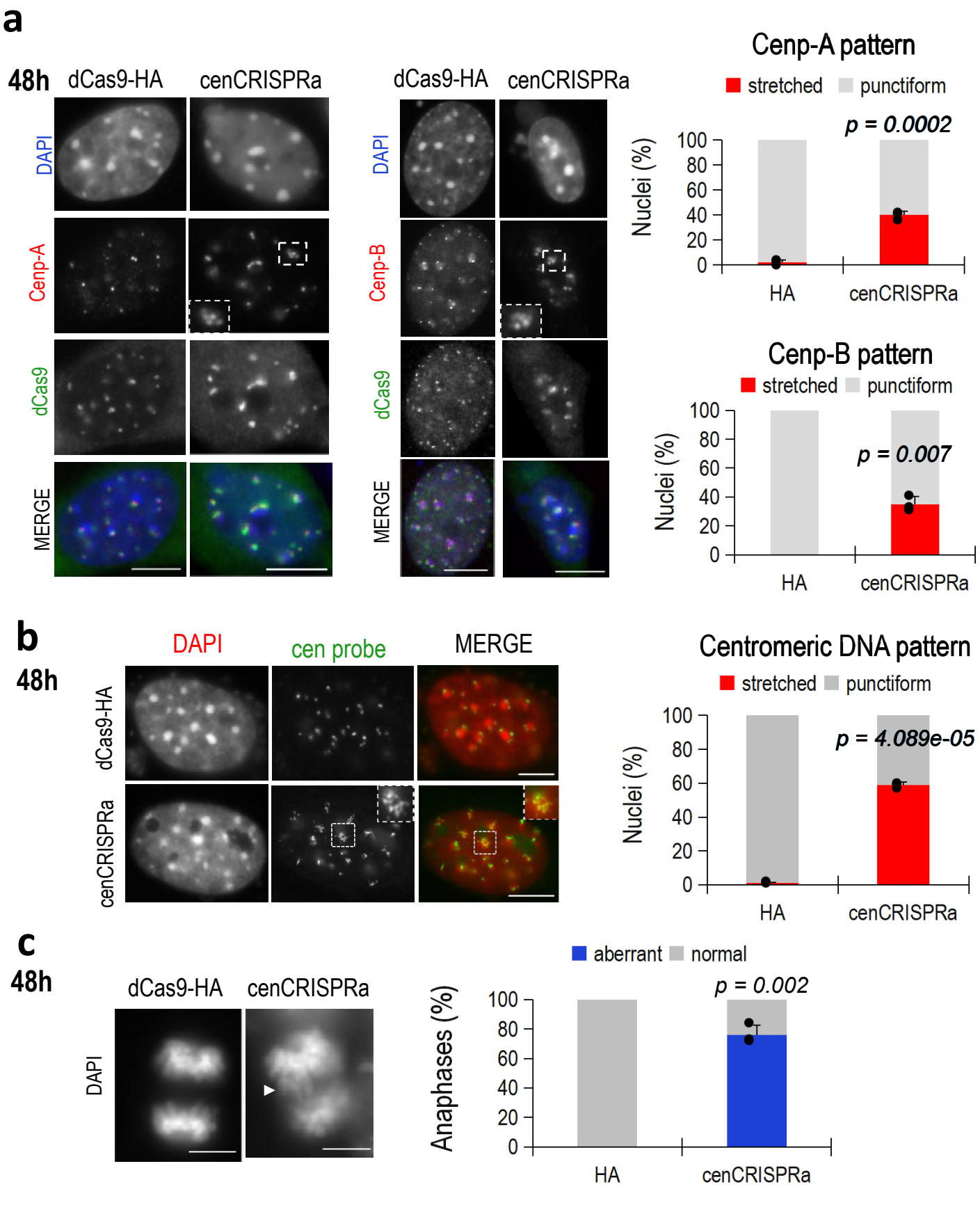

a
48h
dCas9-HA
cenCRISPRa
DAPI
Cenp-A
dCas9
MERGE
dCas9-HA
cenCRISPRa
DAPI
Cenp-B
dCas9
MERGE
Cenp-A pattern
stretched
punctiform
p = 0.0002
Nuclei (%)
100
80
60
40
20
0
HA
cenCRISPRa
Cenp-B pattern
stretched
punctiform
p = 0.007
Nuclei (%)
100
80
60
40
20
0
HA
cenCRISPRa
b
48h
DAPI
cen probe
MERGE
dCas9-HA
cenCRISPRa
Centromeric DNA pattern
stretched
punctiform
p = 4.089e-05
Nuclei (%)
100
80
60
40
20
0
HA
cenCRISPRa
c
48h
dCas9-HA
cenCRISPRa
DAPI
aberrant
normal
p = 0.002
Anaphases (%)
100
80
60
40
20
0
HA
cenCRISPRa

Figure 3

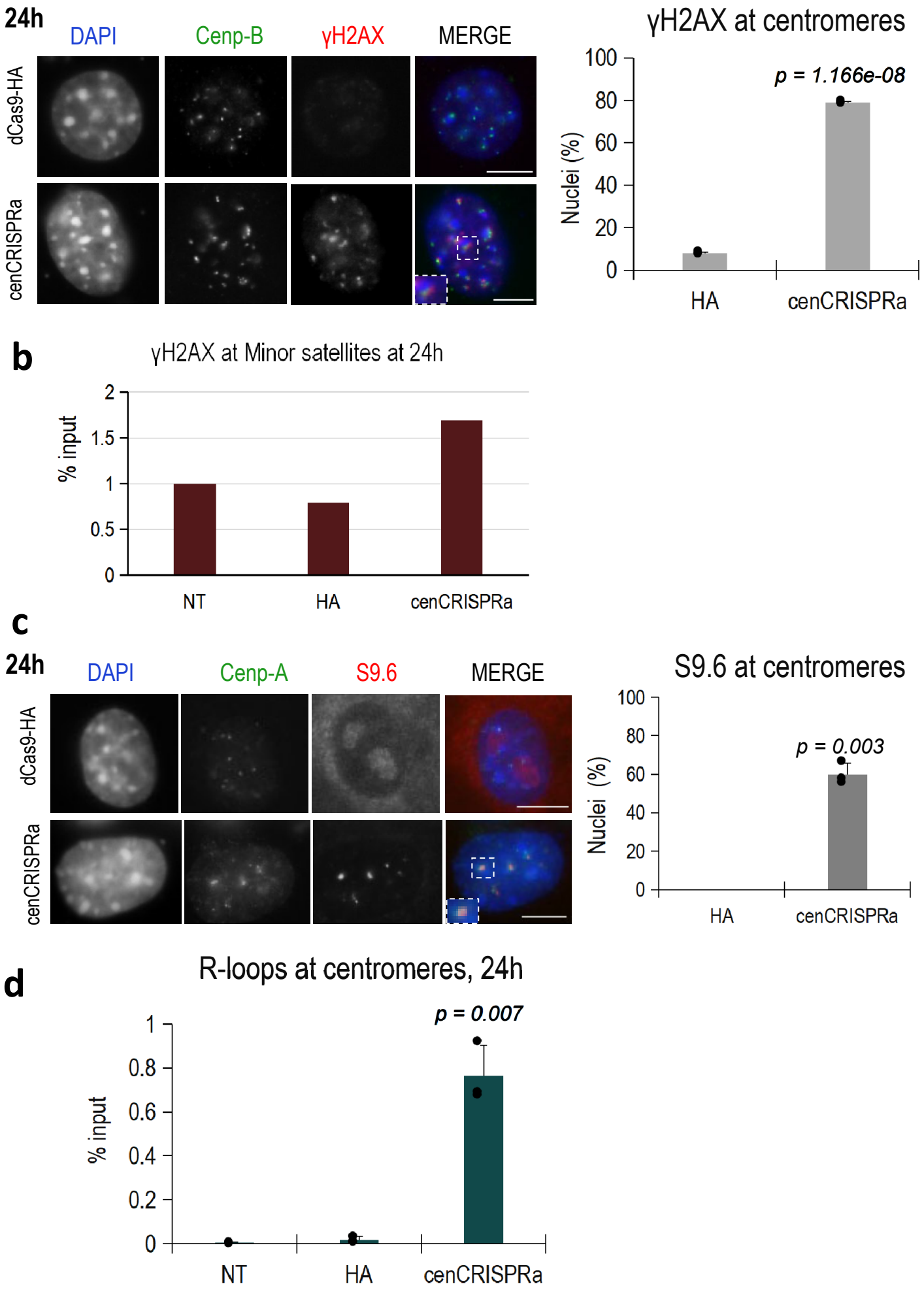
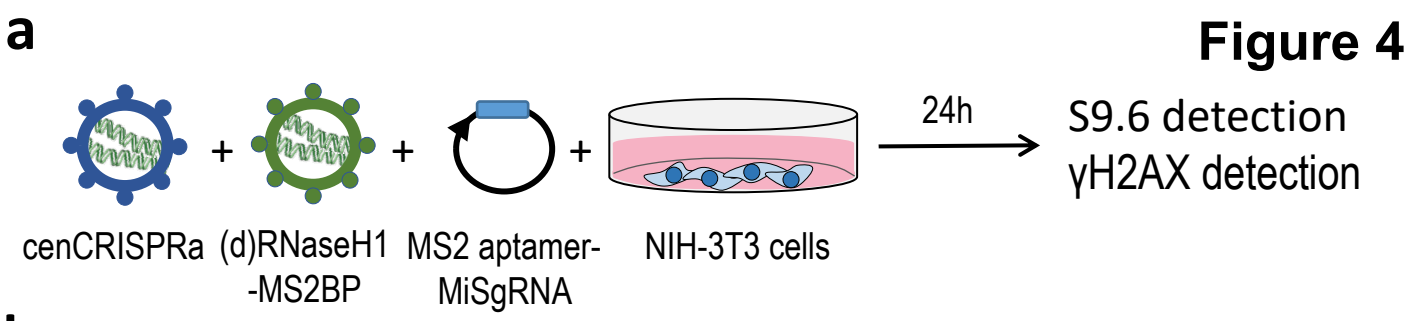
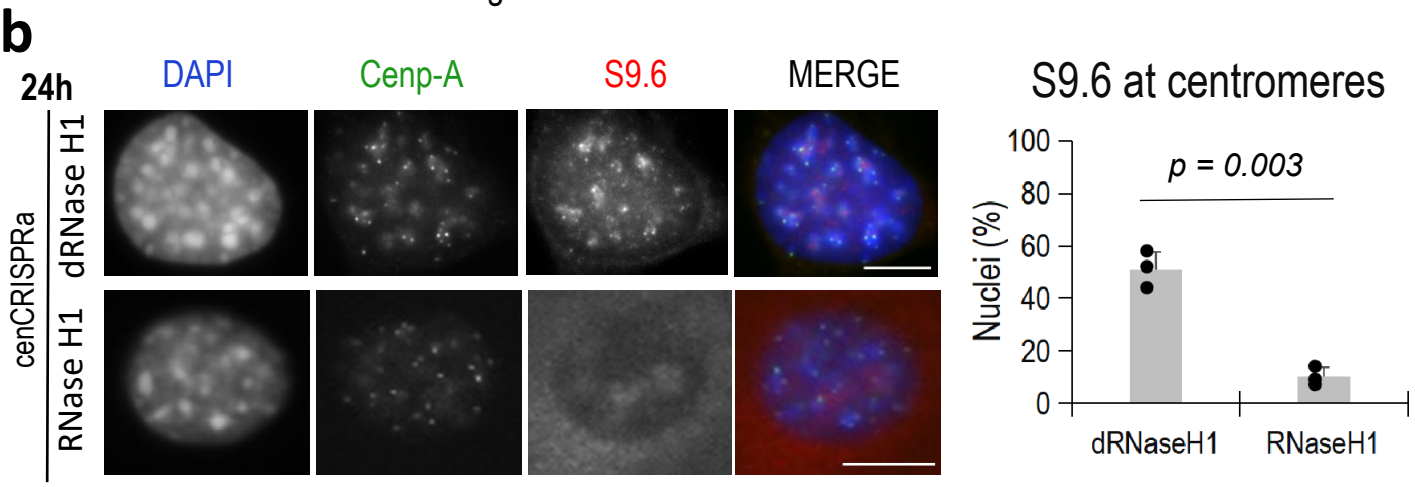
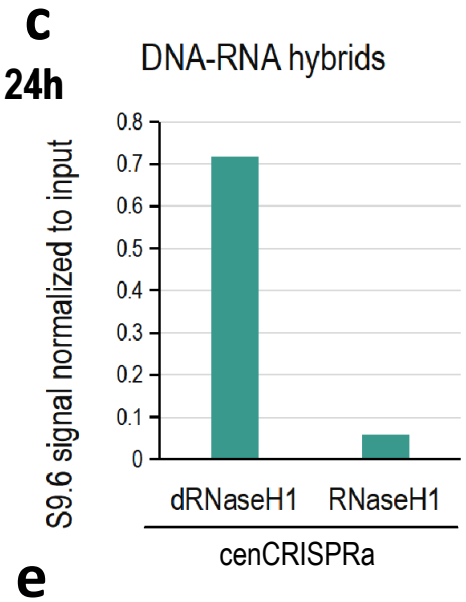
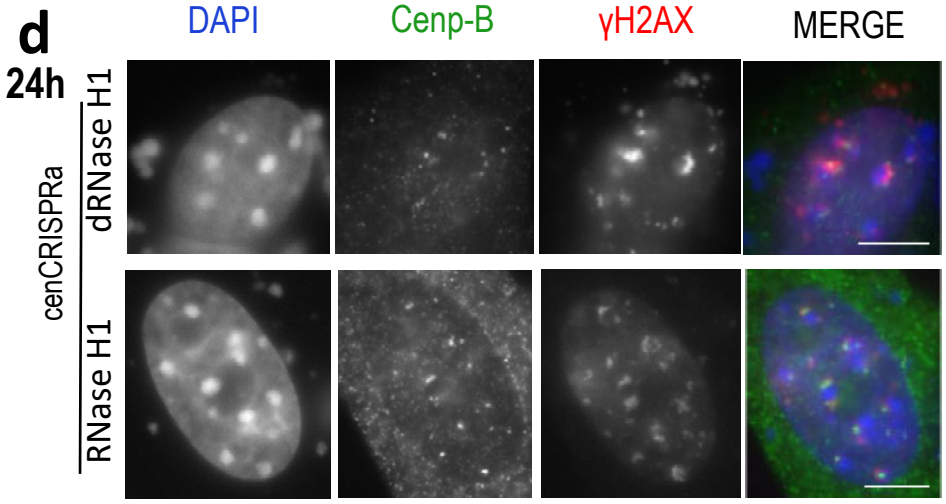

a
24h
DAPI
Cenp-B
γH2AX
MERGE
dCas9-HA
cenCRISPRa
γH2AX at centromeres
p = 1.166e-08
Nuclei (%)
100
80
60
40
20
0
HA
cenCRISPRa
b
γH2AX at Minor satellites at 24h
% input
2
1.5
1
0.5
0
NT
HA
cenCRISPRa
c
24h
DAPI
Cenp-A
S9.6
MERGE
dCas9-HA
cenCRISPRa
S9.6 at centromeres
p = 0.003
Nuclei (%)
100
80
60
40
20
0
HA
cenCRISPRa
d
R-loops at centromeres, 24h
p = 0.007
% input
1
0.8
0.6
0.4
0.2
0
NT
HA
cenCRISPRa

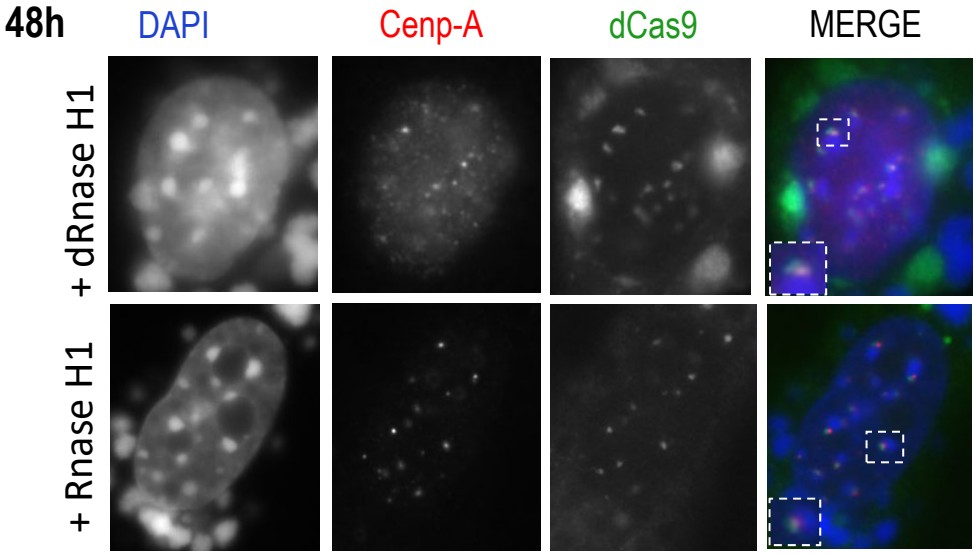
Figure 4


a
cenCRISPRa
(d)RNaseH1
-MS2BP
MS2 aptamer-
MiSgRNA
NIH-3T3 cells
24h
S9.6 detection
γH2AX detection


b
24h
DAPI
Cenp-A
S9.6
MERGE
cenCRISPRa
dRNase H1
RNase H1
S9.6 at centromeres
p = 0.003
Nuclei (%)
100
80
60
40
20
0
dRNaseH1
RNaseH1


c
24h
DNA-RNA hybrids
S9.6 signal normalized to input
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
dRNaseH1
RNaseH1
cenCRISPRa


d
24h
DAPI
Cenp-B
γH2AX
MERGE
cenCRISPRa
dRNase H1
RNase H1


e
48h
DAPI
Cenp-A
dCas9
MERGE
+ dRNase H1
+ Rnase H1

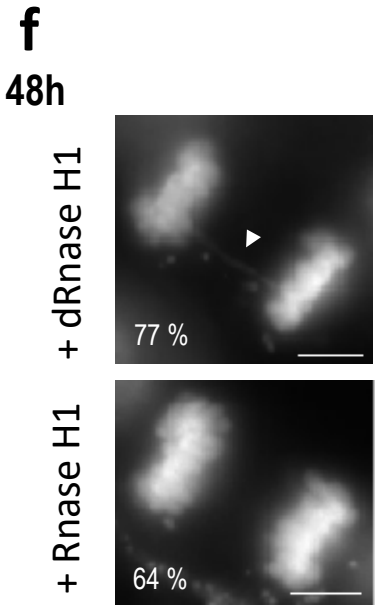

f
48h
+ dRNase H1
+ Rnase H1
77 %
64 %

Figure 5

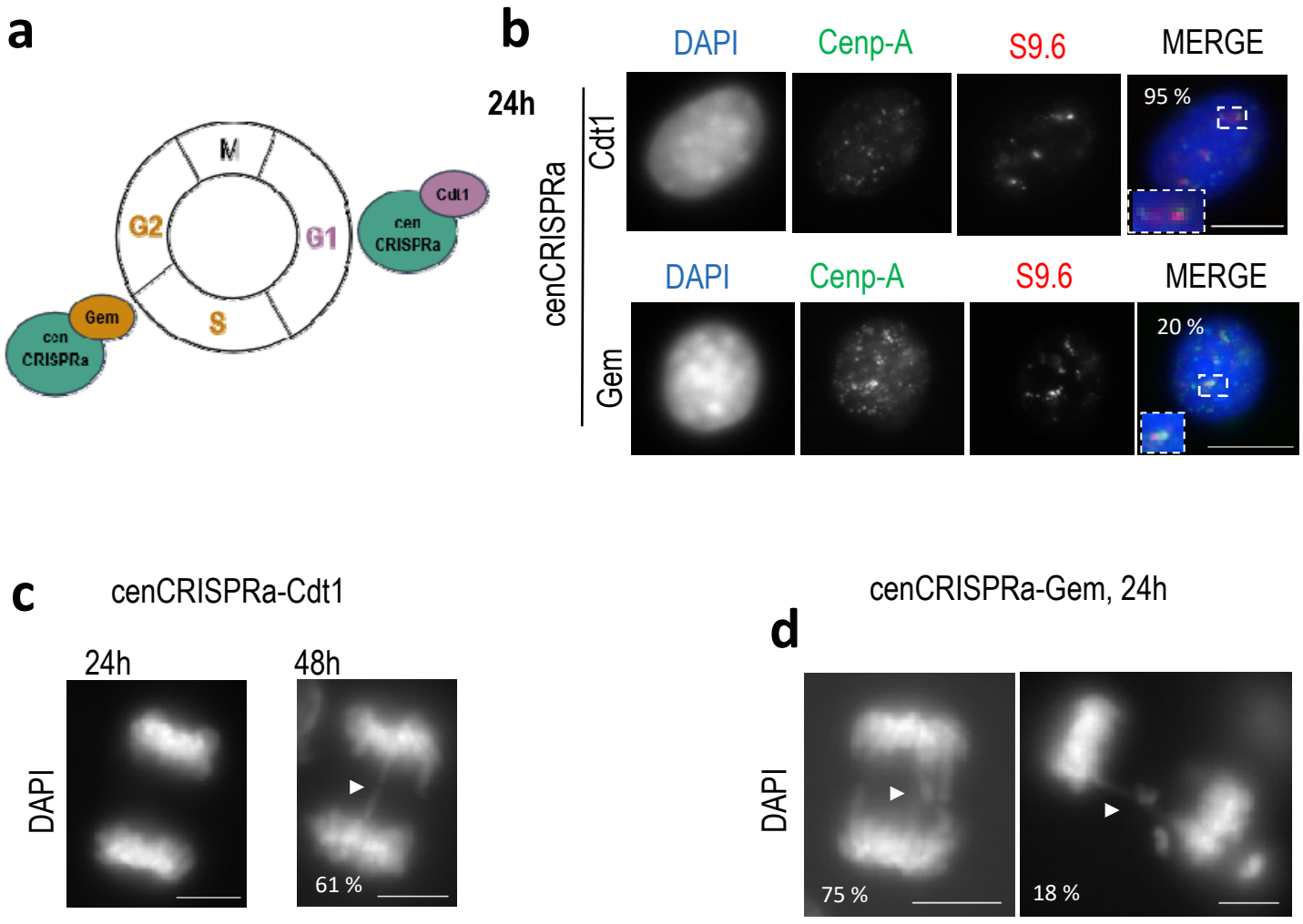
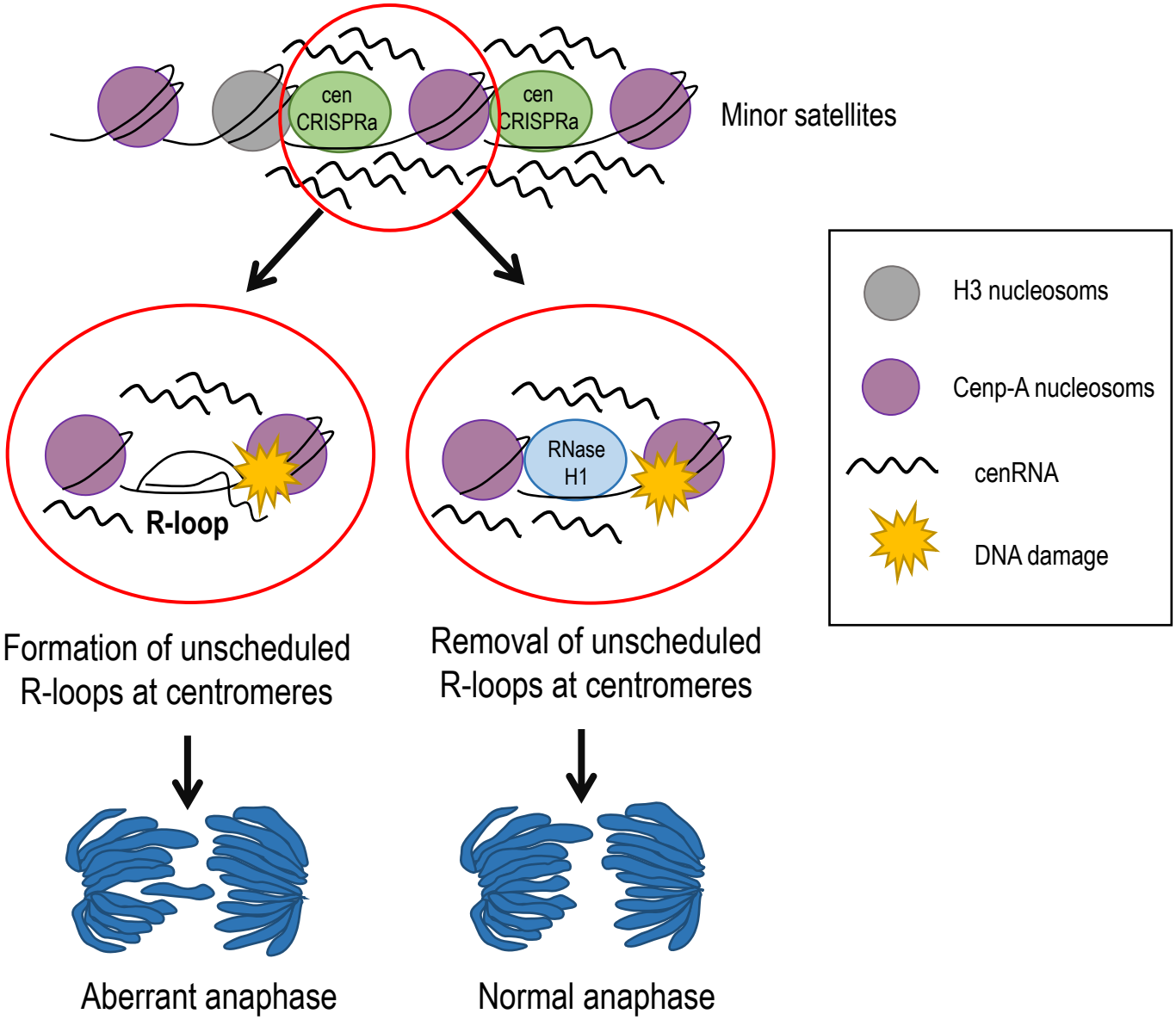
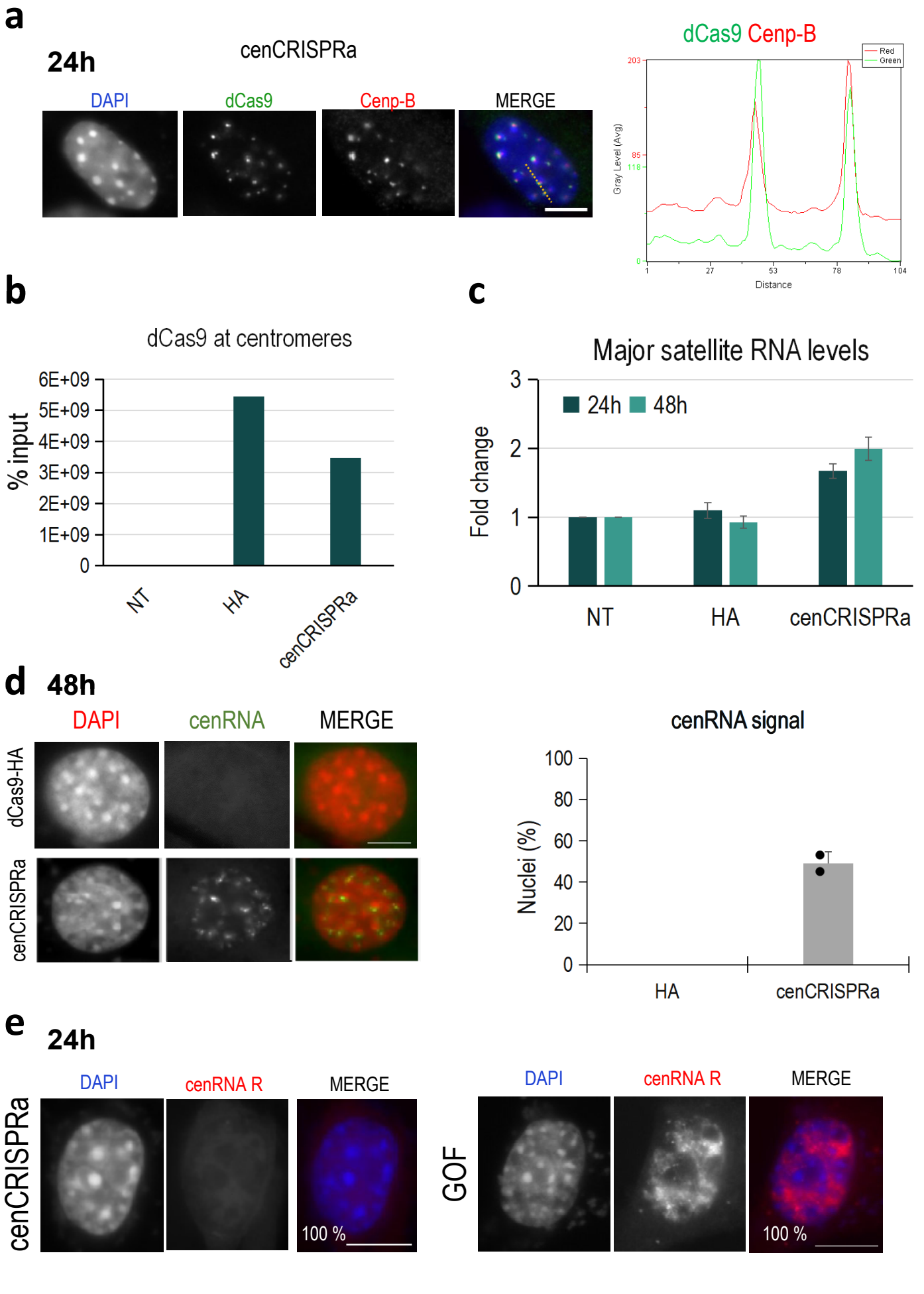

a
M
G2
G1
S
Cdt1
cen CRISPRa
Gem
cen CRISPRa
b
24h
DAPI
Cenp-A
S9.6
MERGE
cenCRISPRa
Cdt1
95 %
DAPI
Cenp-A
S9.6
MERGE
Gem
20 %
c
cenCRISPRa-Cdt1
24h
48h
DAPI
61 %
d
cenCRISPRa-Gem, 24h
DAPI
75 %
18 %

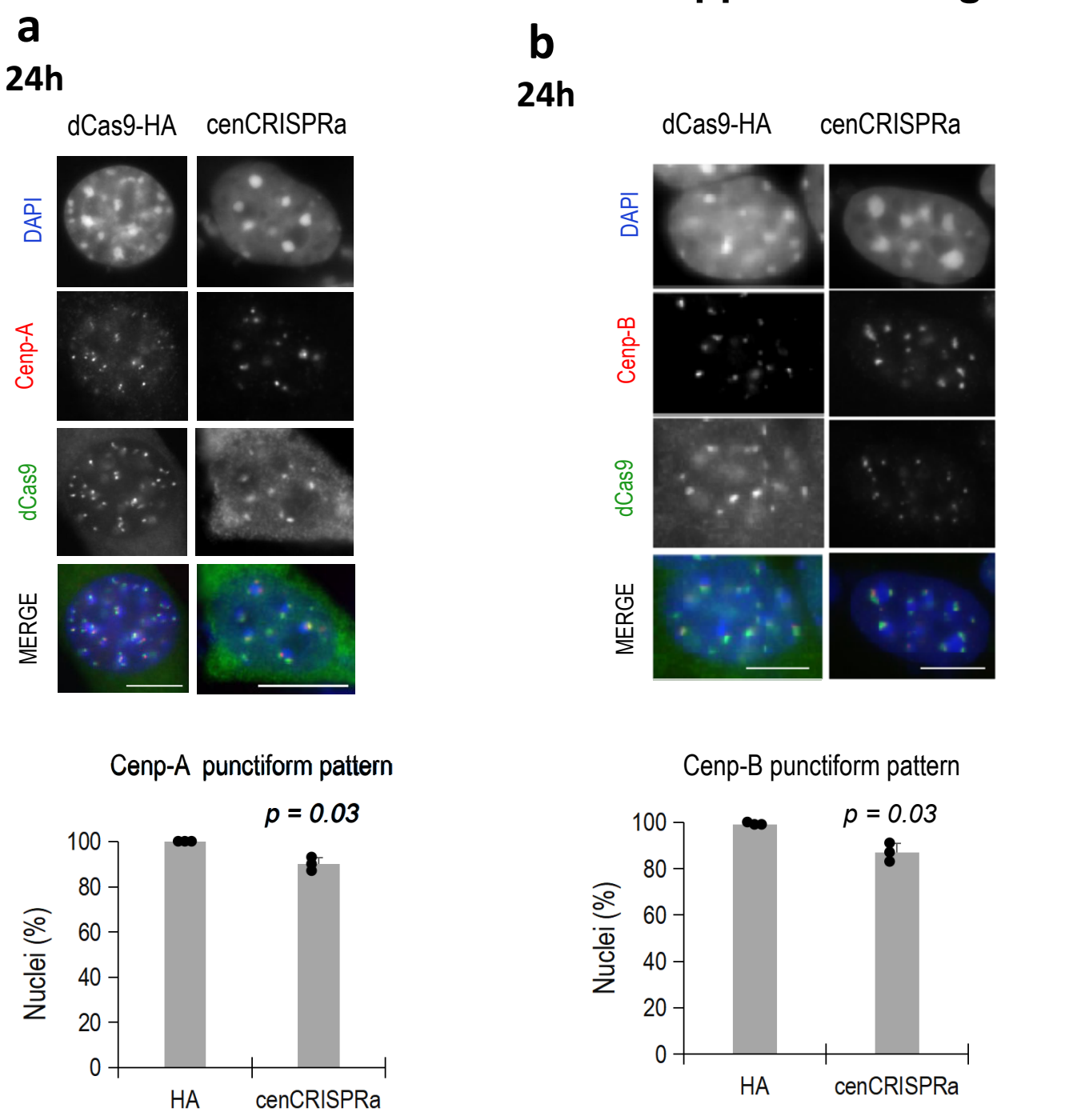
Figure 6


cen CRISPRa
cen CRISPRa
Minor satellites
R-loop
RNase H1
H3 nucleosoms
Cenp-A nucleosoms
cenRNA
DNA damage
Formation of unscheduled R-loops at centromeres
Removal of unscheduled R-loops at centromeres
Aberrant anaphase
Normal anaphase

Supplemental figure 1

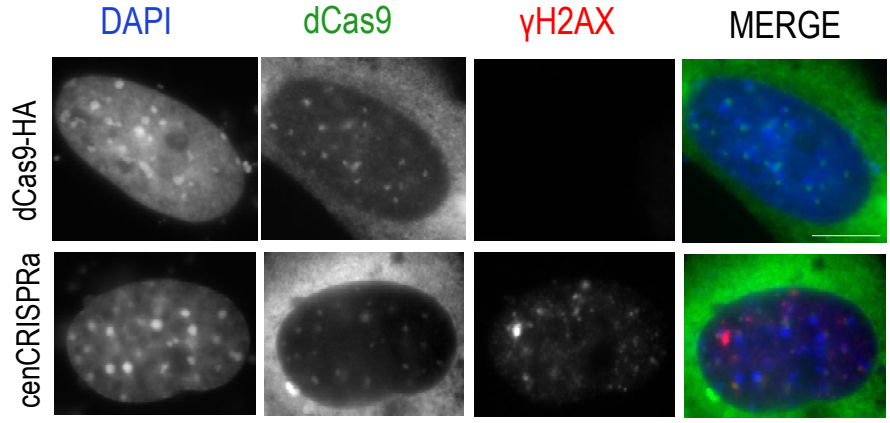


Supplemental figure 2

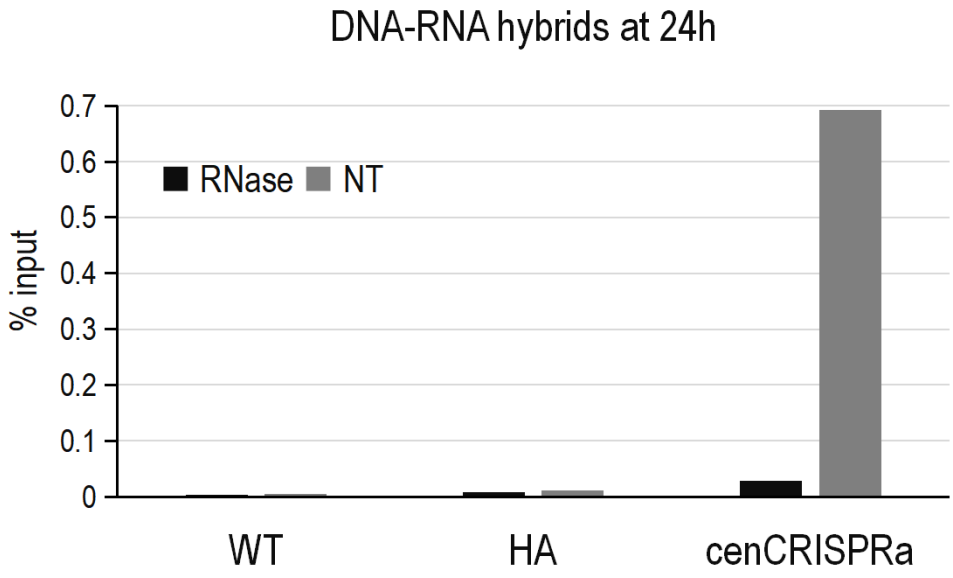


Supplemental figure 3

a 48h

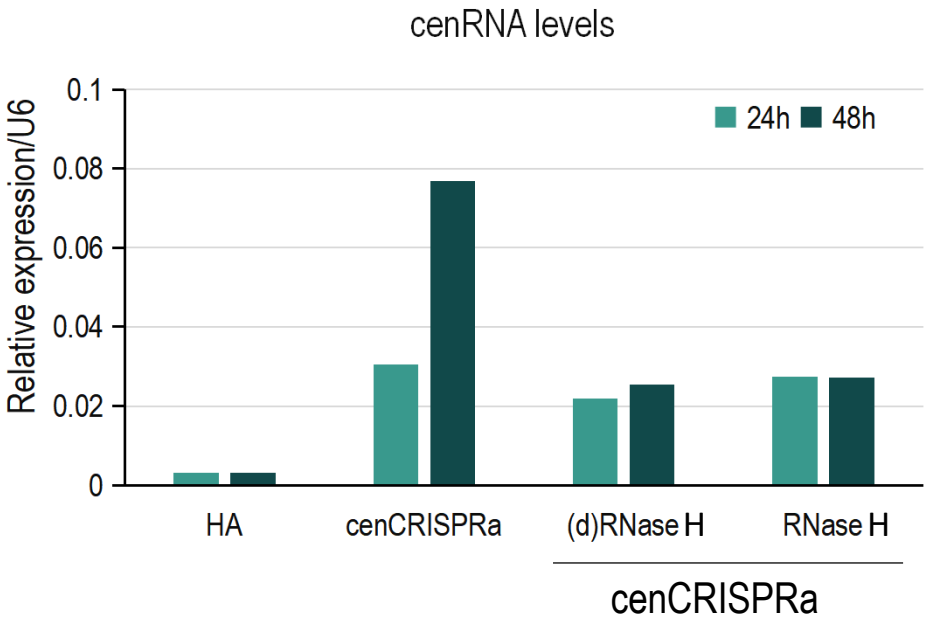


b

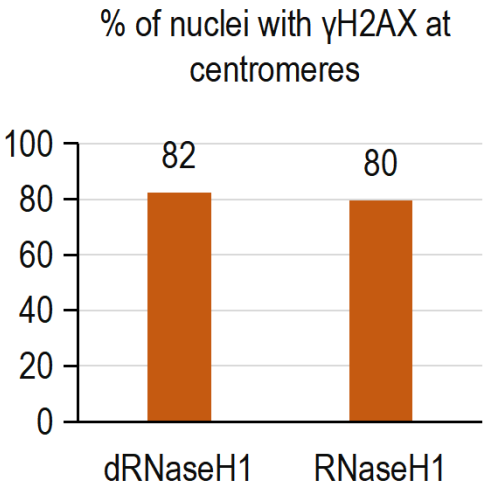


Supplemental figure 4

a



b



Part II: unpublished results

1. Characterization of centromeric transcripts

Transcripts from (peri)centromeric regions are very heterogeneous. Their size and direction of transcription vary according to the cellular context (Vourc'h and Biamonti, 2011). These differences potentially reflect specific transcriptional regulatory mechanisms and multiple roles in response to different stimuli or developmental contexts. Thus, their characterization, in terms of orientation, size and stability, could provide clues on their function in physiologic and pathologic conditions.

The following section illustrates my work on the characterization of cenRNAs, most notably in determining their prevalent orientation and their stability, in order to be able to manipulate their levels for further functional studies (results in the manuscripts and below).

1.1. Murine Forward cenRNA is more abundant than Reverse cenRNA

Tandem repeats do not have canonical promoter regions and have been shown to be transcribed from both the forward and the reverse strand (Probst 2010; Casanova 2013). Previous work by my host team has shown that murine Minor satellite repeats are transcribed in both orientations, i.e. towards the pericentromeric heterochromatin (Forward cenRNAs) or towards the telomere (Reverse cenRNAs) (Bouzinba Segard *et al.*, 2006), as schematized in **Figure 29.a**.

We then questioned whether there was a difference in the levels of Forward cenRNAs or Reverse cenRNAs. We analyzed the strandedness of cenRNA by a strand-specific RT-qPCR, using a primer that will amplify either the Forward or Reverse cenRNAs (relative to a reference sequence published by Kalitsis *et al.*, 2006). In the same reverse transcription reaction, we used U6 primers for normalization of cDNA levels between different samples, as schematized in **Figure 29.b**.

Since cenRNAs are expressed at low levels in a physiological context, we took advantage of a dynamic system that has been previously shown by the team to lead to DNA-damage induced high levels of cenRNAs (Hedouin *et al.*, 2017). For this purpose, we treated NIH-3T3 cells with the genotoxic agent Etoposide (Etop) for 8h, and we analyzed the levels of Forward or Reverse cenRNAs. The qPCR results show that Forward cenRNAs are 3 times more abundant than the Reverse cenRNAs in untreated NIH-3T3 cells (**Figure 29.c**). This difference in

abundance was even more striking after 8h of Etop treatment, where Forward cenRNAs showed 37 times higher levels compared to Reverse cenRNAs (**Figure 29.d**).

However, this difference may reflect the different efficiency of the sense or the antisense primers in the reverse transcription reaction. To validate the predominance of Forward cenRNAs with an orthogonal approach, we performed RNA-FISH with either a probe complementary to the Forward or Reverse cenRNAs. While the cenRNAs are undetectable by RNA-FISH in untreated cells, concordantly with the relatively low levels of cenRNA detected by qPCR (data not shown; Bouzinba-Segard *et al.*, 2006a; Hédouin *et al.*, 2017), in Etop treated cells the signals obtained with both probes form punctiform signals at the periphery of the chromocenters (**Figure 29.e**). However, Etop treated cells show more numerous signals for the Foward cenRNAs compared to Reverse cenRNAs. This again suggests that the Forward cenRNAs are more abundant compared to the Reverse. Interestingly, the less numerous punctiform signals for Reverse compared to the Forward cenRNAs might suggest that cenRNAs are transcribed at the same rate and in both orientations, however not from the same number of centromeres.

In sum, these results indicate that cenRNAs are transcribed in both orientations, in physiological or DNA damage inducing conditions, and in both conditions the Forward cenRNAs are more abundant in the cells.

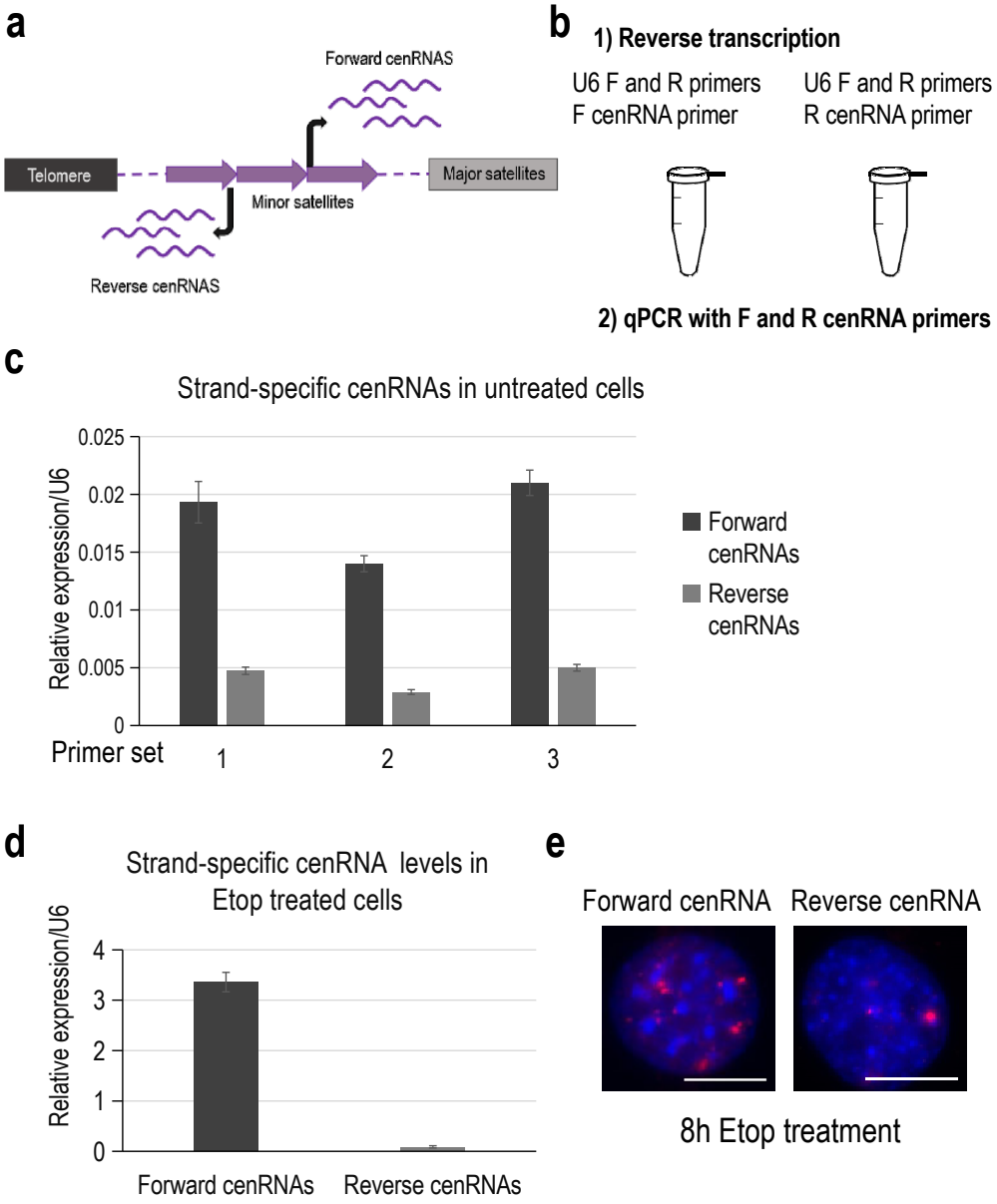


Figure 29. Forward cenRNAs are more abundant than Reverse cenRNAs in murine NIH-3T3 cells

a. Schematic representation of Minor satellite DNA (purple arrows) and cenRNAs (purple wavy lines) that originate from both strands. Forward cenRNAs are transcribed in the direction going towards the Major Satellites (gray rectangle), while Reverse cenRNAs are transcribed in the direction towards telomeres (black rectangle).

b. Schematic representation of a strand-specific RT-qPCR reaction. Reverse transcription

c. Forward or Reverse cenRNA relative levels in untreated (left) or Etop treated (right) NIH-3T3 cells were analysed with strand-specific RT followed by a qPCR, and normalized to U6 expression. n=3

d. Strand-specific RT-qPCR as in c., on NIH-3T3 cells treated with 10uM of Etop for 8h. n=3

e. RNA-FISH for Forward or Reverse cenRNA in NIH-3T3 cells treated with 10uM Etop for 8h. cenRNA probes were labeled with Alexa-Fluor 594, and nuclei were counterstained with DAPI. Images represent one focal plan; n=5.

1.2. Forward and Reverse cenRNAs don't have the same RNA stability

The preliminary data that we obtained suggest that the Forward cenRNAs could be transcribed at higher levels than the Reverse cenRNAs, although this is difficult to address by run-on experiments in the context of tandem repeats. Alternatively, the different abundance of Forward or Reverse cenRNAs could be controlled at a post-transcriptional level, whereby Forward cenRNAs could be more stable than the Reverse ones. To explore the different inherent stability of cenRNAs, we first needed to induce high levels of cenRNAs in both orientations to be able to monitor their decrease in time. For this purpose, we treated cells with Etop for 2h, after which we inhibited transcription using the RNA Pol II inhibitor Actinomycin D, in order to follow the stability of these transcripts. The results suggest that the Forward cenRNAs, while being more highly expressed, are less stable compared to the Reverse cenRNAs, after the first 2h of transcriptional inhibition (**Figure 30.a**). Interestingly, after an initial rapid loss of Forward cenRNA, the levels of Forward and Reverse cenRNAs became equivalent and decreased at the same rate for the rest of the kinetics (4h, 6h and 8h of inhibited transcription). In sum, these results suggest that there is an initial difference in stability where Forward cenRNAs are more abundant, while at 2h of transcription block their levels rapidly decrease and match those of Reverse cenRNAs. This could be explained if Forward and Reverse transcripts formed double-stranded RNAs (dsRNAs). dsRNAs have been shown to have relative stability and affinity for certain proteins and RNA modifications (Barik, 2004; George *et al.*, 2014), which could maintain the stability of cenRNAs over time.

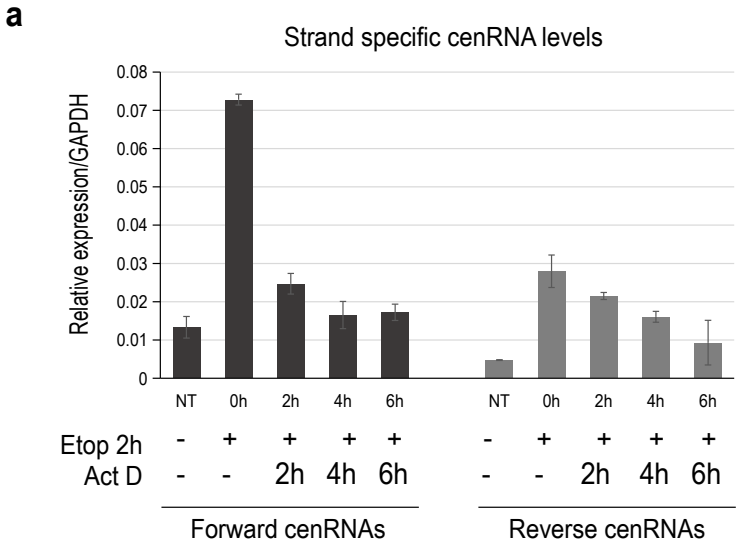


Figure 30. Forward and Reverse cenRNAs don't have the same RNA stability
a. Forward or Reverse cenRNA relative levels were analysed with strand-specific RT followed by a qPCR, and normalized to GAPDH expression as an invariant control. NIH-3T3 cells were treated with 10uM Etop for 2h, after which RNA pol II was blocked using 10uM Actinomycind D (Act D). Cells were collected for RNA extraction in a kinetics of 0h, 2h, 4h and 6h. n=3

2. Determining the functional consequences of cenRNA accumulation on cellular phenotypes

High levels of cenRNAs have been reported in pathological situations with chromosomal instability (Ting *et al.*, 2011). Previous work in the lab has linked the aberrant accumulation of cenRNAs to genome instability in transformed cells (Bouzinba-Segard *et al.*, 2006) while global DNA damage-induced high levels of cenRNAs correlate with permanent cell-cycle arrest in primary MEFs (Hédouin *et al.*, 2017a). High levels of cenRNAs are also an intrinsic characteristic of senescent cells, where the DDR is persistantly activated (Fumagalli *et al.*, 2014; Hédouin *et al.*, 2017). However, the causal link between high levels of cenRNAs and the emergence of pathological phenotypes has not been formally established.

During my PhD, I was interested in answering whether perturbations of cenRNA levels *per se*, without affecting the endogenous centromere transcription, would be detrimental for cells. For this question, I performed the gain of function (GOF) of Forward or Reverse cenRNAs in primary murine embryonic fibroblasts (MEFs) at early passages which feature low levels of endogenous cenRNAs. Conversely, I knocked-down (KD) Forward or Reverse cenRNAs in $p53^{-/-}DNMT1^{-/-}$ murine immortalized fibroblasts which display high levels of endogenous cenRNAs and address the consequences of the manipulations of cenRNAs levels in these cellular models.

2.1. Replicative senescence correlates with high levels of cenRNAs

Since increased levels of cenRNAs correlated with permanent cell-cycle arrest and senescence (Hédouin *et al.*, 2017a), we used the well-established system of replicative senescence observed in primary MEFs that enter senescence around passage 5-6 (Parrinello *et al.*, 2003). Since reactive oxygen species (ROS) in culture are known to contribute to this progressive entry into senescence (Colavitti, 2015), primary MEFs were cultured in presence or absence of β-Mercaptoethanol (β-Me), a reducing agent that acts as a biological antioxidant by scavenging hydroxyl radicals and prevents the accumulation of ROS in culture. At each passage, cells were collected for RNA extraction and cenRNAs levels were analyzed by RT-qPCR. We confirmed that culturing cells in the presence of β-Me decreased the expression $p16^{ink4a}$, a marker of senescence, which were globally lower in cells cultured with β-Me. This was the case even at passage 8 of cells cultured with β-Me, where the levels were comparable to those at passage 3 in the absence of β-Me (**Figure 31.a**). Next, we assessed

cenRNA levels and found that, although they increased in both conditions and with the increasing number of passages, their levels became twice higher in the absence of β-Me at late passages (p7, p8) (**Figure 31.b**).

These results therefore support a correlation between the onset of replicative senescence and increased levels of cenRNAs in primary MEFs.

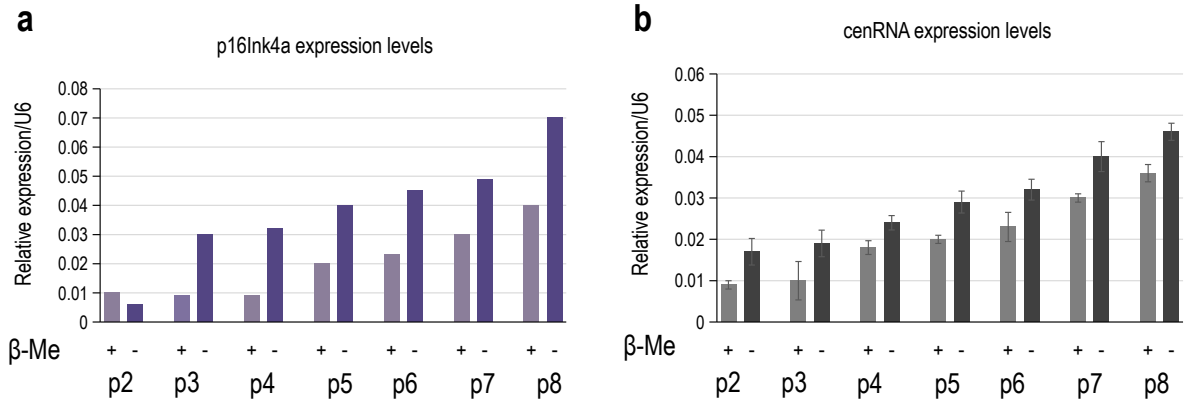


Figure 31. Replicative senescence correlates with increase of cenRNA levels
a. p16INK4a relative expression levels from primary MEFs from cell culture passage 2 to passage 8 (p2,p8), were analysed with RT-qPCR, and normalized to U6 expression. n=1
b. cenRNA relative expression levels analysed from the same samples as Figure.3.a using Minor Satellite primers. n=3

2.2. Overexpression of ectopic cenRNAs results in an increase of $p16^{Ink4a}$ expression levels

Given the correlation between high cenRNA levels and senescence, a question that arises is whether the overexpression of cenRNAs could be sufficient to promote permanent cell-cycle arrest. To test this hypothesis, we transfected primary MEFs at passage 2 (p2) with vectors driving the expression of either Forward or Reverse cenRNA, under the control of the CMV promoter, or vector driving expression of GFP as a control.

Knowing that entry into senescence takes at least five days for murine cells, and since primary MEFs show low transfection efficiency, we performed three transfection rounds over 8 days, with the kinetics shown in **Figure 32.a**.

The levels of Forward or Reverse cenRNAs increased at the same rate after each transfection round (d1, d3, d5), reaching 800 times higher levels compared to cells transfected with GFP, which exhibit only the endogenous cenRNAs (**Figure 32.b**). To test whether ectopic

overexpression of cenRNAs leads to premature senescence, I assessed the transcript levels of $p16^{Ink4a}$ (**Figure 32.c**). As expected, at d1 the levels of $p16^{Ink4a}$ are low in all samples. After two transfection rounds (d3), a 2 fold increase of $p16^{Ink4a}$ was observed in cells transfected with Forward cenRNA compared to the GFP transfected control primary MEFs at the same passage. Finally, after three transfection rounds (d5), there is more than a 2-fold increase of $p16^{Ink4a}$ in cells transfected with either Forward or Reverse cenRNAs, compared to the control MEFs, suggesting that accumulation of exogenously expressed cenRNAs might promote entry into senescence of primary MEFs. However, despite the increase in $p16^{Ink4a}$ expression, the cells do not appear senescent according to the β-galactosidase staining, the gold standard assay for senescence, indicating that probably a longer culturing time is needed to observe senescence (data not shown). Moreover, $p16^{Ink4a}$ is a late marker of senescence and is required for the maintenance of the proliferation arrest, while the $p21^{Cip1}$ protein is an early marker and is considered necessary for the initiation of senescence (Stein *et al.*, 1999), which makes it a more suitable marker to analyze in the kinetics we performed.

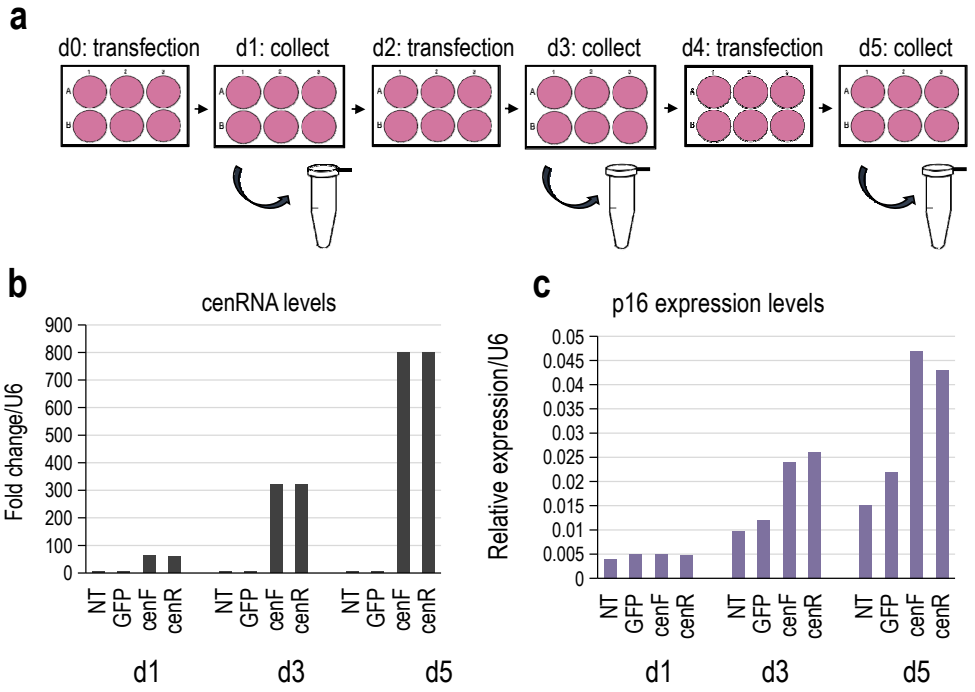


Figure 32. Overexpression of ectopic cenRNAs leads to increase in p16Ink4a expression levels
a. schematic representation of the kinetics of transfection. Cells were transfected on d0 and passaged and collected at d1. The second round of transfection was performed on d2, and the cells were passaged and collected again on d3. The last round of transfection was performed on d4, and all of the cells were collected on d5. **b.** Transcript levels of cenRNAs from primary MEFs after one transfection (d1), two transfections (d3) and three transfection (d5) rounds with vectors driving the expression of non treated (NT) cells or cells transfected with either GFP, Forward cenRNA (cenF) or Reverse cenRNA (cenR) were analysed by RT-qPCR and normalized with U6 expression. n=2
c. Transcript levels of p16Ink4a were analyzed by RT-qPCR from the same samples as in **b**. n=2

2.3. Knockdown of Forward or Reverse cenRNA impacts cellular phenotypes

To question whether cells which highly express cenRNAs could survive a decrease in their levels, we performed the knock-down of Forward or Reverse cenRNAs using antisense oligonucleotides called LNA-gapmers (Alfeghaly *et al.*, 2021). When they hybridize with complementary RNA, they form a substrate recognized by RNase H, hence providing an efficient mechanism of knocking down transcripts both in the nucleus and the cytoplasm. We used two different LNA-gapmers to target the Reverse cenRNAs (LNA_R1 and LNA_R2), and one LNA-gapmer to target Forward cenRNAs (LNA_F), or a scrambled LNA as control (schematized in **Figure 33.a**). We performed the KD of cenRNAs in $p53^{-/-}DNMT1^{-/-}$ immortalized fibroblasts (iMEFs) because they display high levels of cenRNAs observed in RT-qPCR, compared to other murine cells available in the team **(Figure 33.b**). We transfected the LNA-gapmers in $p53^{-/-}DNMT1^{-/-}$ iMEFs and assessed cenRNA levels 24h after transfection by strand specific RT-qPCR. Knockdown with LNA_F was very efficient and it led to an 85% decrease of Forward cenRNAs compared to control cells (**Figure 33.c**). Similarly, LNA_R1 and LNA_R2 promote a 90% decrease of Reverse cenRNAs (**Figure 33.d**). Surprisingly, knocking down the Forward cenRNAs had the biggest phenotypic effect, where cells exhibited abnormal morphology compared to untreated cells (**Figure 33.e i**). Loss of Forward cenRNA resulted in accumulation of vesicles in the cytoplasm (**Figure 33.e ii**) and ultimately led to cell death. This outcome is not entirely surprising since these cells lack the p53 protein and in theory cannot undergo senescence. Interestingly, cells transfected with LNA_R1 were dying to a lesser extent **(Figure 33.e iii**). This observation, along with the previous finding of the different activation of Forward and Reverse cenRNAs in response to genotoxic stress (Figure 1.d), suggests that Forward and Reverse cenRNAs might have different functions. It is tempting to imagine that Forward cenRNAs could act as sensors of cellular or centromeric stress, like the pericenRNAs from Sat III of chromosome 9 in response to heat-shock in human cells (Jolly *et al.*, 2004; Biamonti and Vourc'h 2010). While nuclear stress bodies, that form on Sat III DNA in human cells, do not form in murine cells (Hédouin *et al.*, 2017a), Forward cenRNAs could function as soluble sentinels and recruit opportune proteins to centromeres, for example in case of DNA damage, DNA repair factors which are abundant at this locus (Di Paolo *et al.*, 2014; Zhu *et al.*, 2018c). Consequently, a rapid loss of cenRNAs (24h) could result in the lack of recruitment of these factors and leave centromeres unprotected.

The cellular defects observed upon Forward cenRNA KD might be explained in light of a study that has linked the loss of cenRNA, over 6 days in culture, with impaired CENP-A

loading, mitotic errors and abnormal morphology of human cells (Quénet and Dalal, 2014). Therefore it would be interesting to analyze the localization of Cenp-A upon KD of cenRNAs. Moreover, it would help elucidate whether Forward and Reverse cenRNAs have different functions, where one, and not the other, could be essential for Cenp-A targeting to centromeres.

The more drastic phenotypic effect was caused by the KD of Forward cenRNA which affected cell morphology, whereby cells appeared more flat and This is reminiscent of the outcome of cenRNA KD in human cells, where cells appeared less dense (Quénet and Dalal, 2014). Since centromeres of human cells have distinct sequences on different chromosomes (Miga, 2019), the KD might not impact all of the cenRNA population (cenRNAs with distinct sequences), hence hampering the impact on centromere loss of identity and function. Conversely, murine centromere sequences are homogeneous between chromosomes (Kalitsis *et al.*, 2006), and it is possible that the KD with the LNAs we used abolished Forward (or Reverse) cenRNAs derived from all of the chromosomes, therefore greatly impacting centromeres. Yet, the appearance of cellular defect after 24h of Forward cenRNA KD is striking. While LNA_F was predicted not to have off-target effects, we cannot exclude this possibility. Importantly, LNA-gapmers have been reported to hybridize with DNA (Vester and Wengel, 2004). Therefore this could lead to indirect effects that perturb cellular phenotypes. Moreover, since RNaseH activity is also involved in the resolution of DNA-RNA hybrids (Hyjek *et al.*, 2019), the sequestration of RNaseH by LNA-gapmer-RNA hybrids may then hamper clear conclusions of causal links. Hence, to study the effect of downregulation of Forward or Reverse cenRNAs, a more suitable approach would be to inhibit specifically the transcription of endogenous centromeric repeats using strand-specific gRNAs and dCas9 fused to the transcriptional repressor Kruppel-associated box domain (KRAB).

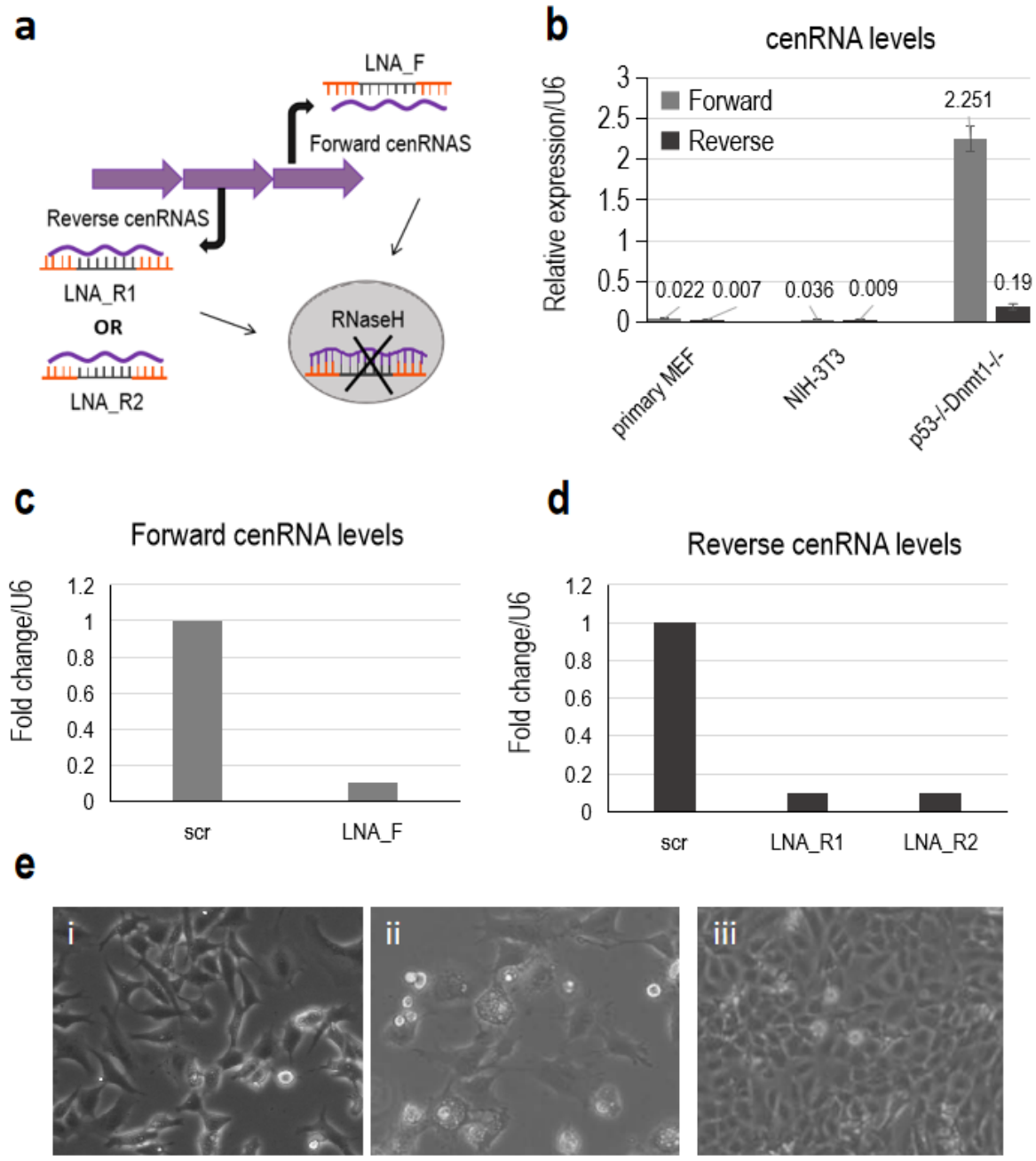


Figure 33. Knockdown of Forward cenRNAs perturbs cell viability

a. Schematic representation of LNA-gapmeres with LNA nucleotides in orange and DNA nucleotides in black. LNA-gapmers hybridize with target RNA and are recognized and cleaved by RNaseH.

b. Strand-specific analysis of cenRNAs expression levels in different mouse cell lines. Strand-specific RT-qPCR was performed to detect Forward or Reverse cenRNA levels in primary MEFs at passage 2 (p2), NIH-3T3 cells or p53-/-Dnmt1-/- cells. cenRNA expression was normalized to U6. n=3

c, d. Knockdown efficiency of strand-specific LNA-gapmeres. Knockdown of Forward cenRNA with LNA_F gapmers (c) or Reverse cenRNA with two different LNA gapmers (d) was assessed with strand-specific RT followed by a qPCR, and normalized to U6 expression. n=1

e. Effect of the stand-specific knockdown on cell viability. Phase contrast light microscopy images of p53-/-Dnmt1-/- fibroblasts. Magnification 10x. i. untransfected cells. ii. transfected with LNA_F for 24h. iii. transfected with LNA_R1 for 24h. n=1

3. Determining the functional consequences of increased centromere transcription on centromere architecture and function

Low transcriptional activity is a characteristic of centromeric repeats in normal somatic cells and it is required for proper centromere function which includes correct Cenp-A incorporation (Bobkov *et al.*, 2018) and correct segregation of sister chromatids in mitosis (Chan *et al.*, 2012; Kabeche *et al.*, 2018). Conversely, strongly elevated centromeric transcription, achieved by targeting of a strong trans-activation domain from Herpes simplex virus (VP16) to centromeres of human artificial chromosomes (HACs), is not compatible with centromere function as it leads to CENP-A eviction (Bergmann *et al.*, 2012). Hence, even though transcription through centromeric repeats is necessary for centromere homeostasis, increased transcription at this locus is detrimental for centromere identity and function. However, the molecular mechanisms that bridge increased transcription and loss of centromere function have not been fully explored. Hence, we specifically targeted increased centromere transcription in NIH-3T3 cells and we assessed its impact on centromeres (described in detail in the published results section). Briefly, we have shown that increased centromere transcription in interphase leads to the formation of DNA-RNA hybrids and DNA damage at centromeres at early timepoints (24h of increased transcription). At later timepoints (48h), this culminates in the perturbation of centromere architecture (disorganization of centromeric DNA and centromeric proteins Cenp-A and Cenp-B) and loss of centromere function (appearance of mitotic errors). However, it is the presence of the DNA-RNA hybrids, but not DNA damage, which is a major cause of centromere disorganization and mitotic errors, since their resolution restores normal phenotypes and mitotic function.

3.1. Setting up the CRISPRa system to activate transcription specifically at centromeres

To target increased transcription at centromeres, we used the CRISPR activation (CRISPRa) strategy, which relies on the catalytically dead Cas9 (dCas9) fused to transcriptional activators, whose targeting to a gene of interest leads to upregulated transcription of the locus (Balboa *et al.*, 2015; Chakraborty *et al.*, 2014; Kearns *et al.*, 2014). However, today there are plenty of different CRISPRa systems available (Chen and Qi, 2017) and the choice of which CRISPRa system to use has to be experimentally tested. Hence, to set up this system in NIH-3T3 cells, we tested three different CRISPRa systems: i. dCas9-VP64 (Konermann *et al.*, 2015), ii. dCas9-VP160 (Perrin *et al.*, 2017), and iii. dCas9-VP64/p65-Hsf (Chong *et al.*, 2018).

dCas9-VP64 and dCas9-VP160 consist of four and ten copies of the acidic domain of herpes simplex viral protein 16 (VP16), respectively. On the other hand, the dCas9-VP64/p65-Hsf is a tripartite system that is composed of dCas9-VP64, a gRNA harboring an MS2 stem-loop and the MS2 bacteriophage coat (MS2) protein fused to p65 transactivating subunit of NF-kappa B and human heat-shock factor 1 (HSF) activation domain. Consequently, when dCas9-VP64 is targeted to centromeres with the gRNA, the stem-loop in the gRNA is bound by the MS2 protein-p65-Hsf, bringing the entire complex to centromeres (Gaj *et al.*, 2013).

To target CRISPRa to murine Minor satellite repeats we used the previously published gRNA (MiSgRNA) (Schmidtmann *et al.*, 2016), and we transiently transfected it along with plasmids expressing either one of these systems (VP64, VP160 or VP64/p65-Hsf), using lipofectamin based reagents, which resulted in the efficiency of transfection of 60%, estimated by counting the number of cells with dCas9 localizing at centromeres in IF (**Figure 34.a**). We then assessed cenRNA levels by RT-qPCR at 24h post-transfection. VP64 and VP160 lead to a 10 and 160 fold increase of cenRNA levels compared to non-treated cells (NT), respectively. Strikingly, the VP64/p65-Hsf transfection resulted in more than a 1000 fold increase in cenRNA levels (**Figure 34.b**). Therefore, while all three CRISPRa systems tested resulted in upregulation of cenRNA levels, VP160 and VP64/p65-Hsf resulted in highly increased cenRNA levels, similar to those observed in conditions of genotoxic stress that are detrimental for centromere identity and cell viability (Hédouin *et al.*, 2017a). Next, we assessed centromere identity and architecture by visualizing Cenp-A in immunofluorescence after 24h of CRISPRa transfection (**Figure 34.c**). VP64-induced transcription did not perturb centromere architecture at 24h post-transfection, since Cenp-A presented a punctiform pattern around chromocenters in 60% of the cells (**Figure 34.c, top**). Conversely, VP160-induced transcription led to perturbation of centromere architecture, visualized as stretching of Cenp-A signal in 52% of cells (**Figure 34.c, middle**). Finally, the elevated centromeric transcription levels caused by VP64/p65-Hsf also led to perturbation of centromere architecture and formation of micronuclei in 66% of cells (**Figure 34.c, bottom**). Moreover, 48h of expression of the different CRISPRa systems and MiSgRNA severely impacted cell viability, where cells transfected with VP64 started dying and proliferating less compared to cells transfected only with dCas9-VP64 without MiSgRNA, hence not targeted to centromeres (**Figure 34.d i and ii**). On the other hand, both conditions of highly increased transcription (VP160 and VP64/p65-Hsf) resulted in cell death 24h post-transfection (data not shown). Hence, we concluded that VP64 is the optimal CRISPRa system among the ones we tested

since its use resulted in the increase of cenRNA to levels comparable to those found in pathological conditions *in vivo*, and concurrently it gives us a window of time (48h before mitotic errors and cell death, results presented in the manuscript) to test the functional consequences of increased centromere transcription on centromere function and ultimately cellular outcomes.

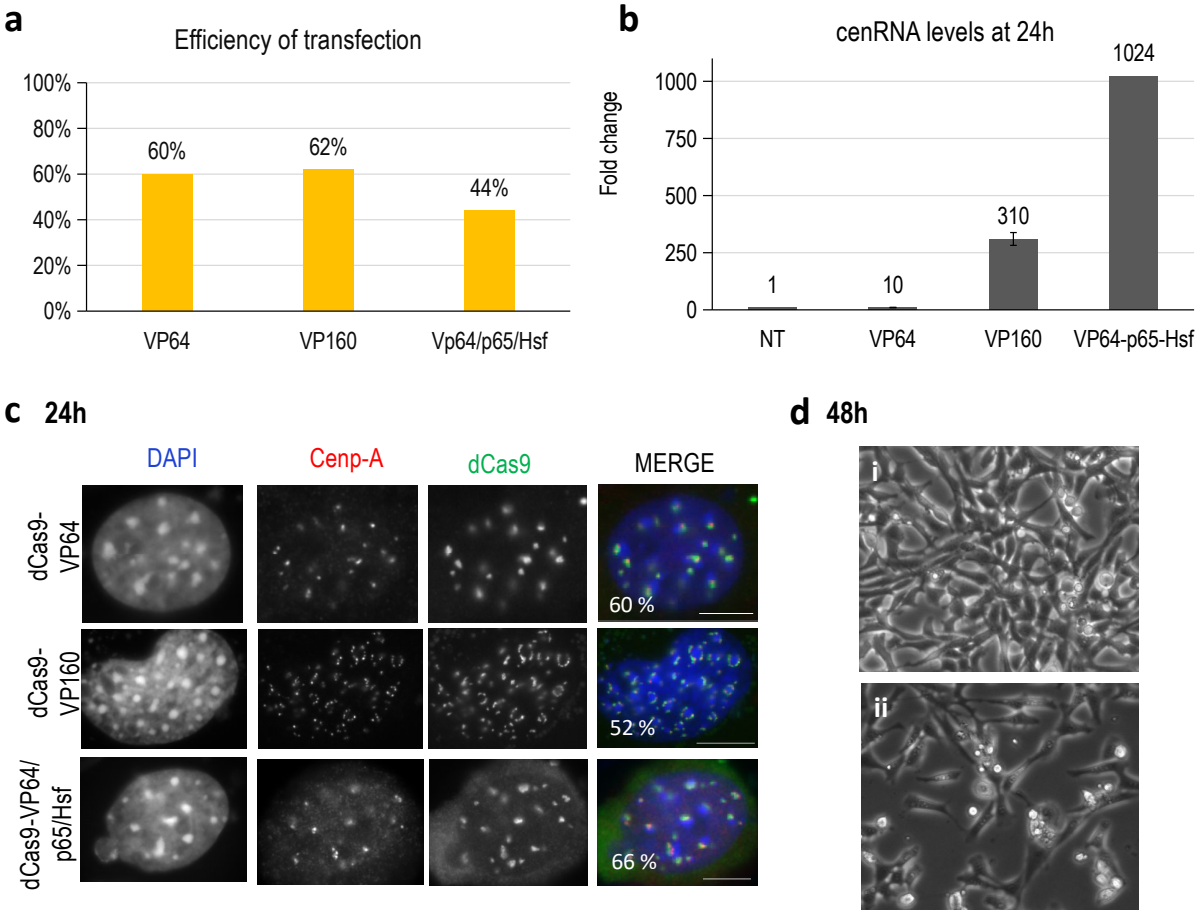


Figure 34. Comparison of different CRISPRa systems.
a. Efficiency of transfection estimated by counting the number of cells displaying CRISPRa at centromeres in IF, as in 34.c.
b. Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA in NIH-3T3, upon activation of centromeric transcription by different systems of CRISPRa (same as in Figure 26.a) for 24h.
c. Representative images of cells expressing different types of CRISPRa systems for 24h: (top) dCas9-VP64, (middle) dCas9-VP160, (bottom) dCas9-VP64/P65/Hsf. Cells were stained with DAPI and antibodies specific for Cas9 and Cenp-A. Quantification indicates the percentage of cells with the represented pattern of Cenp-A. nuclei counted= 50
d. Phase contrast light microscopy images of NIH-3T3 cells transiently transfected for 48h, with dCas9-VP64: i. without MiSgRNA or ii. with MiSgRNA. Magnification 10x.

3.2. Uncoupling the consequences of CRISPRa and cenRNA accumulation, on centromere loss of function

We have shown that increased centromere transcription leads to loss of centromere function (published section results). However, since centromere transcription produces cenRNAs, with this system we were not able to uncouple which of these two components leads to dysfunction of centromeres. Hence, we compared the CRISPRa with a GOF of ectopic cenRNAs in the same cellular system. We transfected NIH-3T3 cells with the plasmid that drives the expression of cenRNAs from a CMV promoter, or a plasmid expressing GFP as a control, as in **Figure 32,** which yielded a transfection efficiency of 67% as assessed by the number of cells positive for cenRNAs seen in RNA-FISH (**Figure 35.b**). We assessed the levels of cenRNAs 48h after. The levels reached 228 fold compared to GFP transfected or non treated cells which exhibit only low levels of endogenous cenRNAs (**Figure 35.a**). We next analyzed cenRNA localization using RNA-FISH. While my host team has previously shown that GOF of cenRNA in transformed Murine erythroleukemia (MEL) cells resulted in their localization at chromocenters (Bouzinba Segard *et al.*, 2006), in NIH-3T3 cells it led to cenRNA accumulation in the entire nucleus (**Figure 35.b**). Interestingly, GOF of ectopic cenRNAs did not result in the appearance of the DNA damage marker γH2AX at centromeres (**Figure 35.c**), in contrast to what has been shown in human cells (Zhu *et al.*, 2011). However, in line with previous studies using GOF of cenRNAs, we observed the appearance of lagging chromosomes in mitosis (**Figure 35.d**), possibly by tethering centromeric and CPC proteins away from centromeres (Bouzinba-Segard *et al.*, 2006). Yet, since cenRNAs accumulate in the entire nucleus, we cannot exclude indirect effects that result in mitotic errors, therefore this system hampers the establishment of causal links.

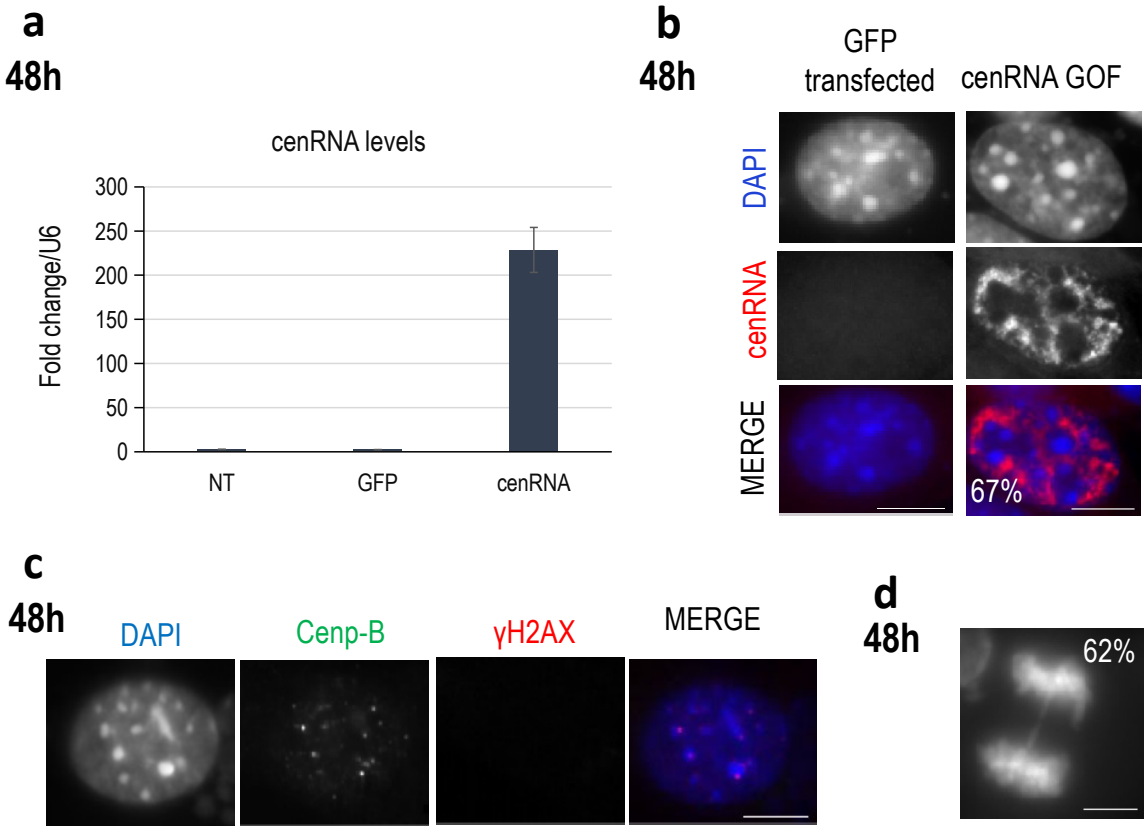


Figure 35. Overexpression of ectopic cenRNAs results in mitotic errors.
a.Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA in NIH-3T3, upon GOF of ectopic cenRNAs for 48h. n=3
b. RNA-FISH for detection of cenRNA in NIH-3T3, 48h after transfection with a plasmid expressing ectopic cenRNAs. Oligo probes were labeled with Alexa-Fluor 594 and nuclei were stained with DAPI. Quantification indicates the percentage of cells with the cenRNA signal. nuclei=50;
c. Immunofluorescence in NIH-3T3 cells transfected with plasmids expressing cenRNAs for 48h. Cells were stained with DAPI and antibodies specific for Cenp-A and γH2AX. n=50;
d. DAPI staining of a mitosis in NIH-3T3 cells transfected with plasmids expressing cenRNAs for 48h. mitoses=15

3.3 Targeting RNAs directly to centromeres leads to DNA damage at this locus

To dissect the contribution of cenRNA accumulation to loss of centromere function, we decided to target cenRNAs directly at centromeres, by adapting the CRISPR-Display method (Shechner *et al.*, 2015). We used dCas9-HA and a MiSgRNA that carries at its 5' end an accessory domain corresponding to either cenRNA of 120 nt (one repeat unit), or a 120 nt fragment of the luciferase RNA (lucRNA) as a control (**Figure 36.a**). As opposed to a gain of function where cenRNAs accumulate in the entire nucleus (**Figure 35.b**), this approach led to the targeting of cenRNA directly to centromeres in 19% of cells, as visualized by RNA-FISH (**Figure 36.b**). Moreover, the localization of cenRNAs at centromeres was assessed indirectly by detecting the co-localization of dCas9-HA and Cenp-A (**Figure 36.c bottom two rows**). Surprisingly, the targeting of lucRNA to centromeres resulted in the accumulation of DNA damage marker γH2AX at centromeres 24 h post-transfection with lucRNA-fused gRNA in 22 % of interphase cells (**Figure 36.c top two rows**). This observation suggested that the excessive presence of non-specific RNA affects centromere physiology. Notably, the presence of non-modified gRNA (**Figure 3.a in the manuscript**) or gRNA presenting a short (20 nt) MS2 loop without the additional RNA fragment (data not shown) did not increase the presence of DNA damage at centromeres, arguing that it is the presence of an additional RNA moiety that is not bound by dCas9 protein (Shechner *et al.*, 2015), that gives the deleterious effect. Interestingly, recruitment of the cenRNA to the centromeres was less deleterious since we detected co-localization of γH2AX with Cenp-A only in 9 % of transfected cells (**Figure 36.c bottom two rows**). The last observation might suggest that the recruitment of the sequence-specific factors has a protective role on centromeres, while the accumulation of general RNA-binding proteins has a negative effect.

Taken together with the previously published observation from my host lab and others (Bury *et al.*, 2020; Ferri *et al.*, 2009; Quénet and Dalal, 2014), these preliminary results are in agreement with the paradigm that cenRNAs are an important structural component of the centromeric chromatin, further validation of these tools and especially the design of appropriate control is needed to reach solid conclusions.

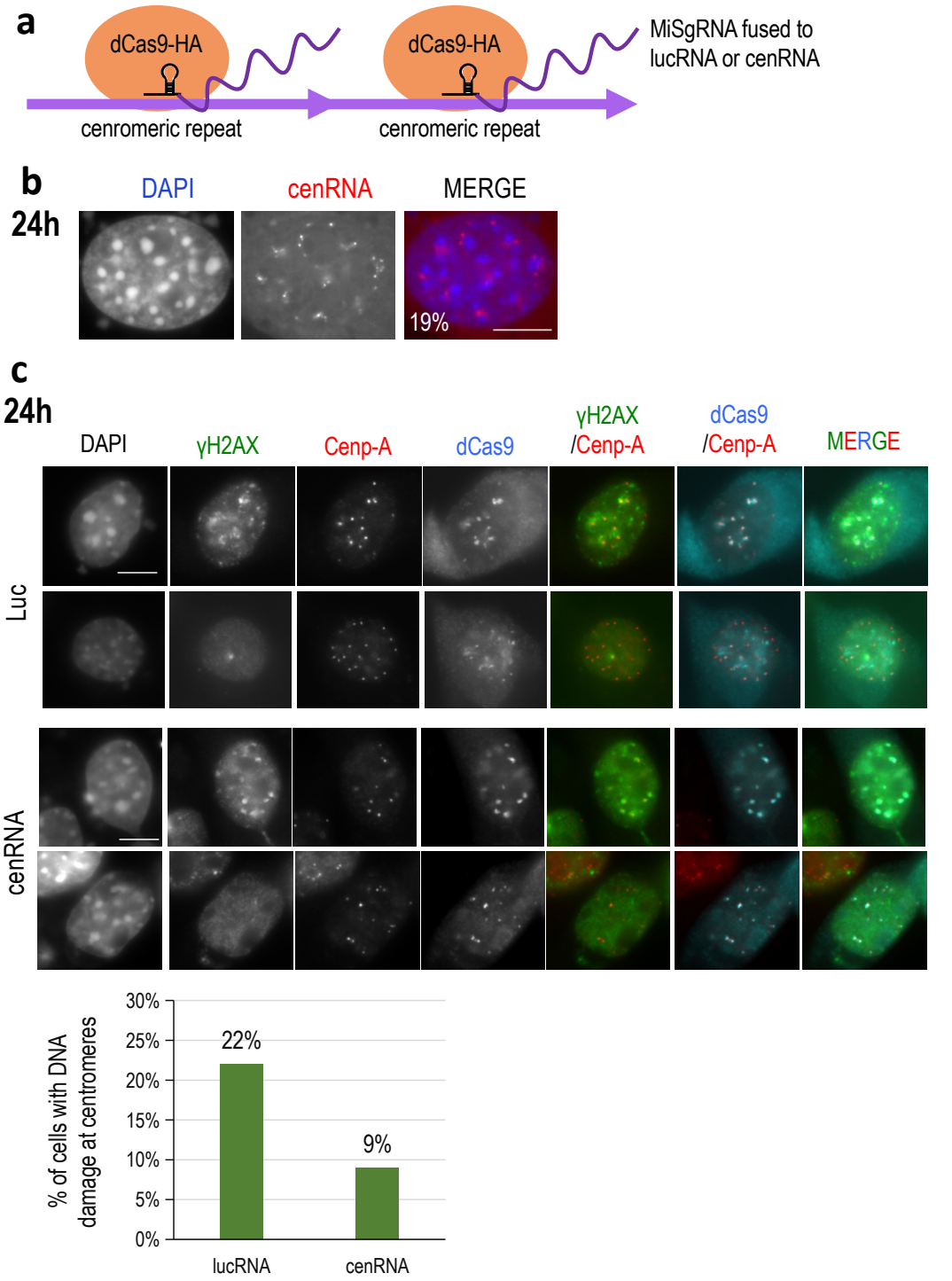


Figure 36. Targeting RNAs to centromeres results in γH2AX appearance at this locus
a. schematic representation of the CRISPR display tool. dCas9-HA protein (orange circle) is recruited to murine Minor satellite repeats (purple arrows) by an engineered MiSgRNA fused to lucRNA (control) or cenRNA. **b**. RNA-FISH for detection of cenRNA in NIH-3T3 expressing dCas9-HA in presence of MiSgRNA fused to cenRNA, at 24h post transfection. Oligo probes were labeled with Alexa-Fluor 594 and nuclei were counterstained with DAPI. Quantification indicates the percentage of cells positive for cenRNA; nuclei=30 **c**. Immunofluorescence in NIH-3T3 cells expressing dCas9-HA and MiSgRNA fused to cenRNA for 24h. Cells were stained with DAPI and antibodies specific for HA (to recognize dCas9-HA), Cenp-A and γH2AX. Histogram represents the percetanges of described signals. nuclei=22 for lucRNA and nuclei=12 for cenRNA.

4. Deciphering the link between DNA hypomethylation and centromere transcription

At DNA repeats, DNA methylation has been implicated in the inhibition of recombination between repeats, in part through maintaining these elements in a silent state (reviewed in Saksouk *et al.* 2014; Francastel and Magdinier 2019; Scelfo and Fachinetti 2019). Loss of DNA methylation at satellite repeats, which is a frequent feature of pathological cases like cancer, also correlates with increased levels of satellite RNAs (Bouzinba-Segard *et al.*, 2006; Ting *et al.*, 2011; Eymery *et al.*, 2009). However, the functional link between the DNA methylation status and centromere transcription has never been fully characterized. During my PhD, I set up to test this functional relationship and to determine whether DNA hypomethylation is necessary or sufficient for increased centromere transcription.

4.1. Increased centromere transcription does not lead to loss of centromeric DNA methylation

DNA repeats are methylated in somatic tissues (Schubeler, 2015), but this does not preclude their timely transcription which is necessary for correct centromere function (Topp *et al.*, 2004; Ferri *et al.*, 2009; Quénet and Dalal, 2014). Moreover, DNA methylation at centromeres does not impede induced centromere transcription and accumulation of cenRNAs (see paper results section). In this context, we observed the formation of R-loops at centromeric repeats. Since R-loops have been shown to protect gene promoters against DNA methylation by recruiting the demethylase TET1 (Arab *et al.*, 2019), and since DNMT1 has a lower affinity for DNA-RNA hybrids compared to dsDNA (Grunseich *et al.*, 2018), we questioned whether the appearance of centromeric R-loops upon increased transcription would promote loss of DNA methylation at these repeats.

To analyze the DNA methylation status at centromeres, we performed a southern blot on genomic DNA from NIH-3T3 cells transduced with dCas9 (fused to either VP64 or HA), or non-treated cells (NT), at 24h and 48h timepoints. As a hypomethylated control for the southern blot, we used genomic DNA from murine embryonic stem cells triple knock out for the three Dnmts (ES TKO). We digested genomic DNA of the samples with HpaII, a methylation sensitive enzyme that cuts only hypomethylated DNA, and we hybridized the blots with probes complementary to murine Minor satellites centromeric repeats. The TKO display the typical ladder of satellite repeats, whose hypomethylated status is visualized by the accumulation of short fragments (bands of size around 120 - 720 bp) (**Figure 37**).

Conversely, NIH-3T3 cells, in all tested conditions, display bands of much lower intensity due to their methylated status at centromeres. Even at 48h of increased transcription, the Minor satellite bands show comparable intensity to the controls, suggesting that increased centromere transcription and DNA-RNA hybrid formation (data presented in the manuscript) do not lead to loss of DNA methylation at this timepoint.

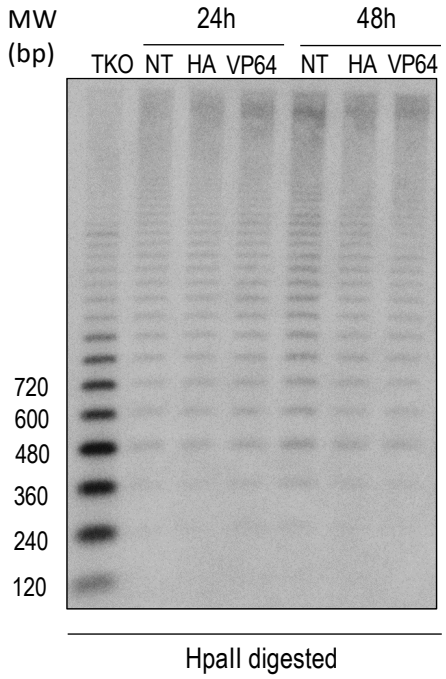


Figure 37. Increased centromere transcription doesn't impact the DNA methylation status at centromeres. Southern blot analysis of DNA methylation at minor satellites in NIH-3T3 cells expressing dCas9 either fused to HA (negative control) or VP64 in the presence of MiSgRNA for 24 or 48 h. Genomic DNA was digested with HpaII enzyme and hybridized with centromere specific probes. Genomic DNA from non transfected NIH-3T3 and murine embryonic stem cells TKO triple knock-out for DNMTs were used as negative and positive controls, respectively. n=1

4.2. The relationships between DNA hypomethylation at centromeres, Zbtb24 as a transcription regulator and centromeric transcription

The ICF syndrome provides a remarkable monogenic context where altered DNAme at tandem satellite repeats is directly linked to compromised centromere integrity (Xu *et al.*, 1999). The ICF2 is caused by mutations in the transcription factor Zbtb24 (de Greef *et al.*, 2011) and it is characterized by DNA hypomethylation at both centromeric and pericentromeric repeats in mouse and human (Ehrlich, 2003; Thijssen *et al.*, 2015; Velasco, Grillo *et al.*, 2018). Interestingly, we have recently shown that Zbtb24 occupies centromeric satellite DNA in wild-type mouse and in human cells (Grillo *et al.*, 2022 - manuscript in preparation, joint as Annex). Moreover, we and others (Grilo *et al.*, 2022; Thompson *et al.*,

2018) have shown that Zbtb24 is a transcription regulator, that mostly functions as a transcriptional activator at unique genomic loci. Given the occupancy of Zbtb24 at centromeres, this opens the question of its possible role in transcriptional regulation of the underlying repeats. To investigate this, I took advantage of mouse embryonic stem cells (mESCs) from *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ blastocysts that were generated in the team (Grillo *et al.*, 2022), due to their hypomethylated status at centromeres. Indeed, as opposed to somatic cells, mESCs show only between 20% and 30% of methylated CpGs in their genome when they are cultured in so-called "2i conditions" (Habibi *et al.*, 2013; Stadler *et al.*, 2011). "2i" refers to the addition of two inhibitors of mitogen-activated protein kinase (MEK) and glycogen synthase kinase-3 (GSK3), and the absence of exogenously added serum and growth factors (Ying *et al.*, 2008), in which mESCs are in a "ground state" of pluripotency (Wray *et al.*, 2010; Ying *et al.*, 2008). Therefore, *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs represent an appropriate cellular context to explore the relationship between loss of the transcriptional factor Zbtb24, DNA hypomethylation at centromeres and cenRNA levels.

We first validated the hypomethylation at centromeres of the *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs cultured in 2i, by performing Southern blot and using the genomic DNA from mESCs TKO as a positive control for hypomethylation. Surprisingly, the Southern blot revealed that hypomethylation of Minor satellites in *Zbtb24*$^{mt/mt}$ mESCs was even more pronounced than in their wild-type counterpart *Zbtb24*$^{wt/wt}$ (**Figure 38.a** and **Figure 5** in the manuscript in the Annexe). We then questioned if the lower levels of centromere methylation would lead to an increase in cenRNA levels, hence we performed RT-qPCR to detect cenRNAs in *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs. Interestingly, we observed that *Zbtb24*$^{mt/mt}$ mESCs display 2.3 fold increase in cenRNA levels compared to the control *Zbtb24*$^{wt/wt}$ mESCs (**Figure 38.b**). Therefore, stronger DNA hypomethylation at centromeres correlates with slightly higher cenRNA levels. Moreover, the higher cenRNA levels in *Zbtb24*$^{mt/mt}$ mESCs suggest that Zbtb24 might act as a transcriptional repressor at centromeric repeats and its loss of function might lead to higher centromere transcription or this increase in cenRNA levels could be the result of the stronger DNA hypomethylation at centromeres (**Figure 38.a**).

To have a more dynamic system to explore the role of Zbtb24 as a transcription repressor of centromeric repeats, we generated a knock-in mESC cell line for the rapid, and reversible, degradation of the protein of interest. To this end, we used genome editing to add the auxin-inducible degron (AID) tag (Holland *et al.*, 2012; Nishimura *et al.*, 2009) to

endogenous Zbtb24 in mESCs to achieve rapid but reversible depletion of Zbtb24, which we recently validated and characterized (Grillo *et al.*, 2022).

Notably, we performed a kinetics of 0h, 24h and 48h of Zbtb24 depletion by auxin treatment in 2i condition and we analyzed cenRNA levels by RT-qPCR. cenRNA levels remained relatively unchanged between conditions (**Figure 38.c**), indicating that Zbtb24 does not function as a transcriptional regulator at centromeric repeats in 2i. This finding further suggests that the slight increase in cenRNA levels seen in *Zbtb24*$^{mt/mt}$ mESCs is a result of DNA hypomethylation at centromere which in this cellular model represents a permissive context for centromere transcription.

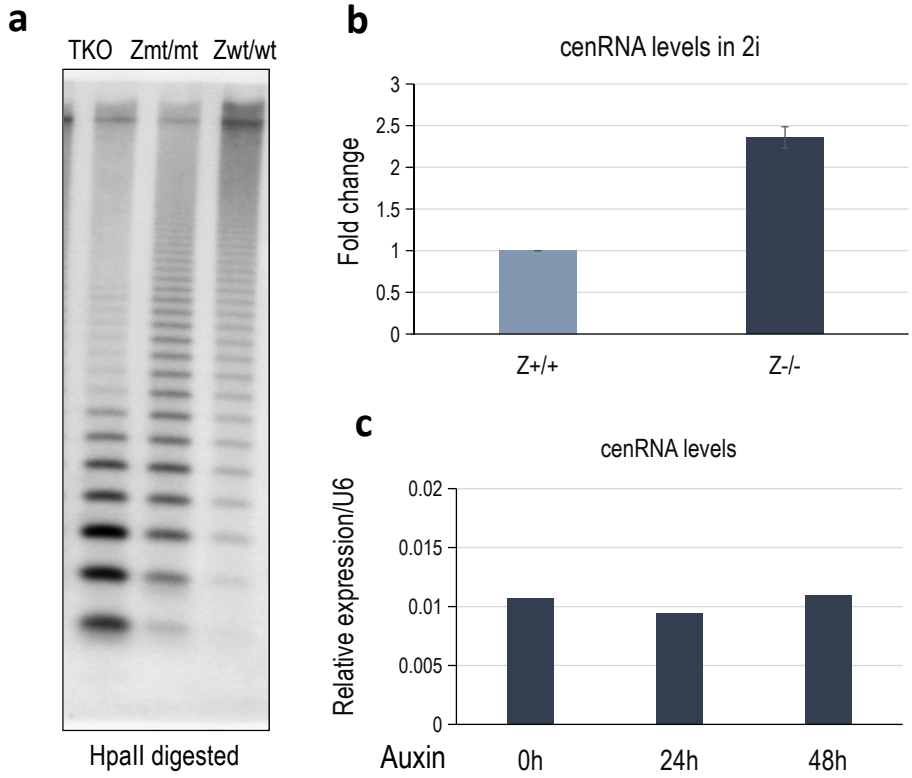


Figure 38. DNA hypomehylation at centromeres correlates with increase of cenRNA levels.
a. Analysis of methylation status in mouse embryonic stem cells (mESCs) cultured in 2i media. Southern blot analysis of DNA methylation at minor satellites in mESCs triple knock-out for DNMTs (TKO), or mESC with mutated Zbtb24 (Zmt/mt) or the wild-type counterpart (Zwt/wt). Genomic DNA was digested with HpaII enzyme and hybridized with centromere specific probes. n=3
b. Analysis of cenRNA levels in Zmt/mt and Zwt/wt mESCs. Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA in Zmt/mt and Zwt/wt mESCs cultured in 2i media. n=3
c. Analysis of cenRNA levels after auxin induced Zbtb24 depletion. Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA after Auxin addition at time 0h, 6h or 24h. n=2

4.3. Hypomethylation at centromeres in ICF1 murine immortalized fibroblasts facilitates increased centromeric transcription

To further explore the permissive hypomethylated context at centromeres found in the ICF syndrome and centromere transcription, I took advantage of another ICF murine cellular model. My host team created knock-in mouse models that reproduce mutations found in the DNMT3b gene of ICF1 patients (Velasco *et al.*, 2010), from which they derived ICF1 immortalized fibroblasts (ICF1 iMEFs), characterized by DNA hypomethlation at centromeric, but not pericentromeric, repeats. Thus, they also represent a good system to address the question of whether DNA hypomethylation at centromeres provides a permissive environment for increased centromere transcription, without the side effects from the hypomethylation of the nearby pericentromeres observed in ICF2/3/4 cells (Velasco *et al.*, 2018).

We first confirmed the hypomethylated status at centromeres of ICF1 iMEFs with Methylation Sensitive Restriction Enzyme assay followed by RT-qPCR (MSRE RT-qPCR), which indicates that ICF1 iMEFs present 60% less DNA methylation at centromeres compared to NIH-3T3 cells (**Figure 39.a**). Interestingly, despite the differences in DNA methylation status at the minor satellites, ICF1 iMEFs and NIH-3T3 share similar low levels of endogenous cenRNA levels (**Figure 39.b**), suggesting that low levels of DNA methylation are not sufficient to lead to high levels of cenRNAs. To assess then if DNA hypomethylation would allow higher levels of targeted transcriptional activation, we transiently transduced dCas9-VP64 and transfected Minor satellite-targeting gRNA (MiSgRNA) in both NIH-3T3 cells (methylated control cells) and ICF1 iMEFs. RT-qPCR result indicate that 24h of dCas9-VP64 targeting leads to increased transcription in both NIH-3T3 and ICF1 iMEFs (**Figure 39.c**). Importantly, ICF1 iMEFs show 30 times more cenRNAs compared to NIH-3T3 cells (**Figure 39.c**). In NIH-3T3 cells cenRNAs accumulate *in cis* after 24h of increased transcription, forming punctiform patterns seen by RNA-FISH using centromeric probes (**Figure 39.d left panels**). On the other hand, in the same conditions, cenRNAs in ICF1 iMEFs seem to diffuse from the locus and start spreading in the nucleus (**Figure 39.d right panels**). This result suggests that DNA hypomethylation, although not sufficient, creates a permissive state for increased centromere transcription in contexts where opportune transcription factors are present.

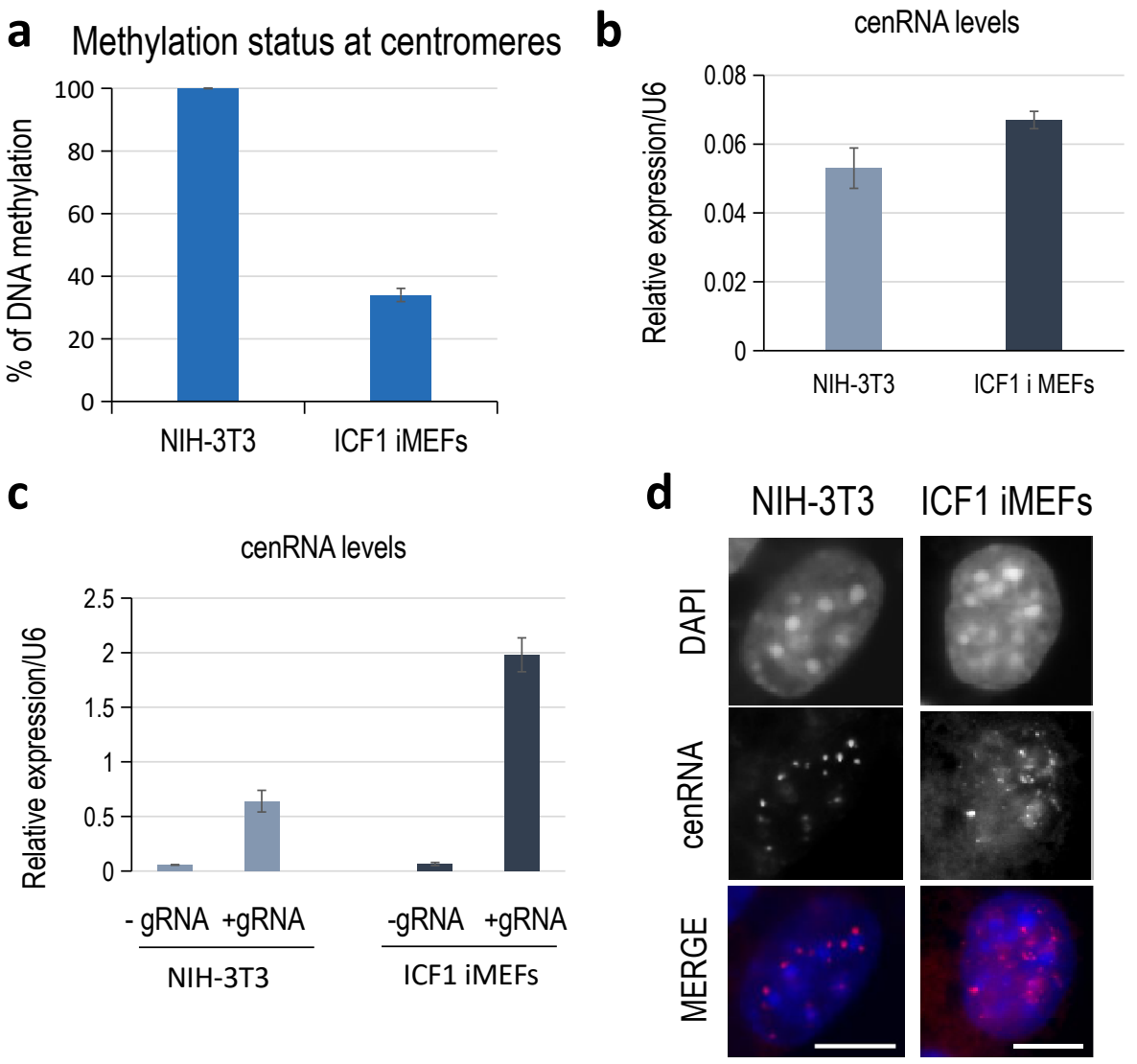


Figure 39. Hypomethylation at centromeres in ICF1 murine immortalized fibroblasts facilitates the induction of centromeric transcription with CRISPRa.

a. Analysis of methylation status of ICF1 murine immortalized fibroblasts (ICF1 iMEFs). Methylation sensitive restriction enzyme (MSRE) digestion followed by a quantitative PCR for murine minor satellite centromeric repeats. Genomic DNA from NIH-3T3 cells was used as a control for methylated centromeres and set at 100 %. n=3

b. Analysis of the levels of endogenous cenRNA levels in ICF1 iMEFs. Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA in ICF1 iMEFs compared to NIH-3T3 cells. n=3

c. Analysis of centromeric transcription activation by CRISPRa. Quantitative PCR with reverse transcription (RT–qPCR) showing relative expression of cenRNA in NIH-3T3 and ICF1 iMEFs upon activation of centromeric transcription by CRISPRa. Cells expressing dCas9-VP64 in presence or absence of MiSgRNA, were analyzed 24 h post transfection. n=3

d. Analysis of cenRNAs localization upon transrciptional activation. RNA-FISH for detection of cenRNA in NIH-3T3 and ICF1 iMEFs expressing dCas9-VP64 in presence of MiSgRNA, at 24h post transfection. Oligo probes were labeled with Alexa-Fluor 594 and nuclei were counterstained with DAPI. nuclei=100

DISCUSSION AND PERSPECTIVES

Centromeres have been long regarded as transcriptionally silent domains. However, research in the last decades has challenged this idea by showing that centromeric repeats can indeed be transcribed as cenRNAs (Biscotti 2015 for review). Although there is no sequence conservation of centromeric repeats between different species, (peri)cenRNAs have been reported in plants (Zhang *et al.*, 2005), yeast (Volpe *et al.*, 2002), insects (Pezer and Ugarković, 2008), and mammals (Bouzinba-Segard *et al.*, 2006a; Bury *et al.*, 2020b; Ferri *et al.*, 2009b; Jolly *et al.*, 2004; Metz *et al.*, 2004; Quénet and Dalal, 2014b; Vourc'h and Biamonti, 2011). Centromere transcription /or the resulting cenRNAs have since been implicated among key players for centromere identity, i.e Cenp-A deposition (Bobkov *et al.*, 2018; Zhu *et al.*, 2018a), and function in mitosis such as the timely targeting of protein complexes to centromeres (Bury *et al.*, 2020; Ferri *et al.*, 2009; McNulty *et al.*, 2017; Quénet and Dalal, 2014; Wong *et al.*, 2007), or for maintaining sister chromatid cohesion by participating in the formation of mitosis-specific centromeric R-loops (Kabeche *et al.*, 2018; Moran *et al.*, 2021).

CenRNAs are present at low levels in cycling cells, which peak at the onset of mitosis in murine and human cells (Ferri *et al.*, 2009; Wong *et al.*, 2007). In contrast, their aberrant accumulation was described in conditions of various cellular stress (Bouzinba-Segard *et al.*, 2006; Giunta *et al.*, 2021; Hédouin *et al.*, 2017), in diseases such as cancer (Eymery *et al.*, 2009; Ting *et al.*, 2011) or the ICF syndrome (Unoki *et al.*, 2020). Moreover, the accumulation of cenRNA in these conditions correlates with chromosomal instability (Bouzinba-Segard *et al.*, 2006; Eymery *et al.*, 2009; Ting *et al.*, 2011), and loss of centromere identity (Hédouin *et al.*, 2017) and function (Bouzinba-Segard *et al.*, 2006).

However, in these pathological contexts, the simultaneous presence of several pathological events at centromeres, such as i) unscheduled and/or increased transcription; ii) DNA hypomethylation that is often observed at satellite repeats; iii) centromeric or global DNA damage, presents a challenge in deciphering the causal link between cenRNA accumulation and chromosomal instability. Hence, during my thesis, I was interested in uncoupling these molecular events to order them and determine their relative contribution to pathological phenotypes, through altered centromere transcription or cenRNA accumulation.

Centromere transcription and cenRNAs: does strand matter?

Despite their essential function, cenRNAs have not been extensively characterized in terms of their transcriptional regulation. In many species, (peri)cenRNAs are transcribed from both strands in yeast (Volpe *et al.*, 2002), murine ESCs (Bulut-Karslioglu *et al.*, 2012; Kanellopoulou *et al.*, 2005; Velazquez Camacho *et al.*, 2017), murine embryos (Casanova *et al.*, 2013, 2013), murine somatic cells (Bouzinba-Segard *et al.*, 2006b; Bulut-Karslioglu *et al.*, 2012; Maison *et al.*, 2002) and HeLa cells (Bury *et al.*, 2020). However, there are not many reports in the literature about the potentially different functions operated by Forward and Reverse satellite RNAs. In some instances, for example during development, the strandedness of the pericenRNAs has important physiological roles. Thus, pericenRNAs are transcribed in both orientations, albeit with different kinetics, only Reverse pericenRNAs are necessary for the major reorganization that accompanies Zygote Genome Activation and developmental progression (Casanova *et al.*, 2013). However, little is known on whether both strands of cenRNAs are subjected to the same regulation and if the two RNA strands would have different functions (discussed further below).

As a first step to answer these questions, we aimed to characterize whether murine Forward and Reverse cenRNAs were present in equal amounts. Interestingly, both in physiological and genotoxic stress conditions, we observed that the Forward cenRNAs are more abundant in the NIH-3T3 cells (**Figure 29.c** and **Figure 29.e**). There are several hypotheses for the predominance of Forward compared to Reverse cenRNAs: i) difference at the level of transcription itself; ii) different stability of the RNAs; iii) technical bias in the detection by RT-qPCR and RNA-FISH. These different possibilities are not exclusive and might jointly impact the amount of cenRNAs observed in a given cellular context. In the following sections I will discuss each of these hypotheses.

I) The case for differences in transcription:

In stress conditions, cenRNA accumulation coincide with and requires transcriptional activation and may not result from a stabilization process of the RNA (Hédouin *et al.*, 2017). This implies that, at least in stress conditions, the discrepancy in levels of Forward and Reverse cenRNAs (**Figure 29.d**) could result from different transcriptional activity on the two centromeric DNA strands. However, many unknowns remain about the transcriptional regulation of centromeric repeats. Due to their nature of tandem repeats without canonical promoters, centromeres have been suggested to be transcribed by read-through from upstream

genes or transposable element promoters (Topp *et al.*, 2004). Thus, previous work in the team has shown that the transcription of murine minor satellites could also be initiated from the truncated LINE L1 (tL1) element located upstream of centromeres, on the side of the short arm of the chromosome (Thesis F. Ferri), which would yield Forward cenRNAs. Moreover, on some chromosomes, minor satellites are preceded by telocentric tandem repeats (TLC) which could also serve as initiating or regulatory sequences for Forward cenRNAs (Kalitsis *et al.*, 2006). Thus, determining the expression levels of tL1 or TLC RNAs in the cells with pathological overexpression of the satellite repeats would give us hints into this hypothesis. Moreover, it would be interesting to block transcription of tL1 or TLC using, for instance, dCas9-KRAB and then determine the levels of cenRNAs, in order to unravel their mode of transcriptional regulation.

Transcription of cenRNAs in one orientation or the other could be controlled by the binding of distinct transcription factors. Indeed, prediction tools for transcription factors (such as PROMO, developed by Messeguer *et al.*, 2002) revealed consensus binding sites for a set of transcription factors, distinct between the two strands. A possible strategy to uncover the transcription factors would be the Apex strategy (Gao *et al.*, 2018). This tool consists of the use of a dCas9 fused to the ascorbate peroxidase enzyme (Apex), which, upon a specific chemical reaction, biotinylates proteins in its vicinity. Subsequent mass spectrometry (MS) would then provide us with a list of possible factors. While this technique would not allow us to discriminate transcription factors that bind one strand or the other of Minor satellites, it would nevertheless provide us with a list of candidates, whose binding to centromeres in a given context could then be validated by ChIP and gel shift assays, as well as their activity as transcription factors at centromeres by GOF experiments. While I managed to set up the dCas9-Apex system, I unfortunately did not have the time to implement it to answer this biological question. Interestingly, some of the predicted transcription factors are tissue specific, such as the erythroid Gata-1 for the forward strand (that would produce Forward cenRNAs) or Pax-2, mostly present in the kidney, for the reverse strand (that would produce Reverse cenRNAs). Hence it would be interesting to analyze if the predominance of Forward compared to Reverse cenRNAs is present in cell types other than the immortalized fibroblasts that we used. Since cenRNAs appear to accumulate in the brain tissue of ICF1 mouse model (unpublished data from the team), we could take advantage of the dynamic mESCs system and differentiate them into the neuronal lineage. By then implementing the dCas9-Apex tool, we could determine the transcription factors of the two cenRNA populations in the brain.

Moreover, tandem repeats are the target of cell context-specific transcription factors that are activated in response to distinct stresses and signaling pathways. This is demonstrated by the transcription of Sat III RNAs in response to heat shock in humans. The Sat III transcription in this context is dependent on the transcription factor HSF1, which binds directly to these sequences (Jolly *et al.*, 2002, 2004). In contrast, Sat III transcription in response to osmotic shock is dependent on the transcription factor TonEPB (Valgardsdottir *et al.*, 2008). Moreover, p53 may directly or indirectly regulate cenRNA transcription in response to genotoxic stress (Hédouin *et al.*, 2017). A better understanding of the transcriptional regulation of centromeric repeats, such as whether both strands are transcribed at the same time, and from the same chromosome would shine light on their functions, both in physiological or in stress conditions.

The origin of oriented transcription of centromeric repeats, and whether it serves any role for centromere function, is currently unknown. Murine Forward cenRNAs are transcribed in the orientation going away from telomeres, towards pericentromeric Major satellite repeats. Given the structural organization of acrocentric mouse chromosomes, it is tempting to speculate that transcription of Reverse cenRNAs might be more controlled and inhibited in order to protect the telomeres from potential pervasive transcription, since unscheduled transcription of subtelomeric TERRA ncRNA leads to DNA damage at telomeres (Silva *et al.*, 2021). This could be verified by targeting increased transcription of one or the other strand of centromeric repeats and then assessing the localization of key telomeric proteins like Trf2 and expression of the subtelomeric TERRA RNA (Cusanelli and Chartrand, 2015).

Reverse cenRNAs could also be solely a byproduct of fortuitous transcription that occurs as a consequence of the open chromatin during DNA replication or transcription of Forward cenRNAs. In that case, it is possible that less repeating units are bound by RNApolII and transcribed, yielding less cenRNA. However, our results (Figure S1, Mihic *et al.*) go against this hypothesis since we do not observe accumulation of Reverse cenRNAs when targeting the transcription on the other strand.

Another possibility is that the two populations of cenRNAs are transcribed at the same levels but from a different number of chromosomes. Our RNA-FISH results on genotoxic stress-induced high levels of cenRNAs (**Figure 29.c**) support this idea. Indeed, in these conditions, Reverse cenRNAs are detected as less numerous punctiform patterns, found around chromocenters, compared to the Forward cenRNAs. While it seems that all or most murine centromeres respond to genotoxic stress by activated transcription of the Forward

cenRNAs (Hédouin *et al.*, 2017), the hypothesis that Reverse cenRNAs would be transcribed only from particular centromeres in response to genotoxic stress merits further exploration.

II) The case of differences in RNA stability:

It is widely accepted that mammalian cells transcribe almost their entire genomes, producing a huge non-coding output (reviewed in (Amaral *et al.*, 2008), which are often undetectable due to their high instability. Indeed, both coding and ncRNAs are known to undergo continuous degradation. To remove superfluous RNA, cells employ different RNA degradation systems, such as, for example, the nonsense-meadiated mRNA decay (NMD, Brogna *et al.*, 2016) or the exosome complex (van Balkom *et al.*, 2015). Indeed, cenRNAs can also be subjected to the degradation by the exosome complex. For example, in *S. pombe*, cenRNAs are undetectable due to their rapid turnover and degradation, but they accumulate in mutants where exonucleases activity is compromised (Choi *et al.*, 2011). Therefore, it is possible to imagine that the predominance of Forward compared to Reverse cenRNAs that we detected, might also be a result of different RNA stability of the two populations, possibly as a result of their preferential degradation.

To get hints into this hypothesis, we tested the stability of Forward and Reverse cenRNAs by inhibiting RNA Pol II-induced transcription and assessed their levels (**Figure 30.a**). In the kinetics of 8h, our results suggest that Forward cenRNAs are initially less stable than the Reverse, since their levels dramatically decrease in the first 2h of transcriptional inhibition. After that, their levels become equivalent to those of Reverse cenRNAs, and hence, the levels of both populations seem to decrease at the same rate. These results do not explain the predominance of the Forward cenRNAs, since they seem less stable, suggesting then that their higher levels result from preferential transcription. However, the different half-life of cenRNAs is interesting in itself, because it hints at the differential regulation of the two populations of cenRNAs. For example, the Forward and Reverse cenRNAs could be subjected to different post-transcriptional modifications, which are known to influence RNA stability (Boo and Kim, 2020; Nachtergaele and He, 2017). One possibility is the adenosine to inosine (A to I) modification mediated by the ADAR machinery (Keegan *et al.*, 2017). Of note, the ADAR recognizes dsRNAs or internal loops and bulges (Wang *et al.*, 2018). Since cenRNAs are transcribed in both directions, it is tempting to speculate that they could form dsRNAs that could be recognized by the ADAR machinery. Interestingly, the vast majority of “A to I” RNA editing sites in humans are in non-coding sequences, such as LINEs or the primate

specific Alu elements (Athanasiadis *et al.*, 2004). Bioinformatics screening for "A to I" editing sites have revealed that they are also abundant on ncRNAs from repeated sequences in the mouse, such as the B1 and B2 SINEs and the L1 LINE (Neeman *et al.*, 2006). However, the authors did not report results for ncRNAs emanating from satellite repeats. While ADAR complexes have not been detected to modify the neighboring pericenRNAs (Lu and Gilbert, 2008), it remains to be tested whether cenRNAs could be subjected to A to I modification.

Another intriguing RNA modification that could alter the stability of cenRNAs is N6-methyladenine (m6A). This modification is the most abundant form of post-transcriptional RNA modification in eukaryotes (Tang *et al.*, 2021). While there are no reports in the literature about m6A decorating cenRNAs, it is intriguing to note that murine Forward, but not the Reverse, cenRNAs harbour the consensus sequence RRACH (in which R represents A or G and H represents A, C or U), on which m6A modification usually occurs (Wang *et al.*, 2021). In addition, it has recently been shown that pericenRNA in mESCs are m6A modified (Duda *et al.*, 2021). These pericenRNAs are a structural component of mouse heterochromatin, where they stay associated with the DNA by forming DNA-RNA hybrids. Importantly, these hybrids are made possible and stabilized by m6A RNA methylation (Duda *et al.*, 2021). Different studies have confirmed the importance of m6A in regulating the stability or resolution of DNA-RNA hybrids (Abakir *et al.*, 2020; Kang *et al.*, 2021; Yang *et al.*, 2019). Since we and others have shown that cenRNA are prone to hybridize with their template DNA and form R-loops, it is tempting to speculate that cenRNAs, at least those involved in the DNA-RNA hybrid, would also be m6A methylated. This could be verified by performing an RNA pull-down of m6A-modified RNA using a specific antibody, and analyzing the levels of Forward and Reverse cenRNAs precipitated.

Stability of RNAs is also influenced by the binding of RNA binding proteins or incorporation into ribonucleoprotein complexes (Oroz and Laurents, 2019). One of the numerous examples is TERRA RNAs, which are bound by an array of hnRNPs, as well as telomere Shelterin factors TRF1 and TRF2, which are required for their stability and/or correct localization (Deng *et al.*, 2009; Lee *et al.*, 2018; López de Silanes *et al.*, 2010). cenRNAs have also been found to associate with a very large number of proteins, predominantly from the centromeric environment (Ferri *et al.*, 2009b; Quénet and Dalal, 2014b; Rošić *et al.*, 2014b). An RNA pull-down of cenRNAs, like those performed on the Forward cenRNA in normal (Ferri *et al.*, 2009) or genotoxic conditions (thesis of S. Hédouin) and subsequent MS had provide us with a list of candidates, whose impact on cenRNA stability need to be tested with LOF of the

proteins. Although it is possible that these factors have many targets and their KD might result in indirect effects. However, given the known role of cenRNAs as protein scaffolds, uncovering their new protein partners would give invaluable insights into their differential function.

III) The case of technical biases in the detection of cenRNAs:

As it is possible that we detect the prevalence of Forward cenRNAs due to technical reasons, such as higher efficiency of the Forward versus the Reverse primers, we designed two other sets of primers and observed the same trend in cenRNA population levels (**Figure 29.c**). In addition, we performed orthogonal approaches such as RNA-FISH with oligos that bind a different region compared to the primers used in qPCR, which corroborated our results (**Figure 29.e**). Even so, RNAs present internal secondary structures, and the oligos might have different efficiencies in hybridizing with the RNA substrate. Therefore, RNA-sequencing could represent an additional approach to confirm the strand predominance of cenRNAs.

However, I would like to point out another possibility that could result in the greater detection of Forward cenRNAs. The murine Minor satellite repeats are composed of 120 bp long repeat units. If, for example, the Forward cenRNAs consisted of several repeat units (we could imagine three repeat units, amounting to 360 nt in length) and Reverse cenRNAs consisted of only one repeat unit (120 nt long), the PCR primers would have more substrate in the case of the Forward cenRNAs, since they could bind to each repeat unit of one molecule of cenRNA, thus skewing the results. This could be easily verified by analyzing the size of cenRNAs with strand-specific probes on a Northern blot, which would tell us the size of cenRNAs as well as show the predominance of one versus the other cenRNA population.

Finally, despite the existence of fosmid clones containing a small number of minor satellite repeats (Kalitsis *et al.*, 2006), it cannot be excluded that the 2,500 repeats that constitute centromeres are only arranged in a "head-to-tail" manner. Presence of “head-to-head” repeats on one strand could result in the detection of Forward and Reverse cenRNAs, but both populations would originate from the transcription of the same DNA strand. Moreover, (centromeric) tandem repeats are prone to recombination events (Jaco *et al.*, 2008) which could result in the creation of “head-to-head” centromeric repeats. Thus long-read sequencing (Chintalaphani *et al.*, 2021) of both the DNA and cenRNAs would elucidate if the Forward and Reverse cenRNAs might originate from the same strand.

In sum, our results point to the predominance of one population of cenRNAs. Nonetheless, many unknowns remain in the characterization of cenRNAs. Why is it then interesting to study the strandedness and stability of the different cenRNAs? The Forward and Reverse cenRNAs could have differential roles, which I will discuss in the next section.

Forward and Reverse cenRNAs: to each their own role?

With increasing interest in the field for ncRNAs of tandem repeat origin, cenRNAs have been assigned many roles, from the correct localization of constitutive components of the centromere such as Cenp-A and Cenp-C (Du *et al.*, 2010; Quénet and Dalal, 2014b; Wong *et al.*, 2007b), the timely recruitment and activity of the CPC (Blower, 2016b; Ferri *et al.*, 2009b; Ideue *et al.*, 2014), and association with DNA repair factors (Kabeche *et al.*, 2018; Racca *et al.*, 2021; Zhu *et al.*, 2018b). It is intriguing to think that different cenRNAs populations, including double-stranded cenRNAs (dsRNAs), or the single-stranded Forward or Reverse ones, may have different roles.

The presence of transcripts from both strands could favor the formation of dsRNAs which could be further processed into siRNAs. The yeast S. pombe is the first organism in which it has been demonstrated that dsRNAs from pericentromeric regions are involved in the formation of heterochromatin through the action of the RNA interference (RNAi) machinery (Volpe *et al.*, 2002). Indeed, deletion of any component of the RNAi machinery, such as RNase III Dicer, RNA-dependent RNA polymerase (RdRp) or RITS components, leads to a loss of heterochromatin. In the mESCs, the knockout of Dicer also greatly affected silencing of centromeric repeats, leading to the accumulation of long (peri)cenRNAs, between 1.4 and 2.4 kb in size while the expression of small dsRNAs was markedly reduced (Kanellopoulou *et al.*, 2005). Strikingly, a recent study has shown that the Dicer KO mESCs have strong chromosome segregation defects and increased transcription of (peri)centromeric repeats, which triggers the interferon response (Gutbrod *et al.*, 2022). Therefore, these studies point to an important role of the RNAi in controlling the transcription and function of satellite repeats.

In contrast to the classical role of miRNAs and siRNAs, which target the degradation of their target RNA, the binding of the RITS complex to pericentromeres, through the association of pericentromeric siRNAs with nascent pericenRNAs, allows the recruitment of chromatin modifying factors such as histone deacetylases and methyltransferases for H3K9, leading to propagation of heterochromatin on this domain (Rea *et al.*, 2000). In addition, the presence of

the RdRp complex maintains siRNA levels and contributes to the establishment of a self-reinforcing loop, allowing propagation of heterochromatin (Volpe *et al.*, 2002). Of note, the RdRp enzyme is absent in mammals. However, dsRNAs can be produced also by bidirectional transcription of the repeats or by folding the RNA molecule onto itself to form secondary stem-loop structures, therefore it is tempting to speculate that centromeric dsRNAs could participate, through the RNAi machinery, in the transcriptional regulation of centromeres.

Although it seems paradoxical that transcription of (peri)centromeric repeats is required for their repression, analysis of its activation during the cell cycle revealed that this transcription is not constant. Indeed, pericentromeric transcription occurs mainly during replication of pericentromeric regions in late S phase and is dependent on RNA polymerase II (Kato *et al.*, 2005; Kloc *et al.*, 2008; Maison *et al.*, 2002). The resulting pericenRNAs, in turn, ensure the restoration of heterochromatin and the maintenance of the newly replicated chromatin in a repressed state via the RNA interference machinery (Chen *et al.*, 2008). Moreover, RNAi is also instrumental in controlling the expression of other ncRNAs, such as LINE1, demonstrated by their accumulation in Dicer KO mESCs (Bodak *et al.*, 2017). However, small double stranded cenRNAs of 20-30 nt in size were not observed in murine somatic cells (Bouzinba-Segard *et al.*, 2006). Hence, if the RNAi pathway is involved in the regulation of cenRNA levels in murine somatic cells remains an exciting question to explore.

In addition to potential roles for double-stranded cenRNAs, single-stranded Forward or Reverse cenRNAs could have differential roles between the two transcript populations. This idea is supported by our cenRNA KD experiments. Indeed, almost complete loss of Forward cenRNAs had a drastic impact on cell viability (**Figure 33.e ii**), suggesting that these transcripts play an important role for the cells, most probably in centromere homeostasis. Conversely, loss of Reverse cenRNAs, while with comparable efficiency, had a lesser impact on the cells. Assessing the protein partners, as mentioned in the previous section, would give us hints into the functional roles of these two cenRNA populations. ncRNAs play many functions within the cell by controlling, for example, the recruitment, stability, or activity of partner protein complexes (Kung *et al.*, 2013). The same is true for cenRNAs, to which several functions and partners have been associated, mainly within centromeres (Scott, 2013). Another indication that goes in the direction of different protein partners for the Forward or Reverse cenRNAs is their different predicted secondary structures [predicted with RNAfold (http://rna.tbi.univie.ac.at/)]. Both of the cenRNAs, 120 nt in size, are predicted to have a

stem-loop structure, but the Forward cenRNAs might display multibranch and internal loops and hairpins, whereas the Reverse seems to harbor an exterior loop. Not surprisingly, the predicted structure varies with the addition of several repeat units. Since RNA binding proteins are highly enriched in intrinsically disordered regions, they mainly recognize RNA structure and not a specific sequence (Zagrovic *et al.*, 2018). This then potentially adds to the complexity of the interactome of cenRNAs, which then could have different functional roles defined by their respective protein binding partners. Thus, a whole new field of study on the role of these RNAs is open.

Besides functioning as protein scaffolds, cenRNAs are also implicated in R-loops formation (our data; (Kabeche *et al.*, 2018; Moran *et al.*, 2021). Our data shows that transcription of the Forward cenRNAs leads to R-loop formation. However, whether that coincides with the strand implicated in the physiological R-loops detected in mitosis (Kabeche *et al.*, 2018; Moran *et al.*, 2021), remains to be explored. It is possible that one orientation of cenRNAs, and not the other, participates in this function. CRISPRi mediated centromere silencing on one or the other strand, and just at the onset of mitosis, would enable us to assess which cenRNAs are involved in R-loop establishment.

The discrepancy in the Forward and Reverse cenRNA levels could have a role also in development or differentiation. Indeed, the level of transcription of (peri)centromeric satellite DNA has been proposed to vary between developmental stages and also tissue types in mammals and insects (Maison *et al.*, 2010; Pezer and Ugarković, 2008). In particular, pericentromere transcription and pericenRNAs are thought to be required for chromocenter formation during early development in the mouse (Casanova *et al.*, 2013; Probst *et al.*, 2010). During mouse early development, pericenRNAs start being detected at the 2-cells stage and are required for the major reorganization of the nucleus that occurs at this stage (Probst *et al.*, 2010). This reorganization correlates with transcription of pericenRNAs that is strand-specific, where Forward transcripts are first to accumulate, followed then by the Reverse ones. Importantly, KD of either Forward or Reverse pericenRNAs at the 2 cell stage, precludes chromocenter formation and leads to developmental arrest. Moreover, after the initial burst of pericenRNA levels at the 2-cells stage, both populations rapidly decrease at the 4-cells stage. These data suggest that pericenRNAs play a role in the establishment of chromocenters and that their strand-specific expression must be finely regulated during development.

On the other hand, in somatic cells, murine (peri)cenRNAs accumulate with terminal differentiation (Bouzinba-Segard *et al.*, 2006; Terranova *et al.*, 2005), with the Forward

transcripts being more abundant in the first days of differentiation (Bouzinba-Segard *et al.*, 2006). However, whether there are different cenRNA populations in terms of orientation, that could emerge during development, and whether they have a role or are just a fortuitous consequence of global changes in genome expression (Bickmore and Zaret, 2010), is an exciting venue to explore.

Besides programmed physiological processes in which cenRNA strand-specificity could have a role, this could also apply to stress conditions. Work from the team and others has shown that (peri)cenRNAs are important actors of the stress response (Hédouin *et al.*, 2017; Jolly *et al.*, 2004; Valgardsdottir *et al.*, 2008). In addition, transcription of human satellite repeats in stress is highly asymmetric, with the G-rich strand being preferentially transcribed. The strand specific transcriptional activation of satellite repeats in response to stress is also confirmed by our RT-qPCR and RNA-FISH results, showing the predominance of the Forward transcripts upon Etoposide treatment. While the functional role of satellite RNAs in response to stress conditions has not been fully elucidated, some hypotheses have been put forward (Eymery *et al.*, 2009). In particular, they play a role in the control of splicing in response to stress *via* the transient sequestration of certain splicing factors. Indeed, several splicing factors such as SRp30c and SF2/ASF proteins co-localize with Sat III RNAs at the 9q12 locus in response to heat shock (Chiodi *et al.*, 2004; Denegri *et al.*, 2001; Metz *et al.*, 2004). Moreover, they could participate in the restoration of the correct chromatin structure during the stress recovery phase. Therefore, it would be interesting to investigate further the importance of Forward or Reverse cenRNAs, or strand-specific transcription, in the response to stress to uncover further their roles in stress recovery.

Although (peri)cenRNAs seem to have a protective role in stressed cells, there is increasing evidence to suggest that, in unstressed cells, the levels of (peri)cenRNAs need to be tightly regulated and that disturbing this balance could have detrimental consequences for the organism.

cenRNAs in disease: passengers or drivers?

Cancers are multifactorial and heterogeneous diseases (Dagogo-Jack and Shaw, 2018) but exhibit a number of invariant features (Senga and Grose, 2021). In particular, gene expression profiles are disrupted and characterized by illegitimate expression of oncogenes and repression of tumor suppressor genes. Furthermore, ncRNAs are also beginning to emerge as

new players in tumor development or progression by participating in oncogenic or tumor suppressor pathways (Cheetham *et al.*, 2013). Interestingly, in the last decade, transcriptomic analyses have highlighted that ncRNAs are highly dysregulated in cancer cells and can even contribute to the metastatic cascade (Liu *et al.*, 2021; Slack and Chinnaiyan, 2019). In particular, (peri)cenRNAs have been observed to accumulate heavily in a variety of human and murine cancers (Eymery *et al.*, 2009; Ting *et al.*, 2011). Furthermore, these same transcripts strongly accumulate in BRCA1-deficient breast cancer cells (Zhu *et al.*, 2011). Likewise, depletion of BRCA1 has been linked to elevated cenRNA levels and impaired chromosome segregation (Di Paolo *et al.*, 2014). Despite the abundant correlations between cenRNA accumulation and disease state, no clear causality has been put forward. This then opens the question: is cenRNA accumulation a cause or just a consequence of disease? A First hint into the causal link came from previous work from the team, which has shown that cenRNA forced accumulation is toxic for the cells and leads to loss of centromere function by aberrant mislocalization of proteins normally associated with centromeres (Bouzinba-Segard *et al.*, 2006). Similarly to Aurora B sequestration from centromeres in the aforementioned study, satellite RNA accumulation can lead to BRCA1 delocalization (Zhu *et al.* 2018). Moreover, ectopic expression or injection of (peri)cenRNAs in cultured human or murine cells leads to mitotic errors and aneuploidy (Bouzinba-Segard *et al.*, 2006; Zhu *et al.*, 2018b), which are characteristics of cancer cells. These data, therefore, suggest that cenRNA accumulation represents a hazard to genome integrity and can potentially represent a driving force in oncogenesis.

To explore this idea, we performed GOF of murine cenRNAs in primary MEFs. Our results show that both Forward and Reverse cenRNA overexpression (**Figure 32.b**) leads to a modest increase in $p16^{Ink4a}$ expression levels in these cells (**Figure 32.c**), suggesting that their accumulation could be necessary and/or sufficient for the onset of senescence, although the kinetics might not be sufficient to observe late markers of senescence. Senescence following ectopic accumulation of cenRNAs is reminiscent of oncogene-induced senescence (OIS), whereby ectopic expression of an oncogene like Ras is sufficient to promote entry of primary MEFs into senescence (Dimauro and David, 2010). Tumorigenesis initiation requires at least two "hits" or mutations in oncogenes or tumor suppressor genes (Knudson, 2001). This is evidenced by overexpression of Ras that in primary fibroblasts leads to OIS (Serrano *et al.*, 1997). However, when Ras is combined with another oncogene, such as the viral Early E1A protein, cells escape senescence and immortalize (Deng *et al.*, 2005; Serrano *et al.*, 1997). It

would then be interesting to explore if cenRNA accumulation could represent one of these "hits" that, in combination with Ras overexpression, would allow cell immortalization.

Given the accumulation of (peri)cenRNAs in cancer (Ting *et al.*, 2011a), another question that arises is: what makes them accumulate in pathological conditions? It is possible that cenRNA accumulation represents a consequence of oncogenesis. For example, high cenRNA levels could result from their activation during the senescence process. Then, if an oncogenic event provokes senescent cells to re-enter the cell cycle, such as mutations in the p53 and pRB genes (Dirac and Bernards, 2003; Sage *et al*., 2003), the newly immortalized cells could maintain high cenRNA levels. Another non-exclusive possibility is that centromeric regions could be directly activated by transcription factors, such as AP-1 family factors, including known oncogenes like cJun and vFos (Lopez-Bergami *et al.*, 2010), whose consensus binding sites are present on the forward strand of the Minor satellites. Finally, loss of the tumor suppressor p53 protein, a hallmark of many cancer cells, could allow for aberrant transcription of repeated sequences. It has been shown that many repeated elements of the genome, such as SINEs or LINEs transposable elements, possess p53 binding sites, and that p53 is required for their transcriptional repression to limit their retrotransposition (Hagan and Rudin, 2007; Harris *et al.*, 2009; Leonova *et al.*, 2013). Notably, p53 participates, in combination with DNA methylation, in preventing the activation of murine SINEs and Major satellite repeats (Leonova *et al.*, 2013). Conversely, in the absence of p53, and with DNA demethylation by the demethylating agent 5-AZA, Major satellites and SINEs are highly expressed. Although the authors did not assess the levels of cenRNAs, this transcriptional dysregulation could also apply to centromeric repeats, since previous work from the team has shown that the p53 protein weakly binds murine Minor satellite repeats (Hédouin *et al.*, 2017).

While most of the correlative studies linking (peri)cenRNAs to disease states focus mainly on their increased levels in these conditions, it still remains to be determined if the accumulation of these RNAs is a result of their stabilization or increased transcription.

On the whole, the establishment of the causal links between accumulation of (peri)cenRNAs and cancer state merits further attention. While it is possible that aberrant transcription of these regions represents a driving force of oncogenesis, by contributing to chromosomal instability characteristic of cancers, it could also be a consequence of transcriptomic perturbations in cancer cells, leading to non-specific de-repression of repeated elements. Indeed, an invariant feature of cancer cells is perturbation of epigenetic states. This can

include DNA hypomethylation at satellite repeats that often correlates with the aberrant overexpression of the underlying repeats.

Satellite DNA hypomethylation and centromere transcription: beyond the binary of the “ON/OFF” model

One of the most characteristic features of repetitive elements is that they are methylated in somatic cells (Papin *et al.*, 2017; Pappalardo and Barra, 2021). A prime example are transposable elements, where DNA methylation represents one of the mechanisms that contributes to repress their mobility (Jursch *et al.*, 2013). Likewise, at satellite repeats, DNA methylation has been proposed to keep these loci in a silent state (Jones, 2012). Indeed, centromeres are enriched in DNA methylation in somatic cells, which is tightly maintained throughout cell cycle progression (Wong *et al.*, 2006). However, this epigenetic mark does not completely preclude low levels of transcription at centromeres, since they are naturally transcribed, albeit at low levels (Bury *et al.*, 2021; Ferri *et al.*, 2009; Quenet and Dalal, 2014). Moreover, we have confirmed that DNA methylation does not impede targeted increased transcription (**Figure 37**), suggesting that DNA hypomethylation is not a prerequisite for centromere transcription. Yet, whether physiological transcription at centromeres occurs only on “pockets of DNA hypomethylation” found on certain repeat units (Logsdon *et al.*, 2021), as is the case of HAC centromeres (Chan and Wong, 2012), awaits formal testing.

In pathological conditions such as cancer, DNA hypomethylation positively correlates with high levels of (peri)cenRNAs (Eymery *et al.*, 2009; Ting *et al.*, 2011). Yet, the causal relationship between DNA hypomethylation and satellite repeat transcription is not clear. In particular, human cells deficient for DNMT1 and DNMT3B are hypomethylated on the Sat III repeats but do not show accumulation of these transcripts compared to WT cells (Eymery *et al.*, 2009). Similarly, hypomethylation of pericentromeric regions in TKO (Dnmt1, Dnmt3A and Dnmt3B deficient) mESCs is not sufficient to induce increased transcription of these regions (Lehnertz *et al.*, 2003; Martens *et al.*, 2005). In contrast, in other cellular contexts, a causal link between methylation and transcription seems to exist. For example, treatment of cells with the DNA demethylating agent 5-azacytidine (5-AZA) leads to increased levels of pericenRNAs in human cells and cenRNAs in murine cells (Bouzinba-Segard *et al.*, 2006), although it’s important to note that these experiments are difficult to interpret as 5-AZA is

known to cause DNA damage, a stress that is linked to the induction of satellite repeats expression (Hédouin *et al.*, 2017).

To better understand the causal relationship between DNA hypomethylation and centromere transcription, we took advantage of the ICF context. While ICF cells feature DNA hypomethylation on a few hundreds of genomic regions (Velasco, Grillo *et al.*, 2018), it is most prominent at tandem repeats in the (peri)centromeric regions (Ehrlich *et al.*, 2006; Velasco, Grillo *et al.*, 2018; Grillo *et al.*, 2022 MS in preparation). ICF1 iMEFs feature similar levels of endogenous cenRNAs as NIH-3T3 cells, which are methylated at centromeres (**Figure 39.a** and **Figure 39.b**). This observation, along with the above cited studies, suggests that DNA hypomethylation is not sufficient to drive increased centromere transcription. However, we have shown that the hypomethylated state of centromeres in ICF1 iMEFs allows for a higher increase in cenRNA levels after targeted transcription (**Figure 39.c**). Furthermore, in mice with a hypomorphic mutation for the *Dnmt3B* gene, whose catalytic activity is greatly reduced resulting in hypomethylation on Minor satellites, only some organs show an increase in cenRNA levels, such as the brain, heart and skeletal muscle (unpublished data, thesis of F. Ferri). This suggests that there may be tissue-specific transcription factors that would explain their tissue-restricted transcription. Possible candidates to test would be factors of the GATA family, for which consensus sites have been found in Minor satellite repeats, and some of those members are specifically expressed in heart, muscle or hematopoietic cells.

On the other hand, the transcription factor ZBTB24, which is also causal to the ICF2, seems to act as a transcriptional repressor at centromeres in the human Lymphoblastoid cell line (LCL) (Grillo *et al.*, 2022, MS in preparation). Loss of ZBTB24 in LCLs from ICF2 patients correlates with an increase of cenRNAs from specific centromeric HORs, that in WT cells are occupied by ZBTB24, thus strengthening the hypothesis that ZBTB24 negatively regulates centromere transcription in somatic cells. However, loss of ZBTB24 is also accompanied by loss of DNA methylation, thus impeding the establishment of clear causal links. To test this, we took advantage of the mESCs cultured in 2i which naturally display low levels of DNA methylation (Habibi *et al.*, 2013; Stadler *et al.*, 2011). In mESCs, rapid loss of the Zbtb24 protein, mediated by the degron system, did not result in an upregulation of cenRNA levels, suggesting that, during corresponding stage of the early development, Zbtb24 might not directly regulate transcription of Minor satellites (**Figure 38.c**). Importantly, in this context, loss of Zbtb24 did not result in additional hypomethylation at minor satellites. Interestingly,

minor satellites of Zbtb24$^{mt/mt}$ mESCs derived from mouse embryos are more hypomethylated than their WT counterparts (**Figure 38.a**) and correlate with slightly higher levels of cenRNAs (**Figure 38.b**). Together, these data suggest that while DNA hypomethylation may favor centromere transcription in specific cellular contexts, it is not necessarily sufficient to promote transcriptional de-repression of satellite sequences. There are several hypotheses on how DNA hypomethylation could, directly or indirectly, regulate centromere transcription.

As stated previously, DNA methylation, or lack thereof, could influence the binding of proteins at centromeres. For instance, both human and mouse centromeres harbour two methylatable CpG di-nucleotides in the Cenp-B box (Tanaka *et al.*, 2005). Cenp-B, the only centromeric protein with sequence-specific binding activity, has been shown in vitro to be sterically hindered by DNA methylation at Cenp-B boxes (Tanaka *et al.*, 2005), although there is controversy since inhibition of DNA methylation leads to Cenp-B spreading over demethylated repeats (Mitchell *et al.*, 1996). Since it was known that DNA (hypo)methylation can influence the binding of transcription factors (Héberlé and Bardet, 2019), this could also apply to centromeric repeats, which could be bound by certain transcription factors in a DNA methylation-dependent manner.

DNA hypomethylation could influence cenRNA transcription by affecting proper centromeric condensation and topology. For example, loss of DNA methylation at pericentromeric repeats has been associated with DNA decondensation (Suzuki and Bird, 2008). Therefore, it is possible to envisage that hypomethylation-induced decondensation of satellite repeats could be more prone to pervasive and unspecific transcription.

Other than the potential role in controlling transcription at centromeres, DNA methylation is important to suppress the recombinogenic potential of DNA repeats (Jones, 2012). Not surprisingly, an important consequence of DNA hypomethylation at (peri)centromeric repeats is the increase in recombination frequency at these loci (Jaco *et al.*, 2008). These crossovers have been linked to chromosome breakage and loss following impaired cohesion of sister chromatids, which cause their premature separation and mis-segregation (Talbert and Henikoff, 2010). DNA hypomethylation is also accompanied by a reduction in the length of Minor satellite repeats, measured by quantitative FISH (Q-FISH) (Jaco *et al.*, 2008). However, the authors did not assess whether cenRNA transcription is altered in these cases.

Surprisingly, DNA hypomethylation is also intimately linked with centromere identity, since Cenp-A was reported to occupy hypomethylated repeats in human and plants (Logsdon *et al.*, 2021; Zhang *et al.*, 2008). This has been confirmed by a recent study, that has shown that

CENP-A localizes on evolutionarily youngest satellite repeats, which coincide with a region of DNA hypomethylation (Altemose *et al.*, 2022). This opens up interesting questions about the relationship between DNA (hypo)methylation and centromere integrity that is determined by Cenp-A presence. Given that a proposed role for centromere transcription is its requirement for Cenp-A deposition, it is tempting to speculate that centromeres are more readily transcribed from hypomethylated repeats, thus determining the arrays that are occupied by Cenp-A. In addition, it would be interesting to assess Cenp-A localization in cells that feature high levels of centromeric hypomethylation, such as physiological contexts of ESCs or pathological contexts of cancer cells.

Further studies are required to fully decipher the relationship between DNA hypomethylation and centromere homeostasis, especially in case of its transcription. A possible way to address this question would be to directly demethylate centromeres using dCas9 fused to the Ten-eleven translocation methylcytosine dioxygenase 1 (TET1) that is involved in active DNA demethylation, and assess cenRNA levels. However, given that tissue-specific transcription factors might be required to trigger transcription, a complementary approach would be to employ a DNA methyltransferase fused to dCas9, in cells that feature demethylated centromeres and high levels of cenRNAs, such as the $p53^{-/-}Dnmt1^{-/-}$ immortalized fibroblasts (**Figure 33.b**).

On the whole, DNA methylation is of crucial importance for the maintenance of centromere identity and stability, by influencing the binding of several key centromeric proteins and preserving this locus from excessive rearrangements. In addition, DNA methylation might have a role in controlling the correct expression of these repeats.

Increased centromere transcription versus cenRNA accumulation: which one is the “lesser of two evils”?

Literature is punctuated with reports of (peri)cenRNAs that accumulate in disease (Eymery *et al.*, 2009; Ting *et al.*, 2011), senescence (Enukashvily *et al.*, 2007; Gaubatz and Cutler, 1990; Hédouin *et al.*, 2017), and various types of cellular stress (Hédouin *et al.*, 2017; Jolly *et al.*, 2004; Pecinka *et al.*, 2010; Pezer and Ugarković, 2008; Tittel-Elmer *et al.*, 2010). The causal link between cenRNA accumulation and cell fate is less clear. In the context of genotoxic stress, cenRNA accumulation results directly from the activated transcription of these repeats, which leads to centromere disassembly and ultimately to proliferative arrest (Hédouin *et al.*,

2017). However, in these conditions, the global DNA damage hinders the establishment of clear causal links between centromere transcription, cenRNAs, and cellular outcomes. Hence, to assess the immediate contribution of the act of centromere transcription, we targeted the dCas9-VP64 activator system (Konermann *et al.*, 2015) to Minor satellite repeats, in the absence of initial DNA damage. First, we have investigated the impact on centromere identity by assessing the localization of its determinant, Cenp-A. We have shown that, at 24h of increased transcription, which results in a 10-fold increase of cenRNAs, centromeric architecture remains intact and Cenp-A exhibits the habitual punctiform pattern. In contrast, 48h of increased transcription, yielding a 30-fold increase of cenRNAs, results in stretched patterns of Cenp-A that still remain present at centromeres. This is in contrast to highly increased centromere transcription at human artificial chromosomes (HACs) that lead to 150-fold increase in cenRNA levels and resulted in the loss of Cenp-A (Bergmann *et al.*, 2012). Similarly, genotoxic-induced centromere transcription, that resulted in 1500-fold increase in cenRNA levels, led to loss of centromere identity by the eviction of Cenp-A from its default location (Hédouin *et al.*, 2017). This suggests that in our study we reach lower levels of transcription compared to the two aforementioned studies. However, cenRNA levels that were reached with our approach are comparable to those found in tumors derived from mouse models of pancreatic, colon, and lung cancers, where satellite RNA levels are increased up to 40 fold compared to normal tissue (Ting *et al.*, 2011).

While the increased transcription that we obtained in our experiments seems compatible with centromere identity since Cenp-A is maintained, it is not without consequences. We have shown that increased centromere transcription in itself can be a precursor of DNA damage at Minor satellite repeats. Importantly, the resulting DNA damage seems to be independent of the aberrant centromeric R-loops, since targeting RNaseH1 (the enzyme responsible for the resolution of DNA-RNA hybrids) concomitantly with the activation of centromere transcription, does not preclude the appearance of γH2AX at centromeres. The DNA damage could then result from the collision of transcription and replication machineries (Reviewd in García-Muse and Aguilera, 2019). Another possibility is that cenRNAs tether DNA repair factors away from centromeres (Racca *et al.*, 2021; Zhu *et al.*, 2018b), which will be discussed further below. Finally, we have shown that sustained centromere transcription leads to decondensation of centromeric domains, which is the likely culprit for the loss of centromere function and mitotic errors we observed. Coupled with DNA damage, the DNA decondensation could fragilize this locus, by leading to illicit recombination at repeated

sequences (Giunta *et al.*, 2021; Yilmaz *et al.*, 2021). Notably, in the timeframe of our experiments, we were not able to detect the changes in the copy number variation of minor satellite repeats (data not shown), however, these might arise later. Moreover, the DNA decondensation could affect the organization of centromeric chromatin, known to play a crucial role in maintaining proper tension between sister chromatids and kinetochores during mitosis (Lawrimore and Bloom, 2019). Aberrant centromeric R-loops could also negatively impact centromeric chromatin, perhaps counter-intuitively given the previous reports of R-loops being essential for proper mitotic function of centromeres (Kabeche *et al.*, 2018; Moran *et al.*, 2021). However, targeting RNaseH1 at centromeres with aberrantly increased transcription restores normal mitoses by removing R-loops, directly implicating the latter in the centromeric loss of function and mitotic errors. It is possible that there are substantial differences between physiological mitotic R-loops and those we observe upon increased centromere transcription in interphase. First of all, we would need to validate that, with our system, R-loops still form at mitotic centromeres. Moreover, physiological R-loops could form on a smaller number of repeat units compared to the instances where we trigger increased transcription. It would be then worth to perform a DRIP assay on cells arrested in mitosis, to determine if levels of R-loops at centromeres are comparable between the two conditions. Dissecting the molecular mechanisms of the R-loops contribution to centromere (dys)function with a particular focus on the cell cycle aspect represents an interesting topic for future investigation. Exploiting further the cell-cycle dependent forms of dCas9-VP64 (used in the manuscript), we could answer how R-loops contribute to centromere (dys)function in the different phases of the cell cycle.

Alongside R-loop formation, non-DNA-bound cenRNAs are also the output of centromere transcription. This raises the question: what is the contribution of centromere transcription, of the resulting R-loop or/and of the soluble cenRNAs, on centromere dysfunction? To answer this, we first performed a GOF with expression plasmids for ectopic cenRNAs. The strong accumulation of cenRNAs did not affect centromere integrity, since we did not observe appearance of DNA damage at this locus. This is in contrast with GOF of cenRNAs in human cells, which leads to specific enrichment of γH2AX at centromeres (Zhu *et al.*, 2011), possibly as a result of sequestering the repair factor BRCA1 away from this locus (Zhu *et al.*, 2018b). However, the GOF resulted in highly increased levels of cenRNAs (more than 200 fold compared to control cells, **Figure 35.a**) which seem to accumulate in the entire nucleus (**Figure 35.b**), we cannot exclude indirect effects. Therefore, to better understand the

contribution of cenRNAs on centromere loss of function, we targeted engineered cenRNA molecules directly to centromeres, in absence of increased transcription. The engineered cenRNAs seem to mimic the local accumulation of cenRNAs that we observed in the case of targeted transcription (**Figure 36.b**). Notably, in physiological untreated conditions, we were not able to detect cenRNAs by RNA-FISH (**Figure 35.b**), which suggests that the engineered cenRNAs accumulate to higher levels than the endogenous cenRNAs. Still, the local accumulation of cenRNAs did not seem to perturb centromeric architecture, since Cenp-A displayed its normal punctiform pattern (**Figure 36.c**), suggesting that centromere disorganization is a consequence of the act of increased transcription. Importantly, we observed the appearance of the DNA damage marker γH2AX at centromeres, both with the targeting of the control lucRNA or cenRNA. Moreover, cells with lucRNA targeted to centromeres were more prone to DNA damage at this locus compared to cells with targeted cenRNAs. This observation is interesting in itself because it suggests that lucRNA might perturb local environment at centromeres by recruiting non-specific protein partners. On the other hand, cenRNAs could bind proteins that normally occupy centromeric chromatin, but possibly at the wrong phase of the cell cycle. Given that levels of physiological cenRNAs peak at the onset of mitosis, and are maintained at relatively low levels outside this phase (Ferri *et al.*, 2009), accumulation of cenRNAs at centromeres outside of G2/M could then interfere with centromeric processes such as deposition of Cenp-A in G1, or even the transcription itself, by possibly hybridizing with the complementary DNA in the open double helix.

When cenRNAs accumulate above a certain threshold, they could function like sponges and sequester proteins from their site of action. Indeed, high levels of cenRNAs have been linked to BRCA1 delocalization away from centromeres, which exposes this locus to unrepaired genotoxic insults (Di Paolo *et al.*, 2014; Zhu *et al.*, 2011). Therefore, this represents another possibility of how untimely or too high accumulation of cenRNAs could jeopardize centromere integrity. It is worth noting that engineered cenRNAs could still exhibit considerable differences compared to the endogenous ones, such as potentially lacking the RNA modifications, length and/or structure as physiological cenRNAs. We therefore cannot exclude that the engineered cenRNAs also tether non-specific proteins to centromeres.

By acting as protein scaffolds, RNAs can provide an increase in the local concentration of proteins. Indeed, RNAs are particularly important for the formation of biomolecular condensates, which are self-organized membraneless subcompartments within the nucleus on

the microscopic scale of 0.1–1 μm (Belmont, 2022; Misteli, 2001, 2020; Spector, 2003). Such subcompartments are associated with a variety of activities and direct genome functions like transcription, DNA replication, recombination, and repair. Although the local enrichment of protein and/or RNA in such condensates is crucial for their function and formation (review in Rippe, 2022), the condensates are destabilized above a certain threshold of RNA due to the increased electrostatic repulsion. Therefore RNAs are key components that can modulate both condensate assembly and dissolution *in vitro* (Banerjee *et al.*, 2017). Hence it is possible that the soluble cenRNAs disrupt proper centromere function and homeostasis by altering the physical characteristics of the locus.

CONCLUSION

Centromeres are key determinants of genome integrity, since they are essential for the correct segregation of genetic material. The importance of centromere homeostasis is highlighted by their dysregulation in disease states, which includes loss of centromere identity and its perturbed architecture, hypomethylation of these repeats, and their aberrant transcription. We have shown that increased centromere transcription and accumulation of the resulting cenRNAs cause centromere dysfunction. However, we propose that they have distinct consequences. Increased centromere transcription disrupts centromere function through the formation of R-loops, which in turn perturbs proper centromere architecture, leading to mitotic errors. Whether these R-loops pose a threat to centromere function because they are formed in an unscheduled manner or at too high levels, remains to be determined. Soluble cenRNAs, on the other hand, could act by altering the physical properties of the centromeric environment, leading to a change in the protein composition that is not compatible with centromere function. Determining the proteins that might be recruited at - or tethered away from - centromeres will help us to understand the molecular mechanisms of how cenRNAs disrupt centromere function. Lastly, we have shown that DNA hypomethylation at centromeres does not invariantly lead to increased transcription, but it creates a favorable environment for it.

While our view of the functional links between the different pathological events that impact centromeres is still incomplete, further investigation promises to yield interesting discoveries.

MATERIAL AND METHODS
(Unpublished results)

Culture of primary cells and immortalized cell lines

Primary murine fibroblasts (MEFs) and $p53^{-/-}DNMT1^{-/-}$ immortalized fibroblasts were cultured in Dulbecco's Modified Eagle Medium/Glutamax supplemented with 20% fetal bovine serum (FBS, Dutscher), 100 U/ml penicillin and 100 µg/ml streptomycin (ThermoFisher) at 37°C and 5% CO_2.

Plasmid transfection

Murine NIH-3T3 and primary MEFs were seeded 24 hours prior to transfection in 12-well or 6-well plates in order to be at ~70% confluence at the time of transfection. The reagent used for transfection (jetPEI®, Polyplus transfection) is a polyethylenimine that compacts DNA into positively charged complexes capable of interacting and fusing with the plasma membrane. Cells are incubated with a solution containing up to 3µg of vector diluted in 150mM NaCl and 6µL of jetPEI® also diluted in 150mM NaCl. The solution containing the vectors and jetPEI® is pre-incubated for 30 min at RT before being placed in the cell culture medium. The transfected cells are incubated at 37°C and 5% CO2 for 24h or 48h before being lysed or fixed for immunofluorescence, DNA- or RNA-FISH.

In vitro transfection of Locked Nucleic Acid (LNA) gapmers

The design and synthesis of LNA gapmers were provided from Exiqon. The knockdown of murine cenRNA was performed on $p53^{-/-}DNMT1^{-/-}$ cells, using the transfection reagent RNAiMAX (Invitrogen). The transfection was performed on 10^5 cells at the day of plating (d0) while the cells were in suspension, using 40 pmol of LNA_R1, LNA_R2 or LNA_F, or scrambled as control (list of LNA-gapmers is listed in Table xx), mixed with 7.5µL of RNAiMAX and 250µL of OptiMEM media. The medium was changed 12 hours after the transfection. The morphology of the cells was assessed under the microscope, 5x magnification. Cells were collected for RNA extraction 24 hours after the transfection (d1).

Quantification of nucleic acids

Genomic DNA and RNA concentrations were measured using the NanoDrop 1000 Overview (Thermo Scientific). Nucleic acid purity was estimated by measuring the ratio of absorbance A260/A280 to measure the presence of proteins and A260/A230 to evaluate the presence of organic solvent (phenol, chloroform, isopropanol) and salts; A260/A280 should be ~1.8 for

pure DNA and >2.0 for pure RNA, and A260/A230 should be in the range of ~1.8-2.0 for both DNA and RNA.

RNA extraction and .DNase treatment of RNA

Total RNA from cell lines was isolated using TRI Reagent® (Sigma) according to manufacturer's instructions. After RNA extraction, a DNase treatment was preformed to eliminate possible contaminations by genomic DNA, using the TURBO DNA-free kit (Invitrogen). Briefly, 3µg of RNA were treated in a total volume of 20µL containing 0.1 volume of 10X DNase Buffer, 1µL of Turbo DNase I and 0.5µL of RNase Out (Life Technologies). After incubation at 37°C for 30 minutes, the enzyme was inactivated with 0.1 volumes of DNase Inactivation Reagent (containing EDTA). After centrifugation at 11,000 x g for 1.5 minutes, the upper phase containing RNA was collected.

Reverse transcription

Reverse transcription was carried out using 1µg DNA-free RNA and 60µM random hexamers (Roche), The mix was incubated at 68°C for 10 minutes in order to break secondary structures, and then incubated with 60µM of random hexamers (Roche), 0.5µM of dNTP (Life Technologies) 20U of RNase inhibitor (New England Biolabs) and 200U of ProtoScript II Reverse Transcriptase (New England Biolabs), for 60 minutes at 42°C, followed by 5 minutes at 95°C to inactivate the enzyme. Control samples were incubated in the same mix but instead of the reverse transcriptase, the same volume of nuclease-free water was added. The cDNAs synthesized were stored at – 20°C or used directly for the PCR reaction. For strand-specific RT, the same protocol was used, but the RT reaction was performed using either the forward or the reverse primer targeting Minor satellites, and both Forward and Reverse U6 primers in the same reaction.

Quantitative Real Time PCR (qPCR)

Amplification was performed in 96-well plates using 25ng of cDNA and 200nM of each primer (final concentration), in a final volume of 15µl containing 1X of LightCycler ® 480 SYBR Green Master Mix (Roche). The PCR conditions were as follows: denaturation (15 seconds at 12 95°C), hybridization (30 seconds at 60°C) and synthesis (20 seconds at 72°C) for a total of 45 cycles. The data was normalized with levels of U6 or GAPDH as an invariant

RNA, and the results were analysed using the comparative Ct method (2– ΔCt for relative expression and 2– ΔΔCt for fold change). Primers are listed in table xx

Ct= cycle threshold

ΔCt = Ct target gene – Ct U6;

ΔΔCt = ΔCt target sample – ΔCt control sample

Genomic DNA extraction

Genomic DNA was extracted and purified from cell lines, using the Monarch® Genomic DNA Purification Kit (New England Biolabs) according to manufacturer's instructions.

Southern blot

Genomic DNA from MEFs (500 ng) was digested with 20 units of HpaII enzyme (New England Biolabs) for 16 h to analyse the DNA methylation patterns of centromeric minor satellite and pericentromeric major satellites repeats respectively. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+ membranes (GE Healthcare) in 20x SSC. After ultraviolet crosslink, the membranes were pre-hybridized in 6x SSC, 5x Denhardt and 0.1% SDS and then hybridized with ^{32}P-labelled probe.

Methylation-sensitive restriction enzyme-coupled qPCR assay (MSRE-qPCR)

Genomic DNA (200ng) was digested at 37°C for 4 hours with 10 units of the methylationsensitive enzyme HpyCH4IV (New England Biolabs). Digestion with HindIII (New England Biolabs), which does not have restriction sites in our regions of interest, served to normalize the data. The endonucleases were subsequently inactivated by incubation at 80°C for 20 minutes and qPCR was performed as described above, using 1ng of digested DNA. Data was analyzed by the comparative CT (ΔCT) method according to the formula:

Methylation (%) = 2(ΔCT) ×100 where ΔCT = CTsample (HpyCH4IV digest) - CTsample (HindIII digest).

LNA gapmer	Target
LNA_F	AACTCACTCATCTAAT
LNA_R1	CGTTGGAAACGGGATT
LNA_R2	ACCACACTGTAGAACA
LNA_scr	CACGTCTATACGC

Table 2. List of LNA gapmers used for cenRNA knockdown

Primer	Orientation	Sequence 5'-3'
U6	Forward	CTCGCTTCGGCAGCACA
U6	Reverse	AACGCTTCACGAATTTGCGT
Gapdh	Forward	CTTCACCACCATGGAGAAGCC
Gapdh	Reverse	GGCATGGACTGTGGTCATGAG
Minor satellites	Forward	GAACATATTAGATGAGTGAGTTAC
Minor satellites	Reverse	GTTCTACAAATCCCGTTTCCAAC
p16Ink4a	Forward	ACGGTGCAGATTCGAACTGC
p16Ink4a	Reverse	TACACAAAGACCACCCAGCG
Zbtb24	Forward	GCTGGAGCACATGAGCCTG
Zbtb24	Reverse	TTGGCACTGGAGTCAGAGAAG

Table 3. List of primers used for qPCR reactions

Probe target	Sequence 5'-3'
Minor satellite	ACATTCGTTGGAAACGGGATTTGTAGAACAGTGTATATCAATGAGTTACAATGAGAAACAT
Major satellite	CACGTCCTACAGTGGACATTTCTAAATTTTCCACCTTTTTCAGTT

Table 4. List of probes used for Southern blot

REFERENCES

Abakir, A., Giles, T.C., Cristini, A., Foster, J.M., Dai, N., Starczak, M., Rubio-Roldan, A., Li, M., Eleftheriou, M., Crutchley, J., *et al.* (2020). N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. Nat Genet *52*, 48–55. https://doi.org/10.1038/s41588-019-0549-x.

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

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*. https://doi.org/10.7554/eLife.05005.

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol Cell *46*, 115–124. https://doi.org/10.1016/j.molcel.2012.04.009.

Aguilera, A., and Gómez-González, B. (2017). DNA-RNA hybrids: the risks of DNA breakage during transcription. Nat Struct Mol Biol *24*, 439–443. https://doi.org/10.1038/nsmb.3395.

Aguilo, F., Zhou, M.-M., and Walsh, M.J. (2011). Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer Res *71*, 5365–5369. https://doi.org/10.1158/0008-5472.CAN-10-4379.

Alfeghaly, C., Aigueperse, C., Maenner, S., and Behm-Ansmant, I. (2021). Non-Coding RNA Silencing in Mammalian Cells by Antisense LNA GapmeRs Transfection. Methods Mol Biol *2300*, 31–37. https://doi.org/10.1007/978-1-0716-1386-3_4.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet *9*, 923–937. https://doi.org/10.1038/nrg2466.

Altemose, N., Logsdon, G.A., Bzikadze, A.V., Sidhwani, P., Langley, S.A., Caldas, G.V., Hoyt, S.J., Uralsky, L., Ryabov, F.D., Shew, C.J., *et al.* (2022). Complete genomic and epigenetic maps of human centromeres. Science *376*, eabl4178. https://doi.org/10.1126/science.abl4178.

Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008). The eukaryotic genome as an RNA machine. Science *319*, 1787–1789. https://doi.org/10.1126/science.1155472.

Amatori, S., Tavolaro, S., Gambardella, S., and Fanelli, M. (2021). The dark side of histones: genomic organization and role of oncohistones in cancer. Clin Epigenetics *13*, 71. https://doi.org/10.1186/s13148-021-01057-x.

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. Nat Genet *51*, 217–223. https://doi.org/10.1038/s41588-018-0306-6.

Aronson, A.I., and McCARTHY, B.J. (1961). Studies of E. coli ribosomal RNA and its degradation products. Biophys J *1*, 215–226. https://doi.org/10.1016/s0006-3495(61)86885-9.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol *2*, e391. https://doi.org/10.1371/journal.pbio.0020391.

Au, S.W.N., Leng, X., Harper, J.W., and Barford, D. (2002). Implications for the ubiquitination reaction of the anaphase-promoting complex from the crystal structure of the Doc1/Apc10 subunit. J Mol Biol *316*, 955–968. https://doi.org/10.1006/jmbi.2002.5399.

Aunan, J.R., Watson, M.M., Hagland, H.R., and Søreide, K. (2016). Molecular and biological hallmarks of ageing. Br J Surg *103*, e29-46. https://doi.org/10.1002/bjs.10053.

Azzaz, A.M., Vitalini, M.W., Thomas, A.S., Price, J.P., Blacketer, M.J., Cryderman, D.E., Zirbel, L.N., Woodcock, C.L., Elcock, A.H., Wallrath, L.L., *et al.* (2014). Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. J Biol Chem *289*, 6850–6861. https://doi.org/10.1074/jbc.M113.512137.

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., and Otonkoski, T. (2015). Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. Stem Cell Reports *5*, 448–459. https://doi.org/10.1016/j.stemcr.2015.08.001.

van Balkom, B.W.M., Eisele, A.S., Pegtel, D.M., Bervoets, S., and Verhaar, M.C. (2015). Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting. J Extracell Vesicles *4*, 26760. https://doi.org/10.3402/jev.v4.26760.

Banerjee, P.R., Milin, A.N., Moosa, M.M., Onuchic, P.L., and Deniz, A.A. (2017). Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. Angew Chem Int Ed Engl *56*, 11354–11359. https://doi.org/10.1002/anie.201703191.

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. https://doi.org/10.1038/35065138.

Barik, S. (2004). Development of gene-specific double-stranded RNA drugs. Ann Med *36*, 540–551. https://doi.org/10.1080/07853890410018817.

Barisic, M., and Geley, S. (2011). Spindly switch controls anaphase: spindly and RZZ functions in chromosome attachment and mitotic checkpoint control. Cell Cycle *10*, 449–456. https://doi.org/10.4161/cc.10.3.14759.

Barra, V., and Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat Commun *9*, 4340. https://doi.org/10.1038/s41467-018-06545-y.

Bartolomé, C., Maside, X., and Charlesworth, B. (2002). On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol *19*, 926–937. https://doi.org/10.1093/oxfordjournals.molbev.a004150.

Beacon, T.H., Delcuve, G.P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A.J., and Davie, J.R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. Clin Epigenetics *13*, 138. https://doi.org/10.1186/s13148-021-01126-1.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet *32*, 29–41. https://doi.org/10.1016/j.tig.2015.11.001.

Beerli, R.R., and Barbas, C.F. (2002). Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol *20*, 135–141. https://doi.org/10.1038/nbt0202-135.

Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc Natl Acad Sci U S A *95*, 14628–14633. https://doi.org/10.1073/pnas.95.25.14628.

Beerli, R.R., Schopfer, U., Dreier, B., and Barbas, C.F. (2000). Chemically regulated zinc finger transcription factors. J Biol Chem *275*, 32617–32627. https://doi.org/10.1074/jbc.M005108200.

Belmont, A.S. (2022). Nuclear Compartments: An Incomplete Primer to Nuclear Compartments, Bodies, and Genome Organization Relative to Nuclear Architecture. Cold Spring Harb Perspect Biol *14*, a041268. https://doi.org/10.1101/cshperspect.a041268.

Belotserkovskii, B.P., Neil, A.J., Saleh, S.S., Shin, J.H.S., Mirkin, S.M., and Hanawalt, P.C. (2013). Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks. Nucleic Acids Res *41*, 1817–1828. https://doi.org/10.1093/nar/gks1333.

Ben-Porath, I., and Cedar, H. (2001). Epigenetic crosstalk. Mol Cell *8*, 933–935. https://doi.org/10.1016/s1097-2765(01)00399-9.

Berger, S.L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. Genes Dev *23*, 781–783. https://doi.org/10.1101/gad.1787609.

Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., and Earnshaw, W.C. (2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J *30*, 328–340. https://doi.org/10.1038/emboj.2010.329.

Bergmann, J.H., Martins, N.M.C., Larionov, V., Masumoto, H., and Earnshaw, W.C. (2012). HACking the centromere chromatin code: insights from human artificial chromosomes. Chromosome Res *20*, 505–519. https://doi.org/10.1007/s10577-012-9293-0.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol Cell *2*, 437–445. https://doi.org/10.1016/s1097-2765(00)80143-4.

Bickmore, W.A., and Zaret, K.S. (2010). Altered states: how gene expression is changed during differentiation. Curr Opin Genet Dev *20*, 467–469. https://doi.org/10.1016/j.gde.2010.08.003.

Black, B.E., and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Curr Opin Cell Biol *20*, 91–100. https://doi.org/10.1016/j.ceb.2007.11.007.

Black, E.M., and Giunta, S. (2018). Repetitive Fragile Sites: Centromere Satellite DNA As a Source of Genome Instability in Human Diseases. Genes (Basel) *9*, E615. https://doi.org/10.3390/genes9120615.

Blanco-Betancourt, C.E., Moncla, A., Milili, M., Jiang, Y.L., Viegas-Péquignot, E.M., Roquelaure, B., Thuret, I., and Schiff, C. (2004). Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. Blood *103*, 2683–2690. https://doi.org/10.1182/blood-2003-08-2632.

Bloom, K.S. (2014). Centromeric heterochromatin: the primordial segregation machine. Annu Rev Genet *48*, 457–484. https://doi.org/10.1146/annurev-genet-120213-092033.

Bloom, K., and Costanzo, V. (2017). Centromere Structure and Function. Prog Mol Subcell Biol *56*, 515–539. https://doi.org/10.1007/978-3-319-58592-5_21.

Blower, M.D. (2016a). Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep *15*, 1624–1633. https://doi.org/10.1016/j.celrep.2016.04.054.

Blower, M.D. (2016b). Centromeric Transcription Regulates Aurora-B Localization and Activation. Cell Rep *15*, 1624–1633. https://doi.org/10.1016/j.celrep.2016.04.054.

Bobkov, G.O.M., Gilbert, N., and Heun, P. (2018a). Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol *217*, 1957–1972. https://doi.org/10.1083/jcb.201611087.

Bobkov, G.O.M., Gilbert, N., and Heun, P. (2018b). Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. J Cell Biol *217*, 1957–1972. https://doi.org/10.1083/jcb.201611087.

Bodak, M., Cirera-Salinas, D., Yu, J., Ngondo, R.P., and Ciaudo, C. (2017). Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells. FEBS Open Bio *7*, 204–220. https://doi.org/10.1002/2211-5463.12174.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods *89*, 123–130. https://doi.org/10.1016/0022-1759(86)90040-2.

Boo, S.H., and Kim, Y.K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. Exp Mol Med *52*, 400–408. https://doi.org/10.1038/s12276-020-0407-z.

Borg, M., Jiang, D., and Berger, F. (2021). Histone variants take center stage in shaping the epigenome. Curr Opin Plant Biol *61*, 101991. https://doi.org/10.1016/j.pbi.2020.101991.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genetics *42*, 1093–1100. https://doi.org/10.1038/ng.708.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96–99. https://doi.org/10.1038/nature02886.

Bouzinba-Segard, H., Guais, A., and Francastel, C. (2006a). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proc Natl Acad Sci U S A *103*, 8709–8714. https://doi.org/10.1073/pnas.0508006103.

Bouzinba-Segard, H., Guais, A., and Francastel, C. (2006b). Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. Proceedings of the National Academy of Sciences *103*, 8709–8714. .

Braun, S.M.G., Kirkland, J.G., Chory, E.J., Husmann, D., Calarco, J.P., and Crabtree, G.R. (2017). Rapid and reversible epigenome editing by endogenous chromatin regulators. Nat Commun *8*, 560. https://doi.org/10.1038/s41467-017-00644-y.

Brocken, D.J.W., Tark-Dame, M., and Dame, R.T. (2018). dCas9: A Versatile Tool for Epigenome Editing. Curr Issues Mol Biol *26*, 15–32. https://doi.org/10.21775/cimb.026.015.

Brogna, S., McLeod, T., and Petric, M. (2016). The Meaning of NMD: Translate or Perish. Trends Genet *32*, 395–407. https://doi.org/10.1016/j.tig.2016.04.007.

Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-cycle-dependent structural transitions in the human CENP-A nucleosome in vivo. Cell *150*, 317–326. https://doi.org/10.1016/j.cell.2012.05.035.

Bulut-Karslioglu, A., Perrera, V., Scaranaro, M., de la Rosa-Velazquez, I.A., van de Nobelen, S., Shukeir, N., Popow, J., Gerle, B., Opravil, S., Pagani, M., *et al.* (2012). A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol *19*, 1023–1030. https://doi.org/10.1038/nsmb.2382.

Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., and Cheeseman, I.M. (2020a). Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. Elife *9*, e59770. https://doi.org/10.7554/eLife.59770.

Bury, L., Moodie, B., Ly, J., McKay, L.S., Miga, K.H., and Cheeseman, I.M. (2020b). Alpha-satellite RNA transcripts are repressed by centromere-nucleolus associations. Elife *9*. https://doi.org/10.7554/eLife.59770.

Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol *18*, 299–314. https://doi.org/10.1038/nrm.2016.166.

Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. J Bacteriol *194*, 5728–5738. https://doi.org/10.1128/JB.01184-12.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., *et al.* (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524–15529. https://doi.org/10.1073/pnas.242606799.

Carmona, J.J., and Michan, S. (2016). Biology of Healthy Aging and Longevity. Rev Invest Clin *68*, 7–16. .

Carpenter, N.J., Filipovich, A., Blaese, R.M., Carey, T.L., and Berkel, A.I. (1988). Variable immunodeficiency with abnormal condensation of the heterochromatin of chromosomes 1, 9, and 16. J Pediatr *112*, 757–760. https://doi.org/10.1016/s0022-3476(88)80698-x.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655. https://doi.org/10.1016/j.cell.2009.01.035.

Casanova, M., Pasternak, M., El Marjou, F., Le Baccon, P., Probst, A.V., and Almouzni, G. (2013). Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. Cell Rep *4*, 1156–1167. https://doi.org/10.1016/j.celrep.2013.08.015.

Casola, C., Hucks, D., and Feschotte, C. (2008). Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol Biol Evol *25*, 29–41. https://doi.org/10.1093/molbev/msm221.

Castellano-Pozo, M., Santos-Pereira, J.M., Rondón, A.G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., and Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell *52*, 583–590. https://doi.org/10.1016/j.molcel.2013.10.006.

Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. Nat Commun *9*, 4346. https://doi.org/10.1038/s41467-018-06677-1.

Chan, F.L., and Wong, L.H. (2012a). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res *40*, 11178–11188. https://doi.org/10.1093/nar/gks921.

Chan, F.L., and Wong, L.H. (2012b). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res *40*, 11178–11188. https://doi.org/10.1093/nar/gks921.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012a). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984. https://doi.org/10.1073/pnas.1108705109.

Chan, F.L., Marshall, O.J., Saffery, R., Kim, B.W., Earle, E., Choo, K.H.A., and Wong, L.H. (2012b). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A *109*, 1979–1984. https://doi.org/10.1073/pnas.1108705109.

Chang, E.Y.-C., and Stirling, P.C. (2017). Replication Fork Protection Factors Controlling R-Loop Bypass and Suppression. Genes (Basel) *8*, E33. https://doi.org/10.3390/genes8010033.

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N.M.C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B.J., Wu, C.-T., *et al.* (2019). Islands of retroelements are major components of *Drosophila* centromeres. PLoS Biol *17*, e3000241. https://doi.org/10.1371/journal.pbio.3000241.

Chardon, F., Japaridze, A., Witt, H., Velikovsky, L., Chakraborty, C., Wilhelm, T., Dumont, M., Yang, W., Kikuti, C., Gangnard, S., *et al.* (2022). CENP-B-mediated DNA loops regulate activity and stability of human centromeres. Mol Cell *82*, 1751-1767.e8. https://doi.org/10.1016/j.molcel.2022.02.032.

Charlet, J., Duymich, C.E., Lay, F.D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., and Jones, P.A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. Mol Cell *62*, 422–431. https://doi.org/10.1016/j.molcel.2016.03.033.

Charpentier, M., Khedher, A.H.Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S., *et al.* (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. Nat Commun *9*, 1133. https://doi.org/10.1038/s41467-018-03475-7.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen *58*, 235–263. https://doi.org/10.1002/em.22087.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature *430*, 992–998. https://doi.org/10.1038/nature02821.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. Nat Methods *12*, 326–328. https://doi.org/10.1038/nmeth.3312.

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet *32*, 828–838. https://doi.org/10.1016/j.tig.2016.10.002.

Cheetham, S.W., Gruhl, F., Mattick, J.S., and Dinger, M.E. (2013). Long noncoding RNAs and the genetics of cancer. Br J Cancer *108*, 2419–2425. https://doi.org/10.1038/bjc.2013.233.

Chiang, H.-C., Zhang, X., Li, J., Zhao, X., Chen, J., Wang, H.T.-H., Jatoi, I., Brenner, A., Hu, Y., and Li, R. (2019). BRCA1-associated R-loop affects transcription and differentiation in breast luminal epithelial cells. Nucleic Acids Res *47*, 5086–5099. https://doi.org/10.1093/nar/gkz262.

Chintalaphani, S.R., Pineda, S.S., Deveson, I.W., and Kumar, K.R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. Acta Neuropathol Commun *9*, 98. https://doi.org/10.1186/s40478-021-01201-x.

Chiodi, I., Corioni, M., Giordano, M., Valgardsdottir, R., Ghigna, C., Cobianchi, F., Xu, R.-M., Riva, S., and Biamonti, G. (2004). RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res *32*, 4127–4136. https://doi.org/10.1093/nar/gkh759.

Choi, E.S., Strålfors, 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. https://doi.org/10.1074/jbc.M111.228510.

Choi, E.S., Cheon, Y., Kang, K., and Lee, D. (2017). The Ino80 complex mediates epigenetic centromere propagation via active removal of histone H3. Nat Commun *8*, 529. https://doi.org/10.1038/s41467-017-00704-3.

Chong, Z.-S., Ohnishi, S., Yusa, K., and Wright, G.J. (2018). Pooled extracellular receptor-ligand interaction screening using CRISPR activation. Genome Biol *19*, 205. https://doi.org/10.1186/s13059-018-1581-3.

Choo, K.H.A. (2001). Domain Organization at the Centromere and Neocentromere. Developmental Cell *1*, 165–177. https://doi.org/10.1016/S1534-5807(01)00028-4.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol Cell *40*, 179–204. https://doi.org/10.1016/j.molcel.2010.09.019.

Ciferri, C., Pasqualato, S., Screpanti, E., Varetti, G., Santaguida, S., Dos Reis, G., Maiolica, A., Polka, J., De Luca, J.G., De Wulf, P., *et al.* (2008). Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. Cell *133*, 427–439. https://doi.org/10.1016/j.cell.2008.03.020.

Cimmino, L., and Aifantis, I. (2017). Alternative roles for oxidized mCs and TETs. Curr Opin Genet Dev *42*, 1–7. https://doi.org/10.1016/j.gde.2016.11.003.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223.

Clark, M.B., and Mattick, J.S. (2011). Long noncoding RNAs in cell biology. Semin. Cell Dev. Biol. *22*, 366–376. https://doi.org/10.1016/j.semcdb.2011.01.001.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell *33*, 717–726. https://doi.org/10.1016/j.molcel.2009.01.026.

Cleveland, D.W., Mao, Y., and Sullivan, K.F. (2003). Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell *112*, 407–421. https://doi.org/10.1016/s0092-8674(03)00115-6.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823. https://doi.org/10.1126/science.1231143.

Cooper, J.L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. Mol Biol Evol *21*, 1712–1718. https://doi.org/10.1093/molbev/msh179.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat Struct Mol Biol *11*, 1037–1043. https://doi.org/10.1038/nsmb851.

Costantino, L., and Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol Cell *71*, 487-497.e3. https://doi.org/10.1016/j.molcel.2018.06.037.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599–601. https://doi.org/10.1038/31275.

Cowan, R.K., Hoen, D.R., Schoen, D.J., and Bureau, T.E. (2005). MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol Biol Evol *22*, 2084–2089. https://doi.org/10.1093/molbev/msi202.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138–163. .

Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep *23*, 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025.

Crosio, C., Fimia, G.M., Loury, R., Kimura, M., Okano, Y., Zhou, H., Sen, S., Allis, C.D., and Sassone-Corsi, P. (2002). Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. Mol Cell Biol *22*, 874–885. https://doi.org/10.1128/MCB.22.3.874-885.2002.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019a). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019b). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., *et al.* (2006). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. Nat Genet *38*, 910–916. https://doi.org/10.1038/ng1842.

Cusanelli, E., and Chartrand, P. (2015). Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. Front Genet *6*, 143. https://doi.org/10.3389/fgene.2015.00143.

Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol *15*, 81–94. https://doi.org/10.1038/nrclinonc.2017.166.

Davalos, V., Moutinho, C., Villanueva, A., Boque, R., Silva, P., Carneiro, F., and Esteller, M. (2012). Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. Oncogene *31*, 2062–2074. https://doi.org/10.1038/onc.2011.383.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development *25*, 1010–1022. https://doi.org/10.1101/gad.2037511.

de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., *et al.* (2011). Mutations in ZBTB24 Are Associated with Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome Type 2. The American Journal of Human Genetics *88*, 796–804. https://doi.org/10.1016/j.ajhg.2011.04.018.

Deininger, P.L., and Batzer, M.A. (2002). Mammalian retroelements. Genome Res *12*, 1455–1465. https://doi.org/10.1101/gr.282402.

Denegri, M., Chiodi, I., Corioni, M., Cobianchi, F., Riva, S., and Biamonti, G. (2001). Stress-induced nuclear bodies are sites of accumulation of pre-mRNA processing factors. Mol Biol Cell *12*, 3502–3514. https://doi.org/10.1091/mbc.12.11.3502.

Deng, Q., Li, Y., Tedesco, D., Liao, R., Fuhrmann, G., and Sun, P. (2005). The ability of E1A to rescue ras-induced premature senescence and confer transformation relies on inactivation of both p300/CBP and Rb family proteins. Cancer Res *65*, 8298–8307. https://doi.org/10.1158/0008-5472.CAN-05-0054.

Deng, Z., Norseen, J., Wiedmer, A., Riethman, H., and Lieberman, P.M. (2009). TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. Mol Cell *35*, 403–413. https://doi.org/10.1016/j.molcel.2009.06.025.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & Development *15*, 2940–2944. .

Di Paolo, A., Racca, C., Calsou, P., and Larminat, F. (2014). Loss of BRCA1 impairs centromeric cohesion and triggers chromosomal instability. FASEB J *28*, 5250–5261. https://doi.org/10.1096/fj.14-250266.

Dimauro, T., and David, G. (2010). Ras-induced senescence and its physiological relevance in cancer. Curr Cancer Drug Targets *10*, 869–876. https://doi.org/10.2174/156800910793357998.

Dinant, C., and Luijsterburg, M.S. (2009). The emerging role of HP1 in the DNA damage response. Mol Cell Biol *29*, 6335–6340. https://doi.org/10.1128/MCB.01048-09.

Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput Biol *4*, e1000176. https://doi.org/10.1371/journal.pcbi.1000176.

Dirac, A.M.G., and Bernards, R. (2003). Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. J Biol Chem *278*, 11731–11734. https://doi.org/10.1074/jbc.C300023200.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol *32*, 1262–1267. https://doi.org/10.1038/nbt.3026.

Drouet, J., Delteil, C., Lefrançois, J., Concannon, P., Salles, B., and Calsou, P. (2005). DNA-dependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks. J Biol Chem *280*, 7060–7069. https://doi.org/10.1074/jbc.M410746200.

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. https://doi.org/10.1371/journal.pgen.1000835.

Du, Z., Zhang, K., and Xie, W. (2022). Epigenetic Reprogramming in Early Animal Development. Cold Spring Harb Perspect Biol *14*, a039677. https://doi.org/10.1101/cshperspect.a039677.

Duda, K.J., Ching, R.W., Jerabek, L., Shukeir, N., Erikson, G., Engist, B., Onishi-Seebacher, M., Perrera, V., Richter, F., Mittler, G., *et al.* (2021). m6A RNA methylation of major satellite repeat transcripts facilitates chromatin association and RNA:DNA hybrid formation in mouse heterochromatin. Nucleic Acids Res *49*, 5568–5587. https://doi.org/10.1093/nar/gkab364.

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. https://doi.org/10.1007/BF00328227.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat Rev Mol Cell Biol *19*, 436–450. https://doi.org/10.1038/s41580-018-0008-z.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., *et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447. .

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin Immunol *109*, 17–28. https://doi.org/10.1016/s1521-6616(03)00201-8.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709–2721. https://doi.org/10.1093/nar/10.8.2709.

Ehrlich, M., Jackson, K., and Weemaes, C. (2006). Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet Journal of Rare Diseases *1*, 1. .

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet *29*, 569–574. https://doi.org/10.1016/j.tig.2013.05.010.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev *24*, 1546–1558. https://doi.org/10.1101/gad.573310.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., *et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973. https://doi.org/10.1126/science.1237973.

Enukashvily, N.I., Donev, R., Waisertreiger, I.S.-R., and Podgornaya, O.I. (2007). Human chromosome 1 satellite 3 DNA is decondensed, demethylated and transcribed in senescent cells and in A431 epithelial carcinoma cells. Cytogenet Genome Res *118*, 42–54. https://doi.org/10.1159/000106440.

Esteller, M. (2008). Epigenetics in cancer. N Engl J Med *358*, 1148–1159. https://doi.org/10.1056/NEJMra072067.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet *12*, 861–874. https://doi.org/10.1038/nrg3074.

Ewing, A.D., Smits, N., Sanchez-Luque, F.J., Faivre, J., Brennan, P.M., Richardson, S.R., Cheetham, S.W., and Faulkner, G.J. (2020). Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol Cell *80*, 915-928.e5. https://doi.org/10.1016/j.molcel.2020.10.024.

Eymery, Horard, B., El Atifi-Borel, M., Fourel, G., Berger, F., Vitte, A.-L., Van den Broeck, A., Brambilla, E., Fournier, A., Callanan, M., *et al.* (2009). A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Res *37*, 6340–6354. https://doi.org/10.1093/nar/gkp639.

Fachinetti, D., Folco, H.D., 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*, 1056–1066. https://doi.org/10.1038/ncb2805.

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol *20*, 332–338. https://doi.org/10.1038/nsmb.2495.

Faulkner, G.J., and Carninci, P. (2009). Altruistic functions for selfish DNA. Cell Cycle *8*, 2895–2900. https://doi.org/10.4161/cc.8.18.9536.

Felsenfeld, G. (2014). The evolution of epigenetics. Perspect Biol Med *57*, 132–148. https://doi.org/10.1353/pbm.2014.0004.

Feng, J.X., and Riddle, N.C. (2020). Epigenetics and genome stability. Mamm Genome *31*, 181–195. https://doi.org/10.1007/s00335-020-09836-2.

Ferrand, J., Plessier, A., and Polo, S.E. (2021). Control of the chromatin response to DNA damage: Histone proteins pull the strings. Semin Cell Dev Biol *113*, 75–87. https://doi.org/10.1016/j.semcdb.2020.07.002.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell *53*, 49–62. https://doi.org/10.1016/j.molcel.2013.10.030.

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009a). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res *37*, 5071–5080. https://doi.org/10.1093/nar/gkp529.

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009b). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res *37*, 5071–5080. https://doi.org/10.1093/nar/gkp529.

Francastel, C., and Magdinier, F. (2019). DNA methylation in satellite repeats disorders. Essays Biochem *63*, 757–771. https://doi.org/10.1042/EBC20190028.

Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740–741. https://doi.org/10.1038/171740a0.

Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V.K., Patil, N., and Cox, D.R. (2003). Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. Genome Res *13*, 341–346. https://doi.org/10.1101/gr.554603.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol *3*, 722–732. https://doi.org/10.1038/nrmicro1235.

Fryns, J.P., Azou, M., Jaeken, J., Eggermont, E., Pedersen, J.C., and Van den Berghe, H. (1981). Centromeric instability of chromosomes 1, 9, and 16 associated with combined immunodeficiency. Hum Genet *57*, 108–110. https://doi.org/10.1007/BF00271181.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev *15*, 490–495. https://doi.org/10.1016/j.gde.2005.08.002.

Fumagalli, M., Rossiello, F., Mondello, C., and d'Adda di Fagagna, F. (2014). Stable Cellular Senescence Is Associated with Persistent DDR Activation. PLoS ONE *9*, e110969. https://doi.org/10.1371/journal.pone.0110969.

Gaillard, H., and Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. Annu Rev Biochem *85*, 291–317. https://doi.org/10.1146/annurev-biochem-060815-014908.

Gaillard, M.-C., Roche, S., Dion, C., Tasmadjian, A., Bouget, G., Salort-Campana, E., Vovan, C., Chaix, C., Broucqsault, N., Morere, J., *et al.* (2014). Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology *83*, 733–742. https://doi.org/10.1212/WNL.0000000000000708.

Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol *31*, 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev *25*, 2041–2056. https://doi.org/10.1101/gad.17010011.

Gao, X.D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S.A., Zhu, L.J., Wolfe, S.A., *et al.* (2018). C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat Methods *15*, 433–436. https://doi.org/10.1038/s41592-018-0006-2.

García-de-Teresa, B., Rodríguez, A., and Frias, S. (2020). Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences. Genes (Basel) *11*, E1528. https://doi.org/10.3390/genes11121528.

García-Muse, T., and Aguilera, A. (2019a). R Loops: From Physiological to Pathological Roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.

García-Muse, T., and Aguilera, A. (2019b). R Loops: From Physiological to Pathological Roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.

Garcia-Perez, J.L., Marchetto, M.C.N., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet *16*, 1569–1577. https://doi.org/10.1093/hmg/ddm105.

García-Pichardo, D., Cañas, J.C., García-Rubio, M.L., Gómez-González, B., Rondón, A.G., and Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. Mol Cell *66*, 597-609.e5. https://doi.org/10.1016/j.molcel.2017.05.014.

Gartel, A.L., and Radhakrishnan, S.K. (2005). Lost in transcription: p21 repression, mechanisms, and consequences. Cancer Res *65*, 3980–3985. https://doi.org/10.1158/0008-5472.CAN-04-3995.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586. https://doi.org/10.1073/pnas.1208507109.

Gaubatz, J.W., and Cutler, R.G. (1990). Mouse satellite DNA is transcribed in senescent cardiac muscle. J Biol Chem *265*, 17753–17758. .

van Gent, D.C., Mizuuchi, K., and Gellert, M. (1996). Similarities between initiation of V(D)J recombination and retroviral integration. Science *271*, 1592–1594. https://doi.org/10.1126/science.271.5255.1592.

George, C.X., John, L., and Samuel, C.E. (2014). An RNA editor, adenosine deaminase acting on double-stranded RNA (ADAR1). J Interferon Cytokine Res *34*, 437–446. https://doi.org/10.1089/jir.2014.0001.

Ghosh, S., and Ghosh, A. (2021). Activation of DNA damage response signaling in mammalian cells by ionizing radiation. Free Radic Res *55*, 581–594. https://doi.org/10.1080/10715762.2021.1876853.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814–825. https://doi.org/10.1016/j.molcel.2012.01.017.

Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., *et al.* (2021a). CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc Natl Acad Sci U S A *118*, e2015634118. https://doi.org/10.1073/pnas.2015634118.

Giunta, S., Hervé, S., White, R.R., Wilhelm, T., Dumont, M., Scelfo, A., Gamba, R., Wong, C.K., Rancati, G., Smogorzewska, A., *et al.* (2021b). CENP-A chromatin prevents replication stress at centromeres to avoid structural aneuploidy. Proc Natl Acad Sci U S A *118*, e2015634118. https://doi.org/10.1073/pnas.2015634118.

Goll, M.G., and Bestor, T.H. (2005). EUKARYOTIC CYTOSINE METHYLTRANSFERASES. Annual Review of Biochemistry *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721.

Gollin, S.M. (2004). Chromosomal instability. Curr Opin Oncol *16*, 25–31. https://doi.org/10.1097/00001622-200401000-00006.

Gómez-González, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc Natl Acad Sci U S A *104*, 8409–8414. https://doi.org/10.1073/pnas.0702836104.

Goutte-Gattat, D., Shuaib, M., Ouararhni, K., Gautier, T., Skoufias, D.A., Hamiche, A., and Dimitrov, S. (2013). Phosphorylation of the CENP-A amino-terminus in mitotic centromeric chromatin is required for kinetochore function. Proc Natl Acad Sci U S A *110*, 8579–8584. https://doi.org/10.1073/pnas.1302955110.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol *20*, 590–607. https://doi.org/10.1038/s41580-019-0159-6.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331–337. https://doi.org/10.1038/337331a0.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res *35*, W52-57. https://doi.org/10.1093/nar/gkm360.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet *6*, e1000869. https://doi.org/10.1371/journal.pgen.1000869.

Grunseich, C., Wang, I.X., Watts, J.A., Burdick, J.T., Guber, R.D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A.B., *et al.* (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. Mol Cell *69*, 426-437.e7. https://doi.org/10.1016/j.molcel.2017.12.030.

Guenatri, M. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. The Journal of Cell Biology *166*, 493–505. https://doi.org/10.1083/jcb.200403109.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res *22*, 633–641. https://doi.org/10.1101/gr.130997.111.

Guldner, H.H., Lakomek, H.J., and Bautz, F.A. (1984). Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. Clin Exp Immunol *58*, 13–20. .

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., *et al.* (2013). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Neuroscience *17*, 215–222. https://doi.org/10.1038/nn.3607.

Gurard-Levin, Z.A., and Almouzni, G. (2014). Histone modifications and a choice of variant: a language that helps the genome express itself. F1000Prime Rep *6*, 76. https://doi.org/10.12703/P6-76.

Gutbrod, M.J., Roche, B., Steinberg, J.I., Lakhani, A.A., Chang, K., Schorn, A.J., and Martienssen, R.A. (2022). Dicer promotes genome stability via the bromodomain transcriptional co-activator BRD4. Nat Commun *13*, 1001. https://doi.org/10.1038/s41467-022-28554-8.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H.D., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., *et al.* (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369. https://doi.org/10.1016/j.stem.2013.06.002.

Hackett, J.A., and Surani, M.A. (2013). DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond B Biol Sci *368*, 20110328. https://doi.org/10.1098/rstb.2011.0328.

Hagan, C.R., and Rudin, C.M. (2007). DNA cleavage and Trp53 differentially affect SINE transcription. Genes Chromosomes Cancer *46*, 248–260. https://doi.org/10.1002/gcc.20406.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. Science *319*, 1352–1355. https://doi.org/10.1126/science.1140735.

Hall, L.E., Mitchell, S.E., and O'Neill, R.J. (2012). Pericentric and centromeric transcription: a perfect balance required. Chromosome Res *20*, 535–546. https://doi.org/10.1007/s10577-012-9297-9.

Hall, L.L., Byron, M., Carone, D.M., Whitfield, T.W., Pouliot, G.P., Fischer, A., Jones, P., and Lawrence, J.B. (2017). Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. Cell Rep *18*, 2943–2956. https://doi.org/10.1016/j.celrep.2017.02.072.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786.e19. https://doi.org/10.1016/j.cell.2017.07.043.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

Hansen, R.S., Stöger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D'Esposito, M., Feil, R., Gimelli, G., *et al.* (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol. Genet. *9*, 2575–2587. .

Hara, M., and Fukagawa, T. (2017). Critical Foundation of the Kinetochore: The Constitutive Centromere-Associated Network (CCAN). Prog Mol Subcell Biol *56*, 29–57. https://doi.org/10.1007/978-3-319-58592-5_2.

Harashima, H., Dissmeyer, N., and Schnittger, A. (2013). Cell cycle control across the eukaryotic kingdom. Trends Cell Biol *23*, 345–356. https://doi.org/10.1016/j.tcb.2013.03.002.

Harris, C.R., Dewan, A., Zupnick, A., Normart, R., Gabriel, A., Prives, C., Levine, A.J., and Hoh, J. (2009). p53 responsive elements in human retrotransposons. Oncogene *28*, 3857–3865. https://doi.org/10.1038/onc.2009.246.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chédin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol *430*, 272–284. https://doi.org/10.1016/j.jmb.2017.12.016.

Hauer, M.H., and Gasser, S.M. (2017). Chromatin and nucleosome dynamics in DNA damage and repair. Genes Dev *31*, 2204–2221. https://doi.org/10.1101/gad.307702.117.

Hayflick, L. (1965). THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. Exp Cell Res *37*, 614–636. https://doi.org/10.1016/0014-4827(65)90211-9.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp Cell Res *25*, 585–621. https://doi.org/10.1016/0014-4827(61)90192-6.

Héberlé, É., and Bardet, A.F. (2019). Sensitivity of transcription factors to DNA methylation. Essays Biochem *63*, 727–741. https://doi.org/10.1042/EBC20190033.

Hédouin, S., Grillo, G., Ivkovic, I., Velasco, G., and Francastel, C. (2017a). CENP-A chromatin disassembly in stressed and senescent murine cells. Sci Rep *7*, 42520. https://doi.org/10.1038/srep42520.

Hédouin, S., Grillo, G., Ivkovic, I., Velasco, G., and Francastel, C. (2017b). CENP-A chromatin disassembly in stressed and senescent murine cells. Sci Rep *7*, 42520. https://doi.org/10.1038/srep42520.

Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R.A. (2008). DNA repair pathways as targets for cancer therapy. Nat Rev Cancer *8*, 193–204. https://doi.org/10.1038/nrc2342.

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

Hergeth, S.P., and Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. EMBO Rep *16*, 1439–1453. https://doi.org/10.15252/embr.201540749.

Hiom, K., Melek, M., and Gellert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell *94*, 463–470. https://doi.org/10.1016/s0092-8674(00)81587-1.

Hoffmann, S., Dumont, M., Barra, V., Ly, P., Nechemia-Arbely, Y., McMahon, M.A., Hervé, S., Cleveland, D.W., and Fachinetti, D. (2016). CENP-A Is Dispensable for Mitotic Centromere Function after Initial Centromere/Kinetochore Assembly. Cell Rep *17*, 2394–2404. https://doi.org/10.1016/j.celrep.2016.10.084.

Holland, A.J., Fachinetti, D., Han, J.S., and Cleveland, D.W. (2012). Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. Proc Natl Acad Sci U S A *109*, E3350-3357. https://doi.org/10.1073/pnas.1216880109.

Howden, S.E., McColl, B., Glaser, A., Vadolas, J., Petrou, S., Little, M.H., Elefanty, A.G., and Stanley, E.G. (2016). A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells. Stem Cell Reports *7*, 508–517. https://doi.org/10.1016/j.stemcr.2016.07.001.

Hoyt, S.J., Storer, J.M., Hartley, G.A., Grady, P.G.S., Gershman, A., de Lima, L.G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., *et al.* (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science *376*, eabk3112. https://doi.org/10.1126/science.abk3112.

Hraiky, C., Raymond, M.A., and Drolet, M. (2000). RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in Escherichia coli. J Biol Chem *275*, 11257–11263. https://doi.org/10.1074/jbc.275.15.11257.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell *12*, 711–721. https://doi.org/10.1016/j.molcel.2003.08.010.

Hyjek, M., Figiel, M., and Nowotny, M. (2019a). RNases H: Structure and mechanism. DNA Repair (Amst) *84*, 102672. https://doi.org/10.1016/j.dnarep.2019.102672.

Hyjek, M., Figiel, M., and Nowotny, M. (2019b). RNases H: Structure and mechanism. DNA Repair (Amst) *84*, 102672. https://doi.org/10.1016/j.dnarep.2019.102672.

Ideue, T., Cho, Y., Nishimura, K., and Tani, T. (2014). Involvement of satellite I noncoding RNA in regulation of chromosome segregation. Genes Cells *19*, 528–538. https://doi.org/10.1111/gtc.12149.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands - 'A rough guide.' FEBS Letters *583*, 1713–1720. https://doi.org/10.1016/j.febslet.2009.04.012.

Innocente, S.A., Abrahamson, J.L., Cogswell, J.P., and Lee, J.M. (1999). p53 regulates a G2 checkpoint through cyclin B1. Proc Natl Acad Sci U S A *96*, 2147–2152. https://doi.org/10.1073/pnas.96.5.2147.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303. https://doi.org/10.1126/science.1210597.

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc Natl Acad Sci U S A *77*, 2450–2454. https://doi.org/10.1073/pnas.77.5.2450.

Iyama, T., and Wilson, D.M. (2013). DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) *12*, 620–636. https://doi.org/10.1016/j.dnarep.2013.04.015.

Jaco, I., Canela, A., Vera, E., and Blasco, M.A. (2008). Centromere mitotic recombination in mammalian cells. J Cell Biol *181*, 885–892. https://doi.org/10.1083/jcb.200803042.

Jakob, B., Splinter, J., Conrad, S., Voss, K.-O., Zink, D., Durante, M., Löbrich, M., and Taucher-Scholz, G. (2011). DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. Nucleic Acids Res *39*, 6489–6499. https://doi.org/10.1093/nar/gkr230.

Jansen, L.E.T., 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. https://doi.org/10.1083/jcb.200701066.

Jeffries, M.A. (2020). The Development of Epigenetics in the Study of Disease Pathogenesis. Adv Exp Med Biol *1253*, 57–94. https://doi.org/10.1007/978-981-15-3449-2_2.

Jenness, C., Giunta, S., Müller, M.M., Kimura, H., Muir, T.W., and Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci U S A *115*, E876–E885. https://doi.org/10.1073/pnas.1717509115.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080. https://doi.org/10.1126/science.1063127.

Jiang, F., and Doudna, J.A. (2015). The structural biology of CRISPR-Cas systems. Curr Opin Struct Biol *30*, 100–111. https://doi.org/10.1016/j.sbi.2015.02.002.

Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hulton, M., Jonveaux, P., Maraschio, P., Mogarban, A., *et al.* (2005). DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Human Mutation *25*, 56–63. https://doi.org/10.1002/humu.20113.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821. https://doi.org/10.1126/science.1225829.

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B.M., Khochbin, S., and Vourc'h, C. (2004). Stress-induced transcription of satellite III repeats. J Cell Biol *164*, 25–33. https://doi.org/10.1083/jcb.200306104.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics *13*, 484–492. https://doi.org/10.1038/nrg3230.

Jursch, T., Miskey, C., Izsvák, Z., and Ivics, Z. (2013). Regulation of DNA transposition by CpG methylation and chromatin structure in human cells. Mob DNA *4*, 15. https://doi.org/10.1186/1759-8753-4-15.

Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018a). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science *359*, 108–114. https://doi.org/10.1126/science.aan6490.

Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018b). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science *359*, 108–114. https://doi.org/10.1126/science.aan6490.

Kaer, K., and Speek, M. (2013). Retroelements in human disease. Gene *518*, 231–241. https://doi.org/10.1016/j.gene.2013.01.008.

Kalitsis, P., Griffiths, B., and Choo, K.H.A. (2006a). Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A *103*, 8786–8791. https://doi.org/10.1073/pnas.0600250103.

Kalitsis, P., Griffiths, B., and Choo, K.H.A. (2006b). Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A *103*, 8786–8791. https://doi.org/10.1073/pnas.0600250103.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489–501. https://doi.org/10.1101/gad.1248505.

Kang, H.J., Cheon, N.Y., Park, H., Jeong, G.W., Ye, B.J., Yoo, E.J., Lee, J.H., Hur, J.-H., Lee, E.-A., Kim, H., *et al.* (2021). TonEBP recognizes R-loops and initiates m6A RNA methylation for R-loop resolution. Nucleic Acids Res *49*, 269–284. https://doi.org/10.1093/nar/gkaa1162.

Kapranov, P., Ozsolak, F., and Milos, P.M. (2012). Profiling of short RNAs using Helicos single-molecule sequencing. Methods Mol Biol *822*, 219–232. https://doi.org/10.1007/978-1-61779-427-8_15.

Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. Mol Biol Evol *35*, 949–962. https://doi.org/10.1093/molbev/msy010.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469. https://doi.org/10.1126/science.1114955.

Kawaji, H., and Hayashizaki, Y. (2008). Exploration of small RNAs. PLoS Genet *4*, e22. https://doi.org/10.1371/journal.pgen.0040022.

Kazazian, H.H. (2011). Mobile DNA transposition in somatic cells. BMC Biol *9*, 62. https://doi.org/10.1186/1741-7007-9-62.

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223. https://doi.org/10.1242/dev.103341.

Keegan, L., Khan, A., Vukic, D., and O'Connell, M. (2017). ADAR RNA editing below the backbone. RNA *23*, 1317–1328. https://doi.org/10.1261/rna.060921.117.

Kiaee, F., Zaki-Dizaji, M., Hafezi, N., Almasi-Hashiani, A., Hamedifar, H., Sabzevari, A., Shirkani, A., Zian, Z., Jadidi-Niaragh, F., Aghamahdi, F., *et al.* (2021). Clinical, Immunologic and Molecular Spectrum of Patients with Immunodeficiency, Centromeric Instability, and Facial Anomalies (ICF) Syndrome: A Systematic Review. Endocr Metab Immune Disord Drug Targets *21*, 664–672. https://doi.org/10.2174/1871530320666200613204426.

Kim, A., and Wang, G.G. (2021). R-loop and its functions at the regulatory interfaces between transcription and (epi)genome. Biochim Biophys Acta Gene Regul Mech *1864*, 194750. https://doi.org/10.1016/j.bbagrm.2021.194750.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126–139. https://doi.org/10.1038/nrm2632.

Kipling, D., and Warburton, P.E. (1997). Centromeres, CENP-B and Tigger too. Trends Genet *13*, 141–145. https://doi.org/10.1016/s0168-9525(97)01098-6.

Kireeva, M.L., Komissarova, N., and Kashlev, M. (2000). Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. J Mol Biol *299*, 325–335. https://doi.org/10.1006/jmbi.2000.3755.

Kisseljova, N.P., Zueva, E.S., Pevzner, V.S., Grachev, A.N., and Kisseljov, F.L. (1998). De novo methylation of selective CpG dinucleotide clusters in transformed cells mediated by an activated N-ras. Int J Oncol *12*, 203–209. .

Kixmoeller, K., Allu, P.K., and Black, B.E. (2020a). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol *10*, 200051. https://doi.org/10.1098/rsob.200051.

Kixmoeller, K., Allu, P.K., and Black, B.E. (2020b). The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. Open Biol *10*, 200051. https://doi.org/10.1098/rsob.200051.

Klawitter, S., Fuchs, N.V., Upton, K.R., Muñoz-Lopez, M., Shukla, R., Wang, J., Garcia-Cañadas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A., *et al.* (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat Commun *7*, 10286. https://doi.org/10.1038/ncomms10286.

Kleinsmith, L.J., Allfrey, V.G., and Mirsky, A.E. (1966). Phosphorylation of nuclear protein early in the course of gene activation in lymphocytes. Science *154*, 780–781. https://doi.org/10.1126/science.154.3750.780.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol *18*, 490–495. https://doi.org/10.1016/j.cub.2008.03.016.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu Rev Biochem *79*, 213–231. https://doi.org/10.1146/annurev-biochem-010909-095056.

Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. Nat Rev Cancer *1*, 157–162. https://doi.org/10.1038/35101031.

Kondo, T., Bobek, M.P., Kuick, R., Lamb, B., Zhu, X., Narayan, A., Bourc'his, D., Viegas-Péquignot, E., Ehrlich, M., and Hanash, S.M. (2000). Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet *9*, 597–604. https://doi.org/10.1093/hmg/9.4.597.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015a). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015b). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. PLoS Genet *7*, e1002384. https://doi.org/10.1371/journal.pgen.1002384.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651–669. https://doi.org/10.1534/genetics.112.146704.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol *8*, R61. https://doi.org/10.1186/gb-2007-8-4-r61.

Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Honda, S., Kobayashi, O., Hatakeyama, K., Ushio, Y., Saya, H., and Hirota, T. (2003). CENP-A phosphorylation by Aurora-A in prophase is required for enrichment of Aurora-B at inner centromeres and for kinetochore function. Dev Cell *5*, 853–864. https://doi.org/10.1016/s1534-5807(03)00364-2.

Kuo, Y.-T., Ishii, T., Fuchs, J., Hsieh, W.-H., Houben, A., and Lin, Y.-R. (2021). The Evolutionary Dynamics of Repetitive DNA and Its Impact on the Genome Diversification in the Genus Sorghum. Front Plant Sci *12*, 729734. https://doi.org/10.3389/fpls.2021.729734.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol *32*, 677–683. https://doi.org/10.1038/nbt.2916.

Kwon, S.H., and Workman, J.L. (2011). The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. Bioessays *33*, 280–289. https://doi.org/10.1002/bies.201000138.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921. https://doi.org/10.1038/35057062.

Lara-Gonzalez, P., Pines, J., and Desai, A. (2021). Spindle assembly checkpoint activation and silencing at kinetochores. Semin Cell Dev Biol *117*, 86–98. https://doi.org/10.1016/j.semcdb.2021.06.009.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240. https://doi.org/10.1038/nature22822.

Lawrimore, J., and Bloom, K. (2019). The regulation of chromosome segregation via centromere loops. Crit Rev Biochem Mol Biol *54*, 352–370. https://doi.org/10.1080/10409238.2019.1670130.

Lee, Y.W., Arora, R., Wischnewski, H., and Azzalin, C.M. (2018). TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. Nat Struct Mol Biol *25*, 147–153. https://doi.org/10.1038/s41594-017-0021-5.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current Biology *13*, 1192–1200. .

Leonova, K.I., Brodsky, L., Lipchick, B., Pal, M., Novototskaya, L., Chenchik, A.A., Sen, G.C., Komarova, E.A., and Gudkov, A.V. (2013). p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. Proc Natl Acad Sci U S A *110*, E89-98. https://doi.org/10.1073/pnas.1216922110.

Li, M., and Fang, Y. (2015). Histone variants: the artists of eukaryotic chromatin. Sci China Life Sci *58*, 232–239. https://doi.org/10.1007/s11427-015-4817-4.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378. https://doi.org/10.1016/j.cell.2005.06.008.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926. https://doi.org/10.1016/0092-8674(92)90611-f.

Li, J., Sun, H., Huang, Y., Wang, Y., Liu, Y., and Chen, X. (2019). Pathways and assays for DNA double-strand break repair by homologous recombination. Acta Biochim Biophys Sin (Shanghai) *51*, 879–889. https://doi.org/10.1093/abbs/gmz076.

Li, M., Dong, Q., and Zhu, B. (2017). Aurora Kinase B Phosphorylates Histone H3.3 at Serine 31 during Mitosis in Mammalian Cells. J Mol Biol *429*, 2042–2045. https://doi.org/10.1016/j.jmb.2017.01.016.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu Rev Biochem *79*, 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131.

Lim, Y.W., Sanz, L.A., Xu, X., Hartono, S.R., and Chédin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. Elife *4*. https://doi.org/10.7554/eLife.08007.

Lindström, M.S., Jurada, D., Bursac, S., Orsolic, I., Bartek, J., and Volarevic, S. (2018). Nucleolus as an emerging hub in maintenance of genome stability and cancer pathogenesis. Oncogene *37*, 2351–2366. https://doi.org/10.1038/s41388-017-0121-z.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322. https://doi.org/10.1038/nature08514.

Liu, S.J., Dang, H.X., Lim, D.A., Feng, F.Y., and Maher, C.A. (2021). Long noncoding RNAs in cancer metastasis. Nat Rev Cancer *21*, 446–460. https://doi.org/10.1038/s41568-021-00353-1.

Locke, D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D., and Eichler, E.E. (2003). Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res *13*, 347–357. https://doi.org/10.1101/gr.1003303.

Logsdon, G.A., Vollger, M.R., Hsieh, P., Mao, Y., Liskovykh, M.A., Koren, S., Nurk, S., Mercuri, L., Dishuck, P.C., Rhie, A., *et al.* (2021). The structure, function and evolution of a complete human chromosome 8. Nature *593*, 101–107. https://doi.org/10.1038/s41586-021-03420-7.

López de Silanes, I., Stagno d'Alcontres, M., and Blasco, M.A. (2010). TERRA transcripts are bound by a complex array of RNA-binding proteins. Nat Commun *1*, 33. https://doi.org/10.1038/ncomms1032.

Lopez-Bergami, P., Lau, E., and Ronai, Z. (2010). Emerging roles of ATF2 and the dynamic AP1 network in cancer. Nat Rev Cancer *10*, 65–76. https://doi.org/10.1038/nrc2681.

Lu, J., and Gilbert, D.M. (2008). Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence? Cell Cycle *7*, 1907–1910. https://doi.org/10.4161/cc.7.13.6206.

Lu, T.X., and Rothenberg, M.E. (2018). MicroRNA. J Allergy Clin Immunol *141*, 1202–1207. https://doi.org/10.1016/j.jaci.2017.08.034.

Luger, K., Mäder, 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. https://doi.org/10.1038/38444.

Luke, B., and Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. EMBO J *28*, 2503–2510. https://doi.org/10.1038/emboj.2009.166.

Ma, J., Setton, J., Lee, N.Y., Riaz, N., and Powell, S.N. (2018). The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nat Commun *9*, 3292. https://doi.org/10.1038/s41467-018-05228-y.

Madaan, M., and Mendez, M.D. (2022). Angelman Syndrome. In StatPearls, (Treasure Island (FL): StatPearls Publishing), p.

Mahaney, B.L., Meek, K., and Lees-Miller, S.P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. Biochem J *417*, 639–650. https://doi.org/10.1042/BJ20080413.

Mailand, N., Falck, J., Lukas, C., Syljuâsen, R.G., Welcker, M., Bartek, J., and Lukas, J. (2000). Rapid destruction of human Cdc25A in response to DNA damage. Science *288*, 1425–1429. https://doi.org/10.1126/science.288.5470.1425.

Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nature Genetics *30*, 329–334. https://doi.org/10.1038/ng843.

Maison, C., Quivy, J.-P., Probst, A.V., and Almouzni, G. (2010). Heterochromatin at mouse pericentromeres: a model for de novo heterochromatin formation and duplication during replication. Cold Spring Harb Symp Quant Biol *75*, 155–165. https://doi.org/10.1101/sqb.2010.75.013.

Maison, C., Bailly, D., Roche, D., Montes de Oca, R., Probst, A.V., Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., *et al.* (2011). SUMOylation promotes de novo targeting of HP1α to pericentric heterochromatin. Nat Genet *43*, 220–227. https://doi.org/10.1038/ng.765.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826. https://doi.org/10.1126/science.1232033.

Malik, H.S., and Henikoff, S. (2009). Major Evolutionary Transitions in Centromere Complexity. Cell *138*, 1067–1082. https://doi.org/10.1016/j.cell.2009.08.036.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656–668. https://doi.org/10.1016/j.cell.2009.01.045.

Malumbres, M. (2014). Cyclin-dependent kinases. Genome Biol *15*, 122. https://doi.org/10.1186/gb4184.

Manic, G., Corradi, F., Sistigu, A., Siteni, S., and Vitale, I. (2017). Molecular Regulation of the Spindle Assembly Checkpoint by Kinases and Phosphatases. Int Rev Cell Mol Biol *328*, 105–161. https://doi.org/10.1016/bs.ircmb.2016.08.004.

Mao, Y.S., Sunwoo, H., Zhang, B., and Spector, D.L. (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol *13*, 95–101. https://doi.org/10.1038/ncb2140.

Maraschio, P., Zuffardi, O., Dalla Fior, T., and Tiepolo, L. (1988). Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. Journal of Medical Genetics *25*, 173–180. .

Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J., and Rauscher, F.J. (1994). Krüppel-associated boxes are potent transcriptional repression domains. Proc Natl Acad Sci U S A *91*, 4509–4513. https://doi.org/10.1073/pnas.91.10.4509.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. The EMBO Journal *24*, 800–812. .

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. https://doi.org/10.1083/jcb.109.5.1963.

Mathew, V., Pauleau, A.-L., Steffen, N., Bergner, A., Becker, P.B., and Erhardt, S. (2014). The histone-fold protein CHRAC14 influences chromatin composition in response to DNA damage. Cell Rep *7*, 321–330. https://doi.org/10.1016/j.celrep.2014.03.008.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum Mol Genet *15 Spec No 1*, R17-29. https://doi.org/10.1093/hmg/ddl046.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272–1278. https://doi.org/10.1038/nsmb.1903.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501–502. https://doi.org/10.1038/35000656.

Maze, I., Noh, K.-M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat Rev Genet *15*, 259–271. https://doi.org/10.1038/nrg3673.

McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol *17*, 16–29. https://doi.org/10.1038/nrm.2015.5.

McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017a). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Dev Cell *42*, 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001.

McNulty, S.M., Sullivan, L.L., and Sullivan, B.A. (2017b). Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. Developmental Cell *42*, 226-240.e6. https://doi.org/10.1016/j.devcel.2017.07.001.

Mellone, B.G., and Fachinetti, D. (2021). Diverse mechanisms of centromere specification. Curr Biol *31*, R1491–R1504. https://doi.org/10.1016/j.cub.2021.09.083.

Mendiola, A.J.P., and LaSalle, J.M. (2021). Epigenetics in Prader-Willi Syndrome. Front Genet *12*, 624581. https://doi.org/10.3389/fgene.2021.624581.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet *10*, 155–159. https://doi.org/10.1038/nrg2521.

Messeguer, X., Escudero, R., Farré, D., Núñez, O., Martínez, J., and Albà, M.M. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics *18*, 333–334. https://doi.org/10.1093/bioinformatics/18.2.333.

Metz, A., Soret, J., Vourc'h, C., Tazi, J., and Jolly, C. (2004). A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. J Cell Sci *117*, 4551–4558. https://doi.org/10.1242/jcs.01329.

Michaux, C., Verneuil, N., Hartke, A., and Giard, J.-C. (2014). Physiological roles of small RNA molecules. Microbiology (Reading) *160*, 1007–1019. https://doi.org/10.1099/mic.0.076208-0.

Miga, K.H., Newton, Y., Jain, M., Altemose, N., Willard, H.F., and Kent, W.J. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res *24*, 697–707. https://doi.org/10.1101/gr.159624.113.

Mihìc, P., Hédouin, S., and Francastel, C. (2021a). Centromeres Transcription and Transcripts for Better and for Worse. Prog Mol Subcell Biol *60*, 169–201. https://doi.org/10.1007/978-3-030-74889-0_7.

Mihìc, P., Hédouin, S., and Francastel, C. (2021b). Centromeres Transcription and Transcripts for Better and for Worse. Prog Mol Subcell Biol *60*, 169–201. https://doi.org/10.1007/978-3-030-74889-0_7.

Mikwar, M., MacFarlane, A.J., and Marchetti, F. (2020). Mechanisms of oocyte aneuploidy associated with advanced maternal age. Mutat Res Rev Mutat Res *785*, 108320. https://doi.org/10.1016/j.mrrev.2020.108320.

Millband, D.N., Campbell, L., and Hardwick, K.G. (2002). The awesome power of multiple model systems: interpreting the complex nature of spindle checkpoint signaling. Trends Cell Biol *12*, 205–209. https://doi.org/10.1016/s0962-8924(02)02276-6.

Miller, J., McLachlan, A.D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J *4*, 1609–1614. https://doi.org/10.1002/j.1460-2075.1985.tb03825.x.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol Cell *41*, 21–32. https://doi.org/10.1016/j.molcel.2010.12.007.

Mishra, P.K., Chakraborty, A., Yeh, E., Feng, W., Bloom, K.S., and Basrai, M.A. (2021). R-loops at centromeric chromatin contribute to defects in kinetochore integrity and chromosomal instability in budding yeast. Mol Biol Cell *32*, 74–89. https://doi.org/10.1091/mbc.E20-06-0379.

Misteli, T. (2001). The concept of self-organization in cellular architecture. J Cell Biol *155*, 181–185. https://doi.org/10.1083/jcb.200108110.

Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell *183*, 28–45. https://doi.org/10.1016/j.cell.2020.09.014.

Mitchell, A.R., Jeppesen, P., Nicol, L., Morrison, H., and Kipling, D. (1996). Epigenetic control of mammalian centromere protein binding: does DNA methylation have a role? Journal of Cell Science *109*, 2199–2206. .

Miyashita, T., Kitada, S., Krajewski, S., Horne, W.A., Delia, D., and Reed, J.C. (1995). Overexpression of the Bcl-2 protein increases the half-life of p21Bax. J Biol Chem *270*, 26049–26052. https://doi.org/10.1074/jbc.270.44.26049.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell *30*, 755–766. https://doi.org/10.1016/j.molcel.2008.05.007.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading) *155*, 733–740. https://doi.org/10.1099/mic.0.023960-0.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res *20*, 899–907. https://doi.org/10.1101/gr.103473.109.

Moran, E.C., Liu, L., Zasadzinska, E., Kestner, C.A., Sarkeshik, A., DeHoyos, H., Yates, J.R., Foltz, D., and Stukenberg, P.T. (2021a). Mitotic R-loops direct Aurora B kinase to maintain centromeric cohesion (Cell Biology).

Moran, E.C., Liu, L., Zasadzinska, E., Kestner, C.A., Sarkeshik, A., DeHoyos, H., Yates, J.R., Foltz, D., and Stukenberg, P.T. (2021b). Mitotic R-loops direct Aurora B kinase to maintain centromeric cohesion. BioRxiv 2021.01.14.426738. https://doi.org/10.1101/2021.01.14.426738.

Moser, S.C., and Swedlow, J.R. (2011). How to be a mitotic chromosome. Chromosome Res *19*, 307–319. https://doi.org/10.1007/s10577-011-9198-3.

Mullenders, L.H.F. (2018). Solar UV damage to cellular DNA: from mechanisms to biological effects. Photochem Photobiol Sci *17*, 1842–1852. https://doi.org/10.1039/c8pp00182k.

Musacchio, A., and Desai, A. (2017). A Molecular View of Kinetochore Assembly and Function. Biology (Basel) *6*, E5. https://doi.org/10.3390/biology6010005.

Musacchio, A., and Hardwick, K.G. (2002). The spindle checkpoint: structural insights into dynamic signalling. Nat Rev Mol Cell Biol *3*, 731–741. https://doi.org/10.1038/nrm929.

Musselman, C.A., Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol *19*, 1218–1227. https://doi.org/10.1038/nsmb.2436.

Nachtergaele, S., and He, C. (2017). The emerging biology of RNA post-transcriptional modifications. RNA Biol *14*, 156–163. https://doi.org/10.1080/15476286.2016.1267096.

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J *31*, 4020–4034. https://doi.org/10.1038/emboj.2012.251.

Nakano, K., and Vousden, K.H. (2001). PUMA, a novel proapoptotic gene, is induced by p53. Mol Cell *7*, 683–694. https://doi.org/10.1016/s1097-2765(01)00214-3.

Nasmyth, K. (2002). Segregating sister genomes: the molecular biology of chromosome separation. Science *297*, 559–565. https://doi.org/10.1126/science.1074757.

Nasmyth, K., Peters, J.M., and Uhlmann, F. (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. Science *288*, 1379–1385. https://doi.org/10.1126/science.288.5470.1379.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953. https://doi.org/10.1126/science.1236083.

Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA *12*, 1802–1809. https://doi.org/10.1261/rna.165106.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77. https://doi.org/10.1038/nature21373.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Mol Cell *65*, 832-847.e4. https://doi.org/10.1016/j.molcel.2017.01.029.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods *6*, 917–922. https://doi.org/10.1038/nmeth.1401.

Nishitani, H., Lygerou, Z., and Nishimoto, T. (2004). Proteolysis of DNA replication licensing factor Cdt1 in S-phase is performed independently of geminin through its N-terminal region. J Biol Chem *279*, 30807–30816. https://doi.org/10.1074/jbc.M312644200.

Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., *et al.* (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. Journal of Human Genetics *58*, 455–460. .

Novo, C.L., Wong, E.V., Hockings, C., Poudel, C., Sheekey, E., Wiese, M., Okkenhaug, H., Boulton, S.J., Basu, S., Walker, S., *et al.* (2022). Satellite repeat transcripts modulate heterochromatin condensates and safeguard chromosome stability in mouse embryonic stem cells. Nat Commun *13*, 3525. https://doi.org/10.1038/s41467-022-31198-3.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci *364*, 99–115. https://doi.org/10.1098/rstb.2008.0168.

Ohtani, N., Zebedee, Z., Huot, T.J., Stinson, J.A., Sugimoto, M., Ohashi, Y., Sharrocks, A.D., Peters, G., and Hara, E. (2001). Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. Nature *409*, 1067–1070. https://doi.org/10.1038/35059131.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257. https://doi.org/10.1016/s0092-8674(00)81656-6.

Ord, M.G., and Stocken, L.A. (1967). Phosphate and thiol groups in histone f3 from rat liver and thymus nuclei. Biochem J *102*, 631–636. https://doi.org/10.1042/bj1020631.

Oroz, J., and Laurents, D.V. (2019). RNA binding proteins: Diversity from microsurgeons to cowboys. Biochim Biophys Acta Gene Regul Mech *1862*, 194398. https://doi.org/10.1016/j.bbagrm.2019.06.009.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475–478. https://doi.org/10.1016/s0960-9822(00)00448-6.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Curr Opin Genet Dev *31*, 12–19. https://doi.org/10.1016/j.gde.2015.03.009.

Pal, S., and Tyler, J.K. (2016). Epigenetics and aging. Sci Adv *2*, e1600584. https://doi.org/10.1126/sciadv.1600584.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Front Genet *6*, 2. https://doi.org/10.3389/fgene.2015.00002.

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. https://doi.org/10.1073/pnas.88.9.3734.

Pannunzio, N.R., Watanabe, G., and Lieber, M.R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J Biol Chem *293*, 10512–10523. https://doi.org/10.1074/jbc.TM117.000374.

Papin, C., Ibrahim, A., Gras, S.L., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S., and Hamiche, A. (2017). Combinatorial DNA methylation codes at repetitive elements. Genome Res *27*, 934–946. https://doi.org/10.1101/gr.213983.116.

Pappalardo, X.G., and Barra, V. (2021). Losing DNA methylation at repetitive elements and breaking bad. Epigenetics Chromatin *14*, 25. https://doi.org/10.1186/s13072-021-00400-z.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809–817. https://doi.org/10.1126/science.2028256.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet *42*, 733–772. https://doi.org/10.1146/annurev.genet.42.110807.091711.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Mittelsten Scheid, O. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell *22*, 3118–3129. https://doi.org/10.1105/tpc.110.078493.

Peeters, S.B., Cotton, A.M., and Brown, C.J. (2014). Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. Bioessays *36*, 746–756. https://doi.org/10.1002/bies.201400032.

Penagos-Puig, A., and Furlan-Magaril, M. (2020). Heterochromatin as an Important Driver of Genome Organization. Front Cell Dev Biol *8*, 579137. https://doi.org/10.3389/fcell.2020.579137.

Perego, M.G.L., Taiana, M., Bresolin, N., Comi, G.P., and Corti, S. (2019). R-Loops in Motor Neuron Diseases. Mol Neurobiol *56*, 2579–2589. https://doi.org/10.1007/s12035-018-1246-y.

Pérez-Martínez, L., Öztürk, M., Butter, F., and Luke, B. (2020). Npl3 stabilizes R-loops at telomeres to prevent accelerated replicative senescence. EMBO Rep *21*, e49087. https://doi.org/10.15252/embr.201949087.

Perrin, A., Rousseau, J., and Tremblay, J.P. (2017). Increased Expression of Laminin Subunit Alpha 1 Chain by dCas9-VP160. Mol Ther Nucleic Acids *6*, 68–79. https://doi.org/10.1016/j.omtn.2016.11.004.

Peters, J.-M. (2002). The anaphase-promoting complex: proteolysis in mitosis and beyond. Mol Cell *9*, 931–943. https://doi.org/10.1016/s1097-2765(02)00540-3.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99–103. https://doi.org/10.1038/nature00862.

Pezer, Z., and Ugarković, D. (2008). RNA Pol II promotes transcription of centromeric satellite DNA in beetles. PLoS One *3*, e1594. https://doi.org/10.1371/journal.pone.0001594.

Pluta, A.F., Mackay, A.M., Ainsztein, A.M., Goldberg, I.G., and Earnshaw, W.C. (1995). The centromere: hub of chromosomal activities. Science *270*, 1591–1594. https://doi.org/10.1126/science.270.5242.1591.

Polo, S.E., and Jackson, S.P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev *25*, 409–433. https://doi.org/10.1101/gad.2021311.

Poon, R.Y.C. (2016). Cell Cycle Control: A System of Interlinking Oscillators. Methods Mol Biol *1342*, 3–19. https://doi.org/10.1007/978-1-4939-2957-3_1.

Potapova, O., Basu, S., Mercola, D., and Holbrook, N.J. (2001). Protective role for c-Jun in the cellular response to DNA damage. J Biol Chem *276*, 28546–28553. https://doi.org/10.1074/jbc.M102075200.

Prendergast, L., McClurg, U.L., Hristova, R., Berlinguer-Palmini, R., Greener, S., Veitch, K., Hernandez, I., Pasero, P., Rico, D., Higgins, J.M.G., *et al.* (2020). Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. Nat Commun *11*, 4534. https://doi.org/10.1038/s41467-020-18306-x.

Presting, G.G. (2018). Centromeric retrotransposons and centromere function. Current Opinion in Genetics & Development *49*, 79–84. https://doi.org/10.1016/j.gde.2018.03.004.

Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344–1354. https://doi.org/10.1016/j.cell.2013.02.011.

Probst, A.V., Okamoto, I., Casanova, M., El Marjou, F., Le Baccon, P., and Almouzni, G. (2010). A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev Cell *19*, 625–638. https://doi.org/10.1016/j.devcel.2010.09.002.

Przewloka, M.R., and Glover, D.M. (2009). The kinetochore and the centromere: a working long distance relationship. Annu Rev Genet *43*, 439–465. https://doi.org/10.1146/annurev-genet-102108-134310.

Qiu, S., and Huang, J. (2021). MRN complex is an essential effector of DNA damage repair. J Zhejiang Univ Sci B *22*, 31–37. https://doi.org/10.1631/jzus.B2000289.

Quénet, D., and Dalal, Y. (2014a). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. Elife *3*, e03254. https://doi.org/10.7554/eLife.03254.

Quénet, D., and Dalal, Y. (2014b). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. ELife *3*. https://doi.org/10.7554/eLife.03254.

Racca, C., Britton, S., Hédouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021a). BRCA1 prevents R-loop-associated centromeric instability. Cell Death Dis *12*, 896. https://doi.org/10.1038/s41419-021-04189-3.

Racca, C., Britton, S., Hédouin, S., Francastel, C., Calsou, P., and Larminat, F. (2021b). BRCA1 prevents R-loop-associated centromeric instability. Cell Death Dis *12*, 896. https://doi.org/10.1038/s41419-021-04189-3.

Rayess, H., Wang, M.B., and Srivatsan, E.S. (2012). Cellular senescence and tumor suppressor gene p16. Int J Cancer *130*, 1715–1725. https://doi.org/10.1002/ijc.27316.

Redon, S., Reichenbach, P., and Lingner, J. (2010). The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic Acids Res *38*, 5797–5806. https://doi.org/10.1093/nar/gkq296.

Ren, W., Gao, L., and Song, J. (2018). Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. Genes (Basel) *9*, E620. https://doi.org/10.3390/genes9120620.

Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. Mol Cell *70*, 650-662.e8. https://doi.org/10.1016/j.molcel.2018.04.001.

Richard, P., and Manley, J.L. (2017). R Loops and Links to Human Disease. J Mol Biol *429*, 3168–3180. https://doi.org/10.1016/j.jmb.2016.08.031.

Rieder, C.L., Cole, R.W., Khodjakov, A., and Sluder, G. (1995). The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. J Cell Biol *130*, 941–948. https://doi.org/10.1083/jcb.130.4.941.

Rinn, J.L. (2014). lncRNAs: linking RNA to chromatin. Cold Spring Harb Perspect Biol *6*, a018614. https://doi.org/10.1101/cshperspect.a018614.

Rippe, K. (2022a). Liquid-Liquid Phase Separation in Chromatin. Cold Spring Harb Perspect Biol *14*, a040683. https://doi.org/10.1101/cshperspect.a040683.

Rippe, K. (2022b). Liquid-Liquid Phase Separation in Chromatin. Cold Spring Harb Perspect Biol *14*, a040683. https://doi.org/10.1101/cshperspect.a040683.

Rizzi, N., Denegri, M., Chiodi, I., Corioni, M., Valgardsdottir, R., Cobianchi, F., Riva, S., and Biamonti, G. (2004). Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. Mol Biol Cell *15*, 543–551. https://doi.org/10.1091/mbc.e03-07-0487.

Roberts, S.A., Strande, N., Burkhalter, M.D., Strom, C., Havener, J.M., Hasty, P., and Ramsden, D.A. (2010). Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. Nature *464*, 1214–1217. https://doi.org/10.1038/nature08926.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem *273*, 5858–5868. https://doi.org/10.1074/jbc.273.10.5858.

Rollins, R.A. (2005). Large-scale structure of genomic methylation patterns. Genome Research *16*, 157–163. https://doi.org/10.1101/gr.4362006.

Rošić, S., Köhler, F., and Erhardt, S. (2014a). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol *207*, 335–349. https://doi.org/10.1083/jcb.201404097.

Rošić, S., Köhler, F., and Erhardt, S. (2014b). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol *207*, 335–349. https://doi.org/10.1083/jcb.201404097.

Roy, D., Yu, K., and Lieber, M.R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. Mol Cell Biol *28*, 50–60. https://doi.org/10.1128/MCB.01251-07.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.-L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Mol Cell Biol *30*, 146–159. https://doi.org/10.1128/MCB.00897-09.

Ruchaud, S., Carmena, M., and Earnshaw, W.C. (2007). Chromosomal passengers: conducting cell division. Nat Rev Mol Cell Biol *8*, 798–812. https://doi.org/10.1038/nrm2257.

Sage, J., Miller, A.L., Pérez-Mancera, P.A., Wysocki, J.M., and Jacks, T. (2003). Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. Nature *424*, 223–228. https://doi.org/10.1038/nature01764.

Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chédin, F., and Selig, S. (2017). Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. Nat Commun *8*, 14015. https://doi.org/10.1038/ncomms14015.

Saha, N., and Muntean, A.G. (2021). Insight into the multi-faceted role of the SUV family of H3K9 methyltransferases in carcinogenesis and cancer progression. Biochim Biophys Acta Rev Cancer *1875*, 188498. https://doi.org/10.1016/j.bbcan.2020.188498.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3. https://doi.org/10.1186/1756-8935-8-3.

Santiago-Frangos, A., Buyukyoruk, M., Wiegand, T., Krishna, P., and Wiedenheft, B. (2021). Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. Curr Biol *31*, 3515-3524.e6. https://doi.org/10.1016/j.cub.2021.05.068.

Sanz, L.A., and Chédin, F. (2019a). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734–1755. https://doi.org/10.1038/s41596-019-0159-1.

Sanz, L.A., and Chédin, F. (2019b). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734–1755. https://doi.org/10.1038/s41596-019-0159-1.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell *63*, 167–178. https://doi.org/10.1016/j.molcel.2016.05.032.

Sarkar, K., Han, S.-S., Wen, K.-K., Ochs, H.D., Dupré, L., Seidman, M.M., and Vyas, Y.M. (2018). R-loops cause genomic instability in T helper lymphocytes from patients with Wiskott-Aldrich syndrome. J Allergy Clin Immunol *142*, 219–234. https://doi.org/10.1016/j.jaci.2017.11.023.

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A *106*, 2525–2530. https://doi.org/10.1073/pnas.0807899106.

Saul, R.L., and Ames, B.N. (1986). Background levels of DNA damage in the population. Basic Life Sci *38*, 529–535. https://doi.org/10.1007/978-1-4615-9462-8_55.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., *et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife *2*, e01749. https://doi.org/10.7554/eLife.01749.

Scelfo, A., and Fachinetti, D. (2019). Keeping the Centromere under Control: A Promising Role for DNA Methylation. Cells *8*. https://doi.org/10.3390/cells8080912.

Schmeer, C., Kretz, A., Wengerodt, D., Stojiljkovic, M., and Witte, O.W. (2019). Dissecting Aging and Senescence-Current Concepts and Open Lessons. Cells *8*, E1446. https://doi.org/10.3390/cells8111446.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., and Leonhardt, H. (2016a). Determination of local chromatin composition by CasID. Nucleus *7*, 476–484. https://doi.org/10.1080/19491034.2016.1239000.

Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F., and Leonhardt, H. (2016b). Determination of local chromatin composition by CasID. Nucleus *7*, 476–484. https://doi.org/10.1080/19491034.2016.1239000.

Schmitz, J., and Brosius, J. (2011). Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie *93*, 1928–1934. https://doi.org/10.1016/j.biochi.2011.07.014.

Schoeftner, S., and Blasco, M.A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol *10*, 228–236. https://doi.org/10.1038/ncb1685.

Schumann, G.G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans *35*, 637–642. https://doi.org/10.1042/BST0350637.

Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M., and Devine, S.E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res *26*, 745–755. https://doi.org/10.1101/gr.201814.115.

Sekulic, N., and Black, B.E. (2012). Molecular underpinnings of centromere identity and maintenance. Trends Biochem Sci *37*, 220–229. https://doi.org/10.1016/j.tibs.2012.01.003.

Senga, S.S., and Grose, R.P. (2021). Hallmarks of cancer-the new testament. Open Biol *11*, 200358. https://doi.org/10.1098/rsob.200358.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593–602. https://doi.org/10.1016/s0092-8674(00)81902-9.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026–1039. https://doi.org/10.1016/j.biochi.2008.04.011.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat Methods *12*, 664–670. https://doi.org/10.1038/nmeth.3433.

Shen, D., and Skibbens, R.V. (2017). Temperature-dependent regulation of rDNA condensation in Saccharomyces cerevisiae. Cell Cycle *16*, 1118–1127. https://doi.org/10.1080/15384101.2017.1317409.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat Cell Biol *13*, 167–173. https://doi.org/10.1038/ncb2157.

Shiio, Y., and Eisenman, R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A *100*, 13225–13230. https://doi.org/10.1073/pnas.1735528100.

Silva, B., Arora, R., Bione, S., and Azzalin, C.M. (2021). TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. Nat Commun *12*, 3760. https://doi.org/10.1038/s41467-021-24097-6.

Silva, S., Camino, L.P., and Aguilera, A. (2018). Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. Proc Natl Acad Sci U S A *115*, 11024–11029. https://doi.org/10.1073/pnas.1807258115.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., *et al.* (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115. https://doi.org/10.1038/nature23875.

Singh, P.P., Shukla, M., White, S.A., Lafos, M., Tong, P., Auchynnikava, T., Spanos, C., Rappsilber, J., Pidoux, A.L., and Allshire, R.C. (2020). Hap2-Ino80-facilitated transcription promotes de novo establishment of CENP-A chromatin. Genes Dev *34*, 226–238. https://doi.org/10.1101/gad.332536.119.

Sinzelle, L., Izsvák, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci *66*, 1073–1093. https://doi.org/10.1007/s00018-009-8376-3.

Sitbon, D., Podsypanina, K., Yadav, T., and Almouzni, G. (2017). Shaping Chromatin in the Nucleus: The Bricks and the Architects. Cold Spring Harb Symp Quant Biol *82*, 1–14. https://doi.org/10.1101/sqb.2017.82.033753.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794–805. https://doi.org/10.1016/j.molcel.2011.04.026.

Slack, F.J., and Chinnaiyan, A.M. (2019). The Role of Non-coding RNAs in Oncology. Cell *179*, 1033–1055. https://doi.org/10.1016/j.cell.2019.10.017.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88. https://doi.org/10.1126/science.aad5227.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol *14*, 1138–1152. https://doi.org/10.1080/15476286.2016.1259781.

Smolka, J.A., Sanz, L.A., Hartono, S.R., and Chédin, F. (2021). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. J Cell Biol *220*, e202004079. https://doi.org/10.1083/jcb.202004079.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol *25*, 514–522. https://doi.org/10.1016/j.tcb.2015.05.003.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell *56*, 777–785. https://doi.org/10.1016/j.molcel.2014.10.020.

Spector, D.L. (2003). The dynamics of chromosome organization and gene regulation. Annu Rev Biochem *72*, 573–608. https://doi.org/10.1146/annurev.biochem.72.121801.161724.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. Mol Cell *52*, 146–152. https://doi.org/10.1016/j.molcel.2013.09.008.

St Clair, S., and Manfredi, J.J. (2006). The dual specificity phosphatase Cdc25C is a direct target for transcriptional repression by the tumor suppressor p53. Cell Cycle *5*, 709–713. https://doi.org/10.4161/cc.5.7.2628.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature https://doi.org/10.1038/nature10716.

Stanley, J.S., Griffin, J.B., and Zempleni, J. (2001). Biotinylation of histones in human cells. Effects of cell proliferation. Eur J Biochem *268*, 5424–5429. https://doi.org/10.1046/j.0014-2956.2001.02481.x.

Stein, G.H., Drullinger, L.F., Soulard, A., and Dulić, V. (1999). Differential roles for cyclin-dependent kinase inhibitors p21 and p16 in the mechanisms of senescence and differentiation in human fibroblasts. Mol Cell Biol *19*, 2109–2117. https://doi.org/10.1128/MCB.19.3.2109.

Stodola, J.L., and Burgers, P.M. (2017). Mechanism of Lagging-Strand DNA Replication in Eukaryotes. Adv Exp Med Biol *1042*, 117–133. https://doi.org/10.1007/978-981-10-6955-0_6.

Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chédin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife *5*, e17548. https://doi.org/10.7554/eLife.17548.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45. https://doi.org/10.1038/47412.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245. https://doi.org/10.1038/nature22989.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell *123*, 1213–1226. https://doi.org/10.1016/j.cell.2005.09.038.

Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., and Felsenfeld, G. (2004). Chromatin remodeling by RNA polymerases. Trends Biochem Sci *29*, 127–135. https://doi.org/10.1016/j.tibs.2004.01.003.

Sugino, A., Hirose, S., and Okazaki, R. (1972). RNA-linked nascent DNA fragments in Escherichia coli. Proc Natl Acad Sci U S A *69*, 1863–1867. https://doi.org/10.1073/pnas.69.7.1863.

Sullivan, B.A., Schwartz, S., and Willard, H.F. (1996). Centromeres of human chromosomes. Environ Mol Mutagen *28*, 182–191. https://doi.org/10.1002/(SICI)1098-2280(1996)28:3<182::AID-EM4>3.0.CO;2-G.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619–621. https://doi.org/10.1126/science.1234848.

Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional *Drosophila* centromere. Genome Res *13*, 182–194. https://doi.org/10.1101/gr.681703.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465–476. https://doi.org/10.1038/nrg2341.

Takatsuka, H., Shibata, A., and Umeda, M. (2021). Genome Maintenance Mechanisms at the Chromatin Level. Int J Mol Sci *22*, 10384. https://doi.org/10.3390/ijms221910384.

Talbert, P.B., and Henikoff, S. (2010). Centromeres convert but don't cross. PLoS Biol *8*, e1000326. https://doi.org/10.1371/journal.pbio.1000326.

Talbert, P.B., and Henikoff, S. (2021). Histone variants at a glance. J Cell Sci *134*, jcs244749. https://doi.org/10.1242/jcs.244749.

Talbert, P.B., Bryson, T.D., and Henikoff, S. (2004). Adaptive evolution of centromere proteins in plants and animals. J Biol *3*, 18. https://doi.org/10.1186/jbiol11.

Tan, J., Duan, M., Yadav, T., Phoon, L., Wang, X., Zhang, J.-M., Zou, L., and Lan, L. (2020). An R-loop-initiated CSB-RAD52-POLD3 pathway suppresses ROS-induced telomeric DNA breaks. Nucleic Acids Res *48*, 1285–1300. https://doi.org/10.1093/nar/gkz1114.

Tanaka, Y., Kurumizaka, H., and Yokoyama, S. (2005). CpG methylation of the CENP-B box reduces human CENP-B binding. FEBS J *272*, 282–289. https://doi.org/10.1111/j.1432-1033.2004.04406.x.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646. https://doi.org/10.1016/j.cell.2014.09.039.

Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., Wei, Z., Su, J., Liu, G., Rong, R., *et al.* (2021). m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. Nucleic Acids Res *49*, D134–D143. https://doi.org/10.1093/nar/gkaa692.

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. https://doi.org/10.1038/nsmb1338.

Terranova, R., Sauer, S., Merkenschlager, M., and Fisher, A.G. (2005). The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. Exp Cell Res *310*, 344–356. https://doi.org/10.1016/j.yexcr.2005.07.031.

Thakur, J., Packiaraj, J., and Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. Int J Mol Sci *22*, 4309. https://doi.org/10.3390/ijms22094309.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., *et al.* (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun *6*. https://doi.org/10.1038/ncomms8870.

Thiru, P., Kern, D.M., McKinley, K.L., Monda, J.K., Rago, F., Su, K.-C., Tsinman, T., Yarar, D., Bell, G.W., and Cheeseman, I.M. (2014). Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. Mol Biol Cell *25*, 1983–1994. https://doi.org/10.1091/mbc.E14-03-0837.

Thomas, M., White, R.L., and Davis, R.W. (1976a). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci U S A *73*, 2294–2298. https://doi.org/10.1073/pnas.73.7.2294.

Thomas, M., White, R.L., and Davis, R.W. (1976b). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci U S A *73*, 2294–2298. https://doi.org/10.1073/pnas.73.7.2294.

Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G.F., and Romano, C. (1979). Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. Hum Genet *51*, 127–137. https://doi.org/10.1007/BF00287166.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., *et al.* (2011a). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. Science *331*, 593–596. https://doi.org/10.1126/science.1200801.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., *et al.* (2011b). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science *331*, 593–596. https://doi.org/10.1126/science.1200801.

Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., and Vaillant, I. (2010). Stress-induced activation of heterochromatic transcription. PLoS Genet *6*, e1001175. https://doi.org/10.1371/journal.pgen.1001175.

Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004a). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci U S A *101*, 15986–15991. https://doi.org/10.1073/pnas.0407154101.

Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004b). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc Natl Acad Sci U S A *101*, 15986–15991. https://doi.org/10.1073/pnas.0407154101.

Toubiana, S., Velasco, G., Chityat, A., Kaindl, A.M., Hershtig, N., Tzur-Gilat, A., Francastel, C., and Selig, S. (2018). Subtelomeric methylation distinguishes between subtypes of Immunodeficiency, Centromeric instability and Facial anomalies syndrome. Hum Mol Genet *27*, 3568–3581. https://doi.org/10.1093/hmg/ddy265.

Trivedi, P., and Stukenberg, P.T. (2016). A Centromere-Signaling Network Underlies the Coordination among Mitotic Events. Trends Biochem Sci *41*, 160–174. https://doi.org/10.1016/j.tibs.2015.11.002.

Trivedi, P., Palomba, F., Niedzialkowska, E., Digman, M.A., Gratton, E., and Stukenberg, P.T. (2019). The inner centromere is a biomolecular condensate scaffolded by the chromosomal passenger complex. Nat Cell Biol *21*, 1127–1137. https://doi.org/10.1038/s41556-019-0376-4.

Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. Mol Cell *63*, 293–305. https://doi.org/10.1016/j.molcel.2016.06.002.

Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell *168*, 644–656. https://doi.org/10.1016/j.cell.2017.01.002.

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., *et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol *11*, 1315–1324. https://doi.org/10.1038/ncb1984.

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., and Sasaki, H. (2020a). CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. Sci Rep *10*, 17865. https://doi.org/10.1038/s41598-020-74636-2.

Unoki, M., Sharif, J., Saito, Y., Velasco, G., Francastel, C., Koseki, H., and Sasaki, H. (2020b). CDCA7 and HELLS suppress DNA:RNA hybrid-associated DNA damage at pericentromeric repeats. Sci Rep *10*, 17865. https://doi.org/10.1038/s41598-020-74636-2.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636–646. https://doi.org/10.1038/nrg2842.

Uruci, S., Lo, C.S.Y., Wheeler, D., and Taneja, N. (2021). R-Loops and Its Chro-Mates: The Strange Case of Dr. Jekyll and Mr. Hyde. Int J Mol Sci *22*, 8850. https://doi.org/10.3390/ijms22168850.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. Nat Rev Genet *19*, 535–548. https://doi.org/10.1038/s41576-018-0017-y.

Valgardsdottir, R., Chiodi, I., Giordano, M., Rossi, A., Bazzini, S., Ghigna, C., Riva, S., and Biamonti, G. (2008). Transcription of Satellite III non-coding RNAs is a general stress response in human cells. Nucleic Acids Res *36*, 423–434. https://doi.org/10.1093/nar/gkm1056.

Varma, D., and Salmon, E.D. (2012). The KMN protein network--chief conductors of the kinetochore orchestra. J Cell Sci *125*, 5927–5936. https://doi.org/10.1242/jcs.093724.

Vaskova, E.A., Dementyeva, E.V., Shevchenko, A.I., Pavlova, S.V., Grigor'eva, E.V., Zhelezova, A.I., Vandeberg, J.L., and Zakian, S.M. (2014). Dynamics of the two heterochromatin types during imprinted X chromosome inactivation in vole Microtus levis. PLoS One *9*, e88256. https://doi.org/10.1371/journal.pone.0088256.

Velasco, G., and Francastel, C. (2019). Genetics meets DNA methylation in rare diseases. Clin Genet *95*, 210–220. https://doi.org/10.1111/cge.13480.

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc Natl Acad Sci U S A *107*, 9281–9286. https://doi.org/10.1073/pnas.1000473107.

Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., and Francastel, C. (2018). Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Hum Mol Genet *27*, 2409–2424. https://doi.org/10.1093/hmg/ddy130.

Velazquez Camacho, O., Galan, C., Swist-Rosowska, K., Ching, R., Gamalinda, M., Karabiber, F., De La Rosa-Velazquez, I., Engist, B., Koschorz, B., Shukeir, N., *et al.* (2017). Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife *6*, e25293. https://doi.org/10.7554/eLife.25293.

Vijg, J., and Suh, Y. (2013). Genome instability and aging. Annu Rev Physiol *75*, 645–668. https://doi.org/10.1146/annurev-physiol-030212-183715.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193. https://doi.org/10.1016/j.cell.2012.09.002.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837. https://doi.org/10.1126/science.1074973.

Vourc'h, C., and Biamonti, G. (2011). Transcription of Satellite DNAs in Mammals. Prog Mol Subcell Biol *51*, 95–118. https://doi.org/10.1007/978-3-642-16502-3_5.

Vydzhak, O., Luke, B., and Schindler, N. (2020). Non-coding RNAs at the Eukaryotic rDNA Locus: RNA-DNA Hybrids and Beyond. J Mol Biol *432*, 4287–4304. https://doi.org/10.1016/j.jmb.2020.05.011.

Waddington, C.H. (2012). The Epigenotype. Int. J. Epidemiol. *41*, 10–13. https://doi.org/10.1093/ije/dyr184.

Wahba, L., Gore, S.K., and Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. Elife *2*, e00505. https://doi.org/10.7554/eLife.00505.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev *30*, 1327–1338. https://doi.org/10.1101/gad.280834.116.

Wang, J., and Davis, R.E. (2014). Programmed DNA elimination in multicellular organisms. Curr Opin Genet Dev *27*, 26–34. https://doi.org/10.1016/j.gde.2014.03.012.

Wang, Y., and Chen, L.-L. (2020). Organization and function of paraspeckles. Essays Biochem *64*, 875–882. https://doi.org/10.1042/EBC20200010.

Wang, H., Perrault, A.R., Takeda, Y., Qin, W., Wang, H., and Iliakis, G. (2003). Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Res *31*, 5377–5388. https://doi.org/10.1093/nar/gkg728.

Wang, K., Peng, J., and Yi, C. (2021). The m6A Consensus Motif Provides a Paradigm of Epitranscriptomic Studies. Biochemistry *60*, 3410–3412. https://doi.org/10.1021/acs.biochem.1c00254.

Wang, Y., Park, S., and Beal, P.A. (2018). Selective Recognition of RNA Substrates by ADAR Deaminase Domains. Biochemistry *57*, 1640–1651. https://doi.org/10.1021/acs.biochem.7b01100.

Wassmann, K., and Benezra, R. (2001). Mitotic checkpoints: from yeast to cancer. Curr Opin Genet Dev *11*, 83–90. https://doi.org/10.1016/s0959-437x(00)00161-1.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738. https://doi.org/10.1038/171737a0.

Weber, M., and Schübeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Current Opinion in Cell Biology *19*, 273–280. https://doi.org/10.1016/j.ceb.2007.04.011.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics *39*, 457–466. https://doi.org/10.1038/ng1990.

Wellinger, R.E., Prado, F., and Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol *26*, 3327–3334. https://doi.org/10.1128/MCB.26.8.3327-3334.2006.

Westhorpe, F.G., and Straight, A.F. (2013). Functions of the centromere and kinetochore in chromosome segregation. Curr Opin Cell Biol *25*, 334–340. https://doi.org/10.1016/j.ceb.2013.02.001.

White, R.L., and Hogness, D.S. (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. Cell *10*, 177–192. https://doi.org/10.1016/0092-8674(77)90213-6.

Wijmenga, C., Hansen, R.S., Gimelli, G., Björck, E.J., Davies, E.G., Valentine, D., Belohradsky, B.H., van Dongen, J.J., Smeets, D.F., van den Heuvel, L.P., *et al.* (2000). Genetic variation in ICF syndrome: evidence for genetic heterogeneity. Human Mutation *16*, 509–517. .

Wilkins, M.H.F., and Randall, J.T. (1953). Crystallinity in sperm heads: molecular structure of nucleoprotein in vivo. Biochim Biophys Acta *10*, 192–193. https://doi.org/10.1016/0006-3002(53)90232-7.

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.* (2007a). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160. https://doi.org/10.1101/gr.6022807.

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.* (2007b). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res *17*, 1146–1160. https://doi.org/10.1101/gr.6022807.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem Soc Trans *38*, 1027–1032. https://doi.org/10.1042/BST0381027.

Wright, S.I., Agrawal, N., and Bureau, T.E. (2003). Effects of recombination rate and gene density on transposable element distributions in Arabidopsis thaliana. Genome Res *13*, 1897–1903. https://doi.org/10.1101/gr.1281503.

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Human Molecular Genetics ddw243. .

Xu, B., and Clayton, D.A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. EMBO J *15*, 3135–3143. .

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187–191. https://doi.org/10.1038/46052.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.-L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov *2*, 16009. https://doi.org/10.1038/celldisc.2016.9.

Yang, X., Liu, Q.-L., Xu, W., Zhang, Y.-C., Yang, Y., Ju, L.-F., Chen, J., Chen, Y.-S., Li, K., Ren, J., *et al.* (2019). m6A promotes R-loop formation to facilitate transcription termination. Cell Res *29*, 1035–1038. https://doi.org/10.1038/s41422-019-0235-7.

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H.S., Rippe, K., Meshorer, E., *et al.* (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. Cell Rep *10*, 1122–1134. https://doi.org/10.1016/j.celrep.2015.01.038.

Yilmaz, D., Furst, A., Meaburn, K., Lezaja, A., Wen, Y., Altmeyer, M., Reina-San-Martin, B., and Soutoglou, E. (2021a). Activation of homologous recombination in G1 preserves centromeric integrity. Nature *600*, 748–753. https://doi.org/10.1038/s41586-021-04200-z.

Yilmaz, D., Furst, A., Meaburn, K., Lezaja, A., Wen, Y., Altmeyer, M., Reina-San-Martin, B., and Soutoglou, E. (2021b). Activation of homologous recombination in G1 preserves centromeric integrity. Nature *600*, 748–753. https://doi.org/10.1038/s41586-021-04200-z.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523. https://doi.org/10.1038/nature06968.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics *13*, 335–340. https://doi.org/10.1016/S0168-9525(97)01181-5.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res *40*, 5569–5576. https://doi.org/10.1093/nar/gks216.

Yu, H. (2002). Regulation of APC-Cdc20 by the spindle checkpoint. Curr Opin Cell Biol *14*, 706–714. https://doi.org/10.1016/s0955-0674(02)00382-4.

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442–451. https://doi.org/10.1038/ni919.

Zagrovic, B., Bartonek, L., and Polyansky, A.A. (2018). RNA-protein interactions in an unstructured context. FEBS Lett *592*, 2901–2916. https://doi.org/10.1002/1873-3468.13116.

Zaitsev, E.N., and Kowalczykowski, S.C. (2000). A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange. Genes Dev *14*, 740–749. .

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., *et al.* (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350. https://doi.org/10.1016/j.cell.2014.11.052.

Zeitlin, S.G., Baker, N.M., Chapados, B.R., Soutoglou, E., Wang, J.Y.J., Berns, M.W., and Cleveland, D.W. (2009). Double-strand DNA breaks recruit the centromeric histone CENP-A. Proc Natl Acad Sci U S A *106*, 15762–15767. https://doi.org/10.1073/pnas.0908233106.

Zeng, C., Onoguchi, M., and Hamada, M. (2021). Association analysis of repetitive elements and R-loop formation across species. Mob DNA *12*, 3. https://doi.org/10.1186/s13100-021-00231-5.

Zhang, W., Lee, H.-R., Koo, D.-H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell *20*, 25–34. https://doi.org/10.1105/tpc.107.057083.

Zhang, W., Mao, J.-H., Zhu, W., Jain, A.K., Liu, K., Brown, J.B., and Karpen, G.H. (2016). Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. Nat Commun *7*, 12619. https://doi.org/10.1038/ncomms12619.

Zhang, X., Chiang, H.-C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S.J., Michalek, J., Jatoi, I., Lautner, M., *et al.* (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nat Commun *8*, 15908. https://doi.org/10.1038/ncomms15908.

Zhao, W., Steinfeld, J.B., Liang, F., Chen, X., Maranon, D.G., Jian Ma, C., Kwon, Y., Rao, T., Wang, W., Sheng, C., *et al.* (2017). BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. Nature *550*, 360–365. https://doi.org/10.1038/nature24060.

Zhou, K., Gebala, M., Woods, D., Sundararajan, K., Edwards, G., Krzizike, D., Wereszczynski, J., Straight, A.F., and Luger, K. (2022). CENP-N promotes the compaction of centromeric chromatin. Nat Struct Mol Biol *29*, 403–413. https://doi.org/10.1038/s41594-022-00758-y.

Zhu, J., Cheng, K.C.L., and Yuen, K.W.Y. (2018a). Histone H3K9 and H4 Acetylations and Transcription Facilitate the Initial CENP-AHCP-3 Deposition and De Novo Centromere Establishment in Caenorhabditis elegans Artificial Chromosomes. Epigenetics Chromatin *11*, 16. https://doi.org/10.1186/s13072-018-0185-1.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011a). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184. https://doi.org/10.1038/nature10371.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011b). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184. https://doi.org/10.1038/nature10371.

Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K.H., Ke, E., Verma, S., Soroczynski, J., *et al.* (2018b). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol Cell *70*, 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023.

Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K.H., Ke, E., Verma, S., Soroczynski, J., *et al.* (2018c). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol Cell *70*, 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023.

Zilfou, J.T., and Lowe, S.W. (2009). Tumor suppressive functions of p53. Cold Spring Harb Perspect Biol *1*, a001883. https://doi.org/10.1101/cshperspect.a001883.

Zou, H., McGarry, T.J., Bernal, T., and Kirschner, M.W. (1999). Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science *285*, 418–422. https://doi.org/10.1126/science.285.5426.418.

(2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. The EMBO Journal *30*, 328–340. https://doi.org/10.1038/emboj.2010.329.

ANNEXE

ZBTB24 is a versatile transcription and DNA methylation factor with dual role at genes and centromeric satellite repeats

Giacomo GRILLO[1,#], Ekaterina BOYARCHUK[1], Pia MIHIC[1], Ivana IVKOVIC[1,5], Mathilde BERTRAND[2], Alice JOUNEAU[3], Thomas DAHLET[5], Michael WEBER[5], Guillaume VELASCO[1,*,&] and Claire FRANCASTEL[1,*,&]

Affiliations

[1] Université Paris Cité, CNRS, Epigenetics and Cell Fate, F-75013, Paris, France.

[2] Bioinformatics and Biostatistics Core Facility, iCONICS, IHU-A-ICM, Institut du Cerveau et de la Moelle épinière, Paris, France

[3]UMR BDR, INRA, ENVA, Université Paris Saclay, 78350, Jouy en Josas, France

[4]University of Strasbourg, Strasbourg, France. 2 CNRS UMR7242, Biotechnology and Cell Signaling, Illkirch, France.

[5] Sanofi Chimie, 30390, Aramon, France

* Correspondence to: Guillaume Velasco, e-mail: guillaume.velasco@u-paris.fr, phone: +33 1 57 27 89 32 and Claire Francastel, e-mail : claire.francastel@gmail.com

&, these authors cosupervised the work

Present address: Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada.

ABSTRACT

Since its discovery as an ICF syndrome-associated gene, ZBTB24 has emerged as a key player in DNA methylation, immunity and development. By extensively analyzing ZBTB24 genomic functions in ICF-relevant cellular models, we revealed here its multiple facets as a transcription factor with key roles in immune response-related genes expression and also in early embryonic development. Importantly, we provide here new insight into ZBTB24 function in DNA methylation and genomic stability. Indeed, we showed that ZBTB24 is recruited to centromeric satellite DNA where it is required for establishing the correct DNA methylation pattern through the recruitment of DNMT3B. Our results further reveal an essential role for ZBTB24 in regulating centromeric satellite repeats expression, more prominently of human alpha satellite RNA. Together, our data unraveled unprecedented functions of ZBTB24 as a key player in centromeric integrity by directly regulating DNA methylation and expression of centromeric satellite repeats.

INTRODUCTION

The study of diseases with abnormal DNA methylation (DNAme) landscapes emphasized the key role of this epigenetic mark in the maintenance of genomic stability (Ehrlich 2002, Robertson 2005; Pappalardo 2021). In particular, a growing body of evidence coming from the etiology of these diseases, including but not limited to cancer, highlighted the existence of a functional link between CpG methylation and centromere integrity, and its major contribution to genome stability (Scelfo 2019; Francastel 2019).

The congenital disorder called the ICF syndrome (Immunodeficiency with Centromeric instability and Facial anomalies syndrome; MIM no. 242860) is one of the most emblematic pathological contexts that underlines the connection between DNAme and centromere integrity. This syndrome is a very rare autosomal recessive disorder, genetically heterogeneous, which is mostly characterized by primary immunodeficiency (Maraschio 1988; Vukic 2019). Although displaying varying degrees of severity, clinical manifestations also include mild facial anomalies, intellectual disability, congenital malformations and developmental delay (Weemaes 2013). The ICF syndrome can be defined as a disease of DNAme and heterochromatin since DNA hypomethylation in patient cells markedly affects constitutive heterochromatin found at pericentromeric satellite repeats: Satellites type II (Sat II) of chromosomes 1 & 16 and, to a lesser extent, Satellites type III (Sat III) of chromosome 9. Hypomethylation of pericentromeric satellite DNA is an invariant hallmark of the ICF patient cells that leads to juxtacentromeric heterochromatin decondensation, chromosomal breaks and rearrangements in the vicinity of centromeres, a molecular feature used to establish diagnosis (Ehrlich 2006). In contrast, the hypomethylation of centromeric alpha satellite DNA is a heterogeneous trait among ICF patients reflecting the genetic heterogeneity of the syndrome.

The ICF syndrome is caused by mutations in the DNA methyltransferase 3B gene (*DNMT3B*, ICF1 OMIM#242860), zinc finger and BTB domain containing 24 gene (*ZBTB24*, ICF2 OMIM#614069), cell division cycle-associated 7 gene (*CDCA7*, ICF3 OMIM#616910), or helicase lymphoid specific gene (*HELLS*, ICF4 OMIM#616911). A comparative methylome analysis of ICF patients showcased a striking similarity between DNAme profiles of ICF2-4 patients, which clearly distinguished them from ICF1 patients (Velasco, Grillo 2018; Velasco 2021). Remarkably, ICF2-4 patient subtypes, which are caused by mutations in the three factors devoid of DNA methyltransferase activity, exhibit hypomethylation of centromeric alpha satellite in addition to the juxtacentromeric satellites common to all ICF subtypes (Jiang 2005, Velasco Grillo 2018). Thus, the ICF syndrome represents a unique pathological context to directly identify molecular actors and pathways linking DNAme to centromere integrity.

In contrast to DNMT3B, the functional connections between ZBTB24, CDCA7 and HELLS in DNAme are not straightforward but begin to be clarified. Indeed, ZBTB24 is a transcriptional activator of *CDCA7* gene (Wu 2016) and CDCA7 protein allows the recruitment of HELLS protein to chromatin and stimulates its remodeling activity (Jenness 2018). Hence, as supported by recent studies, CDCA7/HELLS complex contributes to maintain DNAme by facilitating the recruitment of the DNMT1/UHRF1 DNAme maintenance complex (Unoki 2020; Han 2020).

In light of these data, it appears that ZBTB24 harbor an indirect role in DNAme maintenance. Yet, its contribution to DNAme establishment, in particular at centromeric DNA repeats, and to the maintenance of centromere identity has not been properly addressed.

ZBTB24 belongs to the BTB-ZF family of transcription factors known for their essential role in immune system development (Siggs and Beutler, 2012). ZBTB24 is constitutively expressed, with the highest expression in B cells (de Greef 2011), suggesting its important role in the function of this cell lineage. ZBTB24 harbors a BTB domain which is thought to mediate protein oligomerization and/or dimerization, an AT-hook motif through which it can bind to AT-rich DNA sequences and eight Zinc-finger motifs (ZnF) in its C-terminal part, supporting that ZBTB24 is a transcription factor that binds directly to DNA (Siggs and Beutler, 2012). Indeed, as evidenced by two studies using an exogenous tagged version of the protein either in the HCT116 colon cancer cell line or in mouse embryonic stem cells (mESCs), ZBTB24 is a transcriptional regulator, mostly an activator, that localizes at gene regulatory regions and recognizes a specific DNA motif conserved in mice and humans (Thompson 2018, Ren 2019). Interestingly, its localization at chromocenters in the murine NIH3T3 cells suggests that ZBTB24 also plays an important role at pericentromeric heterochromatin (Nitta et al., 2013; Aktar 2019). As shown in zebrafish, ZBTB24 function would be essential for the DNAme of pericentromeric satellite DNA repeats and their transcriptional repression (Rajshekar 2018).

Considering its loss of function as the cause of the ICF2 syndrome, we explored here the genomic function of ZBTB24 in immune response, embryonic development, DNAme and centromere integrity. We addressed these multiple facets of ZBTB24 by using mouse and human cellular systems and models relevant for the physiopathological features of the ICF syndrome. Notably, we revealed the conserved presence of ZBTB24 at centromeric DNA repeats and provide new insights into its contribution to maintain centromeric integrity.

RESULTS

ZBTB24 pan-genomic binding sites and its transcriptional-dependent landscape in ICF2 patient cells.

To explore the role of ZBTB24 in human cells, and the consequences of its loss of function (LOF) on the transcriptome, we performed endogenous ZBTB24 chromatin immunoprecipitation-DNA sequencing (ChIP-seq) in normal human lymphoblastoid cell lines (LCLs from three healthy donors) and RNA-sequencing (RNA-seq) in LCLs derived from healthy donors and two ICF2 patients (**Fig. S1A and S1B**). We chose this cellular context because i) it is particularly relevant to study the pathophysiological mechanisms of the ICF syndrome and ii) the ZBTB24 protein is expressed at high levels in LCLs compared to other somatic cell types (de Greef et al. 2011).

ChIP-seq data revealed 183 genomic sites occupied by ZBTB24 in LCLs derived from healthy donors (HDs) (**Table S1**, merge peaks, q-value <0.05), which localized predominantly at transcriptional start site (TSS) or within proximal promoters (≤1Kb) and, to a lesser extent, at distal intergenic regions (**Fig.1A and 1B**). The functional annotation of ZBTB24 sites, according to the chromHMM genome segmentation ENCODE data in LCLs (GM12878), showed that it occupies active promoters (n=106, "Active_Promoter" category) and enhancers (n=32, "Strong_Enhancer" category) (**Fig.1A**). ZBTB24 peaks showed a highly significant enrichment of the 5'-CAGGACCT-3' DNA motif (E_value=2.8e-34) (**Fig. 1A**, lower panel), identical to the one recently identified for a FLAG-tagged ZBTB24 protein in the colon cancer cell line HCT116 (Thompson et al., NAR, 2018).

Based on the peak score (MACS2 score) at TSS, we identified the main targets genes of ZBTB24 which are *CDCA7, RIOK2, CNPY4, TAF6, ARID5B, CAMKMT, PREPL, RNF187, RPL32, CDC40, WASF1, OSTC, ACO2, PHF5A, and DTNB* (**Fig.1C**). Notably, for most of them, expression levels were decreased in LCLs from ICF2 patients compared to HDs, indicating that ZBTB24 is a transcriptional activator for this set of genes (**Fig.1C**). These results, combined with those previously reported (Thompson et al., NAR, 2018), strongly suggest that ZBTB24 preferentially targets the TSS of genes in a sequence specific manner to positively regulates their expression.

A more global analysis of the transcriptome of ICF2 LCLs identified 356 up- and 50 downregulated genes compared to normal LCLs (log2FC>0.5, FDR<0.01) with a significant enrichment of genes involved in the interferon response (adjusted p-value= 1.82e-03, **Table S2 and Fig.1D**) and more broadly in immune response pathways (**Fig. S1C**).

As a whole, these results showcase that ZBTB24 acts both directly and indirectly as transcriptional regulator at unique genomic regions in human LCLs, and in particular at innate immune response-related genes reflecting the pathological molecular mechanisms linked to ICF2 immune traits

ZBTB24 is a conserved transcriptional activator of genes in mice and humans.

The molecular and phenotypic features of ICF2 patients clearly pinpoint a critical role for ZBTB24 during development since nearly all patients exhibit developmental delay and growth retardation (Weemaes, Eur J Hum Genet, 2013). To better define this role, we generated a mouse model for the ICF2 syndrome that reproduced the BTB domain mutations reported in Lebanese siblings (Chourey et al. 2012) (**Fig. S1A and S2A-S2D**). Of note, the ZBTB24 protein sequence is highly conserved between mouse and human ranging from 76.9 to 96.9 % of homology for the structural domains (Aktar 2019).

We first derived and characterized mouse embryonic stem cells (mESCs) from wild-type (*Zbtb24*$^{wt/wt}$) and mutant (*Zbtb24*$^{mt/mt}$) blastocysts, isolated at 3.5 dpc (Czechanski, Nat Protoc., 2014), in 2i medium to maintain their stemness (**Fig.S2E-S2G**). We took advantage of the dynamic mESCs cell-state transition to model the first steps of cell differentiation occurring in early embryo. We performed ChIP-seq of ZBTB24 in *Zbtb24*$^{wt/wt}$ mESCs and RNA-seq in *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs maintained in 2i conditions (2i) and after 3 days of differentiation (d3). While the level of ZBTB24 protein was nearly the same in 2i and at d3 (**Fig.S2F**), we identified four times more ZBTB24 genomic binding sites in differentiated mESCs (191 sites, q-value <0.05) compared to mESCs in 2i (47 sites, q-value <0.05) (**Fig. 2A** and **Table S3-S4**). Nearly all the ZBTB24 sites identified in 2i were maintained in differentiated mESCs (44 over 47 peaks). At d3, ZBTB24 binding sites localized mainly at cis-regulatory elements, active promoters and enhancers (**Fig.2A and 2B**).

We next evaluated the conservation of the role of ZBTB24 as a transcriptional regulator in mice and humans (**Fig.1 &2**). Compared to human cells, we identified the same functional and chromatin features, an identical enriched DNA motif 5'-CAGGACCT-3'within ZBTB24 peaks in mESCs and also a conserved genomic distribution, *i.e*. a location of mouse ZBTB24 at promoters and at distal intergenic regions (**Fig.2A**), at TSS (21 peaks in 2i and 55 peaks at d3, **Fig.2C**) and enhancers, mostly at d3 (3 peaks in 2i compared to 35 peaks at d3) (**Fig.2B**). Noteworthy, most of the genes showing the highest ZBTB24 enrichment at TSS were identical to those identified in human LCLs (*Cdca7, Riok2, Cnpy4, Taf6, Arid5b, Camkmt, Prepl, Rnf187, Rpl32, Cdc40, Wasf1, Ostc, Axin2, Aco2, Phf5a and Dtnb*) (**Fig.1C and 2D**),

suggesting thus similar expression regulation in humans and mice. These target genes containing the ZBTB24 specific DNA motif exhibited a reduced expression level in *Zbtb24*$^{mt/mt}$ mESC, strongly supporting a conserved role of ZBTB24 as a transcriptional activator at this conserved set of unique genes (**Fig.2D**).

As a whole, these results reveal a selective and conserved role of ZBTB24 as a transcriptional activator at unique genes, both in humans and mice. These results also validated the use of our ICF2 mouse model to further decipher the conserved genomic function of ZBTB24 between mice and humans.

ZBTB24 is essential for mouse embryonic development.

Crosses of mice heterozygous for ZBTB24 mutation (*Zbtb24*$^{wt/mt}$) did not produce newborns homozygous for the mutation (*Zbtb24*$^{mt/mt}$), indicating the embryonic lethality of *Zbtb24*$^{mt/mt}$ embryos (**Table S5**). We isolated embryos at different developmental stages (6.5, 9.5- and 11.5-days post coitum (dpc)) and noticed that *Zbtb24*$^{mt/mt}$ embryos were growth retarded at all stages and presented developmental delay and morphological abnormalities (**Fig.2E**). These observations highlighted an essential role of ZBTB24 during early mouse development.

Remarkably, the top 5 GO terms of genes linked to ZBTB24 sites in differentiated mESCs, but not in 2i, matched strikingly with the developmental defects reported in the *Zbtb24*$^{mt/mt}$ embryos (**Fig. 2F** and **Table S6**). Indeed, these GO terms, namely pattern specification process, regionalization, epithelial tube morphogenesis, embryonic organ morphogenesis and stem cell population maintenance are child terms connected to the biological processes named Anatomical structure morphogenesis and Multicellular organismal process (**Fig. 2F)**. Nonetheless, and at this early step of differentiation, most of the genes belonging to these GO terms and categorized as developmental or bivalent (**Fig. 2D**, based on ENCODE3 data), were not primary targets of ZBTB24, according to its relatively weak enrichment and the absence of its preferred DNA binding motif. This could explain why we were not able to detect major transcriptional deregulation of this set of genes in our mESCs cellular model at d3 of differentiation (**Fig. 2D**).

Together, these results showed that ZBTB24 is essential for mouse embryonic development possibly through its ability to target and regulate some key developmental genes. Nonetheless, considering the central role of DNAme in mammalian development (Greenberg 2019) and the altered DNAme patterns that featured the ICF2 patients (Velasco&Grillo 2018),

we further explored the role of ZBTB24 in shaping DNAme patterns during early embryonic development.

ZBTB24 regulates CpG methylation mostly away from its genomic sites.

We have previously reported a role of ZBTB24 in DNAme maintenance (Thijssen, 2015), along with a widespread DNA hypomethylation reported in ICF2 patients (Velasco, Grillo et al. 2018). To address the implication of ZBTB24 in the establishment of DNAme, which occurs in the early phases of the mouse embryo development, we profiled the DNA methylome of *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs, in 2i and differentiated state (d3) using Reduced Representation Bisulfite Sequencing (RRBS). Indeed, the dynamic mESCs transition from 2i to d3 is recognized as a powerful model that mimics *in vitro* the embryonic window where DNAme profiles are established *de novo* (Ying et al., 2008, Atlasi and Stunnenberg, 2017, Walter et al. 2016). As expected, and thus validating our dynamic model of DNAme establishment, the DNAme profiles obtained by RRBS allowed us to clearly distinguish mESCs in 2i medium and at d3 with a gain of lowly and highly methylated sequences in the differentiated cells (**Fig.3A and 3B**).

In 2i medium and differentiated mESCs, *Zbtb24*$^{mt/mt}$ cells showed a distinct DNAme profile compared to *Zbtb24*$^{wt/wt}$ mESCs which supports a role of ZBTB24 to shape DNAme profiles in early development (**Fig.3A**). Surprisingly, we noticed a considerable loss of DNAme in *Zbtb24*$^{mt/mt}$ compared to *Zbtb24*$^{wt/wt}$ mESCs in 2i condition (10,974 hypomethylated differentially methylated regions or hypoDMR) (**Fig.3B and Table S7**), a state at which the level of DNAme is known to be already low in comparison to mESCs cultured in serum (Walter et al. 2016). Upon engagement of the differentiation process, the number of hypoDMR in *Zbtb24*$^{mt/mt}$ cells was greatly reduced compared to *Zbtb24*$^{wt/wt}$ cells (195 *vs* 10,974 hypoDMR), suggesting that the wave of DNAme establishment corrected most of the methylation defects identified in 2i state despite ZBTB24 LOF (**Fig.3B and Table S8**). HypoDMRs were mostly annotated to intergenic regions, far away from TSS (**Fig.3C**) and within retrotransposon elements LINE (Long Interspersed Nuclear Elements) and ERV (Endogenous Retroviruses) DNA sequences (**Fig.S3A-S3B and Table S7-S8**). However, the hypomethylation of these retrotransposon elements was not accompanied by their variation in expression (**Fig. S3C**).

We next wondered whether the methylation status of the genomic sites occupied by ZBTB24 was directly impacted in *Zbtb24*$^{mt/mt}$ cells. To this end, we computed the distance between ZBTB24 sites (peak summits) and the position of the DMRs identified in *Zbtb24*$^{mt/mt}$ mESCs. We applied the same analyses to human cells using our previous methylome profiling

in ICF2 patients (Velasco, Grillo et al 2018). In both cases, we found that hypoDMRs in ZBTB24 mutant cells were located far away from ZBTB24 binding sites both in mouse and human cells (**Fig.3D and supplementary Fig.S3D**). Notably, we identified a restricted number of hypermethylated (hyperMe) sites located, mostly detectable in human LCLs, that were closer to ZBTB24 genomic targets compared to the hypoMe ones. Interestingly, these hyperMe sites resided within the promoter region of *RNF187*, *CDC40*, *ZNF717*, and *ARID5B* whose expression levels were downregulated in ICF2 patient cells (**Fig.1C and Fig.S3D**). The hyperMe of CpG within the promoter of *ARID5B* and its downregulation is a conserved feature in *Zbtb24*$^{mt/mt}$ mESCs (**Fig.3D**).

Altogether, it seems that ZBTB24 is not directly implicated in shaping the DNAme landscape nearby its direct target genes, but rather far from its binding sites. However, our methylome shed new light on a potential role of ZBTB24 in the protection of some of its genomic targets against DNAme gain during development.

ZBTB24 targets centromeric DNA and negatively regulates cenRNA levels.

The identification of *ZBTB24* mutations as a genetic cause of the ICF syndrome naturally raised the question of its role in centromere DNAme and integrity. Given the presence of an AT-hook and zinc finger (ZnF) within ZBTB24 structural domains, we hypothesized that ZBTB24 was able to directly target centromeric satellite sequences. To test this hypothesis, we overcame the problem raised by the high sequence homology between satellite sequences and their repeated nature, by using the Repenrich2 method (Criscione, BMC genomics, 2014) to quantify the enrichment of centromeric satellite reads within the regions occupied by ZBTB24 (see Methods section). Of note, the genome organization of mouse and human centromeres are clearly distinct and not conserved. Mouse centromere DNA repeats are relatively homogeneous between all chromosomes and composed of AT-rich tandem repeats of 120 bp called minor satellite (MinSat), annotated as SYNREP_MM and CENSAT_MC in Repeatmasker database (**Fig.S4A and S4B**). Mouse pericentromeric DNA is composed of major satellite DNA repeats that are tandem repeats of 234 bp, annotated as GSAT_MM in Repeatmasker database. In contrast to mouse, the human centromeres are composed by Alpha satellite DNA, monomers of 171 bp arranged head-to-tail and organized into repeated arrays called higher order repeats (HORs) and are chromosome specific.

Based on our ZBTB24 ChIP-seq data obtained from *Zbtb24*$^{wt/wt}$ mESCs, we identified reads mapping to centromeric satellite DNA with a modest enrichment in differentiated ES cells (d3) compared to 2i condition (**Fig.4A**). These analyses were further confirmed by ChIP-qPCR

using primers specific for MinSat repeats (**Fig.4B**). Using the same method in human cells, we were able to identify ChIP-seq reads that mapped to centromeric DNA, namely ALR-Alpha satellite, and also to the pericentromeric satellite repeats HSAT4, HSATII, SAR, SATR1, and SATR2 (**Fig.4C**). Altogether, our results strongly indicate that ZBTB24 occupies centromeric satellite DNA in mouse and in human cells.

Because of its role as a transcription factor, we wondered whether ZBTB24 LOF impacted the level of centromeric transcripts (cenRNA) emanating from these centromeric regions. In mESCs grown in 2i, ZBTB24 LOF led to an increase of MinSat transcripts, although this trend was not observed during mESCs differentiation (**Fig.4D**). This discrepancy may be related to the mESCs differentiation culture conditions with the presence of serum, the heterogeneity of cellular states and epigenome remodeling during the first steps of mESCs differentiation (Marks *Cell* 2012, Kalkan *Development* 2017). In human cells, ICF2 patient cells showed a higher level of Alpha satellite RNA suggesting that ZBTB24 could be a negative regulator of Alpha satellite transcription (**Fig.4E**).

We then performed deeper analyses of our sequencing data of human centromere in order to investigate whether ZBTB24 occupied specific HORs and thus regulated the levels of transcripts generated from these HORs. We found that ZBTB24 is enriched at specific centromeric HORs with a marked preference for cen1_1, cen6, cen7_1, cen8, cen9, cen10_1, cen11_1, cen16_1, cen18_1, cen22_1 and cenX (**Fig. S5**). Interestingly, targeted HORs were mainly the HOR_1 which are known to be the active ones according to the enrichment of CENPA and binding sites of CENPB (CENP-Box) (Dumont et al, EMBO 2020), two essential proteins for centromere function. Accordingly, we identified the prevalence of the CENPB-Box DNA motif within the centromeric sites occupied by ZBTB24 (**Fig. S5**). Furthermore, we found an upregulation of transcripts emanating from HORs occupied by ZBTB24 (**Fig.4F**) in ICF2 patient cells. Among these HORs, we noticed the presence of HOR_1 of chromosomes 1 and 16 which are heavily affected by chromosomal anomalies in the vicinity of their centromeres in ICF lymphocytes cells, representing a hallmark of the syndrome (Tuck-Muller 2000).

Altogether, these results support a direct role of ZBTB24 as a cis-regulator of centromeric transcription, more precisely as a negative regulator of centromeric RNA expression.

ZBTB24 is required for DNMT3B-mediated DNAme establishment at centromeres.

Hypomethylation of centromeric satellite DNA is a molecular feature of ICF2 patient cells, and our data provide evidence for a crucial role of ZBTB24 in controlling the DNAme

status of these repeated DNA sequences (**Fig.S6A and S6B**) (de Greef 2011, Velasco & Grillo 2018). To assess whether such a ZBTB24 function is conserved in mice, we analyzed, using DNAme-sensitive restriction enzymes, the DNAme status of centromeric satellite sequences in *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ embryos. We performed Southern blot, using DNAme-sensitive restriction enzymes. We found that MinSat repeated sequences were hypomethylated in *Zbtb24*$^{mt/mt}$ embryos, as evidenced by the signal intensity profile depicting the presence of low molecular weight DNA fragments that are products of enzymatic digestion (**Fig.S6C**). Altogether, these data converged towards a conserved role of ZBTB24 in regulating DNAme of centromeric DNA repeats. Although there was evidence for ZBTB24 function in MinSat DNAme maintenance in somatic cells (Thijssen 2015), the question of its involvement in DNAme establishment at these satellite repeats was still unclear. To tackle this question, we followed by Southern blots the DNAme status of MinSat during the dynamic mESCs differentiation. In agreement with our RRBS analysis, we observed that MinSat were already strongly hypomethylated in *Zbtb24*$^{mt/mt}$ compared to *Zbtb24*$^{wt/wt}$ mESCs in 2i condition (**Fig.5A**). This hypomethylated state were maintained in *Zbtb24*$^{mt/mt}$ cells at d3 of differentiation, nevertheless less pronounced than in 2i state, reflecting an incomplete establishment of DNAme patterns at MinSat (**Fig.5A**).

We reasoned that the DNA hypomethylation in *Zbtb24*$^{mt/mt}$ differentiated cells may be a direct consequence of their lower DNAme level at baseline (2i). Therefore, in order to unravel in a more straightforward way the role of ZBTB24 in DNAme establishment at MinSat, we generated and characterized an auxin inducible degron (AID) knock-in mESCs lines (called Z_AID) for specifically triggering ZBTB24 degradation (**Fig.5B**). To validate the degradation of ZBTB24 upon auxin treatment in these Z-AID mESCs lines, we followed the levels of the ZBTB24 protein by western blot analysis, as well as the expression level of *Cdca7*, as it is one of the major ZBTB24 target genes, at different times points, as well as during their recovery dynamics after removal of auxin (**Fig. S7A and S7B**). In 2i condition, and after only 1h of auxin treatment, the level of ZBTB24 protein dropped drastically and *Cdca7* expression level significantly diminished 6h after auxin addition (**Fig. S7A and S7B**). Notably, after 24h of auxin treatment in 2i conditions, we could neither detect any change in DNAme levels, nor in the expression levels of MinSat in mESCs (**Fig.S7C and S7D**).

Next, we took advantage of the AID-degron system to either degrade ZBTB24 protein (IAA and Add On/AO conditions) or re-express ZBTB24 (Wash Out/WO) at specific time points during mESCs differentiation (**Fig.5C and 5D**). Concomitantly, we followed the dynamic of DNAme establishment at MinSat. Differentiating mESCs in the presence of auxin

failed to correctly establish DNAme pattern at MinSat which strongly supports the crucial role of ZBTB24 during this process (**Fig.5E**). Moreover, the induced degradation of ZBTB24 at day 1 of differentiation (AO_d1) was sufficient to alter MinSat DNAme profile. Rescuing ZBTB24 expression at day 1 of differentiation (WO_d1) was inefficient in fully fixing defects in DNAme establishment defects at MinSat. These results indicated that ZBTB24 is required all along the process of DNAme establishment at MinSat sequences.

DNA methyltransferase DNMT3B is not only crucial but is the sole DNMT to specifically catalyze the *de novo* methylation of MinSat sequences (Okano et al. 1999) (**Fig.S7E**). We thus wondered whether the presence of ZBTB24 at MinSat was required for targeting DNMT3B activity to centromeric satellite sequences. Whereas the expression level of DNMT3B protein is comparable between *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$, our ChIP-qPCR experiments clearly showed that there was a significant loss of DNMT3B recruitment at MinSat with ZBTB24 LOF **(Fig.5F**).

Therefore, our results demonstrate an essential role of ZBTB24 in guiding *de novo* DNAme at centromeric DNA repeats through DNMT3B recruitment.

DISCUSSION

ZBTB24 has long been envisioned as a ZBTB protein with an “obscure physiological function” (Siggs and Beutler, 2012), until the identification of mutations in *ZBTB24* as a causative factor of the ICF2 syndrome brought into light its important and diverse roles in immune system, centromere biology, DNAme and development (de Greef et al., 2011). By using relevant cellular systems and contexts to address these multiple facets of ZBTB24 functions, we unravel that ZBTB24 is transcriptional regulator implicated in immune response and embryonic development. Moreover, one major finding of our study is the presence of ZBTB24 at centromeres that we showed is conserved between mouse and human, and its requirement for proper establishment of DNAme patterns, through DNMT3B recruitment, at centromeric satellite DNA repeats. As a whole, and to our knowledge, our study provides novel insights of the contribution of ZBTB24 genomic functions into immune response, embryonic development and centromeric integrity.

Conserved and distinct biological functions of ZBTB24

We showed that ZBTB24 genomic sites in human LCLs and in mESCs have conserved (epi)genomic features : this transcription factor targets gene regulatory regions within active chromatin, mainly TSS of a common set of genes and some enhancers and acts as a transcriptional activator for a restricted number of genes marked by the presence of its DNA recognition motif bound by ZBTB24. Overall, the results presented here support a conserved role of ZBTB24 as a transcriptional regulator and are consistent with two recent studies exploring the (epi)genomic functions of ZBTB24 (Thompson 2018, Ren 2019). Compared to previous publications (Thompson 2018, Ren 2019, Helfricht 2020), we identified far less ZBTB24 genomic sites, as well as a lower or comparable number of deregulated genes in *ZBTB24* mutant cells. These discrepancies might be due to the differences in the investigated cellular contexts, in the behaviors of the endogenous versus exogenous tagged-ZBTB24 proteins in terms of genome-wide occupancy, or, else, in the knockdown approaches used for the identification of ZBTB24-regulated genes in constitutive mutants. Thus, compared to these previous studies that used cancer cell lines or non-dynamic cellular models, our work brings a comprehensive insight of the biological function of ZBTB24 that is more relevant for the physiopathological context of the ICF syndrome that encompass alteration of centromere integrity, impaired development and compromised immune function.

We reported here the first transcriptomic profiling of ICF2 patient-derived lymphocyte cells, which showcases deregulations of genes involved in signaling pathways linked to B cell activation and to the innate immune response, that is an activation of genes involved in the interferon response. This suggests the implication of ZBTB24 in cellular defense mechanisms, a function of the protein that is also supported by three other studies reporting an activation of interferon response-related genes either upon knockdown of *ZBTB24* in human B-cell line Raji and the colorectal cell line HCT116, or in a zebrafish ICF2 model (Liang 2016; Thompson 2018; Rajshekar 2018). Of knowledge, autoimmune signs have been reported for some ICF patients and it is worth mentioning that the activation of the interferon response pathways could account for this clinical trait (Sterlin 2016, von Bernuth 2014). Since these interferon response related genes are not directly targeted by ZBTB24, we conclude that their up-regulation is most likely an indirect consequence of ZBTB24 LOF.

In mice, we demonstrated that ZBTB24 is essential for embryonic development. This result is consistent with a previous study using a mouse model carrying mutations in *ZBTB24* obtained by the large genetic deletion of the BTB domain, AT-Hook and the first zinc fingers (Wu, 2016), showing by an embryonic lethality. Interestingly, we noticed that upon mESCs differentiation, ZBTB24 targets a set of developmental genes whose functions are strikingly related to the phenotype of *Zbtb24*$^{mt/mt}$ embryos. Although we did not detect their deregulated expression in the time-range of our differentiation of *Zbtb24*$^{mt/mt}$ mESCs, it is possible that ZBTB24 regulates their transcription *in vivo* and then could contribute to the phenotypes of *Zbtb24*$^{mt/mt}$ embryos.

In addition, in light of the pivotal role of DNAme in embryonic development (Greenberg 2019) and given the deep impact of the ZBTB24 LOF that we unravel on the DNA methylome (discussed below), it is likely that the developmental defects of *Zbtb24*$^{mt/mt}$ embryos are linked to their abnormal DNAme patterns. Nevertheless, the causal mechanisms underlying the embryonic lethality caused by ZBTB24 LOF need further clarification.

In contrast to mice, although a recent paper showed that a strong decrease in *ZBTB24* expression is associated with recurrent spontaneous abortion phenomena (Ruan, 2022), the viability of ZBTB24 mutations in ICF patients indicates either a discrepancy in the importance of ZBTB24 functions during embryonic development or the existence of compensatory mechanisms that could specifically take place only in humans. Interestingly, an elegant study using the zebrafish as a model for the ICF2 reported viable *Zbtb24* homozygous mutant animals with phenotypes that are much more reminiscent of human patients than mouse models. Indeed, the ZBTB24 mutant zebrafish shows small stature, facial anomalies and lethality in adulthood

(Rajshekar, 2018). Overall, at different degrees of severity depending on the species context, ZBTB24 LOF leads to significant developmental alterations, strongly supporting a conserved role of this protein during early developmental steps, and in line with the developmental aspect of the syndrome

Role of ZBTB24 in shaping DNA methylation patterns

By mimicking "in-dish" the establishment of DNAme patterns, we were able to show a modest contribution of ZBTB24 in this genome-wide process. The genome of *Zbtb24*$^{mt/mt}$ mESCs cultured in 2i was remarkably hypomethylated compared to their wild-type counterpart, supporting a main role of ZBTB24 in genome-wide DNAme maintenance rather than in its establishment. This observation is in line with previous studies showcasing progressive loss of DNAme at (peri)centromeric DNA repeats in HEK293 cells upon *ZBTB24* knock-out (Unoki, 2019) or between larval and adult stages in *Zbtb24* mutant zebrafish (Rajshekar, 2018) or else DNAme defects in HCT116 cells upon *ZBTB24* knockdown (Thompson 2018).

A possible model that could account for a role of ZBTB24 in DNAme maintenance would be its ability to read and then protect the DNAme mark as it is the case for ZFP57, another zinc-finger protein implicated in DNAme imprinting (Quenneville 2011, Shi 2019). While some other members of the BTB-ZF TF family are DNAme readers (Filion 2006, Daniel 2002, Prokhortchouk 2001), the lack of the CG dinucleotide within ZBTB24 DNA motif combined with the conserved chromatin localization of ZBTB24 in cells devoid of DNAme (Nitta, 2013) does not support this model.

Rather, we propose that ZBTB24 ensures faithful DNAme maintenance at unique genomic loci and at some transposable elements through its ability to regulate CDCA7/HELLS complex (Wu 2016, Jenness 2018, Velasco, Grillo 2018), the latter being required for the loading of the DNAme maintenance machinery complex DNMT1/UHRF1 onto chromatin (Unoki 2020, Unoki 2021). As such, in the future, it would be interesting to investigate whether DNAme defects in ICF2 patients worsen over time and whether this phenomenon correlates with an exacerbation or the onset of clinical signs of the disease as it was reported and suggested in a case report (von Bernuth 2014).

We also clarified the role of ZBTB24 in DNAme of centromeric satellite sequences. On its own, the hypomethylated status of MinSat repeats in *Zbtb24*$^{mt/mt}$ embryos is not sufficient to conclude whether DNAme establishment or maintenance or both were perturbed upon the loss of ZBTB24 function. Although we previously provided evidence for a role of ZBTB24 in DNAme maintenance of centromeric repeats in mouse embryonic fibroblasts (Thijssen 2015),

its contribution in DNAme establishment was still unknown. By taking advantage of two genetic mESCs knock-in models (ICF2 mouse model and auxin inducible degron or AID), we provide strong evidence that ZBTB24 is required for the proper establishment of DNAme at MinSat repeats. This function is carried out in a direct manner, through the ability of ZBTB24 to bind to MinSat and recruit DNMT3B. In support of our findings, a functional link between ZBTB24 and DNMT3B in the DNAme of gene bodies has been recently reported with a direct interaction between ZBTB24 and DNMT3B proteins being evidenced in HEK 293 (Thompson 2018). However, and because this interaction has not been confirmed in mESCs (Hardikar 2020), we cannot rule out the possibility that the recruitment of DNMT3B to MinSat is undirect and rather depends on chromatin context known to dictate the binding of DNMTs (Baubec 2015, Weinberg 2019, Greenberg 2019). In sum, we propose that the hypomethylation events resulting from ZBTB24 LOF are due to the loading defect of distinct chromatin complexes, comprising DNMT enzymes, involved in both the establishment and maintenance of DNAme.

Role of ZBTB24 as a guardian of centromeric integrity

Importantly, we present here a novel facet of ZBTB24 function at centromeres. Our data provide strong evidence for a conserved and direct role for ZBTB24 at murine and human centromeric satellite sequences. Since these A/T rich sequences lack the DNA recognition motif of ZBTB24, read through its 6th and 7th zinc finger motifs (Ren 2019), we hypothesize that the binding of ZBTB24 requires the involvement of the AT-hook domain, which is known to interact with the minor groove of A/T rich DNA. However, as the AT-hook domain alone has a low affinity for DNA, it is likely that the affinity of ZBTB24 for centromeric satellite sequences is potentiated by the joint action of both AT-hook and ZnF motifs (Rodriguez 2015). In that sense, the presence of both motifs has been shown to be important for the localization of ZBTB24 to (peri)centromeric heterochromatin (Aktar 2019). In light of the intended role of DNAme in centromeric integrity (Scelfo 2019) and acknowledging the fundamental role of centromeres in ensuring faithful transmission of the genetic information (Barra, 2018), ZBTB24 emerges as a guardian of chromosomal integrity and genomic stability through its role in DNAme. Loss of DNAme upon ZBTB24 LOF is associated with an increased level of centromeric RNA, and in a more prominent manner in human cells, which constitute a threat for genomic integrity. In fact, it has been documented that the loss of DNAme at centromeric sequences can lead to their abnormal transcription and the formation of DNA:RNA hybrids and R loops structures, which are known to altogether cause replication stress, double strand breaks and genomic instability (Unoki 2019, Giunta 2021, Garcia- Muse & Aguilera, 2019).

Interestingly, our thorough analysis of the ChIP-seq data of ZBTB24 in human LCLs highlighted a peculiar enrichment of the protein at specific centromeric HORs, mainly the active one according to the enrichment of CENPA and binding sites of CENPB (CENP-Box) (Dumont et al, EMBO 2020). Among the most enriched in ZBTB24 and deregulated HORs in ICF2 cells, we strikingly identified HOR_1 of the chromosomes 1 and 16 which are affected by chromosomal anomalies in ICF patients. Whether or not the deregulation of specific HORs is due to different degree of DNAme loss in ICF cells is a very exciting question that should be addressed in a near future thanks to the achievement of complete sequence sequencing of the human genome, including centromeric satellite arrays, by the Telomere to Telomere (T2T) consortium (Nurk 2022).

Initially, the identification of three new ICFs factors lacking DNA methyltransferase activity, namely ZBTB24, CDCA7 and HELLS, raised the obvious question of their functional link with the DNAme machinery and centromere biology. Likely because of the known role of HELLS as "a guardian of heterochromatin at repeat elements" in mouse (Muegge 2005), most of the studies designed so far focused on the role of the CDCA7/HELLS complex (Jenness 2018, Unoki 2019, Unoki 2020, Hardikar 2020). Our work now provides a novel and previously unconsidered dimension of ZBTB24 role in maintaining genomic integrity. Combined with our previous findings showing an interaction between ZBTB24, CDCA7 and HELLS proteins (Velasco, Grillo 2018), we propose that ZBTB24, CDCA7 and HELLS form multi-partite complexes that would act as a gatekeeper of the integrity of the (epi)genetic information of centromeres ensuring the maintenance of genomic stability. Hence, our study paves the way for new investigations on the underlying molecular mechanisms of the functional interplay between ZBTB24/CDCA7/HELLS, DNAme of satellite repeats, centromeres transcription and organization.

METHODS

ICF2 mouse model

Mice mutant for *Zbtb24* were generated at the Institut Clinique de la Souris (Illkirch, France, http://www.ics-mci.fr) by homologous recombination using standard gene targeting techniques. The targeting construct was designed to replace exon 2 by mutated sequences (deletion of two nucleotides TA), thus mimicking the human mutations of ICF2 patients of a Lebanese family described in (Chouery et al. 2012) carrying the deletion c.396_397delTA (p.His132Glnfs*19). Homozygous mice for *Zbtb24* mutation (*Zbtb24*$^{mt/mt}$) were generated by intercrossing heterozygous mice (*Zbtb24*$^{wt/mt}$) in a C57BL/6J background. Mice were mated overnight and the presence of a vaginal plug the following morning was designated as E0.5. Generation and use of these mice has been approved by the Comité d'éthique en Expérimentation Animale Buffon (Paris, France) and received the agreement number 5314. The genotyping was performed using genomic DNA extracted either from cultured cells, yolk sac or tail of embryos with the REDExtract-N-Amp Tissue PCR kit (Merck) following the manufacturer's instructions.

Cell lines and culture conditions

Mouse ES cells (mESCs) were derived from WT, *Zbtb24* and *Dnmt3b* (*Dnmt3b*$^{m3/m3}$ Velasco, Hube et al., PNAS, 2010) mutant blastocysts as described in (Czechanski et al., 2014) at 3.5 dpc. All mESCs used in this study were maintained in the absence of feeder cells, on plates coated with 1x Laminin (Merck) as described in (Tosolini and Jouneau, 2016), in serum-free 2i medium containing Dulbecco's modified Eagle's medium (DMEM)-F12 Glutamax (Thermo Fisher Scientific) supplemented with 50% Neurobasal medium (Thermo Fisher Scientific), 1x N2 and B27 Supplements (both Thermo Fisher Scientific), 1% penicillin/streptomycin (Merck), 1x non-essential amino acids (PAA), 100nM 2-mercaptoethanol (Thermo Fisher Scientific), 3µM CHIR99201, 1µM PD032552 inhibitors (both from Cell Guidance Systems) and 1000 U/mL units of Leukemia Inhibitory Factor (Merck). Cells were passaged with 0.05% accutase (Thermo Fisher Scientific) every two days and maintained at 37ºC with 8% CO2 in a humidified chamber. For *in vitro* differentiation, 10^6 mESCs were grown in 10 cm bacteriological Petri dish with differentiation medium containing 25% Advanced DMEM, 25% DMEM/F12 GlutaMAX, 50% Neurobasal medium, 1% of penicilline/streptavidin, 1x β-mercatoethanol and 11% of knock out serum, at 37ºC with 5% CO2 (all products from Thermo Fisher Scientific).

The differentiation medium was changed every day and embryoid bodies were collected by centrifugation for 4 minutes at 700 rpm and resuspended in 10ml of fresh medium.
Lymphoblastoid cell lines (LCLs) from healthy donors and ICF2 patients were cultured in RPMI 1640 supplemented with 15% FCS and antibiotics (Thermo Fisher Scientific) as described in (Velasco, Grillo et al. HMG 2018).

Generation of the Auxin-Induced-Degradation system of ZBTB24 in mESCs

*Os*TIR1-9xMyc sequence linked via cleavable P2A sequence was introduced by CRISPR/Cas9-mediated homologous recombination (guide RNAs (gRNAs) are listed in **Table S9**, all plasmids were a kind gift from by Dr. Alexei Arnaoutov, NICHD, NIH) at the C-terminus of the essential ubiquitously-expressed RCC1 protein (regulator of chromosome condensation 1) into Zbtb24$^{wt/mt}$ mESCs. Transfections were all performed using Lipofectamin2000 (Thermo Fisher Scientific). ESCs were selected using 7 μg/ml blasticidin and single cell clones were isolated by limiting dilution and the *os*Tir1 bi-allelic insertion was validated by PCR, Sanger sequencing, and western blotting. Sequences encoding a minimal functional auxin-inducible degron (mAID) and HA-tag were introduced by CRISPR/Cas9-mediated homologous recombination to the N-terminus of ZBTB24 protein (gRNAs are listed in **Table S9** were cloned in pSpCas9(BB)-2A-GFP (PX458) (Addgene #48138), the plasmids containing HA tag, a micro-AID tag (71–114 amino-acid), and puromycin resistance were kind gift from Dr Alexey Arnaoutov, PMID: 32917881). Following hygromycin selection at 150 μg/ml, single-cell clones were isolated by limiting dilution. These clones were then validated for the mAID insertion on WT allele, followed by the characterization of the AID system.

RNA extraction and RT-qPCR analysis

Total RNAs from cell lines were isolated using TRI Reagent® (Merck) according to manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNA-free Kit (Thermo Fisher Scientific). Reverse transcription was carried out using 500ng DNA-free RNA, 50μM random hexamers (Thermo Fisher Scientific), 10U of RNase inhibitor (New England Biolabs) and 200U of ProtoScript II Reverse Transcriptase (New England Biolabs). Complementary DNA reactions were used as templates for real-time quantitative PCR reactions (qPCR). qPCR was performed using the SensiFAST SYBR NoROX Kit (Bioline) supplemented with 0.2μM of specific primer pairs (**Table S10**) and run on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers

and analyzed by the comparative CT(ΔCT) method using U6 RNA as an invariant RNA. Each data shown in RT-qPCR analysis is the result of at least three independent experiments.

Protein extraction and western blot analysis

Cells were washed in PBS and collected. After centrifugation, cell pellets were resuspended in RIPA buffer (300 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0, complete protease inhibitor cocktail (Merck)). Samples were incubated for 30min on ice and were then sonicated with a Bioruptor (Diagenode) at 4°C. Cell debris were removed by centrifugation for 10 min at 16,000g and supernatants were collected. For Western blot analysis, proteins were resolved by SDS–PAGE, transferred onto PVDF membrane (Thermo Fischer Scientific), incubated with the appropriate primary antibody (**Table S11**) and then with horseradish peroxidase-conjugated secondary antibodies. Protein detection was performed with ECL reagents (Thermo Fisher Scientific).

Immunofluorescence

Cells were grown on glass coverslips until reaching 60% of confluence. Cells were then fixed with 4% paraformaldehyde in PBS (Electron Microscopy Sciences). After permeabilization with 0.1% Triton X-100 in PBS, samples were blocked in 2% donkey serum (Jackson Immunoresearch) in PBS followed by incubation with primary antibodies (**Table S11**) in blocking buffer overnight at 4°C. After incubation with secondary antibodies conjugated to Alexa-Fluor 488 or 594 (Jackson ImmunoResearch), coverslips were mounted in Vectashield medium containing 2µg/ml DAPI (Vector Laboratories). Image acquisition was performed at room temperature on a fluorescence microscope (Axioplan 2; Zeiss) with a Plan-Neofluar 100X/1.3 NA oil immersion objective (Zeiss) using a digital cooled camera (CoolSNAP fx; Photometrics) and METAMORPH 7.04 software (Roper Scientific, Trenton, NJ). Images presented correspond to one focal plane.

Chromatin Immunoprecipitation (ChIP)

Mouse ESCs and LCLs were crosslinked with 1% formaldehyde for 10min at room temperature (RT) and the reaction was quenched by adding glycine at a final concentration of 0.125M for 5min at RT. Fixed cells were washed and adherent cells were harvested with phosphate-buffered saline (PBS). Cell nuclei were isolated using a cell lysis buffer (5mM Pipes pH8.0, 85mM KCl, 0.5% NP40, Complete protease inhibitor cocktail (Merck)) for 20min, at 4°C. Nuclei were then pelleted by centrifugation and lysed in a nuclei lysis buffer (50mM Tris HCl

pH8.1, 150mM NaCl, 10mM EDTA, 1% SDS, Complete protease inhibitor cocktail (Merck)) for 20min, 4°C. Chromatin was subjected to sonication with a Bioruptor (Diagenode) yielding genomic DNA fragments with a bulk size of 150- 300 bp. Supernatant was diluted 10 times in IP dilution buffer (16.7 mM Tris pH8, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X- 100, 0.01% SDS, Complete protease inhibitor cocktail (Merck)). Immunoprecipitations were performed overnight at 4°C on 40µg of chromatin with antibodies pre-bound to Dynabeads protein G (Thermo Fisher Scientific). The antibodies used are listed in **Table S11**. The Beads were then washed once with Low salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8, 150 mM NaCl), once with High salt buffer (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris pH 8; 500 mM NaCl), once with LiCl wash buffer (10 mM Tris pH 8.0, 1% Na-deoxycholate, 1% NP-40, 250 mM LiCl, 1 mM EDTA) and twice with TE (10mM Tris pH 8.0, 1mM EDTA pH 8.0). The elution of chromatin was performed in Elution buffer (50mM Tris pH 8.0, 10mM EDTA pH 8.0, 1% SDS) at 65°C for 45min and crosslink was reversed O/N at 65°C after addition of NaCl to 0.2M final concentration. The eluted material was digested with 40µg of Proteinase K (Thermo Fisher Scientific) at 65°C for 2h. DNA was purified by phenol–chlorofom extraction and ethanol precipitation, resuspended in TE. Quantitative PCRs were performed using the SensiFAST SYBR NoROX Kit (Bioline) and analyzed on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche). ChIP-qPCR results were normalized on input signal (% input). As a negative control, an additional mock ChIP using IgG was systematically added to each experiment. Sequences of primers are listed in **Table S10**.

ChIP-seq experiment and analysis

ChIP-seq libraries were prepared either from ZBTB24 ChIP and input samples in LCLs derived from healthy subjects (n=3) or in *Zbtb24*$^{wt/wt}$ (n=2), *Zbtb24*$^{mt/mt}$ (n=2) mESCs cultured in 2i conditions and at d3 of differentiation, using the MicroPlex Library Preparation Kit v2 (Diagenode) according to the manufacturer's protocol. In brief, 1 to 25 ng were used as input material. After adapters ligation, fragments were amplified with Illumina primers for 10 cycles. Libraries were purified using AMPure XP beads protocol (Beckman Coulter, Indianapolis, IN) and quantified using the Qubit fluorometer (Thermo Fisher Scientific). Libraries' size distribution was assessed using the Bioanalyzer High Sensitivity DNA chip (Agilent Technologies). Libraries were normalized and pooled at 2nM and spike-in PhiX Control v3 (Illumina) was added. Sequencing was done using a NextSeq 500 instrument (Illumina) in paired-end mode (2x75 cycles). After sequencing, a primary analysis based on AOZAN

software (ENS, Paris, France) (1) was applied to demultiplex and control the quality of the raw data (based on bcl2fastq v2.20 and FastQC modules / v0.11.5). Quality of raw data has been evaluated with FastQC [1]. Poor quality sequences have been trimmed or removed with Trimmomatic [2] software to retain only good quality paired reads. Reads were mapped to the reference genomes (mm10/hg38) using Bowtie2 v2.3.4.3 [3] with the option -k 2. Then, peaks were called with Macs2 v2.2.4 [4] using the input as control, an FDR of 0.01 and keeping all the duplicates. Peaks from multiple replicates were combined by taking regions covered by at least 2 replicates. Peak annotation was performed with ChiPseeker v1.18.0 [5] and Bedtools intersect v2.28.0 was used to annotate peaks with repeated elements obtained on the repeatmasker website [6], enhancers obtained on FANTOM5[7] and genome segmentation from ENCODE data (ENCODE Project Consortium, Nature, 2012 and Gorkin et al. Nature 2020). Finally, motif detection was performed with RSAT [8]. The Gene ontology enrichment analysis was performed using the clusterProfiler R package. The significant GO categories were identified with Benjamini-Hochberg adjusted p-value (p.adjust) ≤ 0.05.

Bioinformatic analysis of ZBTB24 sites in DNA repeats and HORs

ZBTB24 occupancy on DNA repeats was determined using RepEnrich2 (Criscione 2014) with the following protocol available in https://github.com/nerettilab/RepEnrich2. Briefly, the reads were first uniquely aligned with Bowtie2 (-k 1) on the mm10 and hg38 genomes. Single and multiple reads were then separated and multiple reads were assigned to a FASTQ file. In parallel, annotation of the repeats was performed using RepeatMasker annotations of the repeated elements. Overlap of single reads with repeated elements was tested. Multiple reads were aligned against the repeat assembly using Bowtie2. Finally, the number of reads aligning against families of repeated elements was determined by combining the counts of uniquely mapping reads and multi-mapping reads (fractional counts). Read counts were normalized by the library size and the input. Read mapping to HOR arrays was carried out described in (Dumont et al, 2020). The human ChIP_seq and RNA-seq FASTQ files were filtered against a library containing 6,455,351 sequences of 18 mers alpha satellite seed sequences (Miga Journal of Heredity 2017). Reads that contained at least one 18mers were retrieved and mapped with Bowtie2 on a reference composed of 64 centromeric HOR array consensus sequences (Nechemia-Arbely *et al*, 2019).

RNA-seq experiment and analysis

RNA-seq was performed on independent clones of mESCs (2 clones of *Zbtb24*$^{wt/wt}$ cells and of *Zbtb24*$^{mt/mt}$ cells) and on LCLs derived from healthy (n=2) and ICF2 (n=2) subjects. RNA quality was verified using the Bioanalyzer RNA 6000 Nano kit (Agilent). The libraries were prepared following the TruSeq Stranded Total RNA protocol from Illumina, starting from 1 µg of high-quality total RNA (RIN > 7). Paired end (2 × 75) sequencing was performed on an Illumina NextSeq 500 platform, demultiplexing and quality control were performed with the AOZAN software (ENS, Paris, France) [1] (based on bcl2fastq v2.20 and FastQC modules / v0.11.5). The quality of raw data has been evaluated with FastQC [1]. Poor quality sequences have been trimmed or removed with Trimmomatic [2] software to retain only good quality paired reads. Star v2.5.3a [3] has been used to align reads on reference genome (mm10/hg38) using standard options. Quantification of gene and isoform abundances has been done with rsem 1.2.28, prior to normalisation on library size with DESEq2 [5] bioconductor package. Finally, differential analysis has been conducted with edgeR [6] bioconductor package. Multiple hypothesis adjusted p-values were calculated with the Benjamini-Hochberg [7] procedure to control FDR.

DNA methylation of satellite repeats by Southern blot

Genomic DNA was extracted from mESCs and LCLs using the Monarch Genomic DNA Purification Kit (New England Biolabs) according to the manufacturer's instructions. Genomic DNA from mESCs and LCLs (500 ng) was digested respectively with 20 units of HpaII or HhaI enzymes (New England Biolabs) for 16 h to analyze the DNA methylation patterns of centromeric minor and alpha satellite repeats, respectively. Genomic DNA from *Dnmt1/Dnmt3a/Dnmt3b* triple-knockout (TKO) mESCs (Tsumura, 2006) and from *Dnmt1* knock-out (*Dnmt1-/-*) mouse embryonic fibroblasts (Lande-Diner, 2007) were used as controls of low CpG methylated cellular contexts. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+ membranes (GE Healthcare) in 20 × SSC. After ultraviolet crosslink, the membranes were pre-hybridized in 6 × SSC, 5 × Denhardt and 0.1% SDS and then hybridized with ^{32}P- labelled minor satellite (5′-ACATTCGTTGGAAACGGGATTTGTAGAACAGTGTATATCAATGAGTTACAATGAG AAACAT-3′) or alpha satellite (5'-ATGTGTGCATTCAACTCACAGAGTTGAAC-3'). The membranes were washed 3 times in 6X SSC and 0.1%SDS at 37°C and then subjected to phosphorimaging using FLA 7000 phosphorimager (Fuji). The quantification of radiolabeled

signals intensity was determined using ImageJ software and normalized by the sum of the total signal per lane.

Methylome analysis by RRBS

RRBS libraries were prepared by MspI digestion from 50 to 100 ng genomic DNA44 and sequenced in paired-end 2 × 75 bp on an Illumina HiSeq4000 at Integragen SA (Evry, France). We trimmed RRBS reads to remove low-quality bases with Trim Galore v0.4.2 and aligned reads to the mm10 genome with BSMAP v2.74 (parameters -v 2 -w 100 -r 1 -x 400 -m 30 -D C-CGG -n 1). We calculated methylation scores using methratio.py in BSMAP v2.74 (parameters -z -u -g) and filtered CpGs covered by a minimum of eight reads. DMRs were identified using eDMR from the methylKit R package with a minimum of seven differentially methylated CpGs, a difference in methylation >20% and a q-value <0.001.

Visualization of data

For ChIP-seq data visualization, the plots of the distribution of reads density across all transcription start sites RefSeq genes were generated using deepTools (https://deeptools.readthedocs.io/). Bar plots, dot plots, Hierarchical clustering representations, Heatmaps, violin plots, were generated using functions in R software environment for statistical computing. Samples clustering and heatmap visualizations were performed using the ComplexHeatmap and the gplots **R** packages (https://www.bioconductor.org/). For hierarchical clustering, the hclust function were used with “euclidean” and “ward.D2” method parameters respectively for distance and agglomeration. Basic stacked bar plots and Violin were build using the ggplot2 R package.

Data availability

The ChIP-seq, RNA-seq and RRBS data generated in this study are available in the NCBI Gene Expression Omnibus (GEO) under the accession number XXXXXXX.

The following published datasets were used: GSE111689 (Thompson 2018) and GSE95040 (Velasco, Grillo 2018)

REFERENCES

ACKNOWLEDGEMENTS

The authors would like to thank Drs Florence Larminat, Marie-Cécile Gaillard, Daniele Fachinetti, Catalina Salinas for insightful discussions about this work, Dr Riccardo Gamba for support and advises in centromeric data analysis, Laure Ferry and the Epigenomic Core Facility of the UMR7216 for help with pyrosequencing, Franck Letourneur and Juliette Harmoune of the sequencing platform core facility Genom'IC, Damien Ulveling and Justine Guegan of the bioinformatics and biostatistics core facility ICONICS for help and support with sequencing data analysis. We acknowledge all the patients and their family members for their participation in this study.

FUNDING

This study was funded by the Agence Nationale pour la Recherche (grant ANR-19-CE12-0022), the Fondation Maladies Rares (GenOmics of Rare Diseases, call 2017), Fondation Jérôme Lejeune. GG was supported by the French Ministry of Research and Fondation ARC (http://www.fondation-arc.org/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

G.G., E.B., P.M., I.I., A.J., T.D. and G.V., performed the experiments. G.G., M.B, T.D., M.W. and G.V. performed bioinformatics analysis of sequencing data. G.G., G.V. and C.F. designed the study and supervised the experiments. G.G., E.B., C.F. and G.V. wrote the manuscript. G.V. and C.F. acquired the fundings.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

FIGURE LEGENDS

Figure 1. **ZBTB24 acts as transcriptional regulator modulating innate immune response-related genes in patients-derived LCLs.**

(A) Stacked bar charts showing the distribution of ZBTB24 sites (183 peaks) in lymphoblastoid cell lines (LCLs) across genomic (left) and functional annotations (right). The functional annotation was based on the chromatin states established in GM12878 LCLs (ChromHMM track from ENCODE/Broad Institute). Predicted binding motif of ZBTB24 was identified using Regulatory Sequence Analysis Tools (RSAT, http://rsat.sb-roscoff.fr/) based on ChIP-seq data performed in LCLs. Distal_int, Distal intergenic; Prom, promoter. (B) Profile of the average of ChIP_seq signal of ZBTB24, *i.e* the enrichment of ZBTB24 relative to chromatin input (substraction mode, input from IP of bigWig files) in LCLs on genomic loci defined as 3kb upstream and downstream of annotated Transcriptional Start Site (TSS, NCBI RefSeq) of genes (upper panel). The lower panel shows the heatmap of read density around the same genomic loci. High density levels are shown in blue and low-density levels in red. (C) Unsupervised clustering of genes TSS linked to ZBTB24 sites. The heatmap (IPZ) presents the enrichment of ZBTB24 at each TSS based on the peak score (mean log2MACS2_score). The heatmap "Express." shows the gene expression data, presented as a mean of the row Z-score of each gene in LCLs derived from the blood of healthy (CTL) and ICF2 subjects (ICF2). ZBTB24 genomic sites common to LCLs and HCT116 cell line (GSE111689, Thompson et al 2018) are notified in pink. The Motif columns refer to the occurrence of the predicted ZBTB24 binding motif occurrence (Counts). The color code attributed to each category is depicted next right to the figure. (D) Gene Set Enrichment Analysis (GSEA) of deregulated genes in LCLs derived from the blood of ICF2 patients (pD and pV) (left panel). The expression status of genes belonging to the "Interferon_response" HALLMARK category is presented as the row Z-score and illustrated by a heatmap (right panel).

Figure 2. ZBTB24 targets developmental genes in mESCs and is essential for embryonic development.

(A) Stacked bar chart of the distribution of ZBTB24 sites across genomic features. Peaks overlapping between two independent mESC clones derived from distinct embryos cultured in 2i conditions (IPZ_2i) or at day3 of differentiation (IPZ_d3) were merged and annotated to the mm10 genome. The predicted ZBTB24 binding motif identified by Regulatory Sequence Analysis Tools (RSAT, http://rsat.sb-roscoff.fr/) based on ChIP-seq data performed in 2i condition (IPZ_2i) or at day3 of differentiation (IPZ_d3) is reported in the plot below. (B)

Stacked bar chart of the distribution of ZBTB24 sites across functional regions of the mouse genome (mm10). ZBTB24 peaks overlapping between two independent ES clones cultured in 2i conditions (IPZ_2i) or at day3 of differentiation (IPZ_d3) were merged and annotated based on ChiP-seq assays of 8 histone modifications (laboratory of Bing Ren) as part of the ENCODE 3 Consortium. Pr: promoter; Enh: enhancer; Int,Inter: intron, intergenic; Ex: exon; Hc: heterochromatin. (C) Average of ChIP_seq signal profile of ZBTB24, *i.e* the enrichment of ZBTB24 relative to chromatin input (substraction mode, input from IP) in mESCs cultured in 2i conditions (IPZ_2i) or at day3 of differentiation (IPZ_d3) on genomic loci defined as 3kb upstream and downstream of annotated Transcriptional Start Site (TSS, NCBI RefSeq) of genes. (D) Unsupervised clustering of genes linked to ZBTB24 merge peaks at TSS. The heatmap (IPZ) presents the enrichment of ZBTB24 at TSS of genes based on the peak score (mean log2MACS2_score), in two independent ES clones cultured in 2i conditions (IPZ_2i) or at day3 of differentiation (IPZ_d3). For each locus, the gene status ("Bivalent" or "Dev.": developmental genes) and the predicted ZBTB24 binding motif occurrence are indicated, and a color code is attributed to each category next right to the figure. The gene expression data ("Expression") are presented as the mean of the row Z-score in the two independent ES clones WT and mutant (mt) for ZBTB24 cultured in 2i condition (ZWT_2i, Zmt_2i) or at day3 of differentiation (ZWT_d3, Zmt_d3). (E) Pictures of *Zbtb24*$^{wt/wt}$, *Zbtb24*$^{wt/mt}$ and *Zbtb24*$^{mt/mt}$ embryos at post-implantation stages 6.5, 9.5- and 11.5-days post coitum (dpc). The arrows point to the head and limb buds' malformations reported in *Zbtb24*$^{mt/mt}$ embryos at 11.5 dpc. (F) Gene Ontology of genes linked to ZBTB24 sites in mESCs at day3 of differentiation (GO IPZ_d3). The Dot plot represents the enrichment in biological processes. GeneRatio on the x axis refers to the number of genes linked to ZBTB24 sites divided by the number of genes annotated in the corresponding process. The size and colors of the dots indicate gene counts and the adjusted p-value for the GO pathway listed on the y axis, respectively. Only the top 5 GO terms with the highest degree of enrichment are represented (GO:0007389 pattern specification process, GO:0003002 regionalization, GO:0060562 epithelial tube morphogenesis, GO:0048562 embryonic organ morphogenesis, GO:0019827 stem cell population maintenance). These GO terms are child terms connected to the biological processes GO:0009653 Anatomical structure morphogenesis and GO:0032501 Multicellular organismal process.

Figure 3. ZBTB24 moderately contributes to the establishment of DNAme landscape.

(A) Dendrogram showing the clustering of the two mESCs clones wild-type (WT) and mutant (mt) for ZBTB24, cultured in 2i medium (2i) or at day3 of differentiation (d3), based on their

DNAme profiles performed by RRBS. (B) Violin plot of DNAme level in the two ES clones wild-type (WT) and mutant (mt) for ZBTB24, cultured in 2i medium or at day3 of differentiation. The number and the status (Hypo: Hypomethylation; Hyper: Hypermethylation) of Differentially Methylated Regions (DMR) identified in ZBTB24 mutant compared to WT mESCs is shown next above to the plot. (C) Dot plots showing the distance in nucleotides between DMR and TSS of genes in mESCs cultured in 2i medium or at day 3 of differentiation. (D) Dot plots presenting the distance in base pairs (bp) between ZBTB24 peak summits in WT mESCs (IPZ_2i and IPZ_d3) and DMR (hypomethylated in red, hypermethylated in blue) in WT mESCs cultured in 2i medium and at day3 of differentiation (IPZ_2i and IPZ_d3). Dots which correspond to genes whose change in their level of expression is correlated to DMR position are indicated by gene symbol.

Figure 4. ZBTB24 occupies centromeric satellite DNA and regulates cenRNA level.

(A) Bar chart showing the enrichment of normalized read counts in mouse satellite repeats based on ZBTB24 occupancy in mESCs at day3 of differentiation compared to 2i condition. The satellite repeats which show an enrichment of ZBTB24 occupancy in differentiated mESCs are shown in red and labeled. (B) ChIP assays performed on chromatin prepared from WT(Z_WT) and ZBTB24 mutant (Z_mt) mESCs cultured in 2i medium or at day 3 of differentiation, using ZBTB24 antibodies or IgG, and followed by quantitative PCR with primers specific of Minor satellite sequences. Error bars represent standard error (n=3 independent experiments), p : p-value; two-tailed t-test (C) Bar chart showing the enrichment of normalized read counts in human satellite repeats based on ChIP-seq of ZBTB24 and of IgG in LCLs. The human satellite repeats which show an enrichment of ZBTB24 occupancy are shown in red and labelled. (D) Expression levels of Minor satellite RNA in WT and *Zbtb24* mutant (mt) mESCs cells clones (1 and 2) cultured in 2i medium or at day 3 of differentiation were assessed by RT-qPCR, presented as a relative expression to U6 snRNA levels. Error bars represent standard error (n=4 independent experiments), p : p-value; two-tailed t-test . (E) Expression levels of ALR_Alpha satellite RNA in LCLs derived from healthy (CTL1 and CTL2; WT for ZBTB24) and ICF2 subjects (pD, PV; mutant for ZBTB24) were assessed by RT-qPCR, presented as a relative expression to U6 snRNA levels. Error bars represent standard error (n=3 independent experiments), p : p-value; two-tailed t-test (F) Unsupervised clustering of centromeric HORs occupied by ZBTB24. The heatmap (IP) presents the peak score computed from the mean difference between normalized reads count in ChIP_seq of ZBTB24 (IP_Z) and of IgG (IP_Ig) at the considered HOR in LCLs derived from the blood of three

healthy subjects. The bar chart on the right shows the log2 difference of the peak score between IP_Z and IP_Ig. The heatmap (Expression) shows the expression data of each HOR computed from RNA-seq data of two independent LCL clones WT (CTL) and mutant (ICF2) for ZBTB24 derived from the blood of subjects. The bar chart on the right presents the log2 fold change (FC) of HORs expression based on RNA-seq data, comparing the ICF2 to the WT subjects.

Figure 5. ZBTB24 is required for DNAme establishment of MinSat through DNMT3B recruitment.

(A) Density plots of Southern blot band profiles obtained from the digestion of genomic DNA from *Zbtb24*$^{wt/wt}$ and *Zbtb24*$^{mt/mt}$ mESCs cultured in 2i medium (top) and after 3 days of differentiation (bottom) by the methylation-sensitive enzyme HpaII. The plot represents the profile of the radiolabeled signals intensity detected on each lane of the Southern blot after hybridization by radiolabeled minor satellite probes and normalized by the sum of the total signal per lane (from the top toward the bottom of the lane). (B) Schematic representation of CRISPR/Cas9-based tagging of WT allele of ZBTB24 protein at N-terminus in heterozygous ZBTB24 $^{wt/mt}$ mESCs, expressing TIR1 from Rcc1 locus, to generate AID_ZTBT24 mESCs. Mutated *Zbtb24* allele with the position of the loxP site and the mutation is also shown. (C) Experimental scheme for panels D and E. AID_ZBTB24 cells were either differentiated in ADFNK medium without LIF for 2 days (Z_AID) or treated with auxin (IAA) for the indicated time. For Z_ IAA and Z_IAA_wash-out, auxin was added 6h prior the differentiation induction. (D) Western blot analysis of ZBTB24 and γ-Tubulin protein levels in AID_ZBTB24 cells differentiated according to the schematic in C. Non-differentiated WT (Z_WT) and mutant (Z_mt) mESCs were used as controls, n.s. - non-specific band. (E) Density plot of Southern blot band profiles obtained from the digestion of genomic DNA of mESCs AID_ZBTB24 lines treated (Z_AID_IAA) or not (Z_AID) with auxin (IAA), cultured in 2i and at day 1 or day 2 of differentiation (d1, d2) by the methylation-sensitive enzyme HpaII. During the two days of mESCs differentiation, several conditions were used: auxin was either added 6h before the beginning of the differentiation protocol (Z_AID_IAA), at d1 (Z_AID_AO_d1) or remove at d1 (Z_AID_WO_d1). The plot represents the profile of the radiolabeled signals intensity detected on each lane of the Southern blot after hybridization by radiolabeled minor satellite probes and normalized by the sum of the total signal per lane (from the top toward the bottom of the lane). (F) ChIP assays performed on chromatin prepared from *Zbtb24*$^{wt/wt}$, *Zbtb24*$^{mt/mt}$ and DNMT3B mutant (*Dnmt3b*$^{mt/mt}$) mESCs cultured in 2i medium or at day 3 of differentiation, using anti-ZBTB24, anti-DNMT3B or IgG antibodies, followed by quantitative PCR with

primers specific of minor satellite sequence. Error bars represent standard error (n=3 independent experiments), p : p-value; two-tailed t-test. Western blot analysis showing the expression level of DNMT3B and **γ**-Tubulin proteins in *Zbtb24*$^{wt/wt}$, *Zbtb24*$^{mt/mt}$ and *Dnmt3b*$^{mt/mt}$ mESCs cultured in 2i medium and at day3 of differentiation. Molecular weight is indicated on the left of the panel.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. **Analysis of gene expression alterations in ICF2 LCLs.**

(A) Wild-type and mutant version of the human *ZBTB24* gene. The nucleotides mutated in ICF2 patients are indicated in red. (B) Western blot analysis of ZBTB24 protein levels in ICF2 patients carrying the mutation shown in panel A; **γ**-Tubulin is a loading control. (C) Gene ontology terms corresponding to deregulated genes in ICF2 patients' LCL compared to healthy donors.

Figure S2. Analysis of the role of ZBTB24 during embryonic development.

(A) Wild-type and mutant version of the mouse *Zbtb24* gene. The nucleotides mutated in the ICF2 mouse model presented in this study indicated in red. (B) Schematic map of the *Zbtb24* mutant allele showing the position and the type of mutation. (C) Representation of ZBTB24 functional protein domains: BTB (Broad-Complex, Tramtrack and Bric a brac), AT-hook, and the eight C2H2 zinc finger domains. The position of the introduced mutation is indicated on the scheme. (D) Schematic representation of the WT and mutated *Zbtb24* allele. The genetic background, the position of the loxP sites, the mutation, and the primers used for the genotyping of embryos are indicated. (E) Immunofluorescence showing the single cell expression of the pluripotency factors NANOG and OCT3/4 in $Zbtb24^{wt/wt}$ (Z_WT) and $Zbtb24^{mt/mt}$ (Z_mt) mESCs cultured in 2i medium. (F) Western blot analysis of ZBTB24 protein levels in 2i condition and after three days of differentiation (d3) in $Zbtb24^{wt/wt}$ and $Zbtb24^{mt/mt}$ mESCs; **γ**-Tubulin is a loading control. (G) Bar plots showing the relative expression levels of the pluripotency markers REX1 and KLF4, assessed by RT-qPCR and normalized to U6 small nuclear RNA. The graphs represent the mean of three experiments using WT and mutant (Mut) mESCs for ZBTB24 cultured in 2i medium and at d3 of differentiation. Error bars represent the standard error of the mean (SEM). n.s : non significant according to two-tailed t-test.

Figure S3. ZBTB24 LOF impacts DNAme of repetitive DNA elements without impacting their transcriptional status.

(A) Bar charts representing the number of the identified DMRs within LINE and LTR repeat class elements. (B) Bar charts representing the number of the identified DMRs within LINE and LTR repeat families in mESCs. (C) Bar plots showing the relative expression levels of the LINE1A and LINE1T elements, assessed by RT-qPCR and normalized to U6 small nuclear RNA. The graphs represent the mean of three experiments using WT and mutant (Mut) mESCs for ZBTB24 cultured in 2i medium and at d3 of differentiation. Error bars represent the standard

error of the mean (SEM). (D) Dot plot showing the distance in nucleotides between ZBTB24 peak summit identified in LCLs and Differentially Methylated Probes (DMP : $|\Delta\beta| \geq 0.2$, adjPval≤ 0.05, hypomethylated sites in red, hypermethylated sites in blue) identified by HM450K bead chip technology in human ICF2 patient blood cells compared to healthy subjects. Dots which correspond to genes whose change in their level of expression is correlated to DMP are indicated by gene symbol.

Figure S4. Repeatmasker annotation for minor satellite repeats.

(A) Phylogenetic tree of satellite sequences in the mouse including the minor and major satellite sequences. Satellite sequences were retrieved from Dfam (https://dfam.org). Minor (120bp) and Major (234bp) satellite repeat unit sequences were respectively taken from Wong and Rattner (1988) and From Manuelidis (1982). (B) Comparison of Minor satellite sequence with CENSAT_MC and SYNREP_MM using MUltiple Sequence Comparison by Log- Expectation (MUSCLE).

Figure S5. ZBTB24 is enriched at centromeric HORs.

Box plot showing the enrichment of ZBTB24 relative to IgG within each centromeric HOR (Top). The x axis shows the read density in ChIP-seq of ZBTB24 (IP-Z) and of IgG (IP_IgG) within each HOR. (Bottom) Predicted motif using Regulatory Sequence Analysis Tools (RSAT, http://rsat.sb-roscoff.fr/) based on ZBTB24 peaks within Cen_HORs is indicated below the box plot.

Figure S6. Centromeric satellite repeats are hypomethylated in human cells and mouse embryos mutant for ZBTB24.

(A) Violin plot showing the normalized beta values (BMIQ normalization) of CpG probes of the Illumina Infinium methylation 450K (HM450K) bead chip technology, annotated in LINE, LTR and Satellite repeats on autosomes (n=22,821 probes). The data were generated from DNA of human primary blood cells derived from healthy subjects and ICF2 patients (Velasco, Grillo et al. 2018). (B) Density plot of Southern blot band profiles of alpha satellite repeats obtained from the digestion of genomic DNA of LCLs from healthy (CTL1 and CTL2) and ICF2 subjects (pD and pV) by the methylation-sensitive enzyme HpyCH4IV. The plot represents the profile of the radiolabeled signals intensity detected on each lane of the Southern blot after hybridization of radiolabeled alpha satellite probes and normalized by the sum of the total signal

per lane (from the top toward the bottom of the lane). (C) Density plot of Southern blot band profiles obtained from the digestion of genomic DNA of WT and mutant embryos for ZBTB24 (mt) at 11.5 dpc by the methylation-sensitive enzyme HpaII. The plot represents the profile of the radiolabeled signals intensity detected on each lane of the Southern blot after hybridization by radiolabeled minor satellite probes and normalized by the sum of the total signal per lane (from the top toward the bottom of the lane). The DNA from Mouse Embryonic Fibroblasts knocked-out for DNMT1 enzyme was used as a control of hypomethylated signals.

Figure S7. Generation, characterization and validation of the AID-degron system for ZBTB24.

(A) Western blot analysis of ZBTB24 and γ-Tubulin protein levels in 2i condition in WT (Z_WT), mutant (Z_mt), and AID_ZBTB24 (Z_AID) cell lines. AID_ZBTB24 cells were treated with Auxin (IAA) for the indicated time (upper panel) or treated with IAA for one day (0 h) followed by release into auxin-free 2i medium for the indicated times (wash-out, bottom panel). NT – non-treated with IAA, n.s. - non-specific band. (B) Expression level of *Cdca7* in mESCs AID_ZBTB24 lines cultured in 2i medium, treated or not (Z_AID_NT) for 6H (Z_AID_6h) or 24H (Z_AID_24h) with auxin. RT-qPCR experiments were performed and presented as a relative expression to U6 snRNA levels. Error bars represent standard error (n=3 independent experiments), p : p-value; two-tailed t-test. (C) Density plot of southern blot band profiles obtained from the digestion of genomic DNA of mESC AID_ZBTB24 lines treated or not (NT) for 24 h or 48 h with auxin by the methylation-sensitive enzyme HpaII. The plot represents the profile of the radiolabeled signals intensity detected on each lane of the Southern blot after hybridization by radiolabeled minor satellite probes and normalized by the sum of the total signal per lane (from the top toward the bottom of the lane). The DNA from mESCs triple knocked-out (TKO) for the DNMTs enzymes was used as a control of hypomethylated signals. (D) Density plots of Southern blot band profiles obtained from the digestion of genomic DNA from *Dnmt3b*$^{mt/mt}$ mESCs (D3Bmt) cultured in 2i medium or after 3 days of differentiation by the methylation-sensitive enzyme HpaII. The plot represents the profile of the radiolabeled signals intensity detected on each lane of the southern blot after hybridization by radiolabeled minor satellite probes and normalized by the sum of the total signal per lane (from the top toward the bottom of the lane).(E)Expression level of Minor satellite repeats in mESCs AID_ZBTB24 lines cultured in 2i medium, treated or not (Z_AID_NT) for 6H (Z_AID_6h) or 24H (Z_AID_24h) with auxin. RT-qPCR experiments were performed and presented as a

relative expression to U6 snRNA levels. Error bars represent standard error (n=4 independent experiments), p : p-value; two-tailed t-test.

Figure 1

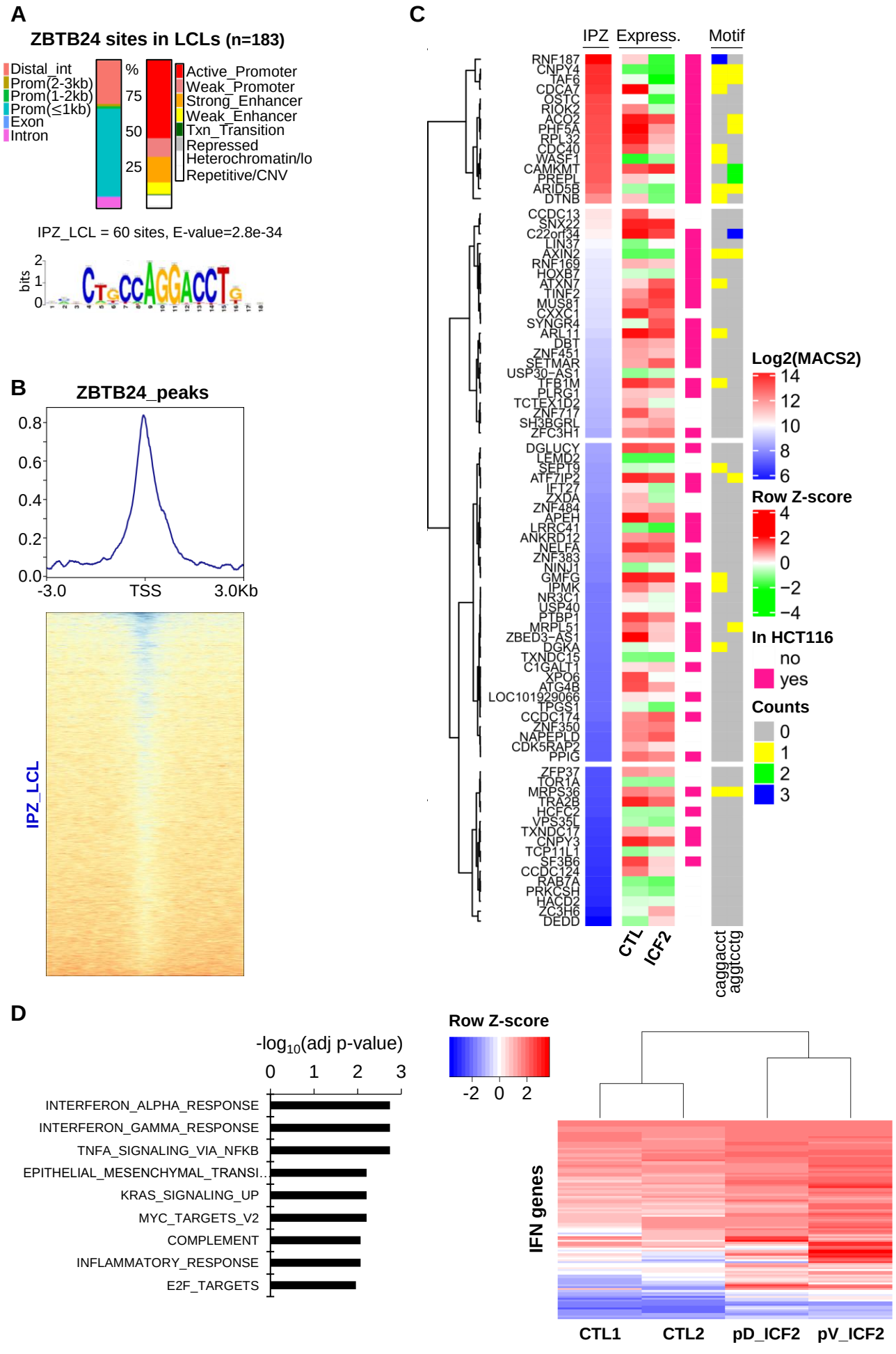

A
ZBTB24 sites in LCLs (n=183)
Distal_int
Prom(2-3kb)
Prom(1-2kb)
Prom(≤1kb)
Exon
Intron
%
75
50
25
Active_Promoter
Weak_Promoter
Strong_Enhancer
Weak_Enhancer
Txn_Transition
Repressed
Heterochromatin/lo
Repetitive/CNV
IPZ_LCL = 60 sites, E-value=2.8e-34
bits
B
ZBTB24_peaks
0.8
0.6
0.4
0.2
0.0
-3.0
TSS
3.0Kb
IPZ_LCL
C
IPZ
Express.
Motif
RNF187
CNPY4
TAF6
CDCA7
OSTC
RIOK2
ACO2
PHF5A
RPL32
CDC40
WASF1
CAMKMT
PREPL
ARID5B
DTNB
CCDC13
SNX22
C22orf34
LIN37
AXIN2
RNF169
HOXB7
ATXN7
TINF2
MUS81
CXXC1
SYNGR4
ARL11
DBT
ZNF451
SETMAR
USP30-AS1
TFB1M
PLRG1
TCTEX1D2
ZNF717
SH3BGRL
ZFC3H1
DGLUCY
LEMD2
SEPT9
ATF7IP2
IFT27
ZXDA
ZNF484
APEH
LRRC41
ANKRD12
NELFA
ZNF383
NINJ1
GMFG
IPMK
NR3C1
USP40
PTBP1
MRPL51
ZBED3-AS1
DGKA
TXNDC15
C1GALT1
XPO6
ATG4B
LOC101929066
TPGS1
CCDC174
ZNF350
NAPEPLD
CDK5RAP2
PPIG
ZFP37
TOR1A
MRPS36
TRA2B
HCFC2
VPS35L
TXNDC17
CNPY3
TCP11L1
SF3B6
CCDC124
RAB7A
PRKCSH
HACD2
ZC3H6
DEDD
CTL
ICF2
caggacct
aggtcctg
Log2(MACS2)
14
12
10
8
6
Row Z-score
4
2
0
−2
−4
In HCT116
no
yes
Counts
0
1
2
3
D
-log10(adj p-value)
0
1
2
3
INTERFERON_ALPHA_RESPONSE
INTERFERON_GAMMA_RESPONSE
TNFA_SIGNALING_VIA_NFKB
EPITHELIAL_MESENCHYMAL_TRANSI..
KRAS_SIGNALING_UP
MYC_TARGETS_V2
COMPLEMENT
INFLAMMATORY_RESPONSE
E2F_TARGETS
Row Z-score
-2
0
2
IFN genes
CTL1
CTL2
pD_ICF2
pV_ICF2

Figure 2

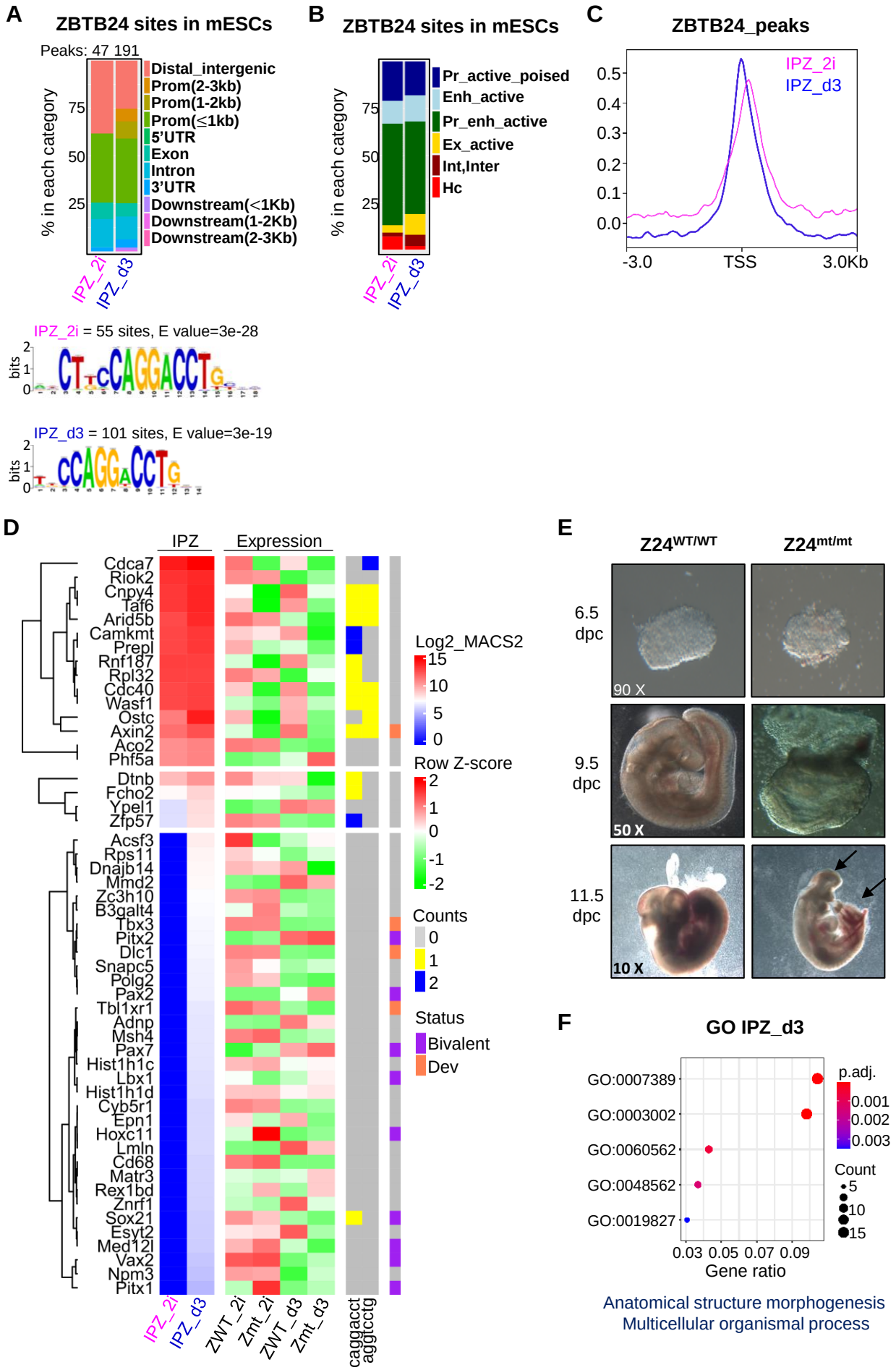

A
ZBTB24 sites in mESCs
Peaks: 47 191
% in each category
Distal_intergenic
Prom(2-3kb)
Prom(1-2kb)
Prom(≤1kb)
5'UTR
Exon
Intron
3'UTR
Downstream(<1Kb)
Downstream(1-2Kb)
Downstream(2-3Kb)
IPZ_2i
IPZ_d3
B
ZBTB24 sites in mESCs
Pr_active_poised
Enh_active
Pr_enh_active
Ex_active
Int,Inter
Hc
C
ZBTB24_peaks
IPZ_2i
IPZ_d3
-3.0
TSS
3.0Kb
IPZ_2i = 55 sites, E value=3e-28
IPZ_d3 = 101 sites, E value=3e-19
bits
D
IPZ
Expression
Cdca7
Riok2
Cnpy4
Taf6
Arid5b
Camkmt
Prepl
Rnf187
Rpl32
Cdc40
Wasf1
Ostc
Axin2
Aco2
Phf5a
Dtnb
Fcho2
Ypel1
Zfp57
Acsf3
Rps11
Dnajb14
Mmd2
Zc3h10
B3galt4
Tbx3
Pitx2
Dlc1
Snapc5
Polg2
Pax2
Tbl1xr1
Adnp
Msh4
Pax7
Hist1h1c
Lbx1
Hist1h1d
Cyb5r1
Epn1
Hoxc11
Lmln
Cd68
Matr3
Rex1bd
Znrf1
Sox21
Esyt2
Med12l
Vax2
Npm3
Pitx1
IPZ_2i
IPZ_d3
ZWT_2i
Zmt_2i
ZWT_d3
Zmt_d3
caggacct
aggtcctg
Log2_MACS2
Row Z-score
Counts
Status
Bivalent
Dev
E
Z24WT/WT
Z24mt/mt
6.5 dpc
9.5 dpc
11.5 dpc
90 X
50 X
10 X
F
GO IPZ_d3
GO:0007389
GO:0003002
GO:0060562
GO:0048562
GO:0019827
p.adj.
Count
Gene ratio
Anatomical structure morphogenesis
Multicellular organismal process

Figure 3

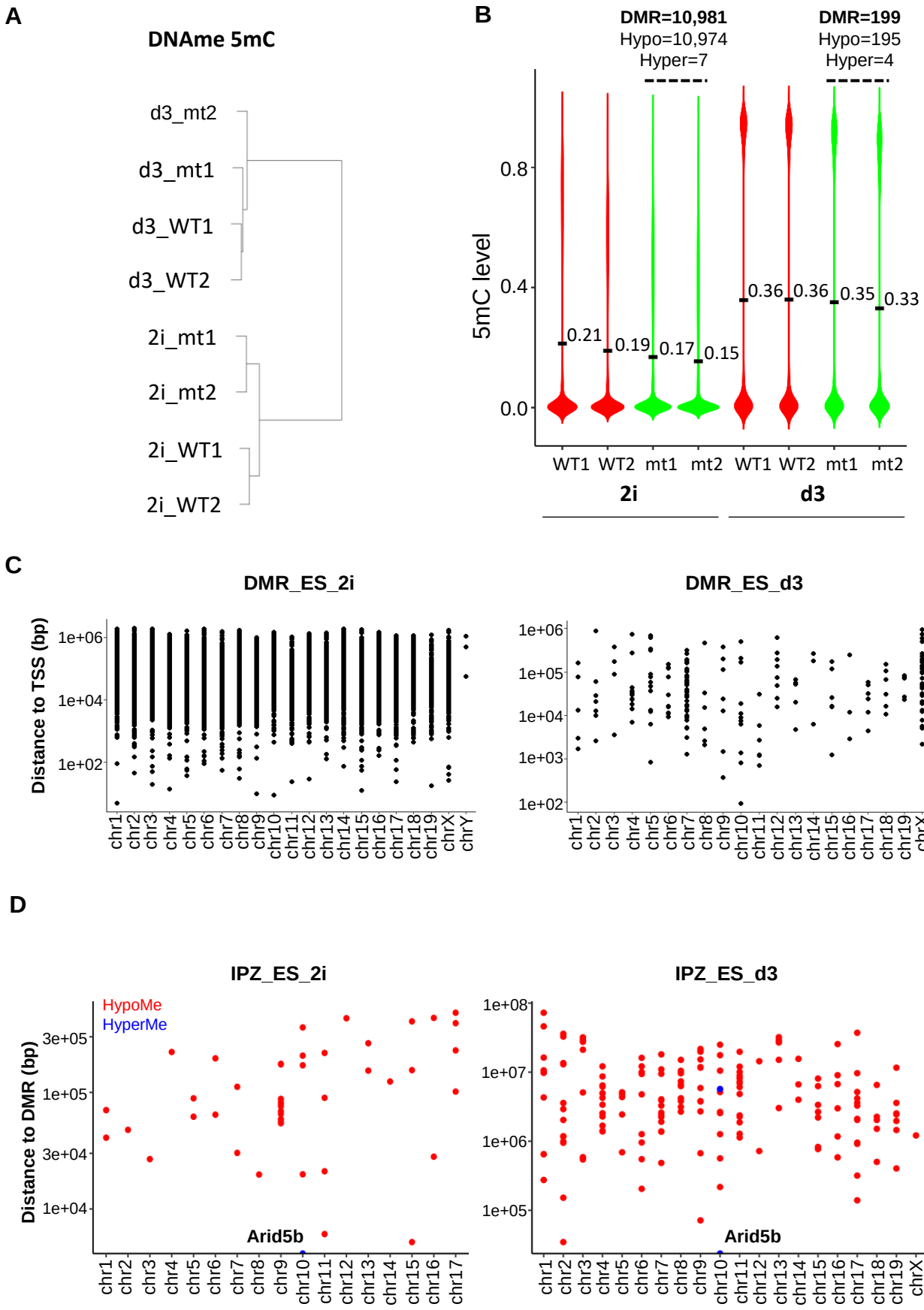
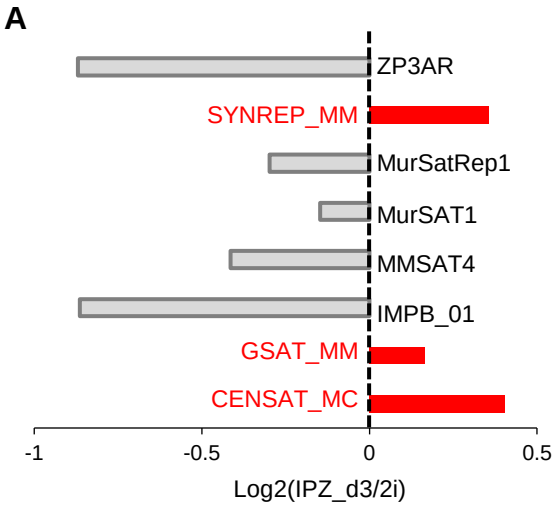
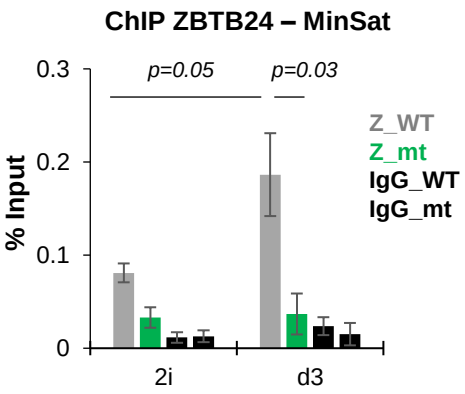
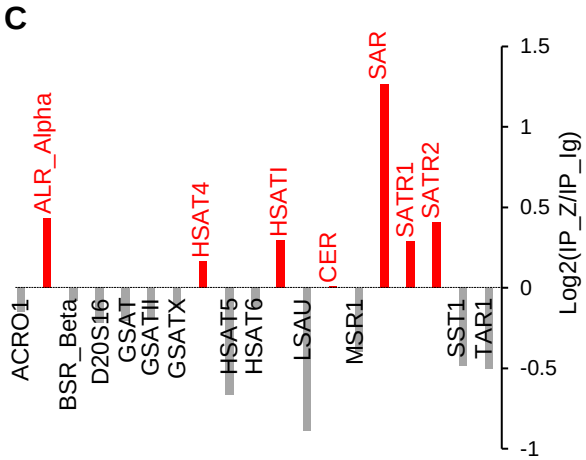
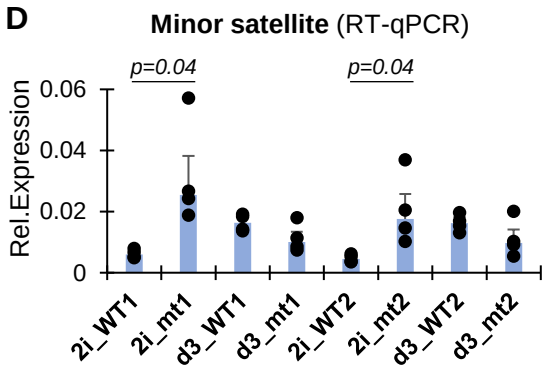

A
DNAme 5mC
d3_mt2
d3_mt1
d3_WT1
d3_WT2
2i_mt1
2i_mt2
2i_WT1
2i_WT2
B
DMR=10,981
Hypo=10,974
Hyper=7
DMR=199
Hypo=195
Hyper=4
5mC level
0.8
0.4
0.0
0.21
0.19
0.17
0.15
0.36
0.36
0.35
0.33
WT1 WT2 mt1 mt2 WT1 WT2 mt1 mt2
2i
d3
C
DMR_ES_2i
DMR_ES_d3
Distance to TSS (bp)
1e+06
1e+04
1e+02
1e+05
1e+03
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chrX chrY
D
IPZ_ES_2i
IPZ_ES_d3
HypoMe
HyperMe
Distance to DMR (bp)
3e+05
1e+05
3e+04
1e+04
1e+08
1e+07
1e+06
Arid5b
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chrX

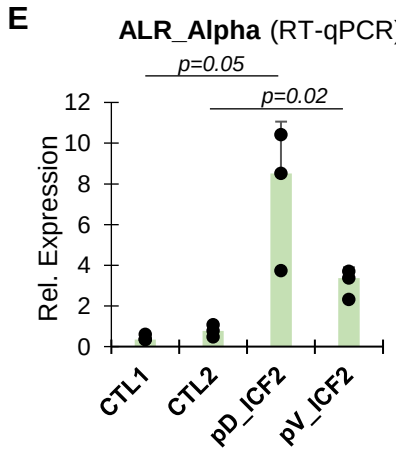
Figure 4


A
ZP3AR
SYNREP_MM
MurSatRep1
MurSAT1
MMSAT4
IMPB_01
GSAT_MM
CENSAT_MC
-1
-0.5
0
0.5
Log2(IPZ_d3/2i)


B
ChIP ZBTB24 – MinSat
p=0.05
p=0.03
% Input
0.3
0.2
0.1
0
Z_WT
Z_mt
IgG_WT
IgG_mt
2i
d3


C
ALR_Alpha
HSAT4
HSATI
CER
SAR
SATR1
SATR2
ACRO1
BSR_Beta
D20S16
GSAT
GSATII
GSATX
HSAT5
HSAT6
LSAU
MSR1
SST1
TAR1
1.5
1
0.5
0
-0.5
-1
Log2(IP_Z/IP_Ig)


D
Minor satellite (RT-qPCR)
p=0.04
p=0.04
Rel.Expression
0.06
0.04
0.02
0
2i_WT1
2i_mt1
d3_WT1
d3_mt1
2i_WT2
2i_mt2
d3_WT2
d3_mt2


E
ALR_Alpha (RT-qPCR)
p=0.05
p=0.02
Rel. Expression
12
10
8
6
4
2
0
CTL1
CTL2
pD_ICF2
pV_ICF2

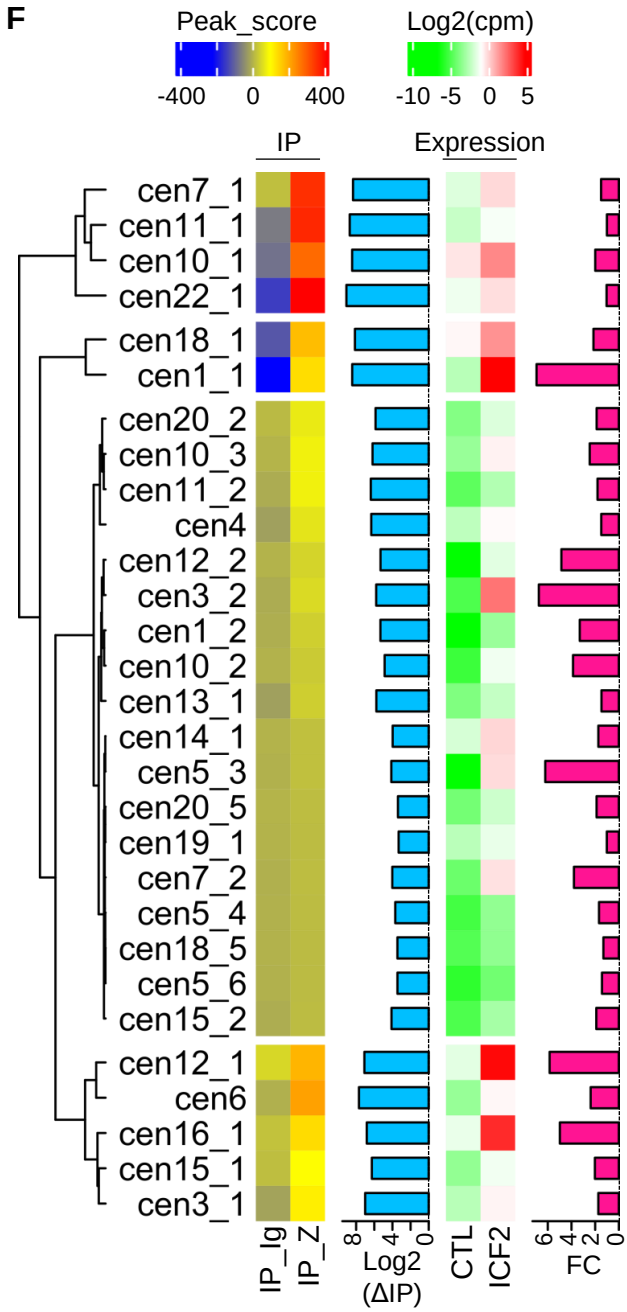

F
Peak_score
-400
0
400
Log2(cpm)
-10
-5
0
5
IP
Expression
cen7_1
cen11_1
cen10_1
cen22_1
cen18_1
cen1_1
cen20_2
cen10_3
cen11_2
cen4
cen12_2
cen3_2
cen1_2
cen10_2
cen13_1
cen14_1
cen5_3
cen20_5
cen19_1
cen7_2
cen5_4
cen18_5
cen5_6
cen15_2
cen12_1
cen6
cen16_1
cen15_1
cen3_1
IP_Ig
IP_Z
Log2 (ΔIP)
CTL
ICF2
FC

Figure 5

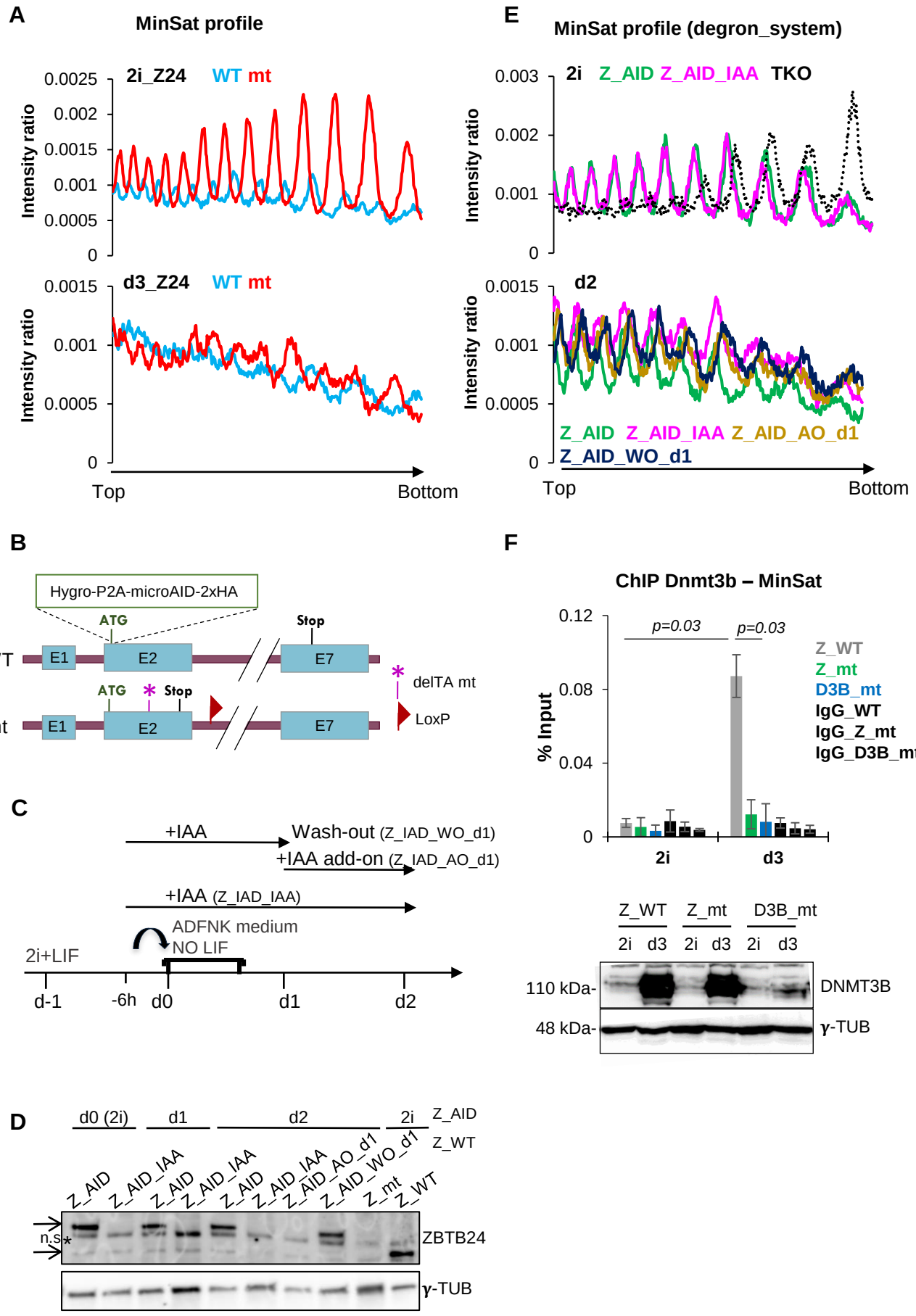
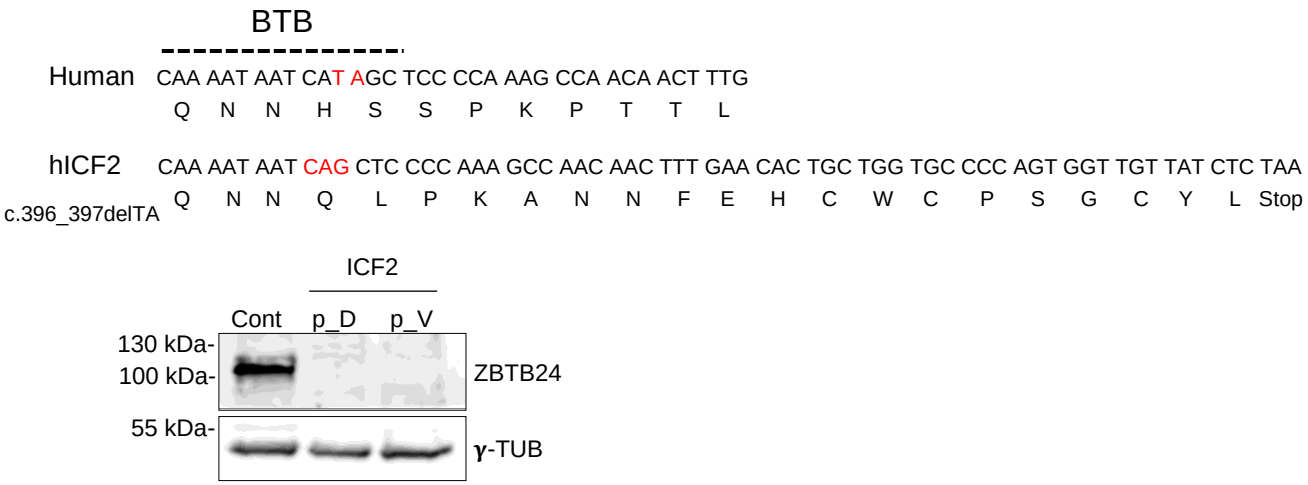

A
MinSat profile
2i_Z24 WT mt
d3_Z24 WT mt
Intensity ratio
Top
Bottom
E
MinSat profile (degron_system)
2i Z_AID Z_AID_IAA TKO
d2
Z_AID Z_AID_IAA Z_AID_AO_d1
Z_AID_WO_d1
B
Hygro-P2A-microAID-2xHA
ATG
Stop
WT
mt
E1
E2
E7
delTA mt
LoxP
C
+IAA
Wash-out (Z_IAD_WO_d1)
+IAA add-on (Z_IAD_AO_d1)
+IAA (Z_IAD_IAA)
ADFNK medium
NO LIF
2i+LIF
d-1
-6h
d0
d1
d2
F
ChIP Dnmt3b – MinSat
% Input
p=0.03
p=0.03
Z_WT
Z_mt
D3B_mt
IgG_WT
IgG_Z_mt
IgG_D3B_mt
2i
d3
Z_WT
Z_mt
D3B_mt
110 kDa-
48 kDa-
DNMT3B
γ-TUB
D
d0 (2i)
d1
d2
2i
Z_AID
Z_WT
Z_AID
Z_AID_IAA
Z_mt
Z_WT
n.s.*
ZBTB24
γ-TUB

Figure S1

A


BTB
Human CAA AAT AAT CAT AGC TCC CCA AAG CCA ACA ACT TTG
Q N N H S S P K P T T L
hICF2 CAA AAT AAT CAG CTC CCC AAA GCC AAC AAC TTT GAA CAC TGC TGG TGC CCC AGT GGT TGT TAT CTC TAA
Q N N Q L P K A N N F E H C W C P S G C Y L Stop
c.396_397delTA

ICF2
Cont
p_D
p_V
130 kDa-
100 kDa-
ZBTB24
55 kDa-
γ-TUB

B

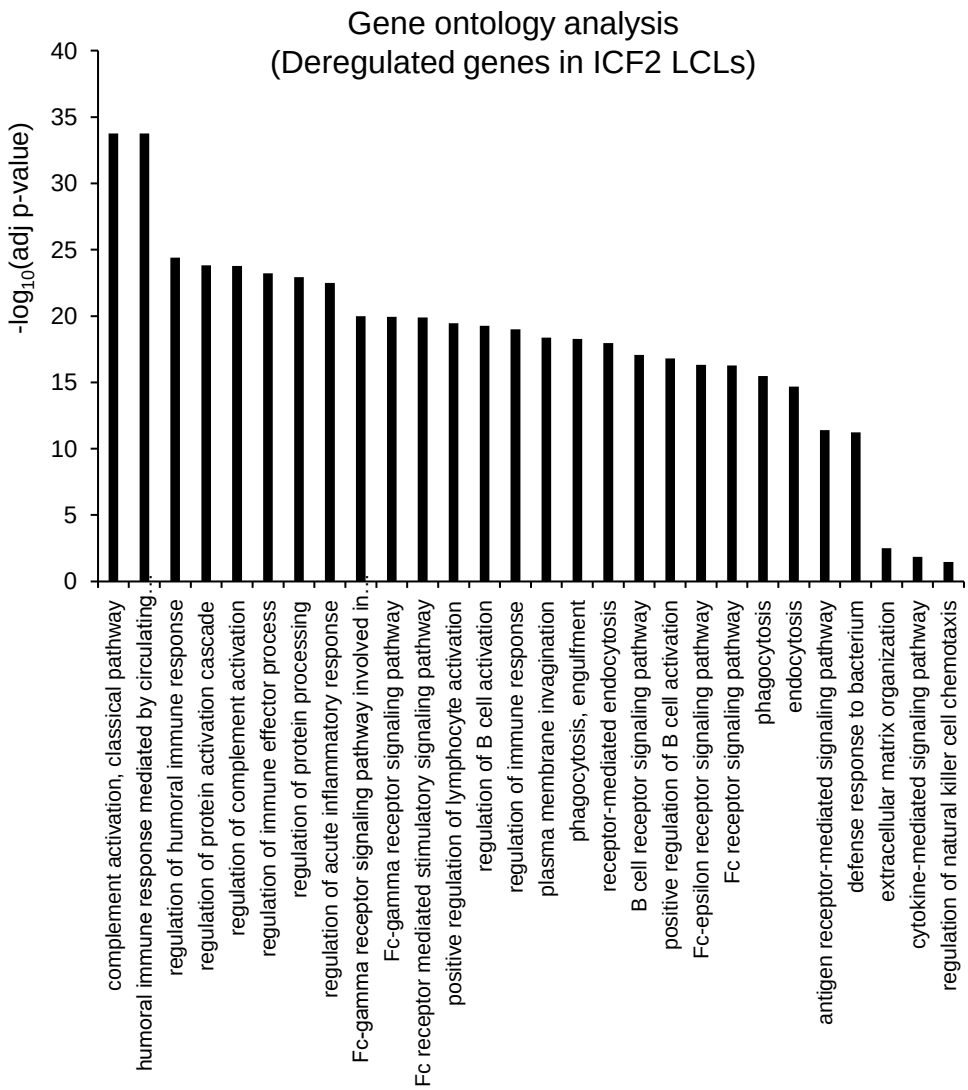

Gene ontology analysis
(Deregulated genes in ICF2 LCLs)
-log10(adj p-value)
40
35
30
25
20
15
10
5
0
complement activation, classical pathway
humoral immune response mediated by circulating...
regulation of humoral immune response
regulation of protein activation cascade
regulation of complement activation
regulation of immune effector process
regulation of protein processing
regulation of acute inflammatory response
Fc-gamma receptor signaling pathway involved in...
Fc-gamma receptor signaling pathway
Fc receptor mediated stimulatory signaling pathway
positive regulation of lymphocyte activation
regulation of B cell activation
regulation of immune response
plasma membrane invagination
phagocytosis, engulfment
receptor-mediated endocytosis
B cell receptor signaling pathway
positive regulation of B cell activation
Fc-epsilon receptor signaling pathway
Fc receptor signaling pathway
phagocytosis
endocytosis
antigen receptor-mediated signaling pathway
defense response to bacterium
extracellular matrix organization
cytokine-mediated signaling pathway
regulation of natural killer cell chemotaxis

Figure S2

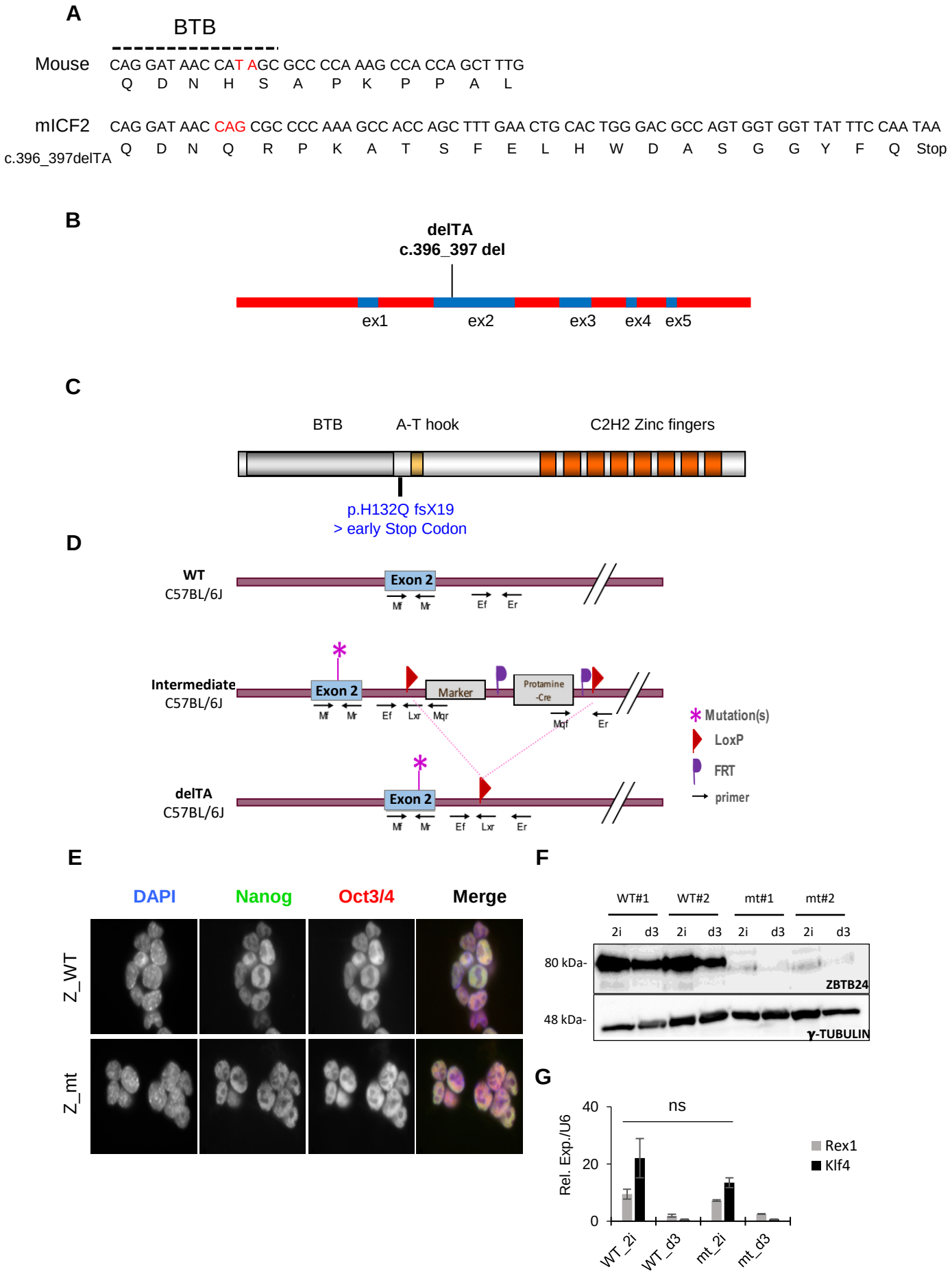

A
BTB
Mouse
CAG GAT AAC CAT AGC GCC CCA AAG CCA CCA GCT TTG
Q D N H S A P K P P A L
mICF2
CAG GAT AAC CAG CGC CCC AAA GCC ACC AGC TTT GAA CTG CAC TGG GAC GCC AGT GGT GGT TAT TTC CAA TAA
Q D N Q R P K A T S F E L H W D A S G G Y F Q Stop
c.396_397delTA
B
delTA
c.396_397 del
ex1
ex2
ex3
ex4
ex5
C
BTB
A-T hook
C2H2 Zinc fingers
p.H132Q fsX19
> early Stop Codon
D
WT
C57BL/6J
Exon 2
Mf
Mr
Ef
Er
Intermediate
C57BL/6J
Marker
Protamine
-Cre
Lxr
Mqr
Mqf
delTA
C57BL/6J
Mutation(s)
LoxP
FRT
primer
E
DAPI
Nanog
Oct3/4
Merge
Z_WT
Z_mt
F
WT#1
WT#2
mt#1
mt#2
2i
d3
80 kDa-
ZBTB24
48 kDa-
γ-TUBULIN
G
ns
Rel. Exp./U6
40
20
0
Rex1
Klf4
WT_2i
WT_d3
mt_2i
mt_d3

Figure S3

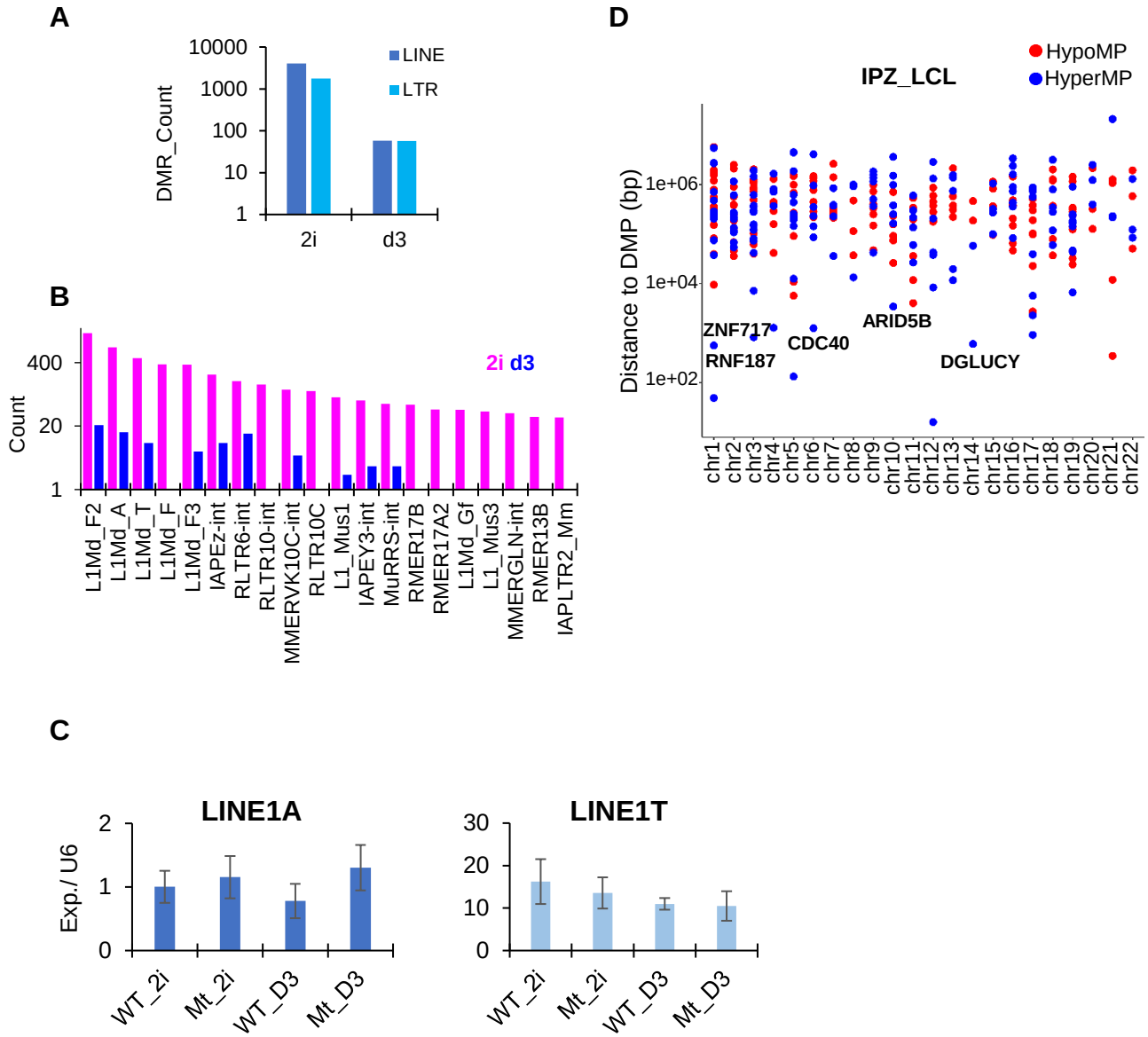
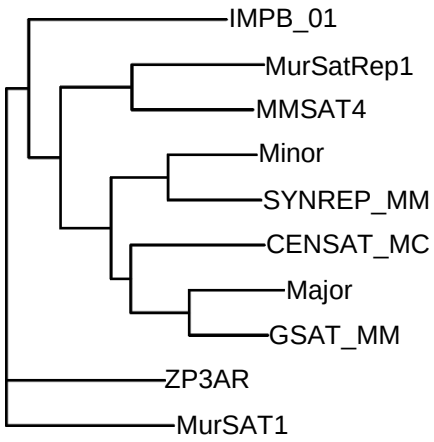

A
DMR_Count
10000
1000
100
10
1
LINE
LTR
2i
d3
B
Count
400
20
1
2i d3
L1Md_F2
L1Md_A
L1Md_T
L1Md_F
L1Md_F3
IAPEz-int
RLTR6-int
RLTR10-int
MMERVK10C-int
RLTR10C
L1_Mus1
IAPEY3-int
MuRRS-int
RMER17B
RMER17A2
L1Md_Gf
L1_Mus3
MMERGLN-int
RMER13B
IAPLTR2_Mm
D
IPZ_LCL
HypoMP
HyperMP
Distance to DMP (bp)
1e+06
1e+04
1e+02
ZNF717
RNF187
CDC40
ARID5B
DGLUCY
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
C
LINE1A
Exp./ U6
2
1
0
WT_2i
Mt_2i
WT_D3
Mt_D3
LINE1T
30
20
10
0
WT_2i
Mt_2i
WT_D3
Mt_D3

Figure S4

A



B

CLUSTAL FORMAT : MUSCLE (3.8) multiple sequence alignment

```
CENSAT_MC   ------------------------------------------------------------
SYNREP_MM   GAACATGTTATATGAGTGAGTTACANTGAAAAACATANAAATNGGAAACGNGAATTGTAG
Minor       ------------------------------------------------------------

CENSAT_MC   ------------------------------------------------------------
SYNREP_MM   AACATCGTATATNAANGAGTTACAATGAAAAACATGGAAAATGAGAAAAACCACACTGTA
Minor       ----------------------------------GGAAAATGATAAAAACCACACTGTA

CENSAT_MC   --ACACGGGATGTTAATGAGTTGCACTGAAAAAC----------------------GNAG
SYNREP_MM   GAACATTGTGTATGAGTGAGTTACAATGAAAAACATANANATNGGAAACGAGANTTGTAG
Minor       GAACATATTAGATGAGTGAGTTACACTGAAAAACACATTCGTTGGAAACGGGATTTGTAG
              ***       * * ****** ** ********                      * **

CENSAT_MC   AA--------------------AATGAGAAACGC----------------ACAGTGTA
SYNREP_MM   AACATCGTATATAAAGGAGTTACAATGAAAAACATGGAAAATGAGAAAAACCACACTGTA
Minor       AACAGTGTATATCAATGAGTTACAATGAGAAACAT-----------------------
            **                      ***** ****

CENSAT_MC   G------------------------
SYNREP_MM   GAACATTGTGTATGAGTGAGTTAGGG
Minor       -------------------------
```

Figure S5

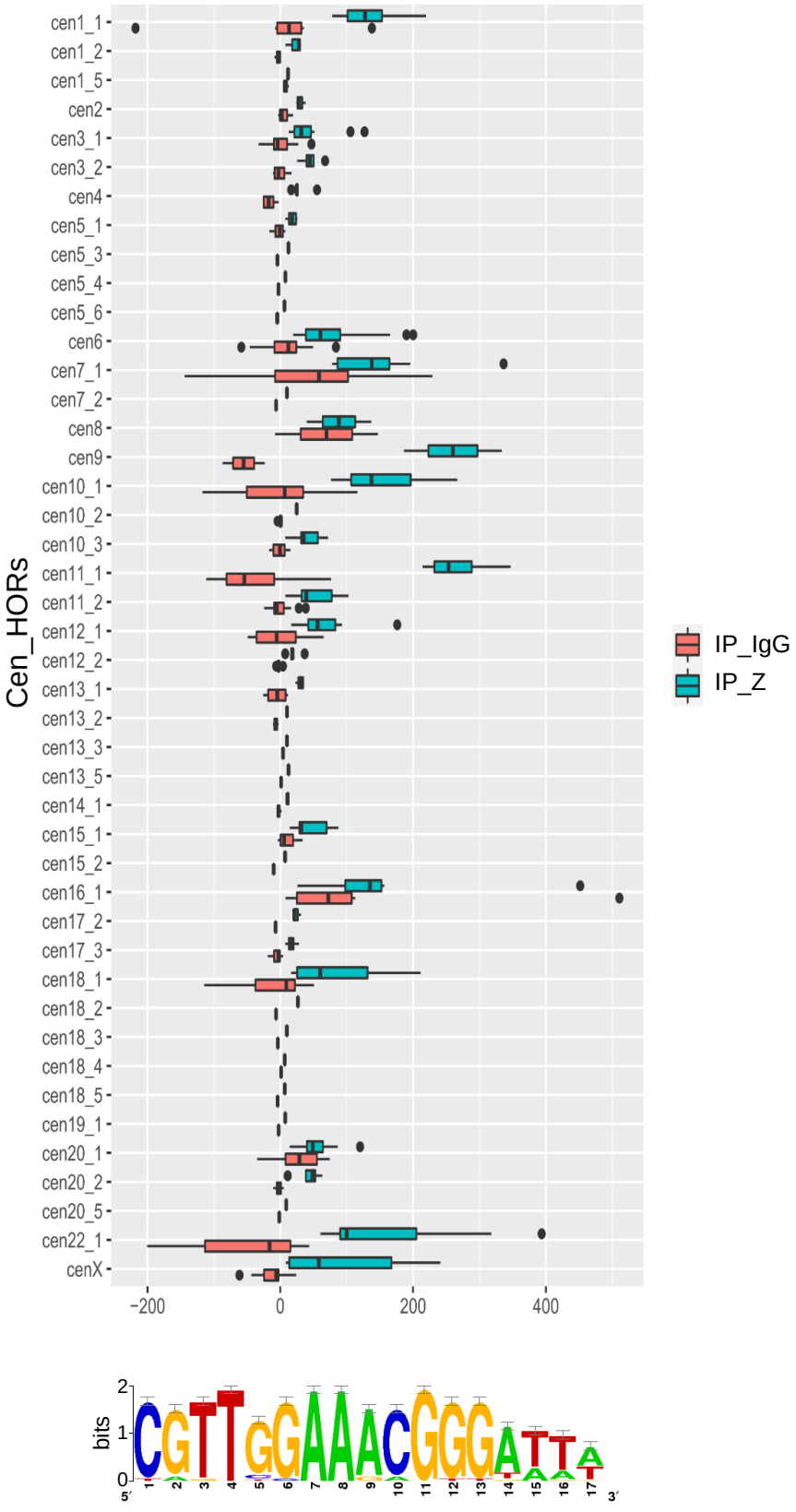

cen1_1
cen1_2
cen1_5
cen2
cen3_1
cen3_2
cen4
cen5_1
cen5_3
cen5_4
cen5_6
cen6
cen7_1
cen7_2
cen8
cen9
cen10_1
cen10_2
cen10_3
cen11_1
cen11_2
cen12_1
cen12_2
cen13_1
cen13_2
cen13_3
cen13_5
cen14_1
cen15_1
cen15_2
cen16_1
cen17_2
cen17_3
cen18_1
cen18_2
cen18_3
cen18_4
cen18_5
cen19_1
cen20_1
cen20_2
cen20_5
cen22_1
cenX
Cen_HORs
-200
0
200
400
IP_IgG
IP_Z
bits
2
1
0
5′
3′
25 sites, evalue=1.4e-17

Figure S6

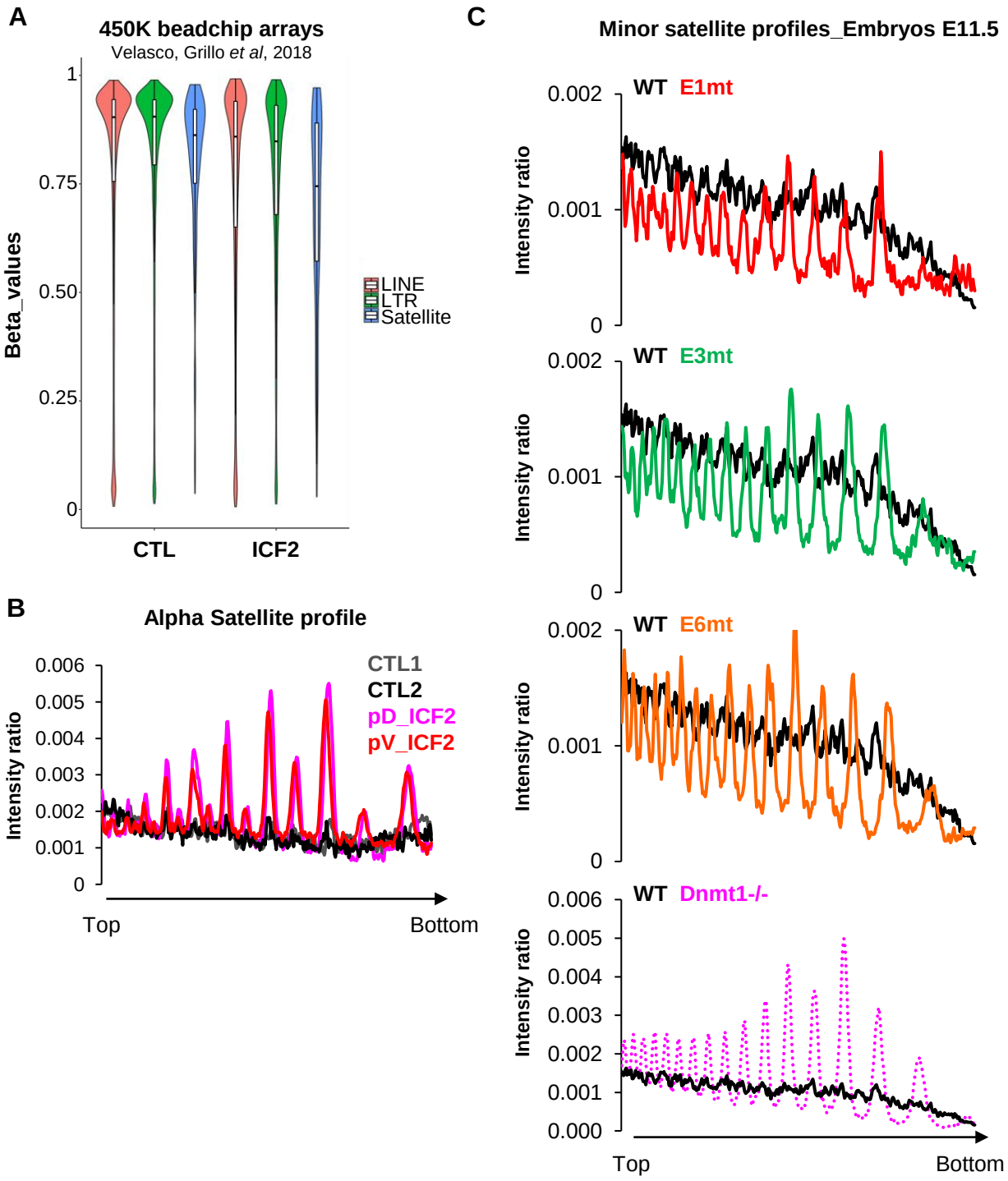
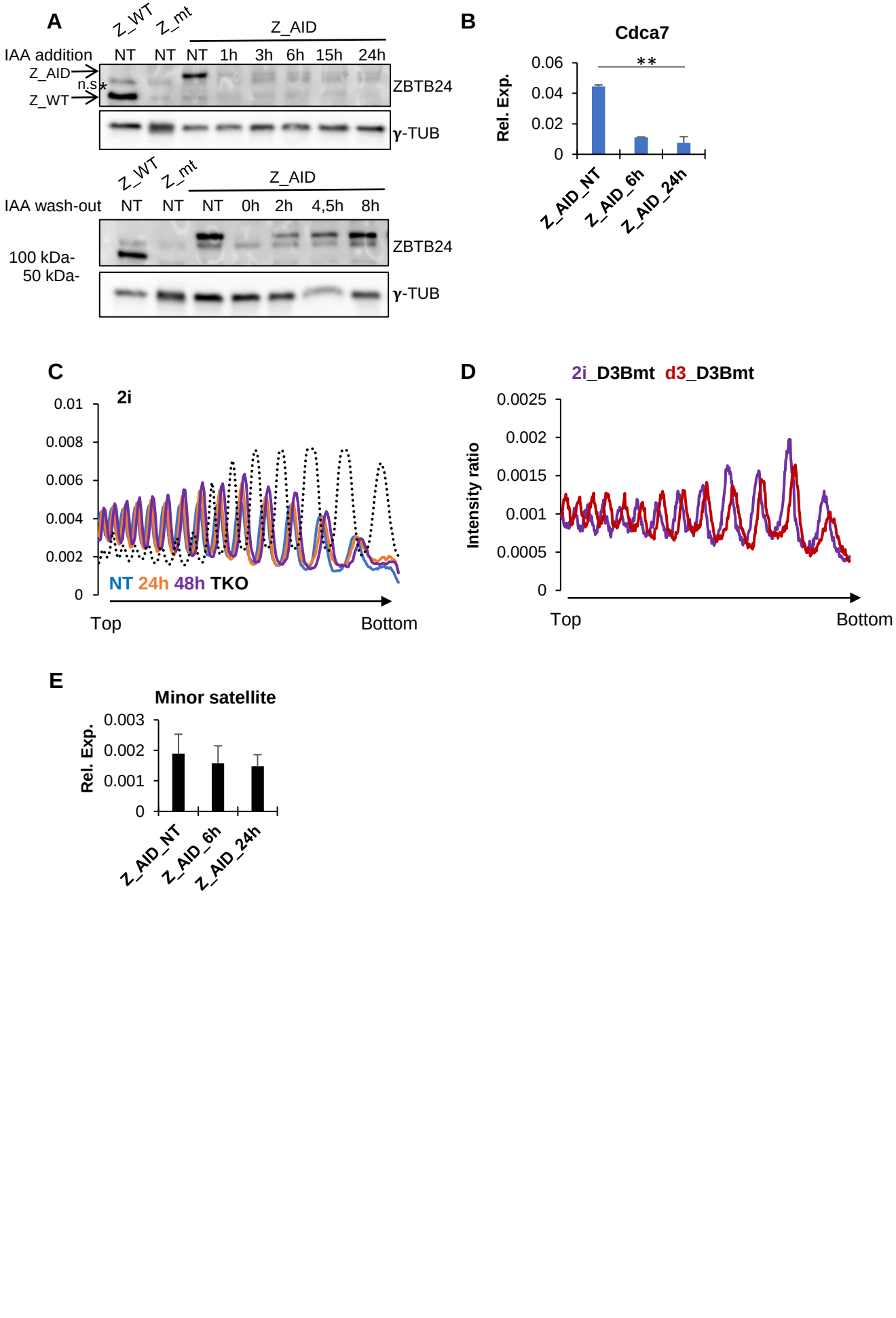

A
450K beadchip arrays
Velasco, Grillo et al, 2018
Beta_values
1
0.75
0.50
0.25
0
CTL
ICF2
LINE
LTR
Satellite
B
Alpha Satellite profile
Intensity ratio
0.006
0.005
0.004
0.003
0.002
0.001
0
CTL1
CTL2
pD_ICF2
pV_ICF2
Top
Bottom
C
Minor satellite profiles_Embryos E11.5
WT E1mt
WT E3mt
WT E6mt
WT Dnmt1-/-
Intensity ratio
0.002
0.001
0
0.006
0.005
0.004
0.003
0.002
0.001
0.000
Top
Bottom

Figure S7


A
Z_WT
Z_mt
Z_AID
IAA addition
NT NT NT 1h 3h 6h 15h 24h
Z_AID
n.s*
Z_WT
ZBTB24
γ-TUB
IAA wash-out
NT NT NT 0h 2h 4,5h 8h
100 kDa-
50 kDa-
B
Cdca7
Rel. Exp.
0.06
0.04
0.02
0
**
Z_AID_NT
Z_AID_6h
Z_AID_24h
C
2i
0.01
0.008
0.006
0.004
0.002
0
NT 24h 48h TKO
Top
Bottom
D
2i_D3Bmt d3_D3Bmt
Intensity ratio
0.0025
0.002
0.0015
0.001
0.0005
0
Top
Bottom
E
Minor satellite
Rel. Exp.
0.003
0.002
0.001
0
Z_AID_NT
Z_AID_6h
Z_AID_24h