Understanding the mechanisms of histone modifications in vivo

Nithyha Parameswaran Kalaivani

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des SCIENCES de la VIE et de la SANTÉ

THÈSE présentée par :

Nithyha PARAMESWARAN KALAIVANI

soutenue le : **16 Dec 2016**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaire et cellulaire de la biologie

Understanding the mechanisms of histone modifications *in vivo*

THÈSE dirigée par:

Prof. Dr. Robert SCHNEIDER — IGBMC,Université de Strasbourg and Functional Epigenetics, Helmholtz Zentrum **München**

RAPPORTEURS:

Dr. Nicola IOVINO — MPI, Freiburg

Dr. Florence CAMMAS — IRCM, Montpellier

AUTRES MEMBRES DU JURY:

Dr. Didier DEVYS — IGBMC, ILLKIRCH

Acknowledgements

I dedicate this thesis to the loving memory of my dad!
I would like to thank

Prof. Dr. Robert Schneider for providing me an opportunity to work in his lab, for his excellent guidance, for his constant support throughout my ups and downs, for his patience, and for the incredible motivation that he provided throughout my PhD. Thanks Rob!

My thesis committee members, Dr. Andrew Bannister and Dr. Romeo Ricci for their scientific guidance.

My thesis defense committee Dr. Didier Devys, Dr. Nicola Iovino and Dr. Florence Cammas for agreeing to evaluate my thesis.

Vincenzo for putting up with me in the first year of my thesis, for guiding me until I got a hang on things, for being patient with me, for being an amazing source of entertainment, and being that friend in the lab that I can always count on to make my otherwise gloomy days positive.

Sylvain for providing me scientific advises throughout my PhD and especially this last year, for the amazing discussions, and for welcoming me in the lab on day 1 and sending me off on day eternity in Strasbourg.

Nina and Poonam for their friendship and for being my personal and scientific to go girls in the lab.

Moyra for her friendship and for her incredible stem cell advises.

Stephanie for being my go to girl for almost everything in the lab and helping me during my rough days.

Adam, Anas, Stephanie and Sylvain for making the last year as much fun for me. We will always be the amazing little left overs of Strasbourg!

All the past and present Schneider lab members for their friendship, scientific advise, support and for sticking with me, cheering me up during my tough times. Thanks a lot guys!

Claudine and Muriel from FACS facility of being incredibly co-operative and supportive throughout my crazy CRISPR phase.

Dr Bernardo Reina San Martin for the gene editing discussions.

Friends from IGBMC for their scientific discussions.

All the facilitites of IGBMC

All of my friends from all over the world, who stayed with me during thick and thin. I have had the most fun and amazing time of my life during these four years, thanks to you all!

My amazing, and incredible family, for providing me constant support and love to sustain tough times and motivating me to find my way through it! For being patient with me and just being there for me. It wouldn't have been possible without you!

Table of Contents

Abstract

Post-translational modifications (PTMs) of histones have emerged as key players in the regulation of gene expression. Ever since the discovery of histone acetylation, there has been a growing list of histone post-translational modifications including phosphorylation, methylation, and ubiquitination. Whilst PTMs are found on all histones, the most studied ones map all to histone N-terminal tails. Histone PTMs are implicated in a various cellular processes including transcription, DNA damage, replication, and cell cycle regulation. Interstingly, histone-modifying enzymes are now drug targets due to their dysregulation in diseases such as cancer. However, little is known to what extent PTMs can impact directly on chromatin function. It has been suggested that PTMs on the N-terminal tails of core histones (H2A, H2B, H3 and H4) have the potential to govern chromatin function according to the so called ‘‘histone code’’ hypothesis by recruiting specific binding proteins. Recently, PTMs have also been detected within the globular domains of histones and map to residues located on the lateral surface of the histone octamer that are in direct or water-mediated contact with the nucleosomal DNA. Within the globular domain of H3 lie four lysine residues that map to the lateral surface, the outer surface of the histone octamer that is close to the DNA: K56, K64, K115, and K122. All four of these lysines can potentially be acetylated. Studies from our lab on H3K122 and H3K64 acetylation suggests that these modifications stimulate transcription, regulate nucleosomal stability and facilitate nucleosome eviction thus, affecting the gene expression.

The goal of my project was to gain insight in the function of acetylation within the globular domain of H3 *in vivo* and to compare these modifications with histone tail modifications by using a CRISPR based system to mutate histones in mouse ES cells. The hypothesis was that histone core modifications H3K56, H3K64, H3K122, and H3K115 acetylation might act synergistically and/or in different combinations, increasing the impact on nucleosome dynamics and hence transcription. In order to study the impact of PTMs *in vivo*, all endogenous WT H3 gene copies have to be replaced with mutant copies that can mimic or prevent PTMs. For the first time in mammals, we have established a “clean” H3.3 mutant system, as an important *in vivo* model to study the effects of histone acetylation on gene expression and chromatin structure without background of endogenous H3.3 histones. These cell lines express mutated H3 where specific lysine residues (in the H3 tails and core) are mutated to arginine (K/R) that cannot be acetylated (mimicking constitutively unacetylated

H3) or by mutating lysine to glutamine (K/Q) to mimic an acetylated lysine. Our results suggest that H3K56, H3K64, H3122, and H3K115 acetylation act synergistically, establishing a strong link between H3K56/64/115/122 acetylation and active transcription *in vivo*. Our data suggest that these modifications have a direct impact on transcriptional activation and potentially influence ES cell differentiation when compared to the tail modifications. In conclusion, these results give us key insights on globular domain modifications that might act in combinations to affect the nucleosome dynamics and hence gene transcription.

Avant propos

L'organisation par l'empaquetage de l'ADN génomique est réalisé grâce à des protéines, parmi elles les histones, qui en s'associant à l'ADN, forment une structure complexe appelée chromatine. Chacune des quatre histones canoniques, H3, H2A, H2B et H4, est constituée de deux domaines structurellement et fonctionnellement distincts : le domaine globulaire qui forme le squelette autour duquel 147 pb d'ADN s'enroulent et les extrémités flexibles ou queues N-terminales non structurées qui dépassent du cœur nucléosomal. Jusqu'à récemment, les études se sont concentrées sur les queues d'histones, qui font office de plateformes pour un grand nombre de modifications post-traductionnelles. Selon le « code des histones », ces modifications ont le pouvoir de réguler les fonctions de la chromatine en recrutant ou repoussant des protéines d'interactions appelées « effecteurs ». Cependant, certaines modifications post-traductionnelles ont aussi été identifiées dans le domaine globulaire des histones, en particulier au niveau d'acides aminés localisés sur la surface latérale de l'octamère d'histones. Ces acides aminés sont particulièrement intéressants car ils interagissent avec l'ADN soit par contact direct, soit par l'intermédiaire d'une molécule d'eau.

De manière intéressante, quatre lysines du domaine globulaire de l'histone H3, K56, K64, K115, and K122, sont localisées au niveau de cette surface latérale de l'octamère d'histones, en contact direct avec l'ADN, et peuvent potentiellement être acétylées. Du fait de leur localisation structurellement importante, Cosgrove *et al.* ont proposé un modèle de « mobilité régulée du nucléosome », selon lequel les modifications sur la surface latérale pourraient directement affecter la dynamique du nucléosome et de la chromatine, et ceci contrairement à celles présentes sur les queues d'histones qui le font par l'intermédiaire de protéines effectrices. Confirmant ce modèle, des travaux de notre laboratoire sur l'acétylation de H3K122 et H3K64 suggèrent que ces modifications régulent la stabilité du nucléosome, facilite son éviction et stimule la transcription, influençant ainsi l'expression génique *in vivo*.
Ces différentes études nous ont conduit à émettre l'hypothèse que les acétylations de H3K122, H3K56, H3K64 et H3K115 pourraient agir toutes ensemble de manière synergique et/ou dans des combinaisons différentes, renforçant l'impact sur la dynamique du nucléosome et ainsi sur la transcription. Afin d'étudier l'impact réel de ces modifications post-traductionnelles *in vivo*, nous considérons que toutes les copies endogènes sauvages du gène

H3 doivent être remplacées par des copies de gènes mutés de telle sorte qu'elles miment ces modifications ou empêchent leur apparition.

C'est la raison pour laquelle la première partie de mon projet de thèse implique l'établissement de lignées cellulaires qui expriment exclusivement une version mutante de H3 dans laquelle les lysines ont été remplacées, soit par des arginines (K/R) ne pouvant pas être acétylées, soit par des glutamines (K/Q) mimant une lysine acétylée. La deuxième partie de mon projet implique l'étude des effets combinés des acétylations de H3122, H3K56, H3K64 et H3K115 sur la différentiation des cellules ES, la transcription, le positionnement des nucléosomes, la compaction de la chromatine ou encore la synergie/crosstalk entre les différentes modifications du domaine globulaire de l'histone H3.

Au cours des premiers 30 mois de ma thèse, j'ai utilisé la lignée cellulaire aviaire lymphoïde B DT40 pour son exceptionnel taux d'intégration ciblé d'ADN par recombinaison homologue par rapport à une insertion aléatoire. Par ailleurs le poulet présente également un moins grand nombre de gènes d'histones avec seulement deux variants H3 en plus de CENP-A : (a) les histones canoniques H3/H3.2 codées par 8 gènes, tous organisés en un groupe unique sur le chromosome 1 et (b) l'histone H3.3 codée par deux gènes . Pourtant durant mes tentatives d'établir une lignée cellulaire mutante pour H3.3, la version finale du génome de Gallus gallus est sortie avec plusieurs ajustements concernant le nombre et la localisation des gènes d'histones. A la même période, l'équipe du Dr. Julien Sale à Cambridge, a généré une lignée cellulaire DT40 déficiente pour H3.3 que nous avons pu obtenir. Grâce à l'obtention de cette lignée cellulaire, j'ai pu adapter mes stratégies pour tenter d'établir des lignées cellulaires mutantes. Malheureusement nous avons rencontré des problèmes additionnels dus aux variations de séquence entre le génome du poulet et celui de la lignée cellulaire DT40.
Malgré l'utilisation de multiples stratégies employant différentes conditions et techniques pour l'édition de gènes dans la lignée DT40, je n'ai pas été en mesure de générer des lignées mutantes pour H3. Ainsi j'ai décidé de changer de système pour passer à la souris et plus précisément aux cellules souches embryonnaires (ES). Même si la souris possède 14 copies du gène H3 incluant les variants H3.1, H3.2 et H3.3, la précision du système d'édition génétique CRISPR/Cas9, bien démontrée dans les cellules ES, va me permettre de manipuler leur génome de manière efficace.

La souris possède deux gènes H3f3a et H3f3b codant pour le variant H3.3, incluant des introns et respectivement longs de 3,5 et 11 kb. La stratégie employée a été de créer une délétion homozygote de H3f3a et hétérozygote de H3f3b, pour ensuite remplacer par knock-in (KI) l'allèle restant de H3f3b par des versions mutantes. J'ai utilisé le vecteur px330-cas9 (U6-gRNA / Cbh / Cas9-3xFLAG / EGFP or mRFP) pour cibler le gène H3f3b dans les régions 5' en amont et 3' en aval de la délétion et réaliser un KO hétérozygote. L'allèle restant, possédant un intron plus petit, présente un avantage pour l'étape de KI des mutants H3f3b taggés 2xHA. Après avoir réussi à supprimer un allèle du gène H3f3b, j'ai utilisé une stratégie similaire pour cibler le gène H3f3a et réalisé un KO homozygote. Je suis en train de générer à l'heure actuelle les clones portant un KI des différents mutants H3f3b.

La souris possède 12 copies de gènes sans introns codant pour H3.1 et H3.2, réparties entre deux groupes de gènes d'histones. J'ai d'abord ciblé 3 gènes H3.1 : H3a, H3b et H3c, qui s'étalent sur 18kb du groupe de gènes hist1. J'ai tout d'abord voulu réaliser une délétion directe en utilisant des vecteurs p330x-cas9 afin de cibler la région de 18kb. Pourtant les clones issus de ce premier ciblage ne présentaient pas de KO. Cela pourrait venir du fait que la délétion de régions de grande taille en utilisant le système CRISPR/Cas9, reste encore difficile. Ainsi j'ai décidé d'utiliser le système CRE/Lox au cours duquel je pourrais insérer par KI des sites LoxP dans la même orientation de part et d'autre de la région à supprimer. Puis par induction de l'enzyme CRE, les sites LoxP pourront recombiner, aboutissant à l'élimination de la région génomique d'intérêt. J'ai donc conçu des oligos contenant les 34 pb constituant le site LoxP ainsi que 6pb du site de restriction EcoRI, le tout flanqué de manière adjacente de chaque côté de la coupure double brin par une séquence homologue de 60pb. J'ai ensuite co-transfecté les vecteurs px330-Cas9-gRNA et les oligos simple-brin ciblant les deux extrémités 5'et 3'. Aucun des clones sélectionnés ne portaient une insertion LoxP aux deux extrémités. Par contre, un clone unique portant une insertion LoxP seulement en 3'a pu être sélectionné puis de nouveau transfecté de la même façon avec le vecteur px330-Cas9-gRNA pour ne cibler que la région 5' et y insérer le site LoxP. Finalement, des clones portant le site LoxP en 5' et 3' ont pu être sélectionnés, puis transfectés avec un plasmide d'expression de CRE, ce qui a entrainé après 48h la délétion de la région comprenant les gènes H3a, H3b et H3c. Forte de cette délétion réussie de 18kb, j'utilise à l'heure actuelle une approche similaire afin d'éliminer d'autres gènes H3.

Lorsque j'aurai réussi à remplacer l'allèle restant du gène H3f3b par les gènes H3.3 sauvage ou mutés suivants : WT, K56/64/115/122R, K56/64/115/122Q, K9/14/18R, K9/14/18Q, tous taggés avec 2xHA, puis éliminé d'autres copies du gène H3 afin d'obtenir un nombre minimal permettant la survie cellulaire, je propose d'effectuer les expériences suivantes : Par séquençage d'ARN je tenterai afin de voir si et comment une mutation unique dans le domaine globulaire ou une combinaison d'entre elles peut impacter l'expression génique. Ces expériences me permettront de comprendre comment les modifications dans le domaine globulaire d'histone peuvent affecter l'état d'équilibre de l'expression génique. J'étudierai en parallèle les effets de mutants d'histone sur l'activation des gènes de réponse rapide au TPA dans les cellules ES. Je cultiverai pour cela les lignées cellulaires WT, les mutants H3R et H3Q avec du TPA et sans sérum de veau fœtal (FCS). J'extrairai les ARN à différents temps d'induction, produirai les ADN complémentaires correspondants par transcription inverse et réaliserai des PCR quantitatives afin d'évaluer l'activation des gènes en question en comparant les différentes lignées cellulaires WT et mutantes. Cela me permettra de voir si les effets combinés de K56/64/115/122ac conduisent à une augmentation de l'activation transcriptionelle plus substantielle que celle obtenue avec une seule modification. Je comparerai également les cinétiques d'activation transcriptionnelle obtenues dans le cas de modifications aux extrémités flexibles des histones avec celles obtenues dans le cas de modifications dans le domaine globulaire. Finalement, pour étudier le crosstalk et la synergie entre les différentes modifications du domaine globulaire, j'analyserai et comparerai les effets de chaque mutation pour chaque position en K56, K64, K115 et K122, sur d'autres modifications d'histone (par exemple acétylation dans les queues ou le domaine globulaire d'histone).

Mon projet va me permettre d'établir pour la première fois un modèle *in vivo* propre et non ambigu, sans le bruit de fond des histones sauvages endogènes, pour étudier les effets de l'acétylation des histones sur l'expression génique et la structure de la chromatine. J'ai jusqu'à maintenant établi les outils nécessaires à cette étude et peux maintenant aller plus loin pendant les trois prochains mois en finissant la création des lignées cellulaires mutantes pour H3 et menant à bien les essais cellulaires adéquats. Cela apportera des éléments de compréhension cruciaux concernant l'importance des modifications globulaires d'histone dans la cellule, tout en complémentant parfaitement les études précédentes realisées *in vitro,* et abordant des questions importantes au sujet des liens de cause à effets des modifications d'histone, en particulier leur rôle dans le cancer.

1. Introduction

1.1 A brief Introduction to Epigenetics

Epigenetics has influenced nearly every aspect of biology and has enormous potential for human health, making it an important field in science. The word epigenetics dates back to 1942 when developmental biologist Conrad Waddington took the Greek word epigenesis, which proposed that the early embryo was undifferentiated, and changed it to epigenetics[1]. The Greek prefix "epi" means "on top of" or "over", so the term "Epigenetics" describes regulation at a level above, or in addition to, those of genetic mechanisms. There are several existing definitions for epigenetics. As an unifying definition, the following has been suggested: "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states"[2].

Much of today's epigenetics research is about studying the covalent and non-covalent modifications on chromatin, and how they influence overall chromatin structure. Chromatin modifications have emerged as a core mechanism for transcriptional regulation of gene expression[3]. But like any other field, epigenetics has its own challenges, particularly to study the heritability of epigenetic information through cellular division, differentiation, and transgenerational inheritance. There are still questions about stability of chromatin modifications during and outside DNA replication[3,4]. Hopefully, in future, novel approaches towards understanding epigenetics will provide us with a far richer understanding of how epigenetic mechanisms shape transcriptional control across an incredibly varied background of developmental stages, tissue types, and disease states[5].

1.2 Chromatin organisation

The Human diploid genome contains approximately 6 billion base pairs of DNA packaged into 22 pairs of chromosomes (except ova and sperm, which are haploid) in each cell nucleus. Each base pair is around 0.34 nanometres making DNA in every cell about 2 metres long. Yet the size of a cell is only around 10-100 micrometres. How this kind of compaction is achieved?

Histones are the proteins that are necessary for the compaction of chromosomal DNA into small nucleus of eukaryotic cells. They are positively charged proteins and comprise of H1, H2A, H2B, H3 and H4[6]. DNA with its phosphate-sugar backbone is negatively charged, thus binding very tightly to the histone proteins. Two copies of histones H3-H4 form a tetramer that combines with two dimers containing H2A-H2B complex to assemble a histone octamer. 147 base pairs of DNA wrap in 1.65 super-helical turns around each histone octamer thus forming a DNA-histone complex called nucleosomal core particle[6,7]. Linker histone H1 binds to the nucleosomal core particle and organises the nucleosome arrays into higher order chromatin structures representing first level of chromatin organization[8,9] and further into higher order chromatin fibres during interphase and 200-300nm chromatid structures in mitotic chromosomes[10,11] (Figure 1.1). Nucleosome is the fundamental repeating functional and structural unit of chromatin[12].

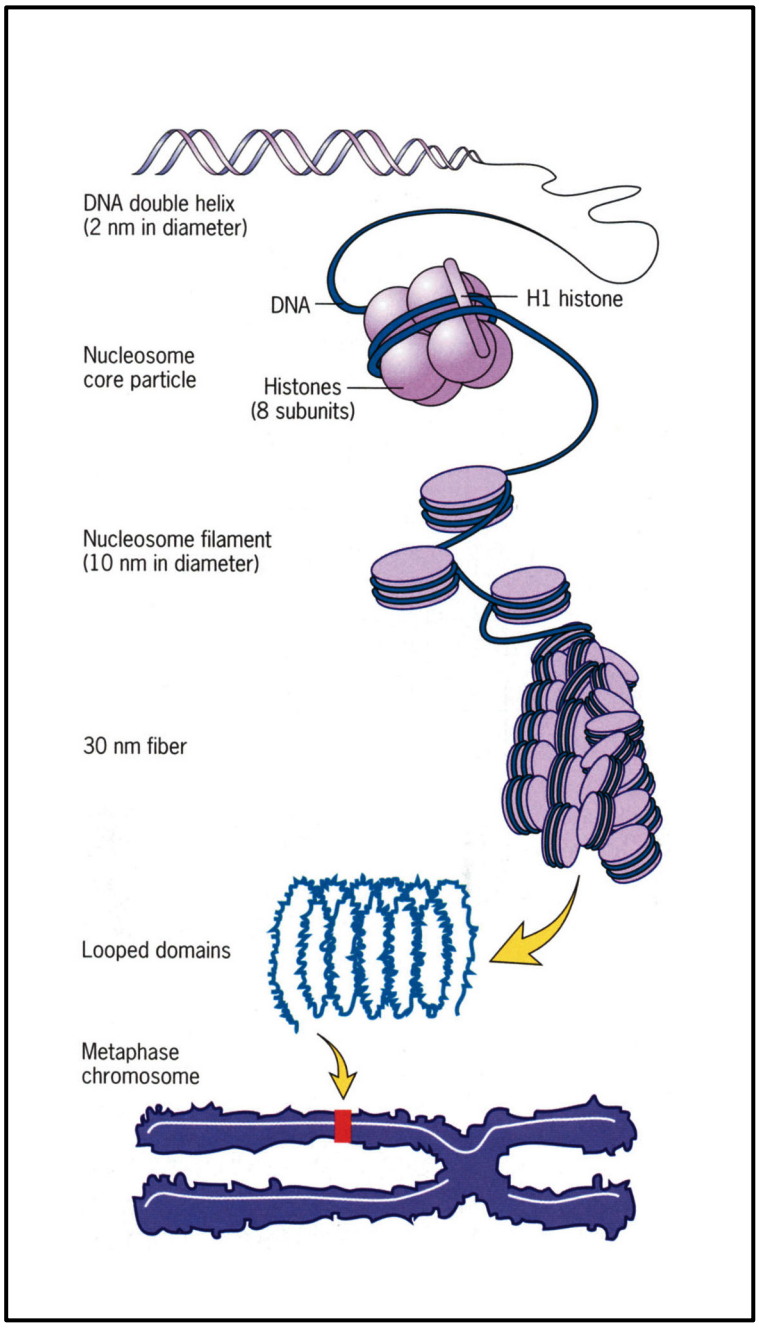


Figure 1.1: Chromosomes are made of DNA-histone protein complexes.
Chromosomal DNA is packaged inside nuclei with the help of histones. These are positively charged proteins that strongly adhere to negatively charged DNA and form complexes called nucleosomes. Each nucleosome is composed of DNA wound 1.65 times around eight histone proteins. Nucleosomes fold up to form a 30nm chromatin fiber, which then forms looped domains. These loops are further compressed, folded and tightly coiled into the chromatid of a chromosome Adapted from Pierce, B. in *Genetics,* 2004 [13].

Microscopy provides the ability to identify specific regions of chromatin thus giving insights into interphase chromatin organisation[14]. Based on the differential compaction of chromatin in interphase, it can be distinguished into either heterochromatin or euchromatin. Heterochromatin is highly condensed, usually inaccessible, and often localized to the

periphery of the nucleus[15]. Constitutive heterochromatin contains chromosomal regions that contain repetitive DNA elements such as major and minor satellite sequences. Facultative heterochromatin contains many repetitive elements that are found in telomeres and centromeres. Euchromatin is less condensed, more accessible and generally more easily transcribed[16].

Even though recent advances in imaging and molecular approaches have postulated significant insights into chromatin organisation, elucidating the intrinsic properties of the chromatin will contribute to a more deeper understanding of the various genomic interactions including DNA replication and gene expression and chromosome assembly[17]

1.3 The core histones

The four core histones are small, basic proteins that are highly conserved among eukaryotes. Each histone consists of a conserved histone fold domain and a unique N-terminal tail (Figure 1.2)[12]. All the four core histones interact in pairs via a "handshake motif" with two H3/H4 dimers interacting together to form a tetramer, while the two H2A/H2B dimers associate with the H3/H4 tetramer in the presence of DNA[18]. The histone fold domain is a hydrophobic core commonly referred to as the globular domain. It is made of two shorter α helices (α1 and α3) 9 to 14 residues in length, which flank a relatively long (29 residue) α helix (α2). The α helices are connected by short loop/β segments[19]. Histone tails are flexible regions protruding from the globular domain of histones and form the nucleosome. These tails are very short comprising of less than 40 amino acids[20].

Many core histone structures including the α helices play important roles in contacting nucleosomal DNA. There are about 120 contact points between the DNA and the histone octamer within a nucleosome[12]. Most importantly, these stable interactions includes the strong contact of the edge of the core DNA made by the α helix within H3 and the beginning of the H3 N-terminal tail domain[12]. This critical point of contact hinders potential enzymes invading the nucleosomes.

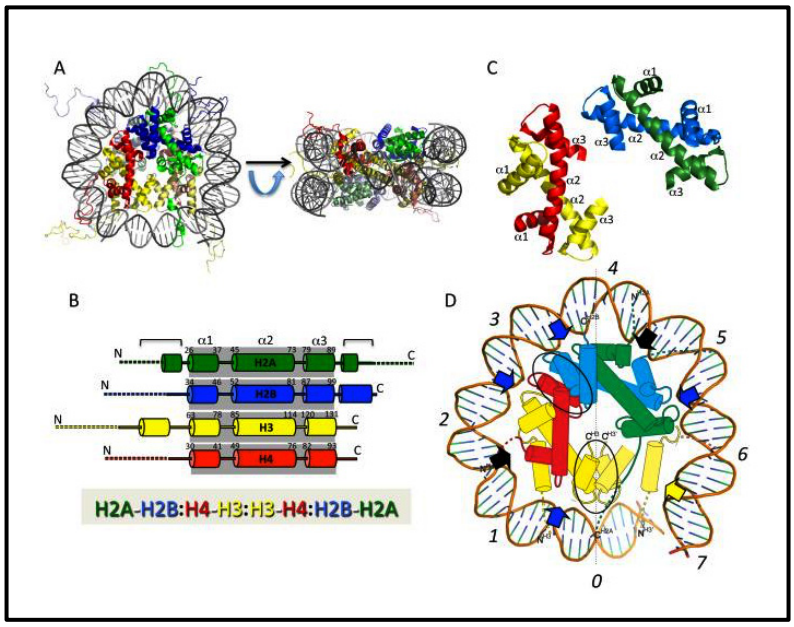


Figure 1.2: Nucleosomal core particle structure.
A. Model of a nucleosomal core particle. H2A, green, H2B, blue, H3, yellow, H4, red. B. Schematic showing secondary structure of the core histone proteins, with α-helices represented by columns. Dashed lines indicate residues within 'tail' domains C. H2A-H2B (green/blue) and H3-H4 (yellow/red) histone dimers. α1, α2 and α3 helices indicated. D. Schematic showing one-half of nucleosomal core, facing down the DNA super-helix axis. Adapted from Cutter, A. R. and Hayes, J. J, 2015[19].

Histone tail domains are unstructured and adopt coil conformations randomly when the proteins are free in solution without DNA or released from their nucleosomal binding sites in high salt solutions[21,22] . Multiple tail residues are tightly associated with nucleosome binding sites in physiological salts[26]. The tail domains contribute slightly to the thermal stability of mononucleosomes[23]. Removing the tail domains increases the unwrapping of DNA and thus providing accessibility of DNA binding factors to nucleosomal DNA[24,25]. Histone tails are easily accessible to enzymes that contribute to posttranslational modifications (PTMs). Some of these PTMs are located in positions that are important for internucleosomal interactions within condensed chromatin structures and are involved in organising higher order chromatin structures[26].

Recent studies suggest that nucleosomes are highly dynamic, capable of being altered in structure, composition and positioning along the DNA[27] (Figure 1.3). These chromatin modulations could be achieved by chromatin remodeling complexes, incorporation of histone variants and PTMs.

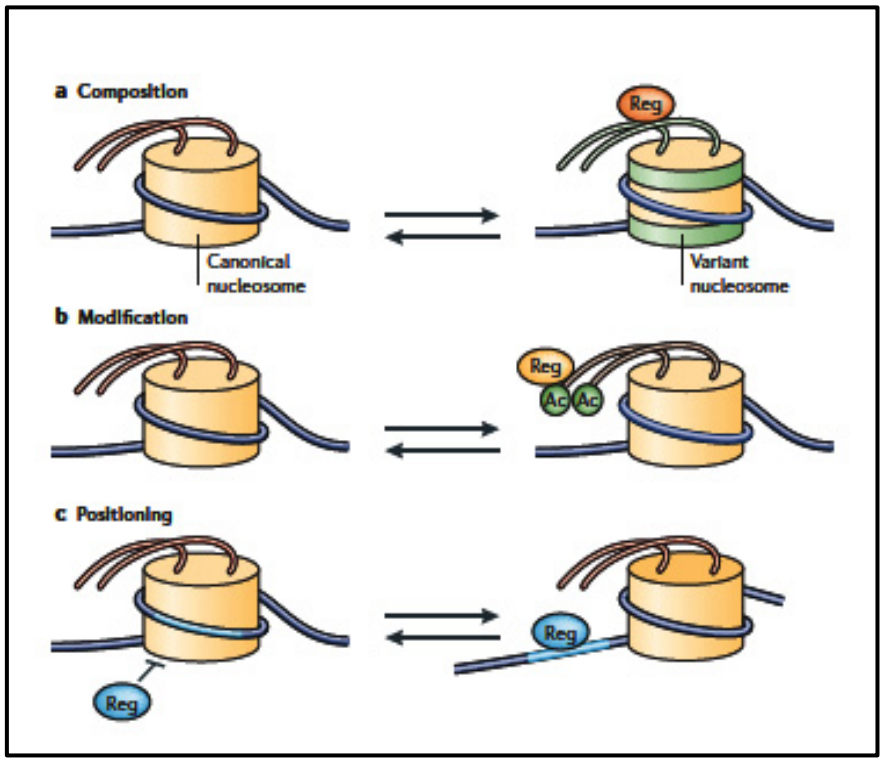


Figure 1.3: Dynamic properties of nucleosomes
a. Remodeling complexes of theSWR1 family can remove the canonical H2A–H2B dimers and replace them with Htz1–H2B dimers (indicated in green), forming a variant nucleosome with unique tails that might bind unique regulatory proteins (Reg). b. Nucleosome modification (only acetylation (Ac) depicted) allows the binding of regulatory factors, which have specialized domains (bromodomains) that recognize acetylated histone tails. c. Nucleosome repositioning allows the binding of a regulatory factor to its site on nucleosomal DNA (light-blue segment). Adapted from Saha, A et al, 2006[28].

1.4 Chromatin remodeling

All eukaryotes contain at least five families of chromatin remodelers: nucleosome remodeling factor (NURF), chromatin accessibility complex (CHRAC), and ATP-dependent chromatin assembly and remodeling factor (ACF) in Drosophila melanogaster, SWI/SFF in yeast and mammals, RSC, ISW1, and ISW2 in yeast, and RSF and WCRF in mammals[29–31]. All remodelers have one subunit in common: a conserved ATPase domain. Apart from this ATPase domain, each remodeler complex contains distinct proteins for the remodeler's unique biological role. Together, remodelers mediate several biological processes.

The SWI/SNF complex is formed by the ATPase subunit and a set of five conserved core members[32]. The SWI/SNF ATPase contains a bromodomain near its C-terminus that binds acetylated histone tails[33]. This motif in S. cerevisiae is important for the binding of the SWI/SNF complex to acetylated chromatin, thus indicating its role in either targeting or

retention of the complex at specific genomic loci[34]. SWI/SNF remodelers are known for its disordering and reorganizing nucleosome positioning to promote transcription factor binding and gene activation[35]. In contrast, ISWI complexes organize and order nucleosome positioning and thus promote gene repression[36]. The translational products are different according to the distinct remodeling mechanism of different chromatin remodeling complexes.

1.5 Histone Variants

In eukaryotes, histone genes could be grouped into two: (a) canonical histones and (b) replacement histones. Canonical histones are those with an expression peak during S phase to provide the main supply of histones during replication[37]. Genes encoding canonical histones are highly similar in sequence, have no introns and are organized in tandem, multicopy clusters. In contrast, replacement histones (histone variants) are expressed throughout the cell cycle. Genes coding for histone variants are represented by single or few gene copies and could have significant differences in their sequence[38]. Their transcripts are often polyadenylated.

Many of the histone variants are thought to alter chromatin structure because of their unique biophysical properties and/or localize to a specific genome[39]. Histone variants are also called replacement histones because of their exchange with the existing histones during development and differentiation. Due to the histone exchange, the histone variants often become predominant in differentiated cells[40]. These observations have led to the suggestion that histone variants have a unique role in chromatin regulation.

Linker histone H1 has several variants (Figure 1.4B) that have different distribution patterns among the genome[41]. Their expression levels highly vary according to the cell type as well as during development and differentiation[42]. Core histone H2A has the highest number of variants including H2A.Z and macroH2A (Figure 1.4A). While some variants like H2A.Z are evolutionarily conserved, others like MacroH2A are found only in mammals[43,44]. H2A.Z has been found to associate with many transcription factor complexes at active transcription start site (TSS) and enhancers[45]. It seems to facilitate active transcription along with transcriptional complexes[43].

Histone H2B has lesser number of variants (Figure 1.4A) and they have been found to have a role in chromatin compaction during gametogenesis[46]. Unlike other variants, sperm specific H2B variant has a long N-terminal tail with high charge that helps the condensation of chromatin fibres[47]. Interestingly, there are no variants described for histone H4.

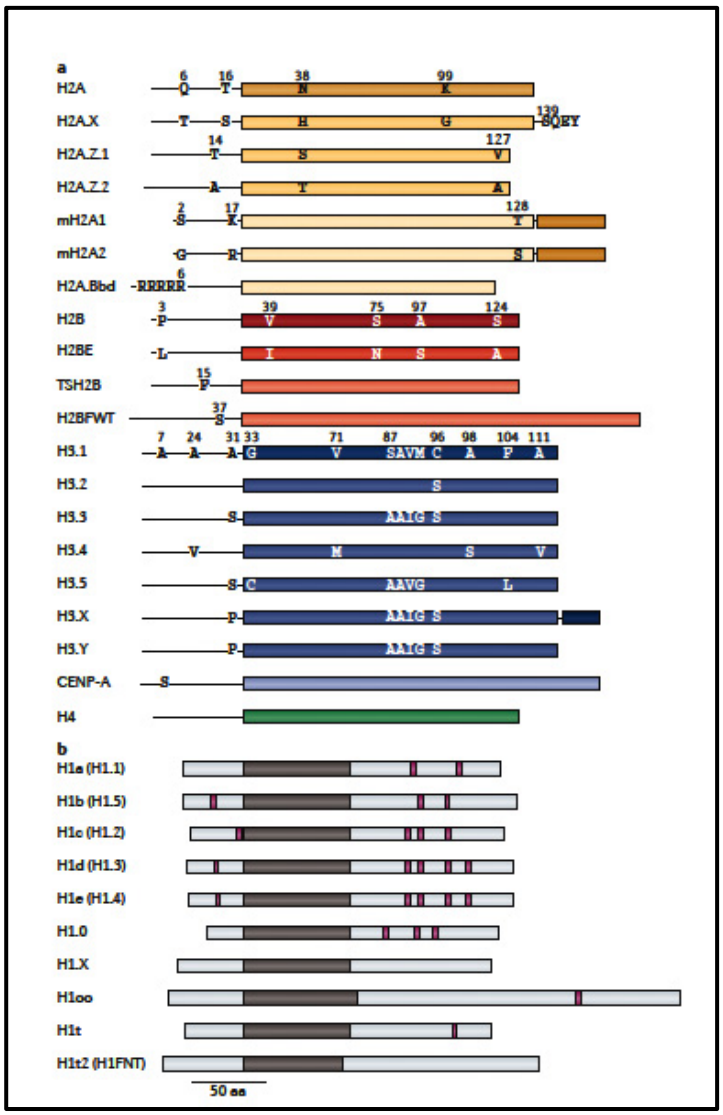


Figure 1.4: Canonical and variant histones
A. Variants of the core histones H2A (yellow), H2B (red), H3 (blue) and H4 (green) are shown. Unstructured amino-terminal tails are shown as black lines. B. Linker histone variants are shown. Unstructured amino-and carboxy-terminal domains are shown in light grey. Globular domains are shown in brown. Adapted from Maze, I et al, 2014[48].

1.5.1 Histone H3 variants

The number of H3 histone subtypes differs greatly among different species (Figure 1.5). In mammals, there are five main H3 subypes: (i) two canonical subtypes H3.1 and H3.2 or H3 and (ii) three replacement variants, H3.3, the centromere-specific variant CENP-A and the testis-specific histone H3t[49]. Two new variants, H3.X and H3.Y, which are primate specific, have been characterized recently[50]. There are sequence specifications for H3 variants (Figure 1.4a).

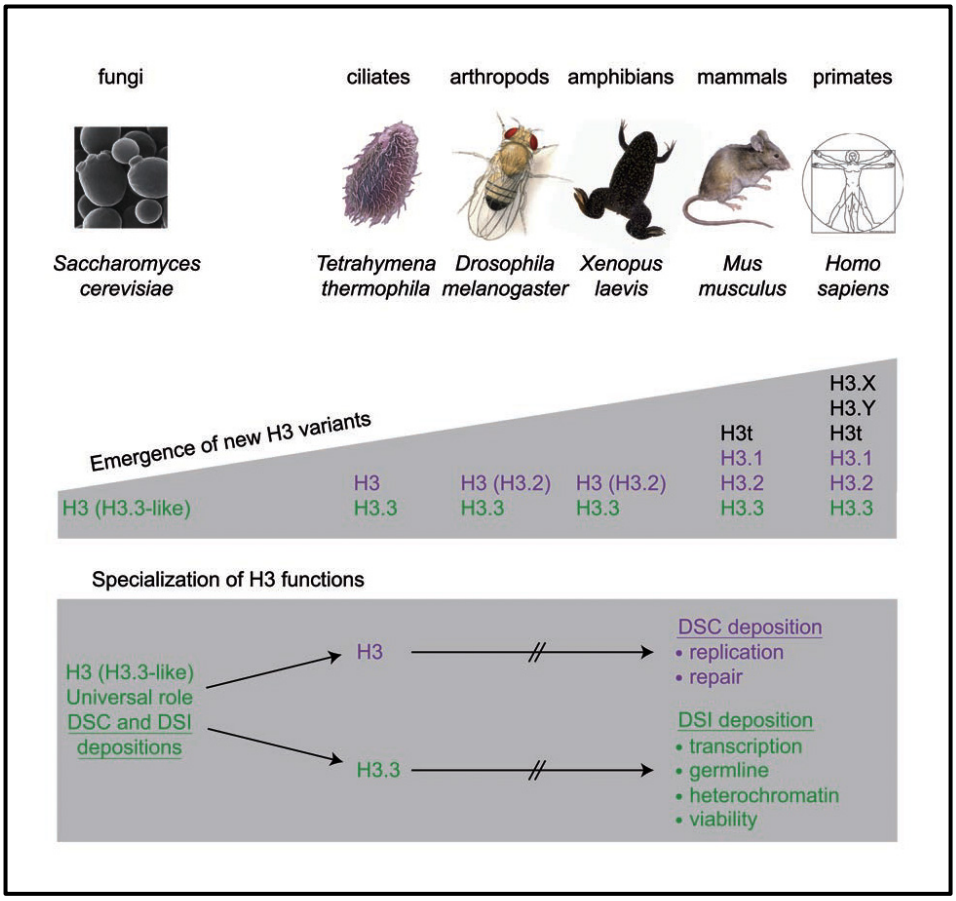


Figure 1.5: Emergence of new H3 variants through evolution.
Schematic representation of H3 and their specialized functions variants among different species. New H3 variants are derived from their ancestral H3.3 histone. Canonical histones H3.1 and H3.2 are represented in purple and variant histone H3.3 in green. Adapted from Szenker, E et al, 2011[37].

Genes encoding the replication-dependent canonical histones H3 (Figure 1.6) lack introns and are organized in multicopy clusters. Their mRNAs are not polyadenylated. Moreover, their translation is tightly regulated by the binding of the stem-loop binding protein (SLBP), and of the U7 small nuclear RNA to the 3' end of the histone RNAs[51]. The unique transcriptional regulation of histone H3 genes allows a huge production at the beginning of the S-phase,

providing a major supply for incorporation of new histones during replication in a DNA synthesis coupled manner[52]. Outside of S-phase, H3.1 can also be incorporated into chromatin at DNA damage and repair sites[53]. Genes encoding the replication independent H3 variants are characterized by single or few copies scattered throughout the genome (Figure 1.5). They have introns and polyadenylated mRNAs. In human, mouse and Drosophila, there are two genes encoding H3.3 protein, H3.3A and H3.3B[54,55]. They both have distinct 5' and 3' untranslated regions[56]. They are expressed throughout the cell cycle allowing histone deposition and exchange through a DNA synthesis-independent pathway during and outside S phase[57].

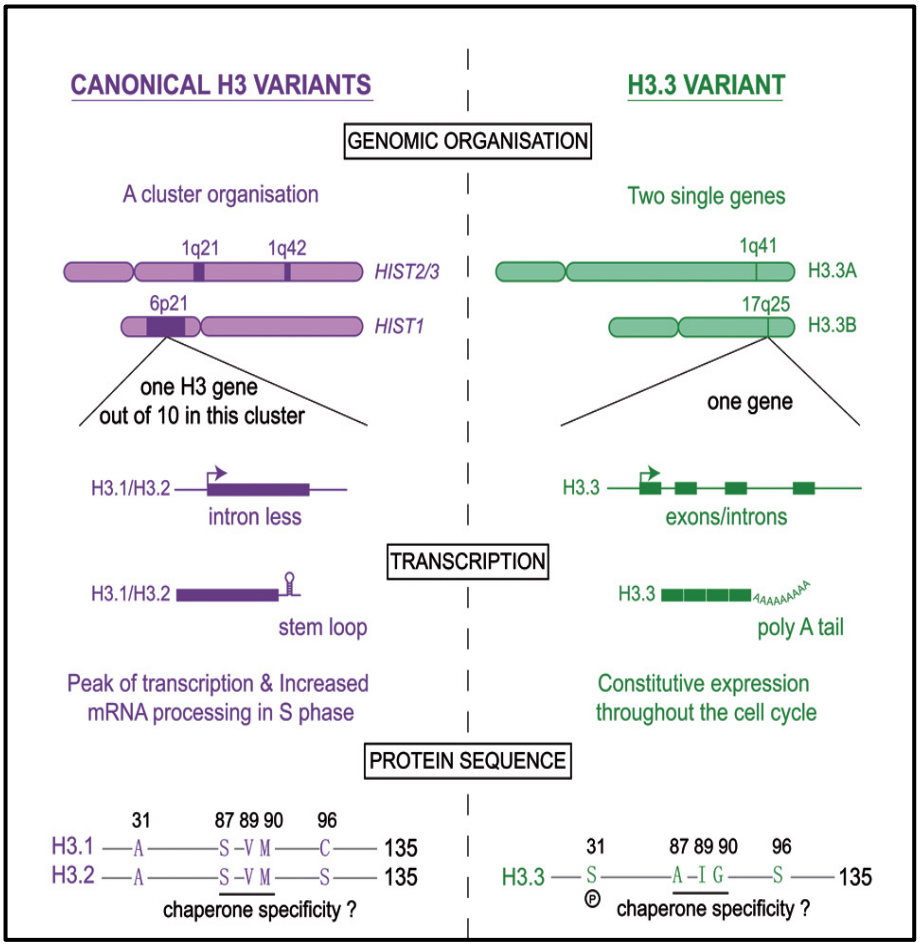


Figure 1.6: Comparison between human H3 and H3.3 variants
Differences between the canonical H3 variants (H3.1 and H3.2) (in purple) and the replacement variant H3.3 (in green) are illustrated. Canonical histone genes are organized in tandem and the cluster HIST1 located on chromosome 6p21 contains core histone genes including 10 histone H3 genes. Canonical histone genes lack introns and are not polyadenylated in contrast to the regular genes coding for H3.3 (H3.3A and H3.3B). The amino acid differences between the canonical H3 and H3.3 are illustrated. Adapted from Szenker, E et al, 2011[37].

In human (Figure 1.6), the two H3.3 genes, H3.3A and H3.3B are located on chromosomes 1 and 17 respectively. The thirteen canonical histone genes encoding for histones H3.1 and H3.2 are localised in three histone gene clusters Hist1, Hist2 and Hist3 on chromosomes 6, 1 and 1 respectively. In mice, the two H3.3 genes, H3f3a and H3f3b are located on chromosomes 1 and 11 respectively. There are twelve genes encoding canonical histones H3.1 and H3.2 located in two histone gene clusters hist1 and hist2 on chromosomes 13 and 3 respectively[51]. The histone gene cluster hist1 contains 51 histone genes in total, which is spread over about 2 Mb[51] (Figure 1.7). Hist1 of human and hist1 of mouse are identical to each other.

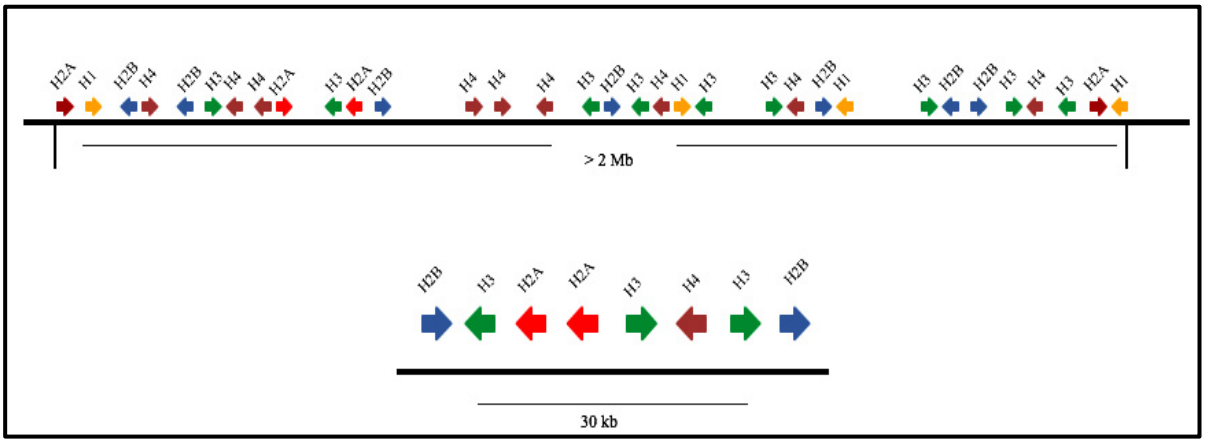


Figure 1.7: Schematic representation of organization of the histone gene clusters
The histone clusters hist1 (upper panel) and hist2 (lower panel) are located on chromosome 13 and 3 respectively**.** The locations of the histone genes are indicated at the top by horizontal arrows with the appropriate designations. These arrows also indicate the directions of transcription.

In contrast to mammals, Gallus gallus has only three H3 variants: (i) H3.2 or canonical histones H3 (ii) H3.3 and (iii) CENP-A, missing mammalian type H3.1. The canonical histone H3 genes are organized in one single histone gene cluster on chromosome 1 (Figure 1.8) spanning 110kb in length. The histone gene cluster comprises of 39 histone genes including seven H3.2 genes. Two H3.3 genes, H3.3a and H3.3b are present on chromosome 3 and 18 respectively[58,59]. Despite considerable divergence of the cDNA sequence of H3.3a and H3.3b, they encode identical proteins, which also have the same sequence as human H3.3.

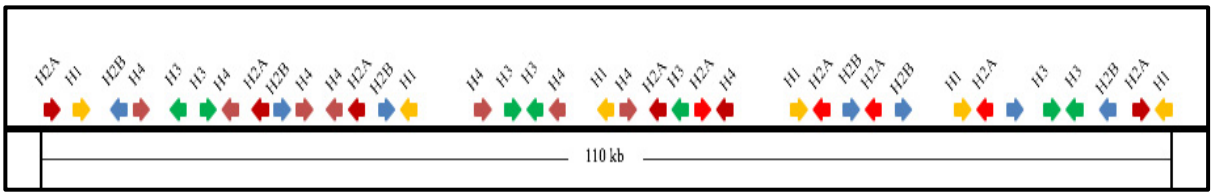


Figure 1.8: Overall organization of chicken genome histone gene cluster
Horizontal arrows with the appropriate designations indicate the locations of the histone genes on chromosome 1. These arrows also indicate the directions of transcription.

The replacement variant CENP-A is an essential protein that binds to centromeres mediating chromosome segregation in eukaryotes[60]. They have conserved histone fold domain and divergent N-terminal tails. Interestingly, H3t, H3.1, H3.2, and H3.3 are the least divergent variants/subtypes with only four amino acid differences among one another.

Several studies reveal that the histone variant H3.3 is enriched at the promoter region as well as throughout the gene body of actively transcribed genes in euchromatin[61–64]. This enrichment of H3.3 at promoters has not only been observed at actively transcribed genes but also at inactive genes[65,66]. The histone regulator A (HIRA) is the histone chaperone mediating the H3.3 deposition in euchromatin. In heterochromatin, H3.3 has been found enriched at pericentromeric heterochromatin and telomeres, though telomeric enrichment has been observed only in ES cells so far[57]. The chromatin remodeler death domain-associated protein (DAXX) complexes together with alpha-thalassemia/mental retardation X-linked syndrome protein (ATRX) assist in the deposition of heterochromatic H3.3[57]. The deposition pathway of H3.3 is completely different from that of the canonical histones that are mediated by the chromatin assembly factor-1 (CAF-1)[57,62]. This might actually have an impact on the fate of histones as well as their metabolism.

Studies on H3.3 genes, H3.3A and H3.3B in D. melanogaster have provided key insight into role of H3.3 in chromatin biology[54,67]. Each of the two genes is independently not required for fly development, likely due to redundancy as H3.3 double mutants show stronger phenotypes that include reduced viability and infertility[67]. Moreover, H3.3 null flies exhibit male meiotic defects involving impaired chromosome segregation[67]. In mice, both single gene knockouts and knockdowns as well as double knockouts have been reported [55,64,68]. Studies on homozygous KO of either H3.3A or H3.3B show reduced viability and male infertility phenotypes[55,68] whereas siRNA knockdown of H3.3 in mouse zygotes show nuclear envelope formation defect and changes in chromosome condensation[69,70]. Interestingly,

studies show that deleting both H3.3A and H3.3B resulting in a total KO of H3.3 shows embryonic lethality and developmental retardation[64]. H3.3 null embryos show increased cell death and reduced cell proliferation whereas H3.3 deletion in ES cells triggered rapid cell death. Also, H3.3 deletion causes dysfunction of heterochromatin structures at telomeres, centromeres, and pericentromeric regions of chromosomes, leading to mitotic defects[64]. This defines a major function for H3.3 in mammalian development is maintaining genome integrity.

In human, mutations on amino acids K27 (K27M) and G34 (G34R/G34V) of H3.3A have been found in pediatric glioblastoma tumors[71–75]. H3.3 related glioblastoma cancers have also been found to possess mutations in DAXX or ATRX[72]. In one of the studies involving sequencing of the exomes of 48 pediatric glioblastoma samples[73], 44% of tumors were identified with H3.3-ATRX-DAXX mutations and 31% with H3.3A K27M and G34R/G34V mutations suggesting that H3.3 dependent defects of the chromatin architecture underlie pediatric and young adult glioblastoma pathogenesis.

1.6 Post-translational modifications

Studies in the chromatin field over the past few decades showed that covalent modifications of core histones, linker histones and DNA could alter chromatin architecture thus regulating transcription, DNA repair and replication. Histones and DNA modifications are set and removed by chromatin modifying enzymes in a regulated manner[76]. These modifications, in addition to cause direct impact on chromatin structure also then serve as docking sites for chromatin "readers", which specifically recognize these modifications[77,78]. The chromatin readers can themselves be chromatin modifiers and remodeling enzymes creating a feed forward mechanism (Figure 1.9). Ever since the discovery of histone acetylation[79], there has been a growing list of histone post-translational modifications (PTMs) including phosphorylation, methylation, ubiquitination, SUMOylation, and GlcNAcylation[66-71].

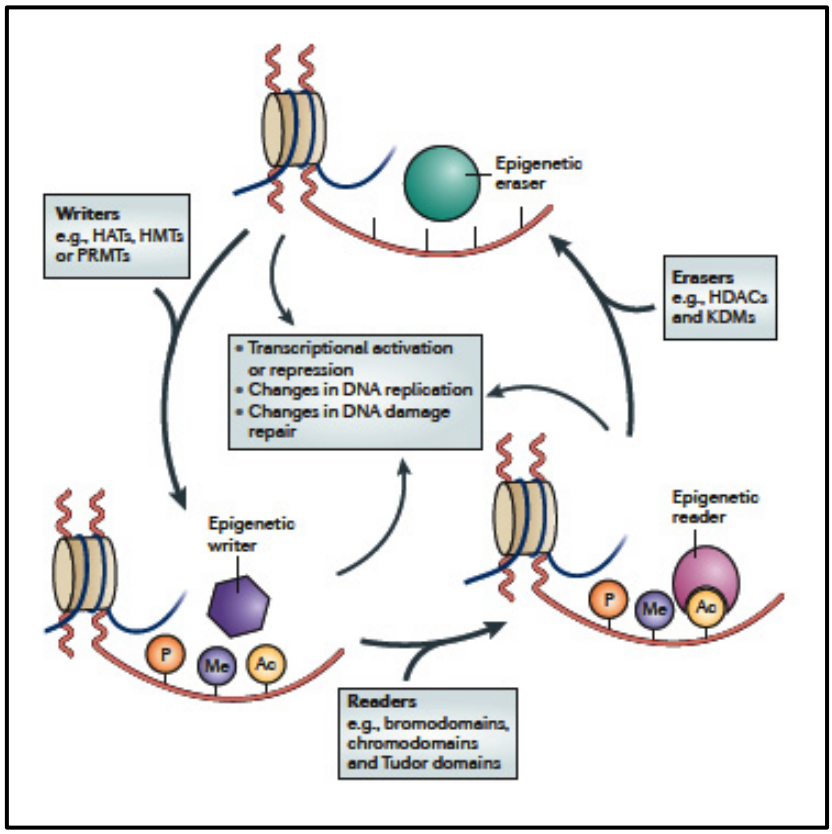


Figure 1.9: "Epigenetic" readers, writers and erasers.
Epigenetic regulation is a dynamic process. Writers such as histone HATs, HMTs, PRMTs marks residues on histone tails. Epigenetic readers such as proteins containing bromodomains, chromodomains and Tudor domains bind to these epigenetic marks. Erasers such as HDACs, KDMs and phosphatases help in removing epigenetic marks[82].

In general, acetylation of histones is associated with transcriptionally active chromatin[83]. So far the list of identified N terminal acetylated residues include H3 lysine residues 9 and 14, and H4 lysine residues 5,8,12 and 16[84]. Acetylation on lysine residues neutralizes its positive charge and weakens the histone-DNA affinity thus creating a more accessible and open chromatin state. It is catalyzed by acetyltransferase(s) (HATs) and removed by deacetylase(s) (HDACs)[85].

The first histone acetyltransferase GCN5 in S. cerevisiae was identified in 1996[86]. Ever since, many HATs have been described including TFIID (transcription factor IID) subunit TAFII250 (TATA-binding protein associated factor) and p300/CBP (CREB binding protein)-associated factor (pCAF)[87,88]. Though HATs are diverse, many of them acetylate histones at enhancers and promoter regions, thus help to induce active transcription. HDACs on the other hand specifically remove acetyl groups from acetylated histones, thus enabling transcriptional

repression of genes. Until now, 18 HDACs have been identified and grouped into four major classes: (a) HDAC class I containing HDACs 1,2,3 and 8 (b) class II with HDACs 4 to 9 (c) class III containing sirtuins (SirT1-7), and (d) class IV with HDAC 11[89]. Targeting these HDACs has emerged as a potential therapeutic strategy for treatment of disease, including cancer, immune disorders, and heart disease.

Methylation of histones is associated with either transcriptional activation or repression[90]. Histone methylation largely occurs on arginine and lysine residues and is regulated by histone methyltransferases (HMTs) and histone demethylases[91]. There are three main families of proteins of HMTs: (a) PRMT1 family (b) SET domain containing protein family, and (c) non-SET domain DOT1/DOT1L[92]. These enzymes can target lysine and arginine residues of histones. Lysine residues can be mono, di and/or tri methylated. H3 lysine 4-tri-methylation (H3K4me3) is spatially enriched at transcription start site (TSSs) of active genes. PTMs linked to gene silencing include H3K9me2, H3K9me3, H3K27me2 and H3K27me3. H3K27me3 is associated with the repression of unwanted differentiation programs during lineage specification7. Histone demethylases are enzymes that specifically demethylate histones. These demethylases include lysine specific demethylase (LSD1) which demethylates H3 lysine 4 residue and the Jumonji C family (JMJD), each of which targets specific methyl lysine groups[93]. In mammals, the silent state of heterochromatin is associated with high levels of H3K9, H3K27 and H4K20 methylation and low levels of acetylation[94]. H3K9 methylation is believed to play an important role in maintaining pericentromeric heterochromatin by interacting with heterochromatin protein HP1[95].

1.7 Histone code hypothesis

Post-translational modifications of histone N terminal tails work require factors that set these modifications (“writers”), “readers” that interprets these marks and factors that remove (“erasers”) the modifications. According to the so called histone code hypothesis[77,85,96–98] these factors regulate together distinct functional outputs such as e.g. transcriptional regulation,.. In Drosophila melanogaster, distinct chromatin states were identified based on 18 histone modifications and grouped into 9 combinatorial states called the 9-state model[99]. Studies suggest that different histone modifications work synergistically to achieve distinct chromatin states.

Several studies provide us the evidence that histone modifications such as methylation, acetylation or phosohporylation can influence each other either synergistically or in an antagonistic way. This could provide a mechanism to generate and stabilize specific imprints or dynamically mark chromatin loci upon events such as DNA damage and repair, mitosis, and gene activation[100].

1.8 PTMs on globular domain of H3

Until now, studies have been centred on histone tails, which serve as platforms for a variety of PTMs[101]. These histone tail PTMs have the potential to regulate chromatin function, by recruiting or repelling specific binding or ‘‘effector’’ proteins. However, PTMs have also been detected within the globular domains of histones[102–104] and multiple map to residues located on the outer, lateral surface of the histone octamer that are in direct or water-mediated contact with the nucleosomal DNA[105]. Within the globular domain of H3 lie four lysine residues that map to the lateral surface: Lys56, Lys64, Lys115, and Lys122. All four lysine residues have found to be acetylated and are in direct contact with the DNA (Figure 1.10). Since lysine acetylation neutralizes its positive charge, this modification has the potential to weaken the histone-DNA interaction[106]. Based on the critical location of these lysine residues, a model of ‘‘regulated nucleosome mobility’’ was proposed[97] according to which modifications on the lateral surface could, in contrast to most PTMs in the histone tails, directly affect nucleosome and chromatin dynamics without depending on recognition by effector proteins.

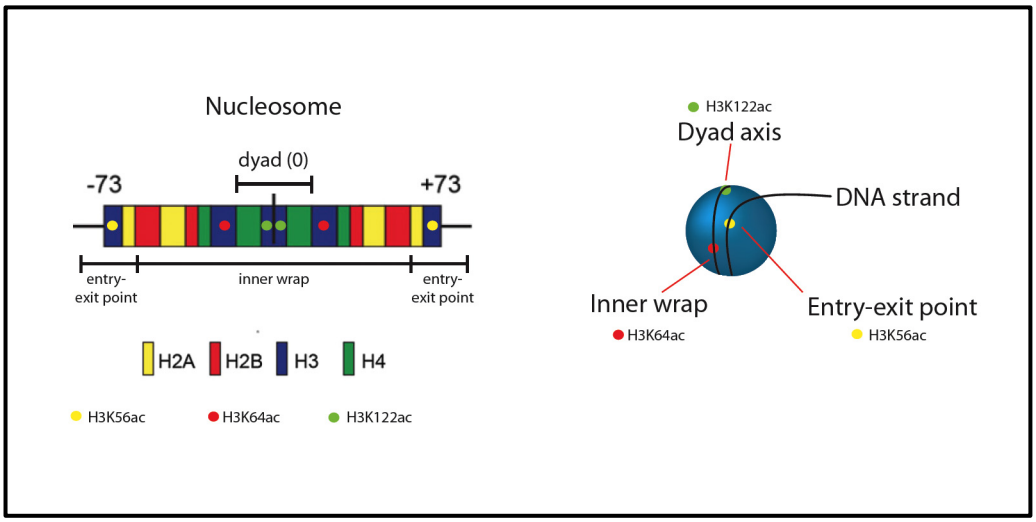


Figure 1.10: Position of lysine residues on the nucleosome octamer lateral surface
On the left is the schematic representation of histone-DNA contacts on nucleosome shown as color-coded rectangle showing the globular domain modifications. On the right is a cartoon of a nucleosome showing H3K56ac, H3K64ac and H3K122ac on the lateral surface of the histone octamer. Adapted from Bintu et el, 2012[107].

One of the first globular domain PTMs identified was acetylation of Lys56 on histone H3 (H3K56)[108–110]. H3K56, localizes to the DNA entry and exit points (Figure 1.10)[12]. It was reported that H3K56 acetylation is highly abundant in *S. cerevisiae*, up to 40% of all H3 molecules are acetylated at K56[109]. In *S. cerevisiae*, H3K56 acetylation is highly dynamic during cell cycle, with its highest abundance during S-phase, and it rapidly decreases when cells enter the G2/M phase of the cell cycle[109]. H3K56ac is involved in transcriptional regulation and DNA damage response[110]. Multiple studies have suggested that H3K56 acetylation might influence nucleosomal stability by enabling the breathing of nucleosome, thus affecting chromatin dynamics[111]. Consistent with this study, H3 is found to be hyper-acetylated at actively transcribed genomic regions and hypo-acetylated in silent regions including repetitive elements[112]. In a recent study[113], mutant nucleosomes containing H3 lysine 56 to glutamine (H3K56Q) substitution (known to mimic the acetylated states of specific histone lysine residues *in vivo)* were crystallized and shown to not from water-mediated contacts between the histone residue and the DNA as found in unmodified nucleosomes. Taken together, these findings suggest that H3K56 acetylation enhances the unwrapping of the DNA close to the DNA entry–exit site (position of H3K56) to regulate DNA accessibility. In mammals, however H3K56ac levels are very low. H3K56 acetylation in mammals is set by p300/CBP and requires the histone chaperone Asf1[108]. H3K56

deacetylation is regulated by SIRT1, member of sirtuin family of NAD+-dependent deacetylases[108].

Acetylation of H3K64 (Figure 1.11), located in proximity of the entry-exit points of a nucleosome) has shown to impact nucleosomal stability[103] . Using fluorescence resonance energy transfer (FRET) approach, which measured the salt dependent stability of nucleosomes, it was demonstrated that completely H3K64 acetylated nucleosomes are less stable than unmodified ones. H3K64 acetylation also reduces histone octamer interactions with DNA. Consistent with the destabilization of nucleosomal structure, this modification is shown to be enriched in euchromatin and associated with active regulatory genomic regions[103]. It can also promote histone eviction thus facilitating transcription. Both *in vitro* and *in vivo*, H3K64 acetylation is set by the histone acetyltransferase p300/CBP[103].

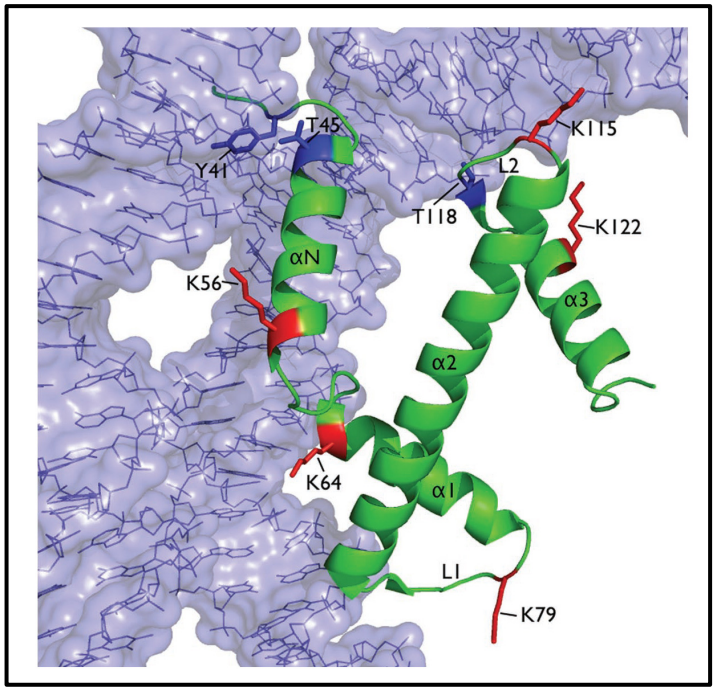


Figure 1.11: Structure of the H3 globular domain within nucleosomal DNA
H3 is shown in green, modifiable residues are highlighted in blue for tyrosine/threonine or red for lysine. The picture was generated with PyMOL41 based on PDB code 1KX3. Adapted from Tropberger, P and Schneider, R, 2010[114].

Acetylation of Lys115 and Lys122 (Figure 1.11) on histone H3 (H3K115 and H3K122) were first identified to occur either individually or in combination based on mass spectrometric analysis of bovine histones[115]. H3K115 and H3K122 are located on the dyad (symmetry) axis of the nucleosome, where the interaction between histones and DNA is strongest (Figure

1.10)[116]. Studies on H3K122 acetylation link this modification directly with nucleosome stability. Acetylation of H3K122 is genome-wide enriched at the TSS of transcribed genes[102]. Biophysical studies on H3K115 and H3K122 acetylated nucleosomes showed that these modifications reduce the affinity of the histone octamer for DNA and slightly increase thermal repositioning of nucleosomes[117]. Several studies suggest that both these PTMs are involved in altering silencing of genes at telomeres, DNA repair, stimulate transcription *in vitro* and regulate nucleosomal stability[102,111,118]. H3K122 acetylation is also shown as a chromatin signature of active enhancers[102]. Mutant nucleosomes containing H3K122Q leads, as acetylation to a loss of water-mediated contacts between the histone residue and the DNA[113]. A recent study demonstrated that H3K115 or H3K122 acetylation could enhance nucleosome disassembly rates of RSC and SWI/SNF complex. Taken together, H3K115 and H3K122 acetylation are suggested to modulate nucleosome structure and dynamics directly. It is shown that H3K122 acetylation is catalyzed by p300/CBP in response to estrogen receptor signaling both *in vitro* and *in vivo*[102].

1.9 Genome editing

In recent years, transcription modulation *in vivo* is achieved by a variety of biotechnological tools that facilitate target genome editing and thus help to elucidate the genetic and epigenetic basis of various biological functions and diseases. It was 1972 in the laboratory of Paul Berg where genetic engineering emerged in the form of recombinant DNA technology, when scientists combined the *E. coli* genome with the genes of a bacteriophage and the SV40 virus[119]. Since then tremendous discoveries have been made to achieve precise gene editing to control gene expression *in vivo*.

Gene editing has taken advantage of the discovery that targeted DNA double strand breaks (DSBs) could be used to stimulate the endogenous cell repair machinery[120]. DNA breaks are typically repaired through either one of the two major pathways (Figure 1.12): (i) Homology directed repair (HDR) or (ii) non-homologous end joining (NHEJ)[120]. In the presence of a homologous sequence of DNA, HDR takes place in a template-dependent manner during G2 and S-phase of the cell cycle[121]. Introducing targeted DSBs along with the supply of exogenous donor template could activate HDR pathway and induce precise gene editing. This exogenous donor template serves as the repair template[122]. Thus, any sequence changes in the donor template including point mutations and sequence insertions could be incorporated at

the targeted endogenous locus. In the absence of a homologous DNA sequence, NHEJ functions to repair DSBs through re-ligation of the cleaved end[121].

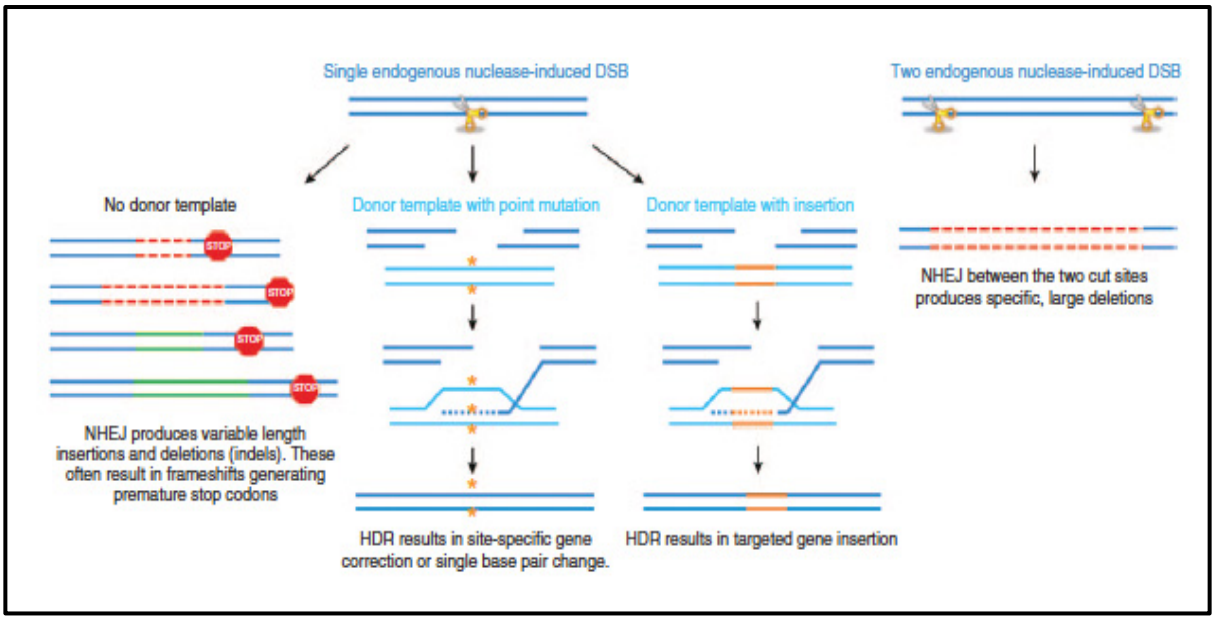


Figure 1.11: Mechanism of Double strand breaks (DSBs).
DSB is repaired by either of two mechanisms: (a) non-homologous end joining NHEJ (on the left) (b) homology directed repair HDR (on the right). Specific donor template with the desired mutation or insertion is provided for the template directed repair to take place. Adapted M Maeder and C A Gersbach[122].

1.10 Gene editing by Homologous Recombination

The development of gene targeting technology in yeast, the chicken DT40 B cell line, mouse, and ES cells has been a major breakthrough leading to the generation of various mutant cell models[123–127]. It uses homologous recombination (HR) technique that involves the exchange of DNA sequence between endogenous locus and exogenous donor template via crossover events[125]. It can be used to create an insertion, mutation or deletion on the endogenous DNA sequence. Mutagenesis strategies including Flp/Frt and Cre/LoxP systems are widely used to provide cell/tissue specific gene deletions[128,129]. Cre/LoxP is a site-specific recombination system where Cre recombinase stimulates recombination events between two DNA recognition sites of 34 bp known as LoxP sites. Genomic regions that are flanked by loxP sites in the same orientation, opposite orientation or on different strands result in deletion, inversion or translocation of DNA sequence in cells expressing Cre recombinase[128].

Though HR is the most widely used gene editing technique, its limitation to certain species and lower efficiency led to the development of new techniques such as ZFNs, TALENs and CRISPR/Cas9.

1.10.1 Gene editing by ZFNs and TALENs

Zinc finger nucleases (ZFNs) facilitate gene editing through the introduction of DSB at the target DNA sequence thus enabling the desired DNA sequence modification during DNA repair[130].

TALEs are natural bacterial effector proteins used by Xanthomonas sp. to modulate gene transcription in host plants to facilitate bacterial colonization. TALEs can be designed in a way that they can be specific for almost any target site within a genome[122]. Like ZFN, TALENs also introduce DSB at desired sites thus inducing either NHEJ or HDR as DNA repair pathway.

1.10.2 Gene editing by CRISPR/Cas9

In addition to ZFNs and TALENs, a more versatile tool known as CRISPR/Cas9 has been developed for genome editing. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) genes are essential in adaptive immunity in select bacteria and archaea, enabling the organisms to defend against foreign genetic elements[122,131]. Three types of CRISPR systems (I, II, and III) have been identified where each system comprises a cluster of CRISPR-associated (*Cas*) genes, noncoding RNAs and a distinctive array of repetitive elements (direct repeats). These repeats are interspaced by short variable sequences derived from exogenous DNA targets known as protospacers, and together they constitute the CRISPR RNA (crRNA) array. Within the DNA target, each protospacer is always associated with a protospacer adjacent motif (PAM), which can vary depending on the specific CRISPR system[132,133]. The Type II CRISPR, derived from *Streptococcus pyogenes* is one of the best studied[131], containing a nuclease Cas9, the crRNA array that encodes the guide RNAs and a required auxiliary trans-activating crRNA (tracrRNA) that facilitates the processing of the crRNA array into discrete units. Each crRNA unit consists of a 20-nt guide sequence and a partial direct repeat, where the former directs Cas9 to a 20-bp DNA target via

Watson-Crick base pairing (Figure 1.12). The only restriction for this system is that the target DNA must immediately precede a 5′-NGG PAM sequence.

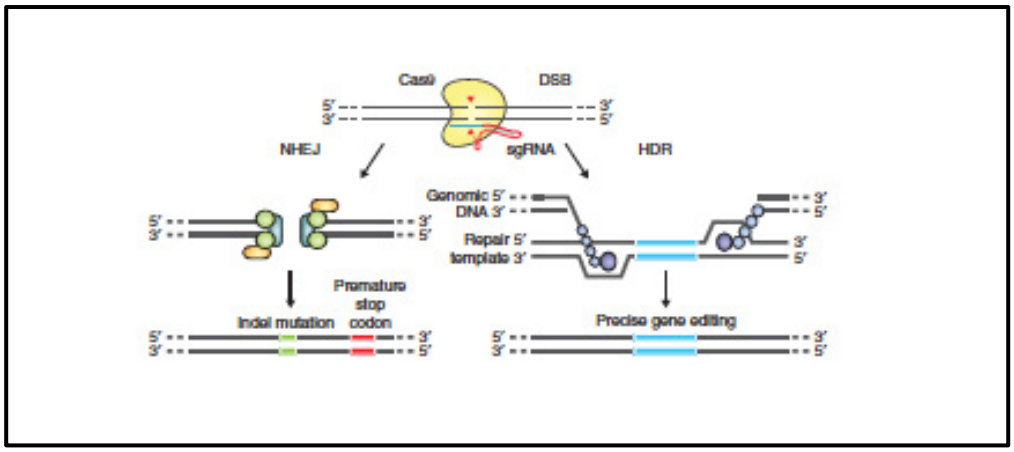


Figure 1.12: Gene editing using CRISPR/Cas9.
DSB repair promotes genome editing. DSB is induced by Cas9 and is repaired by either of two mechanisms: (a) non-homologous end joining NHEJ (on the left) (b) homology directed repair HDR (on the right). Specific donor template with the desired mutation or insertion is provided for the template directed repair to take place. (Adapted from H Yang et al, 2013[131])

CRISPR/Cas9 provides several potential advantages over ZFNs and TALENs such as higher targeting efficiency and the ability to multiplex genome editing[134]. It also does not require the engineering of novel proteins for each target site. Though studies report off-target DNA cleavages by Cas9, the use of modified Cas9 enzyme (containing only one active catalytic domain unlike two in the wild-type), known as the Cas9 nickase (cutting one of the DNA strands, resulting in a “nick”, or single strand break) could significantly reduce them[133]. The emergence of CRISPR and rapidly developing Cas9 hold a great promise in targeted genome editing. The easy design, construction and high efficiency of Cas9 make it a versatile tool for studying the function of underlying genetic and epigenetic mechanisms of candidate genes and transcriptional regulatory elements.

2 Aims

All of the above-mentioned studies on histone modifications are based on *in vitro* approaches or overexpression of mutated histones in a background of endogenous wild type histones. We therefore wanted to establish a "clean" system where it is possible to study the effects of histone modifications in a mammalian cellular system. We decided to focus on novel modifications within the core of the nucleosome and compare them with tail modifications. One of the main open questions in the field is also whether different histone modifications act in combinatorial manner thus affecting gene transcription. I hypothesized that histone core modifications H3K56, H3K64, H3K122, and H3K115 acetylation might act synergistically and/or in different combinations, increasing the impact on nucleosome dynamics and hence transcription. These novel histone core acetylation(s) that I am studying can extend our understanding of chromatin regulation and provide novel mechanistic insights into how chromatin organization is spatially and temporally controlled. The identification of novel pathways regulating chromatin function can pave way for new drug targets in diseases such as cancer in which chromatin states have been distorted.

In order to study the impact of PTMs *in vivo*, all endogenous WT H3 gene copies have to be replaced with mutant copies that can mimic or prevent PTMs. Even though total H3 replacement has been achieved in D. Melanogaster in the recent past[135], low amount of obtainable cells limit its application in biochemical studies. I managed to overcome these limitations and to use mammalian system by employing mouse embryonic stem cells. Latest advances in genome editing such as the CRISPR system has been used to rapidly modify endogenous genes in cells and organisms that so far have been challenging to manipulate genetically[132]. Therefore, establishing H3 mutant ES cell lines without the background of endogenous H3 became feasible.

2.1 Part I – Establishment of histone mutant cell lines

The first main objective of my work was to establish mouse ES cell lines that exclusively express mutated H3 where specific lysine residues (in the H3 tails and core) are mutated to arginine (K/R) that cannot be acetylated (mimicking constitutively un-acetylated H3) or by mutating lysine to glutamine (K/Q) to mimic an acetylated lysine.

2.2 Part II –Impact of histone mutants

The second main objective of my project was to study the combined effects of acetylation of H3K122, H3K56, H3K64, and H3K115 on: (a) gene transcription e.g. whether the combined effects of K56/64/115/122ac enhances transcriptional gene activation than a single modification, (b) ES cell differentiation and (c) to compare them with tail modifications.

3 Results

3.1 Chicken B cell line DT40

To study the impact of individual and combined modifications on the lateral histone octamer surface the aim of my thesis was to establish a system where all histone proteins are mutated at H3K56, H3K64, H3K115, and/or H3K122 without background of endogenous WT H3. For this all the endogenous wild type H3 genes have to be replaced with mutant copies (either K to Q to mimic constitutive acetylation or K to R to preventing acetylation). Such as system had only been established in D. melanogaster cells, a system not allowing biochemical studies[135]. I therefore chose first the chicken B cell line DT40 as a system because of its exceptionally high ratio of targeted to random DNA integration by homologous recombination[136]. Also, there are lesser histone H3 genes in chicken than in mammals.

To first generate DT40 cell lines with a reduced number of endogenous H3 genes, I devised the following two strategies for histone H3.3 and H3.2 respectively (Figure 3.1.1): (a) To knockout (KO) all H3.3 alleles up to one heterozygous allele and in a second step to knock-in (KI) H3 mutants in the remaining copy to study the effects of H3.3 mutations (with an endogenous H3.2 background) (b). To perform a heterozygous KO of all H3.2 alleles on chromosome 1, delete H3.2 on the other allele to minimum number of copies with which the cells can survive and then knock in mutant H3 in remaining H3 copies to study the effects of H3.2 mutations.

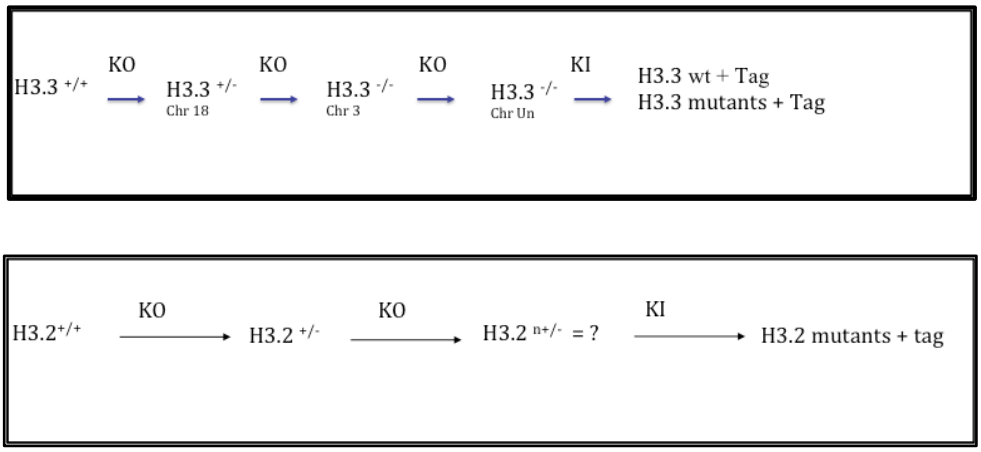


Figure 3.1.1: Strategy for establishing a H3.3 mutant system (upper panel) and H3.2 mutant system (lower panel) using DT40 cell line. $H3.2^{n+/-}$ is the H3.2 KO clone where 'n' represents the minimum number of H3.2 copies with which the DT40 cells will survive.

To generate DT40 knock out cell lines I used a mutant loxP system. This system is composed of two different mutant loxP sites (loxP_RE and loxP_LE). Cre recombinase can still recognize these mutant loxP sites and delete the intervening region. This process creates the loxP_RE+LE site, which is poorly recognised by Cre[137]. This has the advantage that genetic instability associated with re-cutting of the remaining loxP_RE+LE site is minimized.

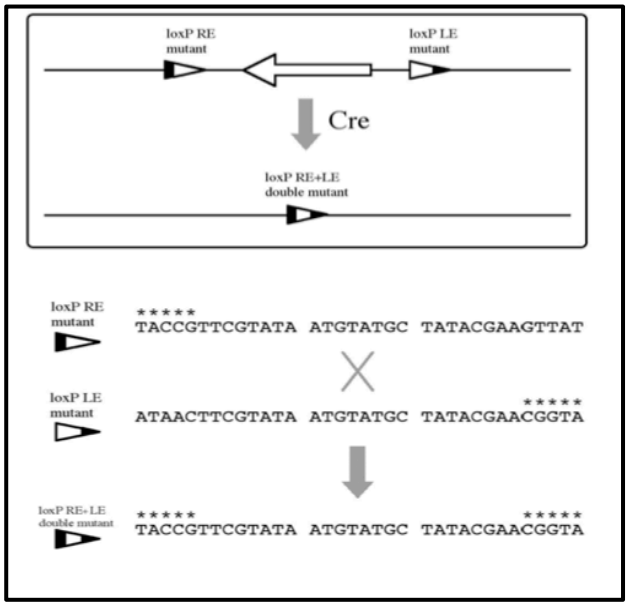


Figure 3.1.2: Mutant loxP system. The scheme of mutant loxP system is shown with sites for the rearrangement of mutant loxP sites. Nucleotide changes in mutant loxP sequences are highlighted with stars Adapted from Arakawa et al, BMC Biotechnology[137].

To establish a H3.3 on chr 18 (H3.3b) heterozygote KO cell lines, DT40 cells were transfected with ΔH3.3b/pLoxNeo targeting vector containing mutant loxP sites using (Figure 3.1.2) and homologous arms to the regions flanking H3.3b. In this targeting vector, neo transcribed by the chicken β-actin promoter is flanked upstream and downstream by the 2 kb and 2.3 kb fragments derived from both the 5' and 3'regions of the target genes respectively (Figure 3.1.3).

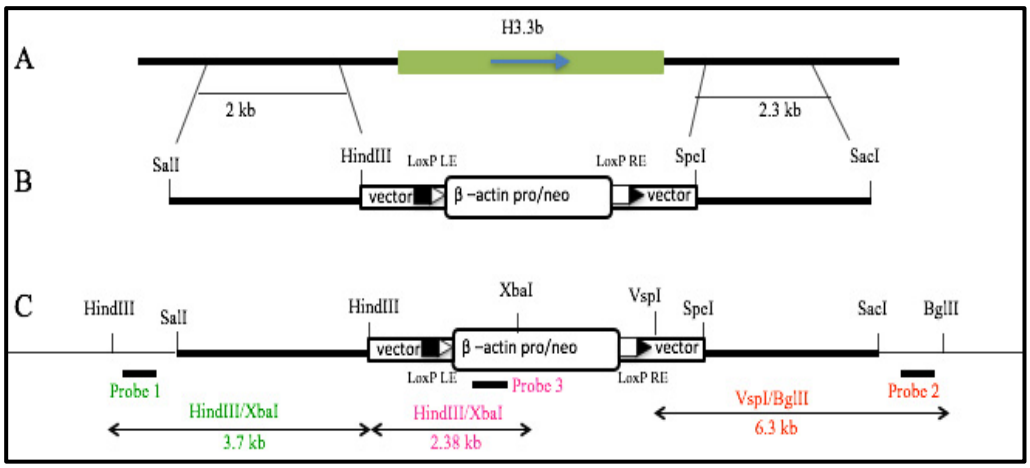


Figure.3.1.3: Schematic diagram of the homologous recombination resulting in deletion of one allele of H3.3b gene.
A. Location of H3.3b gene in DT40 genome along with the direction of transcription. B. Targeting ΔH3.3b/pLoxNeo construct. The boxes represent Neo under the control of chicken β-actin promoter. C. The locus surrounding the H3.3 gene in DT40 clones after targeted integration of the ΔH3.3b/pLoxNeo construct. Bars along with the expected sizes of the resulting fragments indicate the locations of probes 1, 2 and 3 after restriction enzyme digestion.

All the resulting cell clones were expanded and analysed using southern blotting. The membranes was hybridized (Figure 3.1.3) with: (i) probe 1 after HindIII/XbaI digestion resulting in a fragment size of 3.7 kb after the KO and 7.1 kb for the wild-type (WT) (ii) probe 2 after BglII/VspI digestion resulting in a 6.3 kb fragment after the KO and 7.5 kb for the wild type and (iii) probe 3 derived from the vector, that will hybridize to the HindIII/XbaI fragment of 2.38 kb. All the three probes were initially labelled using DIG for southern blotting. The digoxygenin (DIG) system is a non-radioactive labeling method that uses DIG, a steroid hapten, to label DNA probes for hybridization. Several attempts to optimize southern blotting using the DIG system were unsuccessful despite efficient labeling of the

probes (Figure 3.1.4). It might be that DIG labeling method is not sensitive enough to detect the DNA fragments. Therefore, I switched to using radioactively labelled probes.

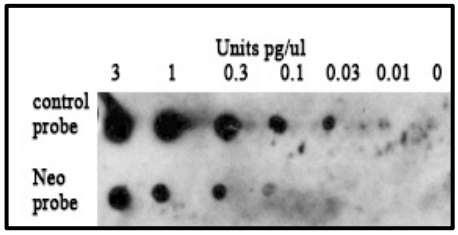


Figure 3.1.4: Dot blot showing yield of DIG-labelled DNA.
A series of dilutions of DIG-labelled DNA is applied to a positively charged nylon membrane. The intensities of the dilution series of labelled control DNA probe (provided with the kit) and NEO probe are compared. Visibility of 0.1 pg dilution spot of the probes suggest the expected labeling efficiency.

Radioactively labelled probe 3 hybridized to the HindIII/XbaI fragment of 2.38 kb in 25 of the 26 clones suggesting that the KO vector has been integrated (Figure 3.1.5).

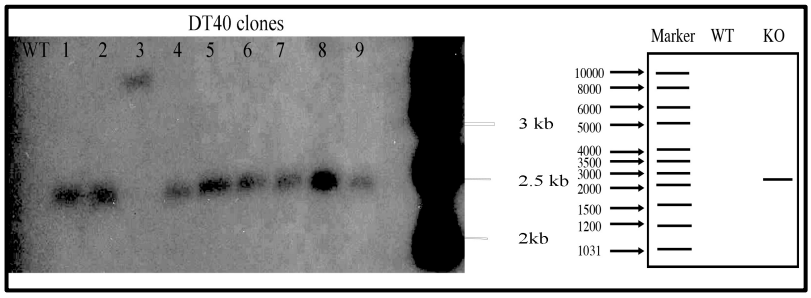


Figure 3.1.5: Analysis of DT40 clones for H3.3b KO using radiolabelled probe 3.
26 clones (only partial data shown) were analysed with probe 3 for Neomycin to check for the vector integration in the DT40 genome. The lane WT with no band represents DT40 WT cell line with endogenous H3.3b and is a negative control. The remaining lanes represent DT40 clonal cells with a band of approximately 2.38 kb suggesting the vector integration. The panel on the right shows the schematic representation of the expected results in WT and KO cells.

In all the 26 clones and the WT control, probe 1 hybridized unexpectedly to a fragment of around 4.2 kb which did not correlate to the expected size of 3.7 kb after the KO or 7.1 kb for the WT (Figure 3.1.6). Because all cell lines tested resulted in same fragment sizes, I analysed the targeting arm sequence for any point mutations, addition or deletion of restriction sites. It was found that a sequence variation in 5'TA of the vector gave rise to an additional XbaI site thus changing the size of the fragment and making it impossible to

distinguish with this strategy between the KO clones and WT. The size difference between expected and resulted fragments might be due to an un-sequenced short region within HindIII/HindIII fragment in the chicken genome.

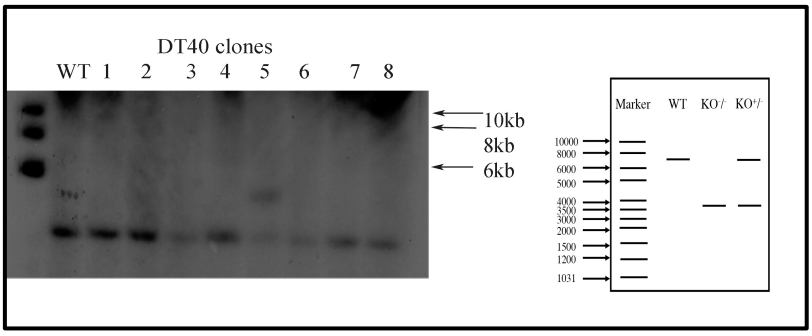


Figure 3.1.6: Analysis of DT40 clones for H3.3b KO using probe 1.
26 clones (only partial data shown) were analysed with probe 1 to check for the targeted vector integration in the DT40 genome. The lane WT represents DT40 WT cell line with endogenous H3.3b. The remaining lanes represent DT40 clonal cells. The panel on the right is the schematic representation of the expected size of 3.7 kb in WT and 7.1 kb in KO cells.

Probe 2 hybridization resulted in a fragment size of more than 10 kb and did not correlate to the expected size of 6.3 kb after the KO or 7.5 kb for the WT (Figure 3.1.7). This might be due to the presence of point mutations in the genome resulting in loss or gain of restriction sites. Also, the H3.3 locus has un-sequenced regions in the first intron that might contribute to size differences.

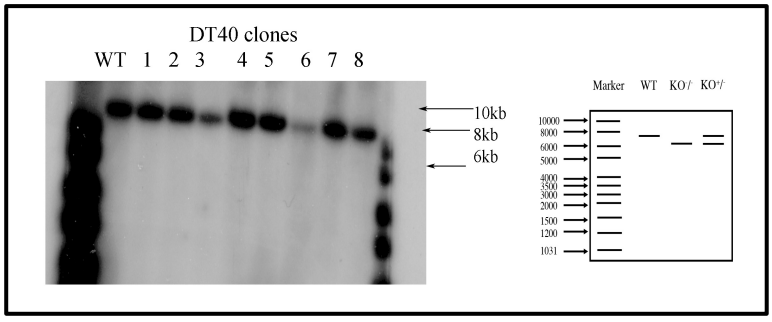


Figure 3.1.7: Analysis of DT40 clones for H3.3b KO using probe 2.
26 clones (only partial data shown) were analysed with probe 2 to check for the targeted vector integration in the DT40 genome. The lane WT represents DT40 WT cell line with endogenous H3.3b. The remaining lanes represent DT40 clonal cells. The panel on the right is the schematic representation of the expected results in WT and KO cells showing 6.3 kb after the KO and 7.5 kb for the WT.

However while trying to establish different KO analysis strategies, the final Gallus gallus Genebuild (Gallus_Gallus-4.0, April/December 2013) was released and there were also several updates on the number of H3 genes. At the same time, Julien Sale group from Cambridge[138] generated total H3.3 KO DT40 cells that we were able to obtain (Figure 3.1.8). Based on this, I revisited the strategies for establishing the mutant cell lines: (a) To KI H3 WT and mutants – with a 2xHA tag in H3.3 deficient DT40 cells obtained from the Sale lab (with an endogenous H3.2 background) (b) To perform a heterozygous KO of all H3.2 genes on chromosome 1 and KO H3.2 on the other allele to minimum number of copies with which the cells can survive in an H3.3 deleted background again in the Sale lab DT40 line and next to perform a KI of H3.2 WT and mutants with a FLAG tag.

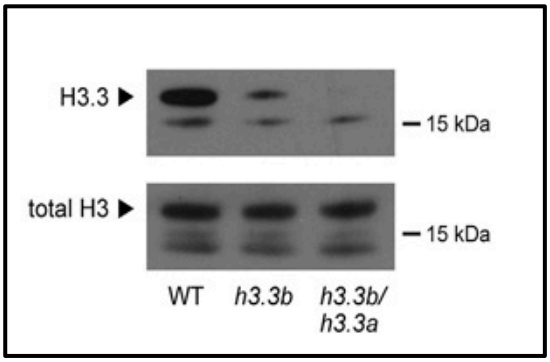


Figure 3.1.8: Confirmation of loss of H3.3 expression.
Western blot of acid-extracted histones for H3.3 and total H3 from WT, cells lacking H3.3b and cells lacking both H3.3a and H3.3b [138].

For this, I transfected the cells obtained from the Sale lab with a two times HA tagged H3.3 construct (H3.3-2xHA) with Puromycin as selection marker for homologous recombination. However, we encountered additional problems due to unexpected sequence variations between the published chicken genome and the actual DT40 genome. By analyzing the recently published DT40 genomic DNA sequence[139], I optimized my constructs and performed the transfection for the KI of H3.3-2xHA. The resultant clones were analysed by PCR using primers DT125 and DT126 across the H3.3 coding sequence for KIs (Figure 3.1.9). The WT lane represents the PCR product of endogenous H3.3 coding region of approximately 1.6 kb in the DT40 wild-type cell line and the remaining lanes represent the 40 clonal cell lines that were derived after transfecting H3.3 KO cells from Cambridge with the H3.3-2xHA construct and Puromycin selected. Absence of the 1.6 kb product confirmed that the clonal cells did not have an H3.3 KI.

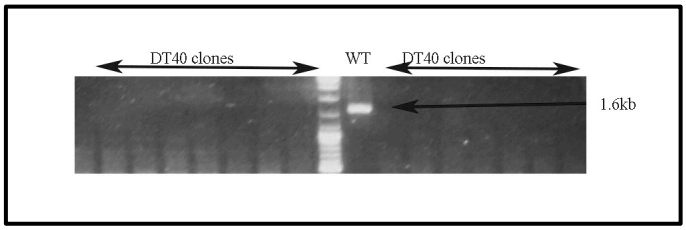


Figure 3.1.9: PCR analysis of H3.3-2xHA KI in DT40 cells.
PCR using primers DT125 and DT126: The WT lane represents DT40 wild-type cell line with endogenous H3.3B of product size 1.6Kb. The remaining lanes represent DT40 clonal cells where the absence of band indicates the absence of KI.

Together this suggests that we might have had problems with efficient recombination in DT40 cells and where therefore aiming for an alternative approach. The use of sequence-specific nucleases such as TALE nucleases and the CRISPR/Cas9 system has recently presented as an alternative method of gene manipulation in cell lines[140]. Hence, I used wild type Cas9 fused with GFP/RFP via the self cleaving P2A sequence (px330) obtained from the Zhang lab[140] (Figure 3.1.10) to try to generate H3.3 KI DT40 cell lines.

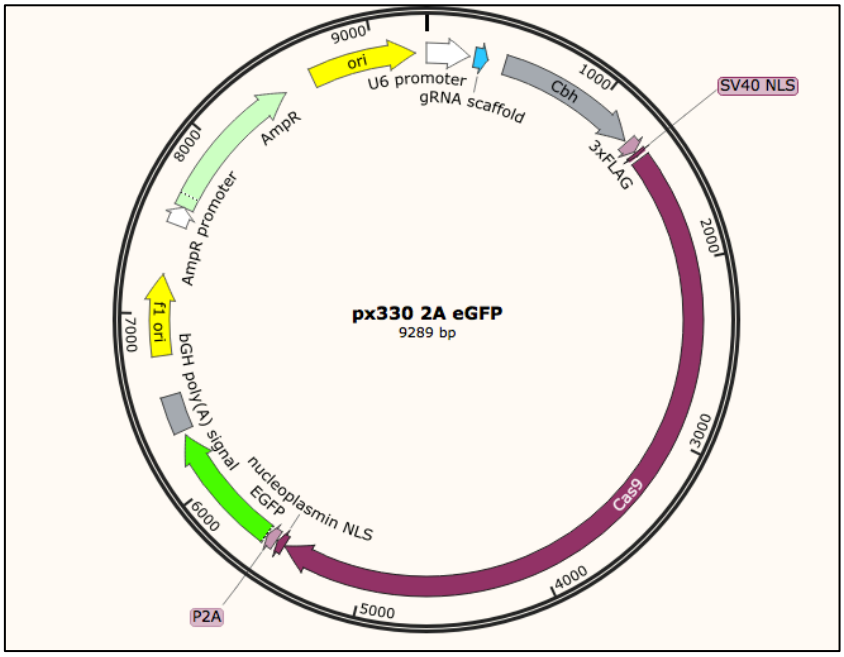


Figure 3.1.10: Illustration of px330 cas9 vector.
The cloned gRNA is driven by U6 promoter and cas9 fused to 3xFLAG by Cbh promoter. Cas9 is also fused to EGFP at the mRNA level with P2A sequence.

After 24, 48 or 72 hours of transfection, there was no expression for GFP or RFP whereas the control cells (HeLa) showed GFP or RFP expression with 60% transfection efficiency (data not shown). This might be due to improper P2A cleavage in DT40 cells, thus resulting in the lack of GFP/RFP expression.

3.2 Mouse embryonic stem (ES) cells

To generate mouse ES cell lines expressing mutated exclusively H3, I devised the following two strategies for histones H3.3 and H3.1/H3.2 respectively (Figure 3.2.1): (a) To delete (knock out) KO all H3.3 alleles up to one heterozygous allele and in a second step to knock in (KI) H3.3 mutants in the remaining copy to study the effects of H3.3 mutations (with an endogenous H3.1/H3.2 background).

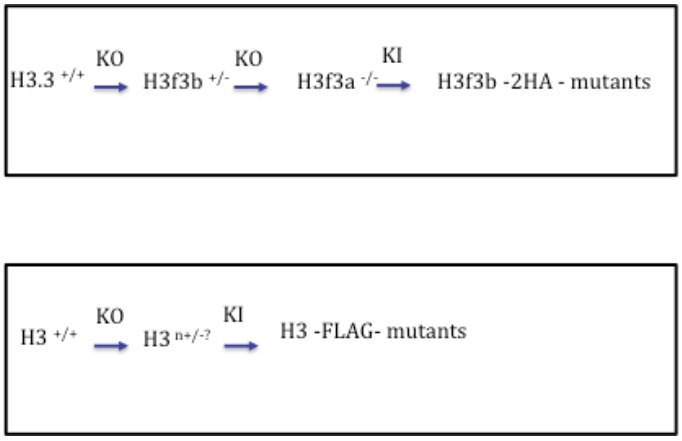


Figure 3.2.1. Strategy for the establishment of H3 mutant system.
Upper panel shows a scheme for establishing H3.3 mutant system and lower panel represents H3.2 mutant system using mouse ES cells. $H3^{n+/-}$ is the H3.1/H3.2 KO cell line where 'n' represents the minimum number of H3 copies with which the cells will survive.

(b) To KO all H3 alleles on the hist1 and hist2 clusters to the minimum number of copies with which the cells can survive and then KI H3 mutants in the remaining copies (with an endogenous H3.3 background) to study the effects of H3.1/H3.2 mutations. Since H3.1 and H3.2 genes are spread over greater than 2 Mb in hist1 cluster and over 40 kb in the hist2 cluster, I grouped ten of the twelve H3 genes according to their proximities into (i) Group I containing four H3 genes spanning over 40 kb in hist1 (ii) Group II containing three H3

genes spanning over 18 kb in hist1 and (iii) Group III with three H3 genes over 30 kb in hist2 (Figure 3.2.2). These grouping will allow me to perform stepwise deletions.

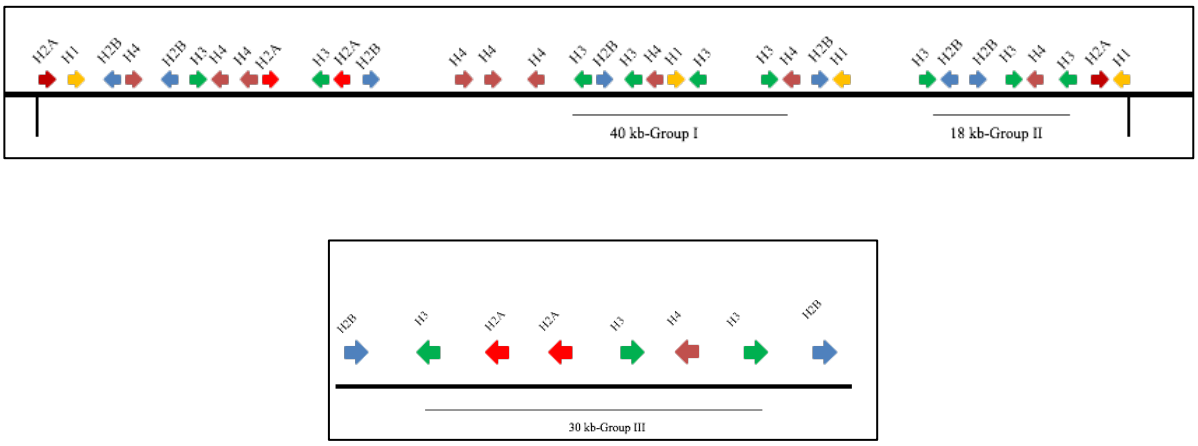


Figure 3.2.2: Schematic representation of organization of the histone gene clusters.
The histone H3 genes are grouped according to their proximities. Group I contains 4 H3 genes over 40 kb, Group II contains 3 H3 genes spread across 18 kb on hist1 cluster on chromosome 13 (upper panel) and Group III of 3 H3 genes spans over 30 kb on hist2 cluster on chromosome 3 (lower panel).

3.2.1 H3.3 mutant system

3.2.1.1 Heterozygous deletion of H3f3b

In mouse, two the genes H3f3a and H3f3b encoding H3.3 are of 3.5 kb and 11 kb size including introns, respectively. First, I targeted H3f3b and disrupted one of the alleles by CRISPR using a targeting strategy that was designed to remove the H3f3b gene totally. In this targeting vector px330 Cas9 from Zhang lab, (MIT) (Figure 3.1.10), Cas9 driven by Cbh promoter is fused to 3xFLAG and to either E-GFP or RFP. The 20-nt gRNA is cloned in the gRNA scaffold and is driven by U6 promoter.

The targeting strategy involved two gRNAs, directed to regions upstream and downstream of H3f3b (Figure 3.2.3 a). The two gRNAs were designed using the Zhang MIT CRISPR tool (http://crispr.mit.edu) and were cloned into the px330-GFP and px330-RFP vectors, respectively. The cells were sorted for both GFP and RFP signals using FACS after 48 hours of co-transfection of two-cas9/gRNA vectors. All the clones were analysed using PCR with primers MS27 and MS28 (expected product size of 1.2 kb) for the presence of wild type H3f3b and primers MS27 and MS29 for 3.6 kb H3f3b KO (expected product size of 650 bp, Figure 3.2.3 b). This PCR analysis identified 2 out of 12 clones (clones 7 and 9) to be heterozygous and 2 out of 12 to be homozygous (clones 4 and 5) H3f3b knock outs. The

absence of H3f3b was confirmed by sequencing (Figure 3.2.1.1.1 c) and heterozygote clone 7 was picked for targeting H3f3a.

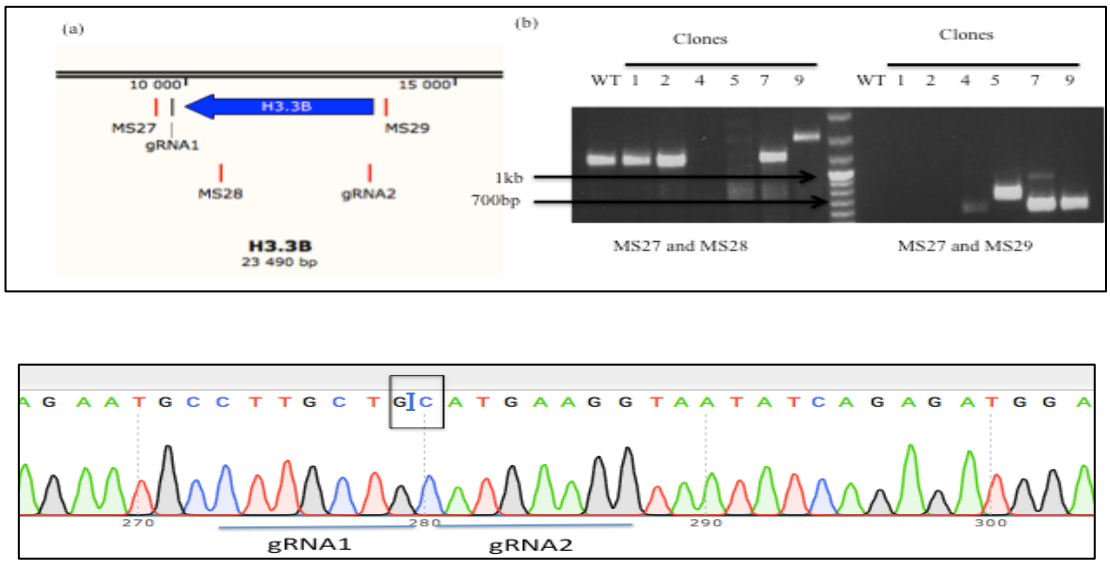


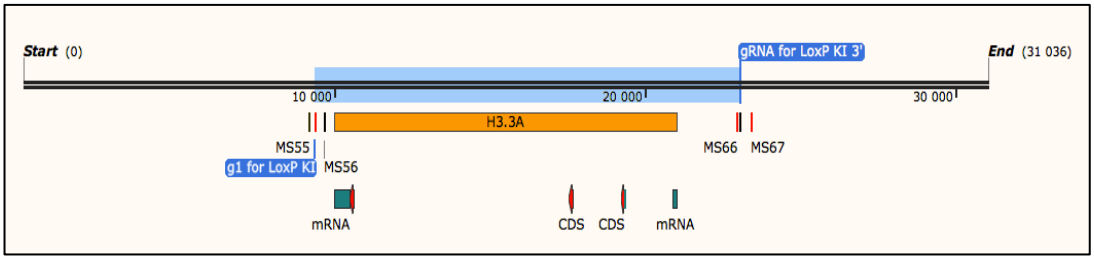
Figure 3.2.3: H3f3b KO strategy in mouse ESC.
(a) Strategy for H3f3b KO with the blue arrow representing the direction of transcription. gRNA1 and gRNA2 are the cas9 target sites. MS27 and MS28 are PCR primers for screening for the presence of H3f3b of size 1.2 kb, MS27 and MS29 are for KO of size 650 bp. (b) PCR screening of clones for the endogenous H3f3b using MS27/MS28 (left panel) and MS27/MS29 for KO (right panel). Clones 1 and 2 have 1.2 kb products only using primers MS27 and MS28 indicating there are no knock outs. Clones 4 and 5 have 650 bp products only using primers MS27 and MS29 indicating they have homozygous KO of H3f3b. Clones 7 and 9 have 1.2 kb and 650 bp products using both MS27 and MS28, and MS27 and MS29 respectively, indicating that they have heterozygote KO of H3f3b. (c) Sequencing result of clone 7 with heterozygous H3f3b KO. The box around GC region highlights the recombined region after deletion.

3.2.1.2 Homozygous deletion of H3f3a

Having successfully deleted H3f3b, I used next a similar strategy to target H3f3a but the resultant clones did not have any H3f3a deletions. This could be because of its 11 kb size and larger deletions using CRISPR could be challenging. So, I decided to apply the Cre/LoxP system where I could integrate LoxP sites 5' upstream and 3' downstream of H3f3a in the same direction and then express CRE for deletion of the sequence between these two sites (Figure 3.2.4 a). I designed DNA oligos containing the 34 bp loxP site and a 6 bp EcoRI site[141] flanked by 60 bps sequences on each side adjoining the DSBs. The two gRNAs targeting 5' upstream and 3' downstream of H3f3a were designed using the Zhang MIT CRISPR tool and were cloned into the px330-GFP and px330-RFP vectors, respectively. The cells were sorted for both GFP and RFP signals using FACS after 48 hours of co-transfection two-cas9/gRNA vectors and the single stranded DNA oligos with the LoxP sites. The resultant clones were analysed using PCR primers MS55 and MS56 for detecting LoxP

insertion on the 5' end and primers MS66 and MS67 for LoxP insertion on 3' end of H3f3a. PCR analysis revealed that 3 out of 15 clones carried LoxP site integrations in the 5' upstream region. All the clones containing LoxP site integration were confirmed using sequencing (Figure 3.2.4 b). This demonstrates that HDR-mediated repair can introduce targeted integration of 40 bp DNA elements efficiently through CRISPR/Cas9-mediated genome editing.

(a)



(b)

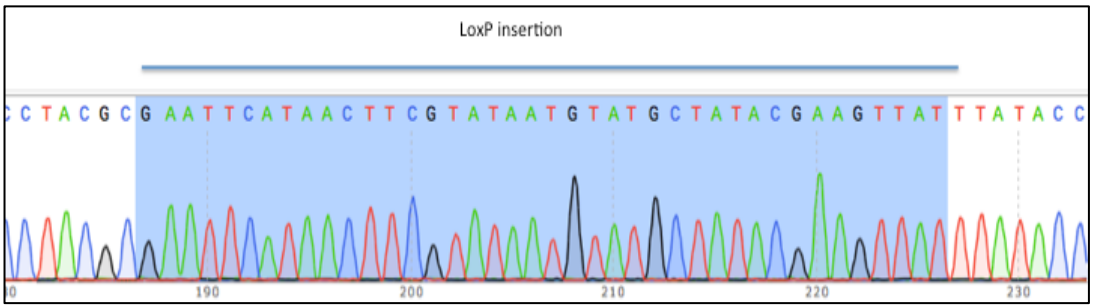


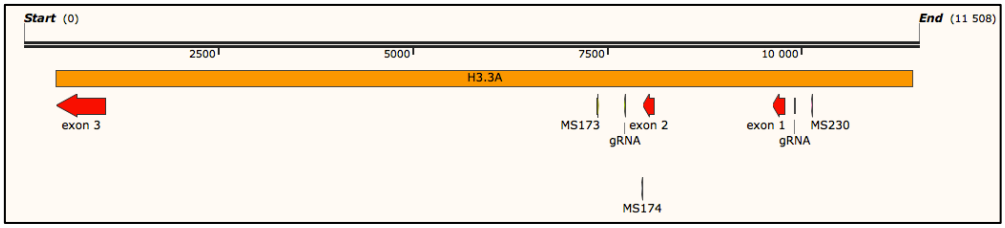
Figure 3.2.4: Targeted integration of LoxP sites for H3f3a KO.
(a) Strategy for H3f3a LoxP KI with gRNAs for cas9 target on 5' and 3' side of H3f3a. MS55 and MS56 are PCR primers for screening the presence of LoxP insertion on 5' end and MS66 and MS67 for 3' end. (b) Sequencing result of clone 15 with LoxP insertion on 5' upstream of H3f3a.

Several attempts with different gRNAs that were tested (data not shown) for introducing a LoxP site 3' downstream of H3f3a were unsuccessful. This could be due to factors such as chromatin accessibility, CpG methylation or DNA flexibility interfering with the Cas9 binding[142].

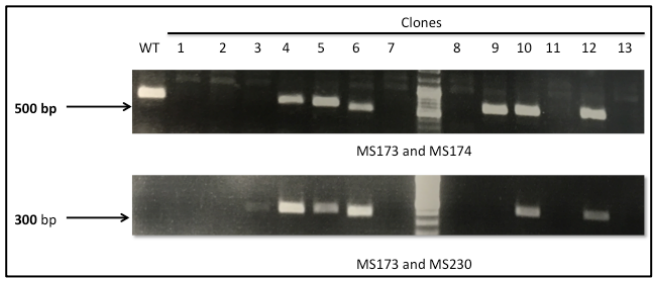
Hence, I devised an alternative targeting strategy where I designed to remove exons 1 and 2 involving the majority of the coding sequence (amino acid residues 1-94) of H3f3a. It involved two gRNAs, directed to regions downstream of exon 2 and upstream of exon1 of H3f3a (Figure 3.2.5 a). The gRNAs were designed using the Zhang MIT CRISPR tool and were cloned into px330-GFP and px330-RFP. The cells were sorted for GFP and RFP

positive cells using FACS after 48 hours of co-transfection of the two-cas9/gRNA vectors. All the clones were analysed using PCR with primers MS173 and MS174 (expected product size of 330 bp) for the presence of H3f3a and primers MS173 and MS230 for 2 kb H3f3a KO (expected product size of 580 bp, Figure 3.2.5 b). PCR identified 9 out of 20 clones to be heterozygous and 6 out of 20 clones to be homozygous. The absence of full length H3f3a mRNA was confirmed by a semi quantitative RT-PCR (Figure 3.2.5 c).

(a)



(b)



(c)

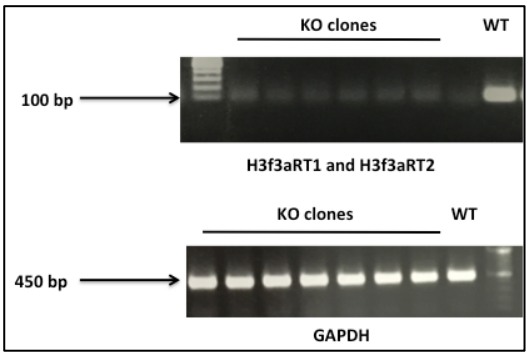


Figure 3.2.5: H3f3a KO strategy in mouse ESC.
(a) Strategy for H3f3a KO. Exons 1,2 and 3 are shown in red. gRNAs targeting downstream of exon 2 and upstream of exon 1 marked along with primers used for screening for presence or absence of H3f3a. (b) PCR screening of clones for the endogenous H3f3a using MS173/MS174 (upper panel) and MS173/MS230 for KO (lower panel). Clones 3,4,5,6,9,10 and 12 have 580 bp product and 330 bp product using both primers MS173/MS174, and MS173/MS230 respectively, indicating they have a heterozygous KO of H3f3a. Clones 1,2,7,8,11 and 13 have 330 bp product only using primers MS173/MS230 indicating there have homozygous KO of H3f3a. (c) RT-PCR screening of clones for the detection of RNA of endogenous H3f3a using primers H3f3aRT1 and H3f3aRT2 (100 bp product size) spanning exons 2 and 3 (upper panel) and GAPDH expression levels (450 bp product size) as control (right panel).

3.2.1.3 KI of mutants on H3f3b

To study the effects of histone core modifications H3K56, H3K64, H3K122, and H3K115 acetylation and compare them with tail modifications I used the ES cell line with homozygous H3f3a deletion and H3f3b heterozygous deletion for targeted mutation of the remaining copy of H3f3b gene. For this ES cells were co-transfected with the two gRNAs vectors (px330-GFP and px330-RFP), with gRNAs directed to regions upstream and downstream of H3f3b (Figure 3.2.6) and a homology directed repair (HDR) donor vector encoding different mutations (Table 3.2.1).

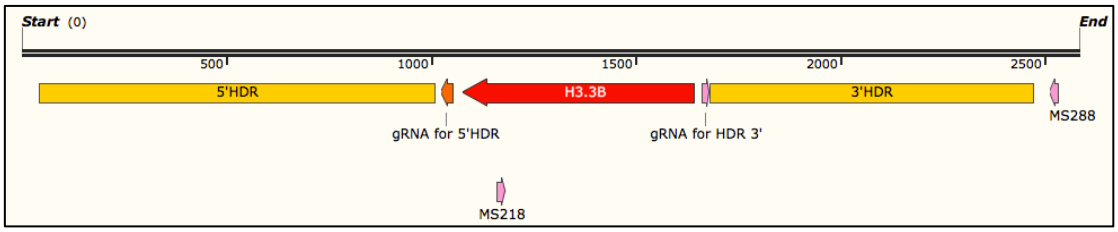


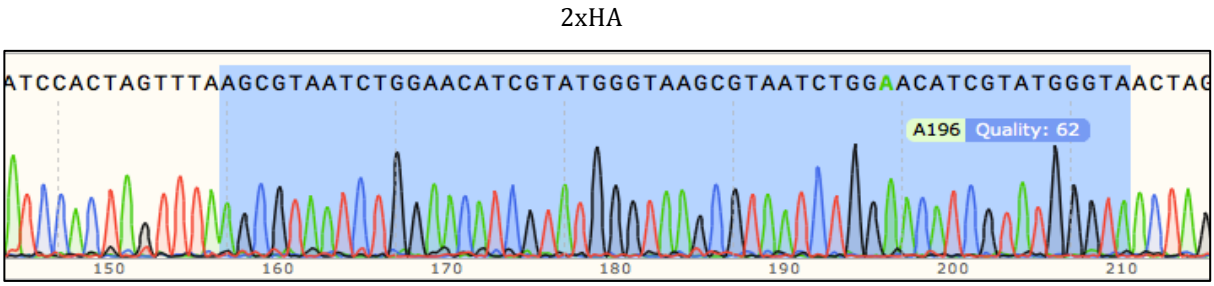
Figure 3.2.6: HDR mediated targeted integration of mutants on H3f3b.
Strategy for H3f3b mutants KI with gRNAs for cas9 target on 5' upstream and 3' downstream of H3f3b. MS218 and MS288 are PCR primers for screening the presence of KI mutants. Homology arms used are indicated in yellow

No.	Vectors encoding mutations
1	H3.3WT-2xHA
2	H3.3K56/64/115122R-2xHA
3	H3.3K56/64/115/122Q-2xHA
4	H3.3K9/14/18R-2xHA
5	H3.3K9/14/18Q -2xHA

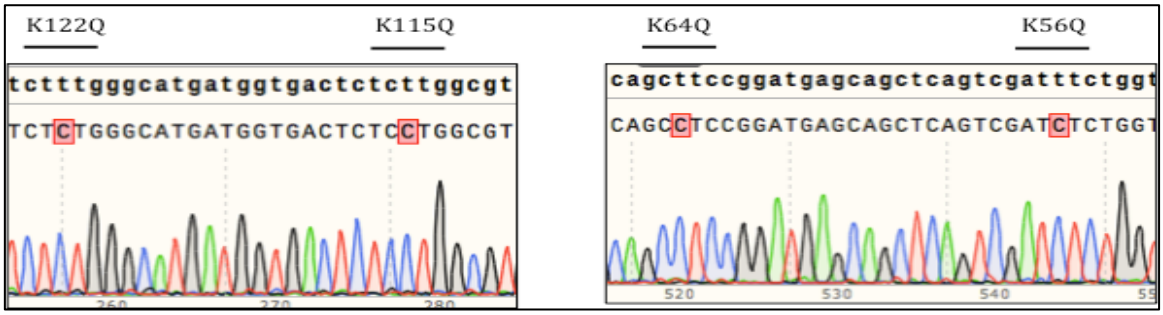
Table 3.2.1: List of HDR vectors encoding different mutations.

The HDR vector was cloned with 1 kb homologous sequence immediately upstream and 1 kb homologous sequence immediately downstream of H3f3b gene (homology arms) flanking the mutated H3f3b sequence and a 2xHA tag into the pLox Neo vector carrying Neomycin resistance. Stable mutant cell lines expressing 2xHA -H3.3WT, -H3.3K56/64/115/122Q, and -H3.3K9/14/18R were established and confirmed using sequencing (Figure 3.2.7).

(a)



(b)



(c)

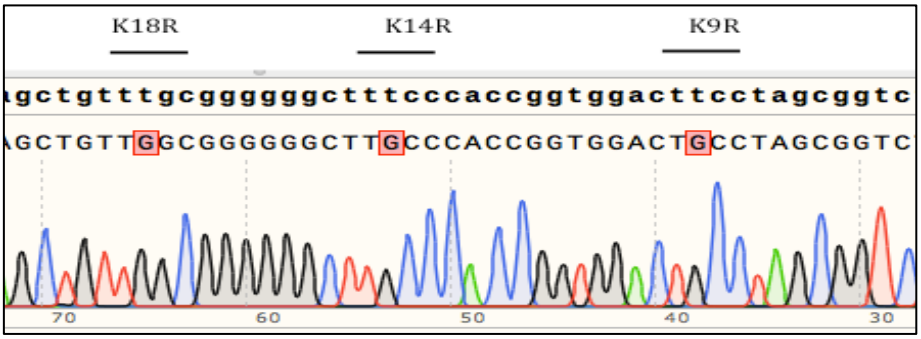


Figure 3.2.7: Sequencing result of H3K56/64/115/122R and H3K9/14/18Q mutant clones.
(a) Sequencing of PCR product of H3.3WT-2xHA showing the 2xHA tag in blue (b) Sequencing of PCR product of H3.3K56/64/115/122R clone shows the presence of arginine residues on K122, K115, K64 and K56 marked in that order from left to right. Mutated codons are highlighted and compared to the WT sequence on the top. (c) Sequencing of PCR product of H3.3K9/14/18Q clone shows the presence of glutamine residues on K18, K14 and K9 marked in that order from left to right. Mutated codons are highlighted and compared to the WT sequence on the top.

To check the expression level of tagged H3 mutants in all the three cell lines I performed Immunoblots on acid extracted histones from these cell lines using an anti HA antibody. Immunoblotting using H3.3 antibody was performed to check the presence of endogenous H3.3WT (Figure 3.2.8).

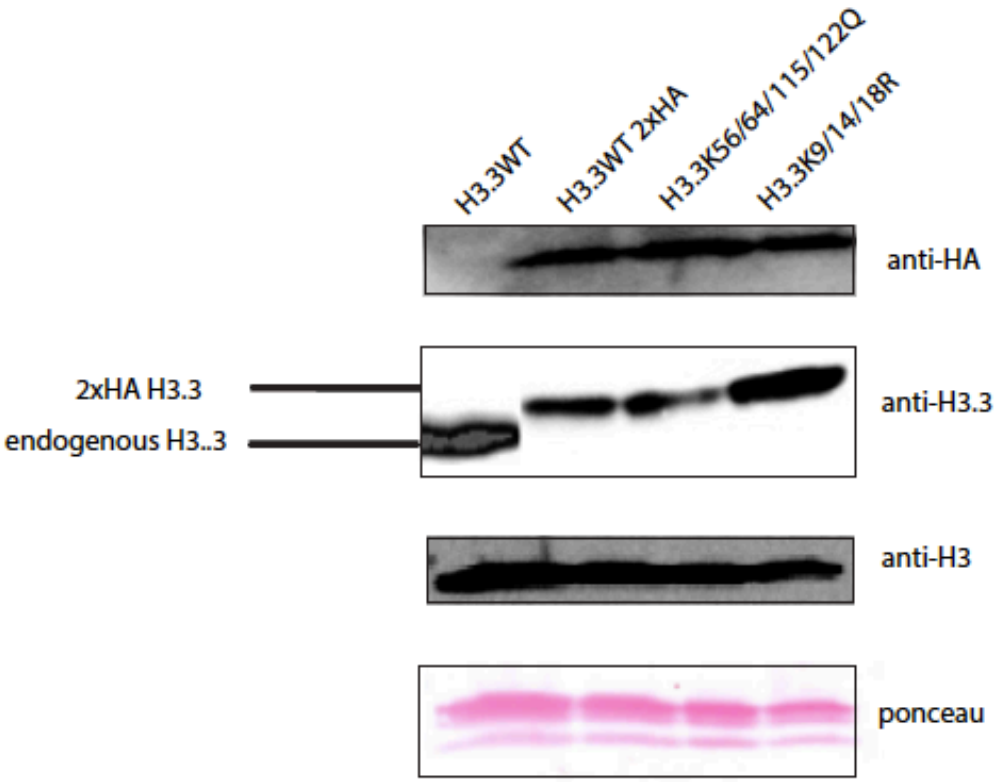


Figure 3.2.8: Immunoblotting on acid extracted histones of 2xHA -H3 mutants.
Histones from ES WT and mutant cells expressing HA -H3.3WT, -H3.3K56/64/115/122Q, -H3.3K9/14/18R were separated using acrylamide gel, transferred to nitrocellulose membrane, stained with ponceau and immunoblotted with anti HA, H3.3 and a general H3 antibody. The upper band on aH3.3 immunoblot represents HA tagged histones and the lower band indicates endogenous H3 histones.

Interestingly the ES cell lines expressing H3.3K56/64/115/122R and H3.3K9/14/18Q were viable but grew extremely slow when compared to ES WT and the other mutants. Both of these knock in cells were forming aggregates like pluripotent cells, appeared smaller and dissociated. They died within several days likely due to slow proliferation and differentiation (Figure 3.2.9)

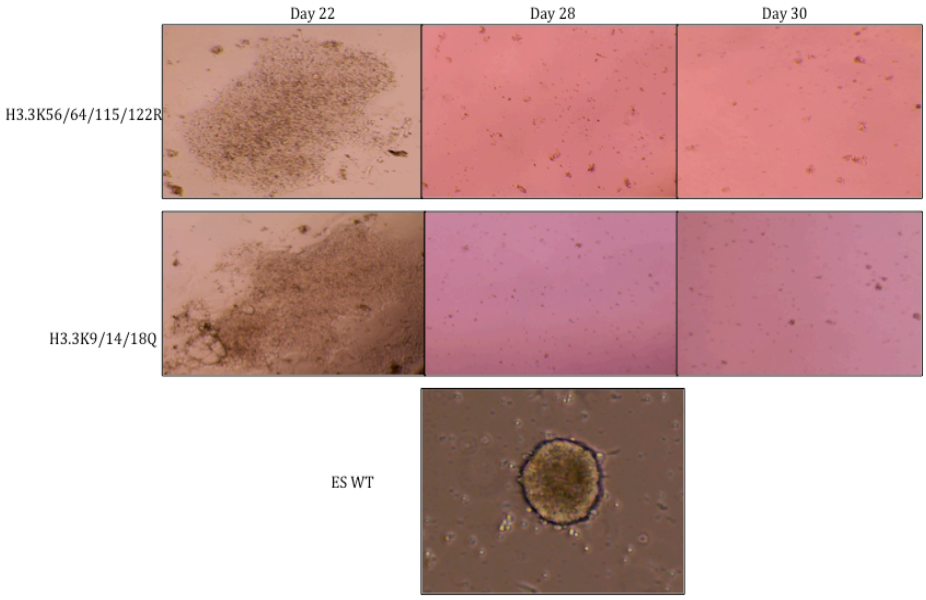
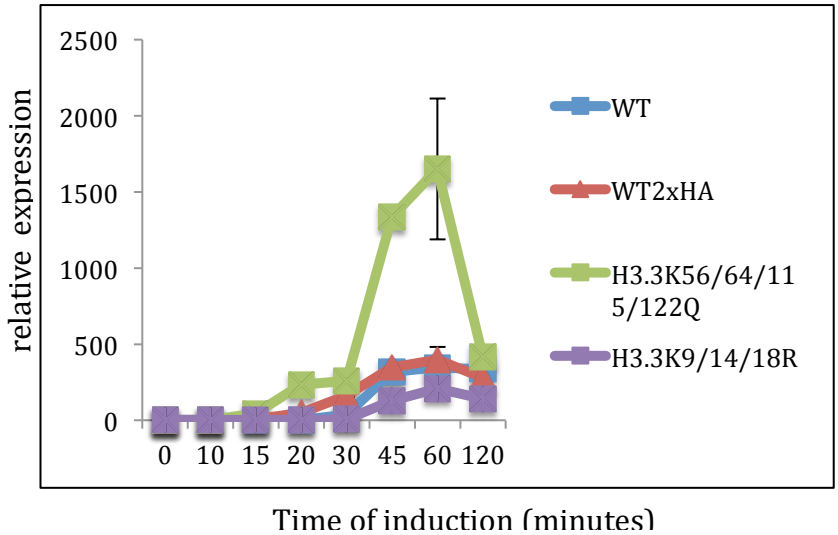


Figure 3.2.9: Images showing phenotypes of H3.3K56/64/115/122R and H3.3K9/14/18Q mutant clones. Upper panel shows the images of H3.3K56/64/115/122R on day 22, 28 and 30 after selection. Lower panel shows the images of H3.3K9/14/18Q on day 22, 28 and 30 after selection. The cells fail to form aggregates that define pluripotent cells and look dissociated. The mutants die within several days later.

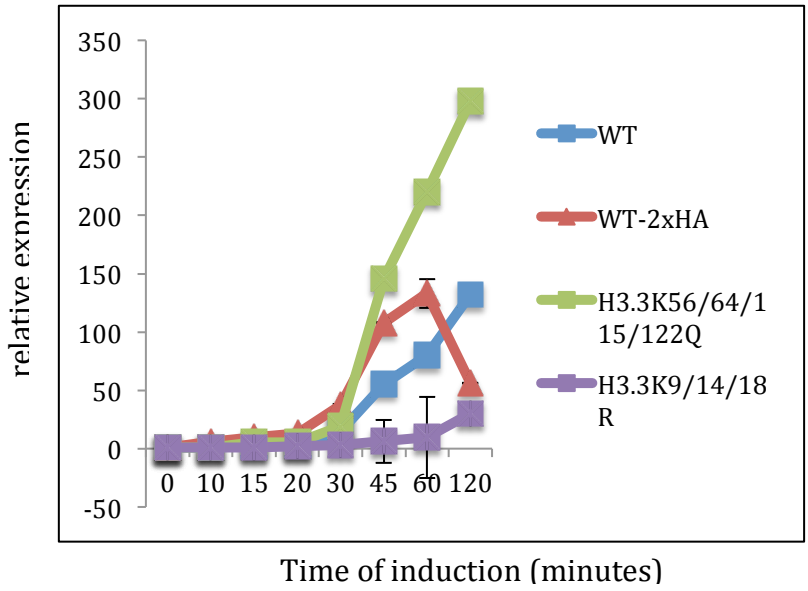
3.2.1.4 H3K56/64/115/122ac affects induction of gene expression

Studies from our lab on H3K64 and H3K122 acetylation indicated a strong link between these modifications and active transcription. To study the effects of all four globular domain acetylation on gene transcription *in vivo*, I subjected all the three mutant ES cell lines to gene induction experiments. I followed the kinetics of the activation of immediate-early (IE) response gens such as c-fos, egr1 and c-myc, which are down regulated by serum starvation and re-induced by 15% FCS and 100 nM TPA (12-O-tetradecanoylphorbol-13-acetate) stimulation[143], a small analogue of diacylglycerol that is able to activate the protein kinase C (PKC) signal transduction cascade (Figure 3.2.10). It is to be noted that c-fos and egr1 genes are not expressed in WT mouse ES cells.

(a)


relative expression
2500
2000
1500
1000
500
0
0 10 15 20 30 45 60 120
Time of induction (minutes)
WT
WT2xHA
H3.3K56/64/115/122Q
H3.3K9/14/18R

(b)


relative expression
350
300
250
200
150
100
50
0
-50
0 10 15 20 30 45 60 120
Time of induction (minutes)
WT
WT-2xHA
H3.3K56/64/115/122Q
H3.3K9/14/18R

(c)

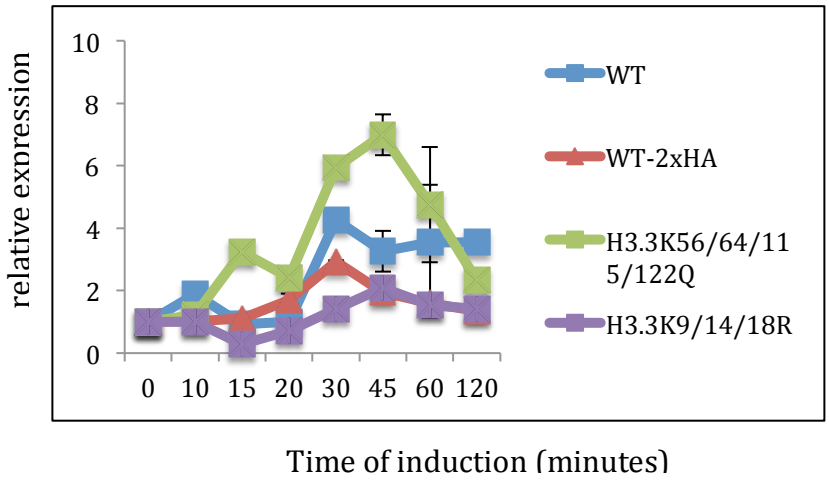

relative expression
10
8
6
4
2
0
0 10 15 20 30 45 60 120
Time of induction (minutes)
WT
WT-2xHA
H3.3K56/64/115/122Q
H3.3K9/14/18R

Figure 3.2.10: Acetylation mimic H3K56/64/115/122Q directly affects gene expression in ES cells.
The graphs represent the mRNA expression kinetics of early response genes upon serum and TPA inducstion (a) c-fos, (b) egr1 and (c) c-myc mRNA levels relative to the housekeeping gene HPRT1 in all four ES cell lines. ES cells were grown normally for 24 hours, starved for 24 hours and re-induced with medium containing 15% FCS + 100 nM TPA and was followed over a time-course. The curves are average of two biological experiments. The value of timepoint 0 was set as 1 and all the values are normalized to HPRT1 gene.

Treatment of ES cells with 15% FCS and TPA induced a maximal expression of IE genes within 30 to 60 minutes (Figure 3.2.10). ES mutant cell line expressing the acetylation mimic H3.3K56/64/115/122Q had an eight-fold increase in c-fos mRNA levels compared to the ES WT cells at 60 minutes after induction (Figure 3.2.10 a). Also, this cell line had a three-fold increase in egr1 mRNA levels at 120 minutes after induction and a two-fold increase in c-myc at mRNA levels at 45 minutes after induction when compared to the ES WT (Figure 3.2.10 b, c). A repressive effect was found in ES cell line expressing acetylation-deficient mutant H3.3K9/14/18R, which follows the same kinetics of the ES WT cells in the initial stage of activation but then tends to decrease the activation at final stages. This suggests that the presence of acetylation on all four globular domain residues H3.3K56/64/115/122, in this case, in the form of the mimic H3.3K56/64/115/122Q enhances transcriptional activation *in vivo.*

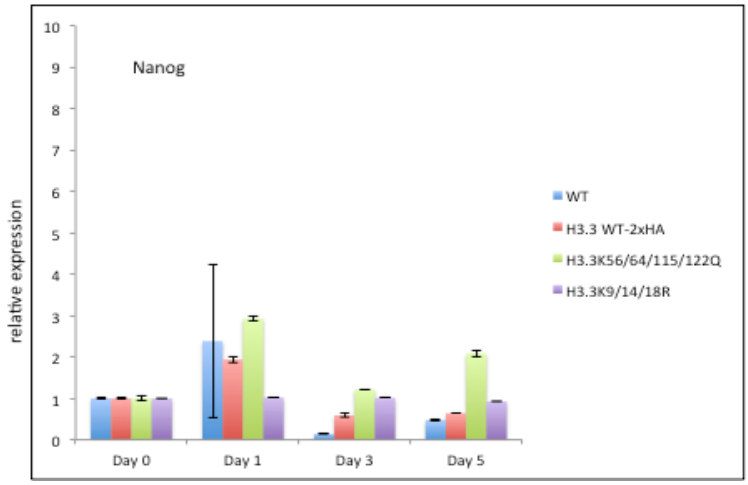
3.2.1.5 H3K56/64/115/122ac affects transcription during ES cell differentiation

In order to study the effects of all four globular domain acetylation on ES cell differentiation, all the three mutant ES cell lines were differentiated into embryoid bodies (EBs) in an experiment in collaboration with Dr. Moyra Lawrence and Julia Schluckebier. Differentiation of ES cells into EB was perfomed by growing them in hanging drop cultures. After 2 days in hanging drops, EBs were placed into suspension for 3 days when they undergo differentiation and cell specification along the three germ lineages – endoderm, ectoderm, and mesoderm. During EB differentiation of ESCs, they turn off pluripotency genes such as Nanog and start activating genes such as T-Brachyury (mesoderm), Fgf5 (ectoderm) and Gata4 (endoderm). RNA extraction and cDNA synthesis were done on day 0, day 1, day 3 and day 5.

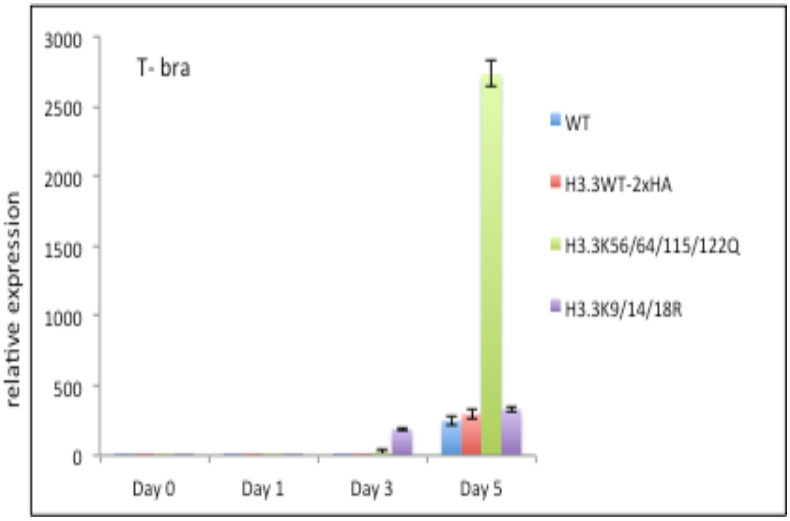
At day 3 during differentiation, qPCR analysis (Figure 3.2.11) in all cell lines showed the marked upregulation of the ectoderm gene Fgf5, endoderm Gata4 gene and mesoderm gene

Brachyury T at day 5. Consistent with the process of ES cell differentiation, the pluripotency marker Nanog is repressed in EBs of all the mutant cell lines and ES WT.

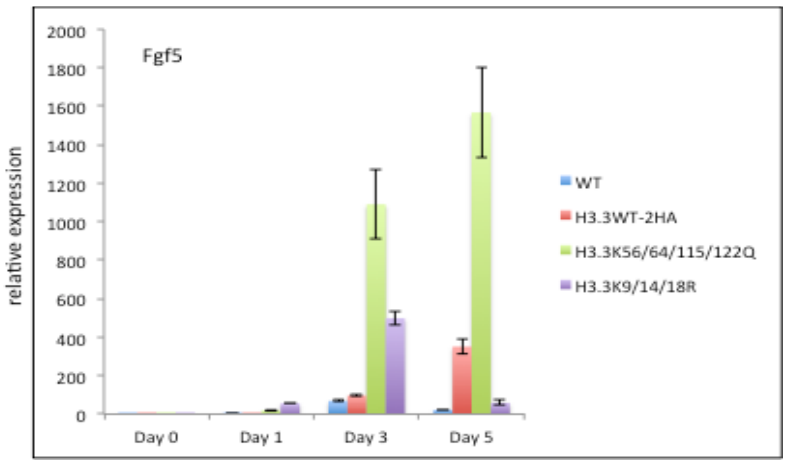
(a)



(b)



(c)



(d)

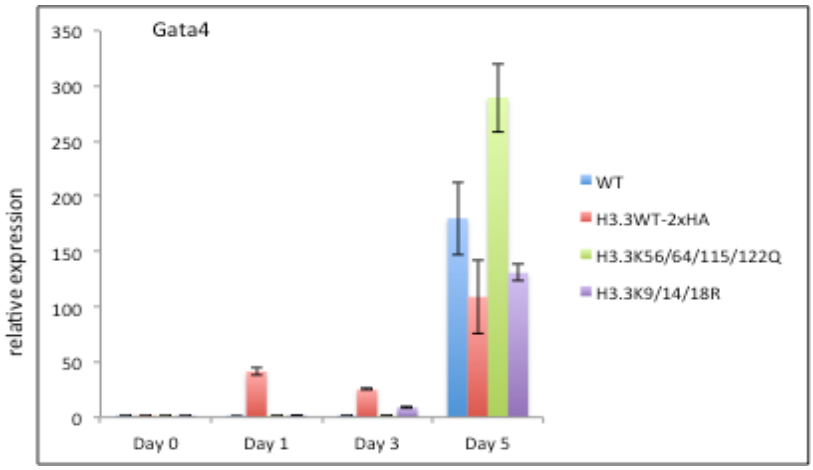


Figure 3.2.11. Acetylation mimic H3K56/64/115/122Q directly affects gene expression in ES cells. Quantitative PCR for expression of (a) Nanog (pluripotent), (b) Brachyury T (mesoderm), (c) Fgf5 (ectoderm) and (d) Gata4 (endoderm) genes in all four ES cell lines from day 0 to 5 of differentiation into cardiomyocytes. Shown are the averages and standard deviations of technical replicates. The value of day 0 was set as 1 and all the values are normalized to HPRT1 gene.

A key regulator of mesodermal specification is the T-box transcription factor, brachyury. Induction of brachyury on day 5 during EB differentiation is enhanced a thousand-fold in EBs acetylation mimic H3.3K56/64/115/122Q (Figure 3.2.11 b) when compared to EBs of ES WT, H3.3WT-2xHA and H3.3K9/14/18R. Similarly, Fgf5, a marker of ectoderm differentiation shows a thousand-fold increase in H3.3K56/64/115/122Q EBs induction on day 3 during differentiation when compared to ES WT (Figure 3.2.11 c). Also, H3.3K56/64/115/122Q EBs showed a two–fold increase in a marker of endodermal differentiation, Gata4 induction compared to ES WT (Figure 3.2.11 d). These results suggest that the presence of acetylation on all four globular domain residues H3.3K56/64/115/122 could affect the expression of ES cell differentiation specific genes.

3.2.2 H3.1/H3.2 mutant system

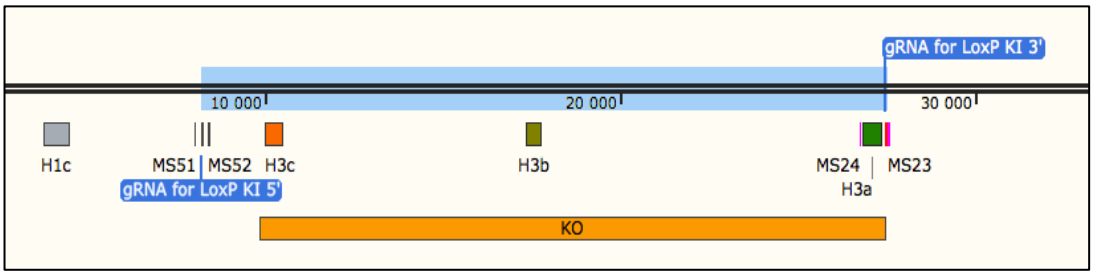
3.2.2.1 Deletion of Group II H3 genes

In mouse, there are twelve genes encoding H3.1 and H3.2 without introns, spread between two histone gene clusters. I first targeted Group II containing three H3 genes, H3a, H3b, and H3c that span across 18 kb on hist1 (Figure 3.2.2). First, I aimed for direct deletion using px330 cas9 vectors to target the 5’ upstream of H3c and 3’ downstream of H3a (Figure

3.2.12). But the resultant clones did not have any KO. This could be because of the 18 kb that should be deleted and such large deletions using CRISPR could be challenging.

Therefore, I decided to use again the Cre/LoxP system where I could integrate LoxP sites at 5' upstream of H3c and 3' downstream of H3a in the same direction and express CRE for deletion. I designed DNA oligos containing the 34 bp loxP site and a 6 bp EcoRI site[144] flanked by 60 bps sequences on each side adjoining the DSBs. Then, I co-transfected px330/gRNA vectors, and the single-stranded DNA oligos targeting both 5' downstream of H3c and 3' upstream of H3a (Figure 3.2.12 a). The resultant clones were analysed using PCR primers MS51 and MS52 for detecting LoxP insertion on the 5' end of H3c and primers MS23 and MS24 for LoxP insertion on 3' end of H3a. PCR identified 6 out of 15 clones carrying LoxP site on 5' end or 4 out of 15 clones having LoxP site integration on 3' end but no clones with both insertions. All the clones containing LoxP site integration were confirmed using sequencing (Figure 3.2.12 b).

(a)



(b)

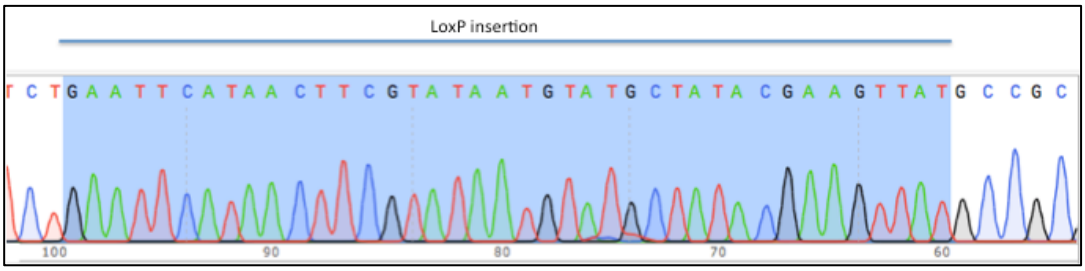
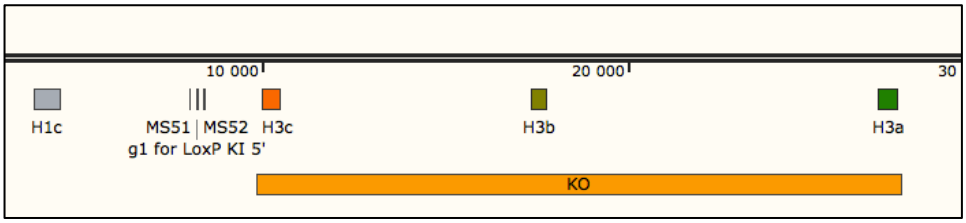


Figure 3.2.12: H3 LoxP KI strategy for H3c, H3b and H3a genes on hist1 cluster.
(a) Strategy for H3 LoxP KI with gRNAs for cas9 target on 5' and 3' side of 18kb KO region. MS51 and MS52 are PCR primers for screening the presence of LoxP insertion on 5' end and MS23 and MS24 for 3' end. (b) Sequencing result of clone 12 with LoxP insertion on 3' side of the 18 kb KO region.

Hence, a clone containing LoxP insertion on 3' end was picked and transfected with gRNA vector and single-stranded DNA oligo targeting the 5' end of H3c (Figure 3.2.13 a). The resultant clones were analysed using PCR primers MS51 and MS52 and indeed 4 out of 16

clones contained the loxP 5' integration. All the clones containing this LoxP site integration were confirmed using sequencing (Figure 3.2.13 b). Next, I used a selected clone for CRE expression and recombination.

(a)



(b)

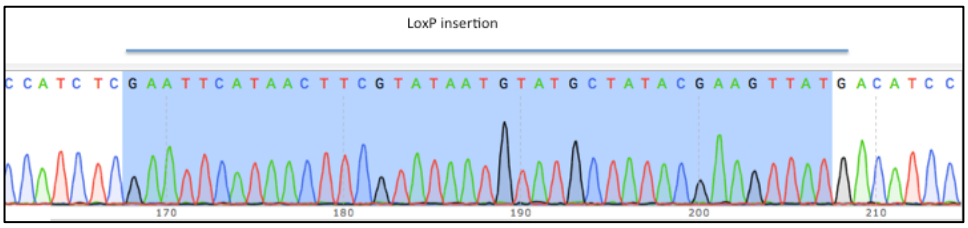
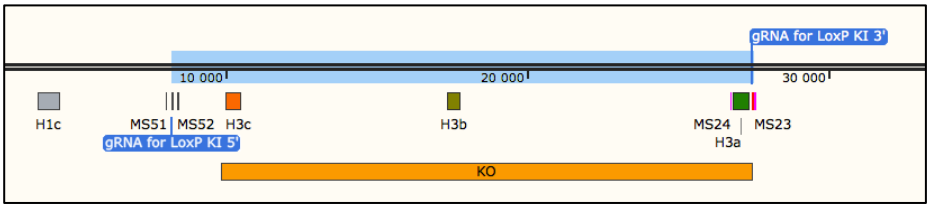


Figure 3.2.13 H3 LoxP KI on 5'end of KO region.
(a) Strategy for H3 LoxP KI with gRNA for cas9 target on 5' side of 18kb KO region. MS51 and MS52 are PCR primers for screening the presence of LoxP insertion on 5' end. (b) Sequencing result of clone 9 with LoxP insertion on 5' side of the KO region.

Clonal selection was made after the expression of CRE for 48 hours. 14 out 20 clones had the deletion of Group I genes, which were confirmed by PCR using primers MS51 and MS23 (Figure 3.2.14 a). All the clones with Group I KO were confirmed using sequencing (Figure 3.2.14 b).

(a)



(b)

Figure 3.2.14. Deletion of Group I containing three H3 genes.
A scheme showing 18kb KO region containing three H3 genes. (b) Sequencing result of a clone confirming the deletion of Group I genes.

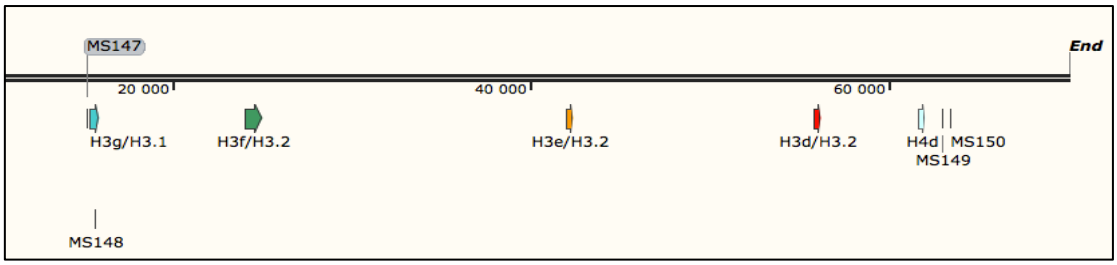
3.2.2.2 Deletion of Group I H3 genes

Having successfully deleted group II containing three H3 genes, I next wanted to delete Group I containing four H3 genesH3d, H3e, H3f and H3g that span across 40 kb on hist1 (Figure 3.2.2). First, I aimed for direct deletion using px330 cas9 vectors to target the 5' upstream of H3d and 3' downstream of H3g (Figure 3.2.15 a). But the resultant clones did not have any KO. This could be because of the 40 kb that should be deleted and such large deletions using CRISPR could be challenging.

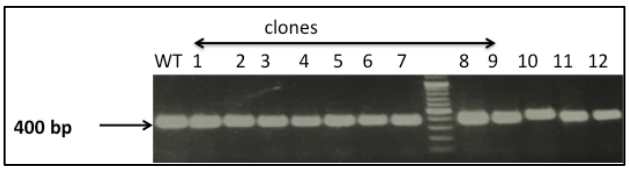
Therefore, I decided to use again the Cre/LoxP system where I could integrate LoxP (Lox2272), where there is a modification in the 8bp spacer region and studied to not recombine between wild type LoxP sites[145] at 5' upstream of H3d and 3' downstream of H3g in the same direction and express CRE for deletion. I used DNA oligos as recombination templates containing the 34 bp loxP site and a 6 bp EcoRI site flanked by 60 bps sequences on each side adjoining the DSBs. The resultant clones were analysed using PCR primers MS147 and MS148 with an expected product size of 440 bp for LoxP insertion and 400 bp on its absence on the 5' end of H3g and primers MS149 and MS150 with an expected product size of 540 bp for LoxP insertion and 500 bp on its absence on 3' end of H3d. PCR analysis using MS147 and MS148 identified 1 out of 12 clones (clone 10) carrying LoxP site on 5' end (Figure 3.2.15 b). Sequencing confirmed a heterozygous insertion of LoxP site on clone 10. Repeating the transfection using the same gRNA on clone 10 did not result in a

homozygous insertion. PCR analysis using MS149 and MS150 showed no clones with insertions on 3' end (Figure 3.2.15 c).

(a)



(b)



(c)

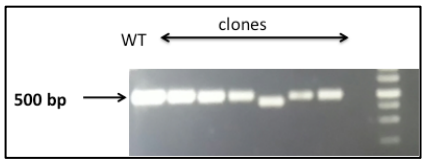


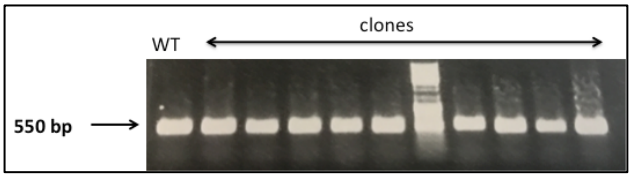
Figure 3.2.15: H3 LoxP KI strategy for H3g, H3f, H3e and H3d genes on hist1.
(a) Strategy for H3 LoxP KI with gRNAs for cas9 target on 5' and 3' side of 40 kb KO region. MS147 and MS148 are PCR primers for screening the presence of LoxP insertion on 5' end and MS149 and MS150 for 3' end. (b) PCR analysis of clones using MS147 and MS148 to screen for LoxP site insertion on 5' end result in a band of 400 bp similar to the wild type indicating the absence of targeted integration of LoxP site. PCR product size of 440 bp on clone 10 indicate targeted insertion. (c) PCR analysis of clones using MS149 and MS150 to screen for LoxP site insertion on 3' end result in a band of 500 bp similar to the wild type indicating the absence of targeted integration of LoxP site.

Several attempts with six different gRNAs that were tested (data not shown) for introducing a LoxP site homozygously on 5' upstream of H3g and 3' downstream of H3d were unsuccessful. This could be due to factors such as chromatin accessibility, CpG methylation or DNA flexibility interfering with the Cas9 binding[146].

Hence, we devised an alternative targeting strategy where we use a donor repair template to activate the HDR pathway to generate an 40 kb KO. For this, ES cells were co-transfected with a px330Cas9 vector with GFP containing two gRNAs directed to regions upstream and downstream of 40 kb (Figure 3.2.15 a) and a homology directed repair (HDR) vector px391.

The HDR vector contains 1 kb homologous sequence immediately upstream and 1 kb homologous sequence immediately downstream of the 40 kb KO region (homology arms) and puromycin resistance flanked by mutant FRT sites for marker recycling. Clones were selected for knock outs using both FACS for GFP and puromycin.

(a)



(b)

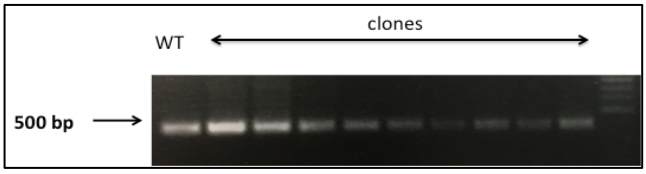


Figure 3.2.16: PCR analysis of H3g, H3f, H3e and H3d genes KO on hist1.
(a) PCR analysis of clones using MS250 and MS251 to screen for Group I genes 40 kb KO on the 5' upstream of H3g gene. All the clones contain the WT allele, indicated by a 550 bp PCR product. (b) PCR analysis of clones using MS252 and MS253 to screen for Group I genes 40 kb KO on the 3' downstream of H3d gene. All the clones contain the WT allele, indicated by a 400 bp PCR product.

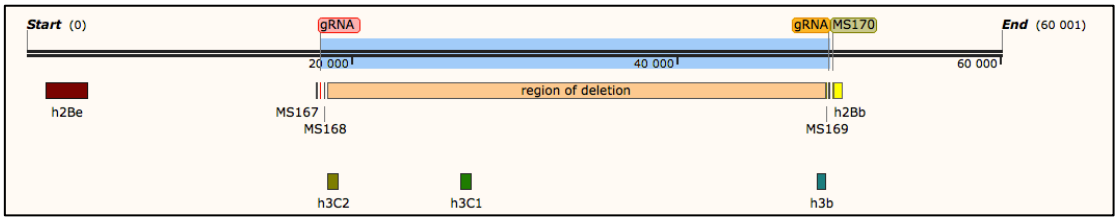
The resultant clones were analysed using PCR primers MS250 and MS251with an expected product size of 550 bp for the presence of WT allele and no product for KO on the 5' end of H3g gene and primers MS252 and MS253 with an expected product size of 400 bp for the presence of WT allele and no product for KO on the 3' end of H3d gene. PCR analysis using MS250 and MS251 identified all the clones carrying the wild type allele (Figure 3.2.16 a) and using MS252 and MS253 also confirmed that all the clones carry the wild type allele (Figure 3.2.16 b). Therefore I decided to focus on the deletion of the Group III H3 genes

3.2.2.3 Deletion of Group III H3 genes

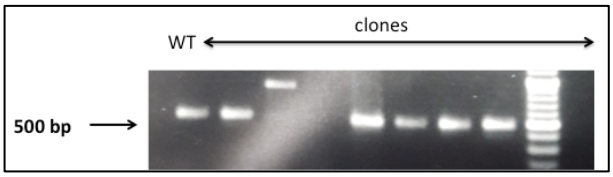
Similar to the strategy used for deletion Group II genes, I targeted Group III containing three H3 genes, h3C1, h3C2, and h3b that span across 30 kb on hist2. First, I aimed for direct deletion using px330 cas9 vectors to target the 5' upstream of h3C2 and 3' downstream of h3b (Figure 3.2.17 a). But the resultant clones did not have any KO.

Therefore, I decided to use again the Cre/LoxP system again, where I could integrate LoxP 5' upstream of h3C2 and 3' downstream of h3b in the same direction and express CRE for deletion. The resultant clones were analysed using PCR primers MS167 and MS168 with an expected product size of 540 bp for LoxP insertion and 500 bp on its absence on the 5' end of h3C2 and primers MS169 and MS170 with an expected product size of 440 bp for LoxP insertion and 400 bp on its absence on 3' end of H3d. PCR analysis using MS167 and MS168 showed no clones with insertions on 5' end (Figure 3.2.17 b). PCR analysis using MS169 and MS170 identified 1 out of 12 clones (clone 10) carrying LoxP site on 3' end (Figure 3.2.17 c). Sequencing confirmed a heterozygous insertion of LoxP site on clone 10. Repeating the transfection using the same gRNA on clone 10 did not result in a homozygous insertion.

(a)



(b)



(c)

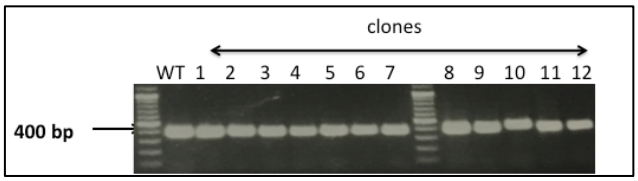


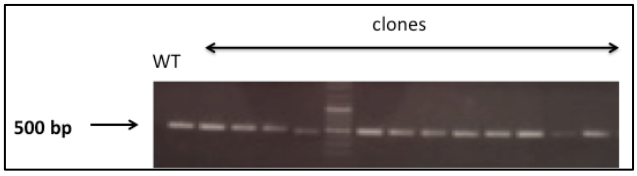
Figure 3.2.17: H3 LoxP KI strategy for h3C2, h3C1 and h3b on hist2.
(a) Strategy for H3 LoxP KI with gRNAs for cas9 target on 5' and 3' side of 30 kb KO region. MS167 and MS168 are PCR primers for screening the presence of LoxP insertion on 5' end and MS169 and MS170 for the 3' end. (b) PCR analysis of clones using MS167 and MS168 to screen for LoxP site insertion on 5' end result in a band of 500 bp similar to the wild type indicating the absence of targeted integration of LoxP site. (c) PCR analysis of clones using MS169 and MS170 to screen for LoxP site insertion on 3' end result in a band of 400 bp similar to the wild type indicating the absence of targeted integration of LoxP site. PCR product size of 440 bp on clone 10 indicates targeted insertion.

Several attempts with different gRNAs that were tested (data not shown) for introducing a LoxP site homozygously on 5' upstream of h3C2 and 3' downstream of h3b were

unsuccessful. This could be due to factors such as chromatin accessibility, CpG methylation or DNA flexibility interfering with the Cas9 binding.

Once again, we devised an alternative targeting strategy where we use a donor repair template to activate HDR pathway for the 30 kb KO. For this, ES cells were co-transfected with Cas9 vector with GFP containing two gRNAs directed to regions upstream and downstream of 30 kb and a homology directed repair (HDR) vector px393. The HDR vector contains 1 kb homologous sequence immediately upstream and 1 kb homologous sequence immediately downstream of the 30 kb KO region (homology arms) and puromycin resistance flanked by mutant FRT sites for marker recycling. Clones were selected for knock outs using both FACS for GFP and puromycin.

(a)



(b)

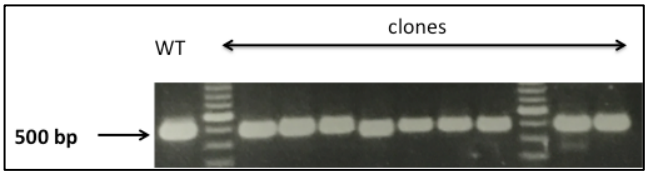


Figure 3.2.18: PCR analysis of h3C2, h3C1 and h3b genes KO on hist2.
(a) PCR analysis of clones using MS260 and MS261 to screen for Group III genes 30 kb KO on the 5' upstream of h3C2 gene. All the clones contain the WT allele, indicated by a 500 bp PCR product. (b) PCR analysis of clones using MS262 and MS263 to screen for Group III genes 30 kb KO on the 3' downstream of h3b gene. All the clones contain the WT allele, indicated by a 500 bp PCR product.

The resultant clones were analysed using PCR primers MS260 and MS261with an expected product size of 500 bp for the presence of WT allele and no product for KO on the 5' end of h3C2 gene and primers MS262 and MS263 with an expected product size of 500 bp for the presence of WT allele and no product for KO on the 3' end of h3b gene. PCR analysis using MS260 and MS261 identified all the clones carrying the wild type allele (Figure 3.2.18 a) and using MS262 and MS263 also confirmed that all the clones carry the wild type allele (Figure 3.2.18 b).

Currently, I am transfecting the Cas9 constructs again for Group I and Group III deletion and analyzing the clones.

4 Discussion

Histone tails provide a surface for interactions with other proteins[147] and are susceptible to a variety of post-translational modifications[148] including acetylation. Acetylation leads to a reduction of positive charges and thereby could diminishes the interaction with the negatively charged DNA backbone, potentially leading to an open chromatin state[85]. According to the histone code hypothesis, distinct histone tail modifications, act sequentially or in combination to form a 'histone code' that is read by other proteins causing distinct downstream events[149]. Although different interpretations of histone code exists, many studies suggest that the existence of such a code should manifest itself such that different modification combinations lead to distinct outcomes[150] More recently, a number of new PTMs have been identified in the globular domain of the histones and these modifications might directly influence nucleosome stability by affecting histone–DNA and histone–histone interactions, as well as the association with histone chaperones[102,103,108]..

Studies on globular domain histone acetylation including the ones from our lab demonstrate that H3K56, H3K64, H3K115 and H3K122 acetylation individually affect nucleosome stability and hence gene transcription[102,103,108,113]. Here, we demonstrate using an acetylation mimic that these modifications, together in a combinatorial manner can have functional implications on the nucleosome function, such as transcriptional activation *in vivo*.

4.1 Histone mutant cell lines

Covalent histone modifications are involved in a multitude of biological processes. Yet, genetic approaches to define the functional significance of histone PTMS have been limited. Histone genes are present in numerous copies in the genomes of all multicellular organisms. This unique circumstance helped histone genes to be among the first eukaryotic genes that were successfully cloned and molecularly characterized[151,152], but it also represented a major obstacle for proper genetic analysis of their function. Although genetic experiments were lacking, biochemical observations on the distribution of histone proteins within genomes and on posttranslational modifications have led to models that still await genetic validation.

Recently, studies have developed new tools using *D. melanogaster* to simultaneously manipulate canonical H3.2 and variant H3.3 *in vivo*[135]. In Drosophila melanogaster, there are

two H3.3 genes, H3.3A and H3.3B and 23 histone gene units (His1-His2B-His2A-His4-His3.2) have been annotated at cytological position 39D-E. Using transgenes, studies showed that canonical H3 and H3 variants can largely compensate for each other's absence and that the major difference resides in their mode of transcriptional regulation. The lack of H3.2 can be overcome by the H3.3 provided by the two wild-type His3.3 genes. Conversely, animals null mutant for both His3.3 genes are not only viable but also if H3.2 is expressed under the His3.3B promoter. However low amount of obtainable cells limit its application in biochemical studies. Many studies have been done using *S. cerevisiae* as model organisms[109,112,153] but it has limited evolutionary conservation with higher eukaryotes concerning the set of histone modifications and variant present, so, the functional data might not match completely. Moreover, heterochromatin largely diverges in budding yeasts, which limits the study of repressive PTMs in this system[154].

4.2 Chicken B cell line DT40

These limitations could be overcome by using chicken B DT40 cell line where enough cells for chromatin biochemistry could be obtained. Moreover, establishing H3 mutant cell lines without the background of endogenous H2 are more feasible in chicken cells due to "only" 7 copies of histone H3 genes, all organised within one cluster. Furthermore, the DT40 cell line has an exceptionally high ratio of targeted to random DNA integration by homologous recombination thus serving a suitable model for knock out and knock in experiments to target histone genes[136]. Studies have been done on histone genes knock outs in DT40 where 21 out 44 histone genes were targeted and deleted successfully[155]. Therefore, I initially decided to use DT40 cell line study the effects of H3K56/64/115/122 acetylation on nucleosome dynamics and hence transcription *in vivo.*

According to the sequence information from assembly Gallus_Gallus-4.0, April 2011, there was one H3.3 gene on chromosome 18 (H3.3B) and six copies of H3.2 on chromosome 1. Here, the strategy was to target H3.3 and make a heterozygote KO and mutate the remaining allele. However while analysing clones the final Genebuild (Gallus_Gallus-4.0, April/December 2013) was released and there were several updates on the histone genes. According to this version, there are two H3.3 genes each on chromosome 3 and 18, H3.3a and H3.3b respectively; eight H3.2 genes on the histone gene cluster, chromosome 1. At the same time the group of Julien Sale, Cambridge generated a H3.3 knock out DT40 cells that

we were able to obtain. Based on this, I revisited the strategies for establishing the mutant cell lines by doing knock in of an H3.3b gene expressing a tagged wild type and mutants in H3.3 deficient DT40 cells (with an endogenous H3.2 background) and to perform (i) a heterozygous KO of all H3.2 genes on Chromosome 1 and (ii) reduce the number of H3.2 genes on the other allele to the minimum number of copies with which the cells can survive and (iii) further do a KI of tagged H3.2 WT and mutants.

I transfected the cells with H3.3b WT tagged construct for homologous recombination. However we encountered an additional problem due to sequence variations between the chicken genome and the DT40 genome. Also, the genebuild contained many un-sequenced regions and gaps that caused problems in designing a targeting vector with endogenous tagged H3.3b. By analysing the recently published DT40 genomic DNA sequence[139], I once again targeted the H3.3 locus resulting in no knock in for tagged H3.3 WT. This could be due to several factors including inefficient recombination and chromatin inaccessibility in the targeted region. Use of CRISPR/Cas9 fused with GFP or RFP with a self-cleaving P2A sequence did not result in expression of GFP or RFP whereas the control HeLa cells showed GFP or RFP expression with 60% transfection efficiency. This might be due to improper P2A cleavage in DT40 cells, thus resulting in the lack of GFP/RFP expression.

Despite using multiple strategies with different conditions and techniques for gene editing in DT40 cells did not result in the generation of H3.3 and H3.2 mutants or knock outs, I decided to change the system. Though mouse has 14 copies of H3 genes including H3.1, H3.2 and H3.3, the precision of CRISPR/cas9 system which has been well demonstrated in mouse Embryonic Stem (ES) cells will allow me to manipulate ES cell genome efficiently.

4.3 Mouse embryonic stem cells (ES)

Homologous recombination applied to mouse embryonic stem (ES) cells has revolutionized the study of gene function in mammals[125]. But the frequency of homologous recombination events is considered to be rare compared to random integration events[124]. Latest discoveries in genome editing including TALENs and CRISPR system have proven to be efficient and faster in achieving precise genomic editing to generate knockout, knock-in and chromosomal rearrangements in cells and animals. Here, the strategy used to study the effects of H3K56/64/115/122 acetylation on gene transcription in ES cells using CRISPR system was to

make a homozygous KO of H3f3a, followed by a heterozygous KO of H3f3b and KI of tagged H3f3b mutants. For the canonical histones H3 that are spread on two histone gene clusters, hist1 and hist2, the strategy was to knock out H3 genes to a minimum number of copies with which the cells can survive and then knock in the tagged H3 mutants.

Using CRISPR/Cas9, efficient deletion of H3f3b of size 3.5 kb was achieved. Indeed, PCR analysis revealed heterozygous deletion in 17% (1/6) of the clones and homozygous deletion in 17% (1/6) of the clones. However, employing CRISPR/Cas9 to make a 11kb deletion of H3f3a gene was not successful. It has been a major limitation for CRISPR genomic editing that involves large DNA fragment insertion, deletion or replacement, that the larger the fragment is, the lower the efficiency [156]. So far, only a few of studies have demonstrated larger deletions (up to 95 kb) using two plasmids expressing sgRNA and Cas9 that were co-injected into embryos of animals including rodent and mice[157,158]. The efficiency also highly depends on several factors including chromatin inaccessibility, CpG methylation and DNA energy and flexibility[159]. Hence, I achieved H3f3a KO with an alternative strategy by deleting exons 1 and 2 of size 2.5 kb, encoding amino acids 1- 94. This deletion ensures that this truncated H3f3a does not get incorporated into the nucleosomes. PCR analysis revealed heterozygous deletion in 45% of the clones and homozygous deletion in 30% of the clones in ES cell lines indicating that the deletion of H3f3a gene occurred with high efficiency.

After reducing H3.3 genes to a single full-length copy, successful KI of HA tagged H3f3b and mutants were done using homology directed repair template (HDR). The mutants used for KI included 2xHA tagged -H3.3WT, -H3.3K56/64/115/122R, -H3.3K56/64/115/122Q, -H3.3K9/14/18R and -H3.3K9/14/18Q. PCR analysis of H3.3WT-2xHA showed that 17% of the clones contained targeted integration at the endogenous locus. Similarly, 2xHA -H3.3K56/64/115/122R, -H3.3K56/64/115/122Q, -H3.3K9/14/18R and -H3.3K9/14/18Q revealed that 11%, 12.5%, 50% and 12.5% clones from each mutants contained the targeted integration. The results reported here show that the sgRNA-mediated DSBs occur at a significantly higher frequency than reported in literatures[141].

Two recent studies reported a high off-target mutation rate in CRISPR/Cas9-transfected human cell lines (Fu et al., 2013; Hsu et al., 2013). Off-target mutations are observed at frequencies greater than the intended mutation, which may cause genomic instability and disrupt the functionality of otherwise normal genes. It is still one major concern when

applying CRISPR/Cas9 system to biomedical and clinical application[160]. Detecting off-target sites in a highly sensitive and comprehensive manner remains a key challenge in the field of gene editing. I initially used a T7 endonuclease I assay to detect off-target mutations, but this assay suffers poor sensitivity (it cannot detect off-target mutations that occur at frequencies <1%)[131], and it is neither practical nor cost-effective for large-scale screening. Deep sequencing methods allow more precise off-targeting detecting (measure off-target mutations at frequencies ranging from 0.01 to 0.1%)[146]. Therefore, the H3f3b mutant cell lines are being analysed for off-target effects using whole genome sequencing for potential indels.

Two H3f3b mutant cell lines, 2xHA -H3.3K56/64/115/122R and -H3.3K9/14/18Q had a different phenotype from the rest of the mutant cell lines. Both of the cell lines grew extremely slowly, looked dissociated and died within 30 days of selection. The H3.3K56/64/115/122R mutant failing to proliferate and die might indicate the importance of all four globular domain modifications for correct chromatin structure. Interestingly, studies from our lab on H3K64 methylation[161] indicated that H3K64me3 is a repressive mark that localizes at pericentromeric heterochromatin and could functionally oppose H3K64 acetylation. Investigations on functional impact of H3K64me3 using H3K64R mutants in NIH3T3 cells seemed to markedly impair normal enrichment of the hallmarks of constitutive heterochromatin, namely, H3K9me3, HP1 and H4K20me3, at pericentromeric heterochromatin foci in mouse NIH3T3 cells, thus severely compromising heterochromatin integrity, implying a key role for H3K64 modifications in reinforcing heterochromatic integrity at the pericentromere. The lack of proliferation of the H3K9/14/18Q mutant could be because the mimic of permanently acetylated lysine residues, which cannot be deacetylated, could perturb histone acetylation turn over that might be required for correct gene expression or because it affects the recruitment of reader proteins[162]. The phenotype of these two mutant cell lines could also be due to CRISPR off-target effects disrupting pluripotent genes such as Nanog and Oct4.

4.4 Expression levels of H3.3 mutants

The expression levels of H3.3-2xHA and H3 were analysed using western blot on acid extracted histones against anti HA, H3.3 and H3 in all three mutant cell lines and ES WT indicating the presence of HA tagged H3.3 with no endogenous H3.3. There seems to be a significant reduction in the levels of H3.3 when compared to ES WT. This result confirms the

presence of tagged H3.3 mutants resulting in the establishment of a clean H3.3 mutant system that has no background of wild type H3.3.

4.5 H3K56/64/115/122 acetylation enhances transcription *in vivo*

Studies from our lab on the role of H3K64 acetylation in gene expression involved overexpression of Flag-HA tagged H3.3WT, H3.3K64R and H3.3K64Q mutants in NIH3T3 cells with a background of wild type H3.3. Upon TPA stimulation, H3.3K64Q showed a two-fold higher transcription at c-fos gene over WT. Similarly, studies from our lab on H3K122 acetylation using S. pombe revealed that H3K122R mutant had delayed gene activation of nmt1+ compared to WT strain, suggesting H3K122 acetylation is important for rapid transcriptional gene activation *in vivo*.

Based on these studies, we hypothesised that the four globular domain residues on H3 might act synergistically or in combination to enhance gene transcription than a single mutation, *in vivo*. Here, all the three mutant cell lines along with the wild type: 2xHA –H3.3WT, H3.3K56/64/115/122Q, -H3.3K9/14/18R and ES WT were assayed for effects in TPA-dependent early-response gene activation in a system without the background of wt H3.3 and without histone overxpression. In this setup, H3.3K56/64/115/122Q enhanced the expression of genes including c-fos, egr1, and c-myc eight-fold, three-fold and two-fold respectively, above the levels obtained with ES WT, H3.3WT-2xHA or the acetylation-deficient H3.3K9/14/18R. The observed transcriptional effects due to the acetylation mimic of four residues in combination within the globular domain strongly suggest that H3K56/64/115/122 acetylation has an intrinsic ability to impact upon transcription mechanisms *in vivo*.

4.6 H3K56/64/115/122 acetylation enhances transcription during ES cell differentiation

Effects of globular domain modifications of H3 on mouse ES cell differentiation have not been extensively studied so far. Until now, studies have shown that H3K56 and H3K64 acetylation marks are strongly enriched at active pluripotency-associated genes such as Nanog, Pou5f1 and Dppa3, whereas after neuronal differentiation this enrichment shifts

towards active differentiation-associated genes such as *Pax6*, correlating these marks with active transcription[103,163].

After showing that all four globular domain H3.3K56/64/115/122 acetylation increases gene transcriptional activation *in vivo,* we wanted to investigate if they could have a similar effect on transcription during ES cell differentiation. Analysis of EB transcriptional profiles of T-Brachyury (mesoderm), Fgf5 (ectoderm) and Gata4 (endoderm) for all the four cell lines including ES WT during differentiation indicated that H3.3K56/64/115/122Q exhibit a dramatically higher induction of these genes when compared to other cell lines. The observed transcriptional effects due to the acetylation mimic of four residues in combination within the globular domain strongly suggest that H3K56/64/115/122 acetylation impact transcription during differentiation and try to moderate the degree, nature, and pace of ES cell differentiation.

4.7 H3 mutant cell lines

Using CRISPR/Cas9, a clean H3.3 mutant system (with endogenous H3 background) was achieved successfully. To target also H3.1 and H3.2 we grouped all the 12 canonical H3 copies into three groups: Group I containing four H3 genes spanning 40 kb on histone cluster 1, Group II containing three H3 genes spanning 18 kb on histone cluster 1 and Group III containing three H3 genes spanning 30 kb on histone cluster 2. We first targeted Group II genes KO, however, employing CRISPR/Cas9 to make a 18kb deletion of was not successful. Therefore, we employed Cre/LoxP system where genomic regions that are flanked by loxP sites in the same orientation result in deletion. PCR analysis at the sites of LoxP insertions revealed the presence of targeted integration on 5' end in 40% of the clones and on 3' end in 27%. Expression of Cre protein in a selected clone containing targeted insertions on both ends revealed that 18 kb deletion occurred in 70 % of the cells.

Having deleted Group II genes successfully, we employed a similar strategy for Group I genes (40 kb). But we did not detect any targeted insertion of LoxP sites in both the alleles. This could be due to various factors including sgRNA efficiency, chromatin inaccessibility, CpG methylation and DNA energy and flexibility[159]. Similarly, Cre/LoxP and CRISPR system were employed for the deletion of Group III genes of size 30 kb. But there was no targeted insertion in both the alleles in any of the clones.

So far, there are no studies on targeting og canonical histones genes in mammalian system. This could be because of the highly repetitive character or a special chromatin organization of these gene clusters of the canonical histone genes that pose a technical challenge for genetic manipulation. Targeting histone H3 genes individually also poses a technical challenge of employing multiple rounds of gene targeting in the same cell line and potentially accumulating a variety of off-target effects, which might misregulate genes that are essential for cell survival.

Also, all the three set of Groups contain not only H3 histones but also other canonical core histones, H2A, H2B and H4 and linker histone H1. So far, deletion of Group II is not found to be lethal to ES cells, but, homozygous deletion of Groups I or III might have a lethal phenotype.

5 Future Outlooks

Covalent histone modifications are involved in a multitude of biological processes. So far, studies have largely focused on histone tail modifications, which indirectly affect chromatin structures by serving as marks for the recruitment or activity of protein complexes that impact on chromatin function. But, recent studies including the ones from our lab showed that histone modifications within the histone fold domains of the nucleosome core have the potential to disrupt histone:histone or histone:DNA contacts and thereby directly alter nucleosome structure or stability. Results from the above experiments indicate that histone acetylation on four globular domain modifications H3K56/64/115/122 together enhance transcription *in vivo more* than a single modification. We are currently performing mRNA profiling using RNA sequencing to know the impact of these modifications on steady state gene expression compared to the tail modifications.

H3K56 and H3K64 are located within the histone fold at the region where the DNA enters and exits the nucleosome whereas H3K115 and H3K122 are located at the histone-DNA interface in the dyad axis of the nucleosome. We hypothesize that this difference could be pivotal for H3K56/64 acetylation and H3K115/122 acetylation to have distinct effects on nucleosome stability and chromatin structure. Therefore, we are currently establishing H3.3 mutant cell lines with H3.3K56/64R, H3.3K56/64Q, H3.3K115/122R and H3.3K115/122Q to address this hypothesis and effects on transcription *in vivo*.

When the presence or absence of one histone modification influences the occurrence of another modification, cross talk between the modifications occurs[164]. Cross talk can reinforce or reverse existing histone modifications. To address the question whether crosstalks occur between globular domain acetylation and other histone modifications, we are employing Mass spectrometric approaches on the H3.3 mutant cell lines. This approach would potentially identify *in situ* crosstalks between globular domain modifications and other histone modifications, and between tail modifications.

With this mutant system generated in this thesis, we will be able to address more specific questions about the effects of histone modifications *in vivo* and their impact on chromatin dynamics and structure. The system we have established here will provide a powerful tool to

mimic histone lysine modifications at specific residues *in vivo* and allow to further biochemically capture the molecular players involved in chromatin signaling pathways.

6 Conclusion

For the first time in mammals, we have established a “clean” H3.3 mutant system, as an important *in vivo* model to study the effects of histone acetylation on gene expression and chromatin structure without background of endogenous H3.3 histones. Our results suggest that H3K56, H3K64, H3122, and H3K115 acetylation act synergistically, establishing a strong link between H3K56/64/115/122 acetylation and active transcription *in vivo.* Our data suggest that these modifications have a direct impact on transcriptional activation and potentially influence ES cell differentiation when compared to the tail modifications. These results give us key insights into the importance of histone modifications *in vivo*, perfectly complementing previous *in vitro* studies from our lab. In future, it will be important to address this new group of modifications, how these modifications together affect chromatin and possibly their role in cancer.

7 Materials

7.1 Primers for gRNA cloning

Name	Gene target	Direction	Sequence 5’ to 3’
MS11	H3f3b 5’	F	CACCGggacccgagtctattcagca
MS12		R	AAACtgctgaatagactcgggtccC
MS13	H3f3b 3’	F	CACCGtgatattaccttcatgcacc
MS14		R	AAACggtgcatgaaggtaatatcaC
MS63	H3f3a 3’	F	CACCGttccaattcatggcgccgct
MS64		R	AAACagcggcgccatgaattggaaC
MS151	H3f3a exon2 5’	F	CACCGacctaaatcctgccgggcgg
MS152		R	AAACccgcccggcaggatttaggtC
MS153	H3f3a exon1 3’	F	CACCGgcttaattagcgctcgacac
MS154		R	AAACgtgtcgagcgctaattaagcC
MS143	H3f3b KI 5’	F	CACCGggagagagcttaagttgaag
MS144		R	AAACcttcaacttaagctctctccG
MS177	H3f3b KI 3’	F	CACCGtgcgagacaagcgacgttca
MS178		R	AAACtgaacgtcgcttgtctcgcaC
MS3	H3a 3’	F	CACCGgaggctaaggtagtctgccg
MS4		R	AAACcggcagactaccttagcctcC
MS32	H3c 5’	F	CACCGgaccactacggatgtcgaga
MS33		R	AAACtctcgacatccgtagtggtcC
MS158	h3b 3’	F	CACCGtcgctcttaggtatcacgct
MS159		R	AAACagcgtgatacctaagagcgaC
MS155	h3C2 5’	F	CACCGcttacggggctcaacgtcgg
MS156		R	AAACccgacgttgagccccgtaagC
MS77	h3g 5’	F	CACCGgcgaacgcaagcgtacttaa
MS78		R	AAACttaagtacgcttgcgttcgcC
MS79	h3d 3’	F	CACCGttgttgaaggacatatcgat
MS80		R	AAACatcgatatgtccttcaacaaC

7.2 Primers for HDR vectors cloning

Name	Gene target	Sequence 5' to 3'
MS6	H3a LoxP	tgacgtttaagactctgattggtttatgtttaagaactgcagttgaggctaaggtagtctga attcataacttcgtataatgtatgctatacgaagttatgccgcggtacttatttttgtatgtgt gtgttttaaagtaaatggaataaaattccagtga
MS34	H3c LoxP	tatggaagcatctccattcctacacctggaaaaggagcaatccggaccactacggatgt cataacttcgtatagcatacattatacgaagttatgaattcgagatggcagtgccctgagg ccttactcagacaacaaggctgtcaatgcagtttgaacct
MS65	H3f3a 3' LoxP	cccgaggatggcaaacgactgtagctaatagaaacaggctttatttccaattcatggcg cgaattcataacttcgtataatgtatgctatacgaagttatcgctaggagagctgctcctg gggggggagaaaaagccttttttcttttcaaggattcttaag
MS111	h3g 3' Lox2272	gtgggaaggaggggtacgagcgcgggtacgtgtgttgcgcgtgtgcgaacgcaagc gtacataacttcgtataggatactttatacgaagttatgaattcttaaaggccaaagtgcgc tacttaggtatctcacttttccctacggttacttgccatggc
MS112	h3d 5' Lox2272	tagaaaaacaaaagaaaagtggatgtgagtattgactgttgagcttgttgaaggacatat ataacttcgtataggatactttatacgaagttatgaattccgattggtaagttacttctcctcc agctaaataagctgattcaagccagattggattata
MS161	h3C2 5'	gacgactgaagattattgttgctcatttaatgcaatgtctgaggcttacggggctcaacga taacttcgtataggatactttatacgaagttatgaattctcggtggtaagtggactatgagt aatatccagtgataaagctttgttctctaacacttga
MS162	h3b 3'	tcgcttacagatccttaaaacaccaggttgaaataagttaactatcgctcttaggtatcaat aacttcgtataggatactttatacgaagttatgaattccgcttggaattttggggctatttttc aggttcacagttctgagaccctaattagaactcc

7.3 Primers PCR

Name	Gene target	Direction	Sequence 5' to 3'
MS27	H3f3b 5'	F	acttggttactataggtagcaac
MS28	H3f3b	R	gctattcacaccagcatc
MS29	H3f3b 3'	R	atgcctgacaacttcttattaac
MS55	H3f3a 5'	F	gcctgtacttagaaatctaacac

MS56	H3f3a 5’	R	gaatacatggtgatgttcaataatg
MS66	H3f3a 3’	F	ttctctagaagcatcgccttg
MS67	H3f3a 3’	R	atcctaattagcggccataattatc
MS173	H3f3a exon2 5’	R	accagggcatttctccatcag
MS174	H3f3a exon2 5’	R	ggtaaatgggtaggtggggaa
MS230	H3f3a exon1 3’	R	gaatgtttgcagacgtttc
MS218	H3f3b KI on HA	F	ctgggctagccctgatcaataac
MS288	H3f3b KI 3’	R	actcttgccgacccaactttc
MS51	H3c 5’	F	gttcagagttcacagaacag
MS52	H3c 5’	R	ctagcttctgttcatctctagc
MS53	H3a 3’	F	gtccagacagaggttcaaac
MS54	H3a 3’	R	cagaaactaaatgaatccttgatg
MS147	H3g 5’	F	tgtgcaggagttaaccaatcgg
MS148	H3g 5’	R	aggtcggtcttgaagtcctg
MS149	H3d 3’	F	cagaggctcttccttatgtaattc
MS150	H3d 3’	R	accctcatacagttcctctct
MS167	h3C2 5’	F	tgatatccagtgtatttcggtc
MS168	h3C2 5’	R	ttattcagactcctagcttgttaac
MS169	h3b 3’	F	gcctaacagttaccagtaatcac
MS170	h3b 3’	R	ttggttgatttacagcagagtc
DT001	H3.3b	F	tgtgcaattcccaaagcacg
DT002		R	gggctgacaggtggctttat
DT003	H3.3b	F	ctgcctcttgctggtgttcta
DT004		R	ttccccctgcgttcttcttc
DT125	H3.3b KI	F	taacaaccaatcagcggccg
DT126		R	ttttttgaattcagctctctctccccgtat

7.4 Primers cloning and site directed mutagenesis

Name	**Gene target**	**Direction**	**Sequence 5’ to 3’**
MS81	H3f3bK56/64R	F	ggaatggcagcctccggatgagcagctcagtcgatctctg

			gtaacga
MS82		R	tcgttaccagaGatcgactgagctgctcatccggaGgctgccattcc
MS83	H3f3bK115/122R	F	aactggatgtctctgggcatgatggtgactctcctggcgtggatg
MS84		R	catccacgccaggagagtcaccatcatgcccagagacatccagtt
MS85	H3f3bK56/64Q	F	ggaatggcagctgccggatgagcagctcagtcgattgctggtaacga
MS86		R	tcgttaccagcaatcgactgagctgctcatccggcagctgccattcc
MS87	H3f3bK115/122Q	F	aactggatgtcttggggcatgatggtgactctctgggcgtggatg
MS88		R	catccacgcccagagagtcaccatcatgccccaagacatccagtt
MS94	H3f3bK9/14/18R	F	tggccagctgtctgcgggggctctcccaccggtggacctcctagcggtc
MS95		R	gaccgctaggaggtccaccggtgggagagccccccgcagacagctggcca
MS96	H3f3bK9/14/18Q	F	tggccagctgttggcgggggcttgcccaccggtggactgcctagcggtc
MS97		R	gaccgctaggcagtccaccggtgggcaagccccccgccaacagctggcca
MS68	5' homology arm for H3f3b HDR KI	F	ttttttaagctttgccagtatagtaaccaggtg
MS69		R	ttttttaagcttgttgaagcggtttttatggc
MS73	3' homology arm + H3f3b before stop codon for H3f3b HDR KI	F	ttttttactagtagctctctcccccc
MS74		R	tttttttctagactattgtagagctttggaacc

MS75	2HA +STOP H3f3b HDR KI	F	ctagtttaagcgtaatctggaacatcgtatgggtaagcgtaatctggaacatcgtatgggtaa
MS76		R	ctagttacccatacgatgttccagattacgcttacccatacgatgttccagattacgcttaaa
DT26		F	aatgtcgacgggagttgcagttc
DT27		R	tggaagcttgcctcacatgcatctc
DT44		F	aaaaggtacctcatgtggtgcgtgtcctcgtagg
DT45		R	aaaaggtaccagctctctctccccgtatcctgcgag

7.5 Primers RT-PCR

Name	Gene target	Direction	Sequence 5' to 3'
H3f3aRT1	H3f3a	F	CAGCGCCGCCTCTCGCTTG
H3f3aRT2	H3f3a	R	CCCTTCGGATTCTCCGTTTCT

7.6 Primers qRT-PCR

Name	Gene target	Sequence 5' to 3'
OL1578	mC-fos F	AGGGGCAAAGTAGAGCAGCTA
OL1579	mC-fos R	CAGTCTGCTTTGTACGAGCCATGGTA
OL1588	mEgr1 F	TATGAGCACCTGACCACAGAG
OL1589	mEgr1 R	GCTGGGATAACTCGTCTCCA
OL1592	mC-myc F	GCTCTCCATCCTATGTTGCGG
OL1593	mC-myc R	TCCAAGTAACTCGGTCATCATC
OL157	mHprt1 F	GCTGGTGAAAAGGACCTCT
OL158	mHprt1 R	CACAGGACTAGAACACCTGC
OL1955	mNanog F	CGGTGGCAGAAAAACCAGTG
OL1956	mNanog R	ATGCGTTCACCAGATAGCCC
OL1957	mFgf5 F	AGTCAATGGCTCCCACGAAG
OL1958	mFgf5 R	TACAATCCCCTGAGACACAGC
OL1959	mGata4 F	TGGAAGACACCCCAATCTCG
OL1960	mGata4 R	ACATGGCCCCACAATTGACA

OL1961	mT brachyury F	CCAAGAACGGCAGGAGGATGTT
OL1962	mT brachyury R	CGTCACGAAGTCCAGCAAGAA

7.7 Plasmids

Vector	**Origin**
ploxNeo-H3f3bWT-2xHA	Self made
ploxNeo-H3f3bK56/64/115/122R-2xHA	Self made
ploxNeo-H3f3bK56/64/115/122Q-2xHA	Self made
ploxNeo-H3f3bK9/14/18R-2xHA	Self made
ploxNeo-H3f3bK9/14/18R-2xHA	Self made
pCAG-NLS-Cre	Raphael Margueron
px391	Cloning facility IGBMC
px392	Cloning facility IGBMC
px393	Cloning facility IGBMC
px394	Cloning facility IGBMC
px395	Cloning facility IGBMC
px396	Cloning facility IGBMC
ΔH3.3b/pLoxNeo	Self made
pLoxNeo-H3.3b-2xHA	Self made

7.8 Antibodies

Number	**Antigen**	**Species**	**Company**	**Catalog number**	**Dilution WB 4% BSA TBST**
152	H3	rabbit	Abcam	ab1791	1:50,000
155	HA	mouse	Dominik Krheuse	12CA5	1:20,000
343	H3.3	rabbit	Millipore	09-838	1:2000

7.9 Chemicals

Name	**Source or Specification**

Acetic acid	VWR, Darmstadt, Germany
Acrylamide 40%	SERVA Electrophoresis, Heidelberg, Germany
Acrylamide (ratio acrylamide to bisacrylamide 37,5 : 1)	Roth, Karlsruhe, Germany
AEBSF 4-(2-Aminoethyl)-benzensulfonylfluorid	Sigma-Aldrich Chemie, Steinheim, Germany
Agarose	Roth, Karlsruhe, Germany
Ammonium sulfate	VWR, Darmstadt, Germany
Ammoniumchloride	Roth, Karlsruhe, Germany
Ammoniumperoxodisulfate APS	Roth, Karlsruhe, Germany
Ampicillin (1000x stock 50 mg/ml)	Roth, Karlsruhe, Germany
Beta-mercaptoethanol	Roth, Karlsruhe, Germany
Bradford reagent	Roth, Karlsruhe, Germany
Bromphenol blue	Sigma-Aldrich Chemie, Steinheim, Germany
BSA	Roth, Karlsruhe, Germany
Calcium chloride	Merck, Darmstadt, Germany
Chloramphenicol (1000x stock 34 mg/ml)	Roth, Karlsruhe, Germany
Colloidal coomassie stock	Roth, Karlsruhe, Germany
Complete-EDTA-free protease inhibitor cocktail (tablet)	Roche, Basel, Switherland
Coomassie brilliant blue	Roth, Karlsruhe, Germany
Cystein-HCL	Roth, Karlsruhe, Germany
Di-potassium hydrogen phosphate	Roth, Karlsruhe, Germany
Di-sodium hydrogen phosphate Na_2HPO_4	Roth, Karlsruhe, Germany
DMSO Dimethyl sulfoxide	Roth, Karlsruhe, Germany
DTT Dithiothreitol	Applichem, Darmstadt, Germany
EDTA Ethylenediaminetetraacetic acid	Roth, Karlsruhe, Germany
Ethanol 100%	Roth, Karlsruhe, Germany
Ethidium bromide	Roth, Karlsruhe, Germany
Formaldehyde 37%	Sigma-Aldrich Chemie, Steinheim, Germany

Gel B	Roth, Karlsruhe, Germany
Glutaraldehyde 25%	Roth, Karlsruhe, Germany
Gluthatione	Sigma-Aldrich Chemie, Steinheim, Germany
Glyerol	Roth, Karlsruhe, Germany
HCl	Roth, Karlsruhe, Germany
HEPES	Roth, Karlsruhe, Germany
NP-40	Sigma-Aldrich Chemie, Steinheim, Germany
Imidazol	Roth, Karlsruhe, Germany
IPTG	Roth, Karlsruhe, Germany
Isoamylalcohol	Roth, Karlsruhe, Germany
Isopropanol	JT Baker, Phillipsburg, USA
Kanamycin (1000x stock 30 mg/ml)	Roth, Karlsruhe, Germany
Magnesium chloride MgCl2	Roth, Karlsruhe, Germany
Methonal	Roth, Karlsruhe, Germany
Nicotinamide	Alfa Aesar, Ward Hill, USA
N^{ε}-acetyllysine (KAc)	Bachem, Bubendorf, Switzerland
Orange G	Sigma-Aldrich Chemie, Steinheim, Germany
Paraformaldehyde (PFA)	Sigma-Aldrich Chemie, Steinheim, Germany
Poly-L-Lysine	Sigma-Aldrich Chemie, Steinheim, Germany
Ponceau Red S	Sigma-Aldrich Chemie, Steinheim, Germany
Potassium chloride Kcl	Roth, Karlsruhe, Germany
Potassium di-hydrogen phosphate KH2PO4	Roth, Karlsruhe, Germany
Silvervitrate AgNO3	Roth, Karlsruhe, Germany
Sodium acetate	Roth, Karlsruhe, Germany
Sodium azide	Roth, Karlsruhe, Germany
Sodium bicarbonate NaHCO3	Roth, Karlsruhe, Germany

Sodium Butyrate NaBut	Alfa Aesar, Ward Hill, USA
Sodium chloride NaCl	Roth, Karlsruhe, Germany
Sodium dihydrogen phosphate NaH3PO4	Roth, Karlsruhe, Germany
Sodium dodecyl sulfate SDS	Roth, Karlsruhe, Germany
Sodium hydroxide NaOH	Roth, Karlsruhe, Germany
Sodium Thiosulfate Na2S2O3	Sigma-Aldrich Chemie, Steinheim, Germany
Spectinomycin (1000x stock 50 mg/ml)	Roth, Karlsruhe, Germany
Sucrose	Roth, Karlsruhe, Germany
TCA Trichloroacetic acid	Sigma-Aldrich Chemie, Steinheim, Germany
TEMED Tetramethylethylenediamine	Roth, Karlsruhe, Germany
Tetracyclin (1000x stock 12,5 mg/ml)	Roth, Karlsruhe, Germany
Tricin	Roth, Karlsruhe, Germany
Tris-Base	Sigma-Aldrich Chemie, Steinheim, Germany
Tris-HCl	Sigma-Aldrich Chemie, Steinheim, Germany
Trition X-100	Roth, Karlsruhe, Germany
Tween-20	Sigma-Aldrich Chemie, Steinheim, Germany

7.10 Buffers

Name	**Composition**
5x Orange G sample loading buffer	50% glycerol 0.1% Orange G 5 mM EDTA pH 8.0
1x TBE buffer	5 mM EDTA pH 8.0 89 mM Tris

	89 mM Boric acid 2mM EDTA pH 8.0
1x TE buffer	10mM Tris-HCl pH 8.0 1mM EDTA pH 8.0
1x PBS	10mM Na2HPO4 2mM KH2PO4 2.7mM Kcl 137mM Nacl adjust to pH 7.2
10x SDS running buffer	247mM Tris 1.9M glycine 0.5% SDS
4x Laemmli sample loading buffer	250mM Tris-HCl pH 6.8 20% B-mercaptoethanol 2% SDS 0.1% bromophenol blue 40% glycerol
Coomassie brilliant blue staining	2% Coomassie brilliant blue 40% Methanol 10% acetic acid
10x Carbonate Buffer	10x Carbonate Buffer 100 mM NaHCo3 30 mM Na2CO3 adjust pH to 9.9 add 20% Ethanol to the 1x buffer
10x Transfer Buffer	250 mM Tris base 1.92 M Glycine adjust pH to 8.6
1x TBST	150 mM NaCl 50 mM Tris-HCl
Depurination Solution	0.25N Hcl
Denaturing Solution	0.5N NaOH 1.5M NaCl

Neutralizing Solution	0.5M Tris-HCl 1.5M NaCl
20X SSC	3M NaCl 300mM Na-Citrate
Low Stringency wash solution	2X SSC 0.1% SDS
High Stringency wash solution	0.5X SSC 0.1% SDS

7.11 Mammalian cell lines

Name	**Specification**
HeLa S3	Human cervix carcinoma derived (Transformed by HPV18) immortalized
MEF	Mouse embryonic fibroblasts
mES cells	Mouse embryonic stem cells W26, MPI Freiburg
H3.3K56/64/115/122Q-2xHA	Stably transfected with ploxNeo-H3f3bK56/64/115/122Q-2xHA
H3.3K9/14/18R-2xHA	Stably transfected with plox-Neo-H3f3bK9/14/18R-2xHA
H3.3WT-2xHA	Stably transfected with ploxNeo-H3f3bWT-2xHA
DT40	Chicken B cell line DT40, MPI Freiburg
DT40 H3.3KO	Total knock out of H3.3a and H3.3b in DT40 cell, from Julien Sale lab, Cambridge

8 Methods

8.1 PCR

DNA fragments were amplified by Polymerase chain reaction (PCR) from a DNA template with sequence-specific oligonucleotide primers. For cloning purposes the phusion polymerase (NEB) with high-fidelity was used.

A typical PCR reaction consisted of the following:

Component	**Final Concentration**
Nuclease-free water	
5X Phusion HF or GC Buffer	1X
10 mM dNTPs	200 µM
10 µM Forward Primer	0.5 µM
10 µM Reverse Primer	0.5 µM
Template DNA	< 250 ng
DMSO (optional)	3%
Phusion DNA Polymerase	1.0 units/ 50 µl PCR

Step	**Temp**	**Time**
Initial Denaturation	98°C	5 minutes
Annealing 33 cycles	98°C 50-72°C 72°C	10 seconds 30 seconds 30 seconds per kb
Final Extension	72°C	10 minutes
Hold	4°C	

The annealing temperature was adjusted according to the Tm of the oligonucleotides used.

8.2 PCR Mutagenesis

The QuikChange II site-directed mutagenesis kit is used to make point mutations, on single or multiple adjacent amino acids. The QuikChange II site-directed mutagenesis method is performed using *PfuUltra* high-fidelity (HF) DNA polymerase. The basic procedure utilizes a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest and two synthetic oligonucleotide primers, both containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by *PfuUltra* HF DNA polymerase, without primer displacement. Extension of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn* I. The *Dpn* I endonuclease (target sequence: 5′-Gm6ATC-3′) is specific for methylated and hemi-methylated DNA and is used to digest the parental DNA template and to select for mutation containing synthesized DNA. The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue super-competent cells.

A typical PCR reaction consisted of the following:

Component	**Final Concentration**
Nuclease-free water	Up to 50 µl
10X Reaction Buffer	1X
10 mM dNTPs	200 µM
10 µM Forward Primer	125 ng
10 µM Reverse Primer	125 ng
Template DNA	20-50 ng
DMSO (optional)	3%
PfuUltra HF DNA polymerase 2.5 U/ µl	1 µl /50 µl PCR

The cycling parameters are as follows:

Step	Cycles	Temperature	Time
1	1	95°C	30 seconds
2	12–18	95°C	30 seconds
		55°C	1 minute
		68°C	1 minute/kb of plasmid length

Following temperature cycling, place the reaction on ice for 2 minutes to cool the reaction to ≤37°C. Then, add 1 µl of the *Dpn* I restriction enzyme (10 U/µl) directly to each amplification reaction. Gently and thoroughly mix each reaction mixture by pipetting the solution up and down several times. Spin down the reaction mixtures in a micro centrifuge for 1 minute and immediately incubate each reaction at 37°C for 1 hour to digest the parental (i.e., the non-mutated) supercoiled dsDNA. $1/10^{th}$ of the PCR reaction was transformed into bacterial colonies.

8.3 Agarose gel electrophoresis

DNA-containing samples were mixed with 6x bromophenol blue loading dye and separated on 1 to 2% agarose gels containing 1µg/ml Ethidium Bromide in 1x TBE, according to the expected size of the DNA fragment. Agarose gels were then analysed under 254 nm UV light. If further processing of the DNA was desired the band of expected size was cut out of the gel and recovered from the agarose gel employing the GenElute Gen Extraction Kit (Sigma) according to the manufacturer's instructions.

8.4 Digestion with restriction endonucleases

All enzymes used for restriction endonuclease digestions were purchased from Thermo Scientific or NEB and used with one of the supplied buffers according to the conditions specified by the manufacturer. For simultaneous digestion with two different enzymes the optimal buffer was determined using an online tool provided by NEB (https://www.neb.com/tools-and-resources/interactive-tools/double-digest-finder).

8.5 Dephosphorylation of linear DNA

To prevent religation of the incompletely digested plasmids the 5' phosphate group was removed with shrimp alkaline phosphatase (SAP, Fermentas). The digested plasmid DNA was incubated at 37°C for 1h with 1U SAP after restriction endonuclease digest. Digested and dephosphorylated DNA was isolated by agarose gel electrophoresis and recovered using GenElute Gen Extraction Kit (Sigma) according to the manufacturer's instructions.

8.6 Ligation

Digested insert DNA was ligated into digested and dephosphorylated plasmid DNA using T4 DNA Ligase (NEB). Insert and vector DNA were mixed in a molar ratio of approximately 2:1. 1U of T4 DNA Ligase (NEB) along with the 1x ligase buffer and was used. A typical ligation reaction was performed overnight at RT in a final volume of 10µl of 1x ligation buffer (NEB).

8.7 Precipitation of nucleic acids

For nucleic acids purifications, samples were supplemented with 3M sodium acetate pH 5.2 (1/10th of sample v/v) and 2.5 volumes of 100% ethanol (v/v) or 2 volume of isopropanol (v/v). Ethanol precipitated samples were incubated at -20˚C for 1 hour to overnight, whereas isopropanol samples were kept for 10-20 min at room temperature to avoid excessive salt precipitation. For samples of low quantity, 10 µg of glycogen (Roche) was added to the solution and ethanol was always used for precipitation. After centrifugation at 20000x g for 30 minutes at 4˚C, the DNA pellet was washed once with cold 70% ethanol, pelleted again and air-dried completely. The pellet was resuspended in water or 1x TE buffer pH 7.5-8.0 or in 1x Blue loading dye for agarose gel electrophoresis.

8.8 Heat shock transformation of E. coli

Chemically competent E. coli cells were thawed on ice 5 µl of a ligation mixture or 1 µl of plasmid DNA were added and incubated for 15 mintues on ice. The bacteria were transformed by applying a heat shock for 45 sec (XL-1Blue) at 42°C. 1000µl of LB medium were added and the bacteria were allowed to recover for 60 minutes at 37°C before plating on the appropriate selection medium.

8.9 Purification of plasmid DNA from E.coli

Plasmid DNA was purified with GenElute™ Plasmid Miniprep/Maxiprep Kit (Sigma). Single colonies of transformed E.coli cells were picked from agar plates to inoculate cultures of LB media containing the appropriate antibiotics. Bacterial cultures were grown and purified as according to the manufacturer's instructions.

8.10 RNA isolation and Reverse Transcription from mammalian cells

Total RNA was extracted using Quick-RNA mini-prep kit from Zymo Research according to the manufacturer's instructions. RNA concentration was measured utilizing a NanoDrop ND) 1000 spectrophotometer (Nanodrop). 1µg of total RNA was used for each reverse transcription reactions. cDNA was synthesized with the RevertAid™ First Strand cDNA Synthesis Kits (Fermentas) using oligodT-primer according to the protocol provided by Fermentas. cDNA was conveniently diluted in sterile water (Sigma) prior to RT-PCR or qPCR analysis.

8.11 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

A semi-quantitative RT-PCR for all the samples was performed for 30 cycles using the 100 ng of synthesized cDNA in a total volume of 20 µl. The reaction mixture contained 0.5 µM of each forward and reverse primer, 0.25 mM dNTP, 0.5 unit of Taq DNA polymerase, 2 µl of 10x PCR Buffer II and 2 mM $MgCl_2$. The RT-PCR reaction is then resuspended in 1x Blue loading dye for agarose gel electrophoresis.

8.12 Quantitative PCR

Quantitative PCR (qPCR) with the 7500 Fast Real-Time PCR System (Applied Biosystems) and SYBR-Green detection method was carried out as follows: A 11 µl reaction consisted of 12.5 µl SYBR Green PCR Master Mix (Applied Biosystems), 1 µl of 1 µM forward and reverse primer mix (100nM final), 100 ng of cDNA as template and up to 25 µl ddH2O. PCR was conducted for 42 cycles and changes in fluorescence measured at the end of each amplification step. Ct-values were calculated from the exponential phase of amplification for each primer pair. Target gene expression was normalized to the expression of HPRT1

housekeeping gene. The fold changes were calculated using the ddct-method. Specificity of PCR products was checked by gel-electrophorsis and/or melting curve analysis.

8.13 SDS-PAGE

SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) is a method for separation of proteins according their molecular weight. The percentage of polyacrylamide in the gel was adjusted according to the size of the protein that was studied and was for the running gel between 6% and 18.7% in 375mM Tris pH8.8 and 0.1% SDS. The stacking gel consisted of 5% polyacrylamide in 125mM Tris pH6.8 and 0.1% SDS. Protein samples were boiled in 1x laemmli loading buffer at 95°C for 5min and electrophoresed in 1xSDS running buffer at 140-200V for 1h-2h, depending on percentage of gel. Proteins were either stained with coomassie brilliant blue or transferred to nitrocellulose membranes for immunoblotting.

8.14 Western Blot transfer

To detect proteins or modifications of proteins by immunostaining the proteins have to be transferred from the gel matrix to a membrane to render antigens accessible for antibodies. Proteins resolved on SDS page were therefore transferred to nitrocellulose membranes. Histones were transferred to membranes of 0.1µm pore size in 1x Transfer buffer at 200mA for 1 hour. Bigger proteins were transferred 0.45µm pore size membranes in 1x Transfer buffer supplemented with 0.05% SDS at 200mA for 2 hours. Efficiency of transfer of proteins to the membranes was checked using ponceau red staining solution.

8.15 Immunostaining of western blots (Immunoblot)

Nitrocellulose membranes were blocked in TBST with 4% BSA at RT for 1 hour. Different dilutions of antibody were incubated with the membrane in TBST with 4% BSA and incubated with gentle agitation at 4°C O/N. The membranes were washed with three changes of TBST for 10 mintues each and incubated with a secondary antibody conjugated to horseradish peroxidase (HRP) in TBST with 4% BSA at RT for 1 hour. The membrane was washed again with three changes of TBST for 10 minutes before antibody binding was detected by incubation of the membrane with a chemiluminescent HRP substrate (Millipore).

8.16 Purification of native histones by acid extraction

Histones are rich in basic amino acids and can therefore be separated from the majority of other nuclear proteins by acid extraction. Cells were harvested and washed in ice-cold PBS once. The cell pellet was resuspended in PBS containing 0.5% Triton-X100 (v/v), protease inhibitors and 10 mM NaBut at a cell density of 10^7cells/ml. The cells were incubated on ice for 10 minutes to lyse the cell membrane, pelleted and washed again in half of the volume of the same buffer. The nuclear pellet was acid-extracted by resuspension in 0.2M HCl and incubation at 4°C O/N on spinning wheels. After acid extraction the sample was centrifuged at 17000xg for 15 minutes and the acidic supernatant containing the histones transferred to new tube. Histone were stored at -20°C or -80°C and were neutralized with 1/2 volume of 1M Tris pH8.8 prior running on SDS-PAGE.

8.17 Cell culture

All cell cultures were maintained at 37°C in a humidified atmosphere containing 5% CO2. DT40 cell were maintained in RPMI supplemented with FBS PAA 10% heat inactivated, chicken serum 1% heat inactivated, 1% Pen Strep 100x, 1:1000 of mercaptoethanol from Millipore for ES cells, at 70-80% confluency. Mouse embryonic stem cells where grown on gelatin-coated plates and were maintained in 500 ml DMEM – glutamax + 4.5g Glucose, 15% ES tested, non heat inactivated FCS (Dutscher P130207ES), 1% Pen Strep 100x, 1%, Non essential amino acids NEAA 100x, 0.1mM BetaME, 2x LIF, at 70-80% confluency.

For gene induction, ES cells were grown at 80% confluency for 24 hours in regular growth medium and starved without FCS for 24 hours. ES cells were stimulated by replacing the starving medium with regular growth medium containing 15% FCS and 100 nM TPA for definite time points.

8.18 Transfection with lipofectamine 2000

Transfections are done in 6 well plates with 1 million cells and 2 ug DNA. Mouse ES cells are grown in complete medium and 30,000 cells are plated in a 6 well plate, the day before transfection. On the day of transfection, 10 µl of lipofectamine 2000 is added to 240 µl of DMEM without any additives in a separate tube. In another tube, DNA of 2 µg is added to DMEM without additives to make up to 250 µl. Both the solutions are mixed and incubated

for 5 minutes as room temperature. The final mixture is slowly added to the cells in suspension and are analysed after 48 hours using immunofluorescence or FACS.

8.19 ES cell differentiation

ESC differentiation was carried out ineither Serum/LIF, Serum (with no added LIF) or N2B27 alone. For embryoid body differentiation, E14tg2a ESCs were trypsinised and washed in PBS, before being resuspended in Serum-containing medium (with no LIF) for counting. 7.5x105 cells were plated in each of 4 uncoated bacterial dishes in 10 mL Serum-containing medium per cell line. Each day, adherent embryoid bodies were blown off the plate using a 5 mL pipette to wash medium over the surface and one plate was collected for RNA extraction by pipetting the embryoid bodies into a Falcon tube and centrifuging for 3 min at 1400 rpm. On day 2 (D2) of the differentiation procedure, the embryoid bodies were transferred into a tube and allowed to settle to the bottom. The medium was aspirated and they were resuspended in 10 mL new medium, using a 10 mL pipette, and transferred to a new uncoated dish. On D3 and D5 each plate was split into two plates. On D7, the embryoid bodies were transferred to a gelatin-coated tissue culture dish and allowed to attach. Beating cardimyocytes were visible within three days.

8.20 CRISPR/Cas9

All gRNA cloning and CRISPR/Cas9 genome editing was done according to the protocol published in Fan Ann et al, *Nature Protocols* **8**, 2281–2308 (2013).

8.21 Southern blotting

On day 1, Digest 10 μg of genomic DNA with 1 μl of the appropriate enzyme making the final volume 40 μl. Run the digest overnight (ON). On day 2, Make sure the Gel box sits level (Use paper towels to adjust until level). Cast a 0.8% Agarose gel in 1x TAE (For the mini-gel boxes, 100 ml are needed and the regular size gel boxes need 150 ml). Heat the solution in the microwave for ~ 3 minutes (about 2’ 30”) until all the agarose is dissolved BUT the solution does not boil. Cool off with cold tap water until the bottom is cool enough to touch with a bare hand. Pour into appropriate gel box and wait for 30 minutes. Gel will be cloudy when it is finished. Move gel (along with the U- shaped caster) to the gel box. Fill the

box with 1x TAE. Make sure to remove all bubbles below the gel caster. Then, remove samples from the 37°C water bath. Add 10 µl of Loading Dye to sample and heat the samples for 1 to two minutes at 65°C. Spin the samples down and load 50 into each well. Run Gel at 30 Volts for about 18 hours. The Dye Front should be at least 2/3 of the way into the gel before stopping. Remove gel from box and caster and put into Staining Solution [50 µl of EtBr in 200 ml of TAE]. - Rock at speed 10 for 30 minutes. Gently rock the gel(s) in Denaturation solution (0.5M NaOH, 1.5M NaCl) once for 30 minutes, wells down. Gently rock the gel(s) in Neutralizing solution (.5M Tris pH 7.5, 1.5M NaCl) twice for 20 minutes each time, wells down. Using a large glass Pyrex baking dish (one for two gels), fill with 10X SSC and place two glasses on the dish. Use Whatman paper as a bridge (use a laminated template to cut bridge width to the longest dimension of the membrane). Fill with more 10X SSC if needed (~1.5L total).

Wearing gloves, write the name of the gel (A, B, 1, 2, etc.) on the upper right corner of the membrane on the DNA side with pencil. Cut the lower right corner for orientation reference. Using a laminated template, cut 3 Whatman layers to the same size as the membrane (12cm x 10cm) for each gel to be blotted. Pour ~50mL 2X SSC into a small rectangle glass baking dish. Pour 10X SSC over the bridge with a transfer pipette to wet thoroughly- place gel(s) wells down on the Whatman paper and remove air bubbles by rolling a 10mL glass pipet across the gel. Dip membrane(s) in 2X SSC, and place on gel(s) with writing side down. Remove air bubbles. Dip 3 Whatman squares/gel in 2X SSC, and place them on top of each membrane. Use strips of Parafilm (3 X 1 sections) to butt up against the gel(s) on all sides to create a barrier between the bridge and the paper towels on top (next step). This will maintain capillary flow through the gel instead of around it. Unfold white paper towels to place on top of (both) gel(s). Use 3/4 pack of paper towels for 2 gels. Place a glass baking dish on top for weight. Leave overnight. The next day, break down blot carefully, discarding the paper towels and 2 Whatman layers. Gently pull off membrane(s) and dip in 2X SSC, then dry blot a little bit with Whatman paper that was on top of blotting paper stack (Figure). Put membranes into the tubes to hybridize or store between damp (with 1X Maleic Acid Buffer or 2X SSC) Whatman paper wrapped in plastic wrap at 4°C.

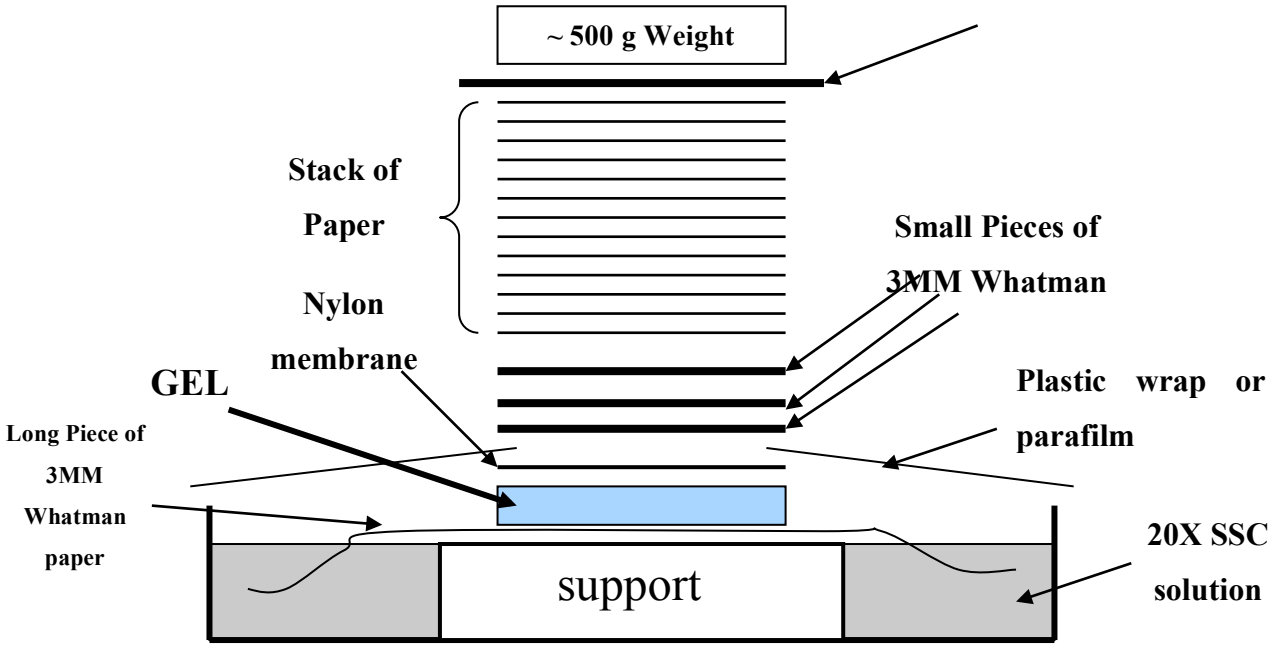


Figure 8.1: Capillary transfer of DNA.
DNA ransfer from an agarose gel to a nylon membrane. 20X SSC solution is drawn from a reservoir and passes through the gel into a stack of paper towels. The DNA is eluted from the gel by the moving stream of 20X SSC solution and is deposited on the nylon membrane. A weight on the top of the paper towels helps to establish a tight connection between the layers of material used in the transfer system. Plastic wrap or parafilm helps to prevent the paper towels from touching the long piece of 3MM Whatman paper below the gel. If such touching occurs, 20X SSC solution is drawn directly from the reservoir to the paper towels. Thus, there would little or no capillary transfer of DNA to the nylon membrane.

8.21.1 Prehybridizing the Blots

1. Thaw the following components that are stored in a -20^{o}C freezer in room 2836A:
   - a dark brown bottle of 100% formamide in a 55^{o}C water bath. Once thawed out, the bottle can be kept at room temperature during preparation of the pre-hybridization and hybridization solutions.
   - a 50-mL Falcon centrifuge tube of 50X Denhardts solution in a 42^{o}C water bath.
   - a 14-mL tube of 10 mg/mL sssDNA in either a 42^{o}C or 55^{o}C water bath. Once thawed out, the tube can be kept on ice.

2. Prepare pre-hybridization and hybridization solutions in two 50-mL Falcon centrifuge tubes by adding components shown in Table 1 to each tube.

Add denatured sheared salmon sperm DNA (or component vii) into the pre-hybridization/hybridization solution immediately before the solution is added to a hybridization bag containing the blot.

- Prepare a denatured sssDNA solution as follows:
    - Heat 150-200 mL of deionized water to boiling in a 250-mL round deep dish using a hot plate.
    - Pipet 0.6 mL of 10mg/mL sssDNA into a sterile 1.5-mL microcentrifuge tube.
    - Put this tube on a small rack floating on the boiling water and keeping the tube in the boiling water for 5 minutes.
    - After 5 minutes of boiling, immediately transfer the tube containing denatured sssDNA on ice for at least two minutes before adding it to the pre-hybridization or hybridization solution. This is called "quenching" step to prevent re-annealing of denatured complement strands of sheared salmon sperm DNA fragments.

*:10 mg/mL sssDNA is added into the pre-hybridization/hybridization solution immediately before the solution is added to a hybridization bag containing the blot.

Warm the pre-hybridization solution in a 42ºC water bath for at least 30 minutes. Store the tube of hybridization solution in either the 42ºC water bath (if hybridization will be set up several hours later) or the fridge until needed (if hybridization will be set up on the next day).

3. Add the warmed pre-hybridization solution to the seal-a-meal bag containing the blot.
4. Squeeze out as much air bubbles as possible from the bag.
5. Seal the open end of the bag with the heat sealer.
6. Put the bag on a plastic container or a Pyrex glass dish.
7. Incubate the bag for at least 1-2 hours at 42ºC in an air incubator with ~100-rpm shaking.

8.21.2 Preparing Radioactively Labeled DNA Probes Using Prime It II Random Prime Labeling Kit (Stratagene)

Randomly Primed DNA Probe Reaction.

Components	Volume
i. 25-50ng DNA template	___ mL
ii. 5X Random primers solution	10 mL

iii. Sterile double-distilled H_20	___ mL
iv. 5X dCTP buffer	10 mL
v. a- ^{32}P dCTP	5 mL
vi. Klenow fragment (Large fragment of DNA polymerase I)	1 mL
Total volume	50 mL

1. Thaw out tubes of components ii and iv that are stored in a -20^{o}C freezer (room 2836A). Once thawed out, keep these tubes on ice.
2. Add components i - iii to a 1.5-mL microcentrifuge tube. Close the lid tightly.
3. Boil the components in the tube for 5 minutes in a round deep dish containing boiling water on a hot plate.
4. After 5 minutes, immediately, quench the tube on ice for at least 2 minutes.
5. Spin down water condensation on the lid of the tube in a microcentrifuge at room temperature for ~30 seconds. Put the tube back on ice.
6. Add components iv-vi into the microcentrifuge tube and mix the contents thoroughly by pipetting up and down several times.
7. Incubate at 37ºC for 10 minutes to generate randomly primed probes.
8. Meanwhile, prepare a Sephadex-100 column to purify labeled DNA probes from unincorporated a- ^{32}P dCTP nucleotides as follows:
   - Get a 1cc or 1 mL syringe from a "General Use" drawer.
   - Remove a cotton filter from a 1-mL disposable pipet using a pointed-end forceps.
   - Stick the loose end of the cotton into the syringe.
   - Use the plunger to push the cotton piece to the bottom of the syringe. Slowly, withdraw the plunger.
   - Wet the cotton piece with TNE buffer using a disposable transfer pipet.
   - Use a disposable transfer pipet to add resin solution (G100 Sephadex) until resin itself is above the 1.0-cc mark on the syringe.
   - After 10 minutes of incubation at 37ºC, add an equal volume (50 mL) of 2X Nick translation dye (containing bromophenol blue and dextran blue dyes) to the labeling reaction.
   - Mix the contents well by pipetting up and down several times. Then add the whole mixture (~100 mL) to the syringe/resin.

- Wait until the dyes completely enterring the resin. Add ~100 mL of TNE buffer at a time.
- The dextran-blue dye (faster migrating dye) co-migrates with the probes. Collect this fraction. Stop the collection when clear liquid starts to come out.
- When done, use a new tube to stop the flow and discard the syringe/resin in a plastic bag sitting in a 1-L plastic beaker dedicated as a radioactive dry waste container.
- Use a P-1000 pipetman to measure the final volume of the probe solution as well as to mix the probe solution.

8.21.3 Hybridizing the Blots with the Denatured Radioactive Labeled Probes

1. Boil the labeled probe and sheared salmon sperm DNA solution (sssDNA) in separate 1.5-mL microcentrifuge tubes for five minutes.
2. Meanwhile, warm the hybridization solution to 42^{o}C in a water bath.
3. Immediately, quench the tubes on ice for at least two minutes. Spin tubes in a microcentrifuge for 30 seconds.
4. Cut a corner out of the pre-hybridization bag. Pour out the pre-hybridization solution into a sink. Squeeze out almost all of the solution in the bag.
5. Add denatured sssDNA to the hybridization solution, then pipet this solution mix into the seal-a-meal bag containing the blot using a 10-mL or 25-mL disposable pipet.
6. Behind a plexi-glass shield, add the probe to the hybridization bag.
7. Seal the cut corner of the hybridization bag.
8. Eliminate bubbles by pushing them away from the blots. Seal the bag to separate bubbles from the blot.
9. Hybridize the blot at 42^{o}C in the air incubator or water bath for at least 16 hours.
10. Record time of starting hybridization.

8.21.4 Washing the Blots

1. Prepare washing solutions in two 1-L Erlenmeyer flasks as shown below:

Final Concentration	Low Stringency (2X SSC/0.1% SDS)	High Stringency (0.2X SSC/0.1% SDS)	Stocks
2X SSC	70 mL	N/A	20X
0.1% SDS	7 mL	7 mL	10%
0.2X SSC	N/A	7 mL	20X
ddH20	up to 700 mL	up to 700 mL	
Total Volume	700 mL	700 mL	

2. Warm the high stringency wash solution to 60^{o}C in a water bath.
3. Keep the low stringency wash solution at room temperature.
4. Pour ~ 350 mL of the low stringency wash solution into a Pyrex glass dish.
5. Write on a 50-mL Falcon centrifuge tube "Name of the hybridization solution", "your initial", and "Date".
6. Remove the hybridization bags from the 42^{o}C incubator or water bath incubator.
7. If the water bath incubator is used, change the temperature from 42^{o}C to 60^{o}C.
8. Behind a plexi-glass shield, cut a corner of the hybridization bag.
9. Carefully, pour the hybridization solution into the labelled 50-mL Falcon centrifuge tube.
10. Store the hybridization solution in a 50 mL Falcon centrifuge tube in a -20^{o}C freezer.
11. Cut the hybridization bag. Remove the blots from the bag. Immediately, place the blots one-by-one into a Pyrex glass dish containing 350 mL of low stringency wash solution.
12. Place the Pyrex dish on an orbital shaker and turn the speed dial to ~ 50-75 rpm. Wash the blots at room temp for 15 minutes.
13. After 15 minutes of washing, use a blunt-end forceps to hold the blots and pour the washing solution into a liquid waste radioactive container behind a plexi-glass shield.
14. Repeat steps 11-13 for the second wash with a fresh 350 mL of low stringency wash solution.
15. While washing the blots in the low stringency wash solution, prepare another Pyrex dish containing ~ 350 mL of high stringency wash solution. Leave this dish in a 60^{o}C water bath incubator.

16. After 15 minutes of the second low stringency wash, transfer the blots into the Pyrex dish containing 350 mL of high stringency wash solution at 60^{o}C using blunt-end forceps. Wash blots in the high stringency wash at 60^{o}C for 30 minutes.
17. While washing the blots, pour off the low stringency wash solution into the liquid waste radioactive container. Rinse the dish with small amount of water using squirt bottle. Use the Geiger counter to monitor the amount of labeled probes that remained on the dish. Wash the dish with a 7x Detergent solution and rinse it with water until radioactive lable cannot be detected with the Geiger counter.
18. Repeat the high stringency wash with a fresh 350 mL of high stringency wash solution.
19. While washing, cut out a piece of 3MM Whatman paper to a size of 8 inch x 10 inch. Wrap the paper with a piece of plastic wrap. (TA will demonstrate).
20. After washing the blots, briefly blot the blots on two or three layers of Kimwipe tissues to remove excess liquid. Place the blots (DNA side facing away from you) on the plastic-wrapped 3MM Whatman paper prepared in step 19.

8.21.5 Exposing the Blots to X-Ray Film

1. In the lab, place the wrapped blots facing an intensifying screen in a vinyl X-ray cassette.
2. Bring the cassette, 2 pieces of small plexi-glass plates, 4 clamps, and a box of X-ray film to the darkroom on the first floor.
3. In the darkroom, remove a piece of X-ray film from the box.
4. Bend an upper right corner of the film to mark the orientation of the film relative to the blots.
5. Put X-ray film between the blots and the intensifying screen in the cassette.
6. Close the cassette.
7. Put the plexi-glass plates on both sides of the cassette.
8. Clamp the plates with clamps.
9. Expose the blot to X-Ray film in a -80^{o}C freezer from several hours to a few days depending on how hot the blot is.

8.21.6 Developing the Exposed X-ray Film (Autoradiography)

1. Remove the cassette from the -80^{o}C freezer and leave it in the lab until the cassette warms up to room temperature (~ 15 - 30 minutes).
2. Take the cassette to the darkroom.

3. Develop the exposed X-ray film (autoradiogram) using the Kodak Film Developer.
4. Align the autoradiogram to the blots. Mark well positions on the blots to the autoradiogram. Write on the autoradiogram positions of the 1-kB ladder, super pool number, etc.
5. Re-expose the blots to X-ray film for longer exposure.

9 Bibliography

1. Speybroeck, L. From Epigenesis to Epigenetics. *Ann. N. Y. Acad. Sci.* **981,** 61–81 (2006).
2. Bird, A. Perceptions of epigenetics. *Nature* **447,** 396–8 (2007).
3. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* **128,** 635–638 (2007).
4. Groth, A., Rocha, W., Verreault, A. & Almouzni, G. Chromatin Challenges during DNA Replication and Repair. *Cell* **128,** 721–733 (2007).
5. Soshnev, A. A., Josefowicz, S. Z. & Allis, C. D. Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol. Cell* **62,** 681–694 (2016).
6. Rill, R. & Van Holde, K. E. Electric dichroism of chromatin. *J. Mol. Biol.* **83,** 459–471 (1974).
7. Wolffe, A. P. & Hayes, J. J. Chromatin disruption and modification. *Nucleic Acids Research* **27,** 711–720 (1999).
8. Woodcock, C. L. F. & Frado, L. L. Y. Ultrastructure of chromatin subunits during unfolding, histone depletion, and reconstitution. *Cold Spring Harb. Symp. Quant. Biol.* **42,** 43–55 (1977).
9. Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science* **183,** 330–2 (1974).
10. Rattner, J. B. & Lin, C. C. Radial loops and helical coils coexist in metaphase chromosomes. *Cell* **42,** 291–296 (1985).
11. Belmont, A. S. & Bruce, K. Visualization of G1 chromosomes: A folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J. Cell Biol.* **127,** 287–302 (1994).
12. Richmond, R. K., Sargent, D. F., Richmond, T. J., Luger, K. & Mader, A. W. Crystal structure of the nucleosome resolution core particle at 2 . 8 A. *Nature* **389,** 251–260 (1997).
13. Pierce, B. in *Genetics: A Conceptual Approach* 833 (2004). doi:10.1017/CBO9781107415324.004
14. Woodcock, C. L. & Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2,** (2010).
15. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **8,** 35–46 (2007).

16. Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu. Rev. Genet.* **43,** 559–99 (2009).
17. Maeshima, K., Ide, S., Hibino, K. & Sasai, M. Liquid-like behavior of chromatin. *Curr. Opin. Genet. Dev.* **37,** 36–45 (2016).
18. Kamakaka, R. T. & Biggins, S. Histone variants: Deviants? *Genes Dev.* **19,** 295–310 (2005).
19. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589,** 2914–2922 (2015).
20. Hacques, M. F., Muller, S., De Murcia, G., Van Regenmortel, M. H. V & Marion, C. Use of an immobilized enzyme and specific antibodies to analyse the accessibility and role of histone tails in chromatin structure. *Biochem. Biophys. Res. Commun.* **168,** 637–643 (1990).
21. Walker, I. O. Differential dissociation of histone tails from core chromatin. *Biochemistry* **23,** 5622–5628 (1984).
22. Smith, R. M. & Rill, R. L. Mobile histone tails in nucleosomes. Assignments of mobile segments and investigations of their role in chromatin folding. *J. Biol. Chem.* **264,** 10574–10581 (1989).
23. Ausio, J., Dong, F. & van Holde, K. E. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone 'tails' in the stabilization of the nucleosome. *J. Mol. Biol.* **206,** 451–463 (1989).
24. Polach, K. J., Lowary, P. T. & Widom, J. Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes. *J Mol Biol* **298,** 211–223 (2000).
25. Vettese-Dadey, M., Walter, P., Chen, H., Juan, L. J. & Workman, J. L. Role of the histone amino termini in facilitated binding of a transcription factor, GAL4-AH, to nucleosome cores. *Mol Cell Biol* **14,** 970–981 (1994).
26. Allan, J., Harborne, N., Rau, D. C. & Gould, H. Participation of core histone 'tails' in the stabilization of the chromatin solenoid. *J. Cell Biol.* **93,** 285–297 (1982).
27. Iizuka, M. & Smith, M. M. Functional consequences of histone modifications. *Curr. Opin. Genet. Dev.* **13,** 154–160 (2003).
28. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **7,** 437–447 (2006).
29. Yudkovsky, N., Logie, C., Hahn, S. & Peterson, C. L. Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. *Genes Dev.* **13,** 2369–

2374 (1999).

30. Corona, D. F. V, Clapier, C. R., Becker, P. B. & Tamkun, J. W. Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep.* **3,** 242–247 (2002).
31. Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: Multiple complexes for many purposes. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1677,** 52–57 (2004).
32. Lusser, A. & Kadonaga, J. T. Chromatin remodeling by ATP-dependent molecular machines. *BioEssays* **25,** 1192–1200 (2003).
33. Winston, F. & Allis, C. D. The bromodomain: a chromatin-targeting module? *Nat. Struct. Biol.* **6,** 601–604 (1999).
34. Hassan, A. H. *et al.* Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111,** 369–379 (2002).
35. Martens, J. A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Current Opinion in Genetics and Development* **13,** 136–142 (2003).
36. Deuring, R. *et al.* The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol. Cell* **5,** 355–365 (2000).
37. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21,** 421–434 (2011).
38. Cheung, P. & Lau, P. Epigenetic Regulation by Histone Methylation and Histone Variants. *Mol. Endocrinol.* **19,** 563–573 (2005).
39. Ausi??, J. Histone variants - The structure behind the function. *Briefings Funct. Genomics Proteomics* **5,** 228–243 (2006).
40. Piña, B. & Suau, P. Core histone variants and ubiquitinated histones 2A and 2B of rat cerebral cortex neurons. *Biochem. Biophys. Res. Commun.* **133,** 505–510 (1985).
41. Brown, D. T., Alexander, B. T. & Sittman, D. B. Differential effect of H1 variant overexpression on cell cycle progression and gene expression. *Nucleic Acids Res.* **24,** 486–493 (1996).
42. Alami, R. *et al.* Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5920–5925 (2003).
43. Jackson, J. D. & Gorovsky, M. A. Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Res.* **28,** 3811–6 (2000).

44. Pehrson, J. R. & Fuji, R. N. Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res.* **26,** 2837–2842 (1998).
45. Greaves, I. K., Rangasamy, D., Devoy, M., Marshall Graves, J. A. & Tremethick, D. J. The X and Y chromosomes assemble into H2A.Z-containing [corrected] facultative heterochromatin [corrected] following meiosis. *Mol. Cell. Biol.* **26,** 5394–405 (2006).
46. González-Romero, R., Rivera-Casas, C., Ausió, J., Méndez, J. & Eirín-López, J. M. Birth-and-death long-term evolution promotes histone H2B variant diversification in the male germinal cell line. *Mol. Biol. Evol.* **27,** 1802–1812 (2010).
47. Zalensky, A. O. *et al.* Human testis/sperm-specific histone H2B (hTSH2B): Molecular cloning and characterization. *J. Biol. Chem.* **277,** 43474–43480 (2002).
48. Maze, I., Noh, K.-M., Soshnev, A. A. & Allis, C. D. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat. Rev. Genet.* **15,** 259–71 (2014).
49. Loyola, A. & Almouzni, G. Marking histone H3 variants: How, when and why? *Trends Biochem. Sci.* **32,** 425–433 (2007).
50. Wiedemann, S. M. *et al.* Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J. Cell Biol.* **190,** 777–791 (2010).
51. Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. & Maltais, L. J. The human and mouse replication-dependent histone genes. *Genomics* **80,** 487–498 (2002).
52. Marzluff, W. F. & Duronio, R. J. Histone mRNA expression: Multiple levels of cell cycle regulation and important developmental consequences. *Current Opinion in Cell Biology* **14,** 692–699 (2002).
53. Polo, S. E., Roche, D. & Almouzni, G. New Histone Incorporation Marks Sites of UV Repair in Human Cells. *Cell* **127,** 481–493 (2006).
54. Akhmanova, A. S. *et al.* Structure and expression of histone H3.3 genes in Drosophila melanogaster and Drosophila hydei. *Genome* **38,** 586–600 (1995).
55. Bush, K. M. *et al.* Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenetics Chromatin* **6,** 7 (2013).
56. Frank, D., Doenecke, D. & Albig, W. Differential expression of human replacement and cell cycle dependent H3 histone genes. *Gene* **312,** 135–143 (2003).
57. Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14075–80 (2010).

58. Brush, D., Dodgson, J. B., Choi, O. R., Stevens, P. W. & Engel, J. D. Replacement variant histone genes contain intervening sequences. *Mol. Cell. Biol.* **5,** 1307–17 (1985).

59. D'Andrea, R. J., Coles, L. S., Lesnikowski, C., Tabe, L. & Wells, J. R. Chromosomal organization of chicken histone genes: preferred associations and inverted duplications. *Mol. Cell. Biol.* **5,** 3108–15 (1985).

60. Zeitlin, S. G. *et al.* Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 15762–15767 (2009).

61. Goldberg, A. D. *et al.* Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140,** 678–691 (2010).

62. Hake, S. B. & Allis, C. D. Histone H3 variants and their potential role in indexing mammalian genomes: the 'H3 barcode hypothesis'. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6428–6435 (2006).

63. Chow, C.-M. *et al.* Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep.* **6,** 354–60 (2005).

64. Jang, C., Shibata, Y., Starmer, J., Yee, D. & Magnuson, T. Histone H3 . 3 maintains genome integrity during mammalian development. *Genes Dev.* **1,** 1377–1392 (2015).

65. Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D. & Banaszynski, L. A. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* **522,** 240–4 (2015).

66. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37,** 1090–1097 (2005).

67. Sakai, A., Schwartz, B. E., Goldstein, S. & Ahmad, K. Transcriptional and Developmental Functions of the H3.3 Histone Variant in Drosophila. *Curr. Biol.* **19,** 1816–1820 (2009).

68. Couldrey, C., Carlton, M. B., Nolan, P. M., Colledge, W. H. & Evans, M. J. A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum. Mol. Genet.* **8,** 2489–2495 (1999).

69. Inoue, A. & Zhang, Y. Nucleosome assembly is required for nuclear pore complex assembly in mouse zygotes. *Nat. Struct. Mol. Biol.* **21,** 609–16 (2014).

70. Lin, C.-J., Conti, M. & Ramalho-Santos, M. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. *Development* **140,** 3624–34 (2013).

71. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet* **44,** 251–253 (2012).
72. Khuong-Quang, D. A. *et al.* K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol.* **124,** 439–447 (2012).
73. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482,** 226–231 (2012).
74. Sturm, D. *et al.* Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* **22,** 425–437 (2012).
75. Voigt, P. & Reinberg, D. Putting a halt on PRC2 in pediatric glioblastoma. *Nat. Genet.* **45,** 587–589 (2013).
76. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128,** 693–705 (2007).
77. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. & Mol. Biol.* **14,** 1025–1040 (2007).
78. Musselman, C. A., Lalonde, M. E., Cote, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19,** 1218–1227 (2012).
79. ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **51,** 786–94 (1964).
80. Sun, Z.-W. & Allis, C. D. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418,** 104–108 (2002).
81. Wierda, R. J. *et al.* Global histone H3 lysine 27 triple methylation levels are reduced in vessels with advanced atherosclerotic plaques. *Life Sci.* **129,** 3–9 (2014).
82. Falkenberg, K. J. & Johnstone, R. W. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat. Rev. Drug Discov.* **13,** 673–91 (2014).
83. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7,** 1395–1402 (1988).
84. Davie, J. R. Covalent modifications of histones: Expression from chromatin templates. *Curr. Opin. Genet. Dev.* **8,** 173–178 (1998).

85. Turner, B. M. Histone acetylation and an epigenetic code. *BioEssays* **22,** 836–845 (2000).
86. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84,** 843–851 (1996).
87. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384,** 641–3 (1996).
88. Parthun, M. R., Widom, J. & Gottschling, D. E. The major cytoplasmic histone acetyltransferase in yeast: Links to chromatin replication and histone metabolism. *Cell* **87,** 85–94 (1996).
89. Gregoretti, I. V., Lee, Y. M. & Goodson, H. V. Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis. *J. Mol. Biol.* **338,** 17–31 (2004).
90. Costello, J. F. & Plass, C. Methylation matters. *J. Med. Genet.* **38,** 285–303 (2001).
91. Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc. Natl. Acad. Sci.* **96,** 14967–14972 (1999).
92. Martin, C. & Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6,** 838–849 (2005).
93. Del Rizzo, P. A. & Trievel, R. C. Molecular basis for substrate recognition by lysine methyltransferases and demethylases. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1839,** 1404–1415 (2014).
94. Gaydos, L. J., Wang, W. & Strome, S. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science (80-.).* **345,** 1515–1518 (2014).
95. Audergon, P. N. C. B. *et al.* Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. *Sci. (New York, NY)* **348,** 132–135 (2015).
96. Corpet, A. & Almouzni, G. A histone code for the DNA damage response in mammalian cells? *EMBO J.* **28,** 1828–1830 (2009).
97. Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nat. Struct. Mol. Biol.* **11,** 1037–43 (2004).
98. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–80 (2001).
99. Kharchenko, P. V *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471,** 480–5 (2011).

100. Gonzalo, S. & Blasco, M. A. Role of Rb family in the epigenetic definition of chromatin. *Cell Cycle* **4,** 752–755 (2005).
101. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15,** 703–708 (2014).
102. Tropberger, P. *et al.* Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* **152,** 859–872 (2013).
103. Di Cerbo, V. *et al.* Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. *Elife* **2014,** (2014).
104. Cocklin, R. R. & Wang, M. Identification of methylation and acetylation sites on mouse histone H3 using matrix-assisted laser desorption/ionization time-of-flight and nanoelectrospray ionization tandem mass spectrometry. *J. Protein Chem.* **22,** 327–334 (2003).
105. Freitas, M. A., Sklenar, A. R. & Parthun, M. R. Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J. Cell. Biochem.* **92,** 691–700 (2004).
106. Casey Research. Going Global. 1–6 (2013). doi:10.2753/CED1061-1932470102
107. Bintu, L. *et al.* Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151,** 738–749 (2012).
108. Yuan, J., Pu, M., Zhang, Z. & Lou, Z. Histone H3-K56 acetylation is important for genomic stability in mammals. *Cell Cycle* **8,** 1747–1753 (2009).
109. Xu, F., Zhang, K. & Grunstein, M. Acetylation in histone H3 globular domain regulates gene expression in yeast. *Cell* **121,** 375–385 (2005).
110. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436,** 294–298 (2005).
111. Hyland, E. M. *et al.* Insights into the Role of Histone H3 and Histone H4 Core Modifiable Residues in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **25,** 10060–10070 (2005).
112. Xu, F., Zhang, Q., Zhang, K., Xie, W. & Grunstein, M. Sir2 Deacetylates Histone H3 Lysine 56 to Regulate Telomeric Heterochromatin Structure in Yeast. *Mol. Cell* **27,** 890–900 (2007).
113. Iwasaki, W. *et al.* Comprehensive structural analysis of mutant nucleosomes containing lysine to glutamine (KQ) substitutions in the H3 and H4 histone-fold domains. *Biochemistry* **50,** 7822–7832 (2011).

114. Tropberger, P. & Schneider, R. Going global: Novel histone modifications in the globular domain of H3. *Epigenetics* **5,** 112–117 (2010).
115. Zhang, L., Eugeni, E. E., Parthun, M. R. & Freitas, M. A. Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma* **112,** 77–86 (2003).
116. Hall, M. A. *et al.* High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.* **16,** 124–129 (2009).
117. Manohar, M. *et al.* Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *J. Biol. Chem.* **284,** 23312–23321 (2009).
118. English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. A. & Tyler, J. K. Structural Basis for the Histone Chaperone Activity of Asf1. *Cell* **127,** 495–508 (2006).
119. Jackson, D. a *et al.* Biochemical Method for Inserting New Genetic Information into DNA of. *Nature* **69,** 2904–2909 (1972).
120. San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77,** 229–257 (2008).
121. Wu, W. *et al.* Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. *DNA Repair (Amst).* **7,** 329–338 (2008).
122. Maeder, M. L. & Gersbach, C. A. Genome-editing Technologies for Gene and Cell Therapy. *Mol. Ther.* **24,** 430–446 (2016).
123. Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Research* **40,** 5795–5818 (2012).
124. Zwaka, T. P. & Thomson, J. A. Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.* **21,** 319–21 (2003).
125. Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244,** 1288–92 (1989).
126. Ma, H., Kunes, S., Schatz, P. J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58,** 201–216 (1987).
127. Windling, P. & Berchtold, M. W. The chicken B cell lind DT40: a novel tool for gene disruption experiments. *Elsevier Sci.* **249,** 1–16 (2001).
128. Araki, K., Araki, M. & Yamamura, K. Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res.* **30,** e103 (2002).
129. Lyznik, L. A., Rao, K. V. & Hodges, T. K. FLP-mediated recombination of FRT sites

in the maize genome. *Nucleic Acids Res.* **24,** 3784–3789 (1996).

130. Hye, J. K., Lee, H. J., Kim, H., Cho, S. W. & Kim, J. S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.* **19,** 1279–1288 (2009).
131. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–308 (2013).
132. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
133. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–1389 (2013).
134. Maeder, M. L. & Gersbach, C. A. Genome Editing Technologies for Gene and Cell Therapy. *Mol. Ther.* **24,** 430–446 (2016).
135. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* **22,** 2253–2257 (2012).
136. Winding, P. & Berchtold, M. W. The chicken B cell line DT40: A novel tool for gene disruption experiments. *J. Immunol. Methods* **249,** 1–16 (2001).
137. Shiao, Y.-H. BMC Biotechnology. *BMC Biotechnol.* **9,** 22 (2009).
138. Frey, A., Listovsky, T., Guilbaud, G., Sarkies, P. & Sale, J. E. Histone H3.3 is required to maintain replication fork progression after UV damage. *Curr. Biol.* **24,** 2195–2201 (2014).
139. Molnar, J. *et al.* The genome of the chicken DT40 bursal lymphoma cell line. *G3 (Bethesda).* **4,** 2231–2240 (2014).
140. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).
141. Ran, F. A. F. A. *et al.* XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell* **154,** 1370–1379 (2013).
142. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32,** 670–676 (2014).
143. Schratt, G. *et al.* Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. *Mol. Cell. Biol.* **21,** 2933 (2001).
144. Yang, H. *et al.* XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell* **154,** 1370–1379 (2013).
145. Chatterjee, P. K. *et al.* Mutually exclusive recombination of wild-type and mutant loxP

sites in vivo facilitates transposon-mediated deletions from both ends of genomic DNA in PACs. *Nucleic Acids Res.* **32,** 5668–5676 (2004).

146. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24,** 132–141 (2014).
147. De La Cruz, X., Lois, S., Sánchez-Molina, S. & Martínez-Balbás, M. A. Do protein motifs read the histone code? *BioEssays* **27,** 164–175 (2005).
148. Beck, H. C. *et al.* Quantitative proteomic analysis of post-translational modifications of human histones. *Mol Cell Proteomics* **5,** 1314–1325 (2006).
149. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403,** 41–45 (2000).
150. Fischer, J. J. *et al.* Combinatorial effects of four histone modifications in transcription and differentiation. *Genomics* **91,** 41–51 (2008).
151. Schaffner, W. *et al.* Genes and spacers of cloned sea urchin histone DNA analyzed by sequencing. *Cell* **14,** 655–671 (1978).
152. Hereford, L., Fahrner, K., Woolford, J., Rosbash, M. & Kaback, D. B. Isolation of yeast histone genes H2A and H2B. *Cell* **18,** 1261–1271 (1979).
153. Ragunathan, K., Jih, G. & Moazed, D. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science (80-.).* science.1258699- (2014). doi:10.1126/science.1258699
154. Bi, X. Heterochromatin structure: Lessons from the budding yeast. *IUBMB Life* **66,** 657–666 (2014).
155. Takami, Y., Takeda, S. & Nakayama, T. An approximately half set of histone genes is enough for cell proliferation and a lack of several histone variants causes protein pattern changes in the DT40 chicken B cell line. *J. Mol. Biol.* **265,** 394–408 (1997).
156. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289,** 21312–21324 (2014).
157. Zhang, L. *et al.* Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* **10,** (2015).
158. Wang, L. *et al.* Large genomic fragment deletion and functional gene cassette knock-in via Cas9 protein mediated genome editing in one-cell rodent embryos. *Sci. Rep.* **5,** 17517 (2015).
159. Hendel, A., Fine, E. J., Bao, G. & Porteus, M. H. Quantifying on- and off-target genome editing. *Trends in Biotechnology* **33,** 132–140 (2015).

160. Gabriel, R., von Kalle, C. & Schmidt, M. Mapping the precision of genome editing. *Nat. Biotechnol.* **33,** 150–152 (2015).
161. Lange, U. C. *et al.* Dissecting the role of H3K64me3 in mouse pericentromeric heterochromatin. *Nat. Commun.* **4,** 2233 (2013).
162. Clayton, A. L., Hazzalin, C. A. & Mahadevan, L. C. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. *Molecular Cell* **23,** 289–296 (2006).
163. Tan, Y., Xue, Y., Song, C. & Grunstein, M. Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11493–8 (2013).
164. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40,** 897–903 (2008).

Nithyha
PARAMESWARAN
KALAIVANI

Understanding the mechanisms of histone modifications *in vivo*

Résumé

Les modifications post-traductionnelles (MPTs) d'histones sont apparues comme un acteur majeur de la régulation de l'expression des gènes. Cependant peu de choses sont connues sur le réel impact des MPTs sur la chromatine. Il a été suggéré que les MPTs d'histones (H2A, H2B, H3 et H4) ont le potentiel de moduler la fonction chromatinienne selon un « code histone » en recrutant des protéines spécifiques de liaison. L'objectif de mon projet est d'approfondir la fonction de l'acétylation du domaine globulaire de l'histone H3 et de comparer cette modification avec celles des queues N-terminale *in vivo* sur une lignée ES cellulaire. Pour étudier l'impact de ces MPTs *in vivo*, toutes les copies endogènes du gène H3 sauvage (WT) doivent être remplacées par des copies mutées. Ainsi la première étape de mon projet est d'établir une lignée cellulaire exprimant seulement H3 mutée (e.g reproduisant une acétylation permanente) afin d'étudier les effets des modifications sur le domaine globulaire de H3 sur (a) l'expression génique, (b) l'architecture chromatinienne mais également pour étudier (c) les effets réciproques et synergiques entre les différentes modifications du domaine globulaire et (d) comparer ces effets avec les modifications sur la queue N-terminale dans un système *in vivo*.

Résumé en anglais

Post-translational modifications (PTMs) of histones have emerged as key players in the regulation of gene expression. However, little is known to what extent PTMs can directly impact chromatin. It has been suggested that PTMs of core histones (H2A, H2B, H3 and H4) have the potential to govern chromatin function according to the so called ‘‘histone code’’ hypothesis by recruiting specific binding proteins. The goal of my project is to gain insight in the function acetylation within the globular domain of H3 and to compare these modifications with histone tail modifications, *in vivo by* using the CRISPR in mouse embryonic stem cells (ES). To study the impact of PTMs *in vivo*, all endogenous wild type (WT) H3 gene copies have to be replaced with mutant copies. Hence, the primary focus of my project is to establish cell lines that exclusively express mutated H3 (e.g. mimicking acetylation) in order to study effects of H3 globular domain modifications on (a) gene expression (b) chromatin architecture as well as to study (c) cross talks and synergisms between globular domain modifications and (d) compare the effects with tail modifications in an vivo system.