Heterochromatin dysregulation in human diseases

Matthias Hahn*, Silvia Dambacher* and Gunnar Schotta#

Munich Center for Integrated Protein Science (CiPS\textsuperscript{M}) and Adolf-Butenandt-Institute, Ludwig-Maximilians-University, Schillerstrasse 44, 80336 Munich, Germany

key words: epigenetics, heterochromatin, epigenetic therapy, FSHD, FRDA, cancer

running title: heterochromatin dysregulation in human diseases

word count: 4778 (w/o references)

abstract: 136 words

*These authors contributed equally to this work

#To whom correspondence should be addressed:

phone: +49-89-2180 75 422

fax: +49-89-2180 75 425

e-mail: gunnar.schotta@med.uni-muenchen.de
Abbreviations

AS - Angelman syndrome
DM1 - myotonic dystrophy type 1
FRDA - Friedreich’s ataxia
FSHD - facioscapulohumeral muscular dystrophy
FXN - frataxin
FXS - fragile X syndrome
HDAC - histone deacetylase
HMTase - histone methyltransferase
HP1 - heterochromatin protein 1
IAP - intracisternal type A particle
ICR - imprinting control region
MBD - methyl binding domain
ncRNA - non-coding RNA
PWS - Prader-Willi syndrome
Rb - Retinoblastoma protein
SAHF - senescence-associated heterochromatic foci
Abstract

Heterochromatin is a repressive chromatin state which is characterized by densely packed DNA and low transcriptional activity. Heterochromatin-induced gene silencing is important for mediating developmental transitions, and in addition, it has more global functions in ensuring chromosome segregation and genomic integrity. Here we discuss how altered heterochromatic states can impair normal gene expression patterns leading to the development of different diseases. Over the last years, therapeutic strategies which aim toward resetting the epigenetic state of dysregulated genes have been tested. However, due to the complexity of epigenetic gene regulation, the “first generation drugs” that function globally by inhibiting epigenetic machineries might also introduce severe side effects. Thus, detailed understanding of how repressive chromatin states are established and maintained at specific loci will be fundamental for the development of more selective epigenetic treatment strategies in the future.
Introduction

Abnormal gene expression patterns are often implicated in the development of different diseases, and thus, detailed insights into the underlying mechanisms of gene regulation will be fundamental in identifying novel interventional treatment strategies. Enormous progress over the last few years, particularly in the field of epigenetic regulation, has contributed to a better understanding of how gene activity is controlled.

In the nucleus DNA is packaged with histones and non-histone proteins in a dynamic structure called chromatin. The grade of chromatin compaction can influence DNA accessibility to the transcriptional machinery. Euchromatin is loosely packed and transcriptionally active, whereas heterochromatin represents a more densely packed chromatin state which is characterized by low transcriptional activity. Heterochromatin can be subdivided into two categories: facultative heterochromatin for silencing of developmental genes and the inactive X chromosome in female mammals, and constitutive heterochromatin which is formed at pericentromeric regions and telomeres.

Five major epigenetic mechanisms are capable of establishing and stabilizing open or closed chromatin structures, thereby regulating transcriptional activity (Figure 1):

1. Nucleosome remodelers can stimulate transcription by removing nucleosomes from promoter regions, allowing transcription factors to gain access to the underlying DNA (47). Additional functions of nucleosome remodelers include histone variant exchange and nucleosome sliding (93).

2. Mammalian cells express three histone H3 variants, H3.1, H3.2 and H3.3. Specific enrichment of H3.3 in transcriptionally active genes and regulatory regions, and H3.1 in repressed or inactive genetic elements suggest a regulatory function of these variants (60). Incorporation of histone variants is also important for many other chromatin-related processes. Faithful DNA damage repair, for example, depends on the presence of a histone H2A variant, H2A.X, which is phosphorylated upon damage detection allowing binding of the DNA damage repair machinery (11).

3. Increasing evidence suggests important functions for long non-coding RNAs (ncRNAs) in transcriptional regulation. Direct interactions of chromatin-modifying enzymes with non-coding RNAs are speculated to facilitate targeting to specific genomic loci. A well-studied example is the histone methyltransferase Ezh2 which was show to interact with different ncRNAs to induce X inactivation and repression of developmental genes (75, 101). Small
ncRNAs, e.g., siRNAs and promoter antisense RNAs can also trigger formation of repressive chromatin structures as will be discussed below.

4. Cytosine bases, preferably in the context of CpG dinucleotides can be methylated to 5-methylcytosine by DNA methyltransferases. DNA methylation is a repressive modification which is enriched at promoter regions of genes and non-coding DNA sequences. MBD (methyl binding domain) proteins and MeCP2 can bind methylated DNA stretches and in turn recruit co-repressor complexes to facilitate transcriptional silencing (78).

5. Histones can be post-translationally modified. The major modifications include phosphorylation, acetylation and methylation. Combinations of different histone modifications represent chromatin signals that are recognized by specific binding proteins which then mediate downstream effects. In the context of transcriptional regulation, histone lysine methylation has been particularly well characterized. This modification generates a high complexity of signals as each lysine position can be mono- (me1), di- (me2) or tri-methylated (me3) and distinct binding proteins for each methylation state can mediate different functions. Transcriptionally active, euchromatic domains are characterized by histone H3 lysine 4 tri-methylation (H3K4me3) at gene promoters and H3K36me3 across gene bodies (2). The two types of heterochromatin carry distinct modification patterns. Facultative heterochromatin is marked by high levels of H3K27me3 (86). In contrast, constitutive heterochromatin features the combinatorial mark H3K9me3 and H4K20me3 (52).

The enzymatic systems for many chromatin-modifying activities have been identified during the last years, however, still very little is known about their targeting to specific genomic regions. For gene silencing and heterochromatin formation across repetitive elements, the production of distinct types of non-coding RNA, antisense RNA and small interfering RNA (siRNA), has been implicated in the targeting mechanism.

In mammals, the promoter regions of many genes are transcribed at a low level, giving rise to promoter-associated RNAs (30). Antisense ncRNAs which are complementary to these promoter transcripts can mediate recruitment of repressive epigenetic modifications, such as H3K27me3 and H3K9me2. This form of transcriptional gene silencing involves the small RNA binding proteins Argonaute 1 and 2, suggesting that epigenetic machineries interact with components of the RNAi pathway (38, 42).

Targeting of repressive epigenetic modifications to repetitive elements at pericentric heterochromatin appears to involve small interfering RNAs. This mechanism is well understood in the budding yeast, *Schizosaccharomyces pombe*: Bidirectional transcription of
repetitive elements generates double stranded RNAs that are cleaved to siRNAs by the endonuclease _dicer_. These small RNAs are then loaded onto _argonaute 1_ which recognizes the DNA sequence from which the small RNA was generated. The histone methyltransferase (HMTase) Clr4 directly interacts with _argonaute 1_ and induces repressive H3K9 methylation (29). It is still controversial if this siRNA mechanism exists in mammals, however, there are parallels. An early step in establishing the combinatorial histone modifications at pericentric heterochromatin is H3K9 tri-methylation by Suv39h1 and Suv39h2, the mammalian Clr4 homologs. H3K9me3 is recognized by heterochromatin protein 1 (HP1) which, in turn, recruits other HMTases, Suv4-20h1 and Suv4-20h2, to mediate H4K20me3 (81, 82). Heterochromatin in mammals features another important epigenetic modification, DNA methylation. Interestingly, there is interdependence between DNA methylation and H3K9me3 (56). However, the mechanisms are not clear.

An important hallmark of constitutive heterochromatin is its ability to spread from a nucleation site into neighboring regions. Position-effect variegation in _Drosophila_ is a well-established model system which highlights this feature (80). In this organism, heterochromatin can eventually expand several hundred kilobases into euchromatic regions. Spreading is hypothesized to work through the following mechanism: H3K9me3 on one nucleosome recruits HP1 which can dimerize and interact with Suv39h that establish H3K9me3 on the neighboring nucleosome. The border between heterochromatic and euchromatic domains is likely to be dynamically regulated by a balance between the antagonizing activities of activating and repressive factors (19). Heterochromatin spreading also exists in mammals. However, it is not clear to what extent gene regulation is affected by its impairment.

Gene silencing through heterochromatin is an important mechanism to ensure establishment of cell-type specific gene expression patterns. Dysregulation of heterochromatin can result in severe developmental defects. In this review we will discuss human diseases which are connected to defective heterochromatin formation. We will distinguish four different categories: 1. gene overexpression due to reduced heterochromatin, 2. gene silencing through aberrant heterochromatin formation, 3. stable heterochromatin at regulatory elements, 4. global dysregulation of heterochromatin.

**Reduced heterochromatin formation in Facioscapulohumeral muscular dystrophy (FSHD)**
Genes are not only controlled by specific transcriptional activators and repressors. Transcriptional regulation is probably always adapted to the genomic environment of a particular gene. In the first part of this review we will highlight a disease which is caused by reduced formation of heterochromatin across a repetitive region, leading to aberrant activation of genes which are in close proximity.

Facioscapulohumeral muscular dystrophy (FSHD) is a frequent (1:20,000) dominant autosomal disease which is characterized by progressive, often asymmetric, weakness and wasting of facial, shoulder and upper arm muscles (45). FSHD is not caused by mutation of a specific disease gene. Instead, increasing evidence suggests a significant role for a complex epigenetic mechanism, resulting in perturbation of heterochromatic gene silencing in the subtelomeric domain of the long arm of chromosome 4 (24). This region (Figure 2A) contains several genes (e.g., FRG1, FRG2, ANTI) and pseudogenes (TUBB4Q, DUX4c) next to a large array of repetitive sequences (D4Z4). Each D4Z4 repeat is around 3.3 kb in length and harbors two classes of the GC-rich repeat sequences hhspm3 and LSau as well as a DUX4 pseudogene. Furthermore, D4Z4 is overall very GC-rich and has characteristics of CpG islands (35).

Reduced D4Z4 repeat number leads to overexpression of genes in FSHD patients

In healthy individuals up to 100 tandem copies of the D4Z4 element generate a heterochromatic domain that silences the nearby genes (Figure 2A). FSHD patients have less D4Z4 repeats (1-10 copies) (89). Interestingly, smaller D4Z4 arrays result in earlier disease onset and enhanced severity in patients (74, 85, 97). Genes proximal to D4Z4 are often inappropriately overexpressed in FSHD patients (24). As several of these genes might contribute to the FSHD phenotype, individual overexpression of FRG1, FRG2 and ANTI was conducted in the mouse model. Notably, overexpression of a single gene, FRG1, was sufficient to induce a muscular dystrophy phenotype which mimics human FSHD (23). Loss of function experiments in Xenopus laevis, using morpholinos against xFrg1, resulted in disrupted organization and inhibited growth of the myotome, and consistent with the mouse data, elevated xFrg1 levels lead to abnormal muscle formation (31). Not much is known about the molecular functions of FRG1. The protein localizes to nucleoli and Cajal bodies and is probably involved in pre-mRNA splicing (90, 91). Interestingly, FRG1 overexpression affects splicing of muscle-specific genes (23).

Recent studies have examined more FSHD patients and the data are very complex. Besides
the classic FSHD type with reduced D4Z4 repeats there is also a phenotypic type where repeat length is not altered. Finally, there are also FSHD patients where FRG1 is rather normally expressed and muscle lineage genes from unrelated loci are dysregulated (39, 46). These studies indicate that the FSHD phenotype can be caused by many different defects, however, in this review we will focus on the mechanism of FRG1 derepression.

Silencing of FRG1 is dependent on heterochromatic histone methylation marks

The close proximity to D4Z4 repeats and the dependence on the repeat number suggest that FRG1 expression is affected by epigenetic mechanisms that act in this environment. Not much is known about epigenetic modifications in the 4q35 region (Figure 2A). In early studies the Polycomb group protein YY1 has been detected at the D4Z4 repeats (24) and at the FRG1 promoter (5). Recruitment of YY1 correlates with establishment of a repressive histone modification, H3K27me3, at both regions. In mammals, H3K27 methylation is controlled by the histone methyltransferases EZH1 and EZH2. Although direct evidence is lacking these enzymes are likely to be responsible for this modification at 4q35. In FSHD patients overexpressing FRG1, H3K27me3 can still be detected at the FRG1 promoter, suggesting that this modification is not sufficient to induce silencing (5, 100).

Two other repressive epigenetic modifications, DNA methylation and H3K9me3, might play important roles in this context: DNA methylation has been detected at both D4Z4 repeats and at the FRG1 promoter, and compared to healthy controls, FSHD patients show significant hypomethylation of these regions (92). Interestingly, hypomethylation is not necessarily dependent on reduced D4Z4 repeats, as patients with normal repeat numbers can also display hypomethylation. In a recent publication Zeng et al. could show that H3K9me3 is highly enriched at D4Z4 repeats (100), suggesting that this domain has features of constitutive heterochromatin. In line with this model, the authors could demonstrate that, just as in pericentromeric heterochromatin, H3K9me3 is induced by SUV39H enzymes and can recruit HP1 proteins.

Regulation of FRG1 expression by D4Z4 heterochromatin: spreading vs. locus interactions

How can changes in the D4Z4 repeat number alter epigenetic modifications across the 4q35 region? We hypothesize that the D4Z4 repeat structure induces a heterochromatin environment, using similar mechanisms as found at pericentric heterochromatin. Bidirectional
transcription across D4Z4 might generate double stranded RNAs (14) which can be processed by the RNAi machinery. The resulting small interfering RNAs could recruit SUV39H enzymes to induce H3K9me3, the binding platform for HP1 proteins. Spreading of this heterochromatic structure may then establish repression of neighboring genes, such as FRG1 (Figure 2B). Reduced D4Z4 repeats might not be able to nucleate enough heterochromatin components to facilitate spreading to very distant loci. This idea is consistent with heterochromatin expansion in Drosophila where small heterochromatic regions can spread only around a few kilobases (36); however, special rearrangements which involve large blocks of pericentric heterochromatin can induce spreading across hundreds of kilobases (19). There is also an alternative explanation for D4Z4-dependent silencing of FRG1. Chromosome conformation capture (3C) techniques have revealed interactions between the heterochromatic D4Z4 repeats and the FRG1 promoter (5, 72). These chromatin interactions might induce repression of FRG1 without the necessity of a spreading mechanism (Figure 2B). If the D4Z4 array is small, interactions with FRG1 might be less frequent and could thus influence the transcriptional status of FRG1.

In summary, the data suggest that FSHD is connected to epigenetic dysregulation of D4Z4 repeats, however, at the current stage of analysis it is not possible to build a consistent model. In future studies it will be crucial to analyze a broader spectrum of epigenetic modifications across the entire 4q35 region. Furthermore it would be very helpful to put these data in context with expression analyses of not only FRG1 but also all other genes in this locus. It is very likely that FRG1 overexpression only correlates with a subset of FSHD patients and other genes might also contribute to disease development. Before specific disease treatment strategies can be developed, the phenotypic FSHD subtypes need to be much better defined on the molecular level.

**Triplet repeat induced heterochromatin formation in Friedreich’s Ataxia**

In the previous section we discussed a disease which is caused by reduced heterochromatin silencing. However, there are also human diseases which can be linked to ectopic heterochromatin formation and aberrant gene silencing. In this section we will describe how repetitive expansion of small nucleotide stretches, often triplets, in promoter, intron or exon regions can lead to heterochromatin formation and silencing of the associated gene, resulting in the development of a disease (59).
Friedreich’s ataxia (FRDA) is an autosomal recessive neurodegenerative disease with a frequency of 1:50,000. Patients suffer from progressive ataxia, muscle weakness, heart disease and eventually diabetes (18, 32). FRDA is caused by an expansion of GAA repeats in intron 1 of frataxin (FXN), resulting in substantially reduced transcription of this gene (4, 10). FXN is a mitochondrial protein which probably works as an iron chaperone. Its precise function is not entirely clear, however, loss of FXN leads to metabolic disturbances, e.g., increased oxidative stress, reduced iron-sulphur clusters and defects in energy metabolism (67).

GAA repeat expansions cause heterochromatin-induced silencing of frataxin

Healthy individuals carry between 6 and 36 GAA repeats in intron 1 of frataxin (Figure 3A), while FRDA patients show expansions ranging from 120-1700 repeats (10, 15, 18, 61). In FRDA patient cells, frataxin is silenced by GAA repeat-dependent formation of heterochromatin (16, 34). Interestingly, repeat length correlates with disease severity (21, 62). This suggests that increased repeats can more easily or more strongly induce silencing. A hallmark of many repeat expansion diseases is somatic instability. With increasing age of the patient, repeats can become longer, which might even accelerate disease progression. Several lines of evidence suggest that this repeat instability is caused by defects during DNA replication, DNA repair or recombination (59).

Transcriptional control of FXN in healthy individuals has not been analyzed in detail, yet. It is not known if binding sites for specific transcription factors exist (Figure 3A). Chromatin immunoprecipitation experiments revealed enrichment of the repressive histone modifications H3K9me3 and H3K27me3 at the FXN promoter and in a region upstream of the GAA repeats (34). Similar to pericentric heterochromatin, there is also enrichment for HP1 and DNA methylation in FRDA patient cells (16, 28).

Exon 1 of the frataxin gene contains a binding site for the chromatin insulator protein CTCF. Interestingly, increased DNA methylation or alterations in the chromatin structure in FRDA correlate with loss of CTCF binding (16). CTCF is speculated to play a role in the regulation of an antisense transcript, FAST-1 (Figure 3A), which may originate from an AluSP sequence more downstream in intron 1 (50), suggesting that antisense transcription is involved in disease development. In agreement with these data FAST-1 cannot be detected in healthy individuals, however, as only a small region of this transcript is known (Figure 3A, black...
arrow), it is formally possible that in healthy individuals a shorter form of FAST-1 is transcribed.

Although, production of the non-coding FAST-1 RNA at the frataxin locus might influence silencing, heterochromatin formation depends mainly on the GAA repeat expansion. This has been demonstrated in a transgenic mouse model where short GAA repeat expansions induce silencing of a linked marker gene (79). Repression of this transgene correlates with decreased promoter accessibility and can be enhanced by HP1 overexpression, typical features of a heterochromatic state. Importantly, silencing of the GAA repeat transgene did not depend on the genomic integration site, suggesting that the heterochromatic state is directly induced by the GAA repeat region.

How can repeat expansions induce heterochromatin?

Different models are discussed to explain how small repeat expansions can trigger heterochromatin formation (Figure 3B), however, there is no ultimate proof for any of them. An interesting feature of long GAA triplet repeats is the formation of non B-DNA structures and sticky DNA in vitro (94). If such structures, e.g. a DNA triplex, form at the FXN locus, normal transcription could be inhibited and RNA polymerase might stall or progress at a much slower rate. In this case cellular checkpoint mechanisms might trigger heterochromatin formation.

Another intriguing possibility to explain heterochromatin formation is antisense transcription across the FXN locus (Figure 3B). In S. pombe bidirectional transcription generates double stranded RNA which is processed by the RNAi machinery and can induce heterochromatin formation across the underlying genomic locus (29). Thus, production of the FAST-1 transcript could be crucial for the silencing mechanism in FRDA. Furthermore there is evidence for antisense transcription in other repeat expansion diseases (13, 53), indicating that this process may be of general importance. A major regulator for antisense transcript production might be CTCF. In FRDA patients, loss of CTCF binding correlates with accumulation of FAST-1 transcripts, and also in a related disease, myotonic dystrophy type 1 (DM1), CTCF limits antisense transcription (13). Further analyses are required to demonstrate whether CTCF directly functions in transcriptional control at these loci.
A third model for induction of heterochromatin at trinucleotide repeat loci was suggested recently (48) and might play a role in distinct repeat expansion diseases (Figure 3B). Transcripts containing expansions of CAG or CUG repeats, as found in fragile X syndrome (FXS) and DM1, can form stable hairpin loops that are substrates for dicer (48). Reduced transcript levels can be explained by dicer-mediated cleavage of the primary transcripts, leading to lower production of mature RNA. In analogy to siRNA triggered formation of heterochromatin, cleavage products of CNG hairpins might resemble siRNAs that potentially trigger heterochromatin formation using a similar mechanism.

Therapeutic strategies for repeat expansion diseases have focused on the development of treatments to minimize the adverse effects of missing an important gene. Mitochondrial dysfunction in FRDA, for example, benefits from an antioxidant therapy. Over the last years it has become clear that frataxin is silenced by local heterochromatin. This has led, more recently, to novel therapeutic strategies being developed which impair heterochromatin formation to allow higher expression of frataxin. This class of therapeutics, histone deacetylase inhibitors, is very promising as frataxin levels could be increased in FRDA patient cells (34) and in a mouse model (73). However, global dysregulation of histone deacetylation might cause severe side effects. Only a better molecular understanding of the epigenetic machinery that specifically targets repeat expansion loci will help to ultimately develop specific compounds for treatment of this class of diseases.

Epigenetic dysregulation of imprinting control regions in Prader-Willi Syndrome and Angelman Syndrome

Heterochromatin at regulatory elements can affect transcription of many genes in a large genomic domain. In this section we discuss diseases which are caused by genetic or epigenetic dysregulation at such regulatory regions.

Genomic imprinting is a mechanism to establish allele-specific differences in gene expression dependent on the parental origin (37). Imprinted genes are only expressed from either the maternal or the paternal chromosome. Regulatory elements, so-called imprinting control regions (ICRs), display parental-specific epigenetic profiles. An ICR on one allele can carry active, euchromatic modifications, whereas the ICR of the other allele shows features of heterochromatin. The different epigenetic profiles of maternal and paternal ICRs consequently affect transcription of multiple neighboring genes in cis by mechanisms which are not completely understood. Notably, epigenetic programming at ICRs is mitotically very stable.
Genetic or epigenetic alterations which lead to impairment of an ICR on one allele can result in dysregulation of many genes on the same chromosome, leading to severe developmental defects and disease.

Two imprinting control regions regulate multiple genes in the 15q11-13 region

In humans, two distinct diseases, Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS) are both caused by aberrant imprinting of genes in the 15q11-13 chromosomal region. Although the same chromosomal region is linked with both diseases, the phenotypic appearance of PWS and AS patients dramatically differs. PWS occurs in 1:15,000-30,000 individuals. Patients display infantile hypotonia, obesity, short stature, small hands and feet, growth hormone deficiency, mental retardation and behavioral problems (9). About 70% of PWS patients carry a deletion of the paternal 15q11-13 region, about 30% have a maternal disomy of chromosome 15. The complex phenotypic appearance of PWS is the result of dysregulation of several genes in 15q11-13 (55). AS patients, in contrast, are characterized by microcephaly, ataxia, mental retardation, jerky arm movements, absence of speech and sleep disorders (95). Similar to PWS about 70% of AS individuals carry 15q11-13 deletions which always have a maternal origin. Around 10% AS patients carry mutations in the ubiquitin protein ligase UBE3A, suggesting that dysfunction of this single gene can cause AS (44, 57).

Small percentages of AS patients have imprinting defects or parental disomy of chromosome 15.

Deletion mapping of AS and PWS patients revealed two imprinting centers in 15q11-13 which are only 35kb apart (8). The PWS-ICR controls numerous genes in this locus, e.g. MKRN3, MAGEL2, NDN, SNURF-SNRPN and a stretch of snoRNA genes, whereas the AS-ICR mainly regulates transcription of two very distant genes, UBE3A and ATPC10 (Figure 4). Genes under control of the PWS-ICR are expressed only from the paternal chromosome. In contrast, UBE3A and ATPC10 which are regulated by the AS-ICR show only maternal expression.

Local heterochromatin formation at the maternal PWS-ICR

How are these allele-specific expression patterns established? The AS-ICR seems central to induce silencing of the maternal PWS-ICR during development (40, 69), however, it might not be involved in direct transcriptional control of the 15q11-13 region which is rather...
regulated by the PWS-ICR. On the paternal allele, PWS-ICR carries active epigenetic modifications, i.e. H3K4 methylation (98) and absence of DNA methylation (20, 25). This stimulates transcription in cis across the entire 15q11-13 region. Notably, in this case UBE3A is transcribed in the antisense direction, which may inhibit the production of sense transcript, and is hypothesized to keep this gene silent on the paternal chromosome (76). On the maternal allele, however, the PWS-ICR features heterochromatic modifications, such as H3K9me3, H4K20me3 and DNA methylation (20, 25, 96, 98). Genes in the 15q11-13 region are consequently silenced by DNA methylation in their promoter regions. The UBE3A gene can be transcribed from the maternal allele as there is no more antisense transcription (Figure 4).

It is not very well known how these epigenetic modifications are established at imprinting control regions. Most data have been obtained from analyses of mouse models, however, it seems plausible that the mechanisms are also conserved in humans. H3K9me3 at the PWS-ICR depends on the histone methyltransferase G9a and, as G9a can only induce H3K9me1 and H3K9me2, another H3K9-specific HMTase is required for the full establishment of this modification (99). H4K20me3 at ICRs is induced by Suv4-20h enzymes using a similar pathway as found in pericentric heterochromatin (68). DNA methylation is established independently of the histone modifications by DNA methyltransferases. The targeting of these enzymes to ICRs is still unclear.

DNA methylation has an important dual role at imprinting centers. It can attract DNA methylation binding proteins, such as MeCP2, and it can also prevent binding of important transcriptional regulators, e.g. CTCF (37). It is still largely unclear how the chromatin state of a relatively small imprinting control region can induce silencing or transcriptional activation of multiple neighboring genes. One possibility is a spreading mechanism of the active or inactive state. Alternatively, formation of chromatin interactions which bring promoter regions of the neighboring genes in close proximity to the imprinting control region might help to transmit epigenetic states onto these target genes.

**Stability of allele-specific expression patterns in imprinting diseases**

How can defective imprinting on only one allele generate severe diseases as, e.g. Angelman Syndrome and Prader-Willi Syndrome? In AS patients, if the AS-ICR is maternally impaired repressive heterochromatin cannot be established on the PWS-ICR (Figure 4). This leads to aberrant activation of the PWS-ICR on the maternal chromosome which prevents proper sense transcription of UBE3A (76). In PWS patients, deletion or impairment of the paternal
PWS-ICR which would normally carry active modifications, prevents transcriptional activity across the entire 15q11-13 region on the paternal chromosome (55). Notably, all these genes are potentially functional on the maternal chromosome, however, they cannot be expressed due to repressive histone modifications at the maternal PWS-ICR (Figure 4). An intriguing possibility to treat this imprinting disease would be to reactivate these silenced alleles. Apparently, specific DNA elements in the PWS/AS ICR are necessary to maintain the allelic imprint by recruiting regulatory proteins which are involved in targeting epigenetic machineries (41). A detailed mechanistic knowledge of these mechanisms is required to develop specific treatment strategies. However, a balanced treatment with general inhibitors of DNA methylation (77) or histone methylation (49) might already be able to improve the situation.

**Heterochromatin dysfunction increases cancer susceptibility**

Heterochromatin has important functions in gene silencing and genomic stability, two processes which are often impaired in cancer cells. Dysregulation of heterochromatin is expected to increase cancer susceptibility by two mechanisms which we will discuss below: aberrant silencing of tumor-suppressor genes and increased genomic instability by reduced formation of pericentric heterochromatin.

**Heterochromatin-induced silencing of tumor-suppressor genes in cancer cells**

Numerous reports have provided evidence for aberrant heterochromatin formation at tumor-suppressor genes leading to their silencing (12, 83). Stochastic establishment of epigenetic silencing at such genes could induce mitotically heritable repression, potentially increasing cancer susceptibility. Recently, an intriguing possibility of how such aberrant silencing may be established has been proposed: During DNA damage repair, repressive modifications accumulate at the break site, and in a minor proportion of events, these modifications are not completely removed after successful repair (66).

A well-studied repressive epigenetic modification in the context of tumorigenesis is DNA methylation, which is often deregulated in cancer cells and connected to aberrant silencing of tumor-suppressor genes (3, 12). Other repressive modifications, e.g. histone lysine methylation, can also be deregulated in tumor cells. There is evidence that epigenetic silencing of tumor-suppressor genes can be mediated by distinct combinations of H3K27me3, H3K9me2 (58), H3K9me3 (54) and H4K20me3 (51). Reactivation of these aberrantly silenced genes might support a conventional tumor therapy. Inhibition of DNA
methyltransferases (5-aza-cytidine) and histone deacetylases (trichostatin A) can indeed be used to remove these repressive modifications and to consequently reactivate the silenced genes. A very general methylation inhibitor, 3-Deazaneplanocin A (DZnep) was shown to be effective in removing H3K27me3 (84). SUV39H1-mediated H3K9me3 in combination with DNA methylation also plays important roles in aberrant silencing of tumor suppressor genes (54). Intriguingly, removal of H3K9me3 by RNAi knockdown of SUV39H1 leads to reactivation of these genes even though DNA methylation was still present at the promoter regions (54). In summary, the data suggest that DNA methyltransferases and histone methyltransferases could be interesting therapeutic targets for cancer treatment.

Reduced formation of pericentric heterochromatin is connected with genomic instability

Heterochromatin has important functions in gene regulation, chromosome segregation and maintenance of genomic stability. Thus, global dysregulation of heterochromatin might also increase cancer susceptibility. In agreement with this hypothesis, a strongly reduced level of H4K20me3 was found to be a hallmark of many human cancers (22, 87, 88) and reduced levels of HP1α have been found in highly invasive and metastatic breast cancer cell lines (43). It is not clear if these findings represent cause or consequence of tumorigenesis, however, several lines of evidence suggest that impairment of heterochromatin can increase cancer susceptibility.

Cause-effect relationships can be best analyzed in the mouse. Pericentric heterochromatin is impaired in mice lacking the two Suv39h HMTases (Suv39h dn mice). Interestingly, these mice are highly susceptible to the development of B cell lymphomas (70). The actual cause of tumorigenesis is not clear, however, four heterochromatin-mediated mechanisms might account for this phenotype.

1.) Formation of a stable heterochromatic structure around centromere regions is important to ensure proper chromosome segregation. Suv39h deficient fibroblasts display increased aneuploidy, and Suv39h dn mice develop lymphomas which sometimes have a hypertetraploid karyotype, characterized by nonsegregated chromosomes that remain attached through their acrocentric regions (70). An intriguing hypothesis of how aneuploidy accelerates tumorigenesis is that it increases the loss of heterozygosity of tumor suppressor genes (1).

2.) Fully established pericentric heterochromatin might be important for the proper repair of DNA damage in satellite regions. Interestingly, DNA damage repair is delayed in Suv39h dn mutants, suggesting that impaired heterochromatin is less efficiently repaired (27). It is also possible that DNA damage repair is faulty in these mutants, which could generate mutations
that support tumorigenesis.

3.) Cells in which DNA damage cannot be fully repaired enter senescence, a stable cell cycle arrest that limits the proliferation of damaged cells and therefore functions as a natural barrier for cancer progression. Senescence is connected to establishment of senescence-associated heterochromatic foci (SAHF) which stably inactivate E2F target genes, thereby preventing cell cycle progression. SAHFs have features of pericentric heterochromatin with enrichment of H3K9me3, H4K20me3 and HP1 (65). Notably, establishment of the senescence program is impaired in absence of Suv39h1, leading to increased tumorigenesis (7).

4.) Silencing of mobile elements is another mechanism to ensure genome stability. Strong activation of the retrovirus-like intracisternal type A particle (IAP) elements, for example, results in catastrophic defects during meiosis (6). A lower level of activation might lead to integration of transposable elements in tumorigenesis-related genes and could activate oncogenes or mutate tumor suppressor genes. Heterochromatic histone modifications H3K9me3 and H4K20me3 play important roles in silencing LINE-1 and probably other transposable elements. These repressive modifications are mediated by Retinoblastoma (Rb) family members that can interact with histone methyltransferases (26). Consequently, dysregulation of Rb function removes H3K9me3 and H4K20me3 from LINE-1 elements resulting in their activation (63). It remains to be tested if genomic instability through activated mobile elements is a major cause of tumorigenesis.

In summary, these connections which relate to defective heterochromatin formation underline the complexity of the tumorigenesis process. Cancer cells always carry genetic and epigenetic mutations, and it seems plausible that epigenetic dysregulation sensitizes cells to accumulate oncogenic genetic mutations. Epigenetic treatment strategies with general histone or DNA methylation inhibitors, which aim for reexpression of tumor suppressor genes, can be very dangerous. Those drugs might kill tumor cells, however, by the same time they sensitize healthy cells to develop a new tumor.

Perspectives

Enormous progress over the last years has improved our understanding of how epigenetic mechanisms contribute to the development of human diseases. These efforts already led to the development of novel therapeutic strategies, which, however, may have side effects. In order to refine these strategies, we need to better understand the interplay between the epigenetic machineries, and we need to learn more about selective targeting mechanisms. Regarding heterochromatin mediated gene silencing, there is evidence from Drosophila that different...
chromosomal regions use distinct combinations of epigenetic machineries (71). We are only beginning to understand this interplay in mammals, where a higher complexity of enzymatic systems induces the different epigenetic modifications. Notably, in the context of human diseases, this high complexity suggests that disease critical pathways may exist, whose inhibition or stimulation would not globally impair epigenetic programming of healthy cells. In order to identify such pathways, a consistent and systematic analysis of epigenetic modifications and transcriptional regulation in healthy vs. patient material is required.

Identification of epigenetic machineries that work in the context of a specific disease-related locus is only the first step. The next question is how these machineries are targeted. Do these epigenetic enzymes interact with proteins that recognize a specific DNA sequence and could an inhibitor for this interaction prevent targeting? Recent data suggest that a different mechanism, binding of non-coding RNAs, assist in targeting. If we understand how the production of these RNAs is controlled we might get novel tools for regulating site-specific targeting of epigenetic modifications (33, 64). Reagents that mediate degradation of these ncRNAs, e.g., antisense RNAs, might allow specific reactivation of aberrantly silenced genes, such as frataxin in FRDA or tumor suppressor genes in cancer.

In this review we discuss how heterochromatin dysregulation can generate severe disease phenotypes. Much less is known about smaller changes in epigenetic programming that would not necessarily cause pathological effects. We hypothesize that during aging, stochastic errors in establishing or maintaining epigenetic programming occur during DNA damage repair (66), or can be triggered by environmental factors. Accumulation of this epigenetic damage may then contribute to the systemic aging phenotype. Treatment strategies that aim toward correcting epigenetic programming in human diseases may also be useful to treat the adverse effects of aging.

Acknowledgements

We apologize to all authors whose work we could not cite due to space limitations. We are grateful to Antonia Jack for critical reading of the manuscript. Research in the laboratory of G.S. is financed by the SFB-TR5 Chromatin, SFB 684, Center of Integrated Protein Science Munich and Friedrich Baur-Stiftung.
References

15. Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL, and...


32. Harding AE. Friedreich's ataxia: a clinical and genetic study of 90 families with an


24


98. Xin Z, Allis CD, and Wagstaff J. Parent-specific complementary patterns of histone


Figure legends

Figure 1: Major epigenetic mechanisms that regulate chromatin structure.

1) Nucleosomes can prevent binding of transcription factors (TF). One function of nucleosome remodeling complexes is to evict nucleosomes to make promoters accessible. 2) Distinct histone H3 variants associate with transcriptionally active or repressive chromatin domains. 3) Non-coding RNAs can recruit histone-modifying enzymes. For example, interaction of the H3K27-specific histone methyltransferase Ezh2 with a ncRNA is important to induce X inactivation. 4) DNA methylation is a repressive modification which is recognized by specific binding factors (MBD proteins) which recruit other co-repressor proteins, e.g., HDACs. 5) Histone modifications can recruit binding partners that induce active or repressive chromatin states. H3K4me3 correlates with transcriptional activity and is recognized by components of the TFIID complex. Heterochromatic chromatin structures feature the combinatorial mark H3K9me3 and H4K20me3.

Figure 2: Reduced heterochromatin formation in Facioscapulohumeral muscular dystrophy (FSHD).

(A) The subtelomeric region of chromosome arm 4q contains a large array of D4Z4 repeats which consists of repetitive elements (LSau and hhspm3) and the DUX4 pseudogene. In healthy individuals, 11-100 tandem repeats of D4Z4 induce a local heterochromatic environment which represses transcription of genes across a region of more than 200kb. DNA methylation, H3K9me3 and HP1, hallmarks of pericentric heterochromatin, can be detected at D4Z4 repeats. There is evidence that DNA specific binding of YY1 recruits EZH2 to induce H3K27me3, another repressive modification. In FSHD patient cells, the reduced D4Z4 array is not able to nucleate a heterochromatic structure. H3K27me3 is still detectable in the locus, however, loss of H3K9me3 and DNA methylation leads to overexpression of genes in the 4q region.

(B) Models for heterochromatin mediated gene silencing in the 4q region. 1) D4Z4 arrays can induce a local heterochromatic domain which spreads across the 4q locus, thereby inactivating the genes in this region. 2) An alternative explanation is the formation of chromatin interactions between the D4Z4 repeats and gene promoters, also leading to repression.

Figure 3: Triplet repeat induced heterochromatin formation in Friedreich’s ataxia (FRDA).

(A) In healthy individuals, intron 1 of the frataxin gene contains only a small number of GAA
repeats. The frataxin gene is transcribed, and CTCF binding in exon 1 may be an important regulatory element for this locus. In FRDA patient cells, expanded GAA repeats lead to formation of a heterochromatic domain, coinciding with loss of CTCF binding and silencing of frataxin expression. Interestingly, repression of frataxin correlates with upregulation of the FAST-1 antisense transcript which may originate from the AluSP sequence in intron 1.

(B) Models for initiation of heterochromatin through triplet repeat expansions. 1) Triplet repeats form unusual DNA structures which may result in stalling of the RNA polymerase. This could trigger a damage signal leading to inactivation of the locus. 2) Bidirectional transcription across the frataxin locus may produce double stranded RNA, triggering siRNA mediated heterochromatin formation. 3) Transcripts containing large CNG triplet repeats form hairpin structures which can be cleaved by dicer. It is currently unclear if the cleavage products function as siRNAs.

Figure 4: Prader-Willi-Syndrome and Angelman Syndrome are caused by epigenetic deregulation of imprinting control regions in chromosomal region 15q11-q13.

The 15q11-13 region (not to scale and not all genes are shown) contains two imprinting control regions, separated by around 35kb. On the maternal chromosome, establishment of H3K4me3 at the AS-ICR triggers heterochromatic modifications at the PWS-ICR leading to repression of most genes in the locus; only UBE3A is maternally expressed. On the paternal chromosome, active modifications at the PWS-ICR stimulate transcriptional activity of 15q11-13 genes. Antisense transcription across UBE3A prevents production of the normal sense transcript, thereby silencing the paternal UBE3A.

In Angelman Syndrome patients, inactivation or deletion of the maternal AS-ICR prevents establishment of repressive modification at PWS-ICR. Consequently, 15q11-13 genes are transcribed and production of the UBE3A antisense transcript leads to complete inactivation of this gene on both alleles. Deletion or inactivation of the paternal PWS-ICR in Prader-Willi Syndrome patients prevents transcriptional activation of 15q11-13 genes.
Epigenetic gene regulation

1. Nucleosome remodeling

2. Histone variants
   - ON: H3.3
   - OFF: H3.1

3. ncRNAs
   - ON: ncRNA
   - OFF: Ezh2

4. DNA methylation
   - ON: Me
   - OFF: DNA methylation

5. Histone modifications
   - ON: TFII D
   - OFF: HP1

- H3K4me3
- H3K9me3
- H4K20me3
- H3K27me3
- DNA methylation
A

D4Z4 repeat unit

healthy

centromere

FSHD

centromere

D4Z4 (11-100)

D4Z4 (1-10)

B

1. spreading of heterochromatin

2. chromatin interaction
A  healthy  

frataxin

(GAA)_{6-36}

CTCF

Exon1

AluSx/AluI

AluSP

Exon2

FRDA

HP1

Me

Me

Me

Me

Me

(GAA)_{65-1700}

Exon1

AluSx/AluI

AluSP

Exon2

FAST-1 antisense transcript

B

1. unusual DNA structures

RNAPol

sticky DNA

2. bidirectional transcription

RNAPol

sticky DNA

3. CNG repeat hairpins (DM1, FXS)

siRNAs

Dicer

siRNAs?
healthy

imprinting diseases

maternal AS-ICR defect: Angelman Syndrome

paternal PWS-ICR defect: Prader-Willi Syndrome