H3K9me is dispensable for *C. elegans* development but essential for genome integrity

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Peter Zeller

aus Freiburg in Deutschland

Basel 2017

Genehmigt von der Philiosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von

Prof. Dr. Susan M. Gasser	Fakultätsverantwortliche und Dissertationsleiterin
Prof. Dr. Marcel Tijsterman	Korreferent

Prof. Dr. M. Spiess

Dekan

Basel, den 21.02.2017

Overview

Rationale

Epigenetic mechanisms as key regulators of chromatin biology have been the focus of intensive research over the past 20 years. It has become clear that epigenetic pathways play a major role in the pathology of numerous diseases ranging from neurodegenerative repeat expansion diseases to certain cases of cancer (Egger et al. 2004).

One of the best-studied epigenetic marks is the methylation of histone 3 on lysine 9 (H3K9me). H3K9 methylation plays a major role in silencing parts of the genome. H3K9me domains encompass a broad variety of sequences, ranging from single genes to repetitive elements (Matsui et al. 2010). In addition to its function in transcriptional repression, H3K9me is implicated in chromosome segregation (Peters et al. 2001) and the maintenance of genome integrity (Peng and Karpen 2009). Lately the H3K9me mark received even more attention due to an identified role in mis-silencing of tumor suppressor genes during the development of cancer (Chen et al. 2010; Hua et al. 2014). Consequently, some of the first drugs to manipulate H3K9me are tested as cancer therapies (Yuan et al. 2013).

Studying H3K9me in complex multicellular organisms has so far proven to be difficult. Both mice and *Drosophila* have at least five histone methyl transferase (HMTs) enzymes that are essential and partially redundant, allowing only for the study of partial reductions in H3K9me. In the nematode *C. elegans* our lab identified the two methyl transferases, essential for all H3K9 methyltaion throughout development (Towbin et al. 2012; Zeller et al. 2016).

Although possessing similar epigenetic complexity and a similar chromosome structure as mammals, *Caenorhabditis elegans* (*C. elegans)* is able to survive, develop and propagate in the absence of H3K9me. Therefore, this model organism gives us the unique opportunity to study the potentially overlooked roles of H3K9me in detail and over multiple generations. Acquiring this knowledge is crucial before H3K9me targeting drugs make it into clinics, as it will allow us to better understand the far-reaching and long-term consequences in the context of an organism.

Chapter 1: Introduction

Nearly every cell in our body contains the same genetic information in the form of DNA. Yet, to fulfill its tissue-specific roles and to adapt to changing environments, each cell needs a specific protein composition. A crucial step therein is the selective expression of coding and non-coding regions of the genome. Transcription is induced by the binding of transcription factors to specific gene cis-regulatory sequences, which stabilizes in turn the binding and processivity of the core transcriptional machinery (Levine and Tjian 2003). Interestingly, the same transcription factors were shown to bind to and activate different target genes in different cell types (Arvey et al. 2012), and the identification of DNA sequence motifs at the cis-regulatory sequences, have been insufficient to predict the actual occupancy of transcription factor sites. Therefore, additional processes must exist that guide transcription.

In eukaryotic organisms, DNA is packaged by proteins in a structure called chromatin that influences its interaction with the transcription machinery by modifications to the DNA itself, as well as the protein composition and their posttranslational modifications of the chromatin complex. The accessibility of DNA sequence to the transcription machinery contributes to gene regulation.

The main DNA modification in mammalian cells is CpG methylation (meCpG, 5-methylcytosine). While the majority of the genome constitutively carries this modification, it is regulated at CpG islands as well as over the body of genes (Weber et al. 2005), where its presence is usually associated with transcriptional inhibition (Razin and Riggs 1980). meCpG does so by the recruitment of silencing proteins to gene bodies (Jones et al. 1998; Bird and Wolffe 1999), or by disabling transcription factor binding sites at CpG inlands (Watt and Molloy 1988; Bell and Felsenfeld 2000). Although *C. elegans* does not have CpG methylation, very rare adenine methylation (0.3% of bulk adenine) was recently discovered as an alternative DNA methylation form and first experiments suggest a silencing function (Greer et al. 2015).

In addition to the regulation by DNA modification, the proteins in chromatin can also be post-translationally modified. The central unit of chromatin is the nucleosome. A nucleosome consists of 147bp of negatively charged DNA wrapped around a positively charged histone octamer (2x H2A, H2B, H3 and H4). Posttranslational histone modifications are especially well-studied at the protruding tails of histones that are easily accessible by enzymes. Depending on the nature and position of the modification on the histone, it can either inhibit or promote transcription. The variety of known chemical modifications has been expanding in recent years. The best studied

modifications include methylation, acetylation, phosphorylation, ubiquitination, SUMOylation and glycosylation (Figure 1A). While acetylation only exists as a single acetyl group, 3 methylation groups can be added to lysine residues forming mono-, di- or tri-methylated forms. Different chromatin states are defined by the combination of these histone modifications that are thought to contain instructive information like a “histone code” (Strahl and Allis 2000).

Histone modifications: establishment, function, removal

Enzymes that modify histones post-translationally are referred to as “writers”. These enzymes possess characteristic protein domains depending on the modification they catalyze. In the case of histone methyl transferases, such as the histone H3K9 methyl transferases, the conserved catalytic domain is called a SET-domain, named for the first letters of Su(var)3-9 (H3K9me), Enhancer of zeste (H3K27me) and Trithorax (H3K4me), the three first characterized histone methyl transferases that share this domain (Tschiersch et al. 1994; Poulin et al. 2005). The SET domain contains the SAM binding site and the catalytic center (Yeates 2002).

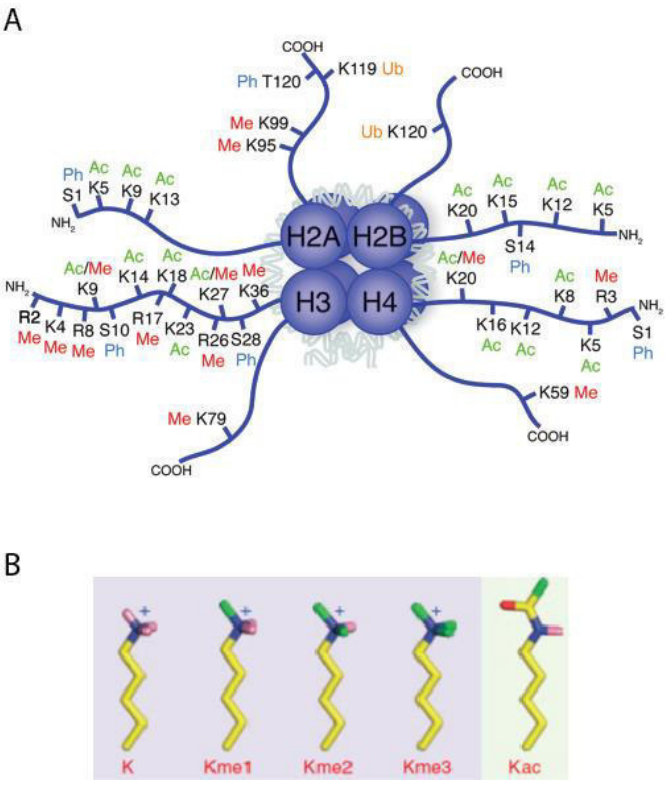


Figure 1: histone modifications

A) Schematic drawing of a nucleosome with the four canonical histones (H2A, H2B, H3 and H4). Amino acids [Lysine (K), Arginine (R), Serine (S) and Threonine (T)]. Posttranslational modifications [methylation (Me), acetylation (Ac), ubiquitination (Ub), and phosphorylation (Ph)] are highlighted on the N- and C-terminal tails of each histone (adapted from (Tollervey and Lunyak 2012)) B) Stick model of different states of Lysine methylation and acetylation. Yellow = carbon, blue = nitrogen, pink = polar hydrogen, red = oxygen, green methyl. The background and sign indicate the charge of the side chain: green = uncharged, blue = positive charge (adapted from (Taverna et al. 2007)).

Histone modifications can have, in principle, two modes of action. They can work directly on nucleosome-nucleosome or nucleosome-DNA interactions, by changing the charge of the highly basic histone tail (Figure 1B). This is especially well studied in case of histone acetylation. In

vitro experiments show that acetylation on H4K16, which is localized in the histone tail, prevents the formation of higher-order chromatin structure (Shogren-Knaak et al. 2006). H3K56 is positioned at the DNA entry/exit position on the histone core. Similar studies have shown that H3K56 acetylation does not affect higher order chromatin but instead leads to enhanced DNA unwrapping from the nucleosome (Neumann et al. 2009). These above effects are considered "cis" effects.

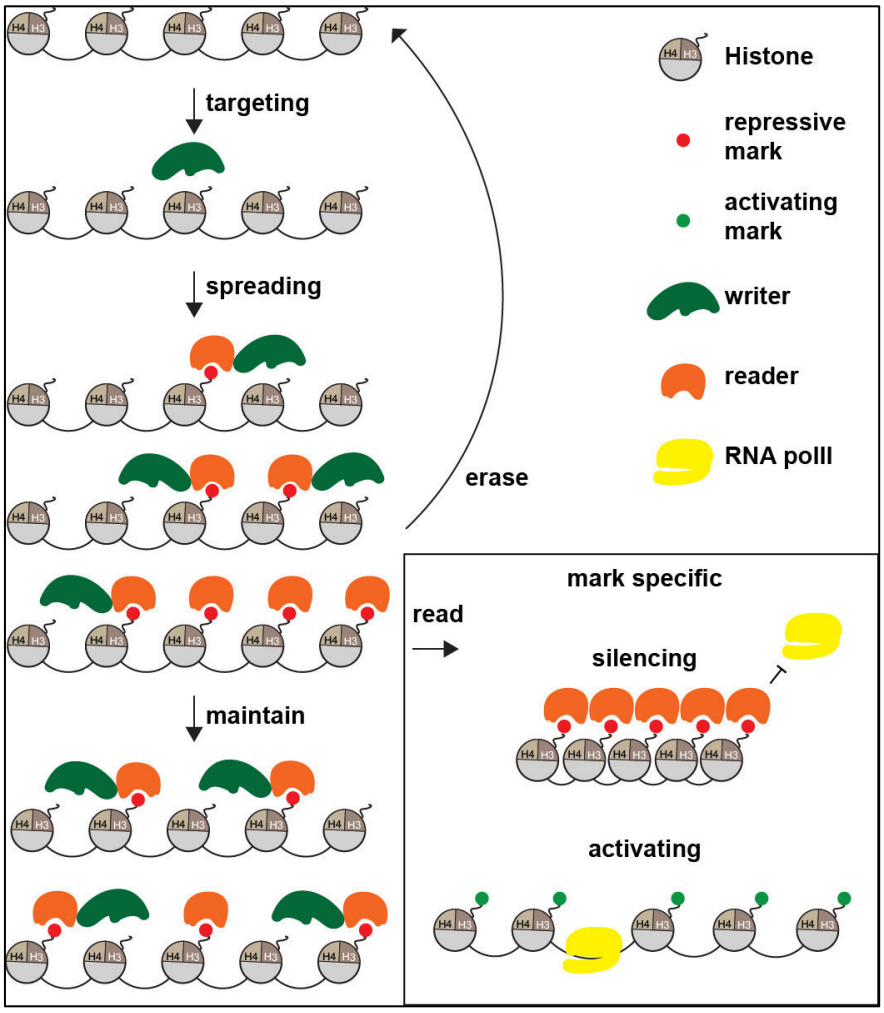


Figure 2: Histone mark dynamics

General and modification-specific steps in the life of a histone modification. For all histone modifications writer, erasers and readers exist enabling their establishment, removal as well as their effect on transcription. For many histone modifications, pathways for their spreading and maintenance over mitosis have been identified.

Alternatively, specific histone modifications can generate site-specific binding sites for proteins that selectively recognize modified lysine or arginine residues. These "readers" of the histone modifications, often act in "trans" by promoting or inhibiting the recruitment of additional regulators either of transcription or chromatin compaction. There is a growing list of domains that allow readers to bind their specific marks, but the most common are Bromo-, Chromo-, Tudor-, MBT-, PHD-, WD40 repeat-, 14-3-3 and BRCT domains (Taverna et al. 2007).

Histone modifications are removed by “eraser” proteins (Figure 2). In addition to a large family of histone deacetylases, there exist two classes of histone lysine demethylases. One is the amine oxidases with LSD1 (homologue of the *C. elegans* H3K4 demethylase LSD-1 and SPR-5) as founding member, that have been shown to be competent in demethylation of mono- and di-methylated lysine residues but are unable to work on tri-methylated lysine. The second are the Jumonji C (JmjC)- domain containing demethylases that are able to demethylate all stages of lysine methylation (Klose et al. 2006).

Two additional steps common to many histone modifications are their ability 1) to spread over a larger genomic region and 2) to be maintained through DNA replication. This is achieved by the combination of a “reader” and “writer” domain for the same histone modification in one protein or in a protein complex, which is referred to as a reader-writer complex. As an example, the association of HP1 with SUV39H1 maintains H3K9me through DNA replication as well as its spreading along the chromatin fiber (Nakayama et al. 2001). A similar process was identified in *C. elegans*, where the methyl transferase SET-25 was shown to be recruited to or stay with its own product H3K9me3 (Towbin et al. 2012).

Impact of histone modifications on transcriptional activity

The exact mode of action for many histone modifications is still under intensive research. For histone modifications, associated with active transcription, some specific mechanisms are identified. Besides its “cis” effects, histone acetylation can recruit bromo domain containing transcriptional regulators (Filippakopoulos et al. 2012). One example is the SWI/SNF chromatin remodeling complexes which are recruited by their bromo-domain containing ATPase components (Hassan et al. 2002). SWI/SNF in turn evicts histones close to the promoter, promoting RNA polymerase II loading (Qiu et al. 2016). Another active mark, H3K4me3, can be found on the promoter and gene bodies of transcribed genes. At promoter regions H3K4me3 can recruit TAF3,. a PHD domain-containing subunit of the basal transcription complex, TFIID, in order to directly promote the pre-initiation complex (Vermeulen et al. 2007). H3K9me3 can also recruit chromatin remodelers such as the PHD finger-containing NURF complex (Wysocka et al. 2006; Musselman et al. 2012). As their name: “remodeler” implies these protein complexes are involved in the remodeling of chromatin compaction and DNA histone interaction.

Determining the exact mechanisms through which repressive histone modifications influence transcription has proven to be very challenging. Early microscopy experiments in moss by Heitz in 1928, distinguished the interphase nucleus into two chromatin “states”. Heterochromatic

regions at the nuclear envelope and around the nucleolus stained strongly during the whole cell cycle, which was interpreted as a constant high level of compaction, while euchromatic regions, that only stain strongly during mitosis, “unfold” and therefore stain lightly during the rest of the cell cycle. This observation was the foundation for the model that the higher order packaging in heterochromatic regions could be refractory for the binding of the transcription machinery. Later on, heterochromatic regions were further separated into constitutive and facultative heterochromatin. Constitutive heterochromatin is condensed on both homologous chromosomes in every cell throughout development; examples are the pericentromeric heterochromatin, telomeres and parts of the rDNA. Facultative heterochromatin, initially referred only to the X-chromosome (Brown 1966), but later was shown to include tissue-specific genes that are differentially repressed depending on cell type. Further experiments showed that these cytologically defined chromosome regions correlate and depend on specific histone modifications, with constitutive and facultative heterochromatin bearing methyl-H3K9 (Noma et al. 2001; Schotta et al. 2002) or methyl-H3K27 (Bernstein et al. 2006; Kalantry et al. 2006), respectively. Experiments in which histone modification-specific antibodies were used to recover sheared chromatin fragments and the sequence of the associated DNA were identified by next generation sequencing (ChIPseq), were used to create high resolution maps of histone modifications on the complete genome. These experiments showed that interspersed repetitive elements all over the genome are also marked by H3K9me in every cell of the body and are part of the constitutive heterochromatin (Pimpinelli et al. 1995; Gerstein et al. 2010; Liu et al. 2011).

The different repetitive elements present in *C. elegans* and their known dependence on H3K9 methylation for silencing in other organisms is crucial for this work and is described in detail in Chapter 2. The distribution of facultative heterochromatin / H3K27me was further refined, showing a striking enrichment on developmentally regulated genes (Orlando 2003). Genome-wide comparisons between transcriptional activity and the presence of histone modifications shows a clear correlation of gene expression with euchromatin, while silent genes are often heterochromatic. The division between H3K9me3 and H3K27me3 is not as clear as originally thought, as H3K9me3 can also be found on developmentally regulated genes (Tachibana et al. 2002; Yamane et al. 2006; Zeller et al. 2016).

Two main approaches have been used to examine if chromatin composition can indeed affect the accessibility of the underlying DNA sequence. One method was to expose chromatin to DNase1. Early in vitro experiments treating isolated nuclei showed a preferential digestion of the actively transcribed albumin gene in liver tissue (Weintraub and Groudine 1976). Combining this approach with whole genome sequencing, it became clear that this method mainly identifies

histone free regions, which are especially found at active enhancer regions and around the transcription start site of active genes (Boyle et al. 2008). Outside of these two regions the sensitivity of this approach turned out to be rather limited, and did clearly distinguish eu- and heterochromatin. The second approach is based on the expression of an *E. coli* DNA methyl transferase (DAM) whose modification on DNA can be quantified as the degree of protection against the methyl-sensitive restriction enzyme *DpnI*. In a genome-wide study, Bell et al. was able to show a small reduction in methylation in H3K27me3 regions, but no difference between H3K9 methylated regions and euchromatic unexpressed loci could be observed (Bell et al. 2010). It has to be noted that heterochromatic chromatin is less readily solubilized in such experiments. To prevent a systematic error the authors therefore normalize to the untreated input material. The fact that both assays show relatively minor differences between eu- and heterochromatin raised the question whether the assays are appropriate for detecting differences between untranscribed, but potentially active and heterochromatic regions (Filion et al. 2010). One caveat might be that the chromatin compaction induced by heterochromatin does not interfere strongly with the temporal interaction of a single protein, but rather hinders the assembly of multiprotein complexes, such as the general transcription machinery. Moreover, these methods mainly measure nucleosome density, which might not be the level of chromatin organization affected in heterochromatin.

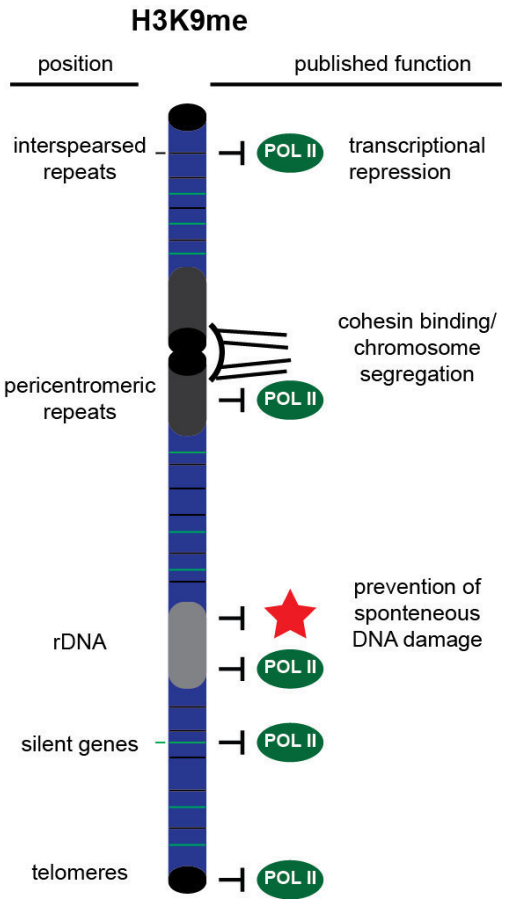
Heterochromatin and H3K9me

This section aims to give an overview of the current state of knowledge on H3K9me starting with its initial discovery as a major component in pericentric heterochromatin in *Schizosaccharomyces pombe* and *Drosophila melanogaster* and covering recent findings on its complex regulation in mammalian gene silencing. Functional roles of H3K9me published so far are summarized in Figure 3.

One of the simplest model organisms used to study H3K9me and heterochromatin is the fission yeast *S. pombe*. In this species, heterochromatic domains are relatively small and are found on the outer centromeric repeats, at the mating type locus and at telomeres. Work in *S. pombe* has been instrumental in the understanding of H3K9me establishment and maintenance at the centromeres, and in its role in chromosome segregation. *S. pombe* contains a single H3K9 HMT, Clr4, and two members of the heterochromatin protein 1 (HP1) family of H3K9me binding proteins (James and Elgin 1986), Swi6 and Chp2 (Ekwall et al. 1995; Ekwall et al. 1996). In the absence of Clr4 and Swi6 cohesin loading at the centromere is impaired, and it is speculated that the lack of sister chromatid cohesion does not allow the kinetochore to withstand the tension

exerted by spindle assembly, leading to chromosome segregation defects (Ekwall et al. 1995). This role has been observed across mono-centromeric organisms (Pidoux and Allshire 2005).

Figure 3: Published function of H3K9me



Summary of known positions of H3K9me over a chromosome as well as its functional relevance.

A second major discovery in *S. pombe* was the ability of the RNA interference (RNAi) machinery to guide H3K9 methylation. RNAi was first discovered in *C. elegans* (Tabara et al. 1998) and describes a biological process in which RNA molecules inhibit gene expression or translation, by neutralizing targeted mRNA molecules. It was later demonstrated in *S. pombe* that in the absence of main components of the RNAi machinery (Dicer or Ago1) heterochromatin formation and spreading around the centromere was perturbed (Volpe et al. 2002). Additionally, a reduction in pericentromere-specific small RNAs in the absence of H3K9me was observed (Hong et al. 2005), demonstrating an intriguing interplay between different silencing machineries, with the RNAi machinery enforcing silencing through and spreading of H3K9me. RNAi-induced H3K9me has now also been reported in *Arabidopsis* (Zilberman et al. 2003), *Drosophila* (Haynes et al. 2006) and in *C. elegans* (Ashe et al. 2012). Although *S. pombe* has been very useful to study pericentromeric heterochromatinization and its function, it does not have the chromatin complexity characteristic of higher eukaryotes.

The bulk of our current understanding of heterochromatin and H3K9 methylation comes from studies in *Drosophila*. This organism contains 3 major sites of heterochromatin: The pericentromeres, telomeres and the small fourth chromosome. *Drosophila* has two major strengths that facilitated early discoveries of the basic machinery involved in heterochromatin formation and spreading. One was the early discovery of chromosomal rearrangements that led to the variegated repression of the eye pigment gene *white* when it was juxtaposed to pericentromeric heterochromatin. *White* expression is important for the normally red eye color of

flies, which become white in the mutant. Flies carrying this rearrangement show a patchy eye color with some cell groups stably expressing and others stably silencing the gene. This stems from epigenetic silencing provoked by the stochastic spread of pericentromeric heterochromatin to the *white* gene (Muller 1930). Combining this rearrangement with mutagenic screens led to the identification of factors promoting *white* expression called enhancers of variegation E(var) and factors important for their silencing named suppressors of variegation Su(var) – proteins. This allowed the early genetic identification of Su(var)3-9 as a pericentromere-specific H3K9me transferase (Tschiersch et al. 1994; Rea et al. 2000; Schotta et al. 2002) and of HP1α as the first H3K9me2, me3 reader (James et al. 1989; Eissenberg et al. 1990), and it demonstrated their roles in gene silencing (Tschiersch et al. 1994). Importantly, loss of Su(var)3-9 led to a specific decrease of H3K9me2 and me3 only at the centromere, suggesting that the other heterochromatic regions in the genome and H3K9me1 at the pericentromere depend on other HMTs. Subsequent experiments identified dSETDB1 as an additional H3K9 HMT responsible for the silencing of the fourth chromosome (Seum et al. 2007).

A second major advantage of *Drosophila*, especially in the pre-sequencing era, was the formation of polytene chromosomes in terminally differentiated salivary gland cells (in particular at larval stage) (Silver and Elgin 1976), allowing for the direct observation of genomic rearrangements, heterochromatization (under-replicated chromosome parts) and in combination with immunofluorescence (IF), direct co-localization of histone modifications and chromatin components along the chromosome. With these tools at hand, it was possible to dissect the prerequisites for H3K9 methylation. Grigliatti et al. showed that euchromatic marks first are removed, and then Su(var)3-9 can act. The loss of histone deacetylase 1 (HDAC1) (H3K9 deacetylase (Mottus et al. 2000)), JIL1 (H3S10 dephosphatase (Ebert et al. 2004)), and LSD1 (H3K4 demethylase (Rudolph et al. 2007)) prevents heterochromatin spreading into bordering euchromatic regions.

Work in *Drosophila* gave the first indication that H3K9me might have a protective function against spontaneous DNA damage, in addition to preventing chromosome segregation defects. Karpen and colleagues could see an increase in γH2AX and Rad51 foci specifically in DAPI bright heterochromatic regions, as well as a distortion of the rDNA locus by DNA-FISH (Peng and Karpen 2009). With the development of sequencing technology, H3K9me was identified at dispersed repetitive elements in the genome that have been shown to be targeted for H3K9 methylation in an RNAi dependent process (Haynes et al. 2006). Furthermore, this advance highlighted a striking limitation in studying higher-order organisms such as the fly and mouse. These genomes contain long stretches of low complexity, repetitive regions creating mega-base

domains, which make it impossible to identify the genomic origin of repetitive sequence recovered from ChIP and RNA extraction experiments (Hoskins et al. 2007; de Koning et al. 2011).

With the increase in complexity of heterochromatic sequences and HMTs involved, work in mice nonetheless advanced our understanding of heterochromatin regulation particularly for its targeting and removal. So far at least eight partially redundant H3K9 histone methyltransferases (HMTs) have been documented in mice (SUV39h1, SUV39h2, G9a, SETDB1, SETDB2, PRDM2, PRDM3 and PRDM16). Some of these are essential by themselves; G9a, SETDB1, PRDM3 and PRDM16 have been shown to be essential for embryonic development (Dodge 2004, Tachibana 2002, Hoyt 1997, Aguilo 2011). This made it very difficult to study the specific function of the H3K9me deposited by them in development. Nonetheless, work in this organism was instrumental for understanding of H3K9me outside the centromere, allowing the characterization of different H3K9me targeting pathways, as well as their regulated targeting and removal by nuclear hormone receptor families (Metzger et al. 2005; Garcia-Bassets et al. 2007). Additionally, the roles observed in fission yeast (silencing and centromere function (Peters et al. 2001)) and *Drosophila* (heterochromatin machinery and transposon silencing (Matsui et al. 2010)) were shown to be conserved in mammals. Thus, despite early genetic identification of heterochromatic components in *Drosophila*, much of their enzymatic characterization was first performed in mice (Rea et al. 2000; Shi et al. 2005).

To understand the role of H3K9me at non-centromeric regions, it is necessary to study their behavior in the absence of H3K9me, something that was difficult in the systems discussed above. *C. elegans* has a huge advantage in this aspect: the worm genome lacks the massive clusters of pericentric repeats, that make mapping of sequencing reads in these chromosomal regions impossible (Treangen and Salzberg 2011). The *C. elegans* genome contains interspersed repetitive elements, with around 80% of them being mapable in chromatin immunoprecipitation (ChIP) experiments. Therefore, it is perfectly suited to study the function of H3K9me beyond its centromeric role.

H3K9me in *C.elegans*

Similar to mammals and flies, H3K9me was found in *C. elegans* on repetitive elements. *C. elegans* contains all classes of repetitive elements (e.g. tandem repeats, DNA transposons and RNA transposons). However, instead of being concentrated in mega-base domains at the

pericentric heterochromatin, repeat sequences in nematodes are organized in smaller clusters, and they are enriched on chromosome arms (Gerstein et al. 2010).

H3K9me writers

There exist multiple enzymes in the *C. elegans* genome for which a role in H3K9 methylation has been postulated: MET-2 (Poulin et al. 2005; Bessler et al. 2010; Towbin et al. 2012), SET-25 (Towbin et al. 2012), SET-26 (Hamilton et al. 2005; Ni et al. 2012; Greer et al. 2014), SET-9 (Ni et al. 2012) and MES-2 (Bessler et al. 2010).

In vivo data argue, however, that the main K9 methyl transferases are MET-2 and SET-25. In their absence, worms lack all detectable H3K9me by mass spectrometry at embryo and L1 larval stages (Towbin et al. 2012). This result was confirmed and extended to L2 stage larvae and dissected gonads from adults by IF (Zeller et al. 2016). MET-2 is the *C. elegans* homologue of SETDB1 and was first described as a potential transcriptional repressor by its involvement in vulva cell fate determination (Poulin et al. 2005). In *C. elegans*, the development of the vulva has proven to be a very powerful system to study cell differentiation, as it is dispensable for survival and defects are readily observed (Horvitz and Sternberg 1991). The study of occurrence of extra vulva revealed an intriguing regulatory network with two main branches of regulators: synMUV A and synMUV B. These two pathways inhibit vulva cell fate in a redundant manner, leading to the occurrence of multiple vulva, when mutants of both branches are combined. These experiments allowed the identification of multiple chromatin regulators including MET-2 (Poulin et al. 2005; Andersen and Horvitz 2007).

Using a fluorescent heterochromatin reporter, it was possible to quantify heterochromatin silencing, nuclear position, as well as chromatin composition using microscopy (Towbin et al. 2010). It could later be shown that heterochromatin depends on MET-2, a homologue of SETDB1, and SET-25, whose SET domain is homologous to EHMT1/G9a and Suv39h1/2, respectively. Both of these enzyme classes target histone H3K9 (Rea et al. 2000; Tachibana et al. 2002). SET-25, however, lacks both the Chromodomain found in Suv39h and the Ankyrin repeats present in G9a. SET-25 and MET-2 work together to silence and tether not only a heterochromatic reporter, but also endogenous chromatin, to the nuclear envelope (Towbin et al. 2012). Dissecting their individual contributions using null alleles, it could be shown that SET-25 is essential for all H3K9me3 in embryos and L1 larvae, and in the absence of MET-2, SET-25 is able to maintain around 20% of H3K9me1, me2 and me3 levels in embryos. MET-2 was shown to be the main H3K9 mono- and di-methyl transferase, being able to maintain wild-type levels in

embryos and L1 larvae in the absence of SET-25, as measured by mass spectroscopy. In embryos or L1 stage larvae of strains lacking both enzymes, no systematic differences in the methylation levels of H3K23, K27 or K36 were scored, arguing for strong specificity in the lysine targeted by these enzymes. Relevant lysine targets in nonhistone proteins have not been examined, however.

Using fusion proteins Towbin et al. suggested that MET-2 was enriched in the cytoplasm of mixed stage embryos, arguing for a role in modifying histone H3 prior to its assembly into nucleosomes (Towbin et al. 2012). The mammalian homologue ESET/SETDB1 shows both a nuclear and cytoplasmic localization (Tachibana et al. 2015). Recent work suggests the creation of the MET-2-GFP fusion protein altered the localization of the enzyme (M. Guidi and J. Padeken unpublished data). In contrast, SET-25 was found strongly enriched in the nucleus, where it binds its own enzymatic product (H3K9me3) in a SET-domain independent manner (Towbin et al. 2012). In other words, once it deposits H3K9me3, it recognizes this mark or else binds another reader, which binds it, and remains associated with heterochromatic domain. Its SET-domain is then available to extend methylation to nearby histone H3 tails. In principle, this provides a mechanism for the self-maintenance of the modification and potentially, for its spread along neighboring nucleosomes.

Two other SET domain proteins, SET-9 and SET-26, were identified in a longevity screen and were further characterized for their potential mode of action (Hamilton et al. 2005; Ni et al. 2012). Due to their high homology, it turned out difficult to separate these two enzymes by RNAi or antibody-based methods. Using an antibody that recognizes both SET-9 and SET-26 in both WT and *set-26* mutant animals, it was shown that SET-9 expression was restricted to the germline, while SET-26 detected in both somatic and germline cells. Quantifying levels of different histone modifications during aging, a decrease of H3K9me3 could be observed in young animals. It is important to note that this decrease results from the normalization of modified H3K9me over total H3 levels, which shows almost a two-fold increase. Considering that the increased complement of histone H3 is most likely not integrated into chromatin, but represents an increase in the free histone pool, H3K9me3 levels on chromatin are probably unchanged. Further experiments using modification-specific immunofluorescence or Western blots on chromatin fractions would be needed to clarify this question. SET-26 was found again in a screen for factors modulating the transgenerational sterility effect observed in mutants of the H3K4 demethylase SPR-5 (Greer et al. 2014). An *in vitro* assay using the purified SET-domain of SET-26, demonstrated a weak in vitro H3K9 methyl transferase activity on calf histones, which already carry a wide range of histone modifications. Other methylation marks on the calf

histones did not change upon incubation with the SET-26 SET domain, suggesting that it is not a nonspecific methyltransferase. However, in this assay, the H3K9me2/3 mark was the weakest pre-existing mark. Considering the lack of robust in vivo data for its relevance in H3K9 methylation as well as the lack of detectable H3K9me in the absence of MET-2 and SET-25, there are two possible explanations. Either SET-26 is active in only a small fraction of the worm's cells at a specific stage, or its function may be dependent on MET-2 and/or SET-25. It is clearly not redundant with these two enzymes, given that deletion of SET-25 eliminates all detectable H3K9me3 in somatic cells, and the deletion of both SET-25 and MET-2 reduces H3K9me below levels detectable by mass spectroscopy.

In early work it was also shown that loss of MES-2, the homologue to Enhancer of Zeste (*Drosophila*) and EZH1/2 in mammals, affected H3K9 methylation levels, specifically in the germline, and not in the soma (Bessler et al. 2010). This result needs additional confirmation, as the antibody used (ab8898), was shown to cross-react weakly with H3K27me3, the primary mark deposited by MES-2 (abcam.com). Still, considering the co-occurrence of H3K27me3 with H3K9me3 in the *C. elegans* genome (Ho et al. 2014), there may be a strong dependency of H3K9me3 on H3K27me3 at certain stages of development, either for its deposition or maintenance. Despite potential cross-dependency, SET-25 and MET-2 are the major H3K9 HMTs in *C. elegans*, and the absence of the two leads to a striking depletion of detectable H3K9me.

H3K9me readers

H3K9 methylation does not change the charge of the histone H3 tail and is thought to mainly work through the recruitment of its readers. *C. elegans* contains 5 published H3K9me binding proteins, 4 of which contain a chromodomain: HPL-1, HPL-2, CEC-3, CEC-4 and the MBT domain containing protein, LIN-61. The highly conserved HP1 family of H3K9me readers is defined by their three domains. A chromodomain connects via a flexible hinge domain to a chromo-shadow domain. The *Drosophila* protein HP1α is the prototype of a H3K9me binding protein. It is essential for centromeric satellite heterochromatin compaction and silencing (Lachner et al. 2001; Nakayama et al. 2001). HP1α is targeted to heterochromatin by its chromodomain that recognizes H3K9me (James and Elgin 1986), where it can dimerize through its chromo-shadow domain.

C. elegans contains 2 homologues of HP1: Heterochromatin Protein like (HPL) -1 and -2. Like HP1, both protein contain a chromo- and a chromoshadow- domain connected by a flexible

linker (Couteau et al. 2002). The exact binding specificity of HPL-1 is not known, but unpublished results from our lab show that it co-localizes with a heterochromatic reporter that is enriched for H3K9 and H3K27 methylation. For HPL-2, pull-down experiments using modified histone tail peptides, and immunoprecipitation experiments, showed strong binding to H3K9me2 and me3 as well as H3K27me3 (Studencka et al. 2012). Recently the genome wide distribution of HPL-2 was characterized by ChIP-seq, showing a slightly different result. In these experiments, HPL-2 distribution correlated best with H3K9me1 and me2 and to a smaller degree also H3K9me3 (Garrigues et al. 2015). No correlation with H3K27me3 was observed in this experiment. As neither of the two studies confirmed the specificity of the antibody used by checking KO animals, it is difficult to assess which result is correct. HPL-1 and HPL-2 fusion proteins show that both proteins are nuclear, with only partially overlapping distribution inside the nucleus (Schott et al. 2006). On an organismal scale, both proteins are expressed in all embryonic cells, while at the young adult stage some cells express primarily one or the other. However, studies were based on fusion proteins integrated into the genome as large arrays, whose expression does not always reflect the expression pattern of the endogenous genes (Schott et al. 2006).

Functionally HPL-2 was shown to be important for the transcriptional silencing of repetitive transgenes in the germline. Additionally it is involved in vulva development, where it showed genetic interactions of a synMUV-B gene (Couteau et al. 2002). At elevated temperature (25°C), loss of HPL-2 leads to partial sterility (Schott et al. 2006). HPL-1 seems to be at least partially redundant with HPL-2, as its deficiency was shown to enhance the sterility effect of HPL-2 in a double mutant, while having no obvious phenotype by itself (Couteau et al. 2002). At elevated temperature (25°C), the *hpl-1 hpl-2* deletion also leads to a developmental arrest. Unexpectedly *hpl-1* only genetically interacts as a synMUV with *hpl-2* and no other tested synMUV A or B mutant (Schott et al. 2006). Depletion of the H3K4 methyl transferase SET-2 was shown to rescue most somatic defects of *hpl-2* and *hpl-1 hpl-2* worms, showing the antagonistic roles of these activating and repressing machineries. Interestingly, germline phenotypes of *hpl-2* were not affected by SET-25 ablation (Simonet et al. 2007).

Other H3K9me readers were shown to have non-overlapping functions to HPL-2. The MBT domain containing protein LIN-61 was shown to bind H3K9me1/me2/me3 in vitro and localizes to H3K9me domains by ChIPseq experiments. Similar to HPL-1 and HPL-2, LIN-61 was identified as a synMUV-B protein (Harrison et al. 2007; Koester-Eiserfunke and Fischle 2011). CEC-3 contains a chromodomain, but no chromo-shadow domain. It was shown to bind H3K9me in vitro and it co-localizes with H3K9me2 and me3 in a MET-2 dependent manner in

vivo (Greer et al. 2014). Functionally CEC-3 was shown to play a non-redundant role with HPL-2 and LIN-61, in restricting somatic gene expression (Zheng et al. 2013). Unclear is its role in the transgenerational sterility observed in worms lacking the H3K4 demethylase SPR-5. These worms accumulate H3K4me and lose H3K9me3 over multiple generations, an effect that could be partially suppressed by the loss of CEC-3 (Greer et al. 2014). Other proteins with similar effects were shown to be H3K4 methyl transferases. A potential explanation would be a role for CEC-3 in K9me-mediated repression, but at the same time it may destabilize or block binding sites for H3K9 HMTs such as SET-25.

CEC-4 has a canonical HP1 alpha-like chromodomain, but with a C-terminal tail that lacks the chromoshadow motif, and which has no homology with any other chromodomain protein. ITC experiments show similar binding affinities of CEC-4 to all states of H3K9 methylation (me1, me2 and me3). CEC-4 localizes independently of histone methylation status to the nuclear envelope. The main role of CEC-4, is the tethering of H3K9 methylated regions to the nuclear envelope in embryos. It also contributes to perinuclear anchoring in muscle and the intestine cells of larvae. Its deletion results in a loss of peripheral localization of a significant number of heterochromatic domains in embryos. Despite this loss of anchoring, *cec-4* mutants only showed minimal alterations in transcription under standard laboratory conditions (Towbin et al. 2012; Gonzalez-Sandoval et al. 2015).

Some of these proteins also seem to play roles outside H3K9-methylated heterochromatin. Indeed, a recent study looking at the genomic distribution of HPL-2 in embryos shows, besides its preferential co-localization with H3K9me1 and me2, that in the complete absence of H3K9me (*met-2 set-25* embryos) a large fraction of HPL-2 is still bound to chromatin, potentially at expressed non-H3K9me positive genes (Garrigues et al. 2015). HPL-1 was shown to bind to innate immune response genes in the absence of an infection, together with the *C. elegans* homologue of the linker histone H1 (HIS-24). The authors speculate that the release of HPL-1 is a method for rapid activation of the immune genes in bulk (Studencka et al. 2012). LIN-61 may also have a second function, as it was shown to play a role in DNA double-strand break repair by homologous recombination (Johnson et al. 2013).

H3K9me erasers

C. elegans contains 13 JmjC-domain-containing proteins (Klose et al. 2006) of which 2 JMJD2a and ceKDM7a are involved in H3K9 demethylation. Depletion of JMJD2A by RNAi was shown by IF to lead to an increase of H3K9me3 on meiotic autosomes, as well as a local increase of

H3K36me3 on the X-chromosome (Whetstine et al. 2006). Phenotypically, the reduction of JMJD2A leads to replication stress, indicated by a replication checkpoint-dependent increase in germ-cell apoptosis, slower DNA replication (CY3-dUTP integration), as well as an accumulation of RAD-51 foci in the germline (Whetstine et al. 2006; Black et al. 2010). Interestingly, all phenotypes could be rescued by deletion of the H3K9me reader HPL-2 (Black et al. 2010). In *Drosophila* it was shown that HP1 modulates replication timing (Schwaiger et al. 2010). Whether H3K9 methylation directly influences replication timing in worms, or if perturbations of H3K9me levels leads to replication stress that arrests replication, is not yet clear.

CeKDM7a is a H3K9me2/H3K27me2 demethylase that co-localizes with H3K4me3 on a genome-wide scale, probably recruited by its additional PHD domain (Lin et al. 2010). Despite its strong H3K9me2/H3K27me2 demethylase activity in vitro, the reduction of both marks in mutants is quite limited, potentially due to its effect being restricted to H3K4me3 positive promoters. At these specific sites, a clear anti-correlation can be seen for ceKDM7a and its targets H3K9me2/H3K27me2. Deletion of ceKDM7a leads to the transcriptional silencing of its targets, most likely by allowing the local acquisition of heterochromatic marks, although this has not yet been shown.

Polycomb

Besides H3K9me the two best studied silencing pathways in *C. elegans* are the Polycomb Repressive Complex (PRC) and the RNAi machinery. The Polycomb group of transcriptional repressors consists of two complexes, PRC1 and PRC2. A main role of PRC2 that seems conserved from worms to mammals is the repression of Hox genes during development (Ross and Zarkower 2003; Deng et al. 2007). Polycomb mediated gene silencing involves the methylation of H3K27 by the PRC2 complex (Cao et al. 2002). H3K27me is then bound by the PRC1 silencing complex that mediates histone H2A ubiquitination (Schwartz and Pirrotta 2013). The *C. elegans* PRC2 complex contains the SET-domain containing H3K27 methyl transferase MES-2 (homologue of Ezh2), the worm specific MES-3 and the WD40 protein MES-6 (homologue of Eed) (Xu et al. 2001). The composition of the PRC1 complex is less well defined in *C. elegans*, but PRC1-related components were identified, including MIG-32 (homologue of Bmi-1) and SPAT-3 (homolgue of Ring1B), that have been shown to be essential for somatic H2A ubiquitination (Karakuzu et al. 2009). The somatic defects of *mig-32*, *spat-3* and *mes-2* mutants have been shown to be very similar, consistent with the idea that they work on a common pathway in *C. elegans*, as in other organisms (Karakuzu et al. 2009).

The *C. elegans* PRC2 complex was first described in a screen that aimed to identify genes causing sterility in the F1 generation after deletion, classifying them as maternal effect sterile (MES) mutations (Capowski et al. 1991). PRC2 was also shown to block artificial induction of somatic transcription programs in the germline. While total loss of the PRC2 complex led to sterility, depletion of PRC2 components by RNAi, allowed ectopic-expressed, cell-type inducing transcription factors in the germline to induce their cell-type specific transcription program (Patel et al. 2012). A similar role for PRC2 was also identified in somatic cells, where MES-2 is required to limit developmental plasticity of embryos (Yuzyuk et al. 2009).

One feature of H3K27me that is relatively well studied is its interaction with other histone modifications and silencing pathways. Early experiments on the MES proteins already showed a mutually exclusive distribution of the PRC2 components (MES-2, 3 and 6) with MES-4 (Fong et al. 2002), a *C. elegans* H3K36 methyltransferase. H3K36 methylation is generally associated with actively transcribed genes and is deposited through a co-transcriptional process. Later ChIP-seq experiments confirmed this mutually exclusive distribution for the histone modifications H3K27me3 and H3K36me3 as well (Ernst and Kellis 2010; Kharchenko et al. 2011; Liu et al. 2011). Early experiments already showed that PRC2 influences the distribution of MES-4. While MES-4 is normally excluded from the majority of the X-chromosome and only found at its left end, it was shown to spread over the complete chromosome in the absence of MES-2, 3 or 6 (Fong et al. 2002). This interaction seems to be bidirectional, as MES-4 seems to limit PRC2 distribution (Gaydos et al. 2012). In mutants lacking MES-4, H3K27me3 redistributes and spreads into formally H3K36me3 positive regions, while it is diluted at its endogenous positions. This is accompanied by transcriptional changes including the de-repression and repression of normally H3K27me3 and H3K36me3 positive regions respectively.

Besides antagonizing H3K36me3, PRC2 co-localizes with other repressive pathways. A substantial subpopulation of H3K9me3 was shown to co-localize with H3K27me3 (Ho et al. 2014). This is especially prominent in *C. elegans*, but can be observed to a lesser degree in mammals as well. The sequence specificity or functional relevance of the co-occurrence is not clear. Mao et al. could recently show that H3K27me3, like H3K9me3, can be targeted by the nuclear RNAi machinery (Mao et al. 2015). A possible hypothesis could therefore be that the nuclear RNAi pool defines the set of sequences that are supposed to be marked by both, H3K9me3 and H3K27me3.

RNA interference

RNA interference is a third very potent inhibitory pathway that is particularly well-studied in *C. elegans* (Lee et al. 2012). Initially RNA interference was identified in *C. elegans*, as a pathway that degrades RNA homologues to an introduced double-stranded RNA (dsRNA) (Fire et al. 1998). The introduced dsRNA was shown to be degraded by the RNase III like enzyme dicer producing primary exo-siRNA (Zamore et al. 2000). Mutants defective for the exo-RNAi pathway, e.g. dicer mutants, also led to defects in seam cell differentiation and fertility, suggesting endogenous functions of the RNAi machinery (Ketting et al. 2001).

With the emergence of deep sequencing, a diverse group of endogenous small RNAs (endo-siRNA) were discovered (Ruby et al. 2006). A large number of studies have focused on classifying these small RNAs according to their length and their most frequent 5`nucleotide (Ruby et al. 2006). There are three main groups of endogenous interfering RNAs: microRNAs, endogenous small interfering RNAs (endo-siRNAs), and Piwi-interacting RNAs (piRNAs). The life of small RNAs contains three phases: they are produced through transcription mainly by RNA polymerase II are processed into their functional form and loaded onto Argonaute proteins to fulfill their regulatory function. MicroRNAs and piRNAs are transcribed from specific loci in the genome, while endo-siRNAs are transcribed from spliced mRNA templates by the RNA-dependent RNA polymerase RRF-3 (Gent et al. 2009). It is still unclear how transcripts are selected for endo-siRNA biogenesis.

The generated pre-miRNA and pre-siRNAs are exported into the cytoplasm, where a process that depends on the helicase DCR-1 further processes them into primary mi- and 26G si-RNA, respectively (Grishok et al. 2001; Knight and Bass 2001). The biogenesis of primary 21U piRNAs is less well understood and does not depend on DCR-1 (Batista et al. 2008; Das et al. 2008). All three types of small RNAs bind to Argonaute effector proteins that, with certain exceptions, possess the ability to cut the complementary mRNA, which is referred to as slicer function (Yigit et al. 2006; Fischer et al. 2011). After their processing, primary siRNA as well as piRNA can additionally amplify their repressive potential by targeting mRNAs for the production of secondary 22G RNAs by a RNA-dependent RNA polymerases dependent process (Sijen et al. 2001; Tijsterman et al. 2002; Gent et al. 2010). These secondary 22G RNAs, were shown to be the main effectors of transcriptional repression (Bagijn et al. 2012; Lee et al. 2012). The generated 22G RNAs bind to a second set of Worm-specific Argonaute proteins called WAGOs (Yigit et al. 2006; Guang et al. 2008; Gu et al. 2009).

The endogenous mRNA targets of small RNAs were determined by containing complementary sequences as well as the accumulation of mRNAs in small RNA biogenesis mutants. MiRNAs are involved in developmental control and physiological processes and target mRNAs in the soma (Reinhart et al. 2000). SiRNAs and piRNA work rather like a surveillance system, silencing harmful RNAs. The siRNAs contain somatic and germline siRNA populations and target mainly mRNAs (Lee et al. 2006). The piRNAs, are specifically expressed in the germline and target mRNAs and transposable elements (Batista et al. 2008; Das et al. 2008); which would occur post-transcriptionally.

The RNAi pathway mediates silencing at two stages: transcriptional and post-transcriptional. At first it was thought that small RNAs would only elicit post-transcriptional regulation, because the introduction of purely intronic dsRNA elicited no silencing response (Fire et al. 1998). But for the majority of secondary siRNA bound to WAGOs, this is thought not to be the case, because WAGO proteins lack the catalytic triad essential for RNA slicing (Yigit et al. 2006). It is therefore thought that WAGO binding leads to the recruitment of other RNA degradation machineries such as the RDE-10/RDE-11 mediated de-adenylation machinery (Yang et al. 2012).

Recent work has now identified a role of RNAi in the control of transcription. A nuclear RNAi pathway links RNAi to the chromatin mediated silencing machinery by targeting H3K9 methylation and heterochromatization. The nuclear RNAi pathway was shown to act in both dsRNA induced silencing (Burkhart et al. 2011), as well as endogenous RNAi pathways such as the piRNA pathway (Luteijn et al. 2012). Both pathways lead to the production of secondary 22G siRNAs that the nuclear RNAi pathway was shown to depend on. If primary siRNAs do not share homology with any mRNA for the production of 22G secondary siRNA, H3K9 methylation is not observed (Gu et al. 2012).

Cytosolic 22G RNAs are bound by the Argonaute protein NRDE-3 (Guang et al. 2008) in the soma and by HRDE-1 (Buckley et al. 2012) in the germline. Both proteins are thought to shuttle the siRNA into the nucleus. This is nicely shown for NRDE-3, where nuclear translocation was shown to depend on siRNA binding. Moreover, the ablation of nuclear localization in an NLS mutant abrogated RNAi function (Guang et al. 2008). In the nucleus NRDE-3 binds to its specific pre-mRNA and recruits the other nuclear RNAi components (Burkhart et al. 2011): NRDE-1, NRDE-2 and NRDE-4, that are shared between the somatic and germline nuclear RNAi pathway. First NRDE-2 is recruited to the pre-mRNA. Its binding in the soma depends only on NRDE-3 (Guang et al. 2010). HRDE-2 binding further enables NRDE-1 recruitment, which in the

soma depends on NRDE-2 and NRDE-3 (Burkhart et al. 2011). After its initial interaction with the pre-mRNA NRDE-1 is thought to relocate to the transcribed DNA region (Burkhart et al. 2011).

Targeting of the nuclear RNAi pathway leads to heterochromatization of the target sequence including H3K9me3 deposition (Burkhart et al. 2011) and perhaps H3K27me3 deposition (Mao et al. 2015). Independent of the initial silencing induction (dsRNA or endo-siRNA) H3K9me3 at the target site was shown to depend on all NRDE components, including NRDE-3 (Burkhart et al. 2011; Gu et al. 2012) in the soma and HRDE-1 in the germline (Buckley et al. 2012). Again, H3K9me3 required the two H3K9 methyltransferases MET-2 and SET-25 (Mao et al. 2015) and transcriptional silencing required in addition the H3K9me reader HPL-2 (Ashe et al. 2012). Similar experiments showed that H3K27me3 depends on NRDE-2, NRDE-3 and on the H3K27 methyltransferase, MES-2 (Mao et al. 2015).The two pathways do not seem to be interdependent. H3K27me3 was not lost in mutants of *met-2* or *set-25*, nor was H3K9me3 lost on MES-2 (Mao et al. 2015). Looking at endogenous targets of HRDE-1 bound siRNAs by ChIP-qPCR Buckley and colleagues found an enrichment of H3K9me3 (Buckley et al. 2012) that at least partially requires the nuclear RNAi pathway and was reduced in a *hrde-1* mutant. In this context, it is interesting to mention that mutants of *nrde-1*, *nrde-2* and *nrde-3* show temperature-dependent sterility, much like the loss of cytoplasmic RNAi factors (Buckley et al. 2012) and H3K9me (Zeller et al. 2016).

The nuclear RNAi pathway also mediates the transgenerational inheritance of the silent state. In the case of the somatic RNAi machinery this only seems to last for one generation (from the treated P0 to the F1 generation) (Grishok et al. 2005), but if established in the germline (HRDE-1 dependent) it can be stably inherited (Grishok et al. 2000). The duration of inheritance seems to depend on the silenced target. Transgenic reporters have been shown to stay repressed for over 20 generations (Ashe et al. 2012), while endogenous sequences lose the silent state in most cases after 4 generations (Gu et al. 2012; Mao et al. 2015). The loss of silencing thereby correlates with a gradual reduction in locus specific 22G siRNA as well as H3K9- and H3K27me3 in the genome (Gu et al. 2012; Mao et al. 2015).

The establishment and the maintenance of the transcriptionally silent state have different protein requirements. The shared silencing agent seems to be the 22G siRNAs. In P0 animals 22G generation depends solely on the cytoplasmic machinery involved in their production. They were shown to depend on ERI-1 in the case of dsRNA and on PRG-1 in the case of piRNA (Luteijn et al. 2012). At this stage (P0 generation), the nuclear RNAi machinery is dispensable for 22G generation (Ashe et al. 2012; Gu et al. 2012) as was the H3K9me reader HPL-2 (Ashe et al.

2012). To study the maintenance of the silent state, the initiating primary RNAi has to be removed. In the case of dsRNA-induced silencing, the primary small RNAs can only be detected in P0 animals, while following generations only have 22G siRNAs (Ashe et al. 2012; Gu et al. 2012). For piRNA initiated inheritance, stably silenced reporters were crossed into a *prg-1* mutant background (Ashe et al. 2012; Luteijn et al. 2012). The first observation was that inheritance of the silent state and of 22G siRNAs became PRG-1 independent from the F1 generation on (Ashe et al. 2012; Luteijn et al. 2012). Interestingly, at this stage the nuclear RNAi and heterochromatin machinery become essential to propagate the silent state (namely NRDE-1, NRDE-2, and HRDE-1 as well as HPL-2 and SET-25) (Ashe et al. 2012; Buckley et al. 2012; Luteijn et al. 2012; Shirayama et al. 2012). Looking specifically at the maintenance of 22G siRNAs, Lutejin et al. could show that they also require all tested components of the nuclear RNAi pathway: *nrde-1* and *hrde-1* (Luteijn et al. 2012). *Set-25* and *hpl-2* were not tested in this context, but it might well be that heterochromatin is able to feed back into the small RNA pathway. At the same time *hrde-1* mutants were shown to lose H3K9me3 at endogenous HRDE-1 target sites, indicating that at least some sites need a constant reinforcement of the silent state to stay heterochromatized (Buckley et al. 2012).

The role of chromatin in genome integrity

DNA damage occurs in the context of chromatin. During DNA damage response heterochromatin seems to play an ambiguous role. On one side, chromatin is thought to be an obstacle for the DNA damage repair machinery that has to be removed. Indeed, following UV damage ubiquitination-mediated histone mobilization has been reported (Wang et al. 2006; Lan et al. 2012) and in the case of DNA double strand breaks histone ChIP experiments showed a local histone depletion around an induced DSB (van Attikum et al. 2004; van Attikum et al. 2007) and a recent study even showed a global histone loss at high levels of Zeocin- or γIR-induced DNA damage (Hauer et al. 2017). The local histone release was shown to depend on the active action of histone remodelers BRG1 and INO80 (van Attikum et al. 2004; van Attikum et al. 2007; Zhao et al. 2009; Jiang et al. 2010). At the same time multiple repressive factors, including HP1 (Luijsterburg et al. 2009), Polycomb (Hong et al. 2008) and HDAC1/2 (Miller et al. 2010), are recruited to sites of DNA damage. Animals lacking HP1 (Luijsterburg et al. 2009) or Polycomb components (Hong et al. 2008) are hypersensitive to genotoxic agents, suggesting a functional relevance of their recruitment. Besides a role for heterochromatin in repair factor recruitment, they are also implicated in the local transcriptional silencing around damage sites, that is meant

to prevent conflicts between the repair and the transcription machinery (Vissers et al. 2012; Ui et al. 2015).

In the absence of the heterochromatin components Su(var)3-9 and HP1α in *Drosophila* the Karpen lab could show an increase of spontaneous RAD-51 foci in DAPI dense regions of the nucleus, suggesting that heterochromatin is important to prevent spontaneous DNA double-strand breaks (Peng and Karpen 2009; Chiolo et al. 2011). Although replication of the genome is a highly regulated process that ensures the fidelity of DNA duplication, mistakes during replication are one of the main sources of mutations in the absence of mutagenic substances. DNA replication initiates at specific sites, defined as replication origins. The selection and activation of DNA replication origins occurs within the context of chromatin. One of the biggest impediments for the replication fork was shown to be transcription (Brewer 1988; French 1992; Liu and Alberts 1995). To avoid such collisions of the replication fork with the transcription machinery, the coordination of replication and transcription is essential. Interestingly, transcriptional active sites were shown to replicate early, while transcriptional silent sites were mostly replicated later during S-phase (Schübeler et al. 2002; Rivera et al. 2014).

The importance of a coordination between transcription and replication was elegantly shown in a study using a method that identifies DNA breaks by ligating sequencing adapters onto the open ends of un-fragmented DNA (Break-seq). The authors could show that fragile sites occurred at sites where replication and transcription both occurred. The shift of collision points by perturbing replication timing, or by inducing unscheduled transcription, resulted in a corresponding change in break position (Hoffman et al. 2015). At the longest human genes, such conflicts seem to be impossible to avoid. Their transcription takes more than one cell cycle, leading to the formation of fragile sites that break in a transcription-dependent manner (Helmrich et al. 2011).

One structure that correlates with conflicts between the replication and transcription machinery are R-loops. They are nucleic acid structures composed of an RNA:DNA hybrid, resulting from the displacement of the second DNA strand by the transcribed RNA. Studies in yeast and mammalian cell culture have shown the accumulation of RNA:DNA hybrids on highly transcribed genes (Wahba et al. 2016). In addition certain sequence features were identified to facilitate hybrid formation e.g. GC content, poly A tracks (Ginno et al. 2012; Wahba et al. 2016). Of particular note is the accumulation of RNA:DNA hybrids on telomeres and the Thy1 transposons (Chan et al. 2014), despite the relatively low expression level of those sequences. Recently numerous studies tried to elucidate the mechanisms that protect cells from RNA:DNA hybrid

accumulation and led to the identification of many more factors involved in transcriptional processivity (Santos-Pereira and Aguilera 2015).

The danger of RNA:DNA hybrids was first shown in cells depleted of certain RNA biogenesis and processing factors such as the THO complex in yeast (Huertas and Aguilera 2003) and *C. elegans* (Castellano-Pozo et al. 2012), or the serine/arginine-rich splicing factor 1 (SRSF1; previously known as ASF and SF2) in vertebrates (Li and Manley 2005). Additionally there is also some evidence for a replication independent role of R-loops in generating DNA damage. Nucleotide excision repair (NER) nucleases XPG and XPF were shown to be able to process R-loops into DSBs in some circumstances (Sollier et al. 2014).

Spatial organization of chromatin

Besides leading to the identification of euchromatin and heterochromatin, the initial cytological approaches also identified a second feature of genome organization – the spatial distribution of chromatin in the nucleus. Open euchromatin was found in the lumen of the nucleus and heterochromatin was enriched at the nuclear envelope and near the nucleolus or in chromocenters (Rae and Franke 1972). Further studies identified an intriguing system in which the nucleus is spatially organized on multiple levels.

In cell lines from many species, Cremer et al. found that chromosomes are not intermingled in an interphase nucleus, but occupy distinct territories (Cremer et al. 1988). Each chromosome territory can be composed of both active and inactive domains. Regions of these territories that tend to be heterochromatic are present at the nuclear periphery and are referred to as lamina associated domains (LADs) (Guelen et al. 2008). The more transcriptionally permissive regions of the territory extend into the nuclear interior and active genes here have been shown to extend into the interchromosomal space (Chambeyron and Bickmore 2004; Lieberman-Aiden et al. 2009) where it is suggested that transcription occurs, possibly in the form of foci or factories (Jackson et al. 1998).

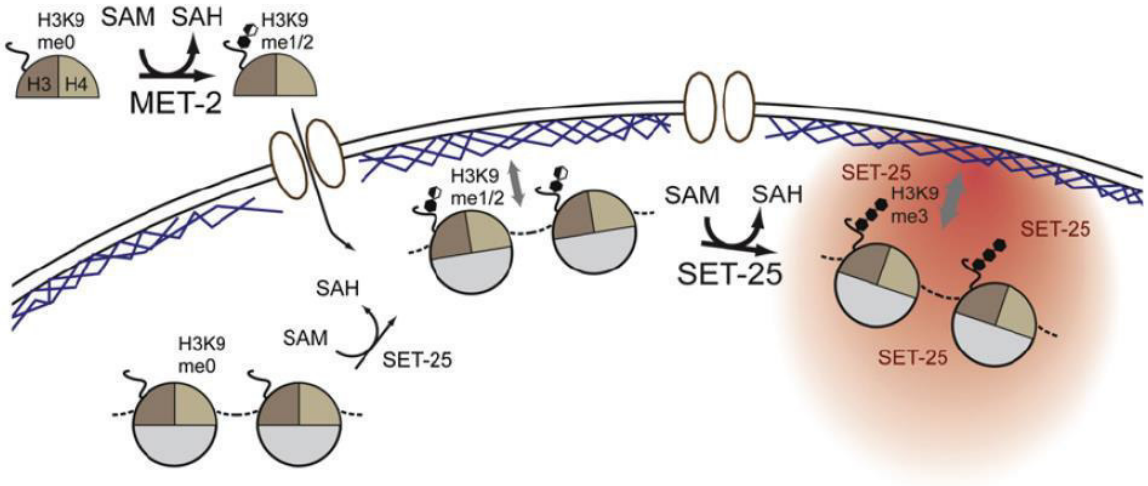
On a more detailed level along the chromosome, regions of preferential topological interaction have been identified. Analyzing the interaction frequencies between different genomic loci using a technique called chromatin conformation capture (Dekker et al. 2002), it was found in flies, mice and humans that commonly regulated genes reside in topologically associating domains (TADs) (Dixon et al. 2012; Sexton et al. 2012). Genomic loci inside a certain TAD can more frequently be crosslinked and are thought to more frequently interact than regions positioned in two different TADs. The biological equivalent of these interactions is still under research, but

they are suspected to represent enhancer-promoter interactions. Interestingly genes inside a TAD often share an activity, which is distinct from a neighboring TAD, suggesting that these structures could be involved in a pathway co-regulating whole chromatin domains. TAD-like self-interacting domains were recently also identified in *C. elegans* (Crane et al. 2015). In this organism TADs have stronger boundaries and are more numerous on the X chromosome, as compared to autosomes. TAD insulation in part depends on the dosage compensation complex (DCC) that is found at the strongest TAD boundaries. Transcriptional changes in DCC mutants are global and show no clear co correlation with the changes in TAD structure.

These levels of organization share a spatial separation of active and inactive chromosome regions. Separating opposing functions would indeed be an efficient way to optimize the usage of involved components by locally concentrating them. In mammals for example HP1 and HDAC3 were found to interact with the nuclear envelope associated proteins LBR and LAP2β, respectively (Mattout-Drubezki and Gruenbaum 2003; Somech et al. 2005), enriching these silencing factor at the nuclear envelope.

Work in *C. elegans* was instrumental to study the dynamic nature of nuclear organization over differentiation. Following several fluorescently tagged reporters containing a limited number of copies of a tissue specific promoter driven mCherry Meister et al. could identify two main forces influencing nuclear positioning of the reporter: heterochromatic tethering to the nuclear periphery and tissue-specific transcription induced release from the envelope (Meister et al. 2010). Much like endogenous repetitive sequences, multi copy transgenes in *C. elegans* acquire heterochromatic marks (namely H3K9me3 and H3K27me3) and were found tethered to the nuclear periphery in all cells, except those in which it was expressed. Using a similar fluorescent reporter bearing additionally a ubiquitously expressed promoter, a whole genome RNAi screen was performed that screened for loss of heterochromatin silencing and anchoring (Towbin et al. 2012). This study identified two H3K9 methyl transferases in *C. elegans* – MET-2 and SET-25 as important for both (Figure 4).

Figure 4: H3K9 methylation in *C. elegans*



Model from (Towbin et al. 2012) depicting the localization and function of the two sole H3K9 methyl transferases in *C. elegans*. It suggests that cytosolic MET-2 di-methylates free histones. Their integration leads to peripheral tethering of target regions. Consecutive function by nuclear SET-25 leads to tri-methylation, which co-localizes with its own product further tightening the silencing effect.

Mass spectrometric analysis of the double mutant argued that there were no other active H3K9 HMTs in *C. elegans*. It was suggested that MET-2 resides largely in the cytoplasm and is responsible for the bulk of H3K9me1 and me2, while SET-25 was nuclear and mediated all of H3K9me3. In the absence of MET-2, SET-25 can also partially take over mono and di-methylation of H3K9. Based on the heterochromatic reporter system, Towbin et al. suggested that both SET-25 and MET-2 are able to anchor heterochromatic sequences to the periphery, only leading to detachment in the double mutant. SET-25 was found strongly enriched in the nucleus, where it binds its own enzymatic product (H3K9me3) in a SET-domain independent manner (Towbin et al. 2012). The model suggested that the formation of SET-25 foci at the periphery might enhance heterochromatin silencing.

Scope

Astonishingly *C. elegans* is viable and fertile in the complete absence of H3K9me. This allowed us to use the *met-2 set-25* mutant to characterize the development of a multicellular organism in the absence of this central heterochromatic mark and ask: what is the main role of H3K9me in *C. elegans* and is there a functional difference between H3K9me3 and H3K9me1, me2 states.

Besides its viability in the absence of H3K9me *C. elegans* has 2 additional major advantages for this study:

1. While mammals possess a multitude of partially redundant partially essential H3K9 methyl transferases, *C. elegans* only has two H3K9me transferases with clear functional separation.
2. The size and distribution of repetitive elements in *C. elegans* allows one to analyze a majority of them through standard sequencing methods, yet the worm still possesses a chromatin complexity similar to mammals.

The major site of H3K9 methylation are repetitive elements (RE). Chapter 2 will therefore give an overview of the current state of knowledge on RE classes found in the genome, how they are controlled and what danger they pose. A special focus is put on the interplay of different epigenetic silencing mechanisms, with H3K9me at its center, that ensure repeat silencing at all stages of development.

Chapter 3 contains the majority of the experimental work characterizing the role of H3K9me in *C. elegans*. Starting with the observation of a striking increase of DNA damage checkpoint dependent apoptosis in *met-2 set-25* germlines, we identify increased mutagenesis specifically in the context of derepressed RE. We suggest that RNA:DNA hybrids that accumulate on derepressed RE drive these mutations by provoking conflicts with the DNA replication machinery.

In the second experimental part (Chapter 4) we closely analyze the distinct contribution of H3K9me2 and me3 to the roles identified in Chapter 3. We identify a partially interdependent role of MET-2 and SET-25, making SET-25 dependent on, and at the same time, redundant with MET-2 at the majority of its target loci. Similarly, the majority of H3K9me associated phenotypes are not, or only mildly observed in worms lacking H3K9me3, arguing for very similar abilities in silencing of H3K9me2 and me3. Besides quantitative differences in the number of transposable elements and genes depending on MET-2 or SET-25 for their transcriptional silencing, tandem repeats depended exclusively on MET-2. Interestingly, the identified silencing pathways occur in

different nuclear sub-compartments. In contrast to our previous model we found that SET-25 dependent silencing occurs all over the genome, while MET-2 repressed regions are enriched at the nuclear periphery. The tethering of endogenous heterochromatic sequences was also found to completely depend on the function of MET-2 and was independent of SET-25. We speculate that peripheral localization is involved in the MET-2 silencing function.

In the same study we also take an additional unbiased approach to identify previously overlooked roles of H3K9me by performing a whole genome synthetic lethality screen with the *met-2 set-25* double mutant. The hits were then further tested in each of the single mutants, *met-2* and *set-25,* showing exclusive genetic interaction with *met-2*. Finding many factors shown or suggested to be involved in RNA:DNA hybrid prevention and DNA damage repair, we conclude that the role of H3K9me in genome integrity described in Chapter 3 as one of its most important roles in *C. elegans*.

This thesis thereby provides evidence how repetitive elements derepression in the absence of H3K9me can lead to the occurrence of spontaneous DNA damage, putting a special emphasis on the danger of tandem repeat expression.

References

Andersen EC, Horvitz HR. 2007. Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. *Development* **134**: 2991-2999.

Arvey A, Agius P, Noble WS, Leslie C. 2012. Sequence and chromatin determinants of cell-type–specific transcription factor binding. *Genome research* **22**: 1723-1734.

Ashe A, Sapetschnig A, Weick E-M, Mitchell J, Bagijn MP, Cording AC, Doebley A-L, Goldstein LD, Lehrbach NJ, Le Pen J. 2012. piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell* **150**: 88-99.

Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, Lehrbach NJ, Simard MJ, Miska EA. 2012. Function, targets, and evolution of Caenorhabditis elegans piRNAs. *Science* **337**: 574-578.

Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S et al. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**: 67-78.

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**: 482-485.

Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB, Schubeler D. 2010. Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol* **17**: 894-900.

Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**: 2560-2569.

Bessler JB, Andersen EC, Villeneuve AM. 2010. Differential localization and independent acquisition of the H3K9me2 and H3K9me3 chromatin modifications in the Caenorhabditis elegans adult germ line. *PLoS Genet* **6**: e1000830.

Bird AP, Wolffe AP. 1999. Methylation-induced repression—belts, braces, and chromatin. *Cell* **99**: 451-454.

Black JC, Allen A, Van Rechem C, Forbes E, Longworth M, Tschop K, Rinehart C, Quiton J, Walsh R, Smallwood A et al. 2010. Conserved antagonism between JMJD2A/KDM4A and HP1gamma during cell cycle progression. *Mol Cell* **40**: 736-748.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311-322.

Brewer BJ. 1988. When Polymerases Collide: Replication and the Transcriptional.

Brown SW. 1966. Heterochromatin. *Science* **151**: 417-425.

Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, Fritz H, Kimble J, Fire A, Kennedy S. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**: 447-451.

Burkhart KB, Guang S, Buckley BA, Wong L, Bochner AF, Kennedy S. 2011. A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. *PLoS Genet* **7**: e1002249.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039-1043.

Capowski EE, Martin P, Garvin C, Strome S. 1991. Identification of grandchildless loci whose products are required for normal germ-line development in the nematode Caenorhabditis elegans. *Genetics* **129**: 1061-1072.

Castellano-Pozo M, García-Muse T, Aguilera A. 2012. R-loops cause replication impairment and genome instability during meiosis. *EMBO reports* **13**: 923-929.

Chambeyron S, Bickmore WA. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* **18**: 1119-1130.

Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. 2014. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* **10**: e1004288.

Chen M-W, Hua K-T, Kao H-J, Chi C-C, Wei L-H, Johansson G, Shiah S-G, Chen P-S, Jeng Y-M, Cheng T-Y. 2010. H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer research* **70**: 7830-7840.

Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV, Karpen GH. 2011. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**: 732-744.

Couteau F, Guerry F, Muller F, Palladino F. 2002. A heterochromatin protein 1 homologue in Caenorhabditis elegans acts in germline and vulval development. *EMBO Rep* **3**: 235-241.

Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**: 240-244.

Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L. 1988. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Hum Genet* **80**: 235-246.

Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R et al. 2008. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell* **31**: 79-90.

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**: e1002384.

Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* **295**: 1306-1311.

Deng H, Sun Y, Zhang Y, Luo X, Hou W, Yan L, Chen Y, Tian E, Han J, Zhang H. 2007. Transcription factor NFY globally represses the expression of the C. elegans Hox gene Abdominal-B homolog egl-5. *Developmental biology* **308**: 583-592.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380.

Ebert A, Schotta G, Lein S, Kubicek S, Krauss V, Jenuwein T, Reuter G. 2004. Su (var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. *Genes & development* **18**: 2973-2983.

Egger G, Liang G, Aparicio A, Jones PA. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**: 457-463.

Eissenberg JC, James TC, Foster-Hartnett DM, Hartnett T, Ngan V, Elgin S. 1990. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **87**: 9923-9927.

Ekwall K, Javerzat J-P, Lorentz A, Schmidt H. 1995. The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269**: 1429.

Ekwall K, Nimmo ER, Javerzat J-P, Borgstrom B, Egel R, Cranston G, Allshire R. 1996. Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *Journal of Cell Science* **109**: 2637-2648.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology* **28**: 817-825.

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ et al. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**: 212-224.

Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, Felletar I, Volkmer R, Muller S, Pawson T et al. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**: 214-231.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806-811.

Fischer SEJ, Montgomery TA, Zhang C, Fahlgren N, Breen PC, Hwang A, Sullivan CM, Carrington JC, Ruvkun G. 2011. The ERI-6/7 Helicase Acts at the First Stage of an siRNA Amplification Pathway That Targets Recent Gene Duplications. *PLoS Genet* **7**.

Fong Y, Bender L, Wang W, Strome S. 2002. Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans. *Science* **296**: 2235-2238.

French S. 1992. Consequences of replication fork movement through transcription units in vivo. *SCIENCE-NEW YORK THEN WASHINGTON-* **258**: 1362-1362.

Garcia-Bassets I, Kwon Y-S, Telese F, Prefontaine GG, Hutt KR, Cheng CS, Ju B-G, Ohgi KA, Wang J, Escoubet-Lozach L. 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **128**: 505-518.

Garrigues JM, Sidoli S, Garcia BA, Strome S. 2015. Defining heterochromatin in C. elegans through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. *Genome Res* **25**: 76-88.

Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S. 2012. Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. *Cell reports* **2**: 1169-1177.

Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, Tao L, Kennedy S, Fire AZ. 2010. Distinct phases of siRNA synthesis in an endogenous RNAi pathway in C. elegans soma. *Mol Cell* **37**: 679-689.

Gent JI, Schvarzstein M, Villeneuve AM, Gu SG, Jantsch V, Fire AZ, Baudrimont A. 2009. A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. *Genetics* **183**: 1297-1314.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**: 1775-1787.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell* **45**: 814-825.

Gonzalez-Sandoval A, Towbin BD, Kalck V, Cabianca DS, Gaidatzis D, Hauer MH, Geng L, Wang L, Yang T, Wang X et al. 2015. Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos. *Cell* **163**: 1333-1347.

Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB, Aristizabal-Corrales D, Chen S, Badeaux AI, Jin Q et al. 2014. A histone methylation network regulates transgenerational epigenetic memory in C. elegans. *Cell Rep* **7**: 113-126.

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizábal-Corrales D, Hsu C-H, Aravind L, He C, Shi Y. 2015. DNA methylation on N 6-adenine in C. elegans. *Cell* **161**: 868-878.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**: 23-34.

Grishok A, Sinskey JL, Sharp PA. 2005. Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes Dev* **19**: 683-696.

Grishok A, Tabara H, Mello CC. 2000. Genetic requirements for inheritance of RNAi in C. elegans. *Science* **287**: 2494-2497.

Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A. 2012. Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44**: 157-164.

Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ et al. 2009. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Mol Cell* **36**: 231-244.

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S. 2010. Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465**: 1097-1101.

Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, Lachowiec J, Kennedy S. 2008. An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**: 537-541.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948-951.

Hamilton B, Dong Y, Shindo M, Liu W, Odell I, Ruvkun G, Lee SS. 2005. A systematic RNAi screen for longevity genes in C. elegans. *Genes Dev* **19**: 1544-1555.

Harrison MM, Lu X, Horvitz HR. 2007. LIN-61, one of two Caenorhabditis elegans malignant-brain-tumor-repeat-containing proteins, acts with the DRM and NuRD-like protein complexes in vulval development but not in certain other biological processes. *Genetics* **176**: 255-271.

Hassan AH, Prochasson P, Neely KE, Galasinski SC, Chandy M, Carrozza MJ, Workman JL. 2002. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111**: 369-379.

Hauer MH, Seeber A, Singh V, Thierry R, Sack R, Amitai A, Kryzhanovska M, Eglinger J, Holcman D, Owen-Hughes T et al. 2017. Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates. *Nat Struct Mol Biol* doi:10.1038/nsmb.3347.

Haynes KA, Caudy AA, Collins L, Elgin SC. 2006. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Current biology* **16**: 2222-2227.

Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell* **44**: 966-977.

Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449-452.

Hoffman EA, McCulley A, Haarer B, Arnak R, Feng W. 2015. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome research* **25**: 402-412.

Hong E-JE, Villén J, Moazed D. 2005. A cullin E3 ubiquitin ligase complex associates with Rik1 and the Clr4 histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. *RNA biology* **2**: 106-111.

Hong Z, Jiang J, Lan L, Nakajima S, Kanno S, Koseki H, Yasui A. 2008. A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. *Nucleic Acids Res* **36**: 2939-2947.

Horvitz HR, Sternberg PW. 1991. Multiple intercellular signalling systems control the development of the Caenorhabditis elegans vulva. *Nature* **351**: 535-541.

Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F. 2007. Sequence finishing and mapping of Drosophila melanogaster heterochromatin. *Science* **316**: 1625-1628.

Hua K-T, Wang M-Y, Chen M-W, Wei L-H, Chen C-K, Ko C-H, Jeng Y-M, Sung P-L, Jan Y-H, Hsiao M. 2014. The H3K9 methyltransferase G9a is a marker of aggressive ovarian cancer that promotes peritoneal metastasis. *Molecular cancer* **13**: 1.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell* **12**: 711-721.

Jackson DA, Iborra FJ, Manders EM, Cook PR. 1998. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* **9**: 1523-1536.

James T, Eissenberg J, Craig C, Dietrich V, Hobson A, Elgin S. 1989. Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. *European journal of cell biology* **50**: 170-180.

James TC, Elgin SC. 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol Cell Biol* **6**: 3862-3872.

Jiang Y, Wang X, Bao S, Guo R, Johnson DG, Shen X, Li L. 2010. INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proc Natl Acad Sci U S A* **107**: 17274-17279.

Johnson NM, Lemmens BB, Tijsterman M. 2013. A role for the malignant brain tumour (MBT) domain protein LIN-61 in DNA double-strand break repair by homologous recombination. *PLoS Genet* **9**: e1003339.

Jones PL, Veenstra GCJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics* **19**: 187-191.

Kalantry S, Mills KC, Yee D, Otte AP, Panning B, Magnuson T. 2006. The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nat Cell Biol* **8**: 195-202.

Karakuzu O, Wang DP, Cameron S. 2009. MIG-32 and SPAT-3A are PRC1 homologs that control neuronal migration in Caenorhabditis elegans. *Development* **136**: 943-953.

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**: 2654-2659.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480-485.

Klose RJ, Kallin EM, Zhang Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics* **7**: 715-727.

Knight SW, Bass BL. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **293**: 2269-2271.

Koester-Eiserfunke N, Fischle W. 2011. H3K9me2/3 binding of the MBT domain protein LIN-61 is essential for Caenorhabditis elegans vulva development. *PLoS Genet* **7**: e1002017.

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

Lan L, Nakajima S, Kapetanaki MG, Hsieh CL, Fagerburg M, Thickman K, Rodriguez-Collazo P, Leuba SH, Levine AS, Rapic-Otrin V. 2012. Monoubiquitinated histone H2A destabilizes photolesion-containing nucleosomes with concomitant release of UV-damaged DNA-binding protein E3 ligase. *The Journal of biological chemistry* **287**: 12036-12049.

Lee HC, Gu W, Shirayama M, Youngman E, Conte D, Jr., Mello CC. 2012. C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell* **150**: 78-87.

Lee RC, Hammell CM, Ambros V. 2006. Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. *RNA (New York, NY)* **12**: 589-597.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147-151.

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365-378.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289-293.

Lin H, Wang Y, Wang Y, Tian F, Pu P, Yu Y, Mao H, Yang Y, Wang P, Hu L et al. 2010. Coordinated regulation of active and repressive histone methylations by a dual-specificity histone demethylase ceKDM7A from Caenorhabditis elegans. *Cell Res* **20**: 899-907.

Liu B, Alberts BM. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science*: 1131-1131.

Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung MS, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P et al. 2011. Broad chromosomal domains of histone modification patterns in C. elegans. *Genome Res* **21**: 227-236.

Luijsterburg MS, Dinant C, Lans H, Stap J, Wiernasz E, Lagerwerf S, Warmerdam DO, Lindh M, Brink MC, Dobrucki JW et al. 2009. Heterochromatin protein 1 is recruited to various types of DNA damage. *J Cell Biol* **185**: 577-586.

Luteijn MJ, van Bergeijk P, Kaaij LJ, Almeida MV, Roovers EF, Berezikov E, Ketting RF. 2012. Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. *Embo j* **31**: 3422-3430.

Mao H, Zhu C, Zong D, Weng C, Yang X, Huang H, Liu D, Feng X, Guang S. 2015. The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. *Current biology : CB* **25**: 2398-2403.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927-931.

Mattout-Drubezki A, Gruenbaum Y. 2003. Dynamic interactions of nuclear lamina proteins with chromatin and transcriptional machinery. *Cellular and Molecular Life Sciences CMLS* **60**: 2053-2063.

Meister P, Towbin BD, Pike BL, Ponti A, Gasser SM. 2010. The spatial dynamics of tissue-specific promoters during C. elegans development. *Genes Dev* **24**: 766-782.

Metzger E, Wissmann M, Yin N, Müller JM, Schneider R, Peters AH, Günther T, Buettner R, Schüle R. 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**: 436-439.

Miller KM, Tjeertes JV, Coates J, Legube G, Polo SE, Britton S, Jackson SP. 2010. Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat Struct Mol Biol* **17**: 1144-1151.

Mottus R, Sobel RE, Grigliatti TA. 2000. Mutational analysis of a histone deacetylase in Drosophila melanogaster: missense mutations suppress gene silencing associated with position effect variegation. *Genetics* **154**: 657-668.

Muller HJ. 1930. Types of visible variations induced by X-rays inDrosophila. *Journal of Genetics* **22**: 299-334.

Musselman CA, Lalonde ME, Cote J, Kutateladze TG. 2012. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**: 1218-1227.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**: 110-113.

Neumann H, Hancock SM, Buning R, Routh A, Chapman L, Somers J, Owen-Hughes T, van Noort J, Rhodes D, Chin JW. 2009. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol Cell* **36**: 153-163.

Ni Z, Ebata A, Alipanahiramandi E, Lee SS. 2012. Two SET domain containing genes link epigenetic changes and aging in Caenorhabditis elegans. *Aging cell* **11**: 315-325.

Noma K, Allis CD, Grewal SI. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**: 1150-1155.

Orlando V. 2003. Polycomb, epigenomes, and control of cell identity. *Cell* **112**: 599-606.

Patel T, Tursun B, Rahe Dylan P, Hobert O. 2012. Removal of Polycomb Repressive Complex 2 Makes C. elegans Germ Cells Susceptible to Direct Conversion into Specific Somatic Cell Types. *Cell Reports* **2**: 1178-1186.

Peng JC, Karpen GH. 2009. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet* **5**: e1000435.

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**: 323-337.

Pidoux AL, Allshire RC. 2005. The role of heterochromatin in centromere function. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **360**: 569-579.

Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C, Gatti M. 1995. Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. *Proc Natl Acad Sci U S A* **92**: 3804-3808.

Poulin G, Dong Y, Fraser AG, Hopper NA, Ahringer J. 2005. Chromatin regulation and sumoylation in the inhibition of Ras-induced vulval development in Caenorhabditis elegans. *Embo j* **24**: 2613-2623.

Qiu H, Chereji RV, Hu C, Cole HA, Rawal Y, Clark DJ, Hinnebusch AG. 2016. Genome-wide cooperation by HAT Gcn5, remodeler SWI/SNF, and chaperone Ydj1 in promoter nucleosome eviction and transcriptional activation. *Genome Res* **26**: 211-225.

Rae PM, Franke WW. 1972. The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. *Chromosoma* **39**: 443-456.

Razin A, Riggs AD. 1980. DNA methylation and gene function. *Science* **210**: 604-610.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593-599.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-906.

Rivera C, Gurard-Levin ZA, Almouzni G, Loyola A. 2014. Histone lysine methylation and chromatin replication. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**: 1433-1439.

Ross JM, Zarkower D. 2003. Polycomb group regulation of Hox gene expression in C. elegans. *Developmental cell* **4**: 891-901.

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127**: 1193-1207.

Rudolph T, Yonezawa M, Lein S, Heidrich K, Kubicek S, Schäfer C, Phalke S, Walther M, Schmidt A, Jenuwein T. 2007. Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU (VAR) 3-3. *Molecular cell* **26**: 103-115.

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**: 583-597.

Schott S, Coustham V, Simonet T, Bedet C, Palladino F. 2006. Unique and redundant functions of C. elegans HP1 proteins in post-embryonic development. *Developmental biology* **298**: 176-187.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. 2002. Central role of Drosophila SU (VAR) 3–9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO journal* **21**: 1121-1131.

Schübeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M. 2002. Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nature genetics* **32**: 438-442.

Schwaiger M, Kohler H, Oakeley EJ, Stadler MB, Schubeler D. 2010. Heterochromatin protein 1 (HP1) modulates replication timing of the Drosophila genome. *Genome Res* **20**: 771-780.
Schwartz YB, Pirrotta V. 2013. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet* **14**: 853-864.
Seum C, Reo E, Peng H, Rauscher III FJ, Spierer P, Bontron S. 2007. Drosophila SETDB1 is required for chromosome 4 silencing. *PLoS Genet* **3**: e76.
Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**: 458-472.
Shi Y-J, Matson C, Lan F, Iwase S, Baba T, Shi Y. 2005. Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell* **19**: 857-864.
Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, Conte D, Jr., Mello CC. 2012. piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. *Cell* **150**: 65-77.
Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, Plasterk RH, Fire A. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**: 465-476.
Silver LM, Elgin S. 1976. A method for determination of the in situ distribution of chromosomal proteins. *Proceedings of the National Academy of Sciences* **73**: 423-427.
Simonet T, Dulermo R, Schott S, Palladino F. 2007. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in C. elegans development. *Developmental biology* **312**: 367-383.
Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular cell* **56**: 777-785.
Somech R, Shaklai S, Geller O, Amariglio N, Simon AJ, Rechavi G, Gal-Yam EN. 2005. The nuclear-envelope protein and transcriptional repressor LAP2beta interacts with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *J Cell Sci* **118**: 4017-4025.
Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41-45.
Studencka M, Konzer A, Moneron G, Wenzel D, Opitz L, Salinas-Riester G, Bedet C, Kruger M, Hell SW, Wisniewski JR et al. 2012. Novel roles of Caenorhabditis elegans heterochromatin protein HP1 and linker histone in the regulation of innate immune gene expression. *Mol Cell Biol* **32**: 251-265.
Tabara H, Grishok A, Mello CC. 1998. RNAi in C. elegans: soaking in the genome sequence. *Science* **282**: 430-431.
Tachibana K, Gotoh E, Kawamata N, Ishimoto K, Uchihara Y, Iwanari H, Sugiyama A, Kawamura T, Mochizuki Y, Tanaka T et al. 2015. Analysis of the subcellular localization of the human histone methyltransferase SETDB1. *Biochemical and biophysical research communications* **465**: 725-731.
Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H. 2002. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes & development* **16**: 1779-1791.
Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**: 1025-1040.
Tijsterman M, Ketting RF, Okihara KL, Sijen T, Plasterk RH. 2002. RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. *Science* **295**: 694-697.
Tollervey JR, Lunyak VV. 2012. Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics* **7**: 823-840.
Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**: 934-947.

Towbin BD, Meister P, Pike BL, Gasser SM. 2010. Repetitive transgenes in C. elegans accumulate heterochromatic marks and are sequestered at the nuclear envelope in a copy-number- and lamin-dependent manner. *Cold Spring Harbor symposia on quantitative biology* **75**: 555-565.

Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* **13**: 36-46.

Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. 1994. The protein encoded by the Drosophila position-effect variegation suppressor gene Su (var) 3-9 combines domains of antagonistic regulators of homeotic gene complexes. *The EMBO journal* **13**: 3822.

Ui A, Nagaura Y, Yasui A. 2015. Transcriptional elongation factor ENL phosphorylated by ATM recruits polycomb and switches off transcription for DSB repair. *Mol Cell* **58**: 468-482.

van Attikum H, Fritsch O, Gasser SM. 2007. Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks. *Embo j* **26**: 4113-4125.

van Attikum H, Fritsch O, Hohn B, Gasser SM. 2004. Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell* **119**: 777-788.

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**: 58-69.

Vissers JH, van Lohuizen M, Citterio E. 2012. The emerging role of Polycomb repressors in the response to DNA damage. *J Cell Sci* **125**: 3939-3948.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833-1837.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30**: 1327-1338.

Wang H, Zhai L, Xu J, Joo HY, Jackson S, Erdjument-Bromage H, Tempst P, Xiong Y, Zhang Y. 2006. Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol Cell* **22**: 383-394.

Watt F, Molloy PL. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development* **2**: 1136-1143.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**: 853-862.

Weintraub H, Groudine M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* **193**: 848-856.

Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M et al. 2006. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**: 467-481.

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P et al. 2006. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**: 86-90.

Xu L, Fong Y, Strome S. 2001. The Caenorhabditis elegans maternal-effect sterile proteins, MES-2, MES-3, and MES-6, are associated in a complex in embryos. *Proc Natl Acad Sci U S A* **98**: 5061-5066.

Yamane K, Toumazou C, Tsukada Y, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. 2006. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**: 483-495.

Yang H, Zhang Y, Vallandingham J, Li H, Florens L, Mak HY. 2012. The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in C. elegans. *Genes Dev* **26**: 846-856.

Yeates TO. 2002. Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* **111**: 5-7.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC. 2006. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**: 747-757.

Yuan Y, Tang A, Castoreno A, Kuo S, Wang Q, Kuballa P, Xavier R, Shamji A, Schreiber S, Wagner B. 2013. Gossypol and an HMT G9a inhibitor act in synergy to induce cell death in pancreatic cancer cells. *Cell death & disease* **4**: e690.

Yuzyuk T, Fakhouri TH, Kiefer J, Mango SE. 2009. The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in C. elegans embryos. *Developmental cell* **16**: 699-710.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48**: 1385-1395.

Zhao Q, Wang QE, Ray A, Wani G, Han C, Milum K, Wani AA. 2009. Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *The Journal of biological chemistry* **284**: 30424-30432.

Zheng C, Karimzadegan S, Chiang V, Chalfie M. 2013. Histone methylation restrains the expression of subtype-specific genes during terminal neuronal differentiation in Caenorhabditis elegans. *PLoS Genet* **9**: e1004017.

Zilberman D, Cao X, Jacobsen SE. 2003. ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science* **299**: 716-719.

Chapter 2: Repeat DNA in genome organization and stability

Jan Padeken, Peter Zeller and Susan M Gasser

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

First two authors contributed equally to this work.

Peter Zeller wrote the introduction, repeat classification, the role of repetitive elements in disease and prepared Fig1
Jan Padeken wrote the parts about control mechanisms in germline and soma, the future perspectives and prepared Fig2

Current Opinion in Genetics & Development 2015, 31:12–19

Summary

Eukaryotic genomes contain millions of copies of repetitive elements (RE). Although the euchromatic parts of most genomes are clearly annotated, the repetitive/heterochromatic parts are poorly defined. It is estimated that between 50 and 70% of the human genome is composed of REs. Despite this, we know surprisingly little about the physiological relevance, molecular regulation and the composition of these regions. This primarily reflects the difficulty that REs pose for PCR-based assays, and their poor map-ability in next generation sequencing experiments. This chapter gives a detailed summary of the nature and classification of REs and their importance in disease pathology. Additionally, we give a detailed overview of the recent advances in understanding the transcriptional regulation of these sequences. We put special emphasis on the differences between somatic and germline specific regulatory mechanisms.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Repeat DNA in genome organization and stability

Jan Padeken[1,3], Peter Zeller[1,2,3] and Susan M Gasser[1,2]

Eukaryotic genomes contain millions of copies of repetitive elements (RE). Although the euchromatic parts of most genomes are clearly annotated, the repetitive/heterochromatic parts are poorly defined. It is estimated that between 50 and 70% of the human genome is composed of REs. Despite this, we know surprisingly little about the physiological relevance, molecular regulation and the composition of these regions. This primarily reflects the difficulty that REs pose for PCR-based assays, and their poor map-ability in next generation sequencing experiments. Here we first summarize the nature and classification of REs and then examine how this has been used in the recent years to broaden our understanding of mechanisms that keep the repetitive regions of our genomes silent and stable.

Addresses
[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland
[2] Faculty of Natural Sciences, University of Basel, Basel, Switzerland

Corresponding author: Gasser, Susan M (susan.gasser@fmi.ch)
[3] These authors contributed equally to this work.

Current Opinion in Genetics & Development 2015, **31**:12–19

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Barbara Panning** and **Eran Segal**

For a complete overview see the Issue and the Editorial

Available online 29th April 2015

http://dx.doi.org/10.1016/j.gde.2015.03.009

Repeat classification

Repetitive elements (RE) are simply defined as sequences that occur multiple times in the genome. This encompasses a huge variety of DNA elements of very diverse structure and origin. REs can be grouped into two broad, but distinct classes. Namely, tandem repeats are small nucleotide stretches repeated in a head to tail orientation, while transposable elements are DNA stretches with the ability to move from one place of the genome to another. Despite their discovery nearly half a decade ago [1] and the advent of whole genome sequencing data, the list of REs is still increasing, as is their calculated contribution to the human genome [2•]. In Figure 1, we summarize the main repeat classes found across species, and their genomic contribution in widely used model organisms.

Tandem repeats

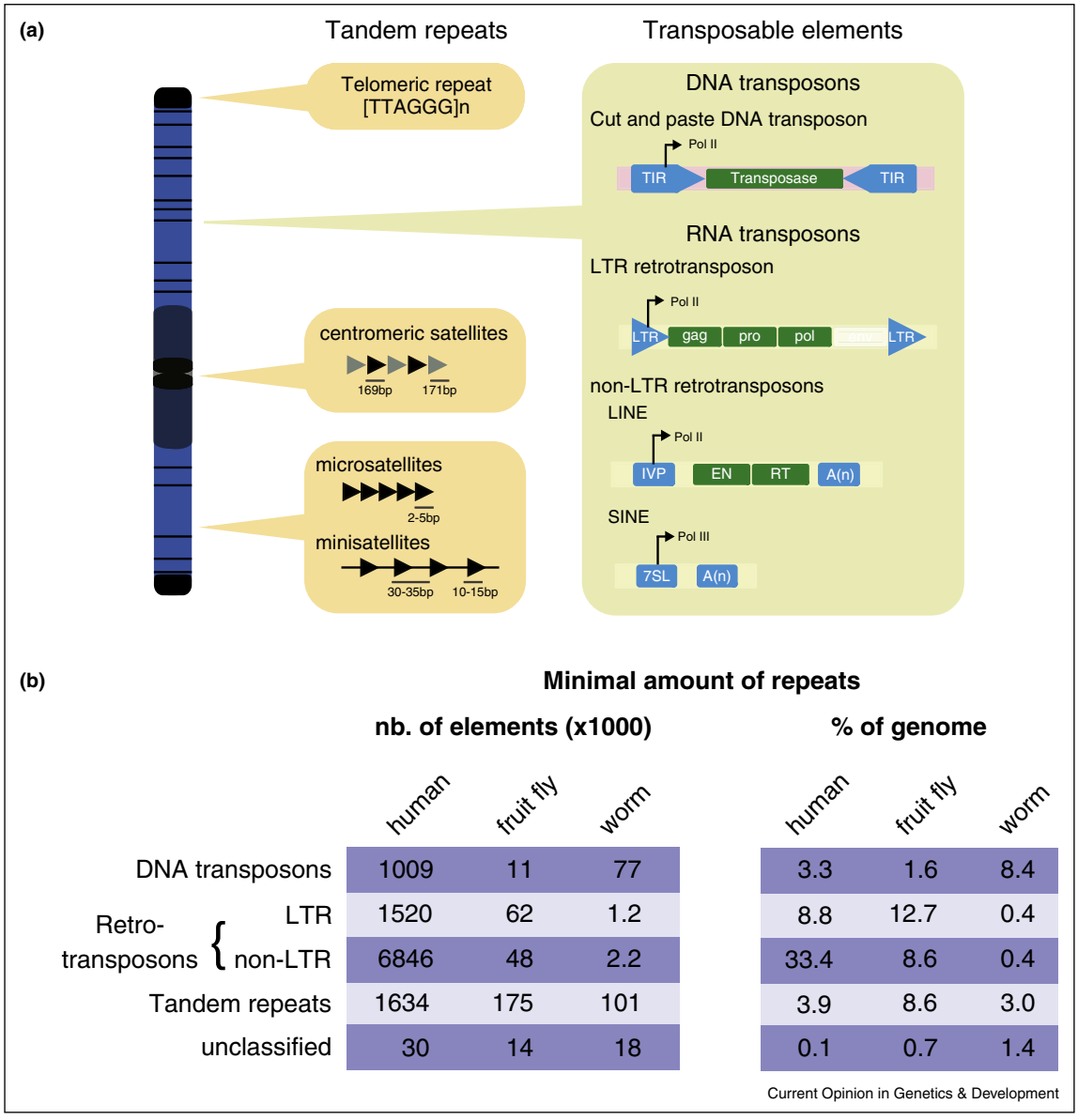
Tandem repeats are short, non-coding sequence stretches that are repeated in a head to tail fashion. They can be categorized according to the size of their building block and total length, forming microsatellite and minisatellite categories. Microsatellites include simple di-nucleotide to penta-nucleotide repeats with a total length of hundreds of basepairs (bp), while minisatellites have a unit length of 30–35 bp with a conserved core sequence of 10–15 bp. The total length of minisatellites ranges from 1 to 15 kb. Not covered by this classification are centromeric and telomeric satellite repeats. At centromeres, organism-specific tandem satellite repeats of ~200 bp building blocks are strongly enriched around the functional kinetochore. In humans, the major centromeric unit is the alpha-satellite sequence, which stretches in a repetitive manner for 3–5 Mb from the centromere. Mouse centromeres are similarly flanked by the major satellite repeat, which form most of the pericentromeric heterochromatin in this organism.

In contrast, tandem repeats are thought to result from improper replication and are highly variable even among genetically related individuals [3]. Even though they are occasionally found in the coding region of genes (e.g., the RUNX-2 and the Huntingtin gene in vertebrates [4]), tandem repeats themselves do not encode for protein and usually have neither promoter nor enhancer function. In most organisms, tandem repeats accumulated in pericentric heterochromatin and at telomeres, where they contribute to chromosome structure. At telomeres they protect the ends of the linear chromosomes and at centromeres they are essential for sister chromatid cohesion and kinetochore function [5]. Finally, as discussed below, trinucleotide microsatellites have garnered much attention, because their expansion is often linked to human disease states [6•].

Transposable elements (class I versus class II autonomous versus non-autonomous)

Transposable elements are sometimes referred to as selfish DNA, and these sequences have the ability to change their position in the genome. In contrast to tandem repeats, transposable elements have a defined structure: they are of specific size, are flanked by short repeats and often encode for a unique set of proteins. On the basis of the intermediate form that is used for transposition they are divided into class I, or RNA transposons, which travel by a 'copy and paste' mechanism through an RNA intermediate, and class II, or DNA transposons, that jump as a DNA molecule that is either cut out from the genome ('cut-and-paste') or copied from the genome

Figure 1



(a) Schematic representation of repeat classes, their distribution and their structural hallmarks. **(b)** Repeat class distribution in mammals, insects and nematodes. The numbers are based on the repeat masker database and due to limitations in map-ability should only be viewed as an estimation.

('rolling circle' and 'self-replicating') as DNA. Both classes are further subdivided based on the presence of terminal repeats, the genes they encode, and their exact mechanism of transposition. If a transposon expresses all of the proteins necessary for its transposition, it is referred to as autonomous. Other families of transposons express only part of this machinery or none at all, and are called non-autonomous. Their transposition depends on the enzymatic machinery expressed by other superfamily members. Recently, genome-sequencing has allowed one to define families of active transposons by the presence of organism-specific copies, and the degree of sequence conservation between copies in the same genome, which is an indication that they have not had time to accumulate mutations.

The two major RNA transposons are LTR and non-LTR retrotransposons. LTR retrotransposons derive from ancient retroviral infections. They are characterized by their long terminal repeats (LTR) at the 5′ and 3′ ends, and are

closely related to retroviruses. Consequently, their autonomous copies encode for proteins closely related to the retroviral gag (structural proteins of the virus core), pol (reverse transcriptase, integrase), pro (protease) and in some cases even functional env (envelope) proteins. The youngest LTR transposons are still transcriptionally active, such as the ERV-I, ERV-II and ERV-K transposons in humans [7], and the Intracisternal A-particle (IAP) retrotransposon in mice [8].

Non-LTR interspersed retrotransposons in their fully autonomous form are also referred to as long interspersed elements (LINE). LINEs have an internal 5′ promoter that drives the expression of the transposition machinery, which consists of a reverse transcriptase and an endonuclease. The LINE transcript is recognized by the transposition machinery through a 5′ polyA tail. Short interspersed elements (SINE), on the other hand, are non-autonomous non-LTR retrotransposons. They depend differentially on LINE elements for their transposition, with specificity being determined by their 5′ tails. Most SINEs are derived from tRNA, 7SL RNA or 5s RNA and therefore possess a RNA-Pol III promoter. The youngest and transpositionally most active non-LTR transposons in humans are the L1 LINE elements and the corresponding SINE Alu elements [9].

Retrotransposition generally involves three steps; the transcription of the genomic element, the reverse transcription of that RNA in 5′–3′ direction, and then the integration of the newly synthesized DNA into a new genomic locus. This has two major consequences: first, only transcriptionally active transposons can transpose and second, interruption of the reverse transcription leads to 5′ truncation, which can result in the loss of the promoter region and transcriptional inactivation [10].

The only DNA transposons that are found active (and only in certain species) are 'cut-and paste' transposons. They express one enzyme — a transposase — that recognizes the terminal inverted repeats (TIR) that flank the transposon, and catalyze the excision of the transposon from its original position, as well as its integration into a new position. Transposon amplification depends on the repair machinery of the host cell and its cell cycle stage. In contrast to retrotransposons, the transposition of a DNA transposon does not require active transcription of the element, but rather on the presence of intact TIRs and a transposase. Thus, a single expressed DNA transposon can result in the active transposition of all its family members, at least in principle. A prominent example of a species with active DNA transposons is the nematode *Caenorhabditis elegans*.

Repeat-linked diseases

An increasing number of diseases have been shown to have links to repetitive sequences. Till date, 22 diseases are correlated with changes in the repeat length, most of which are tandem repeat expansions. These include well-known heritable diseases like Huntington's disease, Friedrich Ataxia and the Fragile X syndrome [11]. A main characteristic of these diseases is that an expansion of the repeat over a crucial threshold (e.g., 200 copies) leads to transcriptional repression of the repeat-linked gene. In rare cases, such as Facioscapulohumeral muscular dystrophy (FSHD), shortening of the disease-linked D4Z4 repeat leads to the transcription of an otherwise silent locus.

Disease-linked changes in repeat element length can occur either near a gene (e.g., FSHD), or with a gene's coding sequence (e.g., Huntington's disease). The molecular mechanisms leading to repeat alterations, and their impact on disease phenotypes, are often obscure. In some cases, expanded tandem repeats have the ability to nucleate heterochromatin at the repeat sequence and/or in the flanking regions [12,13]. Interestingly, it was shown that transcriptionally active repeats are more probably to expand [14], suggesting an important role for repeat-facilitated silencing, in the prevention of disease. In other cases, variations in the exact sequence that is amplified can contribute to dysfunction [15]. In the case of myotonic dystrophy, for example, it was shown that the transcribed repeat forms RNA hairpins that sequester Musclebind1, a protein that normally promotes the splicing of its target RNAs [16].

Comparative studies among different repeat-provoked diseases, showed that each locus may have a specific mode of action. In case of Friedrich Ataxia, for example, even though repeat expansion leads to increased methylation of histone H3 K9, this methylation mark is not crucial for the transcriptional effect. Rather, the repeat itself appears to interfere with transcriptional elongation [12]. Indeed, each specific repeat-linked disease may require an in depth study of its mechanism of action and the correlated pathology. For Huntington's disease alone there are now more than 31 disease models [17]. It therefore becomes important to understand the molecular mechanisms that establish and maintain heterochromatin on tandem repeats during normal development, in order to explain the various ways in which repeat stability can affect heterochromatin in disease.

Repeats in cancer

Besides effects at specific disease genes, both tandem repeats and transposable elements have a major impact on global genome integrity. Palindromic sequences, such as inverted repeats are found to be enriched at common fragile sites in cancer genomes. These are thought to form secondary structures during replication, leading to fork stalling and the formation of double-strand breaks and unscheduled recombination events [18,19]. Replication forks also tend to slow down and stall at GC-rich repeats

possibly due to G-quadruplex formation [20]. In addition to favoring breaks, repeat sequences pose an additional problem during repair, because template switching between repeats at dispersed sites in the genome can lead to chromosomal rearrangements [21].

Transposons provoke specific alterations of gene structure or expression by being excised and subsequently inserted, either directly into the coding sequence or into regulatory elements of genes. Due to the presence of promoter and enhancer elements within the transposon itself, integration near a gene can provoke inappropriate gene expression [22,23]. Transposition events can also introduce genetic information that originates from sequences outside of the transposon. A recent study described this for the long non-LTR retrotransposon L1 [24••], which was shown to be quite potent at introducing extra sequence arising from its locus of origin upon transposition. This event was found in ~25% of analyzed cancer samples [24••], and although most insertions did not lead to cancer-promoting alterations, L1 transduction was nonetheless oncogenic in a subset of hepatocellular carcinomas [25]. Such transpositions, with or without additional DNA, also contributes to the genetic flexibility of cancer genomes, in some cases facilitating the development of therapy-resistant tumors.

The most common cellular mechanism that prevents activation and expansion of REs is the formation of heterochromatin over their sequences. One case is the highly repetitive rDNA locus: in *Drosophila* it was shown that loss of SU(VAR)3-9, an important histone methyl transferase that deposits the repressive H3K9me2/3 mark, leads to a transcriptional activation of REs and the disruption of the nucleolus. This was correlated with an increase in extrachromosomal circular DNA containing rDNA, a result of aberrant recombination [26]. Below we summarize the current state of knowledge on the silencing of non-rDNA repetitive DNA elements, which involves the interplay of multiple repressive pathways in a developmental-stage dependent manner.

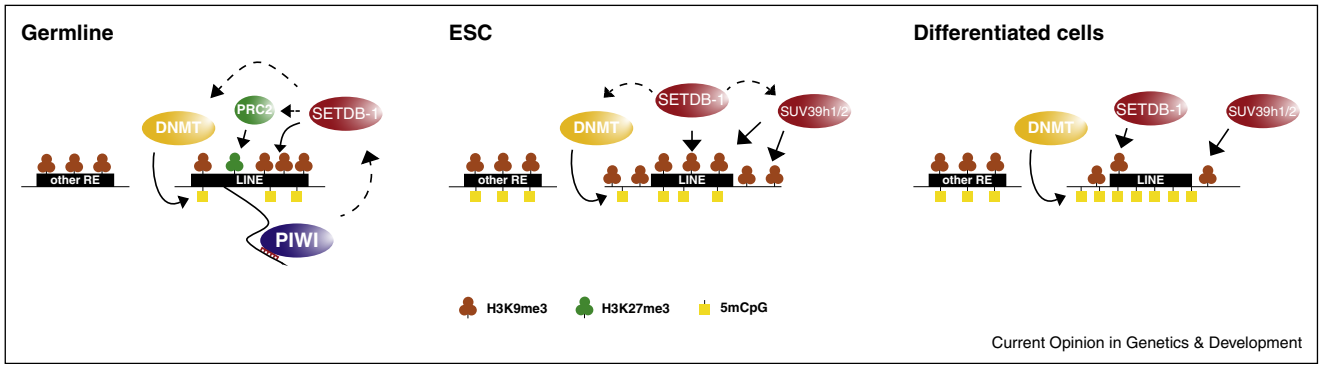
Importance of heterochromatin at RE

Three epigenetic pathways ensure the silencing of RE: methylation of H3K9, DNA methylation and the germ-line specific PIWI pathway. Several recent findings suggest that the three pathways are interconnected and that the importance of each pathway for silencing is dependent on the developmental stage. In somatic cells, as well as in the cells of the germline, transposable elements are enriched for H3K9me3, and — in plants and vertebrates — for DNA CpG methylation as well. Interestingly, the extent of histone methylation seems to differ between individual repeat families and even between individual repeats of the same family. Families that are known to be capable for high transcriptional and tranpositional activity have particular high levels of H3K9me3. In mouse, these are the LTR families IAPEz, IAPEy, and MaLR, and the LINE families L1-Gf, T, and A [27]. A similar pattern can be observed for DNA methylation, especially in differentiated cells (summarized in Figure 2; [28••]).

Silencing of RE in the germline

The threat that RE poses to genomic integrity is particularly crucial in germ cells, as genomic changes in germline cells will be transmitted to the next generation. Moreover gross rearrangements early in meiosis may impair meiotic synapsis. Intriguingly, several transposable elements have evolved to be expressed [29], or transposed, exclusively in germ cells [30]. Cells of the germline are also subject to widespread epigenetic reprogramming during which neither DNA methylation,

Figure 2



Model of the repressive complexes and marks involved in the silencing of repetitive elements during the mammalian development. In the germline LINE elements retain methylation of H3K9 and DNA. They also acquire methylation of H3K27. Both H3K27me3 and DNA methylation are dependent on the methylation of H3K9. The small RNA mediated PIWI pathway seems to be at least partially upstream of H3K9me3. Embryonic stem cells (ESC) show a strong enrichment of H3K9me3 on LINE elements, which is mediated by SETDB-1 and SUV39h1/2. H3K9me3 levels drop in differentiated cells (MEFs), which coincides with an increase in DNA methylation. So far studies mainly focused on the regulation of LINE elements. It remains unclear if other repetitive elements (RE) are regulated in the same way. Dotted lines represent potential recruitment pathways.

nor — in the paternal genome — H3K9me3, are globally maintained (reviewed in [31]). This, of course, raises the question, of how RE silencing is ensured during this crucial developmental time window, and how the various repression pathways crosstalk, to ensure a stable propagation of heterochromatin to the next generation.

During mouse germ-cell development DNA methylation is lost in proliferating pre-migratory and post-migratory primordial germ cells (PGCs) due to reduced levels of expression of DNMT3a and sequestration of DNMT3b and the DNMT1 cofactor NP95 in the cytoplasm [32,33]. DNA methylation on repetitive regions is then re-established during spermatogenesis and during the postnatal growth phase of the oocytes, by DNMT3a and DNMT3b [34]. In parallel, and before DNA methylation is completely reestablished, H3K9me3 is progressively lost during spermatogenesis, and is only re-established during early embryogenesis [31]. Interestingly, a small percentage of the genome retains both DNA methylation and H3K9me3 throughout PGC production, including several LINE and LTR element families [33,35,36••].

In contrast to somatic cells, these specific LINE and LTR elements are not only enriched for H3K9me3 and DNA methylation, but also for H3K27me3. The deletion of SETDB1, a major H3K9me3 methyltransferase, results in the loss of all three silencing marks specifically at LINE and LTR elements. This leads to the transcriptional activation of a subset of those transposable elements [36••,37]. This cascade of events suggests that in PGCs, H3K9me might function upstream of DNA methylation and Polycomb-mediated silencing at transposable elements. Why transposable elements are additionally marked by H3K27me3 in PGCs remains unclear. It is, however, interesting that after fertilization the paternal pericentric heterochromatin is highly enriched for H3K27me3 and Polycomb repressive complex 1 (PRC1) [38].

Another important pathway that silences transposable elements in a germ-line specific manner, is the PIWI pathway. It mediates silencing through a complex of small RNAs, called piRNAs, and the PIWI family of Argonaut proteins (PIWI, AUBERGINE and AGO3 in *Drosophila* and MIWI, MIWI2 and MILI in mouse). The importance of this RNA-mediated silencing mechanism is clear: the deletion of various subunits of the piRNA biogenesis pathway results in sterility, and this correlates with the reactivation of IAP and/or LINE1 elements in postnatal male germ cells, both in mouse and *Drosophila* [39].piRNAs share homology with transposable elements, and partially derive directly from their transcripts. The sequence homology allows for the specific slicing of transposon mRNA by the Argonaut proteins [40]. Loss of MILI, or MIWI2 also results in the loss of *de novo* DNA methylation of IAPs and LINE elements [41], indicating that epigenetic modifications may lie downstream of the PIWI pathway. Indeed, recent studies found that the targeting of actively transcribed transposons by the piRNA/Piwi complex results in an increase of H3K9me3 and DNA methylation over transposable elements in mouse and *Drosophila* [27,42•,43]. Interestingly, in neither *Drosophila* nor mouse, did the disruption of PIWI result in the complete loss of H3K9me3. Instead, H3K9me3 levels were reduced to a level comparable with somatic cells. In some cases this led to the transcriptional activation of RE, while others remained repressed [27]. In mouse, the loss of H3K9me3 was restricted to a subset of LINE1 loci that contain full-length retrotransposon insertions, but not fragmented non-autonomous copies present throughout the genome.

It is striking that in these studies, loss of either DNA methyl transferases, histone methyl transferases (HMTs), or the PIWI pathway only affected a subset of RE's. This suggests that the pathways have distinct subsets of targets and are not entirely epistatic. Particularly in mouse the effect was limited to potentially active transposable elements of the IAP and LINE family. In case of the PIWI pathway, the limited effect on a subset of LINE elements probably stems from the mechanism of RNAi mediated silencing. A comparison between the PIWI-dependent repeat families in *Drosophila* and their transcription status strongly suggests that PIWI requires ongoing transcription in order to silence [42•]. PIWI is therefore restricted to function as an additional layer of silencing, preventing the transposition of active transposons and potentially marking new transposon insertions for repression. On the other hand, heterochromatin coats all repeat families, whether they are full-length and potentially active transposons, or fragments lacking their endogenous promoters, and therefore transcriptionally inactive. How these elements are targeted for chromatin-mediated repression is unclear, nor is it clear how tandem repeats are silenced.

Silencing of RE in somatic cells

A well-studied example of somatic cells is the embryonic stem cell (ESC). These are cultured cells originating from the inner cell mass (ICM) of the blastocyst stage of mammalian embryos. In this system, the silencing of REs is highly dependent on H3K9 methylation, while DNA methylation becomes essential only after further differentiation [28••]. Importantly, the components of the PIWI pathway are not expressed in ESC nor in other somatic cells. On the other hand, studies in *Drosophila* and human tissue culture cells showed that there might be an endogenous siRNA that targets transposable elements in somatic cells of these organisms as well [44,45]. It is not yet clear how these endo siRNA pathways are related to the remaining silencing pathways, or whether they are related to the siRNA pathway described for *Schizosaccharomyces pombe*.

Two recent studies analyzed the distribution of H3K9me3 carefully for the individual repeat families [28••,36••]. This showed that in ESCs, the H3K9me3

distribution is similar to its distribution in germ cells. A particular strong enrichment could be observed at centromeric-satellites and telomeric-satellites, even though the absolute levels of methylation are hard to determine due to the highly repetitive nature of these regions. Also enriched for H3K9me3 are SINE elements, LTR transposons of the ERV1 and ERVK families (IAPs) and LINE elements (e.g., active L1-Gf, T, and A families) [28••,36••]. Within these classes, the enrichment of H3K9me3 was highly specific for some families [36••]. As in germ cells, H3K9me3 enrichment in ESCs was particularly strong at families that encode transcriptionally active transposons.

It remains unclear what mechanisms nucleate H3K9 methylation, and whether or not the distribution observed in ESC is a consequence of what was established in the germline. These questions are further complicated by the redundancy of the four major H3K9 HMTs, namely, G9a, SetDB1, Suv39h1 and Suv39h2. A comparison of the SUV39H1/2 and SETDB1 distribution in mouse ES cells showed that both HMTs are enriched at euchromatic insertion sites of transposable elements. Both enzymes were found on full-length ERVs, and SUV39H1/2 was additionally enriched on full length LINE elements, although they had but a weak enrichment for SETDB1. Elimination of the relevant HMT genes led to the transcriptional up-regulation of these elements in living organisms.

It is striking that in somatic cells, as well as in germ cells, loss of HMTs or components of the PIWI pathway primarily affects full length LINE and ERV elements. As mentioned above, the mechanism(s) that nucleate heterochromatin at incomplete and degenerate copies of transposons, and at tandem simple repeats, are unknown. One hint is given by the following observations: whereas SUV39h1/2 and SETDB-1 co-localized directly on ERV elements, SUV39h1/2 was found to spread into sequences adjacent to the repeats. In agreement with this, deletion of SUV39h1/2 resulted in a complete loss of H3K9me3 on the surrounding sequences [28••], and only reduced H3K9me3 on ERV elements themselves. Together with the observation that REs tend to accumulate in clusters, this suggests the possibility that SETDB1 is targeted to full length and potential transcriptional active transposable elements, where it either directly, or indirectly through H3K9me, recruits SUV39h1/2. Thanks to its interaction with the H3K9me-reader HP1α, SUV39h1/2 has the ability to catalyze the spreading of H3K9me into neighboring sequences. It may, therefore, mediate heterochromatin formation on larger chromatin domains, without requiring a mechanism that recognizes all RE sequences in order to nucleate the HMTs action. This mechanism also implies that the different HMTs may be interdependent at least for a subset of target sites.

Future challenges

Recent studies have collectively shown that H3K9me, the PIWI pathway and DNA methylation are interdependent for the silencing and stabilization of transposable elements in organisms containing CpG methylation. Moreover, H3K9me seems to be the link that connects RNA-mediated silencing and DNA methylation. The current data suggests a degree of selectivity, specificity, as well as redundancy among H3K9 specific HMTs. A second major conclusion is that three pathways of RE silencing are differentially active at different developmental stages. The exact reasons for this are not yet clear. Finally, recent data show that it is technically possible to use next generation sequencing approaches, such as ChIPseq and RNAseq, to analyze the biology of REs. Nonetheless, the interpretation of the data is highly dependent on whether repeat classes, repeat families, or individual repeats are analyzed. For instance, Pezic and colleagues concluded that SINE elements are depleted for H3K9me3, when looking at the methylation state of the complete repeat class [27], yet Bulut-Karslioglu and colleagues showed that around 10% of all annotated SINES are actually enriched for H3K9me3 (n.b. a percentage similar to that of LINE and ERVs enriched for H3K9me3 [28••]). These complications make this important field one that depends largely on new technologies that will improve the mapping of specific REs and help identify their mechanism of control.

Future studies should be able to distinguish regulation by H3K9me3, from regulation by mono-methylation and di-methylation of H3K9. Given that H3K9me2 is able to mediate silencing at some loci [46], it is essential to know whether a reduced H3K9me3 signal results in the complete loss of H3K9 methylation, or in an increase in mono-methylated and di-methylated forms. It will be intriguing to see whether different repeat families are selectively marked by H3K9me2 or H3K9me3 and whether the trimethyl state is due to the action of a single HMT or several HMTs. This can be ideally studied in an organism that does not depend on H3K9me for sister chromatid cohesion during cell division [47,48•].

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. McClintock B: **The origin and behavior of mutable loci in maize**. *Proc Natl Acad Sci U S A* 1950, **36**:344-355.

2. • de Koning AP *et al.*: **Repetitive elements may comprise over two-thirds of the human genome**. *PLoS Genet* 2011, **7**:e1002384.

A re-analysis of the repeat content in the human genome which illustrates that REs are still underrepresented in the current genome assemblies.

3. Lopez Castel A, Cleary JD, Pearson CE: **Repeat instability as the basis for human diseases and as a potential target for therapy**. *Nat Rev Mol Cell Biol* 2010, **11**:165-170.

4. Sears KE *et al.*: **The correlated evolution of Runx2 tandem repeats, transcriptional activity, and facial length in carnivora**. *Evol Dev* 2007, **9**:555-565.

5. Bernard P *et al.*: **Requirement of heterochromatin for cohesion at centromeres**. *Science* 2001, **294**:2539-2542.

6. • Gatchel JR, Zoghbi HY: **Diseases of unstable repeat expansion: mechanisms and common principles**. *Nat Rev Genet* 2005, **6**:743-755.
Comprehensive summary of the most common trinucleotide repeat expansion diseases and their underlying molecular mechanisms.

7. Bannert N, Kurth R: **The evolutionary dynamics of human endogenous retroviral families**. *Annu Rev Genomics Hum Genet* 2006, **7**:149-173.

8. Heidmann O, Heidmann T: **Retrotransposition of a mouse IAP sequence tagged with an indicator gene**. *Cell* 1991, **64**:159-170.

9. Sassaman DM *et al.*: **Many human L1 elements are capable of retrotransposition**. *Nat Genet* 1997, **16**:37-43.

10. Smit AF: **Interspersed repeats and other mementos of transposable elements in mammalian genomes**. *Curr Opin Genet Dev* 1999, **9**:657-663.

11. La Spada AR, Taylor JP: **Repeat expansion disease: progress and puzzles in disease pathogenesis**. *Nat Rev Genet* 2010, **11**:247-258.

12. Punga T, Buhler M: **Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation**. *EMBO Mol Med* 2010, **2**:120-129.

13. Ghorbani M *et al.*: **Comparative (computational) analysis of the DNA methylation status of trinucleotide repeat expansion diseases**. *J Nucleic Acids* 2013, **2013**:689798.

14. Goula AV *et al.*: **Transcription elongation and tissue-specific somatic CAG instability**. *PLoS Genet* 2012, **8**:e1003051.

15. McMurray CT: **Mechanisms of trinucleotide repeat instability during human development**. *Nat Rev Genet* 2010, **11**:786-799.

16. Kanadia RN *et al.*: **A muscleblind knockout model for myotonic dystrophy**. *Science* 2003, **302**:1978-1980.

17. Pouladi MA, Morton AJ, Hayden MR: **Choosing an animal model for the study of Huntington's disease**. *Nat Rev Neurosci* 2013, **14**:708-721.

18. Wang H *et al.*: **CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity**. *Mol Cell* 2014, **54**:1012-1021.

19. Kato T, Kurahashi H, Emanuel BS: **Chromosomal translocations and palindromic AT-rich repeats**. *Curr Opin Genet Dev* 2012, **22**:221-228.

20. Sabouri N, Capra JA, Zakian VA: **The essential *Schizosaccharomyces pombe* Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage**. *BMC Biol* 2014, **12**:101.

21. Anand RP *et al.*: **Chromosome rearrangements via template switching between diverged repeated sequences**. *Genes Dev* 2014, **28**:2394-2406.

22. Yang Z *et al.*: **Apolipoprotein(a) gene enhancer resides within a LINE element**. *J Biol Chem* 1998, **273**:891-897.

23. Feschotte C, Pritham EJ: **DNA transposons and the evolution of eukaryotic genomes**. *Annu Rev Genet* 2007, **41**:331-368.

24. •• Tubio JM *et al.*: **Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes**. *Science* 2014, **345**:1251343.
Comprehensive study about the high contribution of retrotransposition to the genetic background of cancer. The data show how only a hand full of active retrotransposons can spread nearby genetic information (3′ of the transposon) throughout the whole genome.

25. Shukla R *et al.*: **Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma**. *Cell* 2013, **153**:101-111.

26. Peng JC, Karpen GH: **H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability**. *Nat Cell Biol* 2007, **9**:25-35.

27. Pezic D *et al.*: **piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells**. *Genes Dev* 2014, **28**:1410-1428.

28. •• Bulut-Karslioglu A *et al.*: **Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells**. *Mol Cell* 2014, **55**:277-290.
Comparison of the distribution of SUV39h1/2, SETDB1 and H3K9me3 in mouse ESC, indicating that the two HMTs have only a partially overlapping role in the silencing of intact LINE elements.

29. Calvi BR, Gelbart WM: **The basis for germline specificity of the hobo transposable element in *Drosophila melanogaster***. *EMBO J* 1994, **13**:1636-1644.

30. Laski FA, Rio DC, Rubin GM: **Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing**. *Cell* 1986, **44**:7-19.

31. Probst AV, Almouzni G: **Heterochromatin establishment in the context of genome-wide epigenetic reprogramming**. *Trends Genet* 2011, **27**:177-185.

32. Hajkova P *et al.*: **Epigenetic reprogramming in mouse primordial germ cells**. *Mech Dev* 2002, **117**:15-23.

33. Seisenberger S *et al.*: **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells**. *Mol Cell* 2012, **48**:849-862.

34. Kato Y *et al.*: **Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse**. *Hum Mol Genet* 2007, **16**:2272-2280.

35. Kobayashi H *et al.*: **High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice**. *Genome Res* 2013, **23**:616-627.

36. •• Liu S *et al.*: **Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells**. *Genes Dev* 2014, **28**:2041-2055.
Analysis of a germ line specific SETDB1 knockout mouse, showing that SETDB1 is essential for H3K9me3 at LINE elements, which is upstream of DNA methylation and H3K27me3 in germ cells.

37. Leung D *et al.*: **Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1**. *Proc Natl Acad Sci U S A* 2014, **111**:6690-6695.

38. Santenard A *et al.*: **Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3**. *Nat Cell Biol* 2010, **12**:853-862.

39. Ishizu H, Siomi H, Siomi MC: **Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines**. *Genes Dev* 2012, **26**:2361-2373.

40. Watanabe T, Lin H: **Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs**. *Mol Cell* 2014, **56**:18-27.

41. Kuramochi-Miyagawa S *et al.*: **DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes**. *Genes Dev* 2008, **22**:908-917.

42. • Sienski G, Donertas D, Brennecke J: **Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression**. *Cell* 2012, **151**:964-980.
A study in *Drosophila* showing that piRNA targeting of nascent transposon transcripts contributes to the methylation of H3K9 at these loci.

43. Di Giacomo M *et al.*: **Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis**. *Mol Cell* 2013, **50**:601-608.

44. Ghildiyal M *et al.*: **Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells**. *Science* 2008, **320**:1077-1081.

45. Yang N, Kazazian HH Jr: **L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells**. *Nat Struct Mol Biol* 2006, **13**:763-771.

46. Lienert F *et al.*: **Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells**. *PLoS Genet* 2011, **7**:e1002090.

47. Dernburg AF, Sedat JW, Hawley RS: **Direct evidence of a role for heterochromatin in meiotic chromosome segregation**. *Cell* 1996, **86**:135-146.

48. • Peng JC, Karpen GH: **Heterochromatic genome stability requires regulators of histone H3 K9 methylation**. *PLoS Genet* 2009, **5**:e1000435.
A study indicating that H3K9 methylation affects integrity of heterochromatic sequences.

Chapter 3: Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability

Peter Zeller, Jan Padeken, Robin van Schendel, Veronique Kalck, Marcel Tijsterman & Susan M Gasser

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

First two authors contributed equally to this work.

Both wrote the manuscript together and were both involved in most experiments found in figure 3-6 and SupFig1-5

Peter Zeller did the majority of experiments in Fig1, Fig7, SupFig6

Jan Padeken did the majority of experiments in Fig2 and all bioinformatics analysis

Robin van Schendel did the mapping of endogenous mutations

Nature Genetics 48, 1385–1395 (2016)

Summary

Histone H3 lysine 9 (H3K9) methylation is a conserved modification that generally represses transcription. In *Caenorhabditis elegans* it is enriched on silent tissue-specific genes and repetitive elements. Studying H3K9me in developing organisms has so far proven to be difficult. Both mice and *Drosophila* have at least five histone methyl transferase (HMTs) enzymes that are essential and partially redundant allowing only for the study of partial reductions in H3K9me. In the

nematode *C. elegans* our lab identified the two methyltransferases, essential for all H3K9me (Towbin et al. 2012).

Here we show that in *met-2 set-25* double mutants embryos differentiate normally, although mutant adults are sterile owing to extensive DNA-damage-driven apoptosis in the germ line. We give a detailed analysis of H3K9me2 and me3 distribution on genes and repetitive elements in *C. elegans*. We find that transposons and simple repeats are derepressed in both germline and somatic tissues. This unprogrammed transcription correlates with increased rates of repeat-specific insertions and deletions, copy number variation, R-loops and enhanced sensitivity to replication stress. We propose that H3K9me2 or H3K9me3 stabilizes and protects repeat-rich genomes by suppressing transcription-induced replication stress.

nature genetics

Histone H3K9 methylation is dispensable for *Caenorhabditis elegans* development but suppresses RNA:DNA hybrid-associated repeat instability

Peter Zeller[1,2,4], Jan Padeken[1,4], Robin van Schendel[3], Veronique Kalck[1], Marcel Tijsterman[3] & Susan M Gasser[1,2]

Histone H3 lysine 9 (H3K9) methylation is a conserved modification that generally represses transcription. In *Caenorhabditis elegans* it is enriched on silent tissue-specific genes and repetitive elements. In *met-2 set-25* double mutants, which lack all H3K9 methylation (H3K9me), embryos differentiate normally, although mutant adults are sterile owing to extensive DNA-damage-driven apoptosis in the germ line. Transposons and simple repeats are derepressed in both germline and somatic tissues. This unprogrammed transcription correlates with increased rates of repeat-specific insertions and deletions, copy number variation, R loops and enhanced sensitivity to replication stress. We propose that H3K9me2 or H3K9me3 stabilizes and protects repeat-rich genomes by suppressing transcription-induced replication stress.

H3K9 is a common target of methylation *in vivo* and can carry one, two or three methyl groups. H3K9me2 or H3K9me3 mark transcriptionally silent heterochromatin in most eukaryotes[1–3]. In mammals, insects and *Schizosaccharomyces pombe,* H3K9 methylation is highly enriched at telomeres, pericentric heterochromatin and interspersed repetitive elements (REs)[4–7].

Ligands that recognize methylated H3K9, such as heterochromatin protein 1 (HP1), mediate transcriptional repression of reporter genes and chromatin compaction near centromeres[2,8]. H3K9me is also implicated in the silencing of genes both during development[9,10] and in pathological states. For instance, tumor-suppressor genes have been found to be transcriptionally silenced by mistargeted H3K9me in cancers[11,12], and H3K9me marks triplet repeat sequences, whose expansion has debilitating consequences in syndromes such as Huntington's or Fragile X[13,14]. Nonetheless, by reducing levels of H3K9me the efficiency of somatic cell reprogramming can be increased[15,16].

It has been difficult to study the function of H3K9me-mediated repression in complex organisms for several reasons. First, there are at least eight documented and partially redundant H3K9 histone methyltransferases (HMTs) in mammals (SUV39h1, SUV39h2, G9a, SETDB1, SETDB2, PRDM2, PRDM3 and PRDM16 in mice). Second, the vast majority of H3K9 methylation is found on extended stretches of REs that cannot be accurately mapped by standard deep sequencing techniques[17]. In some cases the disruption of individual H3K9me HMTs is embryonically lethal, owing in part to compromised mitotic chromosome segregation[18–20]. The loss of SUV39h1, SUv39h2 or their homologs also results in mitotic defects, aneuploidy and chromosomal rearrangements in mice, flies and fission yeast[7,21,22]. This may have masked phenotypes arising from the loss of H3K9me in transcriptional repression during development.

The holocentric nematode *C. elegans* has only two, nonredundant H3K9me-depositing HMTs, MET-2 and SET-25 (refs. 23,24). Here we exploited the finding that mutants lacking both HMTs have no detectable H3K9 methylation[24], and yet produce viable embryos, to study how the loss of this histone modification impacts a multicellular organism.

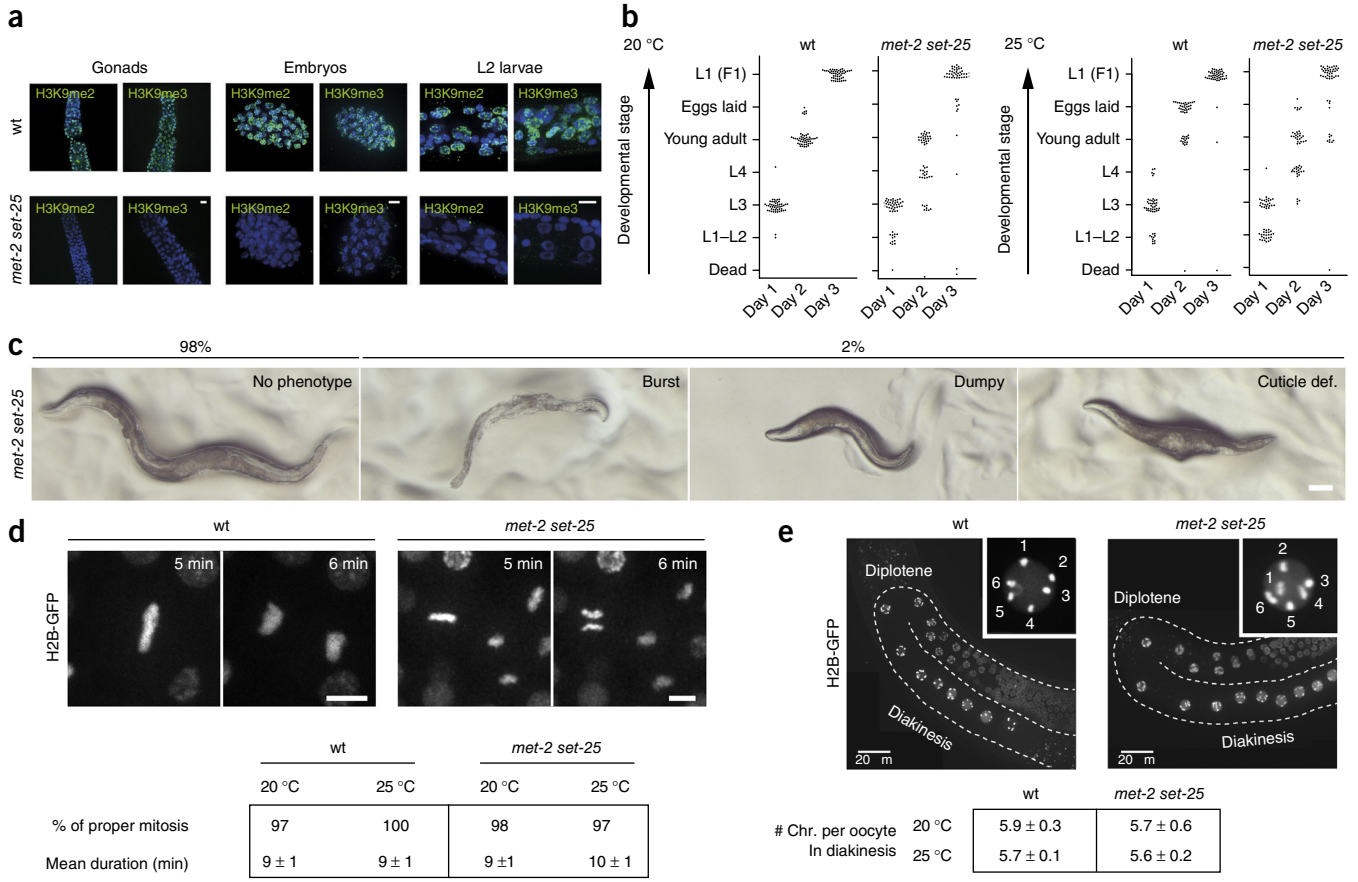
RESULTS

Loss of H3K9me did not impair embryonic differentiation into adult tissues

The HMT MET-2, which catalyzes the mono- and di-methylation of H3K9, is the homolog of mammalian SETDB1, also known as ESET[23]. SET-25, on the other hand, shares considerable SET domain homology with SUV39h1, SUV39h2 and G9a enzymes, and it is the only *C. elegans* enzyme that trimethylates H3K9 (ref. 24). To confirm that *met-2 set-25* double mutant worms lack H3K9 methylation throughout development, we performed immunofluorescence analysis at all stages of worm development (**Fig. 1a**). We found no detectable H3K9me2 or me3 in *met-2 set-25* embryos, second-stage larvae (L2) or gonads of adult worms, confirming our earlier mass spectroscopic analysis of total histones isolated from mutant embryos or larvae[24]. Histone acetylation and other common methylation marks (**Supplementary Fig. 1**) remained intact[24]. Despite this complete absence of H3K9me, the *met-2 set-25* mutant embryos developed into viable adults.

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [2]Faculty of Natural Sciences, University of Basel, Basel, Switzerland. [3]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [4]These authors contributed equally to this work. Correspondence should be addressed to S.M.G. (susan.gasser@fmi.ch).

Received 28 June; accepted 22 August; published online 26 September 2016; doi:10.1038/ng.3672



	wt		met-2 set-25	
	20 °C	25 °C	20 °C	25 °C
% of proper mitosis	97	100	98	97
Mean duration (min)	9 ± 1	9 ± 1	9 ±1	10 ± 1

		wt	met-2 set-25
# Chr. per oocyte In diakinesis	20 °C	5.9 ± 0.3	5.7 ± 0.6
	25 °C	5.7 ± 0.1	5.6 ± 0.2

Figure 1 Worms lacking H3K9me were viable but showed stochastically delayed development. (**a**) Immunofluorescence images using H3K9me2- and H3K9me3-specific antibodies on wild-type (wt) and *met-2 set-25* strains at indicated developmental stages. H3K9me2 and H3K9me3 signals are in green, and DAPI in blue. Scale bars, 5 µm. (**b**) *met-2 set-25* mutation provoked stochastic delays in development from L1 larval stage into fertile adults. Developmental progress of singled mutant and wt L1 larvae monitored every 24 h for 3 d at 20 °C and 25 °C (*N* (number of biological replicates) = 3, *n* (number of animals per replica) = 50). (**c**) Example images of worm morphologies arising in *met-2 set-25* cultures and their frequencies (*N* = 4, *n* = 50). Scale bar, 100 µm. (**d**,**e**) H3K9me2/H3K9me3 was not essential for chromosome segregation in *C. elegans*. Images (**d**) from time-lapse (Δ*t* = 1 min) movies of mitotic cells in embryos expressing H2B-GFP in which mitotic defects were scored (wt, 20 °C *n*= 34; wt, 25 °C *n* = 50; *met-2 set-25*, 20 °C *n* = 45; and *met-2 set-25*, 25 °C *n* = 36). Scale bars, 3 µm. Duration of mitosis reflects minutes from the beginning of chromosome condensation until completion of telophase. The number of bivalent chromosomes (**e**) in wt and *met-2 set-25* worms expressing H2B-GFP counted in oocytes undergoing diakinesis (*N* = 3; wt, 20 °C *n* = 57; *met-2 set-25*, 20 °C *n* = 51; wt, 25 °C *n* = 50; and *met-2 set-25*, 25 °C *n* = 50). Mean and s.d. are shown. Insets, nucleus of an oocyte in diakinesis.

To monitor the kinetics of somatic development, we compared the timing of wild-type (N2) and *met-2 set-25* organisms as they transitioned from the first larvae stage (L1) to the L1 stage of the next generation. This is a highly synchronous cycle that takes 3 d in wild-type strains grown at 20 °C (**Fig. 1b**). In contrast to wild-type worms, 52% of the *met-2 set-25* mutants showed stochastic delays in stage transitions, even though most mutant embryos reached adulthood (88% became mature adults; **Fig. 1b**). These delays were more pronounced at 25 °C than at 20 °C and were not restricted to one specific stage (**Fig. 1b**). Nonetheless, only 2% of the adult offspring displayed a grossly irregular morphology at 20 °C, i.e., 'dumpy' appearance, partially defective cuticles or bursting as adults (**Fig. 1c**). Such aberrant morphologies were below detection level (<0.1%) in wild-type populations[25].

Chromosome missegregation has been suggested to be a main cause for the phenotypes observed in mutants for H3K9 HMTs in other organisms[7,19,26]. Using histone H2B fusion to GFP (H2B-GFP), we tracked the frequency of mitotic chromosome bridges or lagging chromosomes in wild-type and *met-2 set-25* embryos. The frequency of defective mitoses at either 20 °C or 25 °C was similar in wild-type and mutant embryos (**Fig. 1d**). Moreover, the duration of mitosis was identical, which argues against any mutant-specific spindle checkpoint activation (**Fig. 1d**). To monitor meiotic chromosome missegregation, we followed H2B-GFP-tagged oocytes undergoing meiosis in gonads. Thanks to the chromosome condensation and enlarged nuclei that occur in diakinesis, we could determine bivalent chromosome number per cell. Again there was no detectable difference between *met-2 set-25* and wild-type oocytes at either temperature (**Fig. 1e**). Thus, we excluded aneuploidy and spindle checkpoint activation as triggers for the developmental delay or aberrant morphologies of H3K9me-deficient worms.

Temperature-dependent sterility of *met-2 set-25* mutant

Brood sizes were notably smaller upon propagation of the double HMT mutant, and worms became completely sterile after two generations at 26 °C (**Supplementary Fig. 2a**). We determined the number of viable progeny of *met-2 set-25* vs. wild-type worms under controlled growth conditions at 15 °C, 20 °C and 25 °C. Although brood size was equal between the *met-2 set-25* and wild-type worms at 15 °C, mutant adults had significantly fewer viable progeny at both 20 °C and

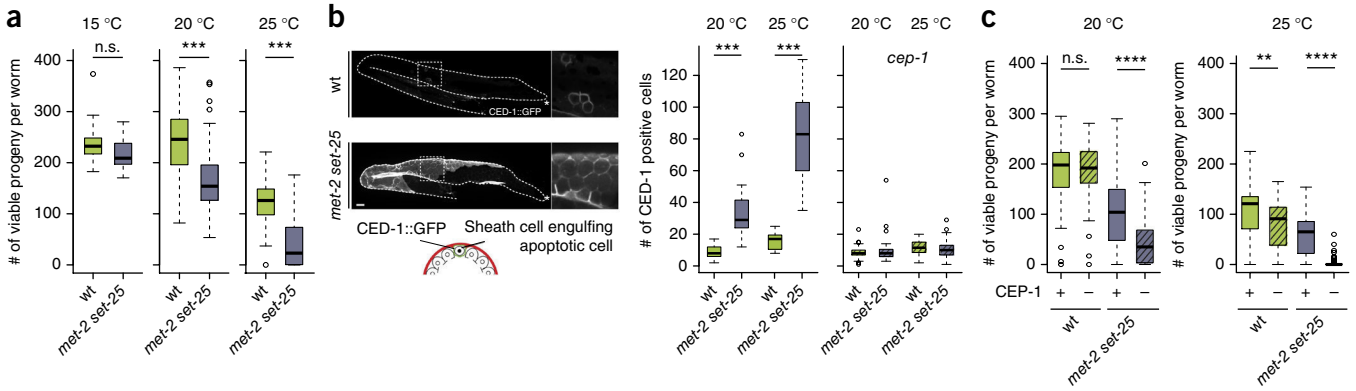


Figure 2 DNA-damage-checkpoint-dependent increase of apoptotic cells in the germ line of *met-2 set-25* worms. (**a**) Number of viable progeny of wt and *met-2 set-25* mutant per worm at 15 °C, 20 °C and 25 °C ($N = 3$, $n = 75$). (**b**) Example image of a gonad and the quantification of the number of apoptotic cells in worms expressing the apoptosis marker CED-1::GFP in wt and *met-2 set-25* background, with and without CEP-1. Apoptosis rate was determined as the number of cells fully engulfed by CED-1::GFP per gonad arm. CED-1 is a phagocytic receptor, which translocates to the plasma membrane during apoptosis. Asterisks indicate gonad tip and boxes mark enlarged section in the overview image ($N = 3$, $n = 75$). Scale bar, 10 µm. (**c**) Number of viable progeny per worm of wt and *met-2 set-25* with or without CEP-1. At both 20 °C and 25 °C ($N = 3$, $n = 75$) *cep-1* and *met-2 set-25* showed a synthetic loss of viable progeny. Boxplots show median, boxes 50% and whiskers 90% of the group. Two-sided Wilcoxon signed-rank test: n.s. indicates not significant, $^{**}P < 0.005$, $^{***}P < 0.0001$ and $^{****}P < 0.00005$.

25 °C (**Fig. 2a**). A similar temperature-dependent loss of fertility has been observed for mutants of the PIWI pathway[27,28] (**Supplementary Fig. 2b**), a germline-specific small RNA pathway that helps to silence transposable elements[29].

Gonad development *per se* was not impaired in the *met-2 set-25* mutant (**Supplementary Fig. 2c**). However, by scoring the expression of the CED-1::GFP phagocytic receptor, which accumulates on the plasma membrane of apoptotic cells[30], we detected a high level of germline apoptosis (**Fig. 2b**). The level increased when we grew worms at 25 °C. In the double mutant an average of 30 cells per germ line were positive for CED-1 at 20 °C (wild-type: 10 cells), and over 80 cells per germ line at 25 °C (wild-type: 18 cells; **Fig. 2b**). Consistently, RNA sequencing (RNA-seq) of *met-2 set-25* gonads showed an increase in mRNA from various other apoptosis-specific genes[31] (**Supplementary Fig. 2d**).

Although *C. elegans* germline cells are known to be particularly sensitive to DNA damage, germline apoptosis can have multiple causes[32]. To see whether apoptosis in H3K9me-deficient gonads is caused by DNA damage, we deleted the mammalian p53 homolog, CEP-1, and scored CED-1::GFP distribution at 20 °C and 25 °C (ref. 33). In the *met-2 set-25 cep-1* triple mutant and in the strain lacking *cep-1* alone, we detected only background levels of germline apoptosis at both temperatures (**Fig. 2b**). This strongly suggests that the germline apoptosis seen in the absence of H3K9me stemmed from DNA damage. The *met-2 set-25 cep-1* triple mutant was synthetic sterile, as expected (**Fig. 2c**). Of embryos laid at 20 °C, hatching rate dropped from above 95% in the *met-2 set-25* mutant to below 80% when coupled with *cep-1* (**Supplementary Fig. 2e**). This is likely due to an increase in DNA damage in the mutant, because the number of RAD-51 foci per cell, a marker of processed breaks, increased significantly ($P < 0.001$, two-sided Wilcoxon signed rank test), as did the number of cells in the mitotic zone of the germ line with RAD-51 foci (3.4% in the wild type and 14.6% in the double mutant; **Supplementary Fig. 2f**). This suggests that germline cells incur enhanced levels of damage in the absence of H3K9me.

H3K9me2 marks REs, whereas H3K9me3 marks REs and silent genes

To understand the link between the loss of H3K9 methylation and the observed increase in DNA damage, we first reexamined the sequences reported to be bound by histones bearing H3K9me2 and H3K9me3. We performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) experiments, not unlike those reported by the modENCODE consortium[34,35]. We found a tenfold enrichment of both H3K9me2 and H3K9me3 along the distal arms of the five worm autosomes in early embryos (**Supplementary Fig. 3a**). We did not observe this distribution for other repressive marks, such as H3K27me3, nor for the active mark, H3K4me3. Chromosome arms were similarly enriched for all types of REs (**Supplementary Fig. 3a**[36]). A detailed analysis of the distribution of H3K9me2 versus H3K9me3 in embryos showed that a high proportion of H3K9me2 was on REs (~34% of all H3K9me2), whereas H3K9me3 was present equally on exons and REs (~26% each, **Fig. 3a**).

Distinct classes of repetitive DNA constitute large fractions of the genomes of complex organisms. These include DNA or RNA transposons, which can generate copies of themselves and integrate into the genome, as well as simple repeats, such as tandemly arranged micro- or minisatellites (**Fig. 3b**). Unlike transposons, these latter repeats lack open reading frames (ORFs) and regulatory sequences. Worm genomes contain all classes of REs, although DNA (rather than RNA) transposons are the most abundant transposable elements[37]. Short repetitive sequences are not found as megabase blocks of pericentric satellite sequence in worms, but as short clusters distributed along the chromosome. As a consequence, 87% of the *C. elegans* REs, or roughly ~60,000 discrete elements, can be uniquely mapped to individual sites of the genome by standard next-generation sequencing.

Plotting the enrichment of H3K9me2 and H3K9me3 on all REs in embryos, we found that 24.3% of REs were exclusively enriched for H3K9me2, and 18.1% had either both marks or exclusively H3K9me3 (**Fig. 3c**). This revealed that 42.4% of mappable REs were enriched for H3K9me, with H3K9me2 and H3K9me3 distributed differentially over the three repeat classes. RNA transposons were most strongly correlated with H3K9me3 (58.5%, with 5.7% bearing H3K9me2 only); tandem or simple repeats were more likely to carry H3K9me2 alone (31.6%), and DNA transposons fell into two groups: 25.5% were uniquely dimethylated whereas 17.7% carried H3K9me3 (**Fig. 3c**).

In embryos H3K9me3 was enriched on transcriptionally silent genes (12.0%), where it coated entire ORFs of loci (**Fig. 3d** and

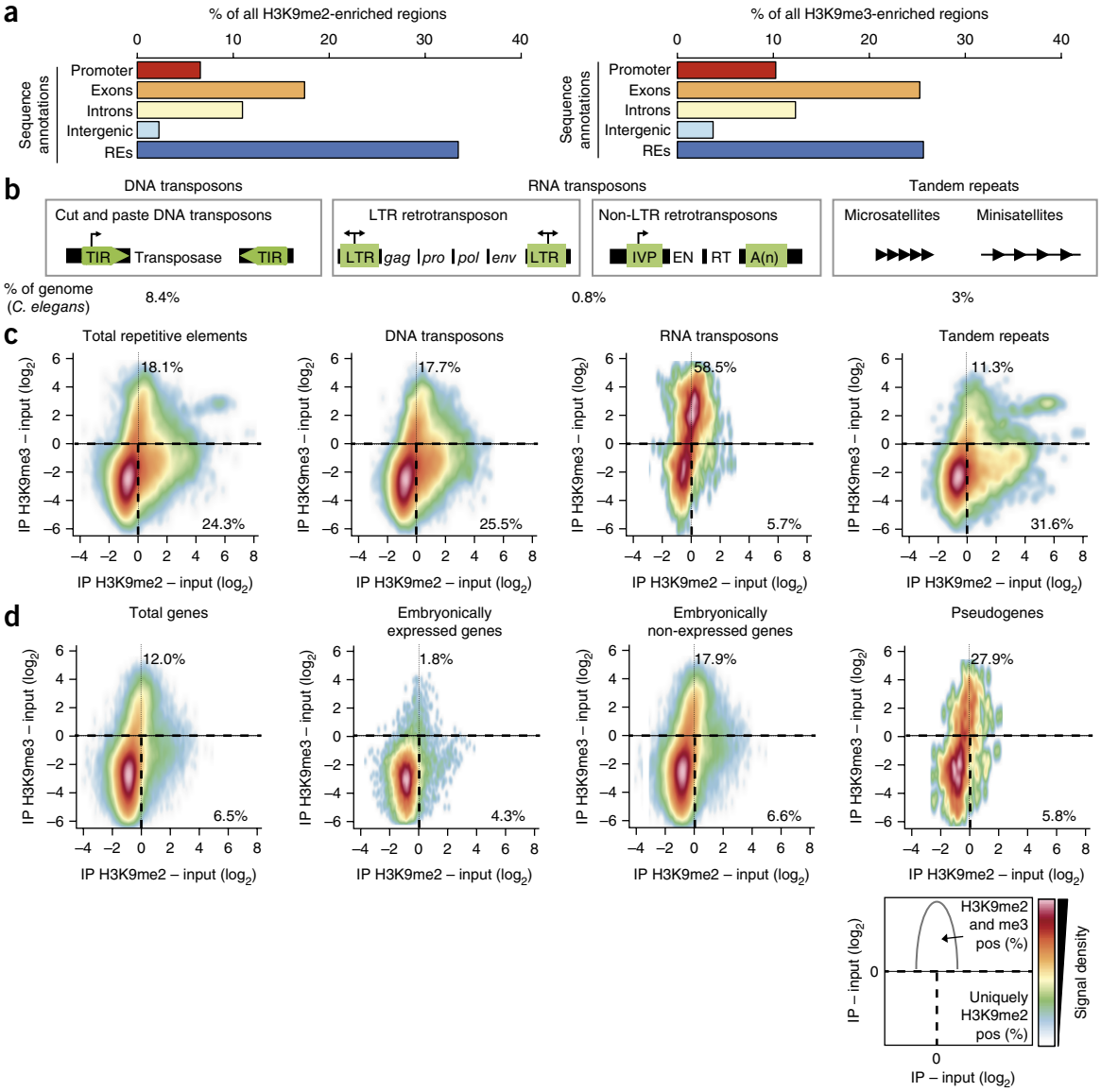


Figure 3 Differential enrichment of H3K9me2 and H3K9me3 on repeat element classes and gene types. (**a**) Percentage of H3K9me2 and H3K9me3 domains covering promoters, exons, introns, unique intergenic sequences or REs ($N = 2$). H3K9me positive regions were determined from genomic bins of sequences recovered after CHIP-seq using H3K9me2- or H3K9me3-specific antibodies with IP/input > 0. (**b**) Schematic representation of the three major repeat classes. DNA transposons encode a single transposase, which catalyzes all the steps of transposition, flanked by two terminal inverted repeats (TIRs). RNA transposons are either long terminal repeat (LTR) or non-LTR retrotransposon types. As derivatives of ancient retrovirus infections LTR retrotransposons encode *gag* (structural proteins of the virus core), *pol* (reverse transcriptase, integrase), *pro* (protease) and *env* (envelope). Non-LTR transposons encode a reverse transcriptase (RT) and an endonuclease (EN). Retrotransposon flanking regions in both cases supply promoter elements. Tandem repeats are short, noncoding sequence stretches that are repeated in a head-to-tail fashion. (**c**) High-density scatterplots show the enrichment of H3K9me2 and H3K9me3 on REs based on CHIP-seq data. IP, immunoprecipitation. RNA transposons were heavily enriched for H3K9me3 (58.5%), whereas 31.6% of tandem repeats had only H3K9me2. Lines indicate the quadrants of single-positive, double-positive and double-negative elements. (**d**) High-density scatterplots of the H3K9me2 and H3K9me3 enrichment on genes. Nonexpressed genes and pseudogenes were enriched for H3K9me3.

Supplementary Fig. 3b), and it was depleted from active genes (1.8%; **Fig. 3d**). Among the H3K9me3-bound genes were many that were expressed only in terminally differentiated tissues and a large fraction of pseudogenes (**Fig. 3d** and **Supplementary Fig. 3c–e**).

Loss of H3K9me led to the derepression of genes and REs

To determine whether loss of H3K9me affects transcription, we performed RNA-seq on RNA isolated from either gonads or early embryos of wild-type and *met-2 set-25* strains, grown at either 20 °C or 25 °C. In embryos cultured at 20 °C, we observed the reproducible derepression (>2-fold compared to wild-type) of 308 genes. Of these 72.2% (234) were marked by H3K9me in wild-type cells, and are therefore likely to be regulated directly by MET-2 and/or SET-25 (**Fig. 4a** and **Supplementary Fig. 4a**). This set of derepressed genes was only a subset (~9.7%) of all genes bearing H3K9me, arguing that the loss of H3K9me is not always sufficient to activate transcription. Derepression of genes was also temperature-sensitive, with 2.2-fold more genes being upregulated at 25 °C, including 83.8% of those

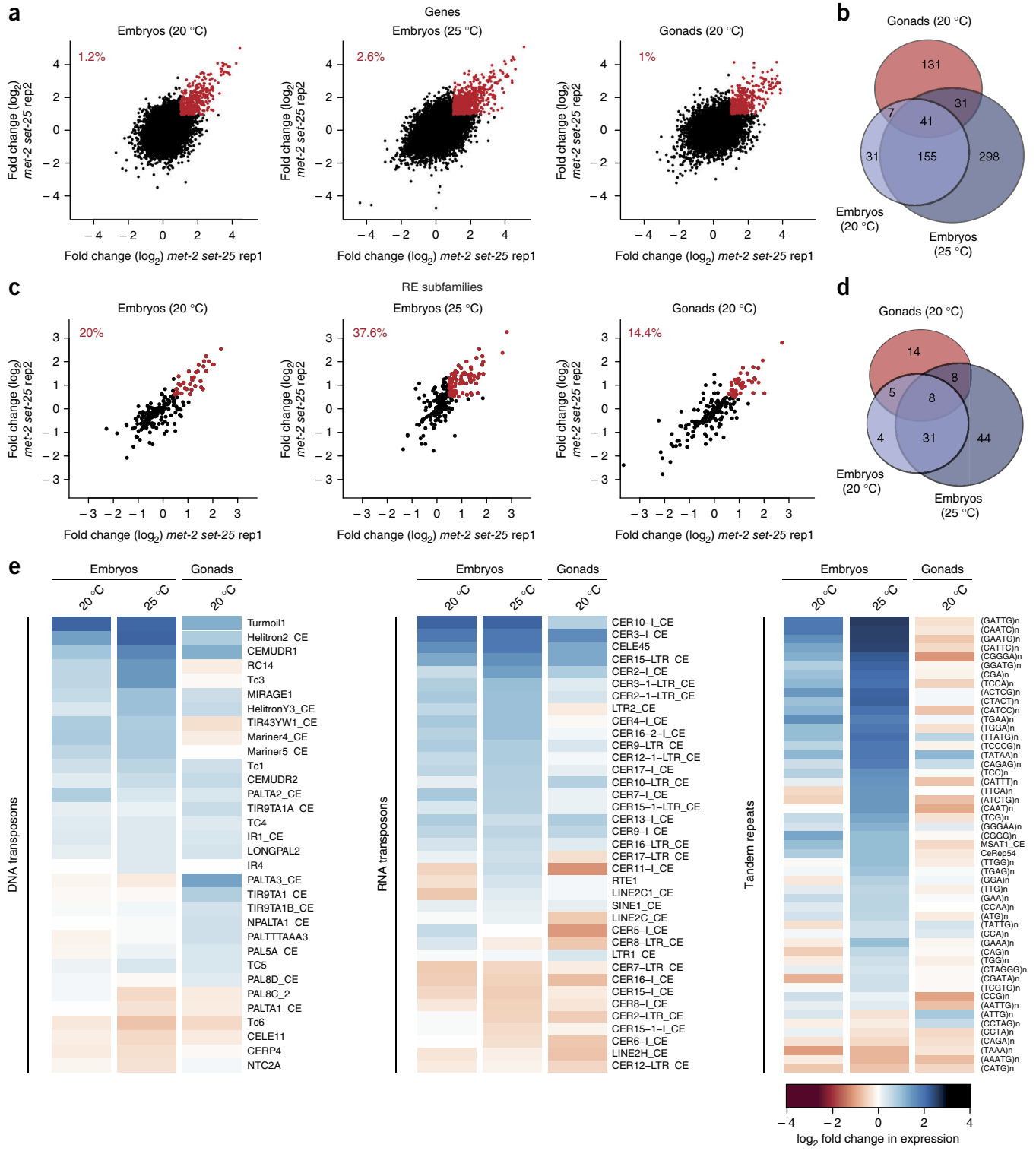


Figure 4 Temperature-dependent derepression of subsets of genes and repeat families in embryos and gonads in *met-2 set-25* worms. (**a**) Fold change ($\log_2$, *met-2 set-25*/wt) in gene expression of two replicas (rep1 and rep2) of RNA-seq data from embryos (20 °C and 25 °C) and from isolated gonads (20 °C) for each strain. The genes marked in red were consistently >2-fold upregulated ($P < 0.05$, FDR < 0.1), and % of total genes is indicated. (**b**) Venn diagram of the derepressed genes shows that genes affected in gonads were distinct from those upregulated in early embryos, and that derepression was temperature-enhanced in embryos ($N = 3$). (**c**) Scatterplot of the expression changes of H3K9me-enriched RE subfamilies in *met-2 set-25* embryos and gonads compared to wt. The REs marked in red were >1.5-fold derepressed in both replicas ($P < 0.05$, FDR < 0.1). (**d**) Venn diagram of the derepressed REs shows that subfamilies affected in gonads were partially distinct from subfamilies derepressed in embryos, and that RE upregulation was temperature-enhanced in embryos ($N = 3$). (**e**) Heat map of the expression changes in all significantly affected RE subfamilies ($P < 0.05$, FDR < 0.05) in embryos (20 °C and 25 °C) and gonads, sorted by repeat class ($N = 3$).

already derepressed at 20 °C (**Fig. 4a**,**b**). Transcription in gonads was elevated by the loss of H3K9me (210 genes). The affected genes were largely distinct from those derepressed in somatic cells (37.6% overlap; **Fig. 4a**,**b**), arguing that transcription factor availability is critical for transcriptional activity in the absence of repressive chromatin. No essential regulators of meiosis were misregulated.

Given that REs were enriched for H3K9me in wild-type worms, we next examined expression changes for REs, which we analyzed as

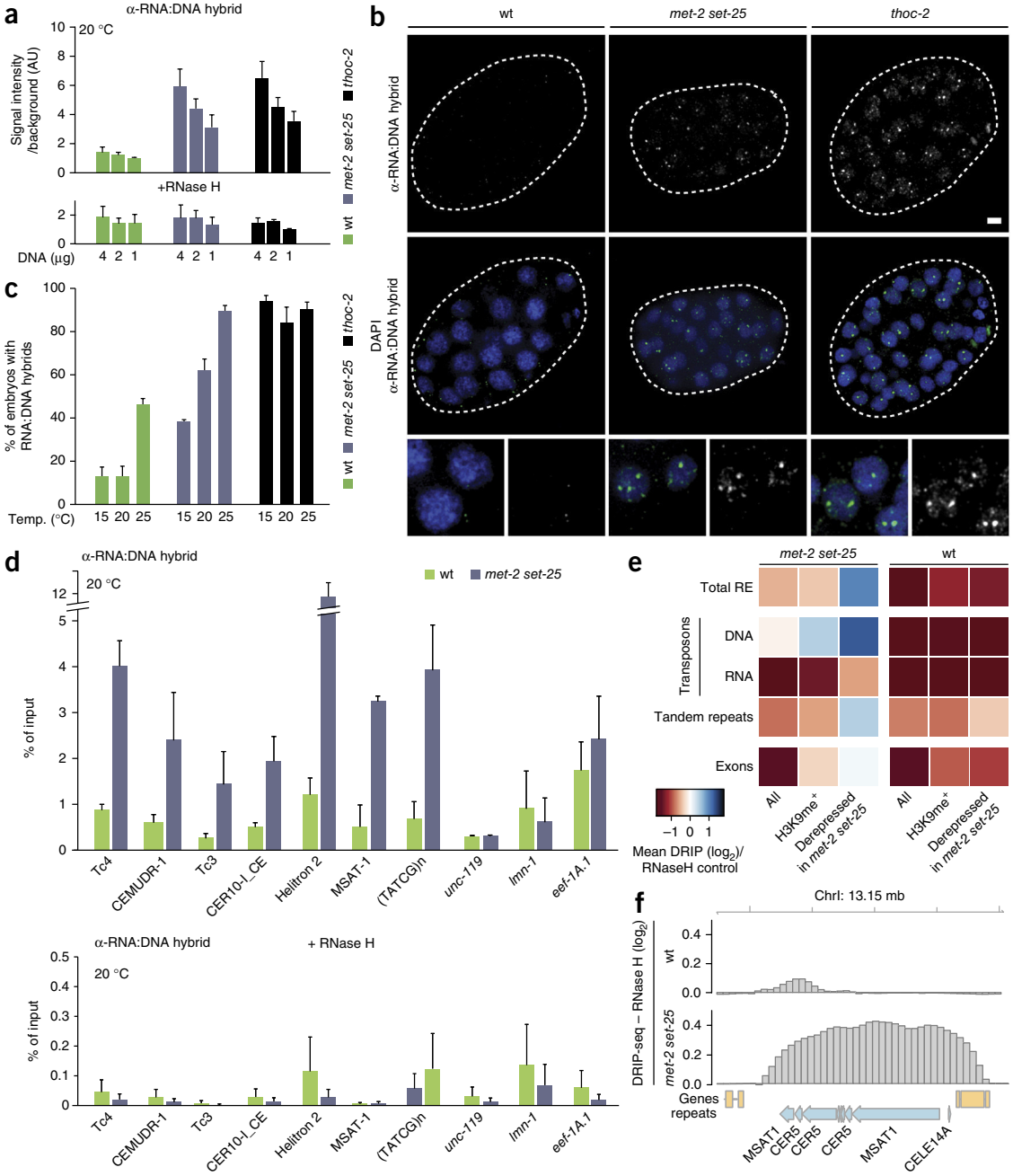


Figure 5 *met-2 set-25* worms accumulate RNA:DNA hybrids at repeat elements. (**a**) Quantification of multiple dot blots against RNA:DNA hybrids (antibody S9.6, HB-8730, ATCC, $n = 3$) in genomic DNA isolated from gravid adults of wt, *met-2 set-25* and *thoc-2* strains grown at 20 °C. 4 µg, 2 µg and 1 µg of nucleic acids were loaded for each strain. Where indicated, genomic DNA was treated with RNase H before blotting (mean + s.e.m.; $N = 3$). (**b**) Immunofluorescence (IF) images of isogenic wt, *met-2 set-25 and thoc-2* mutant embryos grown at 20 °C, stained with antibody S9.6 to visualize RNA:DNA hybrids (green); DAPI is in blue. Scale bar, 5 µm. (**c**) Quantification of IF signals after S9.6 staining from embryos of indicated strains, grown at 15 °C, 20 °C or 25 °C (mean + s.e.m.; $N = 3$, $n = 15$). (**d**,**e**) Genome-wide distribution of R loops determined by DRIP with antibody S9.6, followed by qPCR or deep sequencing of recovered DNA. DRIP-qPCR (**d**) for seven repeat subfamilies upregulated in *met-2 set-25* worms and for three control loci (*unc-119, lmn-1* and *eef-1A.1*), that were not upregulated (mean and s.e.m.; $N = 2$). Heat map of an S9.6 DRIP-seq experiment (**e**) showing mean $\log_2$ enrichment over the corresponding controls treated with RNase H (samples normalized to total number of reads). Loci were segregated based on indicated sequence criteria, and were further subgrouped based on the presence of H3K9me and response to the *met-2 set-25* mutations ($N = 1$). (**f**) DRIP-seq example showing the R-loop signal over a RE cluster. The IP signal was normalized to the input and the RNase H control values were subtracted.

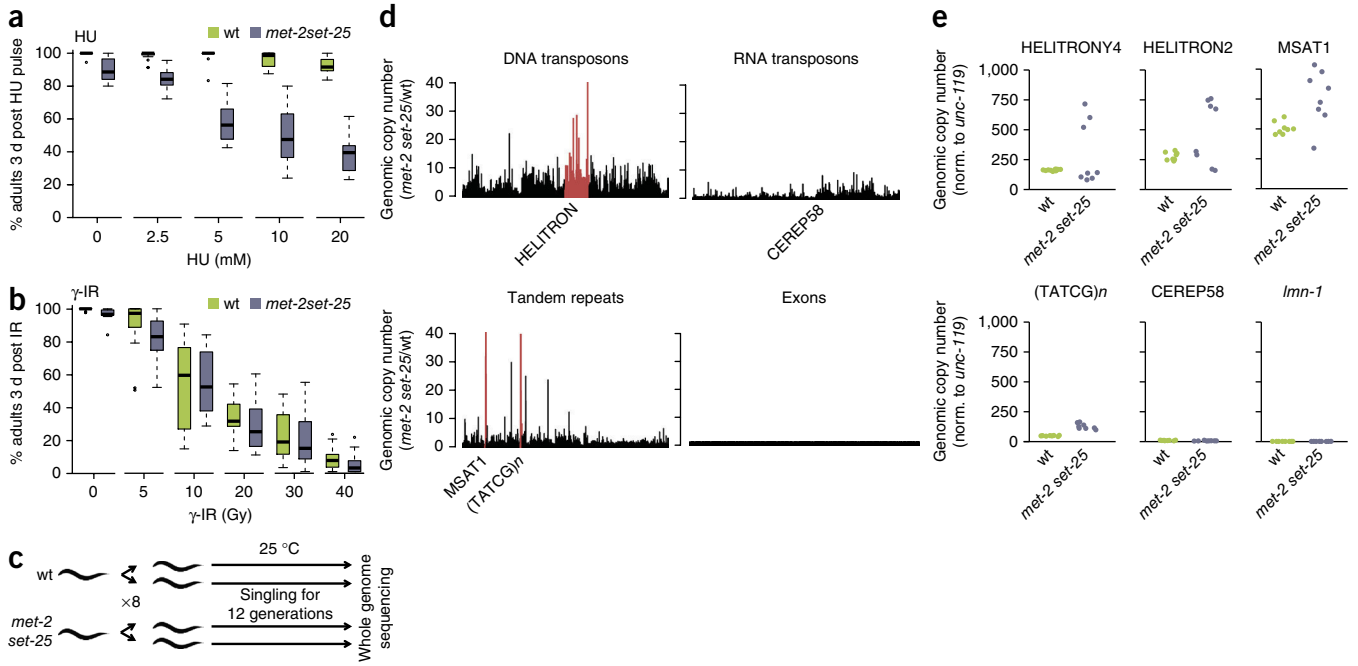


Figure 6 The *met-2 set-25* strain is hydroxyurea sensitive and accumulates mutations in repeat elements and reactivated transposable elements. (**a**) Worms lacking MET-2 and SET-25 were hypersensitive to hydroxyurea (HU). Synchronized populations of wt and mutant L1 larvae were exposed to 20 mM HU for 16 h, and then the numbers of worms that develop into adults after 3 d were quantified. Statistical analysis shows significant differential loss of viability between *met-2 set-25* and wt worms (Wilcoxon test $P < 0.0001$ at all doses; $N = 6$, $n = 25$). (**b**) Synchronized populations of wt and mutant L1 larvae were exposed to sublethal doses of gamma irradiation (γ-IR), and the numbers of worms that developed into adults after 3 d were quantified. Only at 0 Gy was there enhanced *met-2 set-25* lethality (Wilcoxon test $P < 0.005$; $N = 3$, $n = 15$). Boxplots show median, 50% boxes and 90% whiskers. (**c**) Wt and *met-2 set-25* worms were singled for 12 generations at 25 °C, generating 8 substrains per genotype. Genomic DNA of each substrain was sequenced, and only mutations unique to one of the substrains were considered. (**d**) Copy number of REs and exons in *met-2 set-25* determined by genome sequencing and sorted according to repeat class. REs analyzed by qPCR are indicated in red. (**e**) Analysis of the copy number of REs by qPCR from 8 *met-2 set-25* and 8 wt substrains cultured as described in **c**, shows an increase in CNV of DNA transposons, simple repeats but not RNA transposons or single-copy genes.

subfamilies. We characterized ~84% of all annotated repeats (300 subfamilies), and excluded only very-low-complexity repeat sequences or elements with a single annotated occurrence. In *met-2 set-25* mutant embryos at 20 °C, 20% of the H3K9me-enriched repeat subfamilies were derepressed by at least 1.5-fold, and at 25 °C this value increased to 37.6% (**Fig. 4c** and **Supplementary Fig. 4b**). Gonads isolated from double mutant adults showed an increase of transcription in 14.4% of all H3K9me repeat subfamilies (**Fig. 4c**). This lower number of derepressed REs may reflect germline-specific redundant silencing by the PIWI pathway[38,39]. Indeed, different REs were upregulated in gonads and somatic cells (**Fig. 4d** and **Supplementary Fig. 4c**), with tandem repeats being distinctly underrepresented in the germ line (**Fig. 4e**). We note that each class of repeats includes REs that were not derepressed by loss of H3K9me, which may reflect either the existence of other, H3K9me-independent silencing pathways, or a requirement for transcription factors that are tissue-specific or developmental-stage-specific.

We asked whether the transcriptional landscape of genes surrounding a RE might influence its expression upon loss of H3K9me. This is particularly relevant for simple tandem repeats, which lack recognizable promoter or enhancer sequences[40]. To our surprise, ~50% of the derepressed tandem repeats were not in the proximity of an upregulated gene (data not shown).

H3K9me-deficient worms accumulated R loops

We next examined the relationship between aberrant RE transcription and the observed DNA damage. Perturbation of the replication fork is a major driver of DNA lesions[41], and a substantial obstacle for its progression is the transcription machinery, in particular when stalled by RNA:DNA hybrids (R loops)[42–45]. In fission yeast, R loops are enriched at repetitive sequences, such as transposons, telomeres or the rDNA[46], and correlated with genetic instability[47,48]. We therefore checked whether the *met-2 set-25* double mutant accumulated R loops, using multiple approaches based on an antibody specific for RNA:DNA hybrids (S9.6, gift of P. Pasero[49]).

We detected an accumulation of R loops in *met-2 set-25* worms that was not detectable in wild-type worm DNA by performing a dot blot analysis of genomic DNA. We also detected significant R-loop occurrence by immunostaining of mutant, but not wild-type, embryos ($P < 0.001$, Student`s t-test; **Fig. 5a–c** and **Supplementary Fig. 5a**). The level was roughly similar to that scored in a mutant strain deficient for the Tho-Trex complex (*thoc-2*), in which RNA:DNA hybrids accumulate owing to impaired RNA processing and export (**Fig. 5a**)[50,51]. To test for antibody specificity, we treated the isolated DNA with RNase H before blotting, to specifically degrade RNA:DNA heteroduplexes. Quantification showed that 60% of the signal (*met-2 set-25*, loading 4 µg; **Fig. 5a**) was lost after treatment with RNase H. Consistent with the elevated level of RE transcription at higher temperatures, the level of R loops increased with temperature, both in the dot blot analysis of adult worm DNA, as well as in the immunostaining of embryos (**Fig. 5c** and **Supplementary Fig. 5a**). The *thoc-2* mutant, on the other hand, reached R-loop saturation even at 15 °C.

To examine formation of R loops in a sequence-dependent manner, we immunoprecipitated RNA:DNA hybrids from wild-type or *met-2*

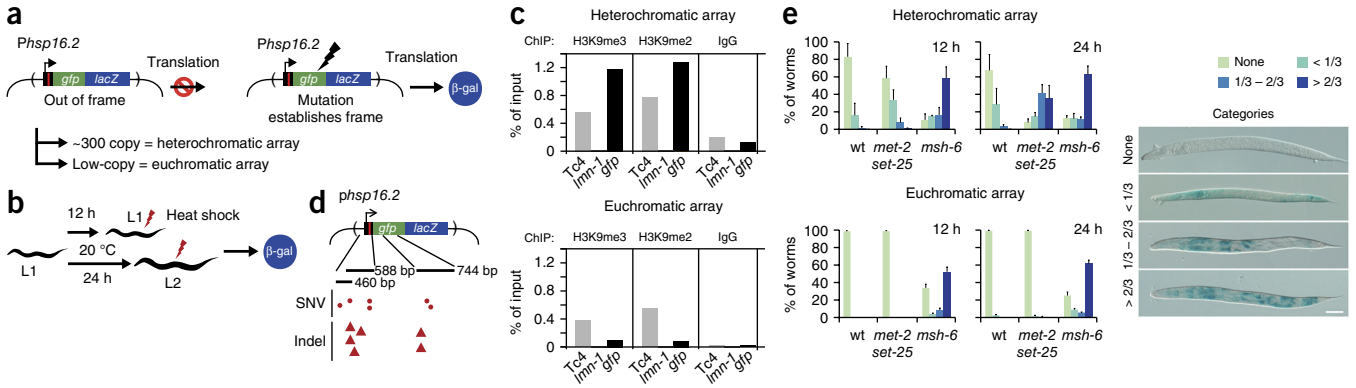


Figure 7 Somatic accumulation of indels leading to frameshift mutations in *met-2 set-25* mutant larvae. (**a**) Schematic of reporter to monitor mutation frequency in single cells in two different chromatin contexts. A *lacZ* construct containing a frame-shift mutation under the control of a heat shock promoter was integrated either as a high copy (~200–300 copies), or low-copy array (~20 copies). The frameshift prevents *lacZ* translation, which can be reestablished by mutation. (**b**) To quantify the accumulation of mutations, L1 larvae were released into development for 12 h or 24 h at 20 °C before a heat shock and subsequent β-gal staining. (**c**) ChIP-qPCR monitored enrichment of H3K9me2/H3K9me3 on the reporter array by PCR for *gfp*. H3K9me was recovered on the heterochromatic (high-copy) array but not the euchromatic (low-copy) array. The genomic copy of *lmn-1* and *Tc4* served as negative and positive controls. (**d**) Genomic DNA of the heterochromatic array was isolated from either *met-2 set-25*, or wt worms grown for 24 h. Indicated fragments were PCR amplified, subcloned and sequenced by Sanger sequencing. Indels and SNVs that restore the ORF are indicated by triangles and dots, respectively in indicated fragments of the construct that were sequenced ($N = 3$, $n = 50$). (**e**) High frequency of LacZ frameshift mutations was recovered in the heterochromatic reporter in *met-2 set-25* and *msh-6* worms, but not in the euchromatic reporter. Results were categorized according to the proportion of β-gal positive cells per worm (mean and s.e.m.; $N = 3$, $n = 50$). Scale bar, 50 µm.

set-25 embryos followed by deep sequencing (DRIP-seq) or qPCR (DRIP-qPCR). By qPCR, we found that specific repeat elements that were derepressed in the absence of H3K9me, were enriched for R loops fourfold to ninefold in mutant over wild-type strains. This was not the case for low- or moderate-level transcribed genes (*unc-119* or *lmn-1)*, nor was there a *met-2 set-25*-dependent increase in DRIP for a highly transcribed gene (*eef-A.1*), although the levels of R loops did increase at highly transcribed genes in both wild-type and *met-2 set-25* strains (**Fig. 5d**). As proof that the antibody was specific for RNA:DNA hybrids, we note that the DRIP-qPCR signal was highly sensitive to treatment with RNase H (**Fig. 5d**).

On a genome-wide level (DRIP-seq), we detected the most pronounced enrichment of RNA:DNA hybrids in *met-2 set-25* embryos on REs that were derepressed in the double mutant (**Fig. 5e**,**f** and **Supplementary Fig. 5e**). RNA:DNA hybrids were particularly enriched on transcribed DNA transposons and tandem repeats but not on RNA transposons (**Fig. 5e**). Confirming R-loop mapping in other organisms, we observed RNA:DNA hybrids more frequently on highly transcribed genes, telomeres and the rDNA locus, even in wild-type cells (**Supplementary Fig. 5b–d**)[52,53], yet these signals showed no further increase in the *met-2 set-25* mutant.

This high level of RNA:DNA hybrids suggests the presence of replication stress in *met-2 set-25* worms. To monitor their sensitivity to fork stalling, we exposed worms to hydroxyurea, a DNA replication inhibitor that reversibly inhibits ribonucleotide reductase, thereby depleting deoxynucleotide pools and exacerbating replication fork stalling[54]. L1 larvae exposed to 20 mM hydroxyurea for 16 h and allowed to recover for 3 d in absence of the inhibitor, yielded 95 ± 3% (mean ± s.d.) viability (resumption of development), whereas only 43 ± 11% of the *met-2 set-25* larvae survived hydroxyurea exposure (**Fig. 6a**). This hypersensitivity was specific to agents causing replication stress, as treating similarly staged larvae with ionizing radiation did not differentially affect wild-type and *met-2 set-25* strains (**Fig. 6b**). Thus hydroxyurea hypersensitivity correlated with the accumulation of R loops, and suggests that both the developmental delays and sterility detected in H3K9me-deficient worms reflect collisions of replication with unscheduled transcription.

In the absence of H3K9me, mutations accumulated in REs

Replication stress and formation of R loops have been correlated with both fork instability and double-strand break hotspots in yeast[44,46,55,56]. To determine whether genomes of H3K9me-deficient worms accumulate mutations at elevated rates, we singled 8 wild-type and 8 *met-2 set-25* worms for 12 generations at 25 °C, thereby creating 8 individual substrains per genotype. Sequencing of the genome of each substrain revealed mutations exclusively in one of the 16 genomes (**Fig. 6c**). This allowed us to score the number, nature and location of changes accumulated owing to the *met-2 set-25* mutation.

We note that the rate and nature of single nucleotide variants (SNVs) did not differ between wild-type and *met-2 set-25* worms, which allowed us to exclude generation time as a confounding factor in the analysis (**Supplementary Fig. 6a**). However, 6 of the 8 *met-2 set-25* sub-strains acquired at least one insertion or deletion (indel) (with a total of 9 different observed indels; **Supplementary Table 1**). In contrast, only one wild-type strain incurred small deletions (3-base pair (bp) and 5-bp). The average indel in the *met-2 set-25* substrains covered 5.3 kilobases (kb) (the largest being 33.5 kb), all *met-2 set-25* indels occurred at sites enriched for H3K9me3, and 8 of the 9 *met-2 set-25* indels occurred in REs whose majority showed enhanced transcription upon loss of H3K9me.

Confirming the existence of large and stable germline changes, we detected a 10-kb inversion flanked on one side by a 1-kb deletion by whole genome sequencing and PCR of *met-2 set-25* worms that had been cultivated for several months. The inversion was immediately adjacent to an excised *Tc3* transposon, and opposite the inversion was a *de novo Tc3* transposon insertion unique to the cultivated H3K9me-deficient strain (**Supplementary Fig. 6b–d**). The excised *Tc3* element carried H3K9me3 in the wild-type strain, and was transcriptionally activated in *met-2 set-25*.

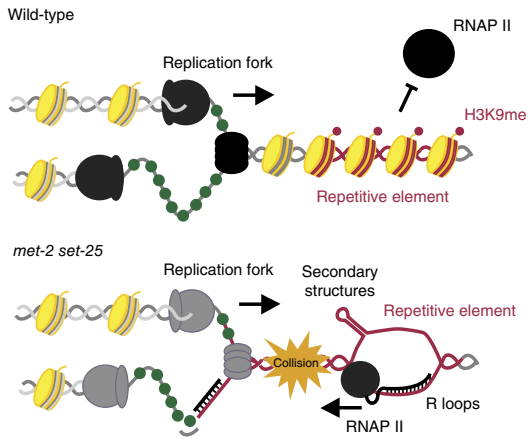


Figure 8 Transcribed REs in H3K9me-deficient strains can exacerbate replication stress provoking genomic instability. A model illustrates how the loss of H3K9me could lead to the formation of secondary DNA structures that engender replication stress specifically at heterochromatic repeats, to perturb genome integrity.

We next checked wild-type and *met-2 set-25* genomes for copy number variations (CNVs) in repeat families with multiple members. Two DNA transposons (HELITRON2 and HELITRONY4) and two tandem repeats (MSAT1 and $(TATCG)_n$) showed high CNV uniquely in the *met-2 set-25* substrains. In contrast, the RNA transposon CEREP58 and the single-copy gene *lmn-1*, like telomeric repeats and the rDNA, remained stable (**Fig. 6d**,**e** and **Supplementary Fig. 6e**,**f**)[57]. RNA transposons, which also failed to accumulate R loops, did not show CNV. We conclude that *met-2 set-25* germ lines accumulated indels at sites bearing H3K9me in wild-type strains as well as changes in DNA-transposon and tandem-repeat copy number.

A reporter incurred frequent indels in H3K9me-deficient somatic cells

This sequence analysis monitored stable germline changes in the worm population, and selected against any mutation that would perturb meiotic genome transmission. To visualize the mutation rate in somatic cells, we used a heterochromatic reporter with a *lacZ* gene placed out of frame to the ATG start codon, generating multiple premature stop codons in the first 100 bp of the transcript. Insertions or deletions between the ATG and the ORF are necessary to enable the translation of the *lacZ* mRNA into a functional β-galactosidase enzyme (**Fig. 7a**)[58]. This allowed us to compare mutation rates of wild-type and *met-2 set-25* worms by microscopy, following a colorimetric stain for heat-shock-induced β-galactosidase expression. By comparing two time points during somatic development (12 h and 24 h after L1) we could differentiate mutations that might have been present in the fertilized egg from mutations incurred during somatic development (**Fig. 7b**). To compare the mutation rate of repetitive heterochromatic and unique euchromatic sequences, we made use of the observation that transgenes integrated as high–copy number arrays induce the formation of H3K9me-containing heterochromatin[59] (i.e., enriched for H3K9me2 and H3K9me3; **Fig. 7c**). We compared this reporter with the same reporter construct integrated as a low–copy number array, which remains unmethylated and euchromatic. We classified phenotypes by the extent of β-galactosidase expression on a worm-by-worm basis, and sequenced the constructs amplified from worms at the 24-h time point (**Fig. 7d**,**e**).

In the wild-type background after 24 h of cultivation, the heterochromatic reporter produced functional β-galactosidase in only around 3 ± 2% (s.e.m.) of the worms (**Fig. 7e**; >1/3 expressing in-frame *lacZ*). In the *met-2 set-25* mutant, the fraction of worms expressing in-frame *lacZ* increased to 78 ± 8% (>1/3 staining blue, $P = 0.01$). In contrast, the euchromatic reporter did not express in-frame *lacZ* in *met-2 set-25* worms (1 ± 1%, any level of in-frame *lacZ*). We used a mutant of the mismatch repair machinery *msh-6* as a positive control[60]. The *met-2 set-25* mutant primarily showed an increase in the β-galactosidase-positive phenotype by 24 h, and not by 12 h, unlike the *msh-6* mutant (**Fig. 7e**), suggesting that the *met-2 set-25*-induced mutations occurred during differentiation. The types of mutations monitored were confirmed by batch-wise cloning of single reporter units and Sanger sequencing. We indeed detected small insertions and deletions in the *met-2 set-25* worms, enabling in-frame translation of β-galactosidase (**Fig. 7d**). Thus, like the germline changes scored by genome sequencing, sequences with H3K9me in wild-type backgrounds accumulated indels at high rates during somatic cell division in H3K9me-deficient worms.

DISCUSSION

H3K9 methylation is the defining histone modification for heritably silent chromatin and is conserved as such from fission yeast to humans. *C. elegans* mutants lacking H3K9me are viable, despite the enrichment of H3K9me2/H3K9me3 on silent tissue-specific genes, on pseudogenes and on RE. In contrast to the case in other species[18,19,61], we found no defects in chromosome segregation upon loss of H3K9me. However, we observed a temperature-dependent sterility, which coincided with an increase in DNA-damage-induced apoptosis and stochastic delays in development. Correlating with these phenotypes, we detected derepression of ~20% RE, from all repeat classes, a value that increased at elevated temperatures. Expression of these RE was accompanied by the accumulation of RNA:DNA hybrids, CNV and a hypersensitivity to replication stress. This correlation suggests that it is either the transcription of the repetitive sequence alone, or transcription coupled with the inherent pairing nature of repeats, that generates insertions and deletions within REs in the absence of H3K9me. Of note, DNA transposons and tandem repeats showed higher levels of R loops and CNV than RNA transposons, although all classes were derepressed upon loss of H3K9me.

The damage incurred in the germ line leads to extensive apoptosis and Rad51 focus accumulation, suggesting that these cells accumulated double-strand breaks as well as indels. There may be additional sources of damage in the germ line, other than those that correlate with replication-fork-associated damage, R loops and indels scored by genome sequencing. We note that RNA polymerase-DNA polymerase collision has been reported to generate fragile sites of breakage[44,62], which in worm germline cells would provoke an apoptotic response[32,63].

It is likely that the genomic mutations we detected in the *met-2 set-25* strain arise from replication fork perturbation. This can be triggered by enhanced stalling of the replication fork generated by R-loop formation, which in turn allows hairpin or fold-back structures to form in repeats as they are being replicated. Hairpin or fold-back structures can also arise from breaks in the single-stranded DNA that accumulate either at R loops or behind the fork, owing to perturbed coordination between leading- and lagging-strand polymerases (**Fig. 8**). The passage of the replication fork through REs itself can lead to hairpin structures[41]. However, we propose that in *met-2 set-25* cells, unprogrammed transcription of REs enhances R loops, which may in turn enhance aberrant structures to such a degree that the cellular machineries that normally relieve such stress, can no longer cope with their abundance.

Unscheduled collisions of the replication and transcription machineries appear to generate breaks as well as other forms of genome instability[42–46,62]. Damage is often attributed to the presence of RNA:DNA hybrids[64–67], yet torsional stress, which can arise from high levels of bidirectional transcription[68,69], may also contribute to genomic instability. We consider it notable not only that the derepression of RE generated genomic mutations and R loops, but that both these events mapped to REs that are normally marked by H3K9me in wild-type cells, and which became derepressed in a temperature-enhanced manner in the *met-2 set-25* mutant. We propose that the crucial role of H3K9 methylation in suppressing transcription on a genome-wide level is not to program cell differentiation, but to stabilize repetitive sequences that accumulate in higher eukaryotic genomes.

Several studies have suggested the use of inhibitors for H3K9me HMTs in the treatment of cancer (for example, lung, prostate, hepatocellular and pancreatic cancer)[19,61], and preclinical studies have been considered promising so far[70]. These same inhibitors have been used to show that hypomethylation of H3K9 increases the rate of induced pluripotent stem cell generation[15,16]. We argue that there are clear drawbacks to such therapies, given the genomic instability provoked by loss of H3K9me shown here. Whereas mammals additionally silence through meCpG, it has been documented that DNA methylation can be targeted by H3K9me or its HMTs[71]. Thus, the findings presented in this study are likely to have implications for protocols that attempt to manipulate the mammalian epigenome.

URLs. http://www.bioconductor.org/packages/3.1/bioc/html/QuasR.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. All data from this study have been deposited in the Sequence Read Archive (SRA) under accession SRP080806.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

A number of strains were provided by the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We thank R. Ciosk and P. Pasero for reagents and materials, I. Katiç and members of the Friedrich Miescher Institute Genomics and Microscopy facilities for advice and discussion, and P. Ginno and L. Constantino for advice on R-loop detection. J.P. is supported by a long-term EMBO fellowship. S.M.G. thanks the Swiss National Science Foundation as well as the Novartis Research Foundation for support.

AUTHOR CONTRIBUTIONS

P.Z. and J.P. planned and executed most experiments, evaluated results and wrote the paper; S.M.G. planned experiments, evaluated results and wrote the paper; R.v.S. and M.T. helped with evaluation of genome mutations and provided the *LacZ* mutagenesis assay; and V.K. provided invaluable technical help.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Bannister, A.J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
2. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
3. Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D. & Grewal, S.I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
4. Allshire, R.C., Javerzat, J.P., Redhead, N.J. & Cranston, G. Position effect variegation at fission yeast centromeres. *Cell* **76**, 157–169 (1994).
5. Cam, H.P. *et al.* Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* **37**, 809–819 (2005).
6. Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
7. Peters, A.H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
8. Eissenberg, J.C. *et al.* Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position–effect variegation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **87**, 9923–9927 (1990).
9. Epsztejn-Litman, S. *et al. De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol.* **15**, 1176–1183 (2008).
10. Wissmann, M. *et al.* Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor–dependent gene expression. *Nat. Cell Biol.* **9**, 347–353 (2007).
11. Chen, M.W. *et al.* H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer Res.* **70**, 7830–7840 (2010).
12. Hua, K.T. *et al.* The H3K9 methyltransferase G9a is a marker of aggressive ovarian cancer that promotes peritoneal metastasis. *Mol. Cancer* **13**, 189 (2014).
13. Kumari, D. & Usdin, K. The distribution of repressive histone modifications on silenced *FMR1* alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* **19**, 4634–4642 (2010).
14. Lee, J. *et al.* Epigenetic regulation of cholinergic receptor M1 (CHRM1) by histone H3K9me3 impairs Ca^{2+} signaling in Huntington's disease. *Acta Neuropathol.* **125**, 727–739 (2013).
15. Loh, Y.H., Zhang, W., Chen, X., George, J. & Ng, H.H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21**, 2545–2557 (2007).
16. Shi, Y. *et al.* A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525–528 (2008).
17. Treangen, T.J. & Salzberg, S.L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
18. Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H. & Li, E. Histone H3-K9 methyltransferase ESET is essential for early development. *Mol. Cell. Biol.* **24**, 2478–2486 (2004).
19. Kondo, Y. *et al.* Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. *PLoS One* **3**, e2037 (2008).
20. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
21. Mellone, B.G. *et al.* Centromere silencing and function in fission yeast is governed by the amino terminus of histone H3. *Curr. Biol.* **13**, 1748–1757 (2003).
22. Peng, J.C. & Karpen, G.H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat. Cell Biol.* **9**, 25–35 (2007).
23. Andersen, E.C. & Horvitz, H.R. Two *C. elegans* histone methyltransferases repress *lin-3* EGF transcription to inhibit vulval development. *Development* **134**, 2991–2999 (2007).
24. Towbin, B.D. *et al.* Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
25. Harris, J. *et al.* Mutator phenotype of *Caenorhabditis elegans* DNA damage checkpoint mutants. *Genetics* **174**, 601–616 (2006).
26. Bernard, P. *et al.* Requirement of heterochromatin for cohesion at centromeres. *Science* **294**, 2539–2542 (2001).
27. Batista, P.J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78 (2008).
28. Ketting, R.F., Haverkamp, T.H., van Luenen, H.G. & Plasterk, R.H. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
29. Sijen, T. & Plasterk, R.H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
30. Zhou, Z., Hartwieg, E. & Horvitz, H.R. CED-1 is a transmembrane receptor that mediates cell corpse engulfment in *C. elegans*. *Cell* **104**, 43–56 (2001).
31. Greiss, S., Schumacher, B., Grandien, K., Rothblatt, J. & Gartner, A. Transcriptional profiling in *C. elegans* suggests DNA damage dependent apoptosis as an ancient function of the p53 family. *BMC Genomics* **9**, 334 (2008).
32. Gartner, A., Milstein, S., Ahmed, S., Hodgkin, J. & Hengartner, M.O. A conserved checkpoint pathway mediates DNA damage–induced apoptosis and cell cycle arrest in *C. elegans*. *Mol. Cell* **5**, 435–443 (2000).
33. Schumacher, B., Hofmann, K., Boulton, S. & Gartner, A. The *C. elegans* homolog of the p53 tumor suppressor is required for DNA damage–induced apoptosis. *Curr. Biol.* **11**, 1722–1727 (2001).
34. Gerstein, M.B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
35. Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **21**, 227–236 (2011).

36. Cangiano, G. & La Volpe, A. Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. *Nucleic Acids Res.* **21**, 1133–1139 (1993).
37. Padeken, J., Zeller, P. & Gasser, S.M. Repeat DNA in genome organization and stability. *Curr. Opin. Genet. Dev.* **31**, 12–19 (2015).
38. Das, P.P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol. Cell* **31**, 79–90 (2008).
39. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
40. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
41. Aguilera, A. & García-Muse, T. Causes of genome instability. *Annu. Rev. Genet.* **47**, 1–32 (2013).
42. Azvolinsky, A., Giresi, P.G., Lieb, J.D. & Zakian, V.A. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell* **34**, 722–734 (2009).
43. Deshpande, A.M. & Newlon, C.S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030–1033 (1996).
44. Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R. & Feng, W. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Res.* **25**, 402–412 (2015).
45. Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D. & Soultanas, P. Co-directional replication–transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).
46. Wahba, L., Amon, J.D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
47. Nakamori, M., Pearson, C.E. & Thornton, C.A. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum. Mol. Genet.* **20**, 580–588 (2011).
48. Prak, E.T. & Kazazian, H.H. Jr. Mobile elements and the human genome. *Nat. Rev. Genet.* **1**, 134–144 (2000).
49. Boguslawski, S.J. *et al.* Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
50. Castellano-Pozo, M., García-Muse, T. & Aguilera, A. The *Caenorhabditis elegans* THO complex is required for the mitotic cell cycle and development. *PLoS One* **7**, e52447 (2012).
51. Castellano-Pozo, M. *et al.* R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583–590 (2013).
52. Balk, B. *et al.* Telomeric RNA–DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* **20**, 1199–1205 (2013).
53. Wahba, L., Costantino, L., Tan, F.J., Zimmer, A. & Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* **30**, 1327–1338 (2016).
54. Rosenkranz, H.S. & Levy, J.A. Hydroxyurea: a specific inhibitor of deoxyribonucleic acid synthesis. *Biochim. Biophys. Acta* **95**, 181–183 (1965).
55. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
56. Zhang, H. & Freudenreich, C.H. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol. Cell* **27**, 367–379 (2007).
57. Collins, J., Saari, B. & Anderson, P. Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. *Nature* **328**, 726–728 (1987).
58. Pothof, J. *et al.* Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi. *Genes Dev.* **17**, 443–448 (2003).
59. Towbin, B.D., Meister, P., Pike, B.L. & Gasser, S.M. Repetitive transgenes in *C. elegans* accumulate heterochromatic marks and are sequestered at the nuclear envelope in a copy-number- and lamin-dependent manner. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 555–565 (2010).
60. Tijsterman, M., Pothof, J. & Plasterk, R.H.A. Frequent germline mutations and somatic repeat instability in DNA mismatch-repair-deficient *Caenorhabditis elegans*. *Genetics* **161**, 651–660 (2002).
61. Wagner, T. & Jung, M. New lysine methyltransferase drug targets in cancer. *Nat. Biotechnol.* **30**, 622–623 (2012).
62. Helmrich, A., Ballarino, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).
63. Vermezovic, J., Stergiou, L., Hengartner, M.O. & d'Adda di Fagagna, F. Differential regulation of DNA damage response activation between somatic and germline cells in *Caenorhabditis elegans*. *Cell Death Differ.* **19**, 1847–1855 (2012).
64. Chan, Y.A., Hieter, P. & Stirling, P.C. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet.* **30**, 245–253 (2014).
65. El Hage, A., Webb, S., Kerr, A. & Tollervey, D. Genome-wide distribution of RNA–DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* **10**, e1004716 (2014).
66. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
67. Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
68. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.* **13**, 204–214 (2012).
69. Kouzine, F., Levens, D. & Baranello, L. DNA topology and transcription. *Nucleus* **5**, 195–202 (2014).
70. Yuan, Y. *et al.* Gossypol and an HMT G9a inhibitor act in synergy to induce cell death in pancreatic cancer cells. *Cell Death Dis.* **4**, e690 (2013).
71. Liu, S. *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28**, 2041–2055 (2014).

ONLINE METHODS

***C. elegans* cultures and strains. Supplementary Table 2** lists the strains used in this study. Strains were made by backcrossing deletion alleles and reporter strains obtained from the *C. elegans* knockout consortium to the GW638 strain (met-2(n4256) III; set-25(n5021) III) at least five times. Worms were grown at 20 °C, except where specifically indicated.

Immunofluorescence analysis, antibodies and live microscopy, including apoptosis assay. IF analysis was carried out as previously described[24] by freeze-cracking and fixation in 1% paraformaldehyde followed by short postfixation in methanol (for embryos and gonads[72]) or methanol followed by acetone (for larval stages). Staining was performed in PBS with 0.1% TritonX-100 and 2% milk powder. For live-cell imaging, larvae were mounted on slides coated with 2% agarose. Microscopy was carried out on a spinning disc confocal microscope (SD1, W1, Visitron, Puchheim). Stacks of images were analyzed using ImageJ.

Antibodies used in this study were mouse anti-H3K9me2, MABI0317 (MBL[73]), mouse anti-H3K9me3, MABI0318 (MBL[73]), mouse anti-RNA:DNA hybrid S9.6, hybridoma HB-8730 (ATCC)[49], rabbit anti-pan-acetyl H4, 06-866 (Merck Millipore) and rabbit anti-RAD-51, 29480002 (Novu Biologics).

Developmental timing, progeny size and hatching rate. Worms of indicated genotype were synchronized through bleaching and were then singled onto plates containing OP50 bacteria. For the developmental timing their stage was determined every 24 h. In order to determine the progeny size, adults were transferred to fresh plates once a day for three days to keep generations separate and their complete progeny size was determined after their hatching at the indicated temperature. To determine the hatching rate singled worms were transferred every 8 h to freshly seeded plates. The number of laid embryos was determined directly after transfer, the number of hatched animals was determined on day 3. If not otherwise indicated, worms were grown at the experimental temperature (transferred from 20 °C) for at least two generations before the experiments.

Chromatin immunoprecipitation experiments. Early embryonic progeny was harvested after synchronization (60–65 h depending on each strain) for wt and *met-2 set-25* mutant strains in two independent biological replicates. H3K9me2 and H3K9me3 ChIP was performed as previously described[74] using the antibodies mentioned above. In brief, 40 µg of chromatin was incubated overnight with 3–6 µg of antibody coupled to Dynabeads Sheep Anti-Mouse IgG (Invitrogen) or Dynabeads Sheep Anti-Rabbit IgG (Invitrogen), in FA buffer (50 mM HEPES/KOH pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 150 mM NaCl)) containing 1% SDS. Chromatin-antibody complexes were washed with the following buffers: 3 × 5 min FA buffer; 5 min FA buffer with 1 M NaCl; 10 min FA buffer with 500 mM NaCl; 5min with TEL buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) and twice for 5 min with TE. Complexes were eluted at 65 °C in 100 µl of elution buffer (1% SDS in TE with 250 mM NaCl) for 15 min. Both input and IP samples were incubated with 20 µg of RNAse A for 30 min at 37 °C and 20 µg of proteinase K for 1 h at 55 °C. Crosslinks were reversed overnight at 65 °C. DNA was purified using a Zymo DNA purification column (Zymo Research).

Library preparation and analysis. Libraries were prepared from chromatin IP and genomic DNA samples using the NEBNext ultra DNA library prep kit for Illumina (NEB # 7370) and the NEBNext Multiplex Oligos for Illumina (NEB # E7335), according to the manufacturer's recommendations. No size selection was performed during sample preparation and the libraries were indexed and amplified using 12 PCR cycles, using the recommended conditions. After further purification with Agencourt AmPure XP beads (Beckman # A63881), the library size distribution and concentrations were determined using a BioAnalyzer 2100 (Agilent technologies) and Qubit (Invitrogen) instrument, respectively. The final pools were prepared by mixing equimolar amounts of all individually indexed libraries and then sequenced on a HiSeq 2500 (Illumina) in rapid mode (Paired-End 50). Processing of the LEM-2 ChIP-seq data, all paired-end ChIP-seq data (2 × 50 bp) were mapped to the *C. elegans* genome (ce6) with the R package QuasR[75] using the included aligner bowtie[76]. Definitions of REs were taken from Repbase[77]. Repeat subfamilies were built to allow assignment of multimapping reads to all REs and collapsing single elements according to their Repbase ID into families.

Read density along the genome was calculated by tiling the genome into 200-bp windows (non-overlapping) and counting the number of sequence fragments within each window, using the qCount function of the QuasR package (see URLs). To compensate for differences in the read depths of the various libraries, we divided each sample by the total number of mapped reads and multiplied by the average library size. Log_2 expression levels were calculated after adding a pseudocount of 1 ($y = \log_2(x + 1)$). ChIP-seq signals are displayed as average enrichment of IP − input ($\log_2$).

RNA expression experiments (RNA-seq and qPCR). For embryos and larvae, RNA was isolated by freeze cracking (four times) followed by phenol-chloroform extraction and isopropanol precipitation. Total RNA was depleted for rRNA using Ribo-Zero Gold kit from Epicentre before library production using Total RNA Sequencing ScriptSeq kit. Gonad RNA was extracted from 50 prepared gonads per replica using the Arcturus pico pure RNA isolation kit followed by library production using the Total RNA-seq NuGen Ovation kit. 50-bp single-end sequencing was done on an Illumina HiSeq 2500. Processing of the RNA-seq data, gene and repeat expression levels from RNA-seq data were quantified as described previously[24] using WormBase (WS190) annotation for coding transcripts and Repbase annotations for REs. Primers used for qPCR experiments are listed in **Supplementary Table 3**.

Mutation sequencing experiments. Worms were grown at 25 °C for 1 month, and singled every second generation. Afterward worms were expanded on peptone-rich plates (20 cm) per replica and mixed staged worms were harvested for genomic DNA isolation. DNA was extracted by a standard protocol digesting worms with proteinase K, followed by phenol/chloroform extraction and RNAse treatment. 50-bp paired-end sequencing was done on an Illumina HiSeq 2500. Reads were mapped to the WS190 genome using BWA[78] and converted into BAM files using samtools[79]. Breakpoints were identified with Pindel[80] and SNVs with samtools.

Southern blot. Southern blot was performed following a standard protocol using a digoxigenin-labeled probe produced by PCR with primers listed in **Supplementary Table 3**.

LacZ mutator assay and cloning for somatic mutations. LacZ mutator assay was adapted from ref. 58. Worms were synchronized and grown for indicated durations on Dh5α containing plates. After a heat shock (heterochromatic array: 5 h (2 h at 33 °C, 1 h at 20 °C and 2 h at 33 °C), euchromatic array: 1 h 20 min (20 min at 33 °C, 10 min at 20 °C, 20 min at 33 °C, 10 min at 20 °C and 20 min at 33 °C) and 2 h recovery at 20 °C, worms were stained for β-galactosidase expression. To identify somatic mutations the indicated regions of the reporter were amplified using a Q5 proofreading polymerase (NEB) and primers listed in **Supplementary Table 3**. PCR products were batch clones into pCR2.1-TOPO sequencing vector (Invitrogen) and Sanger sequencing was performed on 20 clones per replica and region.

DNA damage sensitivity assays. Assays were previously described[81]. Recovery from an hydroxyurea (HU) pulse was monitored by soaking L1 larvae in M9 buffer containing indicated concentrations of HU and OP50 bacteria for 16 h before washing and plating on fresh OP50 plates. At day 3, the percentage of viable adults was quantified. To quantify IR sensitivity worms were irradiated (CellRad, Faxitron) at the L1 stage. At day 3, the percentage of viable adults was quantified.

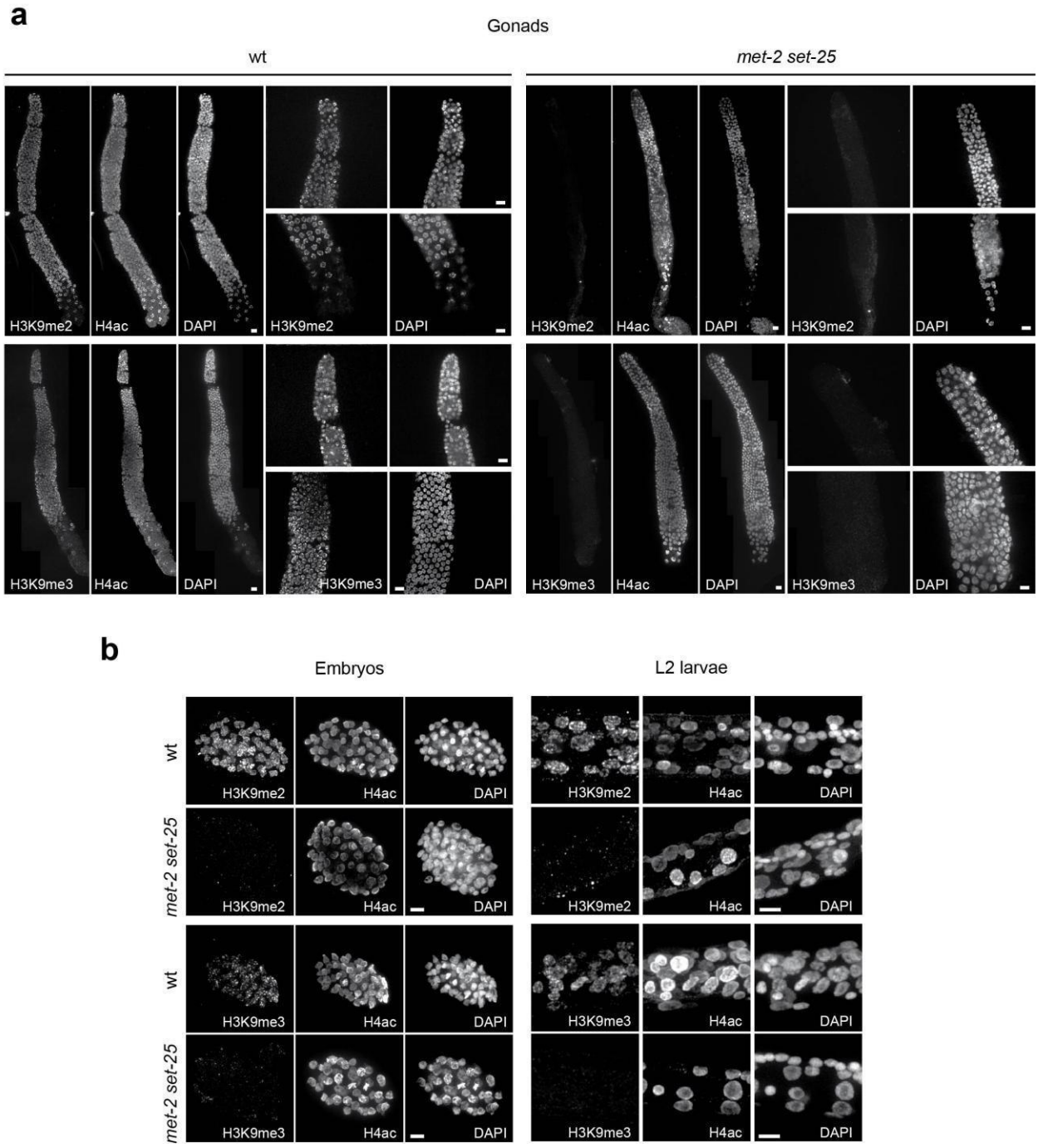
R-loop detection. For dot plots, genomic DNA was isolated using phenol-chloroform extraction followed by ethanol precipitation. DNA concentrations were determined using Nanodrop and the indicated amount of DNA was resuspended to a final volume of 50 µl in nuclease-free water after either a 1 h incubation with 5 µl of RNase H (NEB; +RNAse H), or a 1 h mock incubation at 37 °C, and spotted directly onto a nylon GeneScreen Plus membrane (NEF988; PerkinElmer) using a Bio-Dot Microfiltration Apparatus (Bio-Rad). The membrane was UV-crosslinked and blocked with 5% milk in

doi:10.1038/ng.3672

1 × PBS/0.1% Tween-20 before incubation with primary and secondary antibodies. The mouse S9.6 antibody (HB-8730, ATCC, gift of P. Pasero, Montpellier) was used at a 1:500 dilution, and a 10,000× dilution of goat anti-mouse HRP (Bio-Rad) was used as the secondary. The HRP signal was developed with Clarity Western ECL Substrate (Bio-Rad). Imaging was performed using ImageQuant LAS4000 mini und analyzed using ImageJ. Immunofluorescence staining was performed according to the steps described for the *C. elegans* larval stages using 4XSSC-T (0.1% Tween-20) instead of PBS-T. R loops were stained with the S9.6 antibody, diluted 1:100 in SSC-T and 3% BSA overnight at 4 °C.

DNA:RNA hybridization. Embryos were lysed by bead beating (MP BIOMEDICALS FastPrep-24 5G Instrument) in G2 buffer (80 mM guanidine HCl, 30 mM Tris pH 8.0, 30 mM 5% Tween, 0.5% TritonX). Genomic DNA was isolated by proteinase K digestion, followed by purification using genomic tips (500/G, QIAGEN). DNA was digested over night with AseI and BstUI at 37 °C (ref. 82). For RNase H control samples, RNase H was added in parallel to digestion. 5 µg of digested DNA per IP was incubated with 10 µl of S9.6 antibody overnight in binding buffer (10 mM $NaPO_4$, 140 mM NaCl, 0.5% Triton). Bound DNA fragments were recovered with 50 µl of Protein-A Dynabeads (Invitrogen), followed by four washes with binding buffer and proteinase K treatment. Samples were purified using DNA Clean & Concentrator-5 (Zymo Research) columns. Samples were sonicated to ~400-bp fragments before library preparation, as described above.

72. Meister, P., Towbin, B.D., Pike, B.L., Ponti, A. & Gasser, S.M. The spatial dynamics of tissue-specific promoters during *C. elegans* development. *Genes Dev.* **24**, 766–782 (2010).
73. Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N. & Nozaki, N. The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. *Cell Struct. Funct.* **33**, 61–73 (2008).
74. Ikegami, K., Egelhofer, T.A., Strome, S. & Lieb, J.D. *Caenorhabditis elegans* chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol.* **11**, R120 (2010).
75. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M.B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).
76. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
77. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
79. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
80. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
81. Craig, A.L., Moser, S.C., Bailly, A.P. & Gartner, A. Methods for studying the DNA damage response in the *Caenorhabdatis elegans* germ line. *Methods Cell Biol.* **107**, 321–352 (2012).
82. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).

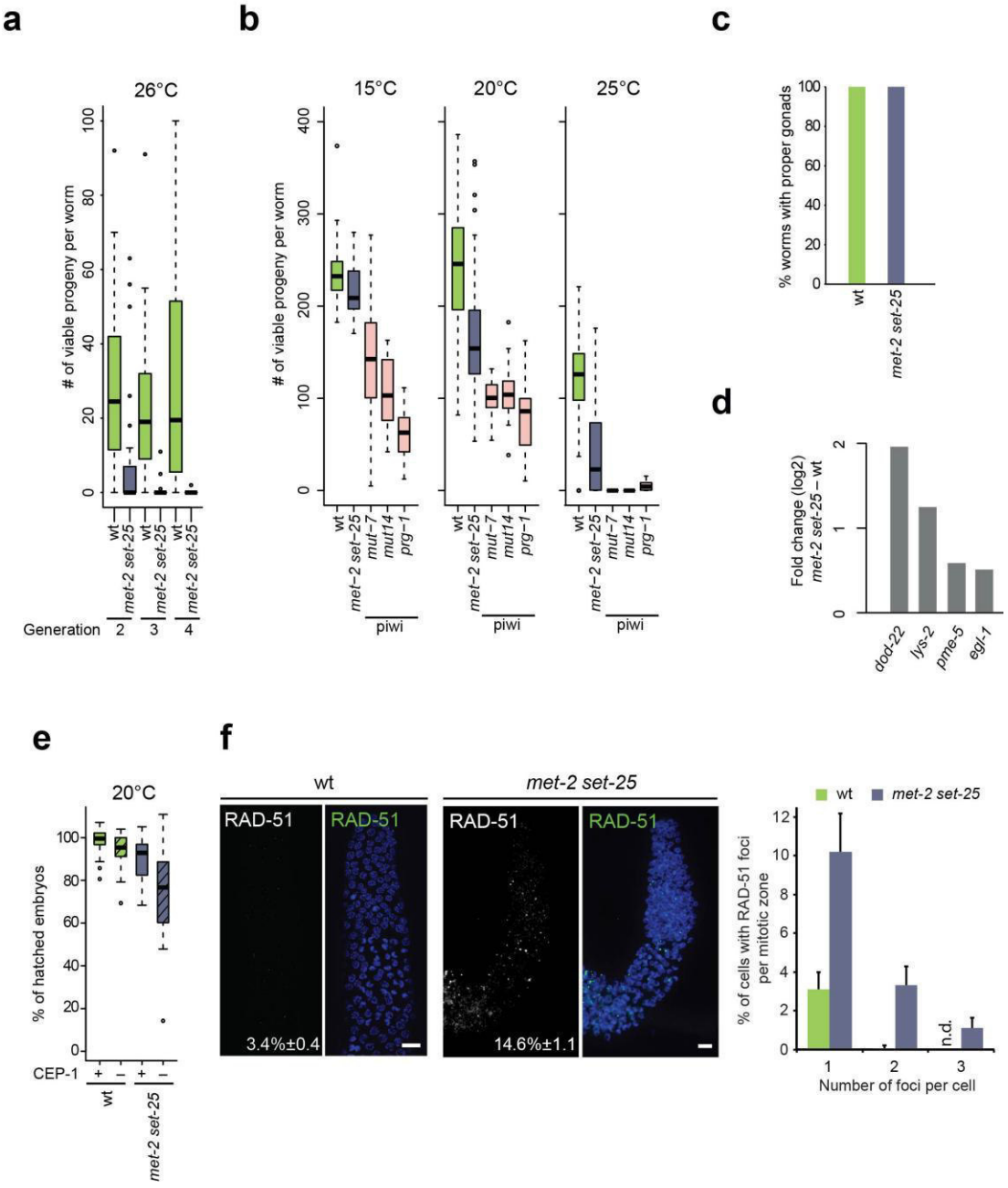


Supplementary Figure 1

Immunofluorescence (IF) confirms absence of H3K9me in *met-2 set-25* worms.

IF images of wild-type (wt) and *met-2 set-25* worms showing the loss of H3K9me2/me3 at the indicated developmental stages grown at 20 °C (Online Methods). The staining of H4 pan-acetyl (H4ac) served as a positive control. (**a**) Gonads (bar, 20 μm). (**b**) Embryos and L2-stage larvae (bar, 20 μm).

Nature Genetics: doi:10.1038/ng.3672

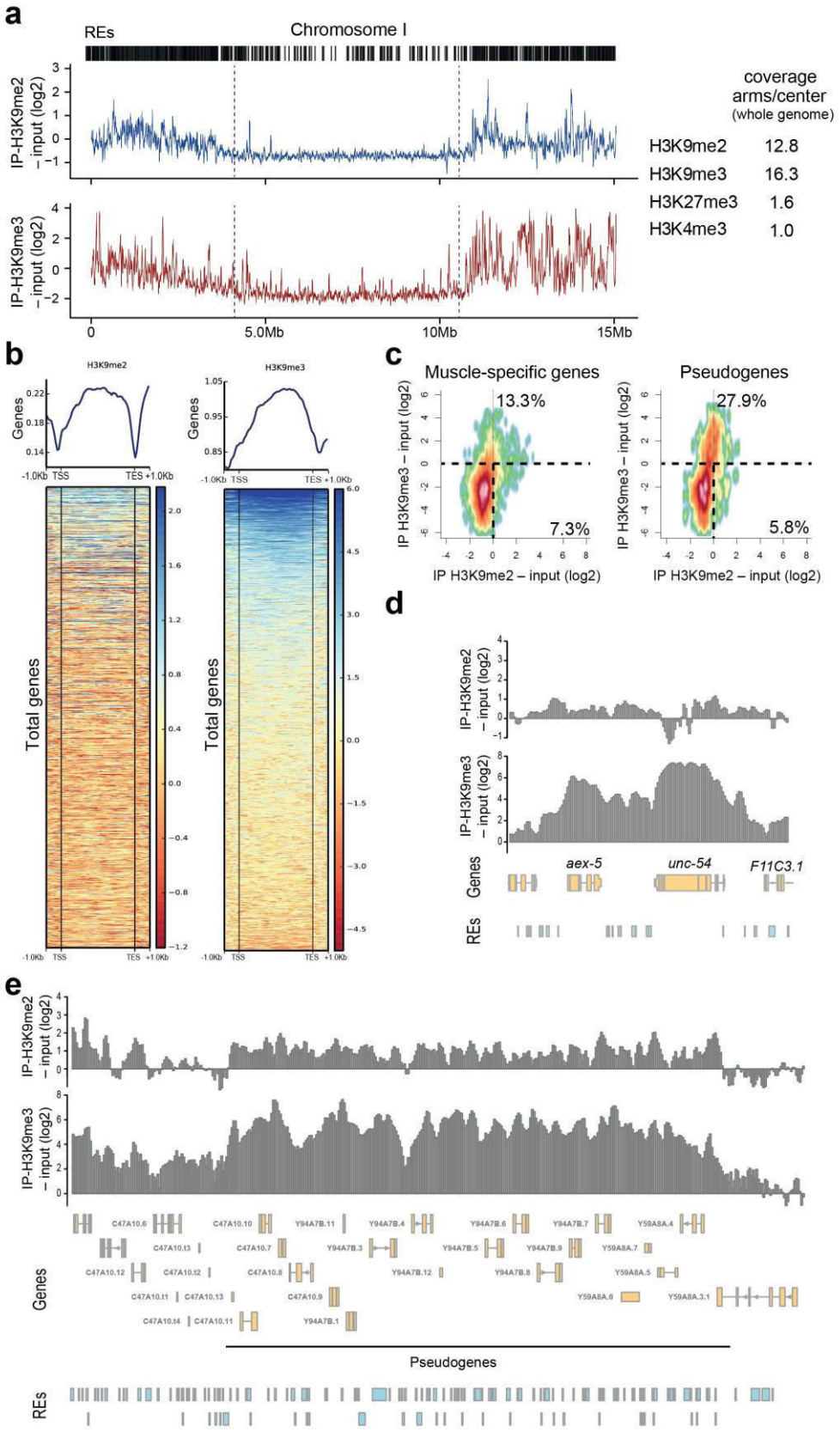


Supplementary Figure 2

Phenotypes of strains losing H3K9me in the germ line.

(**a**) Number of viable progeny of wt and *met-2 set-25* per worm at 26 °C, over three generations, starting at generation 3 after transfer of the worms from 20 °C (number of independent experiments (*N*) = 3, number of worms counted per experiment (*n*) = 60). (**b**) Number of viable progeny of wt, *met-2 set-25* and mutants of the PIWI pathway per worm at 15 °C, 20 °C and 25 °C by generation 3 after transfer from 20 °C to the indicated temperature (*N* = 1, *n* = 25). (**c**) Percentage of worms developing full gonad arms at 20 °C. mex-5:gfp-h2b was used to visualize gonad cells at all stages (*N* = 7, *n* = 2). (**d**) Analysis of RNA-seq data showing average fold change in expression of apoptosis response genes in the gonads of *met-2 set-25* mutant versus wt (*N* = 3, *P* < 0.05 adjusted for multiple testing; FDR < 0.05). (**e**) Percentage of laid eggs hatching at 20 °C. Strains deficient for CEP-1 (p53) additionally expressed the CED-1::GFP apoptosis reporter (*N* = 2, *n* = 80). (**f**) Analysis of RAD-51 foci detected by IF in the mitotic zone of the indicated genotypes to quantify the presence of resected DNA double-strand breaks. The percentage in images is equal to the frequency of mitotic tip cells with detectable RAD-51 foci. The bar graph further segregates positive cells by number of foci per cell (*N* = 3, *n* = 40 gonads).

Nature Genetics: doi:10.1038/ng.3672



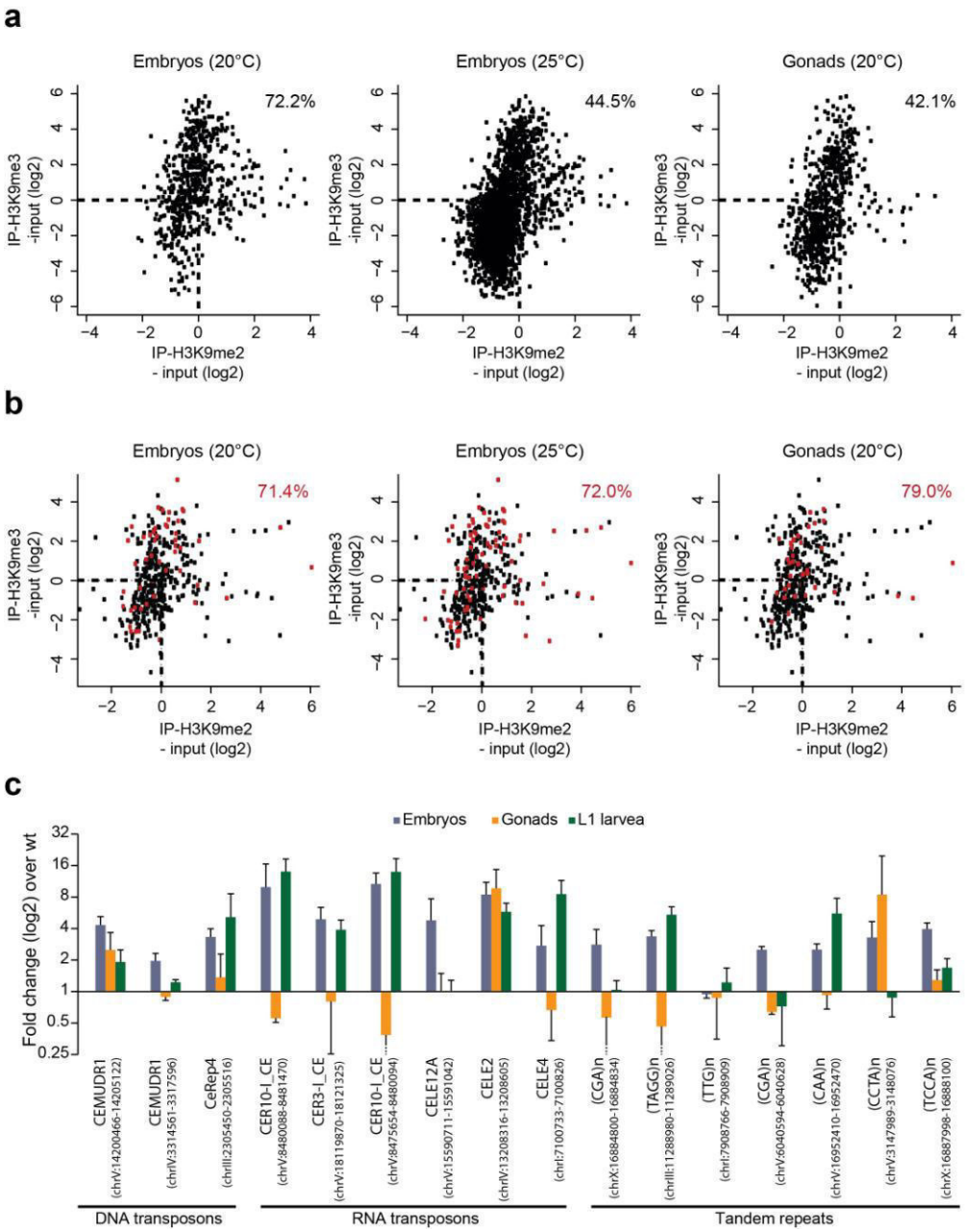
Nature Genetics: doi:10.1038/ng.3672

Supplementary Figure 3

H3K9me distribution on genes and repeats.

(**a**) H3K9me2 and H3K9me3 ChIP-seq were performed on early embryos at 20 °C ($N = 2$) in a wild-type (N2) strain. Distribution along chr. I in relation to repetitive elements (REs) is shown. Quantification to the right shows the ratio of the coverage of the indicated histone modification on chromosome arms (outer two-thirds) versus the center (inner one-third). H3K27me3 and H3K4me3 mapping data are from the modENCODE project based on ChIP performed on mixed-population embryos, 20 °C. (**b**) Metaplot and heat map showing $\log_2$ enrichment of H3K9me2 and H3K9me3 (IP versus input) over gene bodies. Each row represents one gene, displaying the binned coding region plus 1 kb upstream of the transcription start site (TSS) and 1 kb downstream of the transcription termination sites (TES). Blue is most enriched, red is least enriched. (**c**) High-density scatterplot showing H3K9me2 and me3 enrichment over muscle-specific genes and pseudogenes. The upper number indicates the percentage of genes that are H3K9me3 positive (including K9me2 positive and negative), and the lower number indicates the percentage of genes that are H3K9me2 positive but H3K9me3 negative (**d**,**e**) Distribution of H3K9me2 and H3K9me3 determined by ChIP-seq over the gene body of a gene with tissue-spec expression (**d**) neuronal *unc-54*, and a cluster of pseudogenes (**e**).

Nature Genetics: doi:10.1038/ng.3672

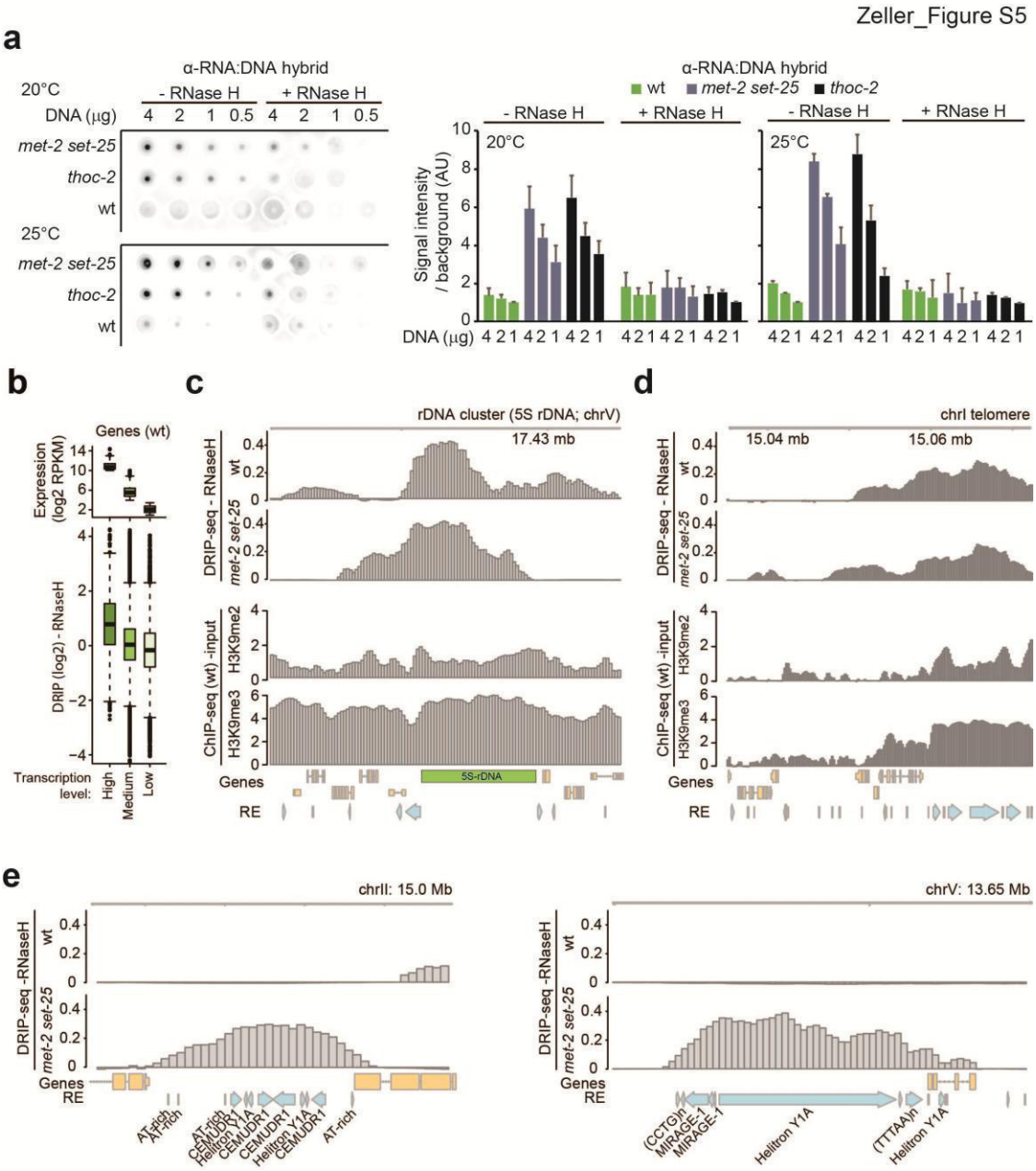


Supplementary Figure 4

Gene and repeat element derepression in the absence of H3K9me.

(**a**,**b**) H3K9me2 and H3K9me3 ChIP-seq was performed on early embryos at 20 °C (*N* = 2) in a wt strain, and gene expression was determined by two replicas of RNA-seq from embryos (20 °C and 25 °C) and from isolated gonads (20 °C) from wt and *met-2 set-25* strains. Scatterplots show H3K9me2 and H3K9me3 enrichment over the genes that are derepressed in early embryos at 20 °C and 25 °C, and in the gonads of *met-2 set-25* animals, versus wt. Number indicates the percentage of derepressed genes enriched for either H3K9me2 or H3K9me3 in wt embryos. (**b**) Scatterplot of H3K9me2 and H3K9me3 enrichment over repeat subfamilies. Red dots mark repeat subfamilies that are derepressed in the indicated tissue and conditions; black dots represent non-affected repeat subfamilies. The red number indicates the percentage of derepressed repeat subfamilies that are either H3K9me2 or H3K9me3 positive. (**c**) qPCR verification of a subset of REs that were detected as derepressed (>2 fold, *met-2 set-25*/wt) in *met-2 set-25* embryos at 20 °C. Expression of the same REs was additionally analyzed in gonads and L1-stage larval RNA. REs of all three main classes are detected (*N* = 3), but clearly there are strong stage-specific expression differences, with larvae and embryos showing more similarity.

Nature Genetics: doi:10.1038/ng.3672

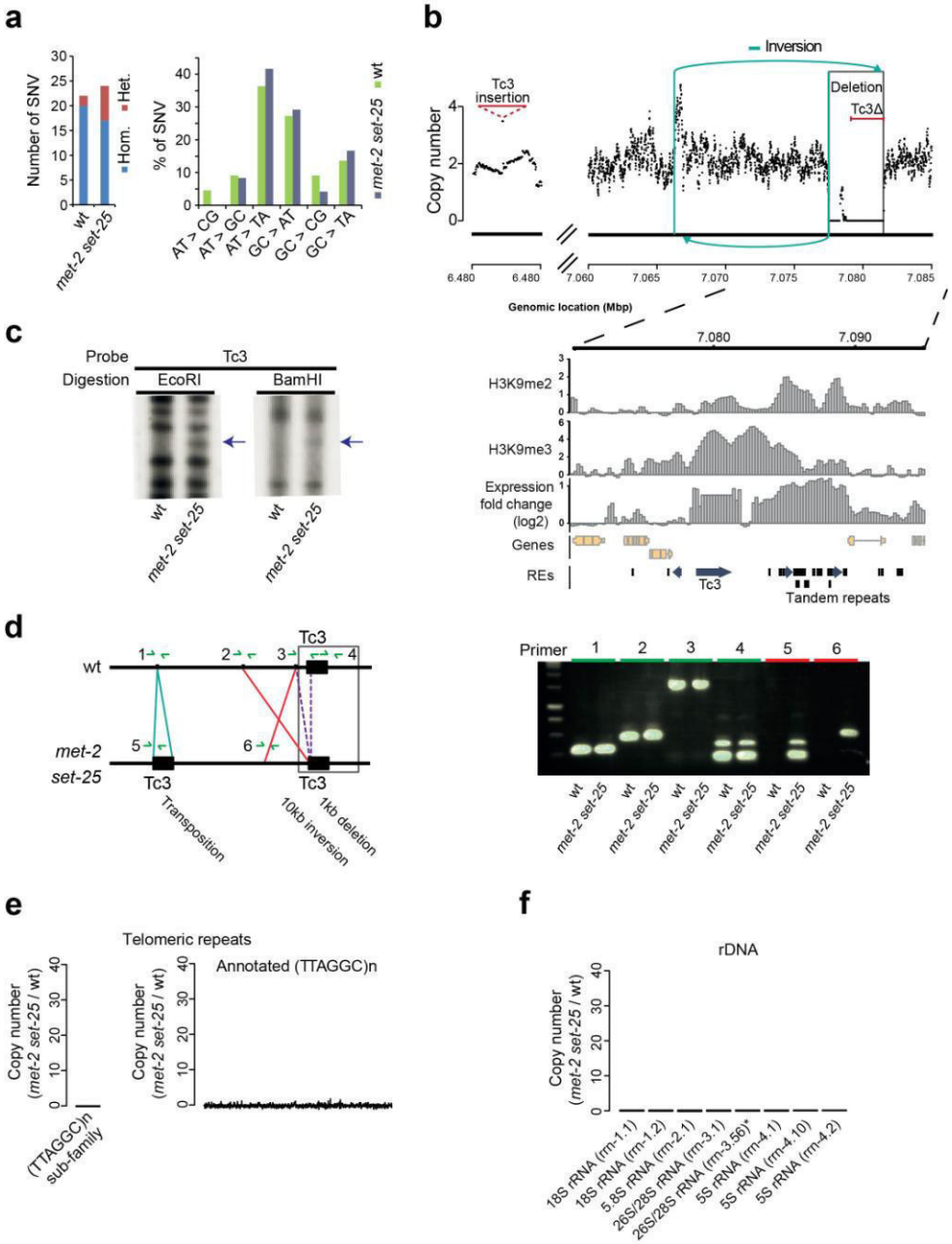


Supplementary Figure 5

R-loop accumulation on repeat elements (REs).

(**a**) RNA:DNA hybrids were detected on dot plots of genomic DNA isolated from the indicated genotypes grown at 20 °C or 25 °C (three blots at each temperature were quantified by scanning for **Fig. 5a**). Equal amounts of DNA (determined by $OD_{260/280}$) extracted from adult worms were spotted with or without RNase H treatment in decreasing concentrations (4, 2 and 1 µg). The nitrocellulose membrane was probed with the RNA:DNA-specific S9.6 antibody (n(20 °C) = 3, n(25 °C) = 3). Quantification on the right side with signals normalized to the background of each blot. (**b**) DRIP-seq signals in wt embryos are shown for genes grouped on the basis of their transcriptional activity. The upper box blot shows the level of transcription of the separate groups (N = 1). This enhancement of R loops on very highly expressed genes has been observed in many organisms. (**c**,**d**) Graphs show the accumulation of RNA:DNA hybrids in wt and *met-2 set-25* embryos relative to the distribution of H3K9me2 and me3 over the rDNA cluster (**c**) or the right telomere of chr. I (**d**), in wt embryos. (**e**) DRIP-seq examples showing the R-loop signal over two RE clusters. The ChIP signal from antibody S9.6, which is specific for RNA:DNA hybrids, was normalized to input, and the RNase H control values were subtracted.

Nature Genetics: doi:10.1038/ng.3672



Supplementary Figure 6

Germline mutations in *met-2 set-25* worms.

Further verification and characterization of the mutations detected in the genome sequencing experiment described in **Figure 6**. (**a**) Number and distribution of single-nucleotide variants (SNVs) or polymorphisms observed in the sequencing experiment described in **Figure 6a**. No dinucleotide preferences were found among SNVs. (**b**) Sketch of a complex rearrangement involving a Tc3 transposon, found exclusively in the *met-2 set-25* genome. The rearrangement was identified by genome sequencing. The graph to the left indicates the precise site of Tc3 insertion. Below are H3K9me2 and H3K9me3 ChIP-seq tracks for the region around the Tc3 transposon that provoked the inversion: it bears high levels of H3K9me3 and showed roughly a twofold change in expression in *met-2 set-25* over wt gonads at 20 °C. (**c**) Southern blotting with a probe against Tc3 shows a novel band detected in the *met-2 set-25* mutant. (**d**) Depiction of the rearrangement shown in **b** indicating the position of the primer pairs used to verify the rearranged genomic context by PCR. PCR confirmation of the rearrangement is shown to the right. (**e**) Copy number ratios of *met-2 set-25* worms relative to wt for the entire telomeric repeat subfamily, and for single telomere repeats. (**f**) Copy number ratios of *met-2 set-25* worms relative to wt for rDNA

Nature Genetics: doi:10.1038/ng.3672

repeats.

Nature Genetics: doi:10.1038/ng.3672

Size [bp]	Type	RE	H3K9me3 enrichment	Fold change expression
33481	Insertion	PALTTAA2	2.0	5.0
2335	Transposition	Tc3	3.4	5.7
2335	Transposition	Tc3	3.4	5.7
2335	Transposition	Tc3	3.4	5.7
1600	Transposion	Tc4	3.9	1.4
536	Deletion	-	1.4	-
23	Deletion	AT-rich	2.1	2.0
3	Deletion	Trinucleotide-rep	3.0	-
2	Deletion/insertion	HELITRON2	2.9	-

Supplementary Table 1

Germline mutations identified in 1 month old *met-2 set-25* substrains.

Characteristics of the insertions (ins), deletions (del) and transpositions (trans) unique to one of the eight *met-2 set-25* sub-strains. H3K9me3 enrichment and fold change in expression is given as the mean enrichment and mean fold change (log2) over the complete domain.

Nature Genetics: doi:10.1038/ng.3672

Name in manuscript	Genotype	International name	Reference
Wild-type (wt)		N2 (bristol)	
met-2 set-25	*met-2(n4256) III; set-25(n5021) III*	GW638	[24]
mex-5:h2b::gfp	*unc+119;rrrSi192[mex-5 prom::gfp: h2b::tbb-2 3'UTR;unc119(+)] II*	gift of R. Ciosk, FMI	[82]
met-2 set-25 mex-5:h2b::gfp	*met-2(n4256) III; set-25(n5021) III; unc+119; rrrSi192[mex-5 prom::gfp: h2b::tbb-2 3'UTR;unc119(+)] II*	GW1028	This paper
ced-1::gfp	*bcIs39 [Plim-7::ced-1::gfp; lin-15(+)] V*	MD701	[83]
met-2 set-25 ced-1::gfp	*met-2(n4256) III; set-25(n5021) III; bcIs39 [Plim-7::ced-1::gfp; lin-15(+)] V*	GW1203	This paper
cep-1 ced-1::gfp	*cep-1(ep347) I; bcIs39[(lim-7)ced-1p::GFP + lin-15(+)] V*	GW1266	This paper
met-2 set-25 cep-1 ced-1::gfp	*cep-1(ep347) I; met-2(n4256) III;set-25(n5021) III; bcIs39[(lim-7)ced-1p::GFP + lin-15(+)] V*	GW1268	This paper
mut-7	*mut-7(pk204) III*	NL917	[28]
mut-14	*mut-14(pk738) V*	NL1838	[28]
prg-1	*prg-1(tm872)*	WM161	[27]
Hc array (heterochromatic array)	*pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	NL3400	[58]
met-2 set-25 hc array	*met-2(n4256) III; set-25(n5021) III; pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	GW1198	This paper
msh-6 hc array	*msh-6(pk2504) I; pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	GW1280	This paper
ec array (Euchromatic array)	*pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	GW1351	This paper
met-2 set-25 ec array	*met-2(n4256) III; set-25(n5021) III; pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	GW1278	This paper

msh-6 ec array	*msh-6(pk2504) I; pkIs1604 [hsp-16.2 ATG(A)17::gfp/lacZ + pRF4(rol-6(su1006))]*	GW1279	This paper
thoc-2	*thoc-2(tm1310) III/hT2[bli-4(e937) let-?(q782) qIs48](I;III)*	GIN101	[84]

Supplementary Table 2

Worm strains used in this study

Repeat qPCR:

	Forward		Reverse	
	Name	Sequence	Name	Sequence
DNA1	SG-7490	AGGCCCATTCCATAGCTTTT	SG-7491	TTTCTGGGATTTCATGCACA
DNA2	SG-7494	CCGTTATCGTGACAAGCAGA	SG-7495	AAGCTTGAGCCGCTGATTAC
DNA3	SG-7521	CACGTTCCCAGCTTCAATTT	SG-7522	TGCAAATCTACAGTAACCCAGAA
RNA1	SG-7498	TACCAACGAGCCGAGTCTTC	SG-7499	TCTTCAGTTTCCTCGCCTGT
RNA2	SG-7502	GAAGCAGTCATTGTGGCTCA	SG-7503	GATTCTCCCTGTGCTCTTGC
RNA3	SG-7509	TACCAACGAGCCGAGTCTTC	SG-7510	TCTTCAGTTTCCTCGCCTGT
RNA4	SG-7523	TGCATGCAAGACTAATTTTCAA	SG-7524	CCAAAAGTTATTGGGCTTTCG
RNA5	SG-7527	GGCACGATGTTTTTGTTGAA	SG-7528	TCTATGAATTTCCCGCAGAA
RNA6	SG-7533	GCCGCTAGACACCTAACGAG	SG-7534	ATTATGGGGACGCAGAAAAA
Tan1	SG-7537	GGATGGATTGGGATGGATG	SG-7536	AACCATGCCAATCCTTTGTT
Tan2	SG-7541	GTGTAGCCGTGGATTGTGTG	SG-7542	ATTTGCCTGCTGGTCCATAG
Tan3	SG-7543	GTTGGGGCGGCTGTAGTT	SG-7544	AGCACCAAAGACGACAACAA
Tan4	SG-7545	ATCGGGATGGCTCGGTAT	SG-7546	ATCCCGATCCTTTGTTGTTG
Tan5	SG-7548	CAGCCACAACTACCACAACG	SG-7550	CAGTCTGTTGGACACGGAAC
Tan6	SG-7551	ACCTACGTGCCTGCCTACAT	SG-7552	TTTTGTCAGGGACATGCGTA
Tan7	SG-7555	CCTGGCACTTACCAGTCCAT	SG-7556	ATCGGGATGGATGGCTTC

DRIP qPCR:

Tc4	SG-7513	GAGTTTCCGTCCCGATTACA	SG-7514	AGAAACGGTTCGAACAATGC
CEMUDR-1	SG-7490	AGGCCCATTCCATAGCTTTT	SG-7491	TTTCTGGGATTTCATGCACA
Tc3	SG-6909	GAAGGATCCGGTGAGCTACG	SG-6910	TACAGGAGTTGGAGGCAGCA
CER10-I_CE	SG-7498	TACCAACGAGCCGAGTCTTC	SG-7499	TCTTCAGTTTCCTCGCCTGT
Helitron 2	SG-8080	GATGTTTGGAGAGATAGTGGG	SG-8082	GACAACATCCATCCACTAACC
MSAT-1	SG-8076	GGAATGTTCCAGAACTTTCTAG	SG-8077	ACATTCCAGACTTTTCCCAA
(TATCG)n	SG-8078	CGTATCGTTTCTAGCTATATCG	SG-8079	CGATACGATATAGCTTTTTACG
unc-119	SG-2869	CCACACCACCTCTAATCTCC	SG-2870	TCATTTCTCTGCGTCTTCCT
lmn-1	SG-2797	CAAGAGAACAACAGACTCCAG	SG-2798	TAATAAGACCACCGCATCAG
eef-1A.1	SG-8975	AGGAATGGTCGTTACCTTCGC	SG-8976	CAGACGGATCCACGACGAATA

Repeat copy number:

HELITRON Y4	SG-8074	TGTTCTGCCAATTTATTTACTC	SG-8075	GGCAGAACATTAGAGTAAATAT
HELITRON 2	SG-8080	GATGTTTGGAGAGATAGTGGG	SG-8082	GACAACATCCATCCACTAACC
MSAT-1	SG-8076	GGAATGTTCCAGAACTTTCTAG	SG-8077	ACATTCCAGACTTTTCCCAA
(TATCG)n	SG-8078	CGTATCGTTTCTAGCTATATCG	SG-8079	CGATACGATATAGCTTTTTACG
CEREP 58	SG-8071	CGGGCAAATCAACAATTGAA	SG-8073	GAAGATGAGGAATTTAATTG
lmn-1	SG-2797	CAAGAGAACAACAGACTCCAG	SG-2798	TAATAAGACCACCGCATCAG

Rearrangement:

PP1	SG-7680	CAAGGTGGAAGACATGTA	SG-7681	TACGAGCAATTCAACTGG
PP2	SG-7682	TTTATATCGGGTCCAACC	SG-7683	GAAGCACGTCGACTGAAG
PP3	SG-7684	AGGGGGGAAGAATTTAAC	SG-7685	TCAAATTAGGGGGGGTCC
PP4	SG-7685	TCAAATTAGGGGGGGTCC	SG-7686	GTGTGGATACTGTCGTTC
PP5	SG-7680	CAAGGTGGAAGACATGTA	SG-7685	TCAAATTAGGGGGGGTCC
PP6	SG-7682	TTTATATCGGGTCCAACC	SG-7684	AGGGGGGAAGAATTTAAC
Tc3-probe	SG-7201	CTGTAAGACGGCAAGAGA	SG-7202	TCTTGTTCTGAGCATACACG

H3K9me2 and H3K9me3 ChIP on LacZ mutator arrays

Tc4	SG-7513	GAGTTTCCGTCCCGATTACA	SG-7514	AGAAACGGTTCGAACAATGC
gfp (array)	SG-6987	GGAGAGGGTGAAGGTGATGC	SG-6988	CATAACCTTCGGGCATGGCA
lmn-1	SG-2797	CAAGAGAACAACAGACTCCAG	SG-2798	TAATAAGACCACCGCATCAG

Cloning of array parts

Part 1	SG-8150	CTACTTTTCCATGTACCG	SG-8258	GTTTCTTTCATCTCACTGGATCCCC
Part 2	SG-8151	TCATCTCACTGGATCCCC	SG-8259	GTTTCTTGTGTCCAAGAATGTTTCC
Part 3	SG-6987	GGAGAGGGTGAAGGTGATGC	SG-2039	GCAGCTGTTACAAACTCAAG

Supplementary Table 3

Primers used in this study

supplement-only references

83. Tocchini, C. *et al.* The TRIM-NHL protein LIN-41 controls the onset of developmental plasticity in Caenorhabditis elegans. *PLoS Genet* **10**, e1004533 (2014).
84. Lu, N., Yu, X., He, X. & Zhou, Z. Detecting apoptotic cells and monitoring their clearance in the nematode Caenorhabditis elegans. *Methods Mol Biol* **559**, 357-70 (2009).
85. Castellano-Pozo, M., Garcia-Muse, T. & Aguilera, A. R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep* **13**, 923-9 (2012).

Chapter 4: Specialized roles of Histone H3 K9me2 and K9me3 in *C. elegans* repeat repression and germline integrity

Jan Padeken, Peter Zeller, Veronique Kalck, Susan M Gasser

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

First two authors contributed equally to this work.

Both were involved in most experiments shown in Fig1-2, Fig5-6 and SupFig2

Peter Zeller did the majority of experiments in Fig3, SupFig1, SupFig3 and wrote the manuscript

Jan Padeken did the majority of experiments in Fig4 and did all bioinformatics analysis

Veronique Kalck did the experiments for Fig3e

Summary

The methylation of lysine 9 of histone 3 (H3K9me) constitutes a major histone modification and a hallmark of constitutive heterochromatin. Lysine can be mono-, di- or tri-methylated, and these modifications provide binding sites for me-lysine readers like HP1 (Lachner et al. 2001). Members of the HP1 family recognize H3K9me, leading to the compaction and silencing of appropriately modified chromatin domains. While most of these readers can recognize more than one methylation state, it has been difficult to identify roles that differentiate the physiological roles of H3K9me2 from those of H3K9me3. Using *C. elegans*, an organism with a single enzyme that can trimethylate H3K9 (SET-25), we were able to differentiate the roles of me2 and me3.

We show that H3K9me3 specifically represses a subset of transposable elements and genes. In contrast, the MET-2-deposited H3K9me2 suppresses transposable elements and tandem repeats. MET-2 suffices to suppress the vast majority of deficiencies that characterize worms lacking all H3K9 methylation. In the complete absence of H3K9me, RNA:DNA hybrids accumulate on

transcribed repeats, forming obstacles for the DNA replication machinery and thereby provoking insertions and deletions at these sites. The loss of MET-2, which reduces H3K9me1, me2, and me3 to 20% of wild-type levels, provokes the same R-loops and delocalizes chromosome arms from the nuclear periphery, leading to sterility, delayed development, and synthetic lethality with a variety of RNA processing enzymes and factors involved in DNA repair. This is not the case in the *set-25* mutant, which loses H3K9me3, but retains H2K9me2. This clearly distinguishes the physiological roles of di- and tri-methylated H3K9.

ChIP-qPCR experiments for H3K9me2 and me3 in single mutants allows us to identify 3 pathways of H3K9 methylation mediated silencing that differ in their dependency on MET-2 and SET-25 as well as the localization of the genomic locus relative to the nuclear envelope.

Specialized roles of Histone H3 K9me2 and K9me3 in *C. elegans* repeat repression and germline integrity

Jan Padeken[1*], Peter Zeller[1,2*], Veronique Kalck[1], Susan M Gasser[1,2†].

1) Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland
2) Faculty of Natural Sciences, University of Basel, Klingelbergstrasse 70, CH-4056 Basel Switzerland

*authors contributed equally

Summary

The methylation of lysine 9 of histone 3 (H3K9me) constitutes a major histone modification and a hallmark of constitutive heterochromatin. Lysine can be mono-, di- or tri-methylated, and these modifications provide binding sites for me-lysine readers like HP1 (Lachner et al. 2001). Members of the HP1 family recognize H3K9me, leading to the compaction and silencing of appropriately modified chromatin domains. While most of these readers can recognize more than one methylation state, it has been difficult to identify roles that differentiate the physiological roles of H3K9me2 from those of H3K9me3. Using *C. elegans*, an organism with a single enzyme that can trimethylate H3K9 (SET-25), we were able to differentiate the roles of me2 and me3 (Supplementary Figure 1a, (Towbin et al. 2012)). We show that H3K9me3 specifically represses a subset of transposable elements and genes. In contrast, the MET-2-deposited H3K9me2 suffices to suppress transposable element, tandem repeat and gene expression as well as the vast majority of deficiencies that characterize worms lacking all H3K9 methylation. In the complete absence of H3K9me, RNA:DNA hybrids accumulate on transcribed repeats, forming obstacles for the DNA replication machinery and thereby provoking insertions and deletions at these sites. Compared to lacking all H3K9me, the loss of MET-2, which reduces H3K9me1, me2, and me3 to 20% of wild-type levels, provokes the same level of R-loops and delocalizes chromosome arms from the nuclear periphery, leading to sterility, delayed development, and synthetic lethality with a variety of RNA processing enzymes and factors involved in DNA repair. These phenotypes are not observed in the *set-25* mutant. A major transcriptional difference is the derepression of tandem

repeats only in the met-2, not in the set-25 mutant. This clearly distinguishes the physiological roles of di- and tri-methylated H3K9.

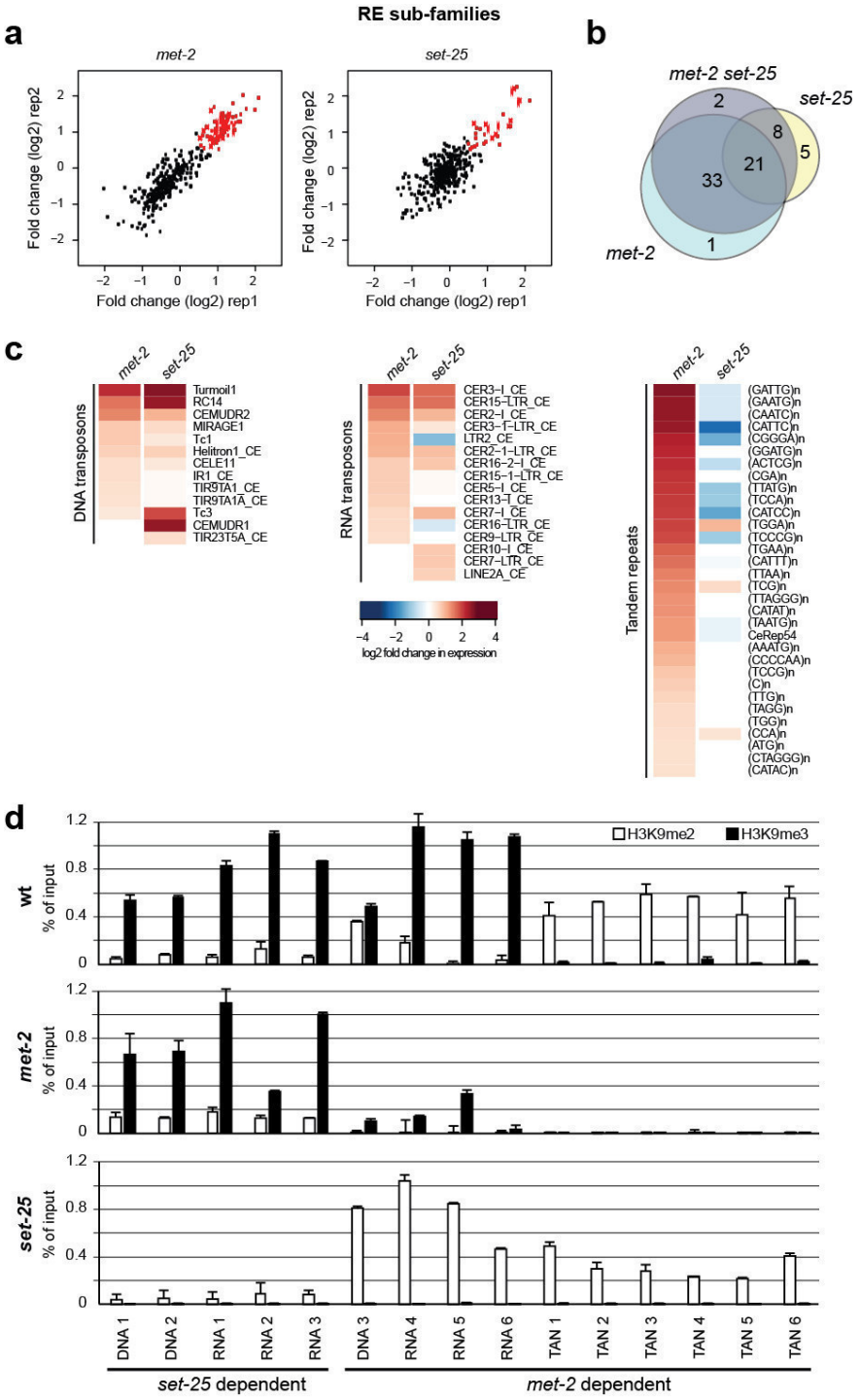
SET-25 activity silences a specific subset of genes and repeats, while MET-2 represses the majority of H3K9 methyl dependent sequences

In *C. elegans*, H3K9me can be found on genes as well as tandem repeats, DNA- and RNA-transposons, which constitute the three main classes of repetitive elements (RE) (Gerstein et al. 2010; Zeller et al. 2016). By RNA-seq analysis of transcripts differentially expressed in *met-2 set-25* double mutant vs wild-type embryos, we observe that a large fraction of RE, as well as tissue-specific genes are derepressed (Zeller et al. 2016). Using highly specific antibodies, we have mapped the distribution of H3K9me2 and me3 to partially overlapping sequence classes (Zeller et al. 2016). Whereas H3K9me2 is primarily enriched on DNA transposons and tandem repeats, H3K9me3 is found over the body of unexpressed genes as well as DNA- and RNA transposons. Their distinct distribution raised the possibility that the two marks are responsible for repressing different sequences.

We therefore performed total RNA sequencing experiments in early embryos of mutants lacking either *met-2* or *set-25*, comparing their transcripts with those recovered from wild type and *met-2 set-25* animals (Zeller et al. 2016). We identified 336 gene transcripts whose levels are significantly increased upon loss of MET-2 (Supplementary Figure 1b, fold change > 2), 75% of these are identical to those detected in worms lacking all H3K9 methylation (Supplementary Figure 1c). In striking contrast, and despite the fact that H3K9me3 is enriched on many genes (Zeller et al. 2016), *set-25* embryos showed derepression of only 30 genes (Supplementary Figure 1b), with over 90% of these genes also abnormally expressed in *met-2 set-25* double mutants (Supplementary Figure 1c). This discrepancy between the distribution of the methylation state in wt embryos and the transcriptional consequence of its loss suggested that H3K9me2 (or another modification) can compensate for H3K9me3 loss to repress most genes.

We next examined the derepression of repeat subfamilies in the single mutants, and compared these to the derepression scored in the double mutant. Very similar to the situation of H3K9me3-marked genes, embryos showed 55 repeat subfamilies derepressed in the *met-2* mutant (Figure 1a, fold change > 1.5; 98% overlapping with *met-2 set-25*, Figure 1b), while in the *set-25* mutant, only 34 repeat subfamilies were significantly derepressed (85% overlapping with RE mis-expressed in *met-2 set-25* embryos; Figure 1a,b).

Figure 1 MET-2 and SET-25 suppress a partially overlapping set of genes and repetitive elements



(a) Scatter blot showing repeat subfamily expression as log2 fold change over wt in early embryos of indicated mutants (2 replicas shown, N = 3). **(b)** Venn diagram showing repeat subfamilies commonly or uniquely regulated by MET-2 and SET-25 (N = 3). **(c)** Heat map showing derepressed repeat subfamilies in *met-2* and/or *set-25* worms separating them into the three major repeat classes (DNA transposons, RNA transposons, Tandem repeats). **(c)** H3K9me2 and me3 ChIP qPCR in early embryos of indicated genotypes on the RE analyzed in (supplementary 1d) (N= 3).

We next examined precisely the classes of repeats that were affected by loss of SET-25 or of MET-2. By separating the REs into the three main classes, it was clear that *met-2* embryos derepressed all three repeat classes, especially tandem repeats, while loss of SET-25 led to

derepression of both transposon classes but not tandem repeats (Figure 1c). This separation of function was confirmed by qPCR of specific examples of each class (Supplementary Figure 1d).

We note that the vast majority of H3K9me3-positive regions do not become expressed in a *set-25* mutant, and that by mass spectroscopy, H3K9me2 levels increase slightly (Towbin et al. 2012). Correlating the enrichment of H3K9me2 and me3 over the genomic regions derepressed in the single mutants (Supplementary Figure1e), we could see that while regions derepressed in the *set-25* mutant are only enriched for H3K9me3, regions dependent on MET-2 could be found enriched for both H3K9me2 and me3, suggesting an interdependency of both marks. To test this hypothesis, we performed H3K9me2 vs H3K9me3 ChIP-qPCR in both *met-2* and *set-25* single mutants as well as wt (Figure 1d). We identify three classes of HMT dependency. First, a group of transposons whose transcriptional silencing depends on SET-25 and that is mainly H3K9me3 positive under wild type conditions. The H3K9me state of this class depended solely on SET-25. Second, different group of transposons that like the first is mainly H3K9me3 positive in wild type animals but depended on MET-2 for its silencing. Indeed, we find that H3K9me2 accumulates on these sequences in the absence of SET-25 and H3K9me is completely lost in *met-2* embryos (Figure 1d). We conclude that at these loci SET-25 mediated H3K9me3 depends on the preceding catalysis of H3K9me1/me2 by MET-2. The third group shows high level of H3K9me2, but not H3K9me3, which is dependent on MET-2. This group encompasses all the tested tandem repeats.

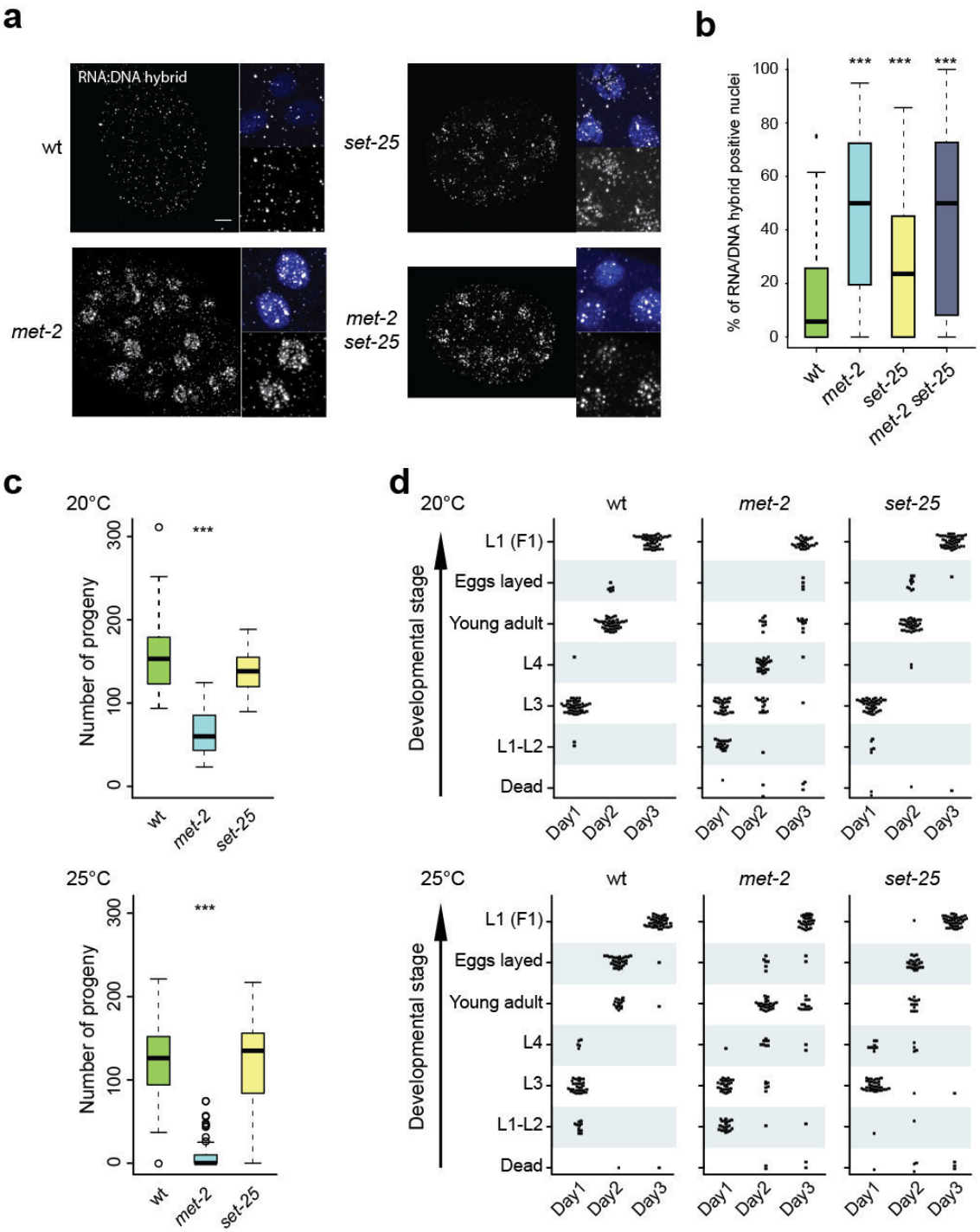
H3K9me2 is sufficient to suppress the accumulation of RNA:DNA hybrids

We recently showed that the transcription of RE leads to a striking accumulation of RNA:DNA hybrids which are specifically enriched on expressed RE (Zeller et al. 2016). These RE classes were enriched for DNA transposons and tandem repeats (and not RNA transposons), much like the transcription patterns detected upon loss of MET-2. We therefore monitored the accumulation of RNA:DNA hybrids in *set-25* vs *met-2* single mutants.

The detection of RNA:DNA hybrids can be performed with the S9.6 antibody (Boguslawski et al. 1986) in early embryos (Figure 2a). As expected, wild-type worms showed very little staining, while the *met-2 set-25* double mutant shows a strong nuclear enrichment for RNA:DNA hybrids (see also Zeller et al. 2016). Both single mutants, *met-2* and *set-25,* are positive for RNA:DNA hybrids , but to different degrees: the *met-2* mutant shows multiple strong foci in the majority of cells, while foci in *set-25* are weaker and less frequent (Figure 2a,b). We examined the synthetic lethality of the single mutants in the presence of transiently applied replication stress (16h HU), and found

that both single mutants (*met-2* and *set-25*) show clear sensitivity to HU (Supplementary Figure 2a).

Figure 2 MET-2 and SET-25 accumulate RNA:DNA hybrids and impair developmental timing and fertility to different degrees



(a) Representative images taken under identical conditions, of early embryos grown at 20°C ofindicated genotypes, stained for RNA:DNA hybrids (antibody S9.6, HB-8730, ATCC). Images on the right show

enlargements, with the lower panel showing the RNA:DNA hybrid signal alone and the upper panel showing the overlay with DAPI (blue, scale bar=5µm). **(b)** Quantification of RNA:DNA microscopy showing the percentage of cells showing nuclear specific staining in indicated genotypes (N=3, n (number of animals per replica) = 20, ***=p-value < 0.0001). **(c)** Fertility was measured by counting the complete progeny of singled worms at indicated temperatures (N = 3, n = 25); **(d)** Developmental timing was analyzed by following singled worms over 3 days of development at indicated temperature. Each dot represents a single worm (N = 3, n = 50).

As previously noted, the complete loss of H3K9me leads to a pronounced loss of fertility, which is detected at 20°C but enhanced at 25°C. This is thought to arise at least in part from the enhanced DNA damage incurred by the high level of RNA:DNA hybrids (Zeller et al. 2016). We therefore scored for fertility in the single mutants by counting the total number of viable progeny produced by singled worms. Whereas *met-2* worms showed impaired fertility at 20°C and became nearly sterile at 25°C (Figure 2c), *set-25* worms were indistinguishable from wild type animals at both temperatures tested. Thus, loss of fertility correlates with the enhanced accumulation of RNA:DNA hybrids scored in *met-2* worms.

In the same context we also monitored a stochastic delay in somatic development in worms that lack all H3K9me (*met-2 set-25).* To test whether somatic development is also differentially affected by the single mutants, like germline infertility, we monitored the timing of somatic development of singled mutants over three days of somatic development. We note that wild-type and *set-25* mutant worms develop in a highly synchronized fashion at 20°C (100% wt and 95% *set-25* young adults at day 2; Figure 2d), whereas *met-2* worms show a strong albeit stochastic delay in development (13% young adult at day 2, 20°C). When exposed to temperature stress (25°C), *set-25* worms also began to show delayed development (98 and 85% reaching adulthood at day 2 in wild type and *set-25* respectively), although it is less pronounced than in *met-2* mutants.

We conclude that H3K9 methylation levels are important for both the robust kinetics of development and for the integrity of oocytes during their maturation in the gonad. MET-2 deposited H3K9me2 is able to compensate for the loss of H3K9me3 under normal conditions, except under temperature stress (25°C), or replication stress (HU). The strong defects in development and germline maturation detected in the *met-2* mutant, correlate with the higher levels of RNA:DNA hybrid accumulation in this mutant, and the transcription of tandem repeats (Figure 1c).

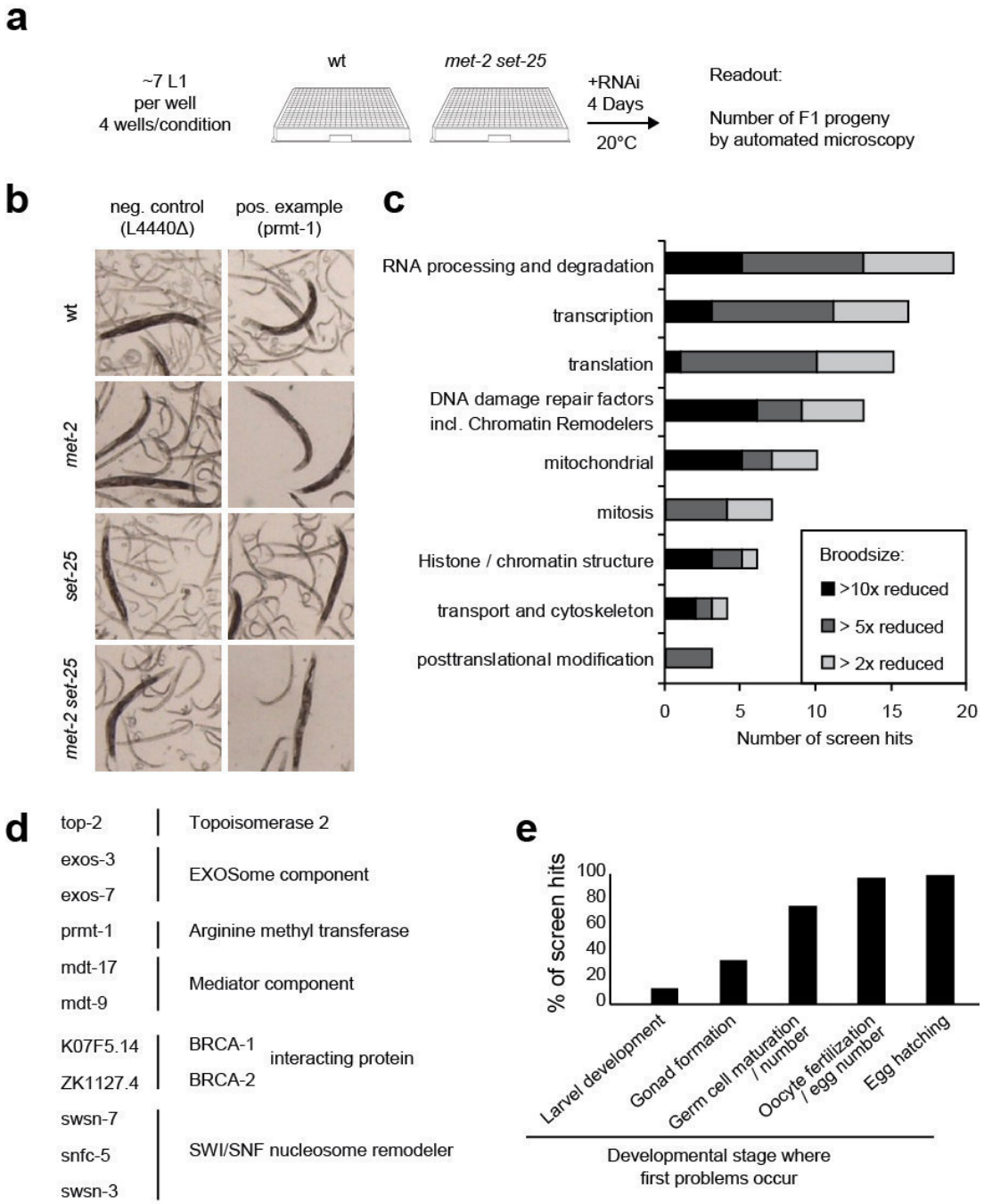
A synthetic lethality screen identifies RNA processing and repair genes that that become essential in the absence of MET-2, but not SET-25

In order to determine whether specific pathways compensate for the loss of MET-2 and SET-25, we combined the genetic strength of *C. elegans* with its ability to survive in the complete absence of H3K9 methylation, to perform a whole genome RNAi synthetic lethality screen. For this an RNAi library based on bacteria expressing small RNAs that target 72% of *C. elegans* coding genome (Kamath et al. 2003), were fed in parallel to wild-type and *met-2 set-25* worms (Figure 3a). The primary screen was performed in quadruplicates, images were taken using the automated IN-Cell microscope system (GE Healthcare) and differential effect of RNAi clones on the survival and reproduction of *met-2 set-25* versus wild type worms was determined both with a by eye (Figure 3b).

Hits having a mutant specific effect in three of 4 replicas were repeated two additional times. RNAi clones of reproducible hits were singled, sequenced and verified on plates, and where possible, clones of a second RNAi library (Vidal (Rual et al. 2004)) were tested, too. Supplementary table 1 contains the complete list of RNAi clones that were reproducibly synthetic sterile or nonviable upon ablation of all H3K9me, and not in wild-type worms. We indicate known or predicted functions, the human homologue, as well as the strength of the mutant specific effect as a fold change in F1 progeny numbers. Classifying them by biological pathways they are involved in, it is striking to see that the majority of hits play roles in transcriptional and repair processes (Figure 3c, RNA processing and degradation, transcription, Chromatin remodelers, Chromatin structure), although effects were also observed for genes involved in translation and mitosis. This supports the proposal that H3K9me is primarily a transcriptional regulator and that misregulation of transcription is at the root of the observed sterility (Zeller et al. 2016). We note that none of these hits was itself transcriptionally affected by loss of H3K9me (Supplementary Figure 3a).

We note that two groups of RNAi targets strongly support the role of H3K9me in preventing RNA:DNA hybrid related DNA damage. Specifically, components of the exosome (Pefanis et al. 2015) and mediator complex (Wahba et al. 2011) as well as arginine methylation (Yang et al. 2014) and topoisomerases (El Hage et al. 2010) are related to the prevention of hybrid accumulation (Figure 3d) (Santos-Pereira and Aguilera 2015), as are proteins involved in DNA damage repair, such as the BRCA1 and BRCA2 interacting factors (Bhatia et al. 2014) and components of the SWI/SNF nucleosome remodeler complex (Figure 3d).

Figure 3 Repair and RNA processing genes that are synthetic lethal with loss of H3K9me are sensitive to *met-2,* but not *set-25* mutation



(a) Synthetic lethality screen concept. L1 worms of wt and *met-2 set-25* worms were fed with RNAi expressing bacteria for 4 days at 20°C before their viability and fertility was assessed based on the number of F1 progeny. **(b)** Representative pictures of worms of indicated genotypes being fed with bacteria expressing either the empty vector (L4440Δ) control or *prmt-1* (positive hit) RNAi. **(c)** Bar-chart showing the different functional categories that RNAi clones synthetic lethal with *met-2 set-25* belong to (N=3, n=2).

(d) Selected screen hits implicated in the prevention of RNA:DNA hybrids, their induction of DNA damage and the repair of DNA damage in general. **(e)** The developmental stage affected by screen hits was further assed using pMEX-5:H2B-GFP to visualize the germline. Graph shows the cumulative percentage of screen hits having encountered development problems at the indicated stage specifically in combination with *met-2 set-25* (N=2, n=20).

Using a H2B-GFP fusion protein to visualize the germline, we characterized the strongest 55 hits in detail to identify the developmental stage at which H3K9me is especially important (Figure 3e and supplementary table 2). 84% of the analyzed hits led to problems in germ cell development, proliferation or maintenance, consistent with the accumulation of germline DNA damage in *met-2 set-25* mutants (Zeller et al. 2016).

Finally, we tested all 103 hits, which reproducibly generated sterility or embryonic death in the background of the double mutant, in two independently derived *met-2* deletion strains (n4256 and ok2307) and a *set-25* mutant (n5021). We found that that all RNAi clones were also synthetic lethal with loss of MET-2 alone (N=3) but none showed any synthetic effect with loss of SET-25 (Figure 3b and Supplementary Figure 3b).

Two pathways control tandem repeat expression: MET-2 mediated H3K9me2 and the BRCA-1 complex

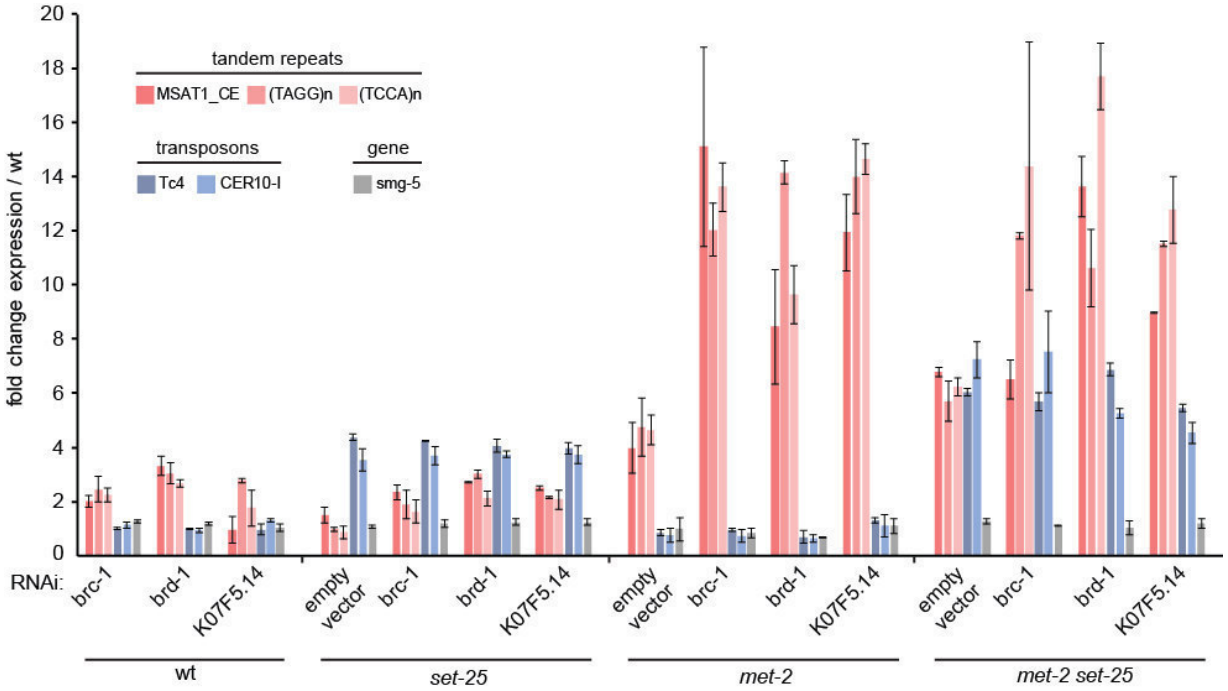
One hit from the synthetic lethality screen that especially caught our attention is K07F5.14. The only identified interactor of K07F5.14 by yeast two hybrid screens is BRC-1 (Li et al. 2004), the *C. elegans* homologue of the human breast-cancer susceptibility gene 1 (BRCA1). BRCA1 is a tumor suppressor with E3 ubiquitin ligase activity that contributes to DNA repair by homologues recombination (Scully et al. 1997; Moynahan et al. 1999) and transcriptional regulation under DNA damage conditions (Chapman and Verma 1996; Anderson et al. 1998). Mutations in BRCA1 lead to a strong predisposition to developing breast and ovarian cancer. *C. elegans* BRC-1, as well as mammalian BRCA-1, forms a heterodimer with BRD-1 and BARD1 respectively (Boulton et al. 2004), which is essential for its role in preserving genome integrity as well as ubiquitination (Polanowska et al. 2006).

New evidence suggests that the repressive function of BRCA1 is of central importance for the preservation of genome integrity. Zhu et al. could show that loss of BRCA1 in mice leads to the derepression of the tandemly repeated satellite DNA, but no other heterochromatic sequences including transposable elements (Zhu et al. 2011). The function of BRCA1 in repeat silencing is mediated by its ability to monoubiquitinate H2A, because the expression of a constitutively

ubiquitinated Histone H2A could restore satellite expression to wt levels. Strikingly, artificial expression of these tandem repeats from a transgene can phenocopy the BRCA1 deletion effect, including loss in centrosome amplification, cell-cycle checkpoint defects, DNA damage and genomic instability (Zhu et al. 2011). This argues that transcriptional control of tandem repeats is one of the main functions of this central and well-studied tumor suppressor.

In order to determine if BRC-1 is also involved in transcriptional tandem repeat repression in *C. elegans* and if BRC-1 and MET-2 work on the same pathway, *met-2*, *set-25* as well as *met-2 set-25* worms were fed bacteria expressing RNAi against BRC-1 or its interacting proteins K07F5.14 or BRD-1. Because K07F5.14 is synthetic sterile with met-2 and *met-2 set-25*, RNA was isolated from young adults of the F0 generation. Relative expression was measured by qPCR of three tandem repeats, two transposons that are transcriptionally derepressed in *met-2 set-25* embryos and one unaffected gene as a negative control (Figure 4). Silencing of all 3 BRC-1 associated proteins in wt worms led to the specific derepression of tandem repeats, while it had no effect on transposable elements or genes, indicating that the role of BRCA1 in tandem repeat silencing is conserved from mammals to worms.

Figure 4 MET-2 and BRC-1 redundantly control tandem repeat repression



QPCR analysis of tandem repeats, transposon and the smg-5 gene expression at the young adult stage. Worms of indicated genetic background (wt=N2, set-25, met-2 and met-2 set-25) were treated from the L1 larval stage on with RNAi against brca-1 and it's interaction partners brd-1 and K07F5.14. Expression is normalized to wt animals grown on empty vector control expressing bacteria (N=2).

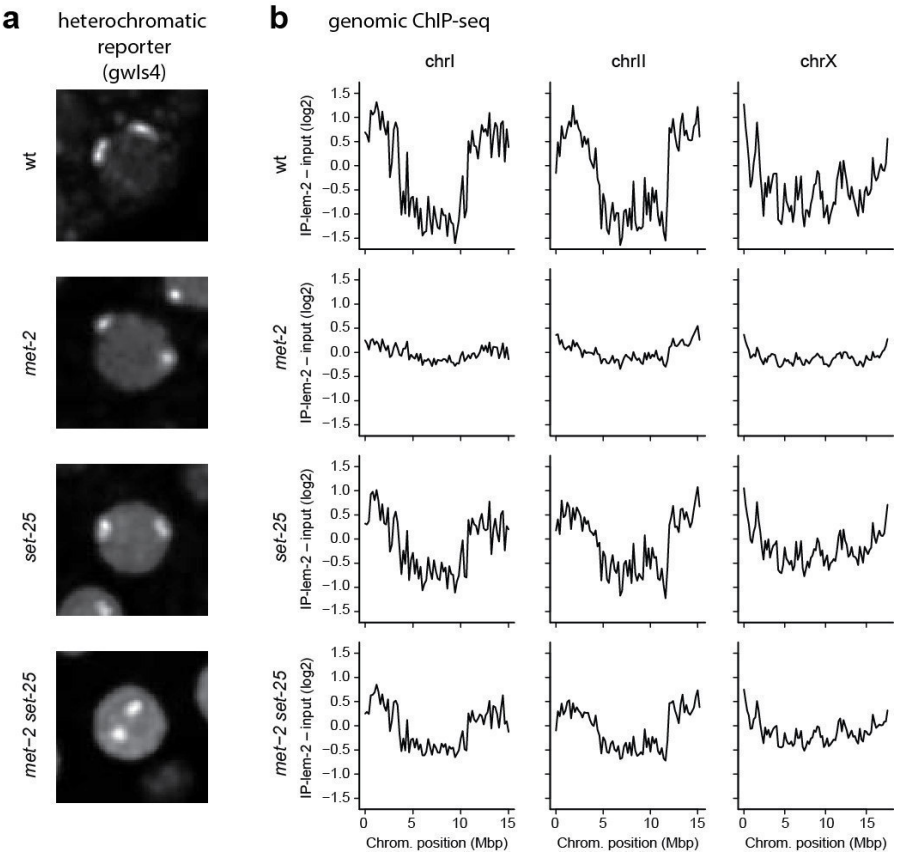
Strikingly, depletion of each of the three BRC-1 associated proteins in *met-2* and *met-2 set-25* animals led to an additive effect specifically in tandem repeat expression. Transcription of H3K9 methylated transposable elements and genes did not increase upon BRC-1, BRD-1 or K07F5.14 knockdown beyond that provoked by *set-25* or *met-2* mutation alone. We conclude that MET-2 and BRC-1 complex-mediated silencing of tandem repeats occurs in a partially redundant manner.

H3K9me2 is able to establish normal peripheral organization in the absence of H3K9me3

We have previously ascribed a function in heterochromatin localization at the nuclear envelope to both H3K9me2 and H3K9me3 (Towbin et al. 2012). This was monitored with a fluorescently tagged, artificial array of genes, bearing a lacO site, which binds lacI-GFP. This array lost its perinuclear localization in *set-25 met-2* mutants, but was anchored peripherally in either single mutant (Figure 5a). Given the parallels of the *met-2* phenotypes with those of the *met-2 set-25* mutant, we examined the perinuclear localization of chromatin genome-wide, using LEM-2 ChIP (Ikegami et al. 2010), to see whether we could score specific roles for H3K9me3 (SET-25 function) in chromatin positioning, as opposed to H3K9me2 (MET-2). To our surprise, the loss of *met-2* caused a complete loss of perinuclear anchoring, even more pronounced than the *met-2 set-25* mutant, while no obvious effect could be observed in the *set-25* single mutant (Figure 5b). This argues that perinuclear anchoring may be primarily mediated by H3K9me2, and that peripheral sequestration may contribute to H3K9me2 mediated silencing.

Indeed, comparing the LEM-2 enrichment under wild type conditions and the transcriptional de-repression in the single mutants, we can see clear correlation for MET-2 dependent, but not SET-25 dependent regions (Figure 6a). This would argue that SET-25 is able to silence RE independent of their nuclear organization relative to the nuclear envelope. Indeed, when we checked the distribution of the SET-25 dependent loci over the chromosome, they were distributed all over with no particular enrichment on the repeat rich chromosome arms (Figure 6b) (Gerstein et al. 2010).

Figure 5 MET-2, but not SET-25, is necessary for the association of the majority of endogenous heterochromatin with the nuclear envelope

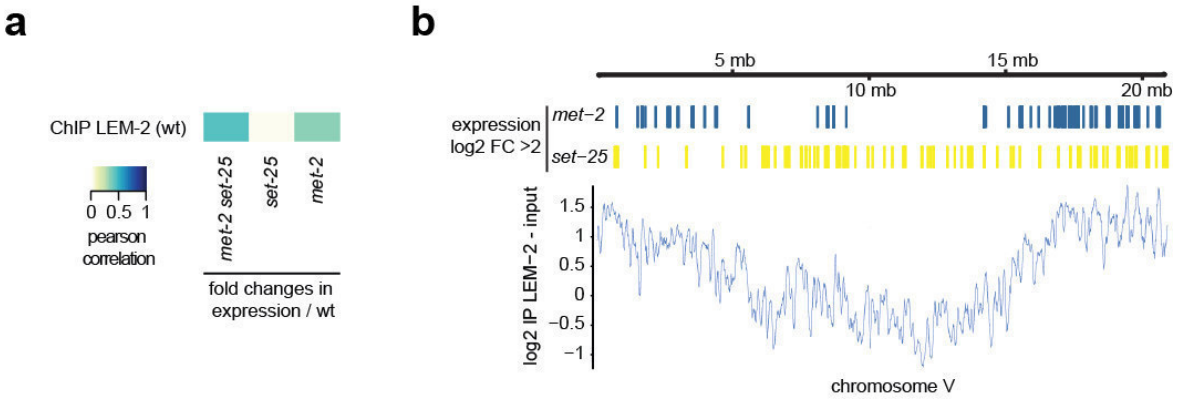


(a) Example pictures of fluorescent microscopy of early embryos of indicated genotypes showing the nuclear localization of the heterochromatic reporter array gwIs-4, which by IF is covered by H3K9me3 and H3K27me3, but no active marks (Towbin et al. 2012). **(b)** LEM-2 ChIP-seq performed in early embryos of the indicated genotypes, grown at 20°C. Graph shows the distribution of LEM-2 signal as a marker for perinuclear positioning over input along chromosome 1,2 and X (N = 2).

In sum, we could show that in the absence of H3K9me3, H3K9me2 fulfills the majority of the functionality ascribed to H3K9 methylation: transcriptional silencing, prevention of the majority of RNA:DNA hybrid accumulation and peripheral anchoring of heterochromatin. At the same time, we do note that there is a specific subset of genes and transposable elements that specifically do depend on SET-25 and H3K9me3, possibly, because they cannot be targeted by MET-2, which seems to be restricted to chromosomal regions that can be tethered to the nuclear envelope (Figure 7).

The replication stress sensitivity we can observe in both *set-25* and *met-2* either points to the fact that already few replication obstacles can be detrimental in combination with replication stress, or an additional role of H3K9me in replication stress response. Furthermore the hits identified in the synthetic lethality screen strengthen our proposition that prevention of RNA:DNA hybrid driven mutagenesis is one of the major functions of H3K9 methylation.

Figure 6 Silencing dependent on MET-2 occurs at the nuclear periphery, while SET-25 dependent silencing occurs also away from the periphery

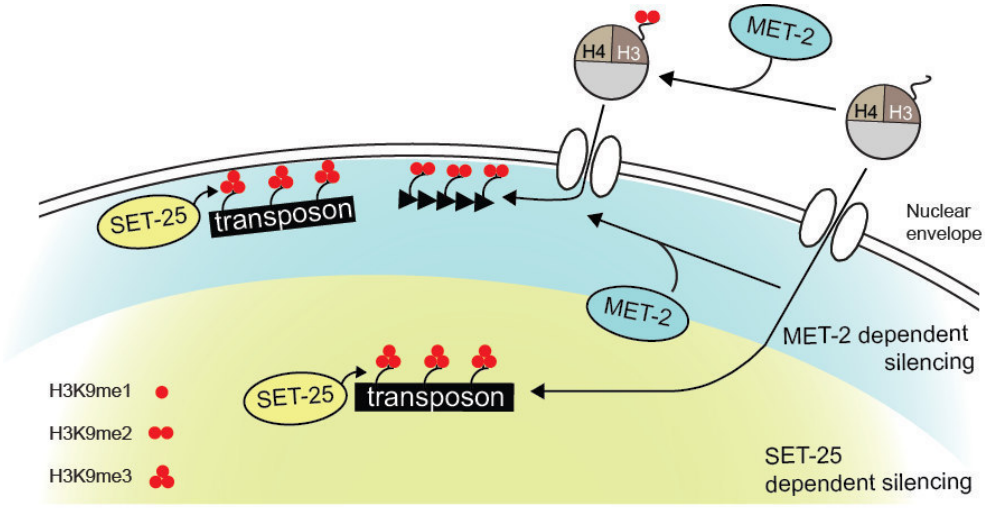


(a) Correlation of the wild type LEM-2 ChIP enrichment with the fold change expression over wt observed in *met-2 set-25* double mutants as well as *set-25* and *met-2* single mutants. Heat map visualizes the Pearson correlation coefficient. **(b)** Visualization of the distribution of genomic region derepressed in *met-2* or *set-25* mutants over chromosome V. Bars indicate regions at least 2 fold derepressed over wt. LEM-2 ChIP enrichment is indicated the chromosome arms tethered to the nuclear periphery.

Additionally, this work highlights the danger of tandem repeat expression, which is the only sequence type exclusively affected by loss of MET-2 and not SET-25. In our previous work focusing on the complete loss of H3K9me we could also show a prominent derepression of tandem repeats in embryos at elevated temperature and in somatic cells compared to the germline (Zeller et al. 2016). This correlated with an increase in RNA:DNA hybrid accumulation accompanied with elevated rates of DNA damage induced germ cell apoptosis at elevated temperature as well as an increased mutation rates of a LacZ-reporter in somatic vs germline cells. Originally, tandem repeats were mainly considered as functionally and structurally important sequences, especially at the telomere (de Lange 2009) and centromere (Haaf et al. 1992), but recent studies also report the potential danger of their derepression (Nakamori et al. 2010). Results by Zhu and colleagues on BRCA1 in mammals (Zhu et al. 2011) as well as our results presented here and in our previous work (Zeller et al. 2016) suggest that transcriptional control of tandem repeats is crucial for the preservation of genome integrity. As BRCA1 deletion was associated with changes in the nuclear distribution of HP1 foci, as well as loss of histone H2A monoubiquitination it was unclear if BRCA1 mediates its silencing role via the H3K9 methylation machinery.

Our results suggest that MET-2 and BRC-1 regulate tandem repeat expression in a partially redundant manner, providing two layers for the protection of the genomes integrity.

Figure 7 Spatial separation of MET-2 and SET-25 dependent silencing



Model visualizing the 3 different pathways to target H3K9me mediated silencing and its localization inside the nucleus. Direct targeting of SET-25 to transposable elements happens independent of the nuclear localization. MET-2 dependent silencing is mainly found close to the nuclear envelope. In can include the additional tri-methylation by SET-25 in the case of transposons or stay in a di-methylated form for tandem repeats.

Methods

C. elegans cultures and strains

Worms were grown on OP50 at 20°C where not otherwise stated. Worms used in this study come from the *c. elegans* knockout consortium or from the Ciosk laboratory and were at least 6 times backcrossed to wt worms. Supplementary table 3 contains a full description of used animals.

Immunofluorescence (IF), antibodies and live microscopy

RNA:DNA hybrid IF was carried out as previously described by freeze-cracking and fixation in methanol followed by acetone. Staining was performed in 4XSSC-T (0.1% Tween-20) and 3% milk BSA. Primary antibody (S9.6, HB-8730, ATCC, gift of P. Pasero, Montpellier) was incubated was performed at a dilution of 1:100 in 4XSSC-T and 3% BSA overnight at 4°C. For live cell imaging larvae were mounted on slides coated with 2% agarose. Microscopy was carried out on a spinning disc confocal microscope (Yokogawa X1 and a Yokogawa W1 scan-head mounted on an Olympus and a Zeiss microscope respectively, Visitron, Puchheim). Stacks of images were analyzed using ImageJ.

Developmental timing and progeny size

For experiments performed at 25°C, animals were adopted for 2 generations to the new temperature and the F3 generations were used for the experiment.

For developmental timing experiments single worms were followed over 3 days of somatic development starting with stage 1 larvae synchronized by starvation. The developmental stage was determined by eye and noted down every 24h.

For progeny size experiments worms were singled as stage 4 larvae. Adults were transferred to fresh plates to keep generations separated and the total number of viable F1 progeny was determined.

RNAi screen

For the primary screen worms were grown in liquid culture (NGM containing NaCl, peptone, CaCl2, MgSO4, KH2PO4 Cholesterol, Carb, IPTG) containing RNAi expressing bacteria of the Ahringer library in 384 well plates from Greiner (781 097) allowing for direct automated microscopy. After incubation of 4d at 20°C in a humid chamber worms were anesthetized by adding Levamisol and imaged on a GE inCell automated microscope. Rescreening was performed in 96 well format adjusting volumes appropriately using sequenced pure RNAi clones from the Ahringer library and where possible also the Vidal RNAi library.

Chromatin Immunoprecipitation (ChIP) experiments

Chromatin was isolated from Early embryos grown after L1 synchronization (60–65 h depending on each strain) in two independent biological replicates. Lem-2 ChIP was performed as previously described (Ikegami et al. 2010) using an antibody from Novus Biologicals (48540002). In brief, 20µg of chromatin was incubated overnight with 3 µg of antibody coupled to Dynabeads Sheep Anti-Rabbit IgG (Invitrogen), in FA-buffer (50 mM HEPES/KOH pH7.5, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 150 mM NaCl)) containing 1% SDS. Chromatin/ antibody complexes were washed with the following buffers: 3 x 5min FA buffer; 5min FA buffer with 1M NaCl; 10 min FA buffer with 500 mM NaCl; 5min with TEL buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) and twice for 5 min with TE. Complexes were eluted at 65°C in 100 µl of elution buffer (1% SDS in TE with 250 mM NaCl) for 15 min. Both input and IP samples were incubated with 20 µg of RNAse A for 30 minutes at 37°C and 20 µg of proteinase K for 1 h at 55°C. Crosslinks were reversed overnight at 65°C. DNA was purified using a Zymo DNA purification column (Zymo Research).

Library preparation and analysis

Libraries were prepared from chromatin IP and genomic DNA samples using the NEBNext ultra DNA library prep kit for Illumina (NEB # 7370) and the NEBNext Multiplex Oligos for Illumina (NEB # E7335), according to the manufacturer's recommendations. No size selection was performed during sample preparation and the libraries were indexed and amplified using 12 PCR cycles, using the recommended conditions. After further purification with Agencourt AmPure XP beads

(Beckman # A63881), the library size distribution and concentrations were determined using a BioAnalyzer 2100 (Agilent technologies) and Qubit (Invitrogen) instrument, respectively. The final pools were prepared by mixing equimolar amounts of all individually indexed libraries and then sequenced on a HiSeq 2500 (Illumina) in Rapid mode (Paired-End 50). Paired-end ChIP-seq data (2x50bp) were mapped to the *C. elegans* genome (ce6) with the R package QuasR (Gaidatzis et al. 2015) using the included aligner bowtie (Langmead et al. 2009). Definitions of REs were taken from Repbase(Jurka et al. 2005).

Read density along the genome was calculated by tiling the genome into 200bp windows (non-overlapping) and counting the number of sequence fragments within each window, using the qCount function of the QuasR package. To compensate for differences in the read depths, libraries were normalized to the total number of reads per library. ChIP-seq signals are displayed as average enrichment of IP – input (log2).

RNA expression experiments (RNAseq and qPCR)

RNA was isolated from early embryos by freeze cracking (4 times) followed by phenol/chloroform extraction and isopropanol precipitation. Total RNA was depleted for rRNA using Ribo-Zero Gold kit from Epicentre before library production using Total RNA Sequencing ScriptSeq kit. 50 bp single-end sequencing was done on an Illumina HiSeq 2500. Processing of the RNA-seq data, gene and repeat expression levels from RNA-seq data were quantified as described previously [9] using WormBase (WS190) annotation for coding transcripts and Repbase annotations for repetitive elements. Log2 expression levels were calculated after adding a pseudocount of 1 (y=log2(x+1)). Repeat sub-families were built to allow assignment of multimapping reads to all repetitive elements and collapsing single elements according to their Repbase ID into families.

DNA damage sensitivity assays HU

Assays were previously described (Craig et al. 2012). Recovery from an HU pulse was monitored by soaking larvae in M9 buffer containing indicated concentrations of HU and OP50 bacteria for 16h before washing and plating on fresh OP50 plates. For somatic HU sensitivity stage 1 larvae were exposed to HU and the percentage of viable adults was quantified at day 3.

Acknowledgments

A number of strains were provided by the Caenorhabditis Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We thank Juliet Leighton-Davies and Christian Parker from NIBR, Novartis for their help with the RNAi screen, R. Ciosk for reagents and materials, I. Katiç and members of the Friedrich Miescher Institute Genomics and

Microscopy facilities for advice and discussion. J.P. is supported by a long-term EMBO fellowship. S.M.G. thanks the Swiss National Science Foundation as well as the Novartis Research Foundation for support.

Disclosure declaration

The authors declare no competing financial interests.

References

Anderson SF, Schlegel BP, Nakajima T, Wolpin ES, Parvin JD. 1998. BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nat Genet* **19**: 254-256.

Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*.

Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* **89**: 123-130.

Boulton SJ, Martin JS, Polanowska J, Hill DE, Gartner A, Vidal M. 2004. BRCA1/BARD1 Orthologs Required for DNA Repair in Caenorhabditis elegans. *Current Biology* **14**: 33-39.

Chapman MS, Verma IM. 1996. Transcriptional activation by BRCA1. *Nature* **382**: 678-679.

Craig AL, Moser SC, Bailly AP, Gartner A. 2012. Methods for studying the DNA damage response in the Caenorhabdatis elegans germ line. *Methods Cell Biol* **107**: 321-352.

de Lange T. 2009. How telomeres solve the end-protection problem. *Science* **326**: 948-952.

El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* **24**: 1546-1558.

Gaidatzis D, Lerch A, Hahne F, Stadler MB. 2015. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**: 1130-1132.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**: 1775-1787.

Haaf T, Warburton PE, Willard HF. 1992. Integration of human α-satellite DNA into simian chromosomes: Centromere protein binding and disruption of normal chromosome segregation. *Cell* **70**: 681-696.

Ikegami K, Egelhofer TA, Strome S, Lieb JD. 2010. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome biology* **11**: 1.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**: 462-467.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M et al. 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**: 231-237.

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

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: 1.

Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T et al. 2004. A map of the interactome network of the metazoan C. elegans. *Science* **303**: 540-543.

Moynahan ME, Chiu JW, Koller BH, Jasin M. 1999. Brca1 Controls Homology-Directed DNA Repair. *Molecular Cell* **4**: 511-518.

Nakamori M, Pearson CE, Thornton CA. 2010. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. *Human Molecular Genetics* **20**: 580-588.

Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu Z-P. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**: 774-789.

Polanowska J, Martin JS, Garcia-Muse T, Petalcorin MI, Boulton SJ. 2006. A conserved pathway to activate BRCA1-dependent ubiquitylation at DNA damage sites. *Embo j* **25**: 2178-2188.

Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S et al. 2004. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res* **14**: 2162-2168.

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**: 583-597.

Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, Livingston DM. 1997. Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* **88**: 265-275.

Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**: 934-947.

Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA: DNA hybrids from generating genome instability. *Molecular cell* **44**: 978-988.

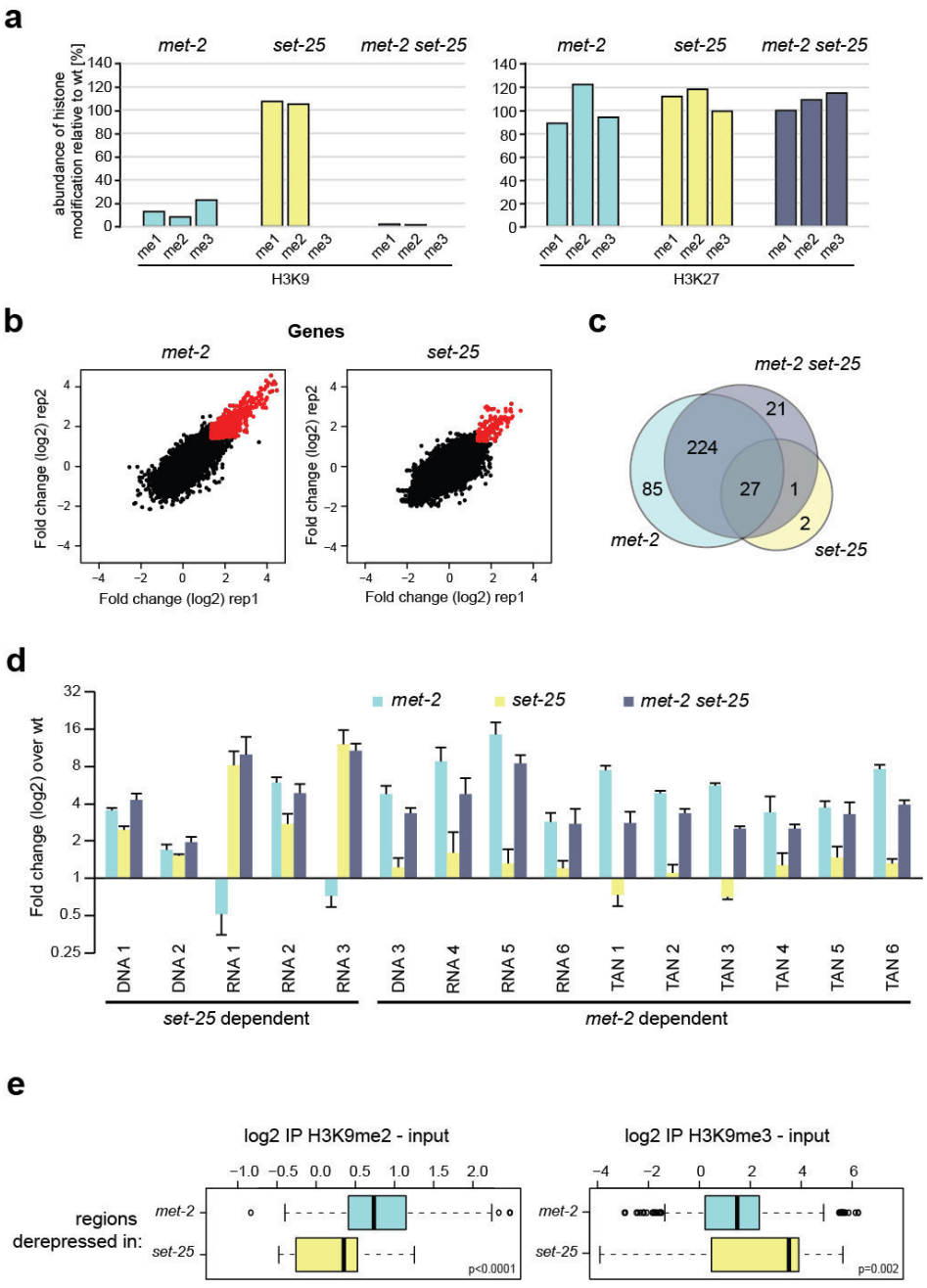
Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Molecular cell* **53**: 484-497.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48**: 1385-1395.

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM. 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**: 179-184.

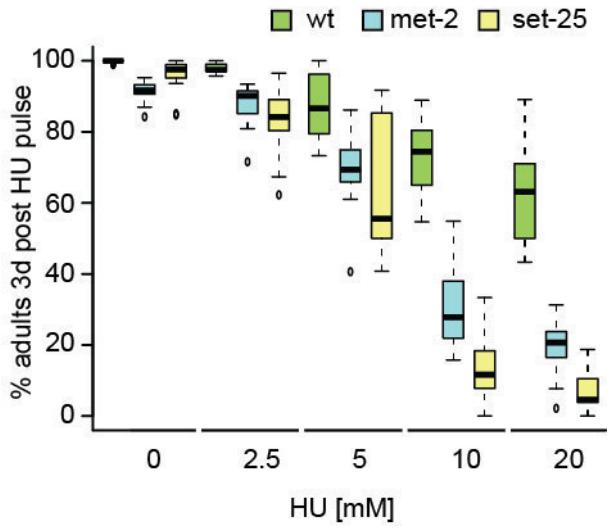
Supplementary Information

Supplementary Fig 1 Transcriptional derepression in the absence of MET-2 or SET-25



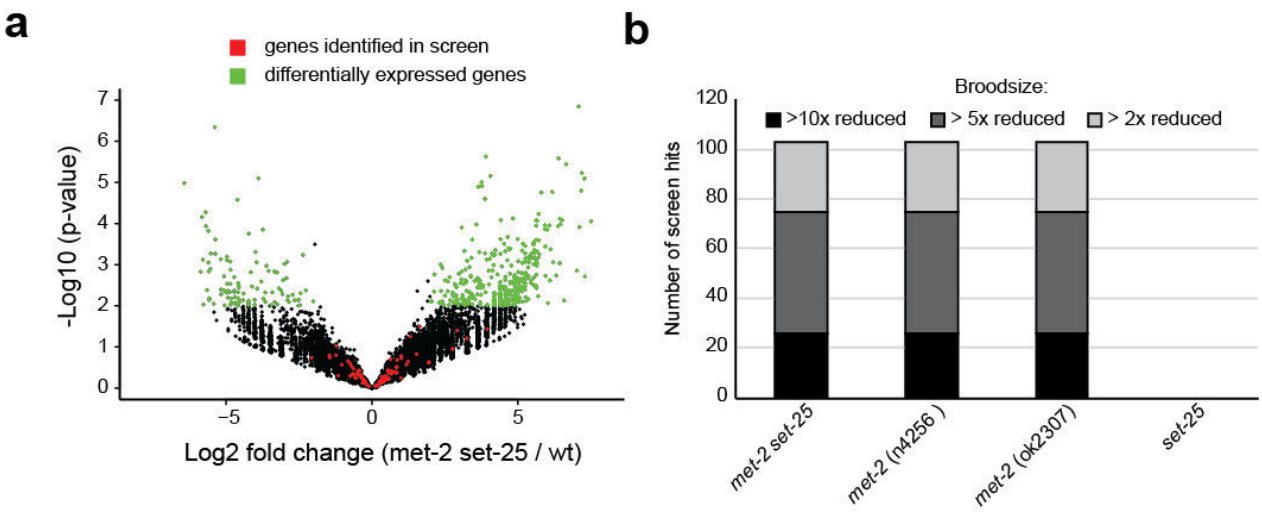
(a) Mass spec quantification of the H3K9me and H3K27me state of *met-2, set-25* or double *met-2-set-25-*mutant used in this publication at early embryonic stage (Values are relative to wt levels (Towbin et al. 2012)). **(b)** Scatter blot showing gene expression as fold change over wild-type early embryos of indicated single mutants (2 replicas shown, N (number of biological replicates) = 3). **(c)** Venn diagram showing genes commonly or uniquely regulated by MET-2 and SET-25 (N = 3). **(d)** qPCR confirmation of the strongest derepressed single repeats and their dependency on *met-2* and *set-25* for silencing in early embryos (N= 3). **(e)** Box-blot, showing the log2 fold enrichment of H3K9me2 and me3 ChIP-seq signal on genomic regions significantly derepressed in *met-2* or *set-25* single mutants.

a



Supplementary Fig 2 HU sensitivity of mutant larvae

(a) Sensitivity to replication stress assayed by recovery after a 16h exposure to hydroxyl urea (HU). somatic cell sensitivity was assayed by exposing stage 1 larvae (L1) to indicated concentrations of HU and quantifying the percentage of worms still developing until adulthood (N = 3, n = 5).



Supplementary Fig 3 MET-2 deposited H3K9me2 is sufficient to protect from the observed synthetic lethality

(a) Vulcano blot displaying transcriptional changes in gonads dissected from *met-2 set-25* mutants relative to wt. Green marks significantly differentially expressed genes (Log2 fold change > 2, -Log10(p-value) > 2) Red marks genes whose downregulation by RNA is synthetic lethal with loss of H3K9me (N= 3). **(b)** The complete list of RNAi clones synthetic lethal with *met-2 set-25* double mutants was retested on 2 *met-2* and 1 *set-25* single mutant. The total number of synthetic RNAi clones is indicated; The fill color indicates the strength of the synthetic effect as fold reduction in progeny numbers (N=3).

Wormbase ID	Gene name	general process	description	human homologue	hit strength
WBGene00001912	his-38	Histone / chromatin structure	Histone H4	HIST1H4H (100 %)	>10X
WBGene00001945	his-71	Histone / chromatin structure	Histone H3	H3F3B (100 %)	>10X
WBGene00001905	his-31	Histone / chromatin structure	Histone H4	HIST1H4H (100 %)	>10X
WBGene00002344	let-70	Histone / chromatin structure	Class I E2 ubiquitin conjugating enzyme.	UBE2D2 (100%)	>5X
WBGene00001884	his-10	Histone / chromatin structure	Histone H4	HIST1H4H (100 %)	>5X
WBGene00001886	his-12	Histone / chromatin structure	Histone H2A	HIST2H2AB (99%)	>2X
WBGene00007433	swsn-7	chromatin remodeller	SWI/SNF nucleosome remodeling complex component	ARID2 (66%)	>10X
WBGene00011111	snfc-5	chromatin remodeller	homolog of SNF5/Ini1, part of the SWI/SNF nucleosome remodeling complex	SMARCB1 (96%)	>10X
WBGene00007029	mys-1	chromatin remodeller	central unit of NuA4 histone acetyl transferase complex	KAT5 (93%)	>10X
WBGene00007030	epc-1	chromatin remodeller	Enhancer of PolyComb-like. Part of NuA4 histone acetyl transferase complex	EPC2 (56%)	>5X
WBGene00013766	prmt-1	chromatin remodeller	Arginine methyl transferase implicated in RNA processing, signal transduction, DNA repair and transcriptional regulation.	HRMT1L2 (95%)	>5X
WBGene00001972	hmg-1.2	chromatin remodeller	High mobility group transcription factor	HMGB2 (74%)	>5X
WBGene00022182	swsn-3	chromatin remodeller	Part of swi/snf complex predicted interaction with snfc-5	SMARCE1 (94%)	>2X
WBGene00000938	dcp-66	chromatin remodeller	Predicted member of the transcriptionally inhibitory nucleosomal remodelling and deacetylase (NuRD) complex.	GATAD2A (54%)	>2X
WBGene00010638	K07F5.14	DNA damage response	interacts with brc-1 (dsb repair)	NOL8 (54%)	>10X
WBGene00015075	B0238.11	DNA damage response	developemental gene	-	>10X

WBGene00016944	uri-1	DNA damage response	Molecular chaperone required for suppressing endogenous genotoxic DNA damage and for development of some somatic structures, such as the vulva, and for RNA interference.	URI1 (93%)	>2X
WBGene00022851	ZK1127.4	DNA damage response	Isoform 3 of BRCA2 and CDKN1A-interacting protein	BCCIP (71%)	>2X
WBGene00010325	exos-3	RNA degradation, DNA damage response	EXOSome (multiexonuclease complex) component	EXOSC3 (77%)	>10X
WBGene00008363	D1046.2	genaral transcription	Zinc finger domain containing protein	-	>10X
WBGene00007017	mdt-17	genaral transcription	Part of the mediator complex	MED17 (82%)	>10X
WBGene00013998	ZK550.4	genaral transcription	general transcription factor TFIIE-alpha	GTF2E1 (82%)	>5X
WBGene00019275	rpac-40	genaral transcription	RNA Polymerase I/III (A/C) shared subunit	POLR1C (94%)	>5X
WBGene00000271	brf-1	genaral transcription	Molecular bridge between TATA-binding protein (TBP)and associated factors that functions in TFIIIB (RNA polymerase III associated transcription factors).	BRF1 (77%)	>2X
WBGene00006542	tbp-1	genaral transcription	TATA-box-binding protein that provides TFIID-like basal transcription activity in human and C. elegans extracts.	TBP-201 (89 %)	>2X
WBGene00014938	mdt-9	genaral transcription	Part of the mediator complex	RABIF (78 %)	>2X
WBGene00006392	taf-10	general transcription	Subunit of TFIID and SAGA complexes, involved in RNA polymerase II transcription initiation and in chromatin modification	TAF10 (55%)	>2X
WBGene00002148	gon-14	Transcription factor	exhibits receptor binding activity and is predicted to have nucleic acid binding activity and protein dimerization activity.	-	>10X
WBGene00018794	attf-3	transcription factor	AT hook Transcription Factor family	HRC (21%)	>5X
WBGene00006386	taf-5	transcription factor	TAF (TBP-associated transcription factor[TATA box binding protein]) family member	TAF5 (93 %)	>5X
WBGene00002696	let-504	Transcription factor	Transcriptional repressor that associates with the NOTCH co-repressor complex.	NKAP (85%)	>5X

WBGene00017742	nfyc-1	Transcription factor	Nuclear transcription Factor Y, C (gamma) subunit	NFYC (47%)	>5X
WBGene00003605	nhr-6	Transcription factor	Nuclear hormone receptor of the NR4A subfamily that can bind the canonical vertebrate NR4A binding sequence.	NR4A2 (60%)	>5X
WBGene00003623	nhr-25	Transcription factor	Nuclear hormone receptor	NR5A2 (54%)	>5X
WBGene00002245	lag-1	Transcription factor	Transcription factor downstream in both the lin-12- and glp-1-mediated Notch-like signaling pathways.	RBPJ (50%)	>2X
WBGene00000466	cel-1	RNA processing	mRNA capping enzyme	RNGGT (92%)	>10X
WBGene00007972	prp-4	RNA processing	part of U4/U6 x U5 tri-snRNP complex	PRPF4 (90%)	>10X
WBGene00015513	snpc-1.1	RNA processing	best human homolo snRNA-activating protein complex subunit 1	SNAPC1 (59%)	>10X
WBGene00011367	snpc-3.4	RNA processing	small nuclear RNA activating complex, polypeptide 3	SNAPC3 (49%)	>5X
WBGene00015104	B0280.9	RNA processing	Ortholog of small subunit (SSU) processome component	UTP18 (98%)	>5X
WBGene00020705	T22H9.1	RNA processing	ribosomal RNA processing	RRP36 (54%)	>5X
WBGene00013343	prp-6	RNA processing	pre-mRNA processing factor	PRPF6 (95%)	>5X
WBGene00000773	cpf-1	RNA processing	human homologe is a pre-mRNA cleavage stimulating factor	CSTF-1 (99 %)	>2X
WBGene00017746	F23F1.5	RNA processing	Functions as an U snRNP-specific nuclear import adapter.	SNUPN (78%)	>2X
WBGene00019264	H35B03.2	RNA processing	Nucleic acid binding activity and zinc ion binding activity, based on protein domain information.	SNRPC (100%)	>2X
WBGene00017238	F08B4.7	RNA processing	Component of the U1 snRNP complex required for pre-mRNA splicing.	SNRPC (100%)	>2X
WBGene00020827	T26A8.4	RNA degradation	Component of the conserved Ccr4-Not deadenylase complex that regulates transcription and RNA degradation;	ZC3H4 (71%)	>10X
WBGene00015486	C05D11.9	RNA degradation	Protein subunit shared by the endoribonucleases RNase MRP and RNAse P.	POP1 (62%)	>10X

WBGene00009289	exos-7	RNA degradation	EXOSome (multiexonuclease complex) component	EXOSC7 (78%)	>5X
WBGene00002804	let-630	RNA degradation	-	PTCD3 (35%)	>5X
WBGene00003218	mep-1	RNA degradation	mep-1 is required for post transcriptional repression by the fem-3 3' UTR.	FOXP2 (22%)	>5X
WBGene00015499	inst-1	RNA degradation	Integrator complex subunit 1	INST1 (40%)	>2X
WBGene00003009	lin-23	RNA degradation	lin-23 encodes an F-box- and WD-repeat-containing protein, components of SCF (Skp1, Cullin, F-box) ubiquitin-ligase complexes that function in ubiquitin-mediated protein degradation.	INTS1 (40%)	>2X
WBGene00003062	lpd-6	translation	predicted rRNA-binding protein required for fat storage and for larval growth and development	PPAN (90%)	>5X
WBGene00004413	rpl-2	translation	ribosomal protein large	RPL8 (100%)	>5X
WBGene00006935	vars-1	translation	Valyl-tRNA synthetase	VARS2 (81%)	>5X
WBGene00013958	ZK265.6	translation	Ribosome biogenesis protein Nop16	NOP16 (48%)	>2X
WBGene00009246	gfm-1	translation	Mitochondrial translation elongation factor G1 (EFG1).	GFM1 (100%)	>2X
WBGene00019693	ostd-1	posttranslational modification	Oligo Saccharyl Transferase	RPN2 (89%)	>5X
WBGene00010700	nipi-3	posttranslational modification	Kinase with functions in the innate immune response.	TRIB1 (26%)	>5X
WBGene00010562	cdc-48.3	posttranslational modification	spermatogenesis associated protein involved in negative regulation of protein kinase activity	SPATA5 (50%)	>5X
WBGene00018961	mrps-16	mitochondrial translation	mitochondrial ribosomal protein, small	MRPS16 (87%)	>10X
WBGene00010458	mrpl-10	mitochondrial translation	Mitochondrial Ribosomal Protein, Large	-	>5X
WBGene00020348	mrpl-23	mitochondrial translation	mitochondrial ribosomal protein large	MRPL23 (93%)	>5X
WBGene00020037	mrps-6	mitochondrial translation	Mitochondrial Ribosomal Protein, Small	MRPS6 (74%)	>5X

WBGene00015133	mrpl-11	mitochondrial translation	Mitochondrial Ribosomal Protein, Large	MRPL11 (97%)	>5X
WBGene00011740	mrpl-51	mitochondrial translation	Mitochondrial Ribosomal Protein, Large	MRPL51 (53%)	>5X
WBGene00012244	mrps-11	mitochondrial translation	Mitochondrial Ribosomal Protein, Small	MRPS11 (65%)	>5X
WBGene00022373	mrpl-15	mitochondrial translation	mrpl-15	MRPL15 (94%)	>2X
WBGene00016142	mrps-18c	mitochondrial translation	Mitochondrial Ribosomal Protein, Small	MRPL3 (62%)	>2X
WBGene00009013	mrps-33	mitochondrial translation	Mitochondrial Ribosomal Protein, Small	MRPS33 (87%)	>2X
WBGene00010809	M01F1.3	mitochondrial	lipoic acid synthase	LIAS (98%)	>10X
WBGene00010042	bcs-1	mitochondrial	BCS1 (mitochondrial chaperone) homolog	BCS1L (94%)	>10X
WBGene00007880	C33A12.1	mitochondrial	NADH dehydrogenase and a few other ones	NDUFA5 (74%)	>10X
WBGene00008147	C47E12.2	mitochondrial	mitochondrial carrier	SLC25A31 (95%)	>10X
WBGene00012094	T27E9.2	mitochondrial	ubiquinol-cytochrome c reductase complex	UQCRH (95%)	>10X
WBGene00003066	lpl-1	mitochondrial	Lipoate ligase homolog	MRPL42 (81%)	>5X
WBGene00019588	K09F6.5	mitochondrial	deoxyribonucleoside diphosphate metabolic process.	-	>5X
WBGene00017508	F16B4.6	mitochondrial	NADH dehydrogenase	NDUFAB1	>5X
WBGene00003061	lpd-5	mitochondrial	part of NADH dehydrogenase	NDUFS4 (73%)	>2X
WBGene00007684	C18E9.4	mitochondrial	part of NADH dehydrogenase	-	>2X
WBGene00013237	Y56A3A.19	mitochondrial	part of NADH dehydrogenase	NDUFAB1 (99%)	>2X
WBGene00009179	kbp-3	mitosis	KNL (kinetochore null) Binding Protein	KIF5C (87%)	>5X
WBGene00006540	tbg-1	mitosis	Gamma-tubulin required for normal bipolar spindle assembly and function.	TUBG1 (98%)	>5X
WBGene00000382	cdc-6	mitosis	Homolog of an origin complex component (CDC6) which controls the start of DNA replication	CDC6 (58%)	>5X
WBGene00022631	nekl-2	mitosis	nekl-2 encodes a putative serine/threonine protein kinase interacts with components of the EGF/RAS signalling pathway.	NEK8 (84%)	>5X
WBGene00010785	top-2	mitosis	Topoisomerase 2	TOP2A (97%)	>2X

WBGene00004953	spd-2	mitosis	Protein essential for centrosome maturation and duplication.	-	>2X
WBGene00004726	sas-4	mitosis	Centriole component required for centriole duplication and spindle assembly.	FILIP1 (7%)	>2X
WBGene00018670	F52C6.13	transport	-	-	>10X
WBGene00017284	sec-6	transport	exocyst complex component	EXOC3 (94%)	>10X
WBGene00010047	F54D1.6	transport	Transmemebrane protein	SUSD2 (21%)	>2X
WBGene00001333	erm-1	cytoskeleton	Cytoskeletal linker	MSN (99%)	>5X
WBGene00016021	C23H3.5	-	developemental gene	-	>10X
WBGene00009365	F33H1.4	-	protein involved in reproduction	TTN (35%)	>5X
WBGene00010044	F54C9.9	-	-	KRI1 (94%)	>5X
WBGene00019400	K04G7.1	-	interacts (pooled two hybrid) with ztf-18 (zink finger protein) and mpk-1 (map kinase)	-	>5X
WBGene00000502	chp-1	-	Protein containing two CHORD domains.	CHORDC1 (98%)	>5X
WBGene00019759	M03F4.6	-	-	TNXB (79%)	>5X
WBGene00019353	K03B4.1	-	-	-	>5X
WBGene00017853	F27C1.3	-	-	-	>5X
WBGene00017355	F10E9.3	-	-	-	>5X
WBGene00019212	H19M22.3	-	metalloendopeptidase	MMP14 (34%)	>5X

supplementary table 1 Complete list of RNAi targets synthetic lethal with loss of H3K9me

Hit strength refers to the fold change difference in F1 progeny number vs wild type.

Larval development *cel-1*, *his-38*, *his71*, *swsn-7*, *epc-1*, T27E9.2, T26A8.4
Gonad formation M01F1.3, *mdt-17*, *snpc-1.1*, *prp-6*, F27C1.3, *let-70*, F54C9.9, *mrlp-23*, *taf-5*, D1046.2, F33H1.4, *kbp-3*, B0280.9, *lpl-1*
Germ cell maturation / number *vars-1*, *his-10*, HMG-1.2, K04G7.1, *let-630*, *bcs-1*, C05D11.9, C23H3.5, *let-504*, *mrps-6*, *nfyc-1*, *mrpl-10*, *mrpl-11*, F16B4.6, *rpac-40*, *rpl-2*, *exos-7*, *ostd-1*, *cdc-48.3*, *lpd-6*, *snfc-5*, *prmt-1*
Oocyte fertilization / egg number *exos-3*, C33A12.1, F52C6.13, *sec-6*, *mep-1*, *mrpl-51*, *snpc-3.4*, C47E12.2, ZK550.4, *mys-1*
Egg hatching K07F5.14

supplementary table 2 categorization of the 55 strongest RNAi clones dependent on the earliest developmental stage they impair.

Name in manuscript	Genotype	Strain designation	Reference
Wild-type (wt)		N2 (Bristol)	
met-2 set-25	*met-2(n4256) III; set-25(n5021) III*	GW638	(Towbin et al. 2012)
met-2 (used in most experiments)	*met-2(n4256) III*	GW907	(Andersen and Horvitz 2007)
met-2 (second allele used to verify screen hits)	*met-2(ok2307) III*	GW502	(Consortium 2012)
set-25	*set-25(n5021) III*	GW641	(Andersen and Horvitz 2007)
mex-5:h2b::gfp	*unc+119;rrrSi192[mex-5 prom::gfp: h2b::tbb-2 3'UTR;unc119(+)] II*	gift of R. Ciosk, FMI	(Tocchini et al. 2014)
met-2 set-25 mex-5:h2b::gfp	*met-2(n4256) III; set-25(n5021) III; unc+119; rrrSi192[mex-5 prom::gfp: h2b::tbb-2 3'UTR;unc119(+)] II*	GW1028	(Zeller et al. 2016)

Supplementary Table 3

Worm strains used in this study

			Forward		Reverse	
Supplementary Figure 1d						
Label	Repeat	Genomic position	Name	Sequence	Name	Sequence
DNA 1	CEMUDR1	chrV:14,200,466 –14,205,122	SG-7490	AGGCCCATTCCATAGCTTTT	SG-7491	TTTCTGGGATTTCATGCACA
DNA 2	CEMUDR1	chrIV:3,314,561 – 3,317,596	SG-7494	CCGTTATCGTGACAAGCAGA	SG-7495	AAGCTTGAGCCGCTGATTAC
DNA 3	CeRep4	chrIII:2,305,450 – 2,305,516	SG-7521	CACGTTCCCAGCTTCAATTT	SG-7522	TGCAAATCTACAGTAACCCAGAA
RNA 1	CER10-I_CE	chrV:8,480,088 – 8,481,470	SG-7498	TACCAACGAGCCGAGTCTTC	SG-7499	TCTTCAGTTTCCTCGCCTGT
RNA 2	CER3-I_CE	chrV:18,119,870 –18,121,325	SG-7502	GAAGCAGTCATTGTGGCTCA	SG-7503	GATTCTCCCTGTGCTCTTGC
RNA 3	CER10-I_CE	chrV:8,475,654 – 8,480,094	SG-7509	TACCAACGAGCCGAGTCTTC	SG-7510	TCTTCAGTTTCCTCGCCTGT
RNA 4	CELE12A	chrV:15,590,711 –15,591,042	SG-7523	TGCATGCAAGACTAATTTTCAA	SG-7524	CCAAAAGTTATTGGGCTTTCG
RNA 5	CELE2	chrIV:13,208,316 –13,208,605	SG-7527	GGCACGATGTTTTTGTTGAA	SG-7528	TCTATGAATTTCCCGCAGAA
RNA 6	CELE4	chrI:7,100,733 –7,100,826	SG-7533	GCCGCTAGACACCTAACGAG	SG-7534	ATTATGGGGACGCAGAAAAA
TAN 1	(CGA)n	chrX:16,884,800 –16,884,834	SG-7537	GGATGGATTGGGATGGATG	SG-7536	AACCATGCCAATCCTTTGTT
TAN 2	(TAGG)n	chrIII:11,288,980 –11,289,026	SG-7541	GTGTAGCCGTGGATTGTGTG	SG-7542	ATTTGCCTGCTGGTCCATAG
TAN 3	(CGA)n	chrV:6,040,594 – 6,040,628	SG-7545	ATCGGGATGGCTCGGTAT	SG-7546	ATCCCGATCCTTTGTTGTTG
TAN 4	(CAA)n	chrV:16,952,410 –16,952,470	SG-7548	CAGCCACAACTACCACAACG	SG-7550	CAGTCTGTTGGACACGGAAC
TAN 5	(CCTA)n	chrV:3,147,989 –3,148,076	SG-7551	ACCTACGTGCCTGCCTACAT	SG-7552	TTTTGTCAGGGACATGCGTA
TAN 6	(TCCA)n	chrX:16,887,998 –16,888,100	SG-7555	CCTGGCACTTACCAGTCCAT	SG-7556	ATCGGGATGGATGGCTTC
Figure 4						
MSAT1_CE		chrX:5,045,029 – 5,054,606	SG-8068	AGTACATTTTCAGCCCTACC	SG-8069	GTACCTACTGGAATATTGTT
(TAGG)n		chrIII:11,288,980 –11,289,026	SG-7541	GTGTAGCCGTGGATTGTGTG	SG-7542	ATTTGCCTGCTGGTCCATAG
(TCCA)n		chrX:16,887,998 –16,888,100	SG-7555	CCTGGCACTTACCAGTCCAT	SG-7556	ATCGGGATGGATGGCTTC
Tc4		chrX:125,694 – 126699	SG-7513	GAGTTTCCGTCCCGATTACA	SG-7514	AGAAACGGTTCGAACAATGC
CER10-I		chrV:8,480,088 – 8,481,470	SG-7498	TACCAACGAGCCGAGTCTTC	SG-7499	TCTTCAGTTTCCTCGCCTGT
Smg-5		chrI:6,734,183 – 6,736,153	SG-8903	CGGACACTTGGATGAGCAGA	SG-8904	CGTTCCATTTCGTTTTGGCG

Supplementary Table 4

Primers used in this study

Supplementary references

Andersen EC, Horvitz HR. 2007. Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. *Development* **134**: 2991-2999.

Consortium CeDM. 2012. large-scale screening for targeted knockouts in the Caenorhabditis elegans genome. *G3 (Bethesda)* **2**: 1415-1425.

Tocchini C, Keusch JJ, Miller SB, Finger S, Gut H, Stadler MB, Ciosk R. 2014. The TRIM-NHL protein LIN-41 controls the onset of developmental plasticity in Caenorhabditis elegans. *PLoS Genet* **10**: e1004533.

Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**: 934-947.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48**: 1385-1395.

Chapter 5: Conclusions and directions

In this thesis I present the first in depth characterization of the role of H3K9me in the multicellular nematode *C. elegans*. We could demonstrate that in the double mutant of the two H3K9 specific HMTs, worms lack all detectable H3K9me as monitored by immunofluorescence and mass spectroscopy. Despite showing temperature-dependent defects, the majority of worms nonetheless complete development from fertilized eggs into fertile adults. In contrast to previous studies in fission yeast (Ekwall et al. 1996), flies (Kellum and Alberts 1995) and mice (Peters et al. 2001), we did not observe an increase in mitotic chromosome segregation defects, which likely reflects the holocentric organization of the *C. elegans* chromosomes (Maddox et al. 2004). Detailed analysis of the genomic distribution of H3K9me2 and me3 in embryos showed a distribution of H3K9me that is comparable to that in mouse and human cells: at telomeres, the rDNA locus, dispersed repetitive elements and non-transcribed genes. This allowed us to characterize the role of H3K9me mediated repression on development and genomic function, without interference phenotypes caused by chromosome missegregation.

We could observe temperature-dependent sterility, which coincided with an increase in DNA-damage-induced apoptosis and stochastic delays in development at 25°C. Correlated with these phenotypes, we detected derepression of ~20% of all mappable repeat elements at 20°C, and this value increased at 25°C. Expression of these RE was accompanied by the accumulation of RNA:DNA hybrids (specifically at DNA transposons and tandem repeats), increased copy number variation (CNV) of RE and a hypersensitivity to replication stress. This correlation suggests that the generation of insertions and deletions within RE is either due to the transcription of the repetitive sequence alone, or is due to transcription coupled with the danger of mispairing events due to the repetitive nature of RE during repair. This suggests a potential mechanism for the previously unexplained observation of spontaneous heterochromatic DNA damage in partially H3K9me deficient flies (Peng and Karpen 2009).

We carefully mapped the distribution of H3K9me2 and me3 over the genome and found that: me2 is mainly found on DNA transposons and tandem repeats, while me3 was enriched on DNA and RNA transposons, and a small fraction of tandem repeats. The clear separation of HMT function in *C. elegans*, with SET-25 being the only H3K9 trimethylase, which does not alter global H3K9me1 or -me2 levels, allowed us to separate their unique roles.

We further showed that in the absence of H3K9me3, H3K9me2 fulfills the majority of the functionality ascribed to H3K9 methylation: transcriptional silencing, peripheral anchoring of

heterochromatin (Towbin et al. 2012) and prevention of the majority of RNA:DNA hybrid accumulation. At the same time, we identify a specific subset of genes and transposable elements that depend on SET-25 and H3K9me3, potentially because they cannot be targeted by MET-2. Interestingly, loci that require MET-2 for H3K9me correlate with chromosomal regions that can be tethered to the nuclear envelope. Finally, we identified a number of RNAi targets that are synthetic lethal with loss of MET-2, which strengthens our proposition that prevention of RNA:DNA hybrid-driven mutagenesis is one of the major functions of H3K9 methylation.

Temperature sensitivity in the absence of H3K9me

In this study we see a very strong effect of increased temperature on transcriptional de-repression, R-loop accumulation and the DNA damage-driven apoptosis in the germline. All three characteristics become more prominent at 25°C vs. 20°C, additionally strengthening the correlation between DNA damage and R-loop accumulation specifically on de-repressed repeats. Our model would suggest that the transcriptional changes induced by the elevated temperature are at the basis of the sensitivity (Zeller et al. 2016).

We do note that even wildtype animals show an increase in global transcription at 25°C vs. 20°C. However, this difference becomes much more pronounced in H3K9me-deficient animals. We speculate that the increase in temperature provides a transcriptionally more active environment, which in the absence of H3K9me-mediated silencing causes a drastic increase in transcription. Interestingly, one sequence type, i.e. tandem repeats, are especially affected in H3K9me-deficient worms at 25°C, relative to H3K9-methylated genes or transposons. The potential danger accompanied with their expression is discussed below.

With our current results, we cannot identify the driver of the temperature-enhanced transcription. We speculate that specific temperature-induced transcription factors are involved, or that RNA polymerase II becomes generally more active. It would be interesting to see if the subnuclear localizations of genomic regions is affected at 25°C, or to see if there is an increase in level or activity of known temperature-induced transcription factors. This could help to identify the components involved.

Repetitive elements: Sequences prone to form higher order structures

Unscheduled transcription appears to drive the formation of RNA:DNA hybrids in cells lacking H3K9me. While we observe the derepression of genes (234 @ 20°C) and all three RE classes in embryos lacking H3K9me, yet the DRIP-seq analysis shows a specific increase of RNA:DNA

hybrids on DNA transposons and tandem repeats, when we compare *met-2 set-25* worms with wildtype. We also observe the occurrence of RNA:DNA hybrids on highly transcribed genes in wildtype worms similar to a previous report (Wahba et al. 2016), yet they do not increase in the absence of H3K9me. It is not clear if replication is necessary for RNA:DNA hybrid formation or whether this stems from the specific character of the repeat RNA and its lack of processing.

Different sequences may be differentially prone to form RNA:DNA hybrids, when transcribed. Indeed, previous studies in yeast and mammalian cell culture have made similar observations. Whereas the accumulation of RNA:DNA hybrids correlates with the strength of expression for most genes (Wahba et al. 2016), sequence features that facilitate hybrid formation are also detected, e.g. GC content, poly A tracks (Ginno et al. 2012; Wahba et al. 2016). Of particular note is the accumulation of RNA:DNA hybrids on telomeres and the Thy1 transposons (Chan et al. 2014a), despite the relatively low expression level of those sequences.

In recent years numerous studies have tried to elucidate the mechanisms that protect cells from RNA:DNA hybrid accumulation and led to the identification of many factors involved in transcriptional processivity (Santos-Pereira and Aguilera 2015), very similar to the factors we found synthetic lethal with loss of H3K9me. This lead to the hypothesis that there are three factors promoting RNA:DNA hybrid formation: first: the affinity of the transcribed RNA to the homologous DNA, which is dictated by the sequence; second, the factors that bind, process and protect the RNA; and third, the processivity and abundance of the RNA polymerase. The coordination of transcription with the replication fork is a fourth factor that we have not been able to examine in worms.

It is intriguing that under the conditions where we score the most striking differences in the level of RNA:DNA hybrids , i.e. high temperature and *met-2* vs. *set-25* single mutant that the major transcriptional difference seems to be a strong derepression of tandem repeats. This class of repetitive elements consists of the linear repetition of a 2-5 base pair sequence, which in theory allows for RNA:DNA pairing in multiple positions. They were neither reported to possess a promoter sequence, an open reading frame or any other sequence elements known to promote the recruitment of RNA processing factors. These characteristics would facilitate the formation of RNA:DNA hybrids in the case of aberrant transcription. A pyrimidine vs purine preference in RNA:DNA hybrid formation, as shown in genes, could not be analyzed due to low resolution of the sequencing data.

Unscheduled collisions of the replication and transcription machineries appear to generate breaks as well as other forms of genome instability (Azvolinsky et al. 2009; Hoffman et al. 2015). In this context damage is often attributed to the presence of RNA:DNA hybrids (Skourti-Stathaki et al. 2011; Chan et al. 2014b), which are thought to stall DNA replication fork progression. Interestingly, tandem repeats have previously been shown to be very fragile if transcription (especially bidirectional) is allowed to pass through them (Wierdl et al. 1996; Lin et al. 2006; Nakamori et al. 2011). So far this was mostly attributed to their potential to form higher order structures such as stem loops (Pearson et al. 2005). The torsional stress produced by bidirectional transcription could favor the formation of stem loops and G-quadruplexes. All these factors have now been shown to also favor RNA:DNA hybrid accumulation (Roy and Lieber 2009; El Hage et al. 2010; Hamperl and Cimprich 2014). Thermodynamically the RNA:DNA hybrid was shown to be more stable than the DNA:DNA structures favoring them as obstacles DNA replication fork progression (Thomas et al. 1976). It is unclear what the predominant harmful structure at transcribed tandem repeats really is. Experiments combining systems for inducible tandem repeat transcription with the depletion of RNA:DNA hybrids through RNAseH overexpression, or specific targeting of a RNAseH fusion protein, may be able to answer this question. Alternatively depleting H3K9me after replication is terminated in L1 larvea would show whether or not replication is required for genome instability.

RNA transposons are also derepressed in the absence of H3K9me, yet they do not accumulate RNA:DNA hybrids. Although *C. elegans* does contain full length RNA transposons of the long terminal repeat (LTR) class (Ganko et al. 2001), no transposition has been reported so far (Bessereau 2006). Besides the possible lack of hybridization-promoting sequence features in the transcript, there may be a link between the absence of RNA:DNA hybrid formation on RNA transposons and their lack of transposition. To prevent co-transcriptional RNA:DNA hybrid formation (Huertas and Aguilera 2003) co-transcriptional RNA degradation would be needed. The 5'-3' exonuclease XRN-2 was shown be involved in the co-transcriptional degradation of nascent RNA as part of transcriptional termination process (West et al. 2004). A later study could additionally show that XRN-2 is involved in the degradation of many endogenous transcripts, if transcriptional processivity is impaired by Spliceostatin A (Davidson et al. 2012). RNA transposons in *C. elegans* are considered to be evolutionary very young (Ganko et al. 2001) and its transcript might not be able to be efficiently processed, thereby leading to XRN-2 recruitment. Unpublished results from our lab (Mattout et al.) additionally showed that the nuclear RNA chaperone complex LSM2-8 targets heterochromatic transcripts for degradation in a H3K27me- and XRN-2-dependent manner. Our study shows that retrotransposons are particularly enriched for H3K9me3 that frequently co-occurs with H3K27me3. Therefore, the specific targeting of the RNA transposon

transcripts for degradation is plausible. This hypothesis could be tested by analyzing sequence specific RNA:DNA hybrid accumulation in worms lacking both H3K27me as well as H3K9me, and in the *lsm-8* and *xrn-2* deletion strains.

Involvement of H3K9me at multiple stages of repeat stability

We hypothesize that the accumulation of RNA:DNA hybrids at RE explains why the genomic mutations scored in the *met-2 set-25* mutant mapped almost exclusively to RE. Considering that REs have multiple copies in the genome, additional factors may influence the pathway of repair. In particular, when the lesion is a DNA double-strand break, H3K9me may play an additional role by sequestering repetitive elements away from the machinery that could repair through ectopic homologous recombination (HR), to prevent the use of a false template during repair. Indeed, the groups of Karpen (Chiolo et al. 2011; Janssen et al. 2016), Chiolo (Ryu et al. 2015) and Soutoglou (Tsouroula et al. 2016) have shown that DNA double-strand breaks within heterochromatic sequences are repaired differently than euchromatic breaks when undergoing repair by homologous recombination. In IF experiments using DAPI as a proxy for heterochromatin, they could describe a HP1a dependent pathway (Chiolo et al. 2011) that allows early steps of the DNA damage response to occur in heterochromatin, while the late steps of homologous recombination (e.g. Rad51 binding) only occurs after the break site has been relocated outside of the heterochromatic region, away from other repeat copies. Non-homologous end joining, which does not need a template, can occur unperturbed within heterochromatic domains (Janssen et al. 2016). The suppression of HR depends on both the H3K9 methyltransferase Su(var)3-9, as well as its reader HP1a, arguing that H3K9me works also in this role through its reader.

Although it is still under debate how exactly those differences are relevant for the repair outcome (Janssen et al. 2016), the mutations we see might be the result of an increase in repeat specific DNA damage coupled with the loss of heterochromatin-specific repair, which assigns two roles to H3K9me in guarding the genome.

How are tandem repeats targeted for H3K9 di methylation?

Taking chapter 3 and 4 together, we identified three pathways that result in the methylation of H3K9. The majority of H3K9me targets appear to depend on the incorporation of H3K9 di-methylated histones. This mark is further processed by SET-25 to H3K9me3. This accounts for H3K9-methylated genes as well as the majority of H3K9-methylated transposons. A second group remains H3K9 dimethylated, apparently lacking the necessary cue to recruit SET-25. This group consists largely of tandem repeats and a subset of DNA transposons. Both groups depend on the

action of MET-2. A third group seems to be directly targeted by SET-25, and not MET-2, as their high levels of H3K9me3 depend solely on SET-25. This group mainly consists of DNA and RNA transposons.

For both transposable elements and genes, the mechanisms that are known to recruit H3K9 HMTs include small RNA pathways like the PIWI pathway in the germline (Haynes et al. 2006; Sienski et al. 2012), dsRNA transcripts at satellite repeats in *S. pombe* (Keller et al. 2012), as well as transcription factors, i.e. the orphan nuclear receptor SHP (Fang et al. 2007; Garcia-Bassets et al. 2007; Bulut-Karslioglu et al. 2012). Considering the potential risk arising from the transcription of tandem repeats, it is relevant to ask what targets this class of RE for H3K9me2 so efficiently? Indeed, beyond centromeric satellite repeats, very little is known about the control of tandem repeats.

One intriguing possibility is the existence of a feedback loop from the DNA damage caused by the transcription of tandem repeats, which might then recruit HMTs which then deposit H3K9me. Several arguments support such an idea. multiple repressive factors, including HP1 (Luijsterburg et al. 2009), Polycomb (Hong et al. 2008) and HDAC1/2 (Miller et al. 2010), are recruited to sites of DNA damage. Besides a role for heterochromatin in repair factor recruitment, they are also implicated in the local transcriptional silencing around damage sites, that is meant to prevent conflicts between the repair and the transcription machinery (Vissers et al. 2012; Ui et al. 2015). In this context the Almouzni group showed a transient transcriptional silencing during DNA damage repair (Adam et al. 2013). In their study, they identified the histone chaperone HIRA as a crucial component to ensure transcriptional reactivation after UVC induced DNA damage repair. But what would happen in a region that does not contain a strong and specific transcription factor to recruit HIRA? Would the heterochromatic state remain, marking the region as potentially dangerous? Would this prevent future breaks from happening? UV damage and replication-fork associated damage are repaired by completely distinct mechanisms, but it cannot be excluded that fork-associated damage also recruits heterochromatin components.

The only other sequence type that accumulates H3K9me2 and not H3K9me3 besides tandem repeats are DNA transposons. The most prominent group of DNA transposons in *C. elegans* works by a cut and paste mechanism leaving a double-strand break behind when it jumps (Vos et al. 1996; Bessereau 2006). This means that both repeat classes that are solely H3K9 dimethylated (DNA transposons and tandem repeats) can potentially cause double strand breaks, once expressed. Coupling the occurrence of DNA damage to the chromatin state in such a manner would allow the silencing and safeguarding of naturally fragile sites.

Crosstalk between HMT pathways

Although H3K9me3 could be found on 12% of all genes (developmentally regulated- and pseudo-genes) while H3K9me2 was only found weakly on 6%, a total of 336 genes were derepressed in the *met-2* and only 30 genes in *set-25* single mutants. Similarly, although we find mainly H3K9me3 on retrotransposons (H3K9me3 pos 58%, H3K9me2 pos 6%), some retrotransposons, i.e. LTR2_CE and Cer16-LTR_CE, are only de-repressed in *met-2* but not *set-25* mutants. Our H3K9me2 and me3 ChIP experiments in wt and *set-25* mutants show that in the absence of SET-25, many normally H3K9me3 positive sites become H3K9me2 positive. This not only suggests their sequential function at these loci, but also shows that MET-2 deposited H3K9me2 is sufficient for their transcriptional silencing. Interestingly, this also shows that the methylation level that dominates under wild type conditions does not need to be the one that is essential for silencing.

Additionally, it should not be forgotten that loss of silencing and transcriptional activation are not identical. Even in the absence of H3K9me, gene expression still depends on their activation by specific transcription factors. Multiple studies show that this might also be the case for RE, as tissue-specific transposon expression was scored (Kim et al. 2001; Singer et al. 2010). Alternatively, a pluripotency specific factor may be needed. For instance, a recent study showed that the transcription of the retrotransposon HERVK is induced during normal human embryogenisis by the transcription factor OCT-4 (Grow et al. 2015). A dependency on transcription factor-mediated activation is thus true for transposons, as for all genes.

A second explanation for the fact that a majority of RE remain transcriptionally silent in the absence of H3K9me would be that H3K9me-mediated silencing is redundant with other repression pathways. As mentioned in the introduction, it was shown that there is an interplay between DNA methylation and H3K9me in mammalian cells (Leung et al. 2014; Liu et al. 2014). In *C. elegans* a substantial proportion of H3K9me3 co-localizes with the repressive H3K27me3 (Ho et al. 2014). In addition, multiple studies could show that H3K27me is able to mark, or spread into regions that lost their original chromatin state (H3K36me (Gaydos et al. 2012; Patel et al. 2012), H3K9me (Saksouk et al. 2014)). This suggests that it can potentially compensate for the loss of another silencing mark, such as H3K9me. If we wish to understand the role of specific epigenetic marks, it will be crucial to consider and monitor adaptive changes brought about by other epigenetic marks or other levels of methylation (me1, me2 and me3), when H3K9me is ablated.

These results also put our previous model concerning the roles of H3K9me2 and me3 into a new light. Based on experiments using a large heterochromatic array, Towbin et al. could show that

either MET-2 (H3K9me2) or SET-25 (H3K9me3) alone are sufficient to anchor chromatin to the nuclear periphery, while SET-25 was more important for the complete transcriptional silencing of the reporter (Towbin et al. 2012). The repression of endogenous sequences is different. Both enzymes can work independently, or else SET-25 depends on, and can be partially compensated by MET-2. This makes MET-2 more crucial for the transcriptional silencing of endogenous sequences than SET-25. Similarly, although SET-25-deposited H3K9me3 is sufficient to tether the heterochromatic reporter in the absence of MET-2, the amount of H3K9 methylation it deposits on the genome may not be sufficient for anchoring to the nuclear periphery, unlike the large repetitive array used as a heterochromatin surrogate. Our study thus has been crucial for refining our understanding of MET-2 and SET-25 roles at a reporter and comparing these with their roles at endogenous sites of repression.

Why have two enzymes, if one can do it all?

The data presented in this thesis shows that H3K9me2 is able to inhibit transcription, and that MET-2 is able to methylate its target sites independently of SET-25. Only a very specific subset of genomic loci depends on the function of SET-25, raising the question what function SET-25 and H3K9me3 have.

A potential explanation might be provided by work on DNA methylation in mammalian cells, where a clear separation into maintenance (DNMT1 (Gruenbaum et al. 1982; Bestor and Ingram 1983)) and de novo methyl transferases exist (DNMT3a and b (Okano et al. 1998; Lyko et al. 1999)). Supporting a similar separation of work, de novo H3K9 trimethylation was identified in *C. elegans* as a SET-25 dependent transgenerational silencing pathway initiated by the addition of exogenous RNAi (Ashe et al. 2012; Buckley et al. 2012). Although the authors did not prove MET-2 independence of this process, it shows the potential of the H3K9me3 machinery to target new sequences for methylation independent of their position in the genome. A pathway crucial for survival in an environment where new retro-transposons can infect at any time, spreading in the host genome, if not transcriptionally silenced.

A second potential reason for the existence of two pathways comes from the localization of the two enzymes, MET-2 is at least partially cytoplasmic, while SET-25 is exclusively nuclear and binds its own product H3K9me3 (Towbin et al. 2012). A cytoplasmic enzyme can modify histones even before they are integrated into the genome and can thereby target any locus in the genome, as long as histones are turned over. However, once the modified histone is integrated, it can be de-methylated if not protected. The nuclear SET-25, on the other hand, by binding its product

could protect H3K9me3, constantly reinforcing and spreading its repressive chromatin state. This ability would be especially crucial when sequences must be repressed in an otherwise euchromatic environment. This fits exactly to the distribution of strongly SET-25 dependent targets along the genome. Unlike MET-2-dependent sequences, SET-25-dependent sequences have no clear correlation with the heterochromatin-rich chromosome arms, nor with association with the nuclear periphery.

A project in progress together with Jan Padeken is aiming to identify the machinery that targets SET-25 to the genome in a H3K9me2-dependent and -independent manner.

Different ways to control heterochromatin propagation: foci formation and the nuclear envelope

As mentioned above, SET-25 might be able to establish and maintain a strong and stable H3K9me3 domain by itself. But how can MET-2, which does not seem to bind its own mark, prevent the loss of H3K9me2? One possibility would be a high histone turnover to integrate freshly methylated histones continuously. On the other hand recent studies of local histone turnover argue against such a hypothesis ((Toyama et al. 2013) (Aygün et al. 2013)). Following the integration kinetics of an inducible tagged histone, heterochromatin was shown to be a zone of especially low histone turnover. A second option would therefore be to prevent the access of demethylating enzymes that could remove the repressive H3K9me (e.g. KDM4b (Tsurumi et al. 2013)), or those that activate transcription like histone acetyl transferases.

In chapter 4 of this thesis we compare the fold-change derepression of sequence windows in *met-2* and *set-25* single mutants with the respective nuclear localization (LEM-2 ChIP signal) under wildtype conditions. This analysis showed a clear correlation of MET-2-dependent silencing and localization close to the nuclear periphery, while sequences that depend on SET-25 for silencing are not enriched at the nuclear envelope. The nuclear envelope has been suggested to facilitate the compartmentalization of chromatin by concentrating certain factors and bringing different loci into close proximity. Peripheral localization of genomic regions in general correlates with an untranscribed state (Pickersgill et al. 2006), silencing factors such as HDAC3 (Somech et al. 2005) and HP1 (Ye and Worman 1996) have been shown to interact with components of the nuclear envelope. Tethering reporter sequences to the periphery can lead to their transcriptional silencing (Finlan et al. 2008; Reddy et al. 2008). However it should be noted though that not every sequence tethered to the periphery is silenced (Ruault et al. 2008).

Trying to put these results together, the most likely hypothesis would be that localization does not force a transcriptional state onto a sequence, but can favor it or its propagation during replication. As mentioned in the introduction, DNA replication can be divided into an early and a late phase, with regions localized at the nuclear periphery replicating late (Guelen et al. 2008; Hansen et al. 2010). Disruption of this separation by depletion of the origin of replication-associated protein (ORCA) was shown to interfere with H3K9me propagation (Wang et al. 2016). This suggests the existence of two main silencing pathways: One that can form locally and reinforcing its own heterochromatic state, in the case of *C. elegans* by direct targeting of SET-25. And a second pathway that depends on the local enrichment of silencing factors at the periphery, but which still contains sequence characteristics that must be targeted for silencing. If this hypothesis would be true, nuclear organization might become more crucial for H3K9me2 maintenance in the absence of self-reinforcing heterochromatic factors such as SET-25.

Future perspectives

This work establishes a previously unreported link between transcriptional derepression of interspersed RE and the accumulation of repeat centered mutations through the accumulation of RNA:DNA hybrids. The consequences of this connection might even be more detrimental in humans, where recent analyses suggest that over two thirds of the genome consists of repetitive sequences (de Koning et al. 2011) compared to at least 12% in *C. elegans* (see Chapter 2 (Padeken et al. 2015)). New therapeutic approaches in cancer therapy (Kondo et al. 2008) and stem cell induction (Chen et al. 2013) on the basis of H3K9 HMT inhibition, should therefore be carefully controlled for mutagenic side effects.

Additionally, this work highlights the potential danger of tandem repeat expression. In the absence of H3K9me they are the main repeat class transcriptionally effected by the elevation of temperature, and thus correlated with sterility (Chapter 3). They represent one of the main differences between germline and somatic repeat derepression in H3K9me-deficient worms, correlating with increased somatic mutation rates of the LacZ-reporter (Chapter 3). Moreover, they are the only sequence type that is affected by loss of MET-2, and not SET-25 (Chapter 4). While repeats were recognized as functionally and structurally important sequences, especially at telomeres and the centromeres, recent studies now also report the potential danger of their derepression (Nakamori et al. 2010). Of particular interest are studies concerning the well-characterized tumor suppressor breast cancer susceptibility gene 1 (BRCA1) (Castilla et al. 1994). Zhu et al. could show that loss of BRCA1 leads derepression of the tandemly repeated satellite DNA. Very strikingly, artificial expression of these tandem repeats from a transgene can

phenocopy the BRCA1 deletion effect, including: loss in centrosome amplification, cell-cycle checkpoint defects, DNA damage and genomic instability (Zhu et al. 2011). This argues that transcriptional control of tandem repeats is one of the main functions of this central and well-studied tumor suppressor. As BRCA1 mutation was associated with changes in the nuclear distribution of HP1a foci, BRCA1 was speculated to mediate this role through the H3K9 methylation machinery. But as we could show combining mutants of the H3K9 HMTs with RNAi against the worm homologue of BRCA1 (BRC-1) shows an additive effect and therefore rather suggest those two pathways work in a partially redundant manner to regulate tandem repeat expression and genome integrity.

An exciting field for future research will be the role of heterochromatin and its maintenance in differentiated cells. In Chapter 3 we showed that the transcriptional derepression as well as the rate of heterochromatic mutation are elevated in differentiated somatic cells, compared to the germline. Counter-selection of highly damaged oocyte genomes by apoptosis, might partially explain the difference in mutation rates, but, considering the absence of the transposon silencing PIWI pathway in somatic cells (Girard et al. 2006), as well as the accumulation of heterochromatin during differentiation (Gifford et al. 2013; Ugarte et al. 2015), these data could also indicate an increased importance of the H3K9me machinery in differentiated cells.

References

Adam S, Polo SE, Almouzni G. 2013. Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. *Cell* **155**: 94-106.

Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J et al. 2012. piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell* **150**: 88-99.

Aygün O, Mehta S, Grewal SIS. 2013. HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin. *Nat Struct Mol Biol* **20**: 547-554.

Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell* **34**: 722-734.

Bessereau JL. 2006. Transposons in C. elegans. *WormBook* doi:10.1895/wormbook.1.70.1: 1-13.

Bestor TH, Ingram VM. 1983. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences* **80**: 5559-5563.

Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, Fritz H, Kimble J, Fire A, Kennedy S. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**: 447-451.

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

Castilla LH, Couch FJ, Erdos MR, Hoskins KF, Calzone K, Garber JE, Boyd J, Lubin MB, Deshano ML, Brody LC et al. 1994. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* **8**: 387-391.

Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. 2014a. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* **10**: e1004288.

Chan YA, Hieter P, Stirling PC. 2014b. Mechanisms of genome instability induced by RNA-processing defects. *Trends in Genetics* **30**: 245-253.

Chen J, Liu H, Liu J, Qi J, Wei B, Yang J, Liang H, Chen Y, Chen J, Wu Y. 2013. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature genetics* **45**: 34-42.

Chiolo I, Minoda A, Colmenares SU, Polyzos A, Costes SV, Karpen GH. 2011. Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**: 732-744.

Davidson L, Kerr A, West S. 2012. Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *The EMBO Journal* **31**: 2566-2578.

de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**: e1002384.

Ekwall K, Nimmo ER, Javerzat J-P, Borgstrom B, Egel R, Cranston G, Allshire R. 1996. Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *Journal of Cell Science* **109**: 2637-2648.

El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* **24**: 1546-1558.

Fang S, Miao J, Xiang L, Ponugoti B, Treuter E, Kemper JK. 2007. Coordinated recruitment of histone methyltransferase G9a and other chromatin-modifying enzymes in SHP-mediated regulation of hepatic bile acid metabolism. *Mol Cell Biol* **27**: 1407-1424.

Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR, Bickmore WA. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* **4**: e1000039.

Ganko EW, Fielman KT, McDonald JF. 2001. Evolutionary history of Cer elements and their impact on the C. elegans genome. *Genome Res* **11**: 2066-2074.

Garcia-Bassets I, Kwon Y-S, Telese F, Prefontaine GG, Hutt KR, Cheng CS, Ju B-G, Ohgi KA, Wang J, Escoubet-Lozach L. 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **128**: 505-518.

Gaydos Laura J, Rechtsteiner A, Egelhofer Thea A, Carroll Coleen R, Strome S. 2012. Antagonism between MES-4 and Polycomb Repressive Complex 2 Promotes Appropriate Gene Expression in C. elegans Germ Cells. *Cell Reports* **2**: 1169-1177.

Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**: 1149-1163.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell* **45**: 814-825.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199-202.

Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY et al. 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**: 221-225.

Gruenbaum Y, Cedar H, Razin A. 1982. Substrate and sequence specificity of a eukaryotic DNA methylase.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948-951.

Hamperl S, Cimprich KA. 2014. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)* **19**: 84-94.

Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences* **107**: 139-144.

Haynes KA, Caudy AA, Collins L, Elgin SC. 2006. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Current biology* **16**: 2222-2227.

Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449-452.

Hoffman EA, McCulley A, Haarer B, Arnak R, Feng W. 2015. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome research* **25**: 402-412.

Hong Z, Jiang J, Lan L, Nakajima S, Kanno S, Koseki H, Yasui A. 2008. A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell. *Nucleic Acids Res* **36**: 2939-2947.

Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell* **12**: 711-721.

Janssen A, Breuer GA, Brinkman EK, van der Meulen AI, Borden SV, van Steensel B, Bindra RS, LaRocque JR, Karpen GH. 2016. A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* **30**: 1645-1657.

Keller C, Adaixo R, Stunnenberg R, Woolcock KJ, Hiller S, Buhler M. 2012. HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. *Mol Cell* **47**: 215-227.

Kellum R, Alberts BM. 1995. Heterochromatin protein 1 is required for correct chromosome segregation in Drosophila embryos. *J Cell Sci* **108 (Pt 4)**: 1419-1431.

Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS. 2001. A Gene Expression Map for Caenorhabditis elegans. *Science* **293**: 2087-2092.

Kondo Y, Shen L, Ahmed S, Boumber Y, Sekido Y, Haddad BR, Issa J-PJ. 2008. Downregulation of histone H3 lysine 9 methyltransferase G9a induces centrosome disruption and chromosome instability in cancer cells. *PloS one* **3**: e2037.

Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, Li Y, Szulwach KE, Jin P, Lorincz MC. 2014. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proceedings of the National Academy of Sciences* **111**: 6690-6695.

Lin Y, Dion V, Wilson JH. 2006. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol* **13**: 179-180.

Liu S, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, Sasaki H, Shinkai Y, Lorincz MC. 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & development* **28**: 2041-2055.

Luijsterburg MS, Dinant C, Lans H, Stap J, Wiernasz E, Lagerwerf S, Warmerdam DO, Lindh M, Brink MC, Dobrucki JW et al. 2009. Heterochromatin protein 1 is recruited to various types of DNA damage. *J Cell Biol* **185**: 577-586.

Lyko F, Ramsahoye BH, Kashevsky H, Tudor M, Mastrangelo MA, Orr-Weaver TL, Jaenisch R. 1999. Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in Drosophila. *Nat Genet* **23**: 363-366.

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

Miller KM, Tjeertes JV, Coates J, Legube G, Polo SE, Britton S, Jackson SP. 2010. Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat Struct Mol Biol* **17**: 1144-1151.

Nakamori M, Pearson CE, Thornton CA. 2010. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)•(CAG) repeats. *Human Molecular Genetics* **20**: 580-588.

Nakamori M, Pearson CE, Thornton CA. 2011. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. *Hum Mol Genet* **20**: 580-588.

Okano M, Xie S, Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* **19**: 219-220.

Padeken J, Zeller P, Gasser SM. 2015. Repeat DNA in genome organization and stability. *Current opinion in genetics & development* **31**: 12-19.

Patel T, Tursun B, Rahe DP, Hobert O. 2012. Removal of Polycomb repressive complex 2 makes C. elegans germ cells susceptible to direct conversion into specific somatic cell types. *Cell Rep* **2**: 1178-1186.

Pearson CE, Nichol Edamura K, Cleary JD. 2005. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet* **6**: 729-742.

Peng JC, Karpen GH. 2009. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet* **5**: e1000435.

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**: 323-337.

Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature genetics* **38**: 1005-1014.

Reddy K, Zullo J, Bertolino E, Singh H. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**: 243-247.

Roy D, Lieber MR. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Molecular and cellular biology* **29**: 3124-3133.

Ruault M, Dubarry M, Taddei A. 2008. Re-positioning genes to the nuclear envelope in mammalian cells: impact on transcription. *Trends in genetics : TIG* **24**: 574-581.

Ryu T, Spatola B, Delabaere L, Bowlin K, Hopp H, Kunitake R, Karpen GH, Chiolo I. 2015. Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat Cell Biol* **17**: 1401-1411.

Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME et al. 2014. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol Cell* **56**: 580-594.

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16**: 583-597.

Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**: 964-980.

Singer T, McConnell MJ, Marchetto MC, Coufal NG, Gage FH. 2010. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in neurosciences* **33**: 345-354.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell* **42**: 794-805.

Somech R, Shaklai S, Geller O, Amariglio N, Simon AJ, Rechavi G, Gal-Yam EN. 2005. The nuclear-envelope protein and transcriptional repressor LAP2beta interacts with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *J Cell Sci* **118**: 4017-4025.

Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* **73**: 2294-2298.

Towbin BD, Gonzalez-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**: 934-947.

Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR, 3rd, Hetzer MW. 2013. Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**: 971-982.

Tsouroula K, Furst A, Rogier M, Heyer V, Maglott-Roth A, Ferrand A, Reina-San-Martin B, Soutoglou E. 2016. Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* **63**: 293-305.

Tsurumi A, Dutta P, Yan S-J, Shang R, Li WX. 2013. Drosophila Kdm4 demethylases in histone H3 lysine 9 demethylation and ecdysteroid signaling. *Scientific reports* **3**: 2894.

Ugarte F, Sousae R, Cinquin B, Martin EW, Krietsch J, Sanchez G, Inman M, Tsang H, Warr M, Passegué E. 2015. Progressive chromatin condensation and H3K9 methylation regulate the differentiation of embryonic and hematopoietic stem cells. *Stem cell reports* **5**: 728-740.

Ui A, Nagaura Y, Yasui A. 2015. Transcriptional elongation factor ENL phosphorylated by ATM recruits polycomb and switches off transcription for DSB repair. *Mol Cell* **58**: 468-482.

Vissers JH, van Lohuizen M, Citterio E. 2012. The emerging role of Polycomb repressors in the response to DNA damage. *J Cell Sci* **125**: 3939-3948.

Vos JC, De Baere I, Plasterk RH. 1996. Transposase is the only nematode protein required for in vitro transposition of Tc1. *Genes Dev* **10**: 755-761.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev* **30**: 1327-1338.

Wang Y, Khan A, Marks AB, Smith OK, Giri S, Lin YC, Creager R, MacAlpine DM, Prasanth KV, Aladjem MI et al. 2016. Temporal association of ORCA/LRWD1 to late-firing origins during G1 dictates heterochromatin replication and organization. *Nucleic Acids Res* doi:10.1093/nar/gkw1211.

West S, Gromak N, Proudfoot NJ. 2004. Human 5[prime] [rarr] 3[prime] exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**: 522-525.

Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD. 1996. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* **143**: 713-721.

Ye Q, Worman HJ. 1996. Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. *Journal of Biological Chemistry* **271**: 14653-14656.

Zeller P, Padeken J, van Schendel R, Kalck V, Tijsterman M, Gasser SM. 2016. Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. *Nat Genet* **48**: 1385-1395.

Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM. 2011. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**: 179-184.

List of abbreviations

CNV	copy number variation
ChIP	chromatin immunoprecipitation
DAPI	4′,6-Diamidin-2-phenylindol (DNA stain used in IF-microscopy)
DNA	deoxyribonucleic acid
DRIP	DNA:RNA hybrid immunoprecipitation
EE	C. elegans early stage embryos
EM	electron microscope
FISH	fluorescence in situ hybridization
GFP	green fluorescent protein
H1,2a,2b,3,4	histone 1,2a…
H3K4me	histone 3 Lysine 4 methylation
H3K9me1,2,3	histone 3 Lysine 9 mono- di- tri- methylation
H3K27me	histone 3 Lysine 27 methylation
HDAC	histone deacetylase
HMT	histone methyl transferase
HU	hydroxyurea
HP1	heterochromatin protein 1
HPL-1	heterochromatin protein like 1
IF	immunofluorescence
L1	C. elegans larvae stage 1
LAD	lamina associated domain
meCpG	DNA methylation on CG dinucleotide

PRC	Polycomb Repressive Complex
ORF	open reading frame
PolII	RNA polymerase II
qPCR	quantitative polymerase chain reaction
rDNA	ribosomal DNA
RE	repetitive element
RNA	ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
SNV	single nucleotide variants
SUV39H1	suppressor of variegation 3-9 homolog 1 (HMT)
TAD	topologically associating domain
TFIID	transcription factor II D

Acknowledgments

I want to thank Susan Gasser for accepting me in her group, encouraging, supporting and “pushing” me through all difficult phases a PhD brings along. Her guidance and the great group of people she recruited made those years an exceptional experience.

I also want to thank my committee members: Iskra Katic, Marcel Tijsterman and Antoine Peters, the insightful discussions during and outside our meetings were key in guiding this thesis in such a fruitful direction. The collaboration with Marcel was instrumental to analyze and accurately interpret our mutation experiments.

Very special thanks go to Jan Padeken as colleague and friend. It was a true pleasure, personally as well as scientifically, to work together with you. The countless hours of discussion in our second office (the corridor) and the follow up of resulting ideas in experiments resulted in a form of discovery driven research that I wish all PhDs would have the chance to enjoy.

In this perspective, I also want to thank every member of the Gasser lab, past and present, for being such a friendly, helpful, smart and just really nice to be around group. I will hold dear every single coffee break we had. In this context, I also want to thank Jennifer, Anna and Jan for proofreading this thesis. I could fill pages for the things that made it such a nice time to work with every single one of them and will therefore do that personally.

The constant support I received from my family was very important for me and helped me overcoming the difficult phases of this PhD.

Last but definitely not least, I want to thank Ekaterina Voronova. She was a crucial support in my life, showing me what living for science really means and constantly encouraging to try a bit harder than one can. I cannot thank her enough. Without her, I would not even have started the amazing challenge of a PhD at the FMI.