A transcription factor–based mechanism for mouse heterochromatin formation

Article *in* Nature Structural & Molecular Biology · September 2012

Impact Factor: 13.31 · DOI: 10.1038/nsmb.2382 · Source: PubMed

15 authors, including:

Aydan Karslioglu
University of California, San Francisco
4 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Valentina Perrera
Ecole Normale Supérieure de Paris
2 PUBLICATIONS **82** CITATIONS
SEE PROFILE

Nicholas Shukeir
Max Planck Institute of Immunobiology an…
14 PUBLICATIONS **647** CITATIONS
SEE PROFILE

Monika Lachner
Max Planck Institute of Immunobiology an…
21 PUBLICATIONS **6,064** CITATIONS
SEE PROFILE

nature structural & molecular biology

A transcription factor–based mechanism for mouse heterochromatin formation

Aydan Bulut-Karslioglu[1,7], Valentina Perrera[1,2,7], Manuela Scaranaro[2], Inti Alberto de la Rosa-Velazquez[1,2], Suzanne van de Nobelen[1], Nicholas Shukeir[1], Johannes Popow[2,6], Borbala Gerle[2,6], Susanne Opravil[2], Michaela Pagani[2], Simone Meidhof[3–5], Thomas Brabletz[3], Thomas Manke[1], Monika Lachner[1] & Thomas Jenuwein[1,2]

Heterochromatin is important for genome integrity and stabilization of gene-expression programs. We have identified the transcription factors Pax3 and Pax9 as redundant regulators of mouse heterochromatin, as they repress RNA output from major satellite repeats by associating with DNA within pericentric heterochromatin. Simultaneous depletion of Pax3 and Pax9 resulted in dramatic derepression of major satellite transcripts, persistent impairment of heterochromatic marks and defects in chromosome segregation. Genome-wide analyses of methylated histone H3 at Lys9 showed enrichment at intergenic major satellite repeats only when these sequences retained intact binding sites for Pax and other transcription factors. Additionally, bioinformatic interrogation of all histone methyltransferase Suv39h–dependent heterochromatic repeat regions in the mouse genome revealed a high concordance with the presence of transcription factor binding sites. These data define a general model in which reiterated arrangement of transcription factor binding sites within repeat sequences is an intrinsic mechanism of the formation of heterochromatin.

Pericentric heterochromatin provides a structural scaffold for centromere formation and is essential for maintaining genome integrity in eukaryotic cells[1]. The DNA sequences at pericentric heterochromatin differ substantially among species[2]; the only common feature among them is their repetitive nature. In mice, the basic unit of pericentric heterochromatin is the major satellite repeat. This AT-rich sequence is 234 base pairs (bp) long and can be divided into four subrepeats[3]. Major satellite repeats are represented about 10,000 times in every chromosome and make up ~10% of the mouse genome[4]. In addition to these repetitive elements, pericentric heterochromatin in mammals has a distinct combination of epigenetic modifications, such as histone H3 Lys9 methylation (H3K9me3), H4K20me3 and DNA methylation[5,6]. Despite this accumulation of repressive epigenetic marks, heterochromatin is transcribed in many organisms, giving rise to noncoding RNAs[7]. The intersection of the RNA-interference pathway and heterochromatin formation has been well documented in *Schizosaccharomyces pombe* and plants[8], but it remains unresolved in mammals. Nonetheless, noncoding RNAs also represent an integral component of mammalian heterochromatin. In mouse cells, RNA might have a structural role in heterochromatin, as cells treated with RNaseA lose heterochromatic marks[9]. Mouse heterochromatin is also a template for the synthesis of double-stranded RNA (dsRNA), which is generated by the bidirectional transcription of major satellite repeats[10]. The synthesis of these heterochromatic transcripts is tightly coupled to the cell cycle[11] and is essential for the establishment of heterochromatin during early mouse development[12,13]. Notably, transcriptional derepression of satellite repeats is a hallmark of pancreatic and epithelial cancers[14].

Although many genetic and epigenetic principles of heterochromatin have been uncovered over the past decade, the mechanisms that initiate heterochromatin formation and maintain its distinction from euchromatin remain elusive. DNA sequence composition, DNA methylation, histone modifications, histone variants and noncoding RNAs might be involved, but none of these seem to be specific to heterochromatin or euchromatin. Given this conundrum, we hypothesized that a distinct arrangement of DNA sequences and/or aberrant transcriptional activity such as a higher probability of dsRNA generation might be an early trigger to discriminate heterochromatin from euchromatin. Because heterochromatin is not transcriptionally inert, we reasoned that it could depend, as any regulatory euchromatic region, on sequence-specific transcription factors that modulate RNA production.

RESULTS

Pax3 binds to major satellite repeats

To examine transcription factors operating at heterochromatic regions, we first investigated transcription factors that had been shown to associate with heterochromatic core components[15]. Pax3 emerged as a particularly notable candidate because it has been reported to interact with both heterochromatin protein-1 (HP1) and the transcriptional regulator Kap1(Trim28)[16]. In addition, subrepeat

[1]Department of Epigenetics, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [2]Research Institute of Molecular Pathology, Vienna, Austria. [3]Department of Visceral Surgery, University of Freiburg Medical Center, Freiburg, Germany. [4]Spemann Graduate School of Biology and Medicine (SGBM), Albert Ludwigs University Freiburg, Freiburg, Germany. [5]Faculty of Biology, Albert Ludwigs University Freiburg, Freiburg, Germany. [6]Present addresses: Institute of Molecular Biotechnology, Vienna, Austria (J.P.) and University College London Genetics Institute, London, UK (B.G.). [7]These authors contributed equally to this work. Correspondence should be addressed to T.J. (jenuwein@ie-freiburg.mpg.de).

Received 9 May; accepted 15 August; published online 16 September 2012; corrected after print 27 December 2012; doi:10.1038/nsmb.2382

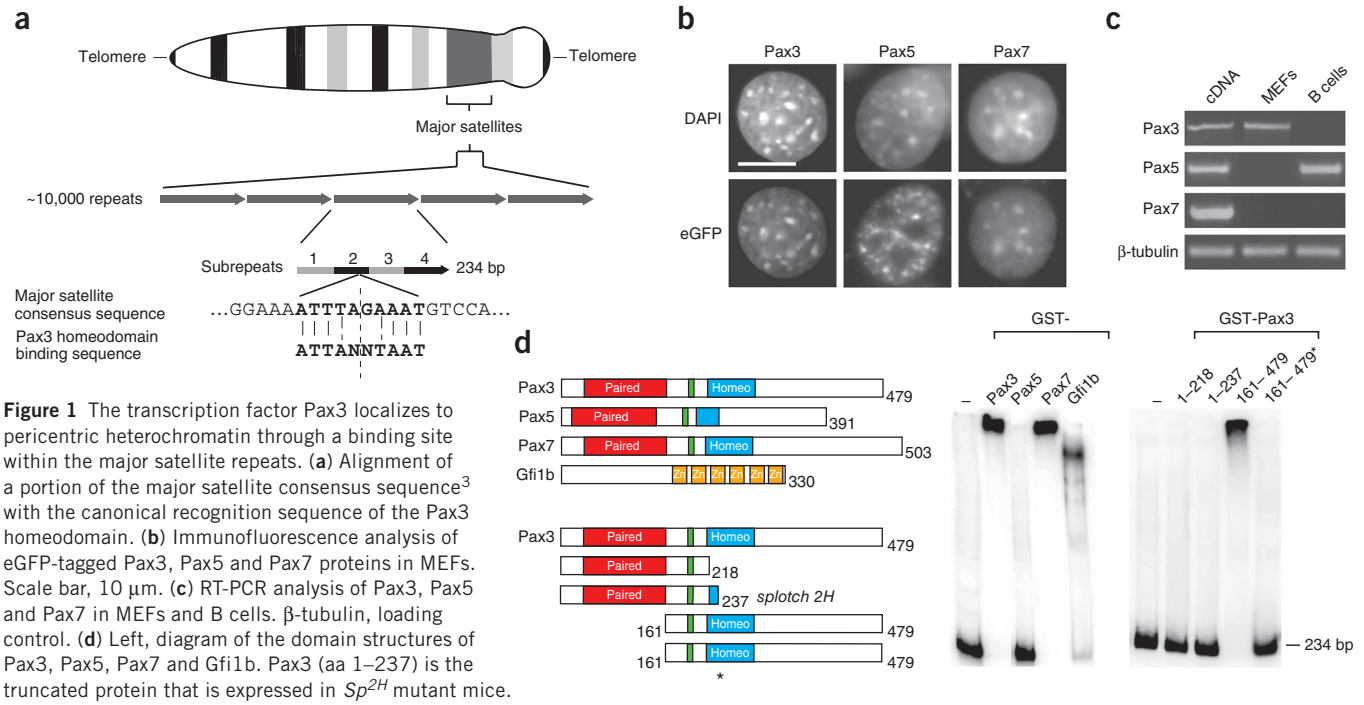


Figure 1 The transcription factor Pax3 localizes to pericentric heterochromatin through a binding site within the major satellite repeats. (**a**) Alignment of a portion of the major satellite consensus sequence[3] with the canonical recognition sequence of the Pax3 homeodomain. (**b**) Immunofluorescence analysis of eGFP-tagged Pax3, Pax5 and Pax7 proteins in MEFs. Scale bar, 10 µm. (**c**) RT-PCR analysis of Pax3, Pax5 and Pax7 in MEFs and B cells. β-tubulin, loading control. (**d**) Left, diagram of the domain structures of Pax3, Pax5, Pax7 and Gfi1b. Pax3 (aa 1–237) is the truncated protein that is expressed in *Sp*^{2H} mutant mice. Pax3 (aa 161–479*) contains a point mutation (S237A) in the recognition helix of the homeodomain. Paired domains are represented in red, homeodomains in blue, octapeptides in green and zinc fingers (Zn) in orange. Right, GST-tagged recombinant proteins were used for EMSA with a radioactively labeled full-length major satellite probe (234 bp). GST alone served as a negative control (–).

2 of the major satellite sequence contains a palindrome (**Fig. 1a**) that is very similar to the consensus sequence recognized by the Pax3 homeodomain[17,18]. Pax3 is a member of the mammalian Pax family, which comprises nine important developmental regulators[19,20]. To examine whether Pax3 constitutes a component of pericentric heterochromatin, we analyzed the subnuclear localization of Pax3 tagged with enhanced green fluorescent protein (Pax3-eGFP) in mouse embryonic fibroblasts (MEFs). Owing to its high similarity to Pax3 (ref. 20), we also included Pax7 in this analysis. Pax5, which does not contain an apparent consensus binding site within the major satellite repeats, served as a negative control. Although ectopic Pax5 also showed a speckled subnuclear pattern, only ectopic Pax3 and Pax7 colocalized with DAPI-dense foci (**Fig. 1b**), which occurred in approximately 30% of cells transfected. The remaining fraction of the cell population showed broad nuclear staining for the ectopic proteins, a distribution similar to that of endogenous Pax3 (**Supplementary Fig. 1a**). This variable staining pattern is consistent with previous reports on the subnuclear localization of Pax3-eGFP in MEFs[21]. Because neither Pax5 nor Pax7 is expressed in MEFs (**Fig. 1c**), we focused our further analysis on Pax3.

We next performed electrophoretic mobility shift assays (EMSAs) to assess direct binding of Pax3 to full-length major satellite repeats (**Fig. 1d**). In addition to glutathione *S*-transferase (GST)-tagged Pax3, we included GST-tagged Pax7, Pax5 and Gfi1b (a zinc-finger protein) as positive controls[22]. Pax3, Pax7 and Gfi1b reduced the mobility of the full-length major satellite probe, whereas Pax5 did not (**Fig. 1d**). Partial deletions of Pax3 revealed that a truncated protein (amino acids (aa) 161–479) containing the homeodomain is sufficient for binding (**Fig. 1d**). A point mutation in the recognition helix of the homeodomain (S273A) abolished DNA interaction (**Fig. 1d**), indicating that an intact homeodomain in Pax3 is required for major satellite binding. This result was in line with previous reports describing pericentric localization of the isolated Pax3 homeodomain[23].

Pax3 represses RNA output from mouse major satellite repeats

Pax3 has been described as an activator and repressor of transcription in euchromatic gene regulation[24]. To elucidate its function at pericentric heterochromatin, we analyzed the RNA output from this chromatin region in wild-type and Pax3-deficient cells. For this, we generated primary MEFs (pMEFs) from *splotch*-mutant (*Sp*^{2H}) mice[25,26] (**Supplementary Fig. 1a**), which carry a radiation-induced 32-bp deletion in the *Pax3* locus that leads to the expression of a truncated protein (aa 1–237) that lacks the homeodomain and is unable to bind to mouse major satellite repeats (**Fig. 1d**). In addition to these '*Pax3*-deficient' pMEFs, we also analyzed *Pax3*-null embryonic stem cells (ESCs) from mice (**Supplementary Fig. 1b**), in which a different *Pax3* mutation, *Sp*^{2G}, abrogates expression of the Pax3 protein as a consequence of insertion of *LacZ* in the first exon of the *Pax3* locus[27]. *Pax3*-null ESCs and *Pax3*-deficient pMEFs showed comparable impairment of pericentric heterochromatin (see below and **Supplementary Fig. 1c**). Because the subnuclear definition of heterochromatin was clearer in the MEFs, we focused the following RNA and immunofluorescence analyses on these cells.

To quantify RNA output, we performed reverse transcription followed by quantitative PCR (qRT-PCR) on total RNA isolated from wild-type and *Pax3*-deficient pMEFs and from MEFs lacking the genes encoding Suv39h1 and Suv39h2 (*Suv39h*–double-null), using a primer pair that amplifies one unit of the 234-bp mouse major satellite repeat. As previously described for *Suv39h*–double-null cells[10], *Pax3*-deficient cells also produced fivefold more major satellite transcripts than wild-type cells (**Fig. 2a**). Next, we asked whether these elevated levels could reflect an accumulation of dsRNA. To answer this, we first digested total RNA with RNaseONE to enrich for a pool of dsRNA[10] and then conducted qRT-PCR analysis. After this enrichment step, the signal for major satellite transcripts in *Pax3*-deficient cells was more than tenfold higher than in wild-type cells and, notably, in *Suv39h*–double-null cells (**Fig. 2b**). To directly

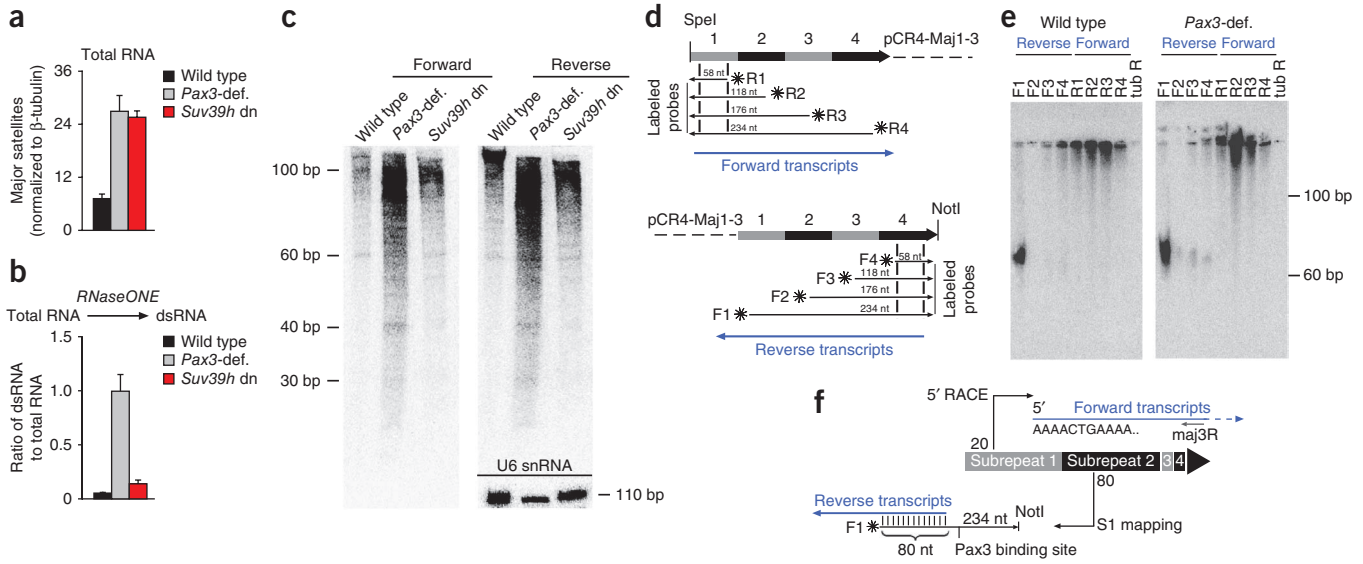


Figure 2 Pax3 represses transcription from major satellite repeats. (**a**) qRT-PCR for major satellite transcripts on total RNA isolated from wild-type, *Pax3*-deficient and *Suv39h*–double-null MEFs. (**b**) qRT-PCR for major satellite transcripts on dsRNA isolated from wild-type, *Pax3*-deficient and *Suv39h*–double-null MEFs, normalized to total RNA. (**c**) Northern blot analysis on total RNA isolated from wild-type, *Pax3*-deficient and *Suv39h*–double-null MEFs using strand-specific probes detecting forward and reverse transcripts. A probe directed against U6 small nuclear RNA (snRNA) served as a loading control. (**d**) Schematic representation of the labeled forward (F) and reverse (R) probes used in S1 nuclease digestion to detect TSSs of forward and reverse transcripts. pCR4-Maj1-3, plasmid containing three major satellite repeats. (**e**) S1 nuclease digestion assay using RNA from wild-type and *Pax3*-deficient MEFs. A tubulin probe (tub R) served as a negative control. (**f**) Diagram of TSSs at major satellite repeats: forward transcripts start within subrepeat 1 at position 20 (as mapped by 5′ RACE), whereas reverse transcripts originate from subrepeat 2 at position 80, close to the Pax3 binding site (as mapped by S1 nuclease digestion). *Pax3*-def., *Pax3*-deficient; dn, double-null; error bars, s.d.

demonstrate transcriptional activity from both DNA strands, we performed northern blots using strand-specific probes. A 15% gel was chosen to allow for detection of fragments shorter than 100 bp, which would not be amplified by qRT-PCR. These northern blots confirmed that both forward and reverse transcripts were substantially more abundant in *Pax3*-deficient than wild-type MEFs (**Fig. 2c**). We also observed an accumulation of smaller transcripts in *Pax3*-deficient cells as compared to *Suv39h*–double-null cells (**Fig. 2c**, see signal below the 100-bp marker). On the basis of these data, we concluded that Pax3 represses transcription from both strands of major satellite repeats.

We next mapped the transcriptional start sites (TSSs) of forward and reverse transcripts by S1 nuclease digestion and 5′ rapid amplification of cDNA ends (RACE) in wild-type and *Pax3*-deficient MEFs. We used two sets of four strand-specific DNA probes annealing with either forward or reverse transcripts (**Fig. 2d**) for the analysis. We found that TSSs were not altered in the absence of Pax3 function, although overall transcript levels were higher in RNA preparations from *Pax3*-deficient cells (**Fig. 2e**). Notably, one forward probe (F1) protected a prominent signal for reverse transcripts of 80 bp (**Fig. 2e**), indicating that the TSS for reverse transcription resides within subrepeat 2 and is in close proximity to the Pax3 binding site (**Fig. 2f**). Although we did not detect unambiguous protection for forward transcription, S1 nuclease mapping suggested that forward transcription could start in subrepeat 1. We verified this by 5′ RACE using the maj3R primer followed by sequencing of the RACE products, which showed that forward transcripts originate at nucleotide position 20 in subrepeat 1 (**Fig. 2f**). Subrepeat 1 has additional Pax3 binding activity (see below), but we could not predict a binding site on the basis of the DNA sequence. Together, these data indicated that Pax3 binding sites are located at the TSSs of both forward and reverse major satellite transcripts. Moreover, the deregulated transcriptional activity of the forward and reverse TSSs could generate dsRNA transcripts that would overlap by at least 60 bp (**Fig. 2f**).

Pax3 and Pax9 safeguard heterochromatin integrity

We next asked whether the upregulation of transcription from major satellite repeats in *Pax3*-deficient cells might also impair the epigenetic signature of pericentric heterochromatin. Indirect immunofluorescence for H3K9me3 and H4K20me3 showed that both marks were considerably reduced at pericentric heterochromatin in *Pax3*-deficient pMEFs (**Fig. 3a**) and *Pax3*-null ESCs (**Supplementary Fig. 1c**). Impaired chromatin signatures at pericentric heterochromatin are known to lead to dysfunctions in chromosome segregation; thus, there is often counterselection or compensation in mutant backgrounds, as exemplified by the redistribution of Polycomb components, such as the histone methyltransferase EZH2 and the associated H3K27me3 mark, to pericentric heterochromatin in *Suv39h*–double-null ESCs[28]. Whereas association of H3K27me3 with the inactive X chromosome was not altered in *Pax3*-deficient pMEFs from females (**Supplementary Fig. 1d**), *Pax3*-deficient pMEFs from both sexes recovered heterochromatic accumulation of H3K9me3 and H4K20me3 after extended cell passaging (**Fig. 3a**).

We reasoned that other Pax family members might be able to rescue impaired heterochromatin during the transition of *Pax3*-deficient pMEFs into immortalized fibroblasts (iMEFs). We therefore examined the expression of all nine Pax factors in wild-type and *Pax3*-deficient iMEFs by qRT-PCR and found substantial transcriptional upregulation of Pax2, Pax4 and, to a lesser extent, Pax9 in *Pax3*-deficient iMEFs (**Supplementary Fig. 2a**). Subsequent analysis by transient, short interfering RNA–mediated knockdown of Pax2, Pax4 or Pax9 (**Supplementary Note**) in *Pax3*-deficient iMEFs revealed redundant functions for Pax3 and Pax9 (but not for Pax2 and Pax4) in protecting heterochromatic H3K9me3 and H4K20me3 marks in iMEFs (**Supplementary Fig. 2b,c**).

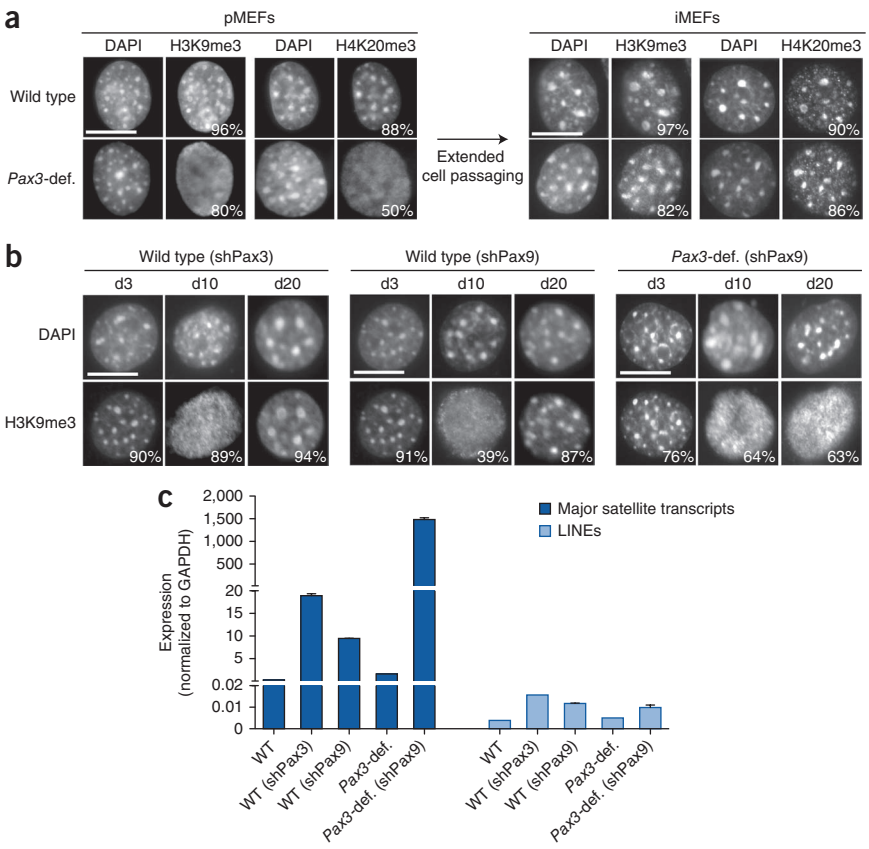


Figure 3 Pax3 and Pax9 have redundant functions in protecting mouse pericentric heterochromatin. (**a**) Immunofluorescence analysis of H3K9me3 and H4K20me3 in wild-type and *Pax3*-deficient pMEFs and iMEFs. Percentages of counted cells that displayed the depicted phenotype are indicated. Scale bars, 10 µm. (**b**) Immunofluorescence analysis of H3K9me3 in wild-type (Pax3) or wild-type (Pax9) iMEFs and in *Pax3*-deficient (shPax9) iMEFs. Representative images of cells at days 3, 10 and 20 after shRNA transduction are shown. Percentages of counted cells that displayed the depicted phenotype are indicated. Scale bars, 10 µm. (**c**) qRT-PCR analysis for transcripts originating from major satellite and LINE repeats in wild-type, wild-type (shPax3), wild-type (shPax9), *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs. Expression was normalized to *Gapdh* mRNA expression. WT, wild type. Error bars, s.d.

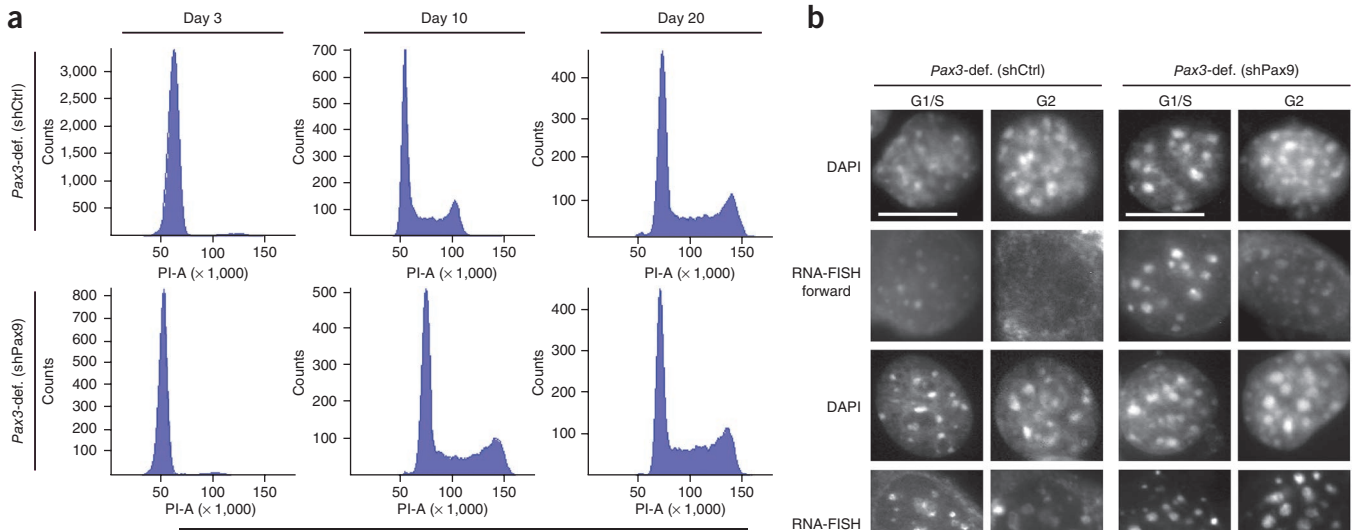
To assess the relative contributions of Pax3 and Pax9 in mammalian heterochromatin maintenance, we generated lentivirus constructs expressing short hairpin RNAs (shRNAs) for stable knockdown of Pax3 and Pax9. Infected wild-type iMEFs were selected for successful transduction, and samples for qRT-PCR (**Supplementary Fig. 3a**) and immunofluorescence analyses (**Fig. 3b** and **Supplementary Fig. 3b**) were taken at day 3, day 10 and day 20 after infection. In both the Pax3- and Pax9-knockdown cells, we observed transient dispersion of H3K9me3 foci at day 10 that had almost fully reverted to a wild-type staining pattern by day 20 (**Fig. 3b**). Notably, the percentage of cells that showed a transient loss of H3K9me3 foci was higher upon knockdown of Pax3 (89%) than Pax9 (39%) (**Fig. 3b**), suggesting that Pax3 is the primary factor in safeguarding heterochromatin integrity. We then generated a stable knockdown for Pax9 (denoted (shPax9)) in *Pax3*-deficient iMEFs. In contrast to our observations in wild-type iMEFs, knockdown of Pax9 in this *Pax3*-deficient background resulted in loss of H3K9me3 at pericentric heterochromatin that persisted at day 10 and day 20 (**Fig. 3b**). This suggested that, over the observed time window, there was no additional compensation, which was also indicated by progressive attenuation of cell viability and an increase in number of chromosome-segregation defects during the cultivation of *Pax3*-deficient (shPax9) iMEFs. These defects were visible as aberrant nuclear morphology, and by day 10, 74% of *Pax3*-deficient (shPax9) iMEFs showed hallmarks of mitotic dysfunction (such as multinucleated and polylobed cells)[29], whereas only 10% of *Pax3*-deficient iMEFs showed such irregularities (**Supplementary Fig. 3c**).

Cell-cycle analysis of major satellite transcription

Whereas single knockdown of Pax3 or Pax9 in wild-type iMEFs resulted in an increase in major satellite transcription—of around ten-fold for shPax9 and 18-fold for shPax3 (**Fig. 3c**)—there was dramatic derepression (a >1,000-fold increase) of major satellite transcripts in *Pax3*-deficient (shPax9) iMEFs (**Fig. 3c**). This result indicated that the functions of Pax3 and Pax9 are not only redundant but also synergistic in safeguarding heterochromatin silencing. The massive RNA output did not reflect general transcriptional dysregulation of repetitive elements, because transcript levels of other repeats, such as long interspersed nuclear elements (LINEs), were only slightly elevated (**Fig. 3c**)—although they were significantly upregulated in *Suv39h*–double-null cells (A.B.K, I.A.d.l.R.-V., B.G. and T.J., unpublished data; ref. 30).

These qRT-PCR experiments were performed on asynchronous cell populations and thus did not allow analysis of cell cycle–specific alterations in major satellite transcription. In wild-type cells, mouse heterochromatin shows the highest transcriptional activity at the G1/S transition and in early S phase and is then silenced during the G2 phase[11]. To demonstrate that the observed increase in major satellite transcripts was due to the loss of Pax3 and Pax9 rather than cell-cycle dysregulation, we analyzed the cell-cycle profile of *Pax3*-deficient cells infected with an empty lentiviral vector (shCtrl) and *Pax3*-deficient (shPax9) iMEFs at days 3, 10 and 20 after virus transduction. Over this time course, *Pax3*-deficient (shPax9) cells showed enrichment in G2/M phase (**Fig. 4a**), consistent with the observed mitotic defects (**Supplementary Fig. 3c**). We also observed fewer *Pax3*-deficient (shPax9) cells in G1/S and S phase than *Pax3*-deficient iMEFs. Because most major satellite transcripts are synthesized at the G1/S transition[11], we could thus conclude that the observed >1,000-fold increase in major satellite transcription was not a consequence of cell-cycle dysregulation.

To further address whether Pax3 and Pax9 contribute to the cell cycle–dependent regulation of major satellite transcription, we analyzed the RNA output from major satellite repeats at distinct cell-cycle stages in *Pax3*-deficient iMEFs and *Pax3*-deficient (shPax9) iMEFs by RNA fluorescence *in situ* hybridization (RNA-FISH). For this, we synchronized cells with aphidicolin and used a combination of four major satellite–specific RNA oligonucleotide probes detecting either forward or reverse transcripts. In *Pax3*-deficient iMEFs, forward transcripts are less abundant than reverse transcripts. In addition, reverse



	Day 3			Day 10			Day 20		
	G0/G1	S	G2/M	G0/G1	S	G2/M	G0/G1	S	G2/M
Pax3-def. (shCtrl)	84.1%	3.5%	12.4%	56.6%	17.4%	26.0%	45.6%	25.6%	28.8%
Pax3-def. (shPax9)	86.1%	3.4%	10.5%	51.1%	12.1%	36.8%	42.0%	16.4%	41.6%

Figure 4 *Pax3*-deficient (shPax9) iMEFs show increased major satellite transcripts in the G1/S and G2 phase of the cell cycle. (**a**) Cell-cycle analysis of *Pax3*-deficient (shCtrl) and *Pax3*-deficient (shPax9) iMEFs. Fluorescence-activated cell sorting profiles of DNA content at days 3, 10 and 20 after shRNA transduction are shown. Percentages of cells in each cell-cycle stage (G0/G1, S and G2/M) are listed in the table below. PI-A, propidium iodide area. (**b**) RNA-FISH detecting forward and reverse major satellite transcripts in synchronized *Pax3*-deficient (shCtrl) and *Pax3*-deficient (shPax9) iMEFs. Cells were arrested with aphidicolin at the G1/S transition and subsequently released into G2 phase. On average, 250 cells were analyzed per sample, and representative images are shown. Scale bars, 10 μm.

transcripts show a more prominent signal at the G1/S transition, and this signal is downregulated in G2 phase (**Fig. 4b**). In *Pax3*-deficient (shPax9) iMEFs, we observed marked derepression of forward transcripts at the G1/S transition and enhanced transcription from the reverse strand in G2 phase (**Fig. 4b**). These data indicate an important role for Pax3 and Pax9 in repression of RNA output from major satellite repeats in the G1/S and G2 stages of the cell cycle.

Intact transcription factor binding sites recruit H3K9me3 to major satellites

The above data demonstrate that the Pax3 and Pax9 transcription factors are components of mouse heterochromatin. Although we observed a distinct heterochromatic accumulation only for eGFP-Pax3 with immunofluorescence (**Supplementary Fig. 4a**), we detected both eGFP-tagged Pax factors at mouse pericentric heterochromatin by chromatin immunoprecipitation (ChIP) (**Supplementary Fig. 4b**). Furthermore, we could also demonstrate that Pax3 and Pax9 bind *in vitro* to full-length major satellite repeats and distinct subrepeats (**Supplementary Fig. 4c**). These results, in combination with previously published results[22,31], indicate that the 234-bp major satellite repeat represents an assembly of various transcription factor binding sites, including Pax3, Gfi1b, Sall1 and Zeb1 (**Fig. 5a**), and it is likely that other transcription factor binding sites are also present. The overlap of some of these transcription factor binding sites suggested that each major satellite repeat could serve as a base for bidirectional transcription, which is then silenced by heterochromatin formation.

It is very difficult to investigate this possibility, however, because the sequence identity of the major satellite repeats defies annotation of individual repeat entities within the large array of pericentric heterochromatin. But there are 42 intergenic major satellite sequences interspersed in the mouse genome, and these can be mapped to distinct chromatin regions (outside pericentric heterochromatin) according to their single nucleotide polymorphisms[32]. Of those intergenic major satellite repeats, seven sequences maintain binding sites for Pax3 and most of the other transcription factors, and 35 sequences show severely permutated transcription factor binding sites. These intergenic major satellite sequences are spread across the mouse chromosomes and are listed in **Supplementary Table 1** as C1–C7 (conserved Pax3 binding site in subrepeat 2) or M1–M35 (mutated Pax3 binding site in subrepeat 2).

We performed ChIP sequencing (ChIP-seq) for H3K9me3 in wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs to identify whether nearly intact or permutated major satellite sequences would differ in the accumulation of H3K9me3. This analysis was done with two distinct H3K9me3-specific antibodies (no. 1926 and no. 4861, generated in our laboratory) covering an average depth of 50 million reads uniquely mapped to the genome (**Supplementary Table 1**). For the major satellite consensus sequence, there was strong enrichment of H3K9me3 in wild-type iMEFs, which was modestly attenuated in *Pax3*-deficient iMEFs but substantially reduced in *Pax3*-deficient (shPax9) iMEFs (**Fig. 5b**). The 20% reduction in H3K9me3 reads that we observed in *Pax3*-deficient (shPax9) iMEFs was much smaller than the >90% reduction in H3K9me3 enrichment observed in directed ChIP experiments (**Supplementary Fig. 5a**,**b**). This difference was probably due to technical limitations of the library preparation, which, even at only 18 PCR cycles, did not allow linear amplification of repetitive sequences.

For conserved intergenic major satellite repeats in wild-type iMEFs, there were between 9,000 and 1,300,000 mapped reads for H3K9me3, whereas the number of H3K9me3 tags for permutated major satellite repeats was generally <50 and only in a few cases reached >500 H3K9me3 reads (**Fig. 5b** and **Supplementary Table 1**). Consistent with

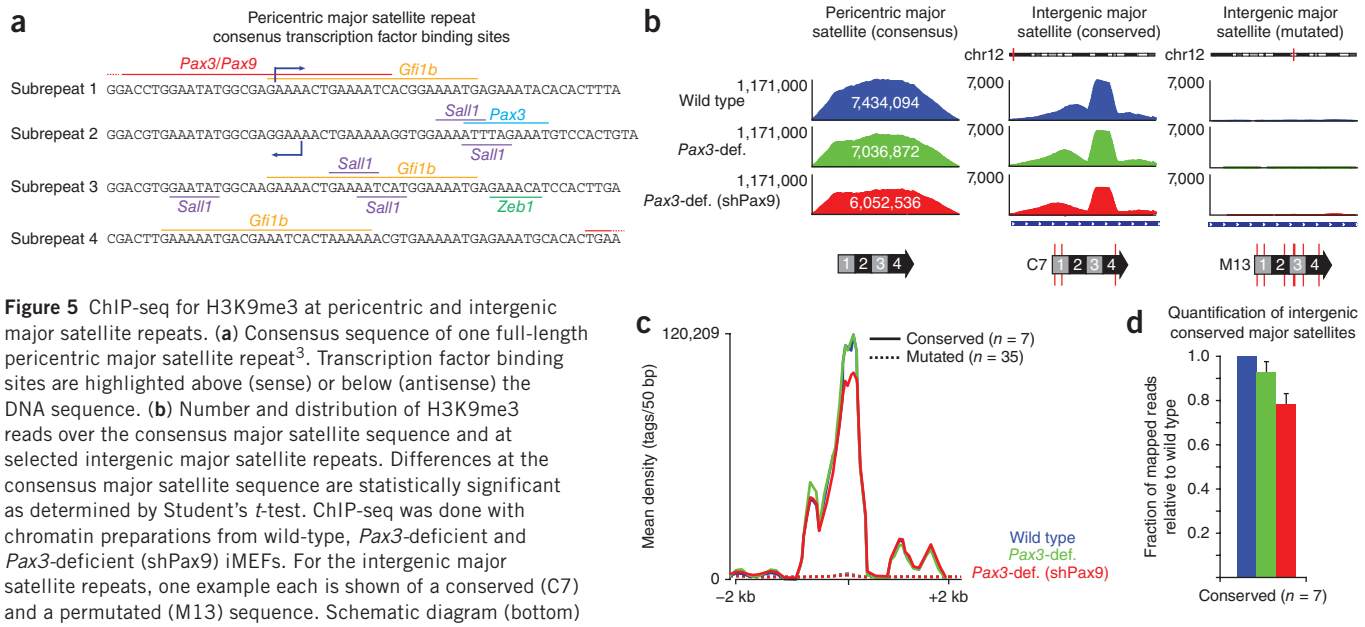


Figure 5 ChIP-seq for H3K9me3 at pericentric and intergenic major satellite repeats. (**a**) Consensus sequence of one full-length pericentric major satellite repeat[3]. Transcription factor binding sites are highlighted above (sense) or below (antisense) the DNA sequence. (**b**) Number and distribution of H3K9me3 reads over the consensus major satellite sequence and at selected intergenic major satellite repeats. Differences at the consensus major satellite sequence are statistically significant as determined by Student's *t*-test. ChIP-seq was done with chromatin preparations from wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs. For the intergenic major satellite repeats, one example each is shown of a conserved (C7) and a permutated (M13) sequence. Schematic diagram (bottom) shows positions of mutated transcription factor binding sites (red bars). (**c**) Meta-analysis of H3K9me3 enrichment in conserved ($n = 7$) or permutated ($n = 35$) intergenic major satellite repeats in wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs. Sequences were aligned and the number of mapped reads (tags) was calculated in bins of 50 bp. (**d**) Quantification of H3K9me3 enrichment at conserved ($n = 7$) intergenic major satellites in wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs. Reads were normalized and data are presented as a fraction relative to the number of wild-type reads, which was set to 1. chr, chromosome. Error bars, s.d.

the general reduction of H3K9me3 enrichment for the consensus major satellite sequence in *Pax3*-deficient (shPax9) iMEFs, there were also substantially fewer (an average reduction of ~25%) H3K9me3 reads at the seven conserved intergenic major satellite repeats (**Fig. 5c**,**d** and **Supplementary Table 1**). These data verified a role for Pax3 and Pax9 in heterochromatin protection and further suggested that initiation of heterochromatin formation depends in part on the number and reiteration of intact transcription factor binding sites within repeat sequences.

Transcription factor binding motifs in Suv39h-dependent heterochromatin

We next asked whether a reiterated distribution of transcription factor binding motifs could constitute a general denominator for the definition of heterochromatin. For this, we extended our analysis to other heterochromatic repeat regions that were identified by genome-wide ChIP-seq for H3K9me3 in wild-type and *Suv39h*–double-null ESCs (A.B.K., I.A.d.l.R.-V. and T.J., unpublished data). We identified 6,387 chromatin regions that showed a substantial decrease in H3K9me3 in *Suv39h*–double-null cells. This Suv39h-dependent heterochromatin was established mainly across DNA sequences containing endogenous retroviruses (primarily long terminal repeats (LTRs)) and LINE elements (see example in **Fig. 6a**).

We assembled the DNA sequences underlying these H3K9me3 peaks and interrogated the data set with the transcription factor affinity prediction (TRAP) algorithm[33–35], which can estimate binding affinities for transcription

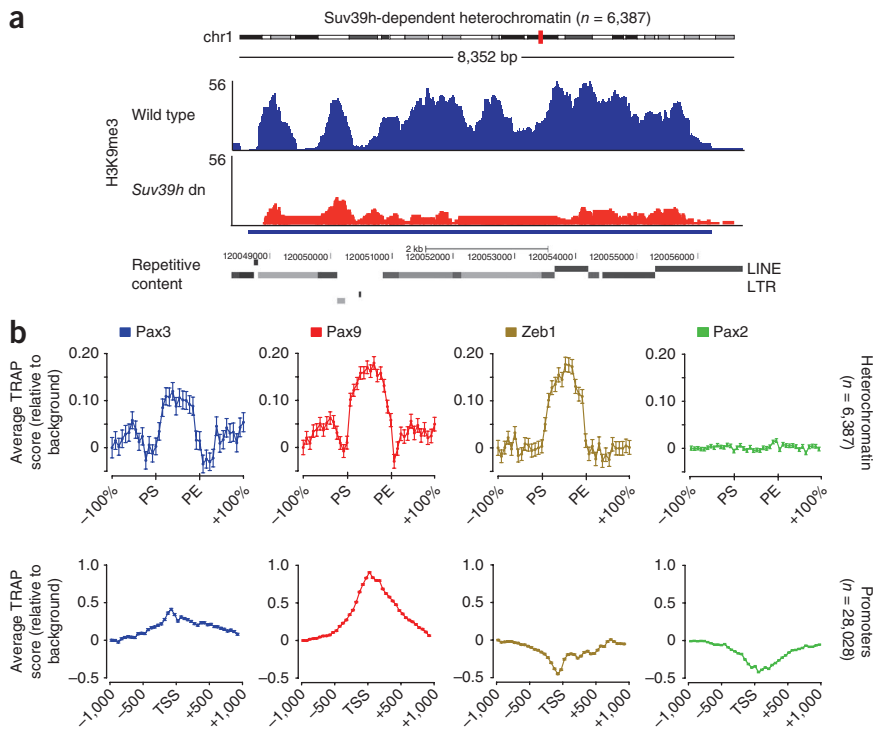


Figure 6 Genome-wide analysis and bioinformatic interrogation of Suv39h-dependent H3K9me3 heterochromatin. (**a**) ChIP-seq for H3K9me3 in wild-type and *Suv39h*–double-null mouse ESCs. One representative example (an 8.4-kb LTR-rich region at position 120 megabases in chromosome 1) from a total of 6,387 identified regions is shown. (**b**) TRAP scores for Pax3, Pax9, Zeb1 and Pax2 binding sites in Suv39h-dependent heterochromatic regions ($n = 6,387$) or at annotated promoter sequences ($n = 28,028$). chr, chromosome; PS, peak start; PE, peak end; dn, double-null. Error bars, s.e.m.

factors on the basis of their annotated binding sites (Genomatix MatBase). We focused on the Pax3, Pax9 and Zeb1 binding sites, as we had already detected these in the major satellite repeats (**Fig. 5a**). We also included the binding motif for Pax2, for which we observed neither a DNA binding sequence in the major satellite repeats nor an accumulation of eGFP-Pax2 at pericentric heterochromatin (**Supplementary Fig. 4a–c**). TRAP scores for Pax3, Pax9 and Zeb1 binding sites showed strong enrichment across the heterochromatic peak regions, extending from the peak start to the peak end (**Fig. 6b**) and remaining high over a domain that spanned from 582 bp to >36 kilobases (kb), depending on the extension of the underlying repeat arrays. There was no such enrichment for Pax2 binding motifs.

We next compared this TRAP profile with a set of DNA sequences corresponding to the 28,028 annotated promoters in the mouse genome (Ensembl version 51) and spanning from 1-kb upstream to 1-kb downstream of the annotated TSSs. There was a clear correlation with the TSSs for all four transcription factor binding sites, but Pax3 and Pax9 motifs were enriched at promoter regions, whereas Zeb1 and Pax2 motifs seemed to be underrepresented (**Fig. 6b**). Notably, and in contrast to the TRAP scores across the set of heterochromatic regions, the TSS association of Pax3 and Pax9 binding sites was concentrated in a narrower window <310 bp from the TSS.

DISCUSSION

Our study reveals a paradigm pathway in which transcriptional repression of pericentric repeats by sequence-specific transcription factors is essential for the integrity of heterochromatin, considerably expanding the role of transcription factors beyond euchromatic gene regulation. This function is not restricted to Pax3 and Pax9, as mouse major satellite repeats contain multiple binding sites for Sall1 (ref. 31), Gfi1b (ref. 22), Zeb1 (**Fig. 5a**) and probably many other transcription factors. Thus, in cell types other than fibroblasts or ESCs, other Pax proteins (such as Pax7 in muscle cells) or cell type–specific transcription factors could mediate heterochromatin silencing.

Moreover, most heterochromatic sequences in other model organisms also comprise transcription factor binding sites that are embedded in repetitive elements. For example, the *S. pombe mat* locus includes binding sites for transcription factors of the CREB family[36], and in *Drosophila melanogaster*, zinc finger–containing proteins such as GAGA[37], Prod[38] and Su(var)3–7 (ref. 39) associate with the heterochromatic chromocenter. In addition, bioinformatic interrogation of human α-satellite sequences revealed several transcription factor binding sites (data not shown) and exposed ZEB1, an important developmental regulator containing zinc fingers and a homeodomain[40], as a potential factor in safeguarding human heterochromatin. Indeed, we found that depletion of Zeb1 in iMEFs also impairs pericentric heterochromatin by a mechanism similar to that described for Pax3 and Pax9 (**Supplementary Fig. 6**). It is conceivable that the great variety of zinc-finger factors (>600 in the mammalian genome) has evolved, at least in part, to restrict transcriptional output from the vast pool of highly divergent repetitive segments in the eukaryotic genome.

Our analysis provides compelling evidence that transcription factors have a function outside of euchromatic gene regulation. What, then, distinguishes euchromatin from heterochromatin? We propose a model in which this distinction ultimately resides in the transcriptional control and RNA output of the underlying DNA sequence. In this model, *any* DNA sequence containing transcription factor binding sites would have the potential to be transcribed. Whereas euchromatic gene transcription has evolved to select a cooperative combination of transcription factor binding sites at regulatory modules such as promoters and enhancers, heterochromatic sequences largely lack

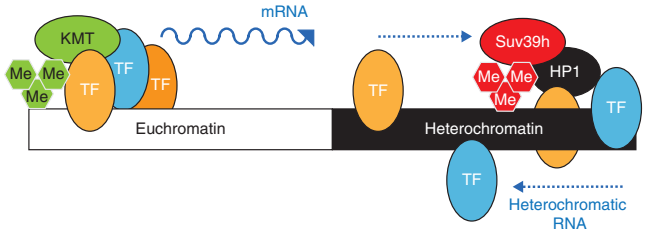


Figure 7 A transcription factor–based model for mouse heterochromatin formation. We propose a model in which the distinction between euchromatin and heterochromatin is based on a synergistic organization, versus a more random distribution, of transcription factor binding sites. TF, transcription factor; KMT, histone methyltransferase; Me, methyl group.

this synergy and present a more uncoordinated and reiterated composition of transcription factor binding sites (see for example, the overlapping binding sites for Pax3, Sall1 and Gfi1b shown in **Fig. 5a**) that would defy the generation of a highly controlled and processed primary RNA transcript. Although heterochromatic repeat sequences are not transcriptionally inert initially, repression mechanisms prevail, resulting in the silencing of the underlying chromatin region (**Fig. 7**). This silencing is mediated by direct transcriptional repression through transcription factors and can be further stabilized by chromatin-based mechanisms such as transcription factor–coupled[16,41] or RNA-guided[42,43] recruitment of histone methyltransferases and other enzymes (**Fig. 7**). Our model would also explain earlier observations, where the quantity and/or quality of RNA transcripts would be a major signal in the definition of chromatin regions as euchromatic or heterochromatic. For example, heterochromatin can be converted into euchromatin by the insertion of strong enhancers and promoters[44,45]. Conversely, if the processing and export of primary transcripts is impaired, euchromatic positions can adopt heterochromatic marks, particularly when there is accumulation of dsRNA[46,47]. Our data suggest that one major factor in the discrimination of euchromatin from heterochromatin is the synergistic organization, versus a more random distribution, of transcription factor binding sites.

METHODS

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

Accession codes. Gene Expression Omnibus: fastq files of H3K9me3 ChIP-seq in wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs have been deposited with the accession number GSE40086.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank A. Copp (University College London Institute of Child Health) for providing *Sp*^{2H} mice, A. Mansouri (Max Planck Institute for Biophysical Chemistry) for *Pax3*-null ESCs and M. Busslinger (Research Institute of Molecular Pathology) for the Pax5 antibody. The Pax3 monoclonal antibody developed by C. Ordahl was obtained from the Developmental Studies Hybridoma Bank, developed under the auspices of the US National Institute of Child Health and Human Development and maintained by the University of Iowa Department of Biological Sciences. We thank D. Gilbert (Florida State University) and R. Agrelo (Institut Pasteur de Montevideo) for advice on RNA-FISH protocols and S. Rudloff (Max Planck Institute of Immunobiology and Epigenetics) for providing the pCAGGs plasmid. ChIP-seq was performed at the Deep-Sequencing/Bioinformatics Unit of the Max Planck Institute of Immunobiology and Epigenetics. In particular, we would like to thank S. Diehl for processing and quality control of the raw data and N. Noureen and F. Ramirez for help with the genome-wide analyses. This work was supported by the Research Institute of

Molecular Pathology through Boehringer Ingelheim (T.J.), the European Union Network of Excellence 'The Epigenome' (T.J., LSHG-CT-2004-503433), the Genome Research in Austria initiative (T.J.), the German Research Foundation (T.J. and T.B., CRC992 Medical Epigenetics) and the Max Planck Society (T.J.) and the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School; S.M.).

AUTHOR CONTRIBUTIONS

A.B.K. and V.P. designed and performed most of the experiments, analyzed the data and contributed equally to this manuscript. M.S. made the initial observation of Pax3 localization to major satellites, generated the *Pax3*-deficient MEFs and contributed to the analysis of RNA output in *Pax3*-deficient MEFs. I.A.d.l.R.-V. did ChIP-seq and contributed to the analysis of intergenic major satellites. S.v.d.N. carried out the Pax9 EMSA experiments, J.P. performed the northern blot analysis and B.G. conducted the dsRNA analysis. N.S., S.O. and M.P. provided excellent technical assistance. S.M. and T.B. shared expertise and reagents for the analysis of Zeb1 in mouse and human cells. T.M. did the TRAP analysis of all Suv39h-dependent heterochromatic regions. M.L. and T.J. advised on experimental design and interpretation of data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2382.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Allshire, R.C. & Karpen, G.H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* **9**, 923–937 (2008).
2. Henikoff, S., Ahmad, K. & Malik, H.S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
3. Vissel, B. & Choo, K.H. Mouse major (γ) satellite DNA is highly conserved and organized into extremely long tandem arrays: implications for recombination between nonhomologous chromosomes. *Genomics* **5**, 407–414 (1989).
4. Garagna, S., Zuccotti, M., Capanna, E. & Redi, C.A. High-resolution organization of mouse telomeric and pericentromeric DNA. *Cytogenet. Genome Res.* **96**, 125–129 (2002).
5. Allis, C.D., Jenuwein, T., Reinberg, D. & Caparros, M.-L. *Epigenetics* Ch. 3, 23–61 (Cold Spring Harbor Laboratory Press, 2007).
6. Lachner, M., Sengupta, R., Schotta, G. & Jenuwein, T. Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 209–218 (2004).
7. Grewal, S.I. & Elgin, S.C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406 (2007).
8. Martienssen, R.A., Kloc, A., Slotkin, R.K. & Tanurdzic, M. Epigenetic inheritance and reprogramming in plants and fission yeast. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 265–271 (2008).
9. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
10. Martens, J.H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
11. Lu, J. & Gilbert, D.M. Proliferation-dependent and cell cycle–regulated transcription of mouse pericentric heterochromatin. *J. Cell Biol.* **179**, 411–421 (2007).
12. Probst, A.V. *et al.* A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* **19**, 625–638 (2010).
13. Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat. Cell Biol.* **12**, 853–862 (2010).
14. Ting, D.T. *et al.* Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* **331**, 593–596 (2011).
15. Fodor, B.D., Shukeir, N., Reuter, G. & Jenuwein, T. Mammalian *Su(var)* genes in chromatin control. *Annu. Rev. Cell Dev. Biol.* **26**, 471–501 (2010).
16. Hsieh, M.J., Yao, Y.L., Lai, I.L. & Yang, W.M. Transcriptional repression activity of PAX3 is modulated by competition between corepressor KAP1 and heterochromatin protein 1. *Biochem. Biophys. Res. Commun.* **349**, 573–581 (2006).
17. Chalepakis, G., Wijnholds, J. & Gruss, P. Pax-3–DNA interaction: flexibility in the DNA binding and induction of DNA conformational changes by paired domains. *Nucleic Acids Res.* **22**, 3131–3137 (1994).
18. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. & Desplan, C. Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.* **7**, 2120–2134 (1993).
19. Robson, E.J., He, S.J. & Eccles, M.R.A. PANorama of PAX genes in cancer and development. *Nat. Rev. Cancer* **6**, 52–62 (2006).
20. Lang, D., Powell, S.K., Plummer, R.S., Young, K.P. & Ruggeri, B.A. PAX genes: roles in development, pathophysiology, and cancer. *Biochem. Pharmacol.* **73**, 1–14 (2007).
21. Corry, G.N., Hendzel, M.J. & Underhill, D.A. Subnuclear localization and mobility are key indicators of PAX3 dysfunction in Waardenburg syndrome. *Hum. Mol. Genet.* **17**, 1825–1837 (2008).
22. Vassen, L., Fiolka, K. & Moroy, T. Gfi1b alters histone methylation at target gene promoters and sites of γ-satellite containing heterochromatin. *EMBO J.* **25**, 2409–2419 (2006).
23. Corry, G.N. *et al.* The PAX3 paired domain and homeodomain function as a single binding module *in vivo* to regulate subnuclear localization and mobility by a mechanism that requires base-specific recognition. *J. Mol. Biol.* **402**, 178–193 (2010).
24. Chi, N. & Epstein, J.A. Getting your Pax straight: Pax proteins in development and disease. *Trends Genet.* **18**, 41–47 (2002).
25. Chan, W.Y., Cheung, C.S., Yung, K.M. & Copp, A.J. Cardiac neural crest of the mouse embryo: axial level of origin, migratory pathway and cell autonomy of the *splotch* (*Sp*^{2H}) mutant effect. *Development* **131**, 3367–3379 (2004).
26. Epstein, D.J., Vekemans, M. & Gros, P. *Splotch* (*Sp*^{2H}), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* **67**, 767–774 (1991).
27. Mansouri, A., Pla, P., Larue, L. & Gruss, P. Pax3 acts cell autonomously in the neural tube and somites by controlling cell surface properties. *Development* **128**, 1995–2005 (2001).
28. Peters, A.H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
29. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell-division genes. *Nature* **464**, 721–727 (2010).
30. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
31. Yamashita, K., Sato, A., Asashima, M., Wang, P.C. & Nishinakamura, R. Mouse homolog of *SALL1*, a causative gene for Townes-Brocks syndrome, binds to A/T-rich sequences in pericentric heterochromatin via its C-terminal zinc-finger domains. *Genes Cells* **12**, 171–182 (2007).
32. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
33. Manke, T., Roider, H.G. & Vingron, M. Statistical modeling of transcription factor binding affinities predicts regulatory interactions. *PLoS Comput. Biol.* **4**, e1000039 (2008).
34. Roider, H.G., Kanhere, A., Manke, T. & Vingron, M. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* **23**, 134–141 (2007).
35. Thomas-Chollier, M. *et al.* Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nat. Protoc.* **6**, 1860–1869 (2011).
36. Jia, S., Noma, K. & Grewal, S.I. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
37. Tsukiyama, T., Becker, P.B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **367**, 525–532 (1994).
38. Suso Platero, J., Csink, A.K., Quintanilla, A. & Henikoff, S. Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in *Drosophila*. *J. Cell Biol.* **140**, 1297–1306 (1998).
39. Cléard, F. & Spierer, P. Position-effect variegation in *Drosophila*: the modifier Su(var)3–7 is a modular DNA-binding protein. *EMBO Rep.* **2**, 1095–1100 (2001).
40. Vandewalle, C., Van Roy, F. & Berx, G. The role of the ZEB family of transcription factors in development and disease. *Cell. Mol. Life Sci.* **66**, 773–787 (2009).
41. Delattre, M., Spierer, A., Jaquet, Y. & Spierer, P. Increased expression of *Drosophila* Su(var)3–7 triggers Su(var)3–9-dependent heterochromatin formation. *J. Cell Sci.* **117**, 6239–6247 (2004).
42. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
43. Tsai, M.C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
44. Ahmad, K. & Henikoff, S. Modulation of a transcription factor counteracts heterochromatic gene silencing in *Drosophila*. *Cell* **104**, 839–847 (2001).
45. Festenstein, R. *et al.* Locus control region function and heterochromatin-induced position effect variegation. *Science* **271**, 1123–1125 (1996).
46. Reyes-Turcu, F.E., Zhang, K., Zofall, M., Chen, E. & Grewal, S.I. Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat. Struct. Mol. Biol.* **18**, 1132–1138 (2011).
47. Zofall, M. *et al.* RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* **335**, 96–100 (2012).
48. Peters, A.H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
49. Schotta, G. *et al.* A silencing pathway to induce H3–K9 and H4–K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
50. 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).

ONLINE METHODS

Cell lines and tissue culture. *Sp*^{2H} (*Pax3*-deficient) heterozygous mice were obtained from the laboratory of A. Copp[25] and crossed to homozygosity. Wild-type and *Pax3*-deficient pMEFs were derived and immortalized following the standard 3T3 protocol[48]. *Pax3*-null ESCs were obtained from A. Mansouri[27] and cultured on a feeder layer in standard embryonic stem cell medium. Feeder-independent embryonic stem cell lines (wild type and *Suv39h* double-null) were grown on dishes coated with 0.2% gelatin.

Transient transfection of eGFP-tagged Pax factors. MEFs were transfected with plasmids encoding Pax3-eGFP, Pax5-eGFP and Pax7-eGFP under the control of a CMV promoter using PLUS-Reagent and Lipofectamine (Invitrogen). For eGFP localization studies, cells were plated on glass chamber slides, fixed in 2% PFA and analyzed 48 h after transfection.

Immunofluorescence. Immunofluorescence analysis was performed as described[28] and with the following antibodies and dilutions: antibody to H3K9me3 (anti-H3K9me3; 1:1,000; ref. 28), anti-H3K27me3 (1:1,000; ref. 28), anti-H4K20me3 (1:1,000; ref. 49), anti-Pax3 (1:100; the Developmental Studies Hybridoma Bank, Iowa University), anti-GFP (1:500; Invitrogen A11122).

Expression profiling by qRT-PCR. Total RNA was extracted with TRIzol (Invitrogen) and reverse transcribed with random hexamers and SuperScript II reverse transcriptase (Invitrogen). First-strand cDNA was then subjected to qRT-PCR using specific primer sets. Primer sequences are listed in the **Supplementary Table 2**.

Electrophoretic mobility shift assay. We amplified full-length major satellite repeats (234 bp) using pCR-Maj1-3 (ref. 50) as a template and radioactively end-labeled with ^{32}P by T4 polynucleotide kinase (T4-PNK). For the analysis of subrepeats, DNA oligos representing subrepeat 1 and subrepeat 2 were hybridized and radioactively end-labeled with ^{32}P. A probe derived from the *Drosophila* even-skipped promoter (e5) served as a positive control. Sequences of all probes and primers are listed in the **Supplementary Table 2**.

GST-tagged proteins were expressed and purified as described[49] and incubated with labeled probes. Binding reactions had a total volume of 20 µl containing 4 µl 5× binding buffer (20% glycerol, 5 mM $MgCl_2$, 2.5 mM EDTA, 250 mM NaCl, 50 mM TRIS-HCl (pH 7.5)), 0.2 µl 0.1 M DTT, 1 µl poly(dI-dC) (300 ng/µl) and 20,000 c.p.m. of the labeled probe, 200 ng recombinant protein, and loaded on a 5% polyacrylamide native gel. The gel was dried and exposed to Phosphor Imager overnight.

RNaseONE treatment. Total RNA (5 µg) was digested with 2 units RNaseONE (Promega, M4261) for 30 s, extracted with RNeasy (Qiagen) and converted into cDNA, which was subsequently amplified with primers specific for the major satellites (**Supplementary Table 2**; RT-PCR primers).

Northern blot. We purified total RNA (10 µg) using a MirVana kit (Ambion, AM1560), separated it on a 15% polyacrylamide-urea gel (National Diagnostic) and transferred it to a nylon membrane (GeneScreen Plus HybridizationTransfer Membrane, Perkin Elmer). RNA probes were synthesized by *in vitro* transcription of pCR4-Maj1-3 (ref. 50) with either T3 or T7 RNA polymerase and digested in alkaline buffer to generate 15-nt-long fragments. Prehybridization (2 h) and hybridization (overnight) were done at 50 °C with 5× saline-sodium citrate (SSC), 20mM Na_2HPO_4, 7% SDS, 1× Denhardt's solution, 30% formamide and 1 mg sheared salmon sperm (Ambion, AM9680). The membrane was washed once with 2× SSC, 1% SDS and once with 0.5× SSC, 1% SDS and exposed to Phosphor Imager overnight.

S1 nuclease digestion assay. pCR4-Maj1-3 (ref. 50) was linearized with NotI for the synthesis of the forward probes and with SpeI for the synthesis of the reverse probes. Primers used for the synthesis of the single-stranded probes are listed in **Supplementary Table 2**. We end-labeled probes with ^{32}P by T4-PNK and used 50,000 c.p.m. of each probe for hybridization. Hybridization, S1 nuclease digestion (Sigma Aldrich, N5661-50KU) and recovery of the undigested DNA probe were performed according to a standard protocol[51]. Dried gel was exposed for 48 h to a Phosphor Imager cassette.

5′ RACE. Nuclear RNA was purified with a MirVana kit (Ambion, AM1560). cDNA was synthesized with the major3 reverse primer and polyadenylated with terminal transferase (Roche, 03 333 566 001). Amplification was achieved with the major3 reverse and adaptor-polyT primers (**Supplementary Table 2**). The resulting PCR products were phosphorylated, cloned and sequenced.

Stable shRNA-mediated knockdown of Pax3, Pax9 and Zeb1. For the initial characterization, we designed five shRNAs against Pax3 using the Invitrogen RNAi tool and cloned them into pLenti4-SV40-mCherry. Constructs were packaged into lentiviral particles for efficient transduction of wild-type iMEFs. Cells were sorted for mCherry expression. For Pax9 and Zeb1, five shRNAs were purchased from the Sigma Mission RNAi library, packaged into lentiviral particles and transduced into wild-type and *Pax3*-deficient iMEFs. Transduced cells were selected with puromycin. The shRNA sequences resulting in the best knockdown efficiency are listed in **Supplementary Table 2**. Cells were collected at three distinct time points (day 3, day 10 and day 20 after transduction), analyzed for knockdown efficiency by qRT-PCR and stained for H3K9me3 by immunofluorescence. Heterochromatic staining as well as cell morphology was scored on an average of 500 cells.

Cell-cycle analysis. A single suspension of 10^7 cells was fixed in 70% ethanol for 2 h at −20 °C, then resuspended in 1 ml of propidium iodide–Triton X-100 solution with RNaseA and incubated at room temperature for 30 min before performing flow cytometry.

Cell synchronization. Exponentially growing cells were first arrested with 0.05 µg/ml nocodazole for 4 h and harvested by mitotic shake-off. Cells were washed twice with PBS and resuspended in normal growth medium. After 5 h, aphidicolin was added to the medium to a final concentration of 10 µg/ml and cells were incubated for 10–12 h. Synchronized cells were subsequently released into aphidicolin-free medium and harvested at time points corresponding to G1/S (0 h) and G2 (10 h). Synchronized cell-cycle progression was monitored by FACS analysis of DNA content.

RNA-FISH. For RNA-FISH, cells were plated on chamber slides and fixed with 4% PFA for 10 min, permeabilized with triton X-100 for 5 min and fixed again with 4% PFA for 10 min. The slides were then washed for 3 min each with 70% EtOH, 80% EtOH, 95% EtOH and 100% EtOH. For detection of major satellite transcripts, we hybridized a mixture of the specific probes containing locked oligonucleotides (sequences are listed in the **Supplementary Table 2**) overnight at 37 °C in 2× SSC and 50% formamide. Slides were washed twice with 2× SSC, 50% formamide and 0.1% SSC. Cells were then blocked with 2.5% BSA diluted in 4× SSC and 0.1% Tween, incubated with avidin–Alexa 488 for 30 min at 37 °C and washed twice with 0.1% Tween in PBS. On average, 250 cells were analyzed for the RNA-FISH signal.

Chromatin immunoprecipitation. ChIP was performed according to standard protocols[10] and immunoprecipitated DNA was analyzed by qRT-PCR with primers specific for major satellites (see **Supplementary Table 2**, RT-PCR primers). H3K9me3-specific antibodies (5 µl crude serum per 25 µg chromatin; no. 4861 and no. 1926, Jenuwein laboratory) were used for ChIP in wild-type, *Pax3*-deficient and *Pax3*-deficient (shPax9) iMEFs. Chromatin from iMEFs overexpressing eGFP-tagged Pax proteins was precipitated with a ChIP-grade anti-GFP antibody (4 µg antibody per 10 µg chromatin; Invitrogen A11122).

ChIP-seq and data analysis. Wild-type, *Pax3*-deficient and Pax3-deficient (shPax9) iMEFs and wild-type and *Suv39h*–double-null ESCs were subjected to the ChIP protocol as described above. We used 10 ng DNA from each ChIP sample for library preparation, following the instructions of the ChIP-seq sample-prep kit from Illumina. The Illumina GAxII platform was used for sample sequencing. Bowtie was used to map reads to the mouse genome (mm9), and the resulting bam files were displayed in the Integrative Genomics Viewer genome browser[32].

We identified heterochromatic regions (n = 6,387) by differential peak calling using wild-type and *Suv39h*–double-null cells. The bed file HET_6387 is available upon request.

Consensus major satellite sequence was defined on the basis of previous publications[3,50] and confirmed by our own sequence data. Annotated major

satellite repeats outside of pericentric heterochromatin were retrieved from the University of California, Santa Cruz (Repeatmasker GSAT_MM). We used the density array method in SeqMiner[52] for the meta-analysis of H3K9me3 enrichment on annotated major satellite repeats. Respective fastq files are available at the GEO database.

TRAP analysis. To assess the overall sequence properties of heterochromatic regions, we adapted the TRAP physical binding model[34] for a selected set of transcription factors (Pax2, Pax3, Pax9 and Zeb1). To each heterochromatic region we added equidistant upstream and downstream flanking sequences. The combined region was divided into 40 equal bins. For each bin we calculated the logarithmic binding affinity and normalized it with respect to different sequence lengths, as detailed in refs. 33–35. For each bin, the resulting score, also called TRAP score, was averaged over all (6,387) heterochromatic regions. Similarly, for the analysis of promoter regions, we defined a flanking sequence of ±1 kb around all annotated TSSs (Ensembl) and divided their sequences into 40 equal bins. After aligning 28,028 promoter regions, the average TRAP score was calculated for each selected transcription factor. We selected as controls random regions from the genome that did not overlap with the identified heterochromatic regions (HET_6387) or annotated promoters. This provided the background distribution of affinities, which could also be inferred from sufficiently large flanking regions. Our approach relied on prior information about the binding preferences of transcription factors obtained from alignments of known binding sites available in the Genomatix MatBase.

51. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual* 3rd edn, Vol. 1, 7.51–7.62 (Cold Spring Harbor Laboratory Press, 2001).
52. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).

doi:10.1038/nsmb.2382

Corrigendum: A transcription factor–based mechanism for mouse heterochromatin formation

Aydan Bulut-Karslioglu, Valentina Perrera, Manuela Scaranaro, Inti Alberto de la Rosa-Velazquez, Suzanne van de Nobelen, Nicholas Shukeir, Johannes Popow, Borbala Gerle, Susanne Opravil, Michaela Pagani, Simone Meidhof, Thomas Brabletz, Thomas Manke, Monika Lachner & Thomas Jenuwein

***Nat. Struct. Mol. Biol.* 19, 1023–1030 (2012); published online 16 September 2012; corrected after print 27 December 2012**

In the version of this article initially published, Simone Meidhof's affiliations should have included Spemann Graduate School of Biology and Medicine (SGBM), Albert Ludwigs University Freiburg, Freiburg, Germany and the Faculty of Biology, Albert Ludwigs University Freiburg, Freiburg, Germany. Acknowledgement of the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School) was omitted. The errors have been corrected in the HTML and PDF versions of the article.