Principles of nucleation of H3K27 methylation during embryonic development

Simon J. van Heeringen, Robert C. Akkers, Ila van Kruijsbergen, M. Asif Arif, Lars L.P. Hanssen, Nilofar Sharifi, and Gert Jan C. Veenstra

Radboud University Nijmegen, Department of Molecular Developmental Biology, Faculty of Science, Nijmegen Centre for Molecular Life Sciences, Nijmegen 6500 HB, The Netherlands

During embryonic development, maintenance of cell identity and lineage commitment requires the Polycomb-group PRC2 complex, which catalyzes histone H3 lysine 27 trimethylation (H3K27me3). However, the developmental origins of this regulation are unknown. Here we show that H3K27me3 enrichment increases from blastula stages onward in embryos of the Western clawed frog (Xenopus tropicalis) within constrained domains strictly defined by sequence. Strikingly, although PRC2 also binds widely to active enhancers, H3K27me3 is only deposited at a small subset of these sites. Using a Support Vector Machine algorithm, these sequences can be predicted accurately on the basis of DNA sequence alone, with a sequence signature conserved between humans, frogs, and fish. These regions correspond to the subset of blastula-stage DNA methylation-free domains that are depleted for activating promoter motifs, and enriched for motifs of developmental factors. These results imply a genetic-default model in which a preexisting absence of DNA methylation is the major determinant of H3K27 methylation when not opposed by transcriptional activation. The sequence and motif signatures reveal the hierarchical and genetically inheritable features of epigenetic cross-talk that impose constraints on Polycomb regulation and guide H3K27 methylation during the exit of pluripotency.

[Supplemental material is available for this article.]
Based recruitment mechanisms include interactions with sequence-specific TFs such as REST (Dietrich et al. 2012; Arnold et al. 2013) and RUNX1 (Yu et al. 2012). Finally, several long noncoding RNAs (IncRNAs) have been identified that influence Polycomb-dependent H3K27 methylation (Rinn et al. 2007; Zhao et al. 2008; Tsai et al. 2010). These IncRNAs may serve as scaffolds by providing binding surfaces for assembly of specific histone modification complexes (Tsai et al. 2010).

To study the earliest dynamics of regulation by Polycomb in the context of vertebrate development, we studied the dynamics and activity of PRC2 in Xenopus tropicalis embryos. H3K27 is mostly newly methylated at promoter-distal positions, with a dramatic increase in enrichment from blastula to gastrula. H3K27 methylation is constrained to preexisting DNA methylation-free domains, which can be accurately predicted from the primary DNA sequence using a Support Vector Machine (SVM) algorithm. Specific regulatory sequence motifs can further distinguish between active and repressed methylation-free regions. These conserved, cell-type-independent signals dynamically guide Polycomb repression during the exit of pluripotency at the beginning of gastrulation.

**Results**

The repressive H3K27me3 modification is newly deposited from blastula stages onward

To characterize the dynamics of H3K27me3 nucleation and spreading in X. tropicalis, we generated epigenetic profiles using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) at four developmental stages. We assayed H3K4me3, H3K27me3, the enhancer mark H3K4me1, the Polr2a subunit of RNA Polymerase II (RNAPII), Ezh2, Jarid2, and a ChIP input track as control (see Methods) (Table 1; Supplemental Table 1).

We identified three clusters of temporal H3K27 methylation patterns (Fig. 1A,B; Supplemental Figs. 1, 2). Cluster 1 is promoter-associated, with H3K27me3 constrained to a ~10-kb region centered at the gene promoter (Fig. 1B, left panel). Cluster 2 shows large domains (up to ~100 kb) completely covering one or more genes (Fig. 1B, right panel). Cluster 3 consists of a set of small sequences with no clear association with genes (Supplemental Fig. 1).

Genes regulated by H3K27me3 are enriched for developmental transcription factors and functions in development, pattern formation, and morphogenesis (Supplemental Table 2). Promoters show very low levels of H3K27me3 at the blastula stage, with a strong increase at gastrulation (Fig. 1C), whereas H3K4me3 is already enriched at the blastula stage. Indeed, a genome-wide analysis of all 1862 promoters that gain H3K27me3 confirms the developmental hierarchy of permissive and repressive modifications (Fig. 1D; Akkers et al. 2009). However, we do observe some sites of early H3K27me3 nucleation. In many broad H3K27me3 domains, methylation of H3K27 is initiated at a local site. The majority of these sites do not correspond to gene promoters and show no H3K4me3 enrichment (Supplemental Figs. 3, 4). We did not observe a significant overlap with noncoding RNA genes (Supplemental Table 4), and TF motif analysis yielded no enriched motifs. The promoter-distal locations of early H3K27 methylation sites, however, raised the question as to how H3K27 methylation relates to both PRC2 and enhancer sites.

**Enhancers recruit PRC2**

We therefore assessed H3K27me3 dynamics at PRC2 subunits Ezh2 and Jarid2 binding sites. Surprisingly, only a small subset of the PRC2 binding sites gains H3K27me3 during subsequent development to the tailbud stage (Fig. 2A; Supplemental Fig. 5). This suggests that binding of the methyltransferase Ezh2 is either very unstable or not sufficient for H3K27 methylation at many loci. The lack of chromatin-associated H3K27me3 is not explained by depletion of histones, as most regions bound by both Ezh2 and Jarid2 binding have the enhancer mark H3K4me1 (Fig. 2A).

To assess the relationship with enhancers, H3K4me1 peaks with RNAPII (Kim et al. 2010), but not H3K4me3, were selected, and the enrichment for H3K27me3, Ezh2, and Jarid2 was determined (Fig. 2B). The results show that PRC2 is recruited to the majority of active enhancers at the pluripotent blastula stage. Consistently, these sites contain motifs of important developmental regulators (Supplemental Fig. 6). The POU, Forkhead, Homebox, HMG-box, and T-box motif families are all significantly enriched, indicating that enhancers from both pluripotency and lineage specification networks are bound by PRC2.

**Comparative analysis of sequence features of vertebrate H3K27me3 domains**

To determine the involvement of DNA sequence in establishing H3K27 methylation, we evaluated TF motifs and repeat content in broad H3K27me3 domains that are established by the gastrula stage and, moreover, compared nucleosome positioning signals, GC content, and CpG islands in Xenopus, zebrafish, and human in a comparative analysis.

First, from known TF motifs, only the REST motif was significantly enriched (Supplemental Fig. 7). Although REST is known to associate with Polycomb complexes (Dietrich et al. 2012; Arnold et al. 2013), this motif is present in only ~1% of H3K27me3 domains. No other TF binding motifs were significantly associated with H3K27me3 domains.

Second, identification of overrepresented repeats shows that large H3K27me3 domains (Cluster 2) are highly enriched for specific simple sequence repeat types (more than 50 times) such as...
TAGA/TCTA and CA/TG (Supplemental Material; Supplemental Table 3). This may reflect a tendency of increased DNA polymerase slippage in Polycomb-repressed chromatin.

Third, to test the involvement of nucleosome positioning signals, we used a recent model for prediction of nucleosome positions from DNA sequence (van der Heijden et al. 2012). We observed a significantly higher predicted positioning signal in H3K27me3 domains compared with random genomic sequence (Fig. 3A, upper panels) in a pattern that closely follows the H3K27me3 enrichment profile (Fig. 3A, lower panels).

Finally, we analyzed the GC% and CpG enrichment of H3K27me3 domains. In human, as reported (Ku et al. 2008; Mendenhall et al. 2010), the GC% distribution of H3K27me3 domains compared with random genomic sequence (Fig. 3A, upper panels) in a pattern that closely follows the H3K27me3 enrichment profile (Fig. 3A, lower panels).

Pan-vertebrate DNA sequence signals of H3K27me3 domains

To use an alternative approach to analyze the DNA sequence properties of H3K27 methylation, we applied an SVM algorithm (Lee et al. 2011). This is a supervised machine learning approach that, in this particular case, uses all sequences of length \( k \) (k-mers) to distinguish a specific set of sequences from genomic background. It is trained with a subset of the genomic data (positive and negative sequences) to produce a classifier that can be used to predict the status of new sequences. The rest of the data that was not used for training can be used to test the performance of the SVM. We first trained an SVM on Xenopus H3K27me3 domains. Even though we found no enriched TF motifs in these regions except for the REST motif, and even though the GC% is similar to genomic background, an SVM trained on part of the genome is able to accurately classify H3K27me3-marked regions in the...
remaining part (Fig. 4A; Supplemental Fig. 9). Next, we analyzed the performance of the SVM in a cross-species analysis (Fig. 4B). Shown is the prediction performance as a receiver operating curve (ROC) that plots the fraction of true positives versus the fraction of false positives; a high area under the curve (AUC (ROC) that plots the fraction of true positives versus the fraction of false positives; a high area under the curve (AUC) corresponds to high accuracy and sensitivity. An SVM trained in any of the three species can predict H3K27me3 domains to a certain degree in all species, uncovering pan-vertebrate sequence conservation in H3K27 methylation (Fig. 4B). The performance of the SVM is not based purely on GC% and CpG content as it also can distinguish natural H3K27me3 sequences from random regions with an identical dinucleotide background or from regions with a similar GC% sampled from the genome (Supplemental Fig. 10). By use of the SVM output from the three species, we identified motifs that are consistently enriched or depleted in H3K27me3 domains (Supplemental Material; Supplemental Figs. 11, 12). Remarkably, the unique sequence properties captured by the SVM trained on human H1 ES cells also identify cell-type-specific H3K27me3 domains that are not present in H1 ES cells (Supplemental Fig. 13), showing that the SVM detects a sequence-based, cell-type-independent propensity for H3K27 methylation.

TF motifs distinguish unmethylated regions that do or do not gain H3K27me3

The H3K27me3-marked sequences form a subset of the NMIs in all three species. More precisely, H3K27me3 decorates DNA methylation-free regions at many developmentally regulated genes, whereas unmethylated islands near other genes, including housekeeping genes, gain H3K4me3 but not H3K27me3 (Fig. 5B). We therefore trained an SVM on DNA methylation-free regions and find it performs very well (ROC AUC 0.953) (Fig. 6A). Only 19% of the NMIs in Xenopus gain H3K27me3. However, an SVM specifically trained to distinguish between NMIs with and without H3K27me3 is able to separate them with high performance (ROC AUC 0.853) (Fig. 6B). Interestingly, the k-mers with the highest and lowest weights of the SVM correspond to important regulatory motifs (Fig. 6C,D). NMIs without H3K27me3 are characterized by motifs of housekeeping TFs such as ETS, NFY, CREB, and YY1 (Fig. 6C). These correspond to the motifs that were previously identified in mouse CpG islands without H3K27me3 (Ku et al. 2008). However, we find that NMIs that gain H3K27me3 show enrichment of motifs for developmental regulators such as Sox and Homeobox TFs (Fig. 6D). Therefore, TF motifs distinguish unmethylated DNA elements that do or do not acquire H3K27me3.

Blastula-stage H3K27me3 peaks function as repressive elements in Xenopus and mouse

We wondered if regions carrying the SVM sequence signature could function as repressive elements. We selected sites with a positive SVM score that showed H3K27me3 enrichment in blastula-stage embryos. These sites (~1 kb) were cloned into a pGL3 reporter vector and used for injection in Xenopus embryos and for transfection of mouse ES cells (Fig. 7A). In both experimental systems, these sites significantly repress luciferase expression (P < 0.001) and show H3K27me3 enrichment (P < 0.001), but no difference in H3K4me3.
enrichment (Fig. 7B). This confirms that the pan-vertebrate conserved sequence signature we identified indeed directs H3K27 methylation and repression across species. None of the selected sites are CpG islands, and the mean GC% is 38%, lower than the overall GC% of both the Xenopus and the mouse genome (~40%).

H3K27 methylation is influenced by nucleosome density and by allosteric activation of Ezh2 (Margueron et al. 2009; Yuan et al. 2012). To assess the influence of sequence length, we tested the effect of multimerization on four H3K27me3 initiation sites in mouse ES cells. Figure 7C shows the H3K27me3 ChIP-qPCR recovery of the same 1-kb fragment multimerized one, two, or four times. The recovery increases almost linearly with the length of the inserted fragment, though the background signal of the vector backbone shows no such increase.

Discussion

Our interspecies analysis of Polycomb regulation uncovered principles of H3K27 methylation during embryonic development involving epigenetic cross-talk, sequence features, and a hierarchy of activation and repression. The developmental timing of H3K27 methylation from the pluripotent blastula stage onward in Xenopus embryos is consistent with a function in stable differentiation and lineage commitment (Bogdanović et al. 2012). The dynamics
support reestablishment of repressive chromatin rather than inheritance from sperm or oocytes. This has previously been suggested for zebrafish (Lindeman et al. 2011) and is supported by analysis of histone modifications by mass spectrometry (Schneider et al. 2011).

The PRC2 complex is present at the majority of active developmental enhancers, possibly recruited by binding of PRC2 to enhancer RNAs (eRNAs) that are expressed at enhancers (Kanhere et al. 2010; Kim et al. 2010). However, only a subset of these enhancers gain the H3K27me3 mark. This implies additional regulation, such as activation or inhibition of Ezh2 through posttranslational modifications (Cha et al. 2005; Lan et al. 2007; Wei et al. 2011). Also, PRC2 may be ineffective in H3K27 methylation at these sites because of H3K27 acetylation, a hallmark of active enhancers.

Our results indicate that Polycomb recruitment in vertebrates is not necessarily mediated through a PRE, defined by a combination of repressive binding motifs. We find that overall susceptibility to H3K27 methylation is linked to specific sequence properties, absence of DNA methylation, and binding sites for cell-type-specific transcriptional activators. The conserved sequence signature, however, marks the Polycomb-regulated subgenome in a cell-type-independent fashion and may reflect a propensity for H3K27 methylation. Part of the sequence component reflects the presence of constitutive DNA methylation-free domains. This is very significant as binding of PRC2 to nucleosomes is counteracted by methylated DNA (Bartke et al. 2010). PRC2 binding and H3K27 methylation are also inhibited by gene activation, especially by the presence of the H3K4me3 permissive mark (Schmitges et al. 2011).

In humans, the conserved sequence signature correlates with CpG islands, which previously have been implicated in Polycomb recruitment (Ku et al. 2008; Mendenhall et al. 2010; Lynch et al. 2012). In contrast, Polycomb-regulated sequences in Xenopus and zebrafish do not exhibit a strong CpG or GC% enrichment. However, they are devoid of DNA methylation and can be accurately identified using an SVM. In addition, Xenopus H3K27me3 nucleation sites show repressive capacity in mouse ES cells, even though these sequences do not correspond to CpG islands.

Interestingly, we find that constitutive and cell-type-specific TF motifs distinguish between different DNA methylation-free regions that do or do not acquire H3K27 methylation. Supporting this model, a single point mutation in a binding site of the activating SP1 TF can result in repression by Polycomb (Caputo et al. 2013). The data suggest a genetic default mechanism for this type of facultative heterochromatin, in which default H3K27 methylation in domains of unmethylated DNA can be counteracted by dynamic and cell-type-dependent gene activation. The default nature of this repression is significant in light of the hierarchy between activation and repression, as H3K4me3 is acquired before localized H3K27 methylation represses multilineage gene expression (Akkers et al. 2009).

Our results indicate that it is not GC richness by itself that causes H3K27 methylation but, rather, associated sequence features and absence of DNA methylation in combination with an absence of activating signals. After initial recruitment of PRC2, the H3K27me3 modification spreads on unmethylated susceptible sequences due to the activating effect of nucleosome density (Yuan et al. 2012) and an allosteric positive feedback loop (Margeron et al. 2009). Indeed, the predicted nucleosome occupancy is significantly higher in H3K27me3 domains compared with the genomic average. During subsequent development and differentiation, Polycomb action may be complemented by gene-specific targeting mechanisms involving REST or other repressors.

Binding of PRC2 is reduced by DNA methylation, and H3K27me3 is also inhibited by other histone modifications (Bartke et al. 2010; Pasini et al. 2010; Schmitges et al. 2011), overriding the genetic-default state in different cell types. The underlying genomic sequence signature may function as a genetically inheritable constraint on Polycomb target selection, and the interplay with other epigenetic modifications may elegantly explain the stability of cell lineage commitment. An H3K27 methylation-default state within DNA methylation-free regions that are not transcriptionally activated may establish two major features of epigenetic regulation of
development: (1) de novo reconstitution of the epigenetic landscape during early development and (2) locking cell lineage commitment by raising the activation threshold of multilineage gene expression after initial lineage specification at the onset of gastrulation.

Methods

Published data
All previously published, public data sets used in this study are summarized in Supplemental Table 6.

Animal procedures
*X. tropicalis* and *Xenopus laevis* embryos were obtained by in vitro fertilization, dejellied in 3% cysteine, and collected at the following Nieuwkoop-Faber stages: nine (blastula), 12 (gastrula), 16 (neurula), and 30 (tailbud).

Chromatin immunoprecipitation and antibodies
Chromatin for chromatin immunoprecipitation was prepared as previously described (Jallow et al. 2004; Akkers et al. 2012). The following antibodies were used: anti-H3K4me3 (Abcam ab8580), anti-H3K27me3 (Upstate/Millipore 07-449), Jarid2 (Abcam ab48137), EZH2 (Active Motif 39103), and POLR2A (Diagenode AC-055-100).

For all ChIP-seq samples, three independent biological replicates of different chromatin isolations were pooled.

Sequencing and alignment
Sequencing samples were prepared according to the manufacturer’s protocol (Illumina). Shortly, adapter sequences were ligated, the library was size-selected (200–300 bp), and amplified. The sequencing (36 cycles) was carried out on a Genome Analyzer IIx (Illumina) except for the Input track, which was sequenced on a HiSeq 2000 (Illumina). Reads passing the Illumina chastity filter were aligned to the *X. tropicalis* genome (version JGI 7.1) using BWA (Li and Durbin 2009). For processing and manipulation of SAM/BAM files, SAMtools was used (Li et al. 2009). All duplicate reads and reads mapping to repeat regions were removed.

Detection of enriched regions
We used PeakRanger version 1.15 (Feng et al. 2011) with the following parameters: FDR $1 \times 10^{-6}$ (H3K27me3, H3K4me3) or $1 \times 10^{-4}$ (Ezh2, Jarid2, RNAPII, H3K4me1), ext_length 300. All peaks were called relative to an input control track. Only scaffolds $>2$ Mb were included in the analysis.

Generation of profiles and heatmaps
All heatmaps and bandplot profiles were generated using fluff (http://simonvh.github.com/fluff/). This Python package uses pysam (http://code.google.com/p/pysam/), pybedtools (Dale et al. 2011), and HTSeq (http://www-huber.embl.de/users/anders/HTSeq/) for BAM and BED manipulations and Pycluster (de Hoon et al. 2004).
for k-means and hierarchical clustering. For all heatmap clustering, the Euclidian distance metric was used. For hierarchical clustering, we used the pairwise complete-linkage function.

Analysis of sequence features

As input for all sequence feature analysis, we used H3K27me3 domains of at least 1 kb from gastrula embryos (stage 12, X. tropicalis; shield, zebrafish) or ES cells (H1-hESC, human). To identify enriched motifs, we used “maxenr” from the GimmeMotifs package (van Heeringen and Veenstra 2011) to scan for all vertebrate motifs from JASPAR. All motifs present in at least 1% of the sequences with an enrichment of at least three times compared with random background sequences were reported. For the repeat analysis, see Supplemental Material. We calculated the average maximum predicted nucleosome occupancy using an empirical statistical mechanics model (Supplemental Material; van der Heijden et al. 2012). An equal number of regions was randomly selected from the same genome and analyzed in the same manner.

Bio-CAP analysis

Peaks were called using MACS 2 (Zhang et al. 2008) relative to the input. The SVM was trained on 1-kb peaks centered around the Bio-CAP peak summit. For motif analysis, the 30 8-mers with either highest or lowest weight were clustered and matched to known motifs using GimmeMotifs (van Heeringen and Veenstra 2011). After clustering, enrichment was determined, and only motifs with an enrichment of at least 1.5 times are shown in Figure 6.

SVM analysis of H3K27me3 domains

For SVM analysis, we employed the method previously described for enhancer prediction (Lee et al. 2011). We used 1-kb H3K27me3 regions as positive sequences and randomly selected genomic sequences with no H3K27me3 ChIP-seq enrichment as negative sequences. For training we used all chromosomes except scaffold 9 (Xenopus) and chr2 (humans, zebrafish), which were used for validation and assessment of (cross-species) performance. In addition,
we performed a 10-fold cross-validation procedure using random partitioning of training (90%) and evaluation (10%) data sets to obtain the average performance (mean ROC AUC) of the SVM on X. tropicalis H3K27me3 regions.

ES cell culture
Male mouse embryonic stem cells (E14) were cultured according to standard SIGTR protocols (http://www.sanger.ac.uk/resources/mouse/sigtr/). Briefly, cells were grown on 0.1% gelatin-coated plates and maintained in high-glucose Dulbecco’s modified Eagle’s medium (Gibco, Invitrogen) supplemented with 15% FBS (PAA Laboratories), 1% penicillin/streptomycin (PS; Gibco), 0.0035% β-mercaptoethanol (Sigma-Aldrich), and 0.2% LIF (Chemicon).

Transient transfection and luciferase assays
Selected regions of ~1 kb in size were amplified from genomic DNA (Supplemental Table 5) and cloned into the pGL3-promoter vector (Invitrogen). Plasmid DNA was purified using the Plasmid Maxi Kit (Qiagen) or the Wizard Plus SV miniprep DNA purification vector (Invitrogen). Plasmid DNA was purified using the Plasmid DNA (Supplemental Table 5) and cloned into the pGL3-promoter backbone control.

Quantitative (RT-) PCR
PCR reactions were performed on a MyiQ single-color real-time PCR detection system (Bio-Rad) using iQSYBR green supermix (Bio-Rad). Primer sequences are available in the Supplemental Material.

Data access
The sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE41161. Visualization tracks are available at our website (http://veenstra.ncmls.nl).

Acknowledgments
We thank H. Stunnenberg for sequencing facilities, E. Gazdag for preparing input DNA, and J. van Noort for access to the nucleosome positioning algorithm. This work was funded by the US National Institutes of Health (grants 1R01HD054356 and 1R01HD069344 to G.J.C.V.) and the Netherlands Organization for Scientific Research (NWO-ALW grant 863.12.002 to S.J.v.H.).

References
van Heeringen et al.


Received April 26, 2013; accepted in revised form November 27, 2013.
Principles of nucleation of H3K27 methylation during embryonic development

Simon J. van Heeringen, Robert C. Akkers, Ila van Krujsbergen, et al.

*Genome Res.* 2014 24: 401-410 originally published online December 11, 2013  
Access the most recent version at doi:10.1101/gr.159608.113

---

**Supplemental Material**  
http://genome.cshlp.org/content/suppl/2014/01/07/gr.159608.113.DC1

**References**  
This article cites 56 articles, 13 of which can be accessed free at:  
http://genome.cshlp.org/content/24/3/401.full.html#ref-list-1

**Open Access**  
Freely available online through the *Genome Research* Open Access option.

**Creative Commons License**  
This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at http://creativecommons.org/licenses/by-nc/3.0/.

**Email Alerting Service**  
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

---

To subscribe to *Genome Research* go to:  
http://genome.cshlp.org/subscriptions

© 2014 van Heeringen et al.; Published by Cold Spring Harbor Laboratory Press