Supplemental Material

Principles of nucleation of H3K27 methylation during embryonic development

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

Supplemental Methods

Motif analysis of PRC2 peaks

GimmeMotifs (van Heeringen and Veenstra, 2011) was run on the overlap of EZH2 and Jarid2 peaks. Default parameters were used, background was set to 'random' and analysis to 'large'. All the resulting motifs were combined with the vertebrate motifs from JASPAR 2010 (Portales-Casamar et al., 2009) and the motifs determined by protein binding microarray from UniPROBE (Newburger and Bulyk, 2009). Maximum enrichment of all motifs was determined and all motifs with an enrichment of at least 3 fold compared to random genomic sequences and occurring in at least 1% of the peaks were clustered (Heeringen et al., 2011). The resulting motifs are shown with their enrichment in Supplemental Figure 6.

Repeat analysis

We downloaded the repeat track from the UCSC Genome Browser for *Xenopus tropicalis* JGI 4.1 (UCSC xenTro2), as the repeat annotation for JGI 7.1 is not available. All sequences for the three different types of H3K27me3 domains, were mapped to xenTro2 using blat. Overlap with the repeats was also calculated for random genomic regions. All repeats occurring with a frequency of at least 1 repeat per 100kb and an over-representation of at least 2 times compared to random genomic background are summarized in Supplemental Table 3.

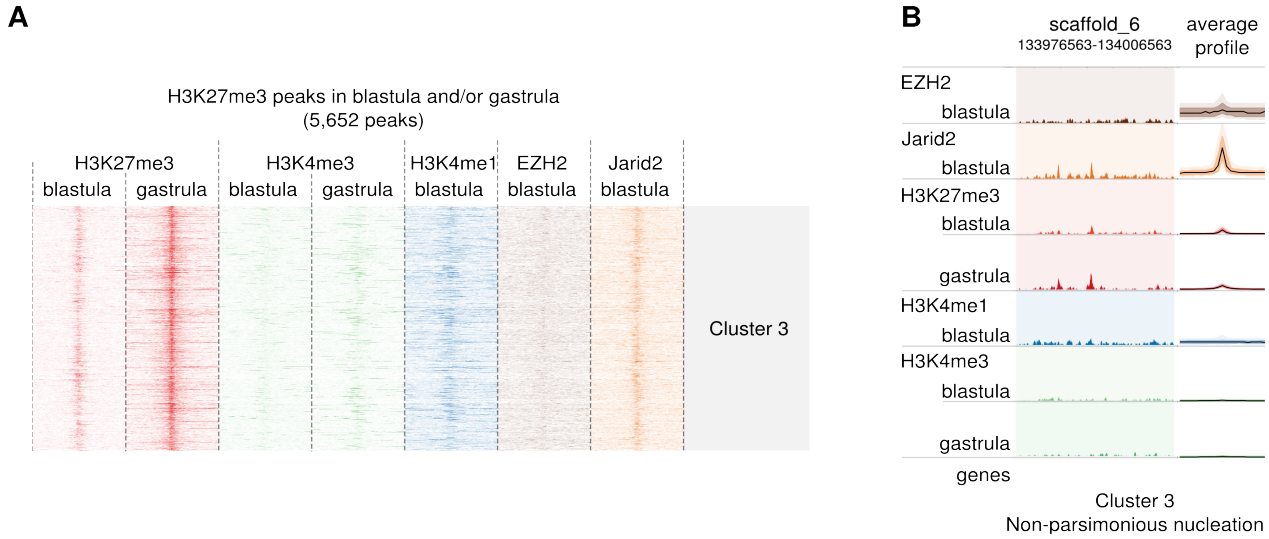
Nucleosome positioning

We calculated the nucleosome occupancy using an empirical statistical model (van der Heijden et al., 2012). This algorithm uses the periodic occurrence of the dinucleotides TA, TT, AA and GC and can identify nucleosome positioning sequences with several base pair accuracy. We scanned a 20kb window around the center of all H3K27me3 domains using a sliding window of 1kb with a step size of 100bp. For each window of 1kb the nucleosome positioning signal was calculated using default parameters and the signal was averaged over 2 consecutive windows per H3K27me3 domain. In figure 3A the mean of all H3K27me3 domains is plotted.

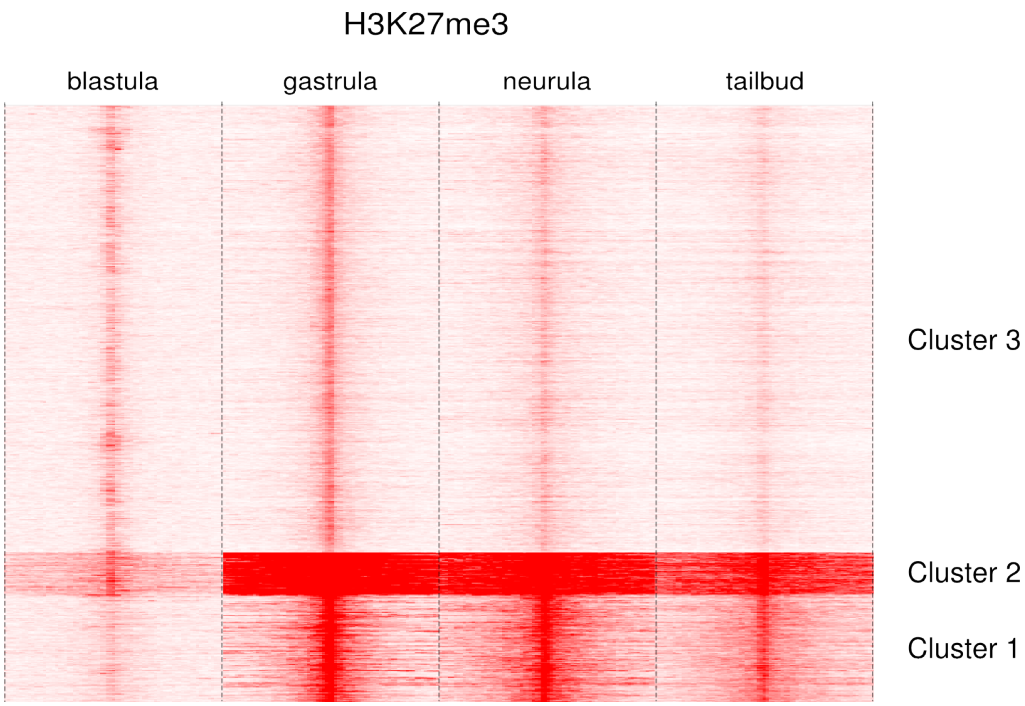
Conserved motifs from the SVMs of three species

We used the k-mer SVM results of all three species (frog, zebrafish, human) to identify conserved motifs that are enriched or depleted in H3K27me3 domains. The k-mer SVM algorithm produces an output file containing the weight of each k-mer. We converted the k-mer weights of all the 8-mers in each species to a ranked list, and combined the ranks of the three species to produce a mean rank per k-mer. Thus, k-mers that are generally positive in different organisms will have a higher rank than k-mers which are specific for one organism. We took the 5% highest and lowest ranked k-mers and selected those that had a different distribution around the H3K27me3 domain border (+/- 3kb) in *Xenopus*. K-mers that exhibited a significantly different distribution (Wilcoxon rank sum test, $p < 0.01$) were clustered using the motif clustering algorithm of GimmeMotifs (van Heeringen and Veenstra, 2011; Heeringen et al., 2011). After clustering we calculated enrichment and significance of the clustered motifs. All significantly enriched (or depleted; $p < 0.05$) motifs are shown in Supplemental Figs. 11 and 12.

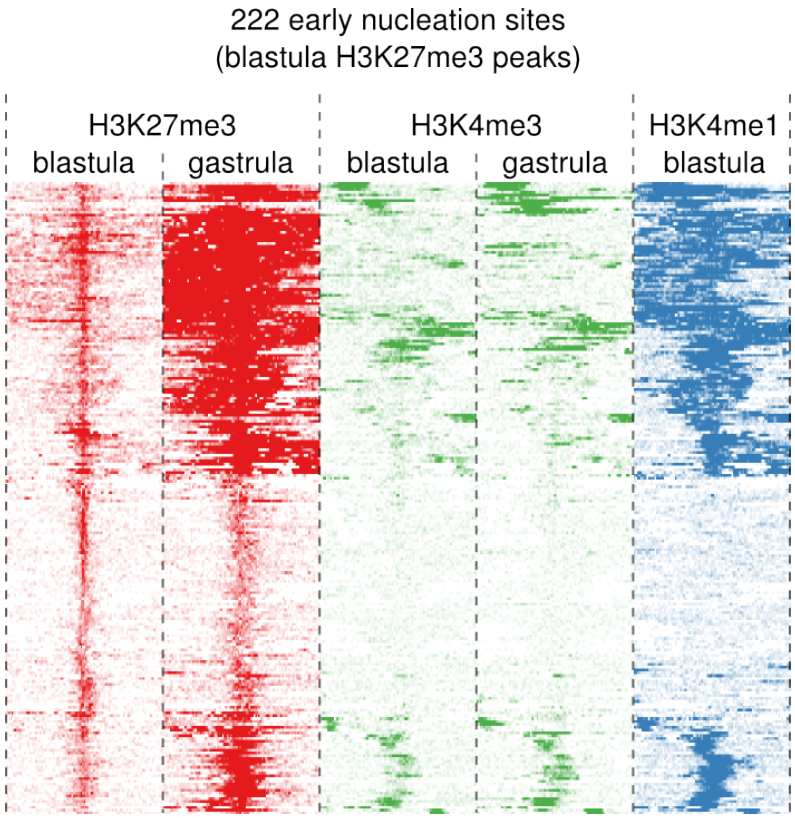
Supplemental Figures 1-13:



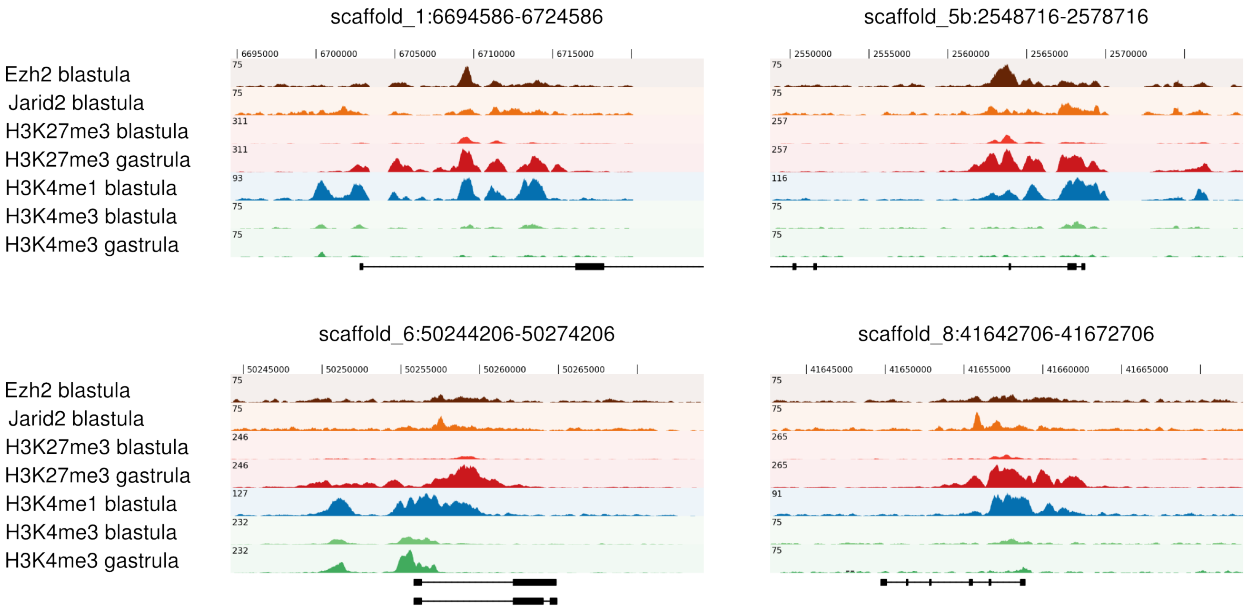
Supplemental Fig. 1. Cluster 3 of H3K27me3 sites. These sites show Jarid2 but no Ezh2 occupancy and are enriched for many different types of repeats (Supplemental Table 3-III). They don't gain H3K27me3 during subsequent development (Supplemental Figure 2), generally do not co-localize with genes and their functional relevance is unclear.
(A) Heatmap of a K-means clustering analysis (k=3, Euclidian distance) of H3K27me3 (blastula and gastrula), H3K4me3 (blastula and gastrula), H3K4me1, Ezh2 and Jarid2 (blastula) in 10 kb regions around H3K27me3 peak summits. The intensity of the color represents the number of reads per 100 bp window. Cluster 3 is visualized, the first two clusters are shown in Figure 1A.
(B) Average profile and representative example of the cluster shown in A. The average profiles show the mean enrichment (black line) and the 50th and 90th percentiles in a dark and lighter color respectively.



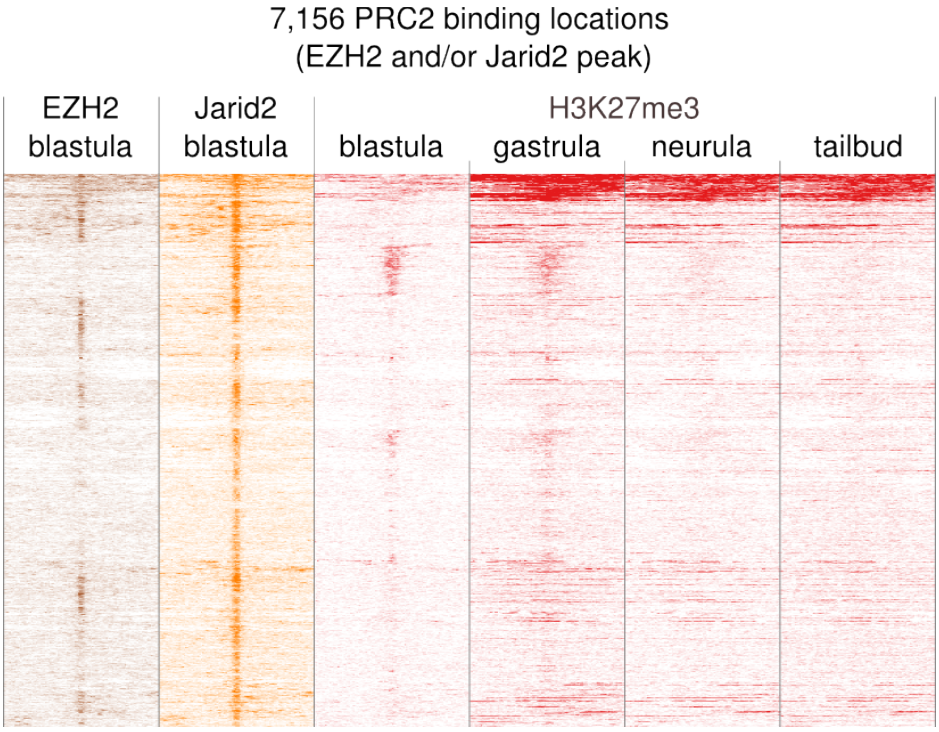
Supplemental Fig. 2. Enrichment of H3K27me3 around H3K27me3 peak summits in blastula, gastrula, neurula and tailbud stages. Enrichment is visualized in 100bp bins, 10kb around the peak summit. The clusters are identical to those shown in Figure 1A and Supplemental Figure 1. In contrast to cluster 1 (promoter-associated peaks) and cluster 2 (broad domains) the regions in cluster 3 do not gain H3K27me3 even in the tailbud stage.



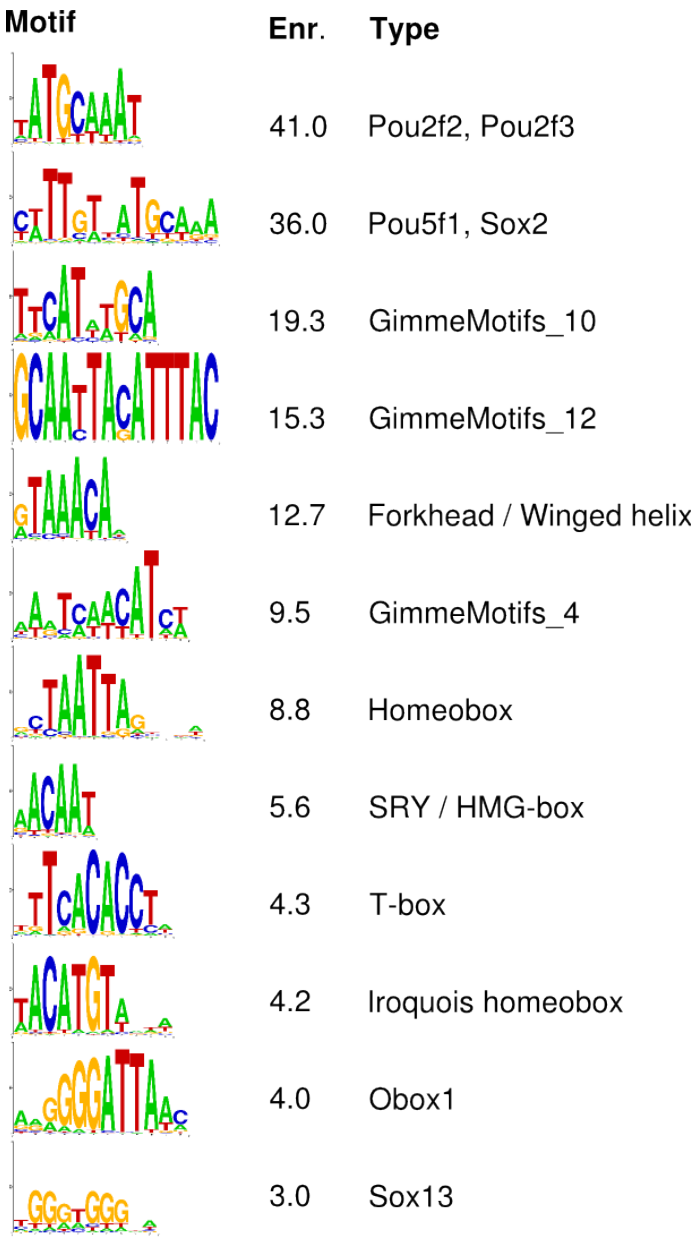
Supplemental Fig. 3. Heatmap of an hierarchical clustering of H3K27me3 in blastula and gastrula (red), H3K4me3 in blastula and gastrula (green) and H3K4me1 in blastula (blue) around H3K27me3 early nucleation sites (+/-5kb). Early nucleation sites were defined as blastula H3K27me3 peaks that were present as a broad domain in the gastrula H3K27me3 data. The intensity of the color represents the number of reads per 100 bp window. In most of the early nucleation sites there is no evidence of H3K27me3, indicating that these sites do not generally occur in promoters.



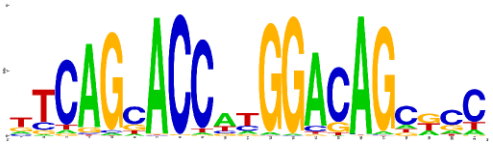
Supplemental Fig. 4. Examples of early nucleation that are not located in promoters. Shown is a 30kb region around the early nucleation sites. The tracks show, from top to bottom: Ezh2 (brown), Jarid2 (orange), H3K27me3 blastula (red), H3K27me3 gastrula (red), H3K4me1 (blue), H3K4me3 blastula (green), H3K4me3 gastrula (green) and gene structure. The Y-axis shows the number of overlapping reads.



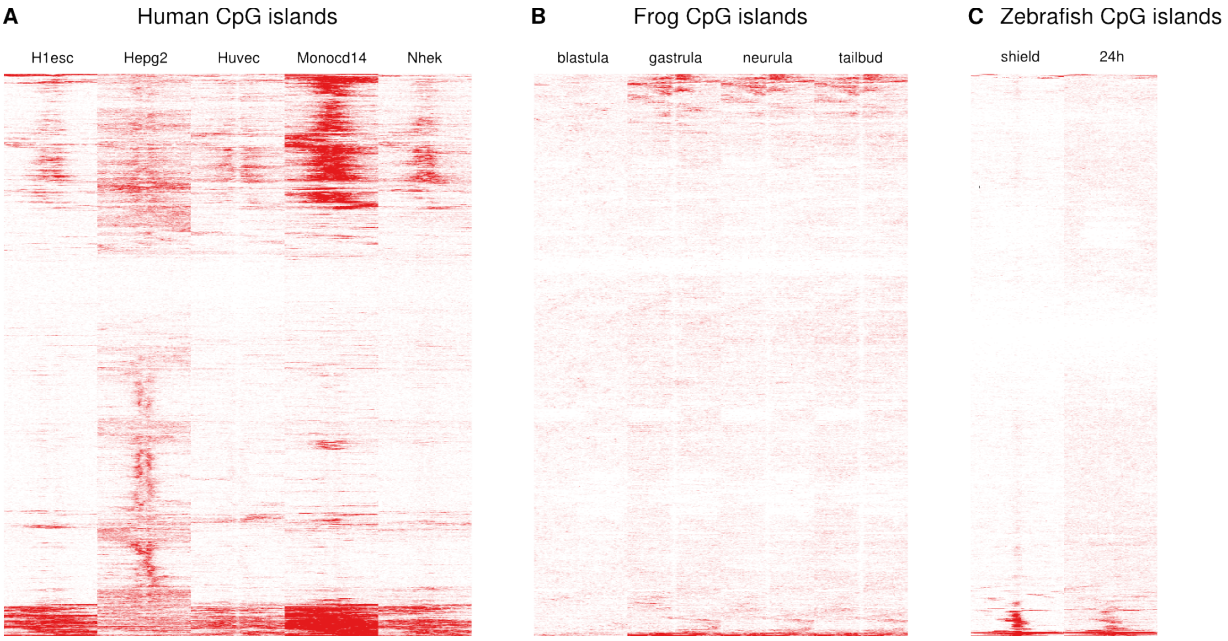
Supplemental Fig. 5. The majority of PRC2 binding sites does not gain H3K27me3. Clustering of EZH2, Jarid2, H3K27me3 (stage 9 blastula, stage 12 gastrula, stage 16 neurula, stage 30 tailbud) around peaks of EZH2 and/or Jarid2. Hierarchical clustering using Euclidian distance of 10 kb regions in 100bp bins.



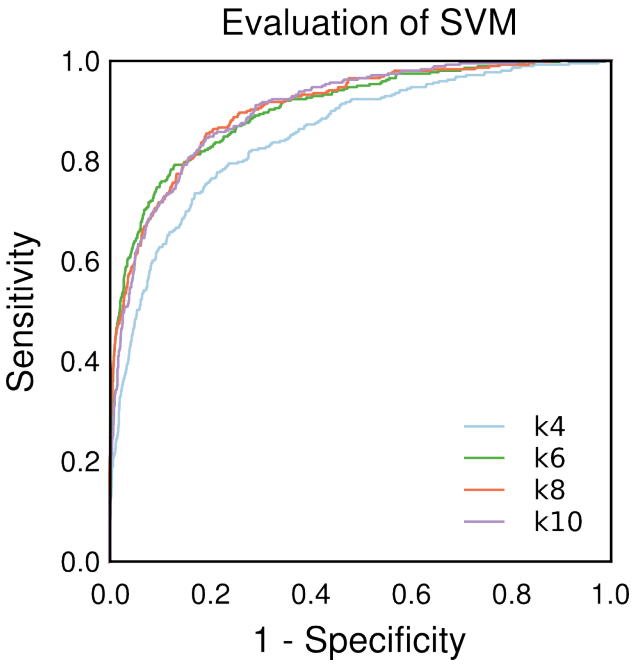
Supplemental Fig. 6. Motifs enriched in PRC2 binding sites. Combined motifs of GimmeMotifs (van Heeringen and Veenstra, 2011), JASPAR (Portales-Casamar et al., 2009) and Uniprobe (Newburger and Bulyk, 2009) are shown with the enrichment relative to random genomic sequences. Only motifs with an enrichment higher than 3 and a minimum occurrence in 1% of the PRC2 peaks are shown.

Motif	Enr.	Type
	22.2	REST

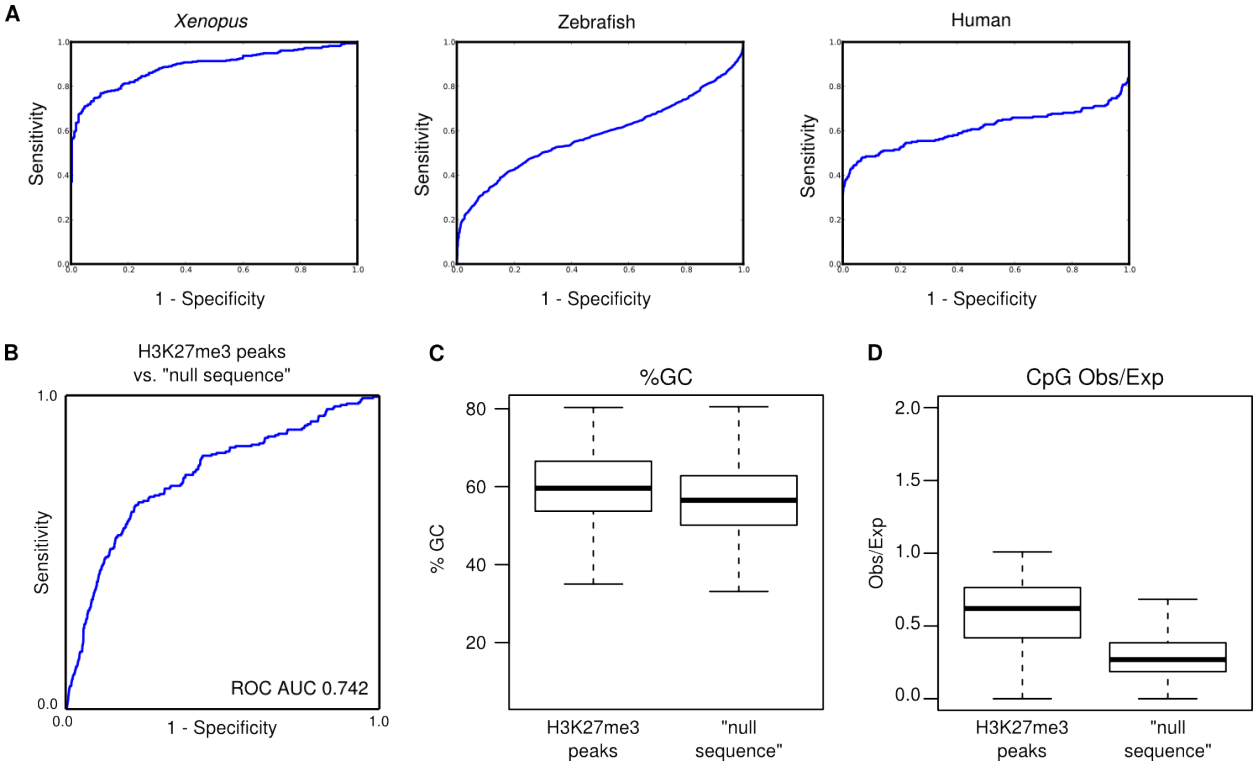
Supplemental Fig. 7. The REST motif is enriched in *Xenopus* gastrula (stage 12) H3K27me3 domains.



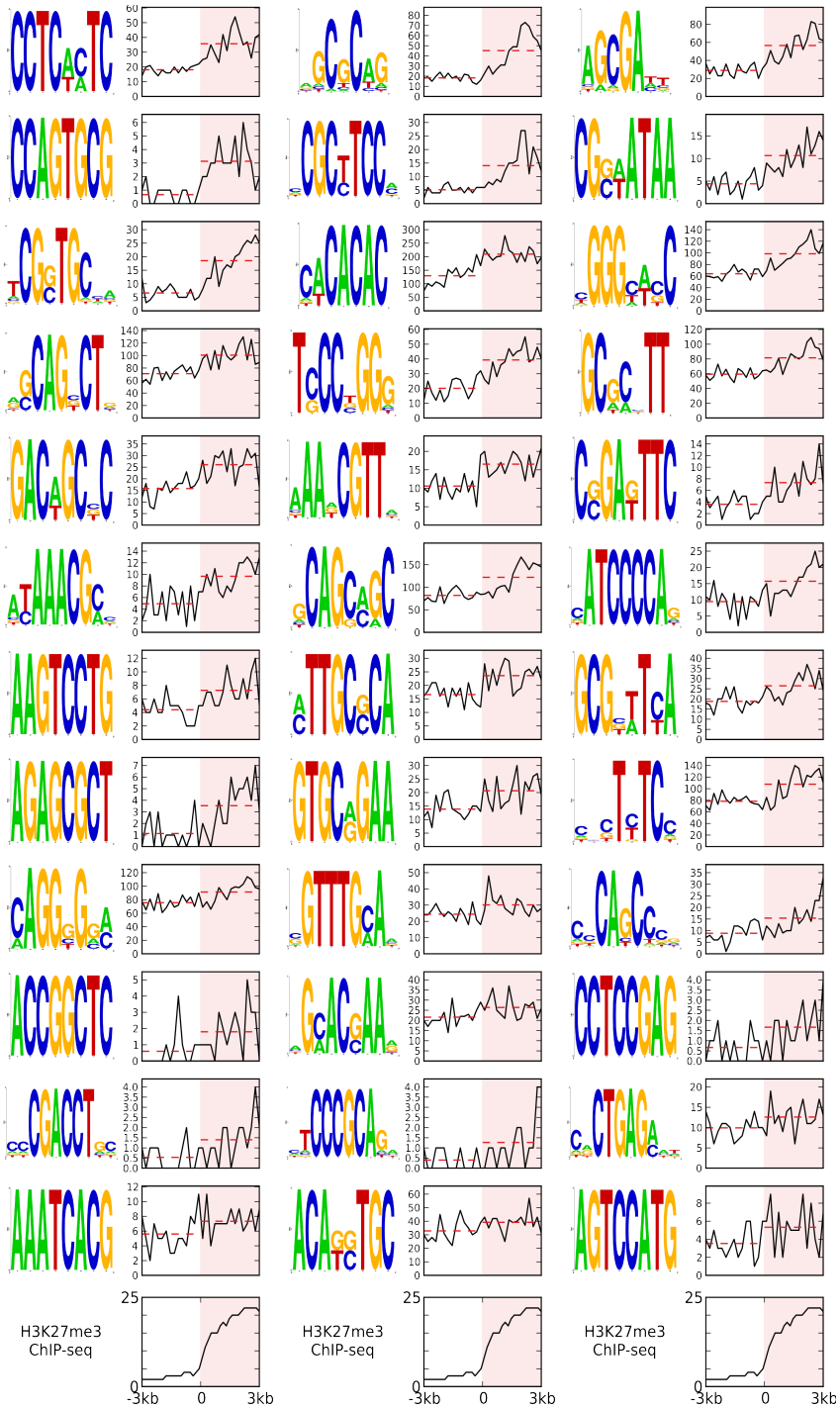
Supplemental Fig. 8. H3K27me3 enrichment around CpG islands in **(A)** 5 different human cell lines from ENCODE (H1-hESC, HepG2, HUVEC, MONOCD14, NHEK), **(B)** four developmental stages (blastula stage 9, gastrula stage 12, neurula stage 16 and tailbud stage 30) in *Xenopus tropicalis* and **(C)** two developmental stagses (shield and 24h) in zebrafish. For human, CpG islands were downloaded from the UCSC Genome Browser. *Xenopus* and zebrafish CpG islands were predicted using the method of Takai and Jones (Takai and Jones, 2002). Shown is the H3K27me3 enrichment in 100 bp bins, 5kb around 5000 randomly selected CpG islands.



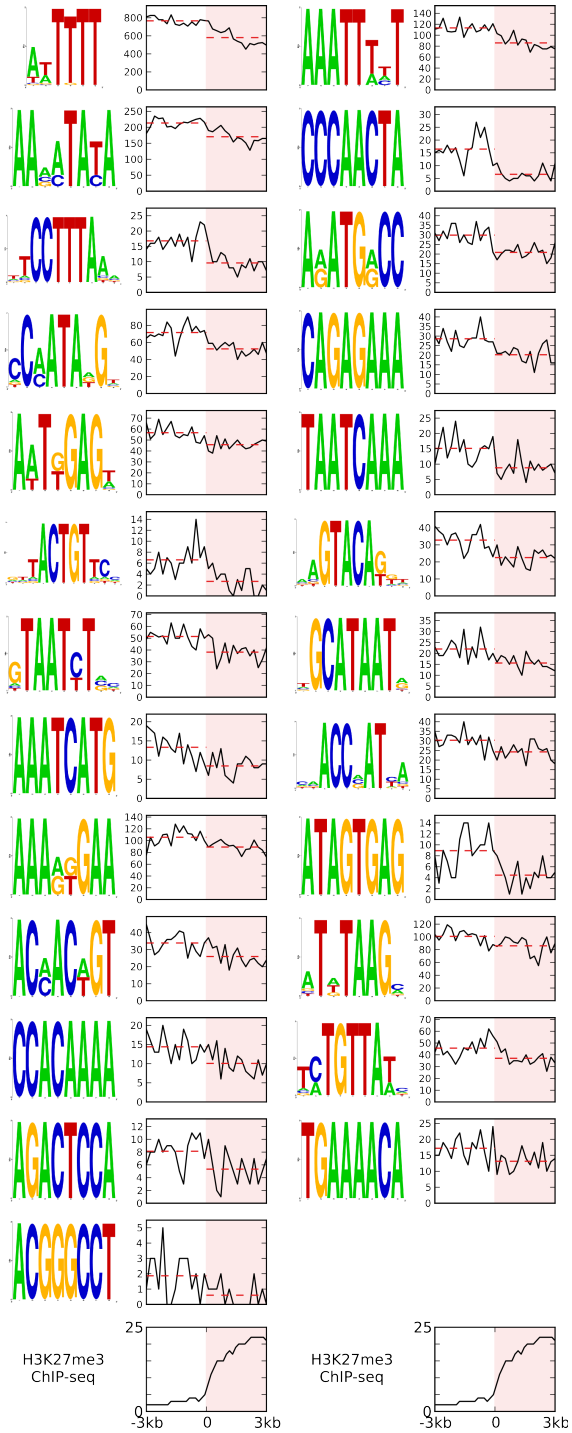
Supplemental Fig. 9. Performance of the k-mer SVM in classifying H3K27me3-positive regions. Different k-mer sizes for SVM training are shown, k-mer 8 is used in all subsequent analysis. Prediction performance as shown as a Receiver Operating Curve (ROC) that plots the fraction of true positives versus the fraction of false positives; a high area under the curve (AUC ~1) corresponds to high accuracy and sensitivity.



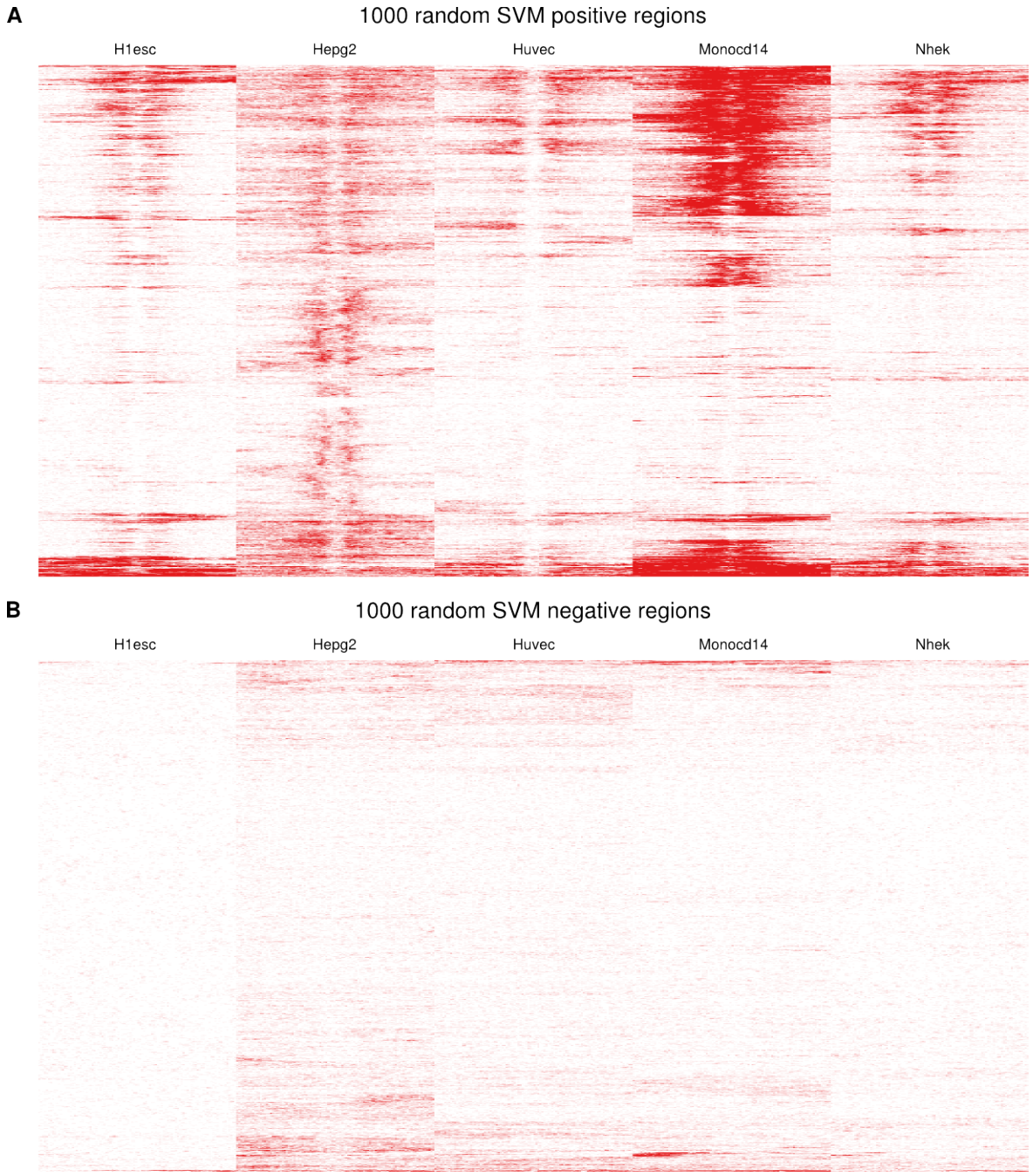
Supplemental Fig. 10. Performance of the k-mer SVM relative to background sequences. **(A)** Performance of the k-mer SVM (k=8) in each species relative to a randomly generated background with the same dinucleotide frequencies. H3K27me3 domains were used as a positive set; a negative set was generated using a 1st order Markov Model resulting in a random sequence set with similar dinucleotide distributions. Prediction performance as shown as a Receiver Operating Curve (ROC) that plots the fraction of true positives versus the fraction of false positives. The SVM can distinguish H3K27me3 domains from random sequence, indicating that dinucleotide composition is not the sole feature that sets apart H3K27me3 regions, even in human where the dinucleotide composition in H3K27me3-enriched regions is very different from that of most of the genome. An equal number of sequences was used for the positive and negative sets *(Xenopus* 338, zebrafish 1,696, human 229) ***(B)*** Performance of the human 8-mer SVM relative to background sequences selected from the genome. Background sequences were selected using the random sampling method implemented as the "Generate Null Sequence" option at http://kmersvm.beerlab.org/. The set of H3K27me3 sequences (229 sequences not used in SVM training) was used as input BED file to generate a set of sequences similar in GC content and repeat fraction. Out of 4,580 regions1,669 (36%) were positive for H3K27me3 in at least one the cell lines shown in Supplemental Fig. 8 and were removed. Prediction performance is shown as a Receiver Operating Curve (ROC) that plots the fraction of true positives versus the fraction of false positives. However, by removing the H3K27me3-positive sequences this set is no longer completely matched in %GC or CpG content, see C) and D) **(C)** %GC of the two sequence sets used in B) (H3K27me3 peaks and generated null sequences). **(D)** Observed number of CpGs divided by expected number of CpGs (Obs/Exp; Gardiner-Garden and Frommer, 1987) of the two sequence sets used in B).



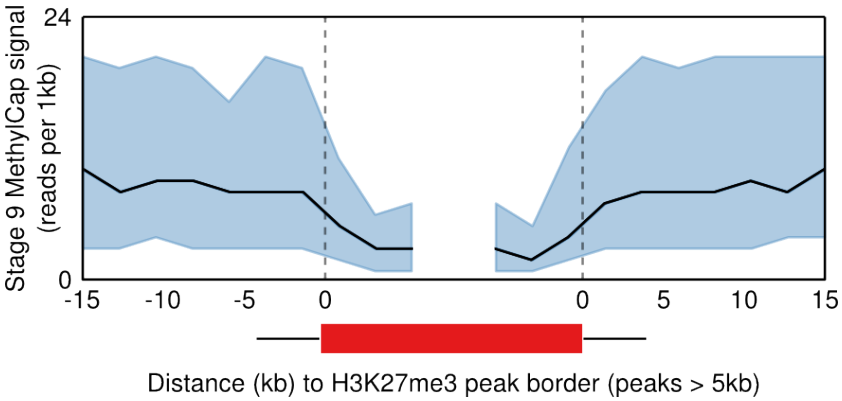
Supplemental Fig. 11. Consistently enriched motifs in H3K27me3 domains. Motifs were obtained from the SVM output of three species (see "Conserved motifs from the SVMs of three species"). Shown is the occurrence of the motifs from -3kb to +3kb around the H3K27me3 peak border in 200 bp bins. For comparison, the H3K27me3 ChIP-seq signal is shown in the bottom graph.



Supplemental Fig. 12. Consistently depleted motifs in H3K27me3 domains. Motifs were obtained from the SVM output of three species (see "Conserved motifs from the SVMs of three species"). Shown is the occurrence of the motifs from -3kb to +3kb around the H3K27me3 peak border in 200 bp bins. For comparison, the H3K27me3 ChIP-seq signal is shown in the bottom graph.



Supplemental Fig. 13. Regions with a positive SVM score in H1-hESC ES cells are positive for H3K27me3 in other human cell lines. 1000 regions with a SVM score >= 2 **(A)** and 1000 regions with a SVM score <= -1 **(B)** were randomly selected. Shown is the H3K27me3 enrichment in 100 bp bins 5kb around the selected region.



Supplemental Fig. 14. Gastrula H3K27me3 peaks are depleted for DNA methylation in blastula-stage embryos. Shown is the median number of blastula MethylCap reads per 1kb (black line) around gastrula H3K27me3 domains larger than 5kb. The blue area contains 50% of the data.

Supplemental Tables 1-6:

Supplemental Table 1. Datasets and statistics

Type	Stage	GEO accession	# mapped reads (million)	# of non-redundant, uniquely mapped reads (million)
H3K4me3 ChIP-seq	9	GSM1009589	13.80	8.57
H3K4me3 ChIP-seq	12	GSM1009590	18.04	10.89
H3K4me3 ChIP-seq	16	GSM1009591	14.01	9.64
H3K4me3 ChIP-seq	30	GSM1009592	15.14	9.94
H3K27me3 ChIP-seq	9	GSM1009593	11.56	5.71
H3K27me3 ChIP-seq	12	GSM1009594	19.04	10.85
H3K27me3 ChIP-seq	16	GSM1009595	14.05	8.85
H3K27me3 ChIP-seq	30	GSM1009596	15.95	11.09
RNAPII ChIP-seq	9	GSM1009597	25.45	13.90
RNAPII ChIP-seq	12	GSM1009598	27.56	19.30
RNAPII ChIP-seq	16	GSM1009599	26.16	19.24
RNAPII ChIP-seq	30	GSM1009600	23.08	16.96
EZH2 ChIP-seq	9	GSM1009601	27.17	12.54
Jarid2 ChIP-seq	9	GSM1009602	27.27	15.57
H3K4me1 ChIP-seq	9	GSM1009603	30.80	19.98
Input	30	GSM1009604	72.45	22.73

Supplemental Table 2 (Heeringen_Supplemental_Table2.xls) Annotation of genes marked by H3K27me3 in *Xenopus tropicalis*. All genes that overlapped with H3K27me3 were submitted to Functional Annotation Clustering at DAVID (http://david.abcc.ncifcrf.gov/; Huang et al., 2009).

Supplemental Table 3 Repeat-analysis of H3K27me3 domains
Shown is the repeat name (according to UCSC Genome Browser annotation), repeat class, repeat type, the number of occurrences in H3K27me3 sites and the overrepresentation as calculated relative to random genomic sequences.

Supplemental Table 3-I. Repeat-analysis of H3K27me3 sites, cluster 1

Repeat	Class	Type	# Occurrences	Over-representation
XLGST3	SINE	tRNA-V-Core-L2	45	9.0
(TA)n	Simple_repeat	Simple_repeat	91	5.3
AT_rich	Low_complexity	Low_complexity	288	4.9
Harbinger-N1_XT	DNA	PIF-Harbinger	37	4.4
POR-1_Xt	DNA	hAT-Charlie	28	4.2
POR	DNA	hAT-Charlie	113	4.1
XBR_Xt	DNA	Kolobok-T2	94	4.0
MuDR1_Xt	DNA	MULE-MuDR	36	3.9
XBR_XL	DNA	Kolobok-T2	67	3.5
TE_ORF_340	Unknown	Unknown	45	3.3
PIR_XL	DNA	Kolobok-T2	33	3.2
DNA1_Xt	DNA	DNA	41	3.0
XLLINE	LINE	CR1	33	2.7
CR1_1b_Xt	LINE	CR1	44	2.3

Supplemental Table 3-II. Repeat-analysis of H3K27me3 sites, cluster 2

Repeat	Class	Type	# Occurrences	Over-representation
(TCTA)n	Simple_repeat	Simple_repeat	194	58.1
(TAGA)n	Simple_repeat	Simple_repeat	155	57.0
(CA)n	Simple_repeat	Simple_repeat	92	56.4
(TG)n	Simple_repeat	Simple_repeat	111	49.3
(TAA)n	Simple_repeat	Simple_repeat	38	20.4
XLGST3	SINE	tRNA-V-Core-L2	109	15.8
(TA)n	Simple_repeat	Simple_repeat	345	14.6
AT_rich	Low_complexity	Low_complexity	587	7.3
DNA4Sat_X	Satellite	Satellite	120	6.5
JH12_XL	DNA	TcMar-Tigger	597	5.4
POR-1_Xt	DNA	hAT-Charlie	41	4.5
TE_ORF_340	Unknown	Unknown	78	4.2
DNA1_Xt	DNA	DNA	78	4.1
POR	DNA	hAT-Charlie	149	3.9
DNA2_Xt	DNA	PIF-Harbinger	186	3.8
REM1_XL	Satellite	Satellite	38	2.9
XBR_Xt	DNA	Kolobok-T2	88	2.8
DNA4_Xt	DNA	TcMar-Tigger	231	2.6
XBR_XL	DNA	Kolobok-T2	65	2.5

Supplemental Table 3-III. Repeat-analysis of H3K27me3 sites, cluster 3

Repeat	Class	Type	# Occurrences	Over-representation
Chap4sat_Xt	Satellite	Satellite	187	24.5
Chap4a_Xt	DNA	hAT-Charlie	595	19.6
Chap4b_Xt	DNA	hAT-Charlie	489	17.3
(CCCTG)n	Simple_repeat	Simple_repeat	154	16.7
(CAGGG)n	Simple_repeat	Simple_repeat	162	16.6
TE_ORF_98	Unknown	Unknown	155	10.5
XLGST3	SINE	tRNA-V-Core-L2	191	8.1
PIR_XL	DNA	Kolobok-T2	380	7.8
TX1	LINE	L1	601	7.5
AT_rich	Low_complexity	Low_complexity	1890	6.9
DNA5_Xt	DNA	TcMar-Tc2	142	6.8
OCR	DNA	hAT-Ac	241	6.6
XLLINE	LINE	CR1	368	6.5
T2_1_Xt	DNA	Kolobok-T2	166	6.3
PIRe_Xt	DNA	Kolobok-T2	140	5.7
TE_ORF_340	Unknown	Unknown	361	5.7
CR1_1b_Xt	LINE	CR1	490	5.5
DNA2_Xt	DNA	PIF-Harbinger	907	5.4
Sat3_Xt	Satellite	Satellite	218	5.3
(TA)n	Simple_repeat	Simple_repeat	415	5.2
CR1_1a_Xt	LINE	CR1	496	5.1
REM1_XL	Satellite	Satellite	222	5.0
POR	DNA	hAT-Charlie	590	4.6
XBR_Xt	DNA	Kolobok-T2	459	4.2
DNA1_Xt	DNA	DNA	265	4.1
Harbinger-N1_XT	DNA	PIF-Harbinger	163	4.1
JH12_XL	DNA	TcMar-Tigger	1548	4.1
DNA4_Xt	DNA	TcMar-Tigger	1190	4.0
MuDR1_Xt	DNA	MULE-MuDR	171	4.0
XBR_XL	DNA	Kolobok-T2	342	3.9
XR_XL	DNA	Kolobok-T2	159	3.7

Supplemental Table 4. No enrichment of ncRNAs at sites of early H3K27me3 nucleation

Dataset	# of sites with ncRNA[1]	p-value[2]	q-value[3]
Gastrula dorsal	27	0.143	0.334
Gastrula ventral	48	0.039	0.273
Adult liver	3	0.575	0.575
Adult skin	7	0.276	0.413
Oocytes stage I,II	20	0.082	0.289
Oocytes stage II, IV	19	0.295	0.413
Oocytes stage V,VI	14	0.359	0.419

[1] Out of 221 regions
[2] One-tailed Fisher's exact test
[3] Benjamini-Hochberg

Supplemental Table 5. Early nucleation sites tested for repressive activity in luciferase vector

Genomic position[1]	Gene
scaffold_51:1047989-1049046	tbr1
scaffold_163:572638-573798	hoxd3
scaffold_55:1284126-1285099	foxl2
scaffold_163:572987-574012	hoxd3
scaffold_63:1612878-1613871	tfap2b
scaffold_68:3214875-3215834	six1
scaffold_163:580186-581153	hoxd4
scaffold_4:5302999-5304022	foxd3
scaffold_11:4266992-4267980	gata3
scaffold_56:1,322,481-1,323,458	evxl
scaffold_4:1,316,132-1,317,077 (Control)	n/a

[1] Genomic position is based on JGI 4.1 (UCSC xenTro2).

Supplemental Table 6. Overview of all public datasets used in this study.

Species	Data type	Cell type / developmental stage	Reference	Source	Remarks
X. tropicalis	short RNA	Gastrula dorsal	Faunes et al., 2011	GSM744253 (1)	
X. tropicalis	short RNA	Gastrula ventral	Faunes et al., 2011	GSM744254 (1)	
X. tropicalis	short RNA	Adult liver	Armisen et al., 2009	GSM372598 (1)	
X. tropicalis	short RNA	Adult skin	Armisen et al., 2009	GSM372601 (1)	
X. tropicalis	short RNA	Oocytes stage I,II	Armisen et al., 2009	GSM372602 (1)	
X. tropicalis	short RNA	Oocytes stage II, IV	Armisen et al., 2009	GSM372603 (1)	
X. tropicalis	short RNA	Oocytes stage V,VI	Armisen et al., 2009	GSM372604 (1)	
zebrafish	H3K27me3 ChIP-seq	shield	Pauli et al., 2012	GSM831521 (1)	mapped to danRer7 using bwa
zebrafish	H3K27me3 ChIP-seq	24h	Irimia et al., unpublished	GSM861348 (1)	mapped to danRer7 using bwa
human	H3K27me3 ChIP-seq	H1-hESC	ENCODE Project Consortium, 2011	(2)	Broad histone modifications
human	H3K27me3 ChIP-seq	HepG2	ENCODE Project Consortium, 2011	(2)	Broad histone modifications
human	H3K27me3 ChIP-seq	HUVEC	ENCODE Project Consortium, 2011	(2)	Broad histone modifications
human	H3K27me3 ChIP-seq	MONOCD14	ENCODE Project Consortium, 2011	(2)	Broad histone modifications
human	H3K27me3 ChIP-seq	NHEK	ENCODE Project Consortium, 2011	(2)	Broad histone modifications
X. tropicalis	MethylCap	stage 9 and 12	Bogdanovic et al., 2011	GSE23913 (1)	mapped to JGI 7.1 using bwa
zebrafish	Bio-CAP	liver, testes	Long et al., 2013	GSE43512 (1)	mapped to danRer7 using bwa
human	Bio-CAP	liver, testes	Long et al., 2013	GSE43512 (1)	mapped to hg19 using bwa
X. tropicalis	Bio-CAP	stage 11-12	Long et al., 2013	GSE43512 (1)	mapped to JGI 7.1 using bwa
zebrafish	Bisulfite sequencing	sphere	Potok et al., 2013	(3)	
human	Bisulfite sequencing	H1-hESC	Lister et al., 2009	GSM432685 (1)	

(1) http://www.ncbi.nlm.nih.gov/geo/
(2) http://genome.ucsc.edu/ENCODE/downloads.html
(3) HCI's GNomEx: https://bioserver.hci.utah.edu/gnomex/gnomexFlex.jsp?topicNumber=10

References

Armisen, J., Gilchrist, M.J., Wilczynska, A., Standart, N., and Miska, E.A. (2009). Abundant and dynamically expressed miRNAs, piRNAs, and other small RNAs in the vertebrate Xenopus tropicalis. Genome Res. *19*, 1766–1775.

Bogdanovic, O., Long, S.W., van Heeringen, S.J., Brinkman, A.B., Gómez-Skarmeta, J.L., Stunnenberg, H.G., Jones, P.L., and Veenstra, G.J.C. (2011). Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. Genome Res. *21*, 1313–1327.

ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.

Faunes, F., Sanchez, N., Moreno, M., Olivares, G.H., Lee-Liu, D., Almonacid, L., Slater, A.W., Norambuena, T., Taft, R.J., Mattick, J.S., et al. (2011). Expression of transposable elements in neural tissues during Xenopus development. PloS One *6*, e22569.

Heeringen, S.J. van, Akhtar, W., Jacobi, U.G., Akkers, R.C., Suzuki, Y., and Veenstra, G.J.C. (2011). Nucleotide composition-linked divergence of vertebrate core promoter architecture. Genome Res. *21*, 410–421.

Van Heeringen, S.J., and Veenstra, G.J.C. (2011). GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. Bioinforma. Oxf. Engl. *27*, 270–271.

Van der Heijden, T., van Vugt, J.J.F.A., Logie, C., and van Noort, J. (2012). Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome occupancy. Proc. Natl. Acad. Sci. U. S. A. *109*, E2514–2522.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Long, H.K., Sims, D., Heger, A., Blackledge, N.P., Kutter, C., Wright, M.L., Grützner, F., Odom, D.T., Patient, R., Ponting, C.P., et al. (2013). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. Elife *2*, e00348.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. *37*, D77–82.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. *22*, 577–591.

Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W., and Sandelin, A. (2009). JASPAR 2010: the greatly expanded

open-access database of transcription factor binding profiles. Nucleic Acids Res. *38*, D105–D110.

Potok, M.E., Nix, D.A., Parnell, T.J., and Cairns, B.R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. Cell *153*, 759–772.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. U. S. A. *99*, 3740–3745.