Large hypomethylated domains serve as strong repressive machinery for key developmental genes in vertebrates

Ryohel Nakamura¹, Tatsuya Tsukahara¹, Wei Qu², Kazuki Ichikawa², Takayoshi Otsuka¹, Katsumi Ogoshi³, Taro L. Saito², Kouji Matsushima³, Sumio Sugano⁴, Shinichi Hashimoto⁵,⁶, Yutaka Suzuki², Shinichi Morishita²,* and Hiroyuki Takeda¹,*

ABSTRACT

DNA methylation is a fundamental epigenetic modification in vertebrate genomes and a small fraction of genomic regions is hypomethylated. Previous studies have implicated hypomethylated regions in gene regulation, but their functions in vertebrate development remain elusive. To address this issue, we generated epigenomic profiles that include base-resolution DNA methylation and histone modification maps from both pluripotent cells and mature organs of medaka fish and compared the profiles with those of human ES cells. We found that a subset of hypomethylated domains harbor H3K27me3 (K27HMDs) and their size positively correlates with the accumulation of H3K27me3. Large K27HMDs are conserved between medaka and human pluripotent cells and predominantly contain promoters of developmental transcription factor genes. These key genes were found to be under strong transcriptional repression, when compared with other developmental genes with smaller K27HMDs. Furthermore, human-specific K27HMDs show an enrichment of neuronal activity-related genes, which suggests a distinct regulation of these genes in medaka and human. In mature organs, some of the large HMDS become shortened by elevated DNA methylation and associate with sustained gene expression. This study highlights the significance of domain size in epigenetic gene regulation. We propose that large K27HMDs play a crucial role in pluripotent cells by strictly repressing key developmental genes, whereas their shortening consolidates long-term gene expression in adult differentiated cells.

KEY WORDS: Epigenetics, DNA methylation, H3K27me3, Vertebrate development, Large hypomethylated domain, Medaka

INTRODUCTION

Cytosine methylation of CpG dinucleotides of the genomic DNA is one of the essential epigenetic modifications in vertebrates. The genomes of all vertebrates studied thus far are globally methylated, and only a small fraction of genomic regions is hypomethylated (Hendrich and Tweedie, 2003; Tweedie et al., 1997). Recent genome-wide analyses have revealed that the majority of gene promoters are hypomethylated (Lister et al., 2009). The hypomethylated promoters are considered as active or permissive for transcription, and DNA methylation at the proximal promoter region is known to be tightly associated with gene silencing (Bird, 2002).

In addition to DNA methylation, histone modifications influence promoter activity (Zhou et al., 2011). Histone H3 lysine 4 (H3K4) methylation is distributed exclusively on hypomethylated DNA regions (Cedar and Bergman, 2009; Hu et al., 2009; Ooi et al., 2007), and positively regulates promoter activity. However, hypomethylated promoters are not exclusively found in genes with active transcription. In embryonic stem cells (ESCs), promoters of developmentally regulated genes are also hypomethylated (Suzuki and Bird, 2008; Xie et al., 2013). These promoters are frequently marked by repressive histone H3 lysine 27 (H3K27) methylation and are proposed to be ‘poised’ for immediate induction during cell differentiation (Bernstein et al., 2006; Zhao et al., 2007). This poised, but not simply silenced, state of developmental genes is thought to be essential for pluripotent cells to maintain the undifferentiated state with pluripotency. H3K27me3 and hypomethylation at developmental gene promoters were also reported in zebrafish and Xenopus early embryos (Akkers et al., 2009; Bogdanovic et al., 2011; Lindeman et al., 2011; Potok et al., 2013; Vastenhouw et al., 2010), suggesting that repression of hypomethylated developmental gene promoters by H3K27me3 is an essential feature in vertebrate development. However, the nature of these hypomethylated promoters marked by H3K27me3 is beginning to be elucidated in the context of development: e.g. the mechanism that regulates the accumulation of H3K27me3.

Recent genome-wide analyses using mammalian cells suggested an antagonistic relationship between the patterns of DNA methylation and H3K27me3 (Brinkman et al., 2012; Lindroth et al., 2008). Importantly, in post-natal mouse brains, DNA methylation at regions flanking proximal promoters was shown to facilitate transcription of neuronal genes by antagonizing H3K27 methylation (Wu et al., 2010). These studies suggest that DNA methylation has diverse functions depending on where it occurs, which in turn indicates the potential role of DNA methylation or hypomethylation outside the proximal promoter in developmental gene regulation. More recently, developmental transcription factor genes were found to be frequently located in conserved large hypomethylated genomic domains, the size of which tends to be much larger than other promoters (Jeong et al., 2014; Long et al., 2013; Xie et al., 2013). However, previous studies have not addressed whether the size of hypomethylated domains affects gene regulation. Furthermore, although global DNA methylation and promoter hypomethylation are conserved among vertebrate genomes, they are not a general feature in other animal species (Suzuki and Bird, 2008). As the role and pattern of DNA methylation vary widely among organisms (Bird, 2002), it is important to ask to what extent their patterns are conserved among vertebrates in order to understand the function and evolution of hypomethylated domains in the vertebrate lineage.
Here, we investigate the epigenomic profiles of medaka blastula embryos (pluripotent cells) and adult tissues by using single-base resolution DNA methylomes from our previous study (Qu et al., 2012) and also by generating genome-wide histone modification maps and additional methylomes. In blastula embryos, large compartmentalized genomic domains with H3K27me3 and DNA hypomethylation (K27HMDs) were found at developmental transcription factor genes. Strikingly, we found that the size of the K27HMD positively correlates with H3K27me3 levels. Thus, whereas hypomethylated promoters usually act as active or permissive domains, K27HMDs provide strong repression when they are broadly hypomethylated. Comparative analyses between medaka and human pluripotent cells revealed both conservation and species-specific changes in K27HMDs, providing insights into the evolution of vertebrate epigenomes. We also found a phenomenon, which we termed ‘HMD shortening’, at genes with sustained expression in adult tissues. We propose that large K27HMDs play a crucial role in pluripotent cells by strictly repressing key developmental genes, while their shortening consolidates long-term gene expression in adult differentiated cells.

RESULTS
Identification of hypomethylated domains in the medaka genome
Cells from medaka blastula embryos are known to retain pluripotency (Yi et al., 2009). To obtain the global genomic distribution of epigenetic modifications in medaka blastula embryos, we performed ChIP-seq analyses using antibodies against H3K27me3, H3K4me1, H3K4me2, H3K4me3 and H3K27ac (supplementary material Table S1), and integrated these results with previously established medaka single-base DNA methylomes (Qu et al., 2012). At the blastula stage, methylation frequency of individual CpG sites has a clear bimodal distribution (supplementary material Fig. S1A) and thus, contiguous regions of a low level of methylation appeared obvious on a genome browser with sharp boundaries (Fig. 1A). By scanning the whole genome, we identified 15,145 regions with more than nine contiguous low methylated CpG sites (methylation frequency <0.4), termed hypomethylated domains or HMDs (supplementary material Table S4). Gene promoters and CpG islands accounted for the majority of HMDs (supplementary material Fig. S1B). H3K4 methylations exclusively distributed to HMDs, as previously reported (Hu et al., 2009; Ooi et al., 2007), and some parts of those regions are co-enriched with other histone modifications (Fig. 1A).

We noticed that HMDs generally had sharp boundaries (Fig. 1B), suggesting that the boundaries may harbor a specific sequence feature. Because HMDs largely overlap with CpG islands, we analyzed the CpG density around the HMD boundary. We found that the CpG density was significantly higher inside the HMD than outside. Interestingly, the CpG density dropped off around the HMD boundary and a CpG-poor region spanned just outside the HMD (Fig. 1C). CpG-poor regions at HMD boundaries have also been reported in human (Molaro et al., 2011), and may have a specific function in the establishment of the boundaries or may contribute to the sharpness. Furthermore, we identified motifs that are significantly enriched around the boundary sequence (supplementary material Fig. S2).

Interestingly, we noticed that one motif showed a high similarity to the CTCF-binding sequence (Fig. 1D). Mapping of CTCF-binding motifs on the medaka genome confirmed that they are indeed highly enriched around HMD boundaries (Fig. 1E). CTCF is known to mediate chromatin looping and functions as a barrier to separate distinct chromatin regions with different modifications (Handoko et al., 2011). Together, these results suggest that HMD boundaries are strictly confined and have specific features in the genomic sequence.

Previous studies reported that nucleosome positioning is affected by CTCF binding and by DNA methylation and histone modifications (Fu et al., 2008; Segal and Widom, 2009; Valouev et al., 2011). To determine the nucleosomal distribution around HMD boundaries bearing a CTCF motif, we re-analyzed the previous data of nucleosome core distribution in medaka blastula embryos (Sasaki et al., 2009). We found clear peaks of nucleosome core signals, and a transition of the periodic pattern of the nucleosomal core position around the CTCF site; the average score of nucleosome core position exhibits a clear 170 bp periodic pattern outside HMDs but the peak becomes low and less defined inside HMDs (Fig. 1F). The binding motif site showed no significant peak, indicating actual CTCF binding at the nucleosome-free region. These results suggest that the nucleosome structure also clearly changes at the HMD boundary from ‘packed’ to ‘loose’.

Large HMDs mark key transcription factor genes for vertebrate development
It is well known that DNA methylation at proximal promoters silences cell type-specific genes, and that permissive promoters are generally hypomethylated (Smith and Meissner, 2013). In some cases, hypomethylated domains extend beyond promoters and cover much larger regions (Bogdanovic et al., 2011; Jeong et al., 2014; Laurent et al., 2010; Long et al., 2013; Xie et al., 2013). However, the significance of such large domain size in gene expression has not been addressed. Interestingly, in medaka genome, although most HMDs were several kb in size and highly enriched with H3K4me2 and other active histone modifications (Fig. 2A,C), we found that a subset of HMDs was extremely large and those regions were frequently enriched with H3K27me3 (Fig. 2B,C). To examine the relationship between the HMD size and H3K27me3, we first classified the HMDs by the existence of H3K27me3 enrichment. 2398 HMDs were found to contain H3K27me3 peaks detected by QuEST software (Valouev et al., 2008) and were classified as K27HMDs (supplementary material Table S4). We also classified the H3K27me3-marked hypomethylated regions using ChromHMM software, which can annotate the regions in a statistically principled manner (Ernst and Kellis, 2012). We confirmed that the K27HMDs largely overlap with the regions identified by ChromHMM (supplementary material Fig. S3). However, regions identified by ChromHMM were frequently divided into smaller fragments, and therefore, we chose K27HMDs that were suitable for the further analysis that focused on domain size. Although H3K27me3-free HMDs (nonK27HMDs) had a high enrichment of H3K4me2, the majority of K27HMDs also had a low but significant enrichment of H3K4me2 (supplementary material Fig. S4).

We next investigated the characteristics of large HMDs. Notably, 12% of K27HMDs were larger than 4 kb, whereas 99.8% of 12,747 nonK27HMDs were less than 4 kb (Fig. 2D). As the majority of HMDs overlapped with promoter regions, we linked medaka genes to promoter-associated HMDs (supplementary material Table S5). Intriguingly, large K27HMD (>4 kb) associated genes (317) and small K27HMDs (<4 kb) associated genes (1295) showed specific features in their functions. Gene ontology analysis showed that terms related to transcription regulation and developmental processes are highly enriched in large K27HMDs (Fig. 2E; supplementary material Fig. S5A). Indeed, 65% of large K27HMD genes encoded DNA-binding factors (supplementary material Table S6), and had crucial functions for embryonic development. By contrast, small K27HMDs showed enrichment of terms for developmental processes, signal transduction and cell communication, with relatively low enrichment.
for transcription factors (Fig. 2E; supplementary material Fig. S5A).
Thus, the large K27HMDs mark a specific set of developmental genes, and may have an important role in vertebrate development. We also confirmed that the HMD size does not reflect gene length, as large K27HMDs tended to have shorter genes than small ones (supplementary material Fig. S5B).

It is known that the teleost underwent whole-genome duplication (Jaillon et al., 2004; Kasahara et al., 2007; Taylor et al., 2003), and we found that some duplicated genes encoding transcription factors showed different HMD size. We picked two duplicated gene pairs [pax6 (ENSORLG00000009913 and ENSORLG00000000847) and tbx2 (ENSORLG00000014792 and ENSORLG00000010011)] for expression analysis because, for each pair, one gene is associated with a large K27HMD and the other with a small K27HMD (supplementary material Fig. S5C,D).

**In situ** hybridization of those gene pairs in medaka embryos demonstrated that the genes with large HMDs are expressed in a tissue-specific pattern (supplementary material Fig. S5E,F), which is nearly identical to the conserved expression pattern in vertebrates, including mouse (Harrelson et al., 2004; Puschel et al., 1992). By contrast, the genes with small HMDs showed no or partial expression of the conserved pattern (supplementary material Fig. S5E,F). These results suggest a conserved function for large K27HMD genes, and strengthen the possibility that the size of the HMD reflects the gene function.

**The size of K27HMD and the number of CpGs in the domain correlate with H3K27me3 levels**
We assumed that the size of the HMDs contributes to the transcriptional regulation of developmental transcription factor...
The fact that most of the large HMDs are enriched with H3K27me3 led us to hypothesize that the size of the HMD is one of the determinants of the level of H3K27me3 accumulation. Comparison of K27HMD size and ChIP signal enrichment or ChIP peak intensity revealed a significant correlation (Fig. 2F,G; supplementary material Fig. S6A,C). By contrast, negative correlation was found between large K27HMD size and ChIP reads per kb for indicated modifications.

Fig. 2. Large HMDs have high H3K27me3 enrichment and mark key transcription factor genes. (A,B) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka blastula embryos are shown for representative nonK27HMD (A) and large K27HMD (B). (C) Heat-map representations of DNA methylation, H3K27me3, H3K4me1, H3K4me2, H3K4me3 and H3K27ac are shown for all HMDs in medaka blastula embryos. HMDs are ordered by size. Dark colors represent high signal; white represents low signal. (D) Size distributions of nonK27HMDs and K27HMDs in blastula embryos. (E) Functional annotations of genes marked by large (top) and small (bottom) K27HMDs. The top three over-represented Gene Ontology (GO) PANTHER biological process terms are shown. The values of x axes (in logarithmic scale) correspond to the P values calculated by the DAVID tool (Huang et al., 2009). (F) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka blastula embryo are shown for large K27HMDs (tbx2 and gata6; top two) and small K27HMDs (jag2 and fz9; bottom two). (G) Boxplots show the correlation between HMD length and ChIP reads per kb for indicated modifications. In the box plots, bottom and top of the boxes correspond to the 25th and 75th percentiles; the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile ranges of the lower and upper quartiles, respectively. (H) Comparison of the number of low methylated CpG sites and the highest H3K27me3 ChIP peak intensity inside K27HMD. Spearman’s rank correlation coefficient ($r_s$) and P values are shown.
domain size and active modification levels (H3K4me1, H3K4me2, H3K4me3 and H3K27ac). Among those active modifications, the negative correlation was most significant for H3K4me2 (Fig. 2F,G; supplementary material Fig. S6A,B). These results indicate that larger K27HMDs have a stronger repressive property for transcription. Given that large K27HMD genes mostly encode transcription factors that are crucial for development, the strong repression of those genes in blastula embryos might prevent improper cell differentiation and maintain stemness.

Previously, Polycomb group (PcG) proteins that mediate H3K27me3 accumulation were reported to preferentially bind to CpG islands (Deaton and Bird, 2011; Ku et al., 2008; Woo et al., 2010), but DNA methylation reduces this binding (Hagarman et al., 2013; Wu et al., 2010). We therefore counted the number of low methylated CpG sites inside the K27HMDs and examined their relationship with H3K27me3 levels. A significant correlation was observed between the low methylated CpG count and the H3K27me3 ChIP peak intensity (Fig. 2H), supporting the possibility that a large K27HMD provides a large number of unmethylated CpGs that can potentially recruit PcG proteins.

Comparison of HMDs between medaka and human pluripotent stem cells

To test whether the importance of the HMD size holds true for other vertebrates, we applied our analysis pipeline to a whole-genome dataset from human ESCs (hESCs) (Lister et al., 2009, 2011). hESCs were chosen because we assumed that they are equivalent to medaka blastula cells in terms of pluripotency. We first tested whether the HMD boundaries in hESCs are also enriched with CTCF. Mapping of CTCF ChIP-seq peaks from the ENCODE Project Consortium (2011) around HMDs showed the highest enrichment at the HMD boundary (supplementary material Fig. S7), suggesting a conserved feature of the HMD boundary between medaka and human. Next, we examined the epigenetic status at promoters of orthologous genes and its conservation. All medaka genes annotated to single human orthologous genes in Ensembl database (13,301 genes) were used for comparison between the two species. We first found that the majority of medaka promoters (59%) have similar DNA and histone modifications in human ES cells (Pearson’s Chi-squared test \( P < 2.2E-16 \); Fig. 3A; supplementary material Table S7). Furthermore, the comparison of HMD size between medaka and human further revealed a conserved tendency: whereas nonK27HMDs largely reside in small size fraction in both medaka and human (Fig. 3B), genes marked by large K27HMDs in medaka embryos are also marked by large K27HMDs in hESCs (Fig. 3B,C). We defined HMDs larger than 8 kb as large HMDs in human, so that the number of genes designated as large K27HMDs is equal to large K27HMD (>4 kb) genes in medaka. We confirmed that the genes associated with conserved large K27HMDs are highly enriched for GO terms related to transcription regulation and developmental processes (Fig. 3D). Furthermore, at the level of protein sequence, genes found in large K27HMDs are more conserved between the two organisms than those in small K27HMDs or nonK27HMDs, and genes with no HMD (i.e. methylated) showed the lowest conservation (supplementary material Fig. S8), indicating that large K27HMD genes are under strong evolutionary constraint.

As in medaka blastula embryos, large K27HMDs in hESCs accumulated higher levels of H3K27me3 (Fig. 3E), and the correlation between the number of low methylated CpG sites and H3K27me3 levels was also significant in hESCs (supplementary material Fig. S9). Thus, large K27HMDs seem to be more repressive than small K27HMDs. We tested this idea by analyzing the transcriptome data set from a previous study (Lister et al., 2011). First, as expected, nonK27HMD genes exhibit high levels of expression, whereas the majority of no HMD (methylated) genes are silenced (Fig. 3F). Consistent with the poised model (Bernstein et al., 2006; Pan et al., 2007), the expression levels of K27HMD genes tend to be low. Importantly, large K27HMD genes showed significantly lower expression than small ones, almost similar levels to methylated genes (Fig. 3F). This tendency was also observed in blastula embryos, but was not so evident (data not shown), probably owing to the presence of maternally derived transcripts and due to unusual transcriptional environment of mid-blastula transition that the blastula embryos had just experienced (Aizawa et al., 2003). Taken together, these results suggest the large size of K27HMD at key transcription factor genes is conserved among vertebrate species, and have strong repressive effect on gene expression in pluripotent cells.

Human-specific K27HMDs mark genes related to neuronal activities

Although the epigenetic status of homologous genes is largely conserved between medaka and human pluripotent cells, a subset of genes has been subjected to differential DNA and histone methylations. We looked into these differences focusing on human-specific K27HMD genes, because they could reflect changes in function that occurred during vertebrate evolution. For this purpose, K27HMD genes in hESCs were classified into three categories according to the epigenetic status of their medaka counterparts. Class 0 is a category of K27HMD genes shared by the two species as described above, whereas genes of the latter two became K27HMDs in human from nonK27HMDs (Class I) or methylated (Class II) in medaka, respectively (Fig. 3A). We then determined GO terms enriched for these three classes of hESC K27HMDs. We first confirmed that developmental genes are highly enriched in Class 0 K27HMDs (Fig. 4A). By contrast, no such high enrichment for developmental genes was observed for class I and class II genes. Furthermore, the large K27HMDs were mostly included in Class 0 (supplementary material Fig. S10), strongly suggesting that this epigenetic machinery is essential for development and thus conserved among vertebrates. Interestingly, class II genes are more closely associated with signal transduction and neuronal activities (Fig. 4B,C; supplementary material Table S8), whereas Class I genes did not show any enrichment of specific terms. We also failed to find any preference of GO terms for genes in medaka-specific K27HMDs (data not shown).

The human-specific HMDs could be due to a different differentiation state between the two types of cells, medaka blastula cells and hESCs, despite their presumed pluripotency. However, sequence conservation of three classes of human K27HMDs among vertebrates revealed that Class I and II HMDs are more divergent than Class 0 (Fig. 4D), suggesting that genetic variations in cis-elements, rather than cellular characteristics of the cells, account for the observed differences in epigenetic modifications between medaka and human.

Dynamics of the epigenetic state of zic1/zic4 genes during development

Next, we sought to determine how the epigenetic state of large K27HMDs changes from embryonic stages to adult. We focused on the zic1 and zic4 (zic1/zic4) genes because they have a particularly large K27HMD (26 kb) in medaka blastula embryos (Fig. 5A) and their function has been extensively studied in medaka (Kawanishi et al., 2013; Moriyama et al., 2012). zic1/
zic4, organized head-to-head with a small intergenic sequence (~3.4 kb), encode zinc-finger transcription factors and function in the neuronal development and specification of dorsal fate in the trunk (Kawanishi et al., 2013; Merzdorf, 2007). They are thought to share cis-regulatory elements and thus their expression is nearly identical. It is known that their expression is activated by secreted factors during embryogenesis, and is autonomously maintained throughout life in the dorsal parts of somite derivatives, i.e. the myotome, dermis and vertebrae (Kawanishi et al., 2013). Thus, these genes served as a good example with which to analyze the change in epigenetic states of key developmental genes from embryo to adult. We isolated dorsal (zic-positive) and ventral (zic-negative) myotome separately from adult fish and generated methylomes and histone modification maps for each half (supplementary material Tables S1 and S3, Fig. S11). Although the overall methylation pattern was similar between dorsal and ventral myotome, we noticed a significant difference in the chromatin state at the zic1/zic4 locus (Fig. 5A). In the dorsal myotome, the zic1/zic4 locus showed low H3K27me3 levels and high H3K4me2 levels, consistent with the state of active transcription. Surprisingly, however, we found large blocks of DNA hypermethylation at regions outside the H3K4me2-enriched promoter regions, which led to the

![Fig. 3. Comparison of HMDs between medaka and human.](image-url)
shortened HMD around the promoter region (Fig. 5A, blue box). We termed this phenomenon ‘HMD shortening’. By contrast, the ventral myotome maintained the blastula-type epigenetic state: high H3K27me3 and low DNA methylation. Given that HMD size correlates with transcriptional repression level, this HMD shortening might promote zic1/zic4 expression. To examine when this HMD shortening occurs during development, we additionally investigated the DNA methylation pattern in dorsal and ventral myotome at the hatching stage, a stage when the dorsal-specific expression of zic1/zic4 is already induced (Kawanishi et al., 2013). ChIP-qPCR analysis confirmed the high enrichment of H3K27me3 remained in the dorsal myotome of Da (Fig. 5B; supplementary material Fig. S12A). By contrast, the dorsal-specific DNA hypermethylation was not detected at this stage (Fig. 5A). Bisulfite sequencing at the larval stages revealed a gradual increase in DNA methylation as larval development progressed (Fig. 5C). Thus, HMD shortening occurs after the establishment of active histone modifications and gene expression in dorsal myotome. These results suggest that DNA hypermethylation in the zic1/zic4 HMD is not necessary for the initial induction but rather for the maintenance of gene expression.

We hypothesized that HMD shortening at later stages depends on the active chromatin state. To explore this possibility, we determined the DNA methylation pattern in the zic1/zic4 mutant Double anal fin (Da). The Da mutant has a large transposon insertion that impairs the mesoderm enhancer of zic1/zic4, leading to a dramatic reduction in myotome expression and ventralized dorsal structures in the trunk (supplementary material Fig. S12A; Moriyama et al., 2012). ChIP-qPCR analysis confirmed the high enrichment of H3K27me3 remained in the dorsal myotome of Da

Fig. 4. Characteristics of human-specific K27HMDs. (A,B) Functional annotation of genes marked by Class 0 (K27HMD in medaka; A) and Class II (methylated in medaka; B) human K27HMDs. The top over-represented GO PANTHER biological process terms are shown. The values of x axes (in logarithmic scale) correspond to the P values calculated by the DAVID tool. (C) Genome browser representations of DNA methylation, H3K27me3 and H3K4me2 in medaka (top) and human (bottom) are shown for Class II genes (grin2a, grin2b, grik2 and cntn4). (D) Average vertebrate PhastCons profiles around the boundaries of Class 0 (K27HMD in medaka; green), Class I (nonK27HMD in medaka; orange) and Class II (methylated in medaka; red) human K27HMDs.
Strikingly, bisulfite-seq showed a complete absence of the dorsal-specific DNA hypermethylation in myotome of Da (Fig. 5A), suggesting the requirement of zic1/zic4 activation for HMD shortening. To further elaborate this possibility, we made Da heterozygous fish and investigated DNA methylation of each allele. wt and Da alleles were distinguished by single base nucleotide variations. In the dorsal myotome of heterozygous fish, only the zic1/zic4 locus of the wt allele is activated and that of the Da allele remained repressive because of the transposon insertion. Bisulfite sequencing revealed that the wt allele was highly methylated, whereas the Da allele remained hypomethylated (Fig. 5D). Thus, we conclude that the activation of the zic1/zic4 locus, but not of Zic1/Zic4 proteins or their downstream factors, induces HMD shortening. Collectively, these results suggest that once zic1/zic4 are induced during embryogenesis, they are autonomously subject to HMD shortening during growth.

HMD shortening associates with sustained gene expression in matured organs

The epigenetic profile of the zic1/zic4 locus suggested that large K27HMDs have a repressive function, while the shortening of the K27HMD associates with their sustained expression. We also noticed that the large K27HMD at Hox gene clusters underwent the remarkable DNA hypermethylation that led to shortened HMDs at the active promoter in adult myotome (supplementary material Fig. S12C). To address whether HMD shortening at active gene loci is a general feature in adult tissues, we compared the epigenomes of the adult dorsal myotome and liver with that of blastula embryos. For liver epigenome data, we generated histone modification maps (supplementary material Table S1), and integrated these with methylome data from a previous study (Qu et al., 2012). We found that the majority of K27HMDs have unchanged methylation levels in adult tissues, but a significant

Fig. 5. Active chromatin-dependent K27HMD shortening at zic1/zic4 locus in adult myotome. (A) Genome browser representation at the zic1/zic4 loci. DNA methylation, H3K27me3 and H3K4me2 are shown for blastula embryos, and for adult dorsal and ventral myotome; DNA methylation is shown for dorsal and ventral myotome of hatching stage larvae and adult Da mutant. Red box indicates a large K27HMD identified in the blastula embryo and blue box indicates a shortened HMD in adult dorsal myotome. Navy bars (labeled A-D) indicate regions examined for ChIP-qPCR and bisulfite sequencing. (B) ChIP-qPCR analysis of dorsal and ventral myotome from wt adult, hatching stage larvae and Da adult for H3K27me3 (top) and H3K4me2 (bottom) at regions indicated in A. Error bars represent s.d. from three technical replicates. (C) Bisulfite sequencing in dorsal myotome at multiple stages after hatching. Graphs indicate the ratio of methylated CpGs at each stage. The examined regions are indicated in A. (D) Bisulfite sequencing in adult Da heterozygous fish at region C in A. Even in the same nucleus, only the wt allele shows DNA hypermethylation.
proportion of HMDs were subjected to DNA hypermethylation in which more than 5% of their CpG sites became highly methylated (>0.4) (Fig. 6A; supplementary material Tables S4 and S5). Indeed, 8.3% of K27HMDs identified at the blastula stage acquired elevated DNA methylation in both adult myotome and liver, and 9.6% and 5.3% showed elevated DNA methylation specifically in adult myotome and adult liver, respectively (Fig. 6B). Large K27HMDs (>4 kb) tend to be more frequently methylated in adult tissues, 15.7% for both, and 20.9% and 5.2% only for myotome and liver, respectively (Fig. 6B). Strikingly, the large K27HMDs with elevated DNA methylation still retained low methylation levels around TSSs in adult tissues, resulting in the HMD shortening toward the promoter region (Fig. 6A,C; supplementary material Fig. S13A). Similar to zic1/zic4, this change in DNA methylation occurred mainly after the hatching stage in myotome (Fig. 6C).

Consistent with the fact that larger K27HMD have higher H3K27me3 enrichment in pluripotent cells, shortened K27HMD in adult tissues showed significantly lower levels of H3K27me3 than unchanged K27HMDs (Fig. 6D; supplementary material Fig. S13B). Lower levels of H3K27me3 associated with elevated DNA hypermethylation at active loci. (A) Genome browser representation of DNA methylation, H3K27me3 and H3K4me2 enrichment are shown for K27HMD with myotome-specific hypermethylation (six2; left) and liver-specific hypermethylation (hnf6; right). Red box indicates large K27HMD identified in the blastula embryo and blue box indicates shortened HMD in adult dorsal myotome. (B) Fraction of K27HMDs with DNA hypermethylation in adult myotome and liver for all K27HMD (left) and large K27HMD (right). (C) Average DNA methylation around TSSs marked by a large K27HMD with elevated DNA methylation in myotome at multiple stages. (D,E) Boxplots show H3K27me3 enrichment (D) and gene expression (E) at hypermethylated (elevated methylation) and unchanged K27HMDs in adult myotome. The number of HMDs (E) and genes linked to each HMD category (F) are shown under the boxes. P values were calculated using non-paired Wilcoxon tests. (F,G) Models for the HMD size-dependent repression of developmental genes in pluripotent cells (F) and for HMD shortening associated with sustained gene expression in adult differentiated cells (G).
DNA methylation because the levels of H3K27me3 accumulation in blastula embryos did not show any difference between the two groups (elevated methylation and unchanged methylation in adult) of K27HMDs (supplementary material Fig. S13D). To examine whether HMD shortening is correlated with gene expression, we performed RNA-seq for adult myotome and liver (supplementary material Table S2). Consistent with the reduced H3K27me3 levels, gene expression levels were significantly higher in the K27HMDs with elevated methylation than in K27HMDs with unchanged methylation (Fig. 6E; supplementary material Fig. S13C). Thus, K27HMD shortening is associated with reduced H3K27me3 levels and with active gene expression. By contrast, DNA methylation occurred in some nonK27HMDs and, in this case, a negative correlation between DNA methylation and gene expression was observed (supplementary material Fig. S14). These results suggest opposing functions of DNA methylation on the expression of nonK27HMDs and K27HMDs.

Importantly, a correlation between the HMD size and H3K27me3 levels was also observed for unchanged K27HMDs in adult myotome and liver (Fig. 6D; supplementary material Fig. S13B, unchanged small versus unchanged large). Furthermore, the expression level was significantly lower for unchanged large K27HMDs than for small K27HMDs (Fig. 6E; supplementary material Fig. S13C). Taken together, these data suggest that the large K27HMDs have a strong repressive effect on gene expression in pluripotent and adult matured cells (Fig. 6F), and the shortening of the large K27HMDs may facilitate persistent gene expression in adult tissues (Fig. 6G).

To further investigate whether HMD shortening occurs in other vertebrate species, we analyzed methylome and transcriptome data from zebrafish sphere stage embryo (an equivalent stage for medaka blastula) and adult myotome (Potok et al., 2013). We identified 237 HMDs larger than 8 kb in zebrafish sphere embryos, and 139 of those showed elevated methylation in adult myotome (supplementary material Table S9). Expression levels of genes associated with elevated methylation were significantly higher than those of unchanged HMD genes (supplementary material Fig. S15). These results suggest that HMD shortening also associates with active gene expression in zebrafish. Therefore, the regulation of HMD size could be an important epigenetic mechanism for the strict and stable regulation of key developmental genes in vertebrates.

**DISCUSSION**

Previous studies have demonstrated that the H3K27me3 domains often occupy several kb around promoters of developmental genes, but at some gene loci, the domains exceed this to cover much larger regions (Zhao et al., 2007). Very recent studies also identified the large genomic domains with DNA hypomethylation at transcription factor gene loci in various vertebrate species (Jeong et al., 2014; Long et al., 2013; Xie et al., 2013). However, the biological significance of the size of these genomic domains has not been addressed. H3K27me3 marked promoters are proposed to keep developmental genes poised for immediate activation in pluripotent cells (Bernstein et al., 2006). Indeed, the transcriptional activity of K27HMD genes in undifferentiated cells was found to be low, but slightly higher than that of methylated genes (Fig. 3F, small K27HMD genes versus methylated genes), suggesting leaky transcription under polycomb-mediated repression. Importantly, we found that the H3K27me3 level correlates with K27HMD size and the transcription level of genes in large K27HMDs is significantly lower than that in small ones in both differentiated medaka tissues and human ESCs. In pluripotent cells, the large K27HMDs preferentially mark transcription factor genes with crucial function in development, whereas the smaller ones tend to mark genes related with signal transduction. Transcription factors in large K27HMDs, therefore, are strictly shut off if not required. A previous study in zebrafish sperm also reported that genes with high levels of H3K27me3 at proximal promoters mostly encode transcription factors (Wu et al., 2011), which are largely overlapped with genes identified by the domain size (large K27HMD genes) in our study. This further emphasizes the generality of size-dependent H3K27me3 accumulation and strong repression of transcription factor genes in vertebrates. The strict repression is important presumably because the derepression of those transcription factors would result in inappropriate cell differentiation (Boyer et al., 2006; Fujikura et al., 2002) or malignancies such as cancer (Darnell, 2002). The precise mechanism regulating the amount of H3K27me3 awaits future studies. However, our data suggest that large HMDs provide a broad platform with unmethylated CpGs that potentially recruit PcG proteins to enrich H3K27me3, while small HMDs have limited capability to bind PcG proteins.

Recent studies have identified the super-enhancer in which large domains occupied by clusters of enhancers drive strong expression of key cell-identity genes (Whyte et al., 2013). The concept of the large K27HMD is reminiscent of that of the super-enhancer, but the two large genomic domains have opposing functions: robust repression versus strong activation. These large domains with dense epigenetic modifications may strengthen the repression or activation effect, and cooperatively contribute to the control of cell identity.

During growth and differentiation, large K27HMDs undergo substantial changes in histone modification and DNA hypermethylation when their genes are activated. Interestingly, Polycomb-target genes in ESCs often acquire hypermethylation after differentiation (Mohn et al., 2008; Rush et al., 2009), but the underlying mechanism is unknown (Deaton and Bird, 2011). The analysis of the large K27HMD harboring the zic1/zic4 genes revealed that DNA hypermethylation is locus activity dependent and gradually occurs after the histone modification pattern has changed. A likely scenario is that the reduction of H3K27me3 allows the accumulation of DNA methylation outside a proximal promoter harboring H3K4me2 and the shortened HMD then protects a region from H3K27me3 accumulation, resulting in consolidation of the long-term expressions of developmental genes. Consistently, we have previously demonstrated two distinct regulations of zic1/zic4 transcription: the expression in somites is induced by secreted signals but is later maintained throughout life in a cell-autonomous manner (Kawashiki et al., 2013). Similar to zic1/zic4, it has been reported that the embryonic Hox expression pattern is epigenetically maintained in fibroblasts of the human adult foot and is required to maintain its site-specific identity (Rinn et al., 2008). Notably in our dataset, very large K27HMDs covering Hox clusters observed at the blastula stage undergo HMD shortening at active loci in the adult myotome. Given that DNA methylation is a stable epigenetic modification (Smith and Meissner, 2013), this K27HMD shortening could serve as cellular memory by marking developmental genes once activated. We propose the two-step regulation of key developmental genes: histone modification-dependent induction and HMD shortening-dependent long-term maintenance.

Our comparative analysis demonstrated that the overall pattern of DNA methylation is conserved between medaka and human. Moreover, the majority of gene promoters shared the same epigenetic states between the two systems. In particular, large K27HMDs were highly conserved; thus, pre-marking of key transcription factor genes by a large K27HMD and its strong
repression are shared features in the vertebrate development. Importantly, our analysis also identified human-specific K27HMD genes and found that those methylated in medaka pluripotent cells (Class II) are related to neuronal activities. Human-specific K27HMDs had lower sequence conservation than those common to both human and medaka. Indeed, regulatory elements under human constraint, but not conserved in mammals, were found to be associated with neuronal activities (Ward and Kellis, 2012). These neuronal genes are not expressed in both medaka and human pluripotent cells, but the repression is mediated by distinct modifications: DNA methylation and H3K27me3, respectively. What, then, is the biological significance of such differential epigenetic marking? One possibility is that K27HMD may enable flexible changes in regulation of those genes; the poised state or sustained expression can be achieved through regulation of histone modification and of HMD size. Interestingly, a significant number of glutamate receptor genes were found at Class II human-specific K27HMDs. Glutamate receptors are known to have a key function in synaptic plasticity and are important for learning and memory (Lancaster and Dalmau, 2012; Salter and Kalia, 2004). The flexible regulation of these genes could be required for the sophisticated neuronal network in human.

MATERIALS AND METHODS
Fish strains
We used medaka d-rR strains as wild type. The Da mutant used in this study was the same strain as previously described (Moriyama et al., 2012). We crossed the Da homozygous mutant and d-rR to generate Da heterozygous mutants. Medaka fish were maintained and raised under standard conditions. Identification of developmental stages was performed as previously described (Iwamatsu, 2004).

Whole-mount in situ hybridization
Whole-mount in situ hybridization was performed as previously described (Takashima et al., 2007). For tbx2 (ENSRG00000014792 and ENSRLG00000010011) primers for DIG-labeled RNA probes are listed in supplementary material Table S10. For pax6 (ENSRG00000009913 and ENSRLG00000008474) 3’UTR was cloned by SMARTer RACE cDNA Amplification kit (Clontech) and used for RNA probe synthesis. Primers are listed in supplementary material Table S10.

RT-qPCR
RNA was isolated from adult wild-type and Da mutant muscle (dorsal and ventral parts were separated) using ISOGEN (Nippon Gene), according to the manufacturer’s protocol. SuperScrip III (Invitrogen) was used for cDNA synthesis. qPCR was performed with the Stratagene Mx3000P system (Agilent Technologies) using the THUNDERBIRD SYBR qPCR mix (TOYOBO). All primers are listed in supplementary material Table S10.

Bisulfite sequence
Genomic DNA was isolated from adult and larva muscle (dorsal and ventral parts were separated) and bisulfite treatment was performed using methyl easy DNA Bisulfite Modification Kit (Human Genetic Signatures). Bisulfite-converted DNA was subjected to PCR using Ex Taq (TaKaRa) and TOPO-TA cloning (Life Technologies). Amplified fragments were sequenced and analyzed using the QUMA software (Kumaki et al., 2008). All primers are listed in supplementary material Table S10.

Whole-genome bisulfite sequence (WGBS)
For adult and larva muscle (dorsal and ventral parts were separated), sample preparation, library construction, sequencing and mapping for WGBS were performed as previously described (Qu et al., 2012). Briefly, genomic DNA from medaka was isolated and sonicated to a desired size range (100-400 bp). The DNA fragments were treated with DNA polymerase to generate blunt ends and were ligated with double-stranded DNA adaptors containing methylated cytosines, which were designed to amplify only those DNA fragments carrying bisulfite-converted adaptor sequences at both ends. Followed by 7-10 cycles of PCR, 250-450 bp fractionated DNA was sequenced using an Illumina GAIIx genome analyzer. We converted all cytosines in reads and in both the Watson and Crick strands of the reference genome to thymines for primary mapping, and used Smith-Waterman alignments between the original sequences of primary best hits. The level of methylation of a particular cytosine was estimated by dividing the number of mapped reads reporting a cytosine (C) by the total number of reads reporting a C or T (thymine).

ChIP
ChIP was performed using the following antibodies: H3K27me3 (Millipore, 07-449), H3K4me1 (Millipore, 07-436), H3K4me2 (Millipore, 07-473) and H3K27Ac (abcam, ab4729). ChIP was performed as previously described with modifications (Lindeman et al., 2009). Cells were dissociated using a 21G needle and fixed with 1% formaldehyde for 10 min at RT then quenched by adding glycine (200 mM final). After washing with PBS containing 20 mM Na-butylate, complete protease inhibitor (Roche) and 1 mM PMSF, cell pellets were suspended in lysis buffer [50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS, 20 mM Na-butylate, complete protease inhibitors, 1 mM PMSF], sonicated ten times using sonifer (Branson) at power five, and centrifuged for collecting chromatin lysates. The chromatin lysates were diluted with ChIP RIPA buffer [10 mM Tris-HCl (pH 8.0), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton-X100, 0.1% SDS, 0.2% sodium deoxycholate, 20 mM Na-butylate, complete protease inhibitors, 1 mM PMSF] and rotated with an antibody/protein A Dynabeads complex overnight at 4°C. Immunoprecipitated materials were washed three times with ChIP RIPA buffer and once with TE buffer, followed by elution with Lysis buffer at 65°C over night. Eluted samples were treated with RNase A for 2 h at 37°C and proteinase K for 2 h at 55°C, and DNA was purified by phenol/chloroform extraction and ethanol precipitation. Input DNA was simultaneously treated from the elution step. Input and immunoprecipitated DNAs were analyzed by quantitative PCR with the Stratagene Mx3000P system (Agilent Technologies) using the THUNDERBIRD SYBR qPCR mix (TOYOBO). All primers are listed in supplementary material Table S10.

ChIP-seq
For ChIP-seq analysis, we used ~10^6 cells for one modification. Using the input and immunoprecipitated samples of ChIP, ChIP-seq templates were prepared using the TrueSeq DNA sample prep kit (version 2; Illumina) and sequenced using Genome analyzer IIX (Illumina). The sequence was read by 36-base single-end sequencing.

RNA-seq
RNA-seq was isolated from adult muscle (dorsal and ventral parts were separated) and adult liver using RNAeasy mini kit (Qiagen) or ISOGEN (Nippon Gene), according to the manufacturer’s protocol. Purified RNAs were treated with Riboninus eukaryote kit for RNA-seq (Life Technologies). RNA-seq analysis was conducted basically following the instructions from the manufacturer. Briefly, the RNA-seq template was prepared using TrueSeq RNA-seq sample prep kit (version 2; Illumina), omitting the poly(A) selection procedure. The double-stranded PCR products were purified and size fractionated using a bead-mediated method (AMPure, Ambion). Sequencing was conducted on Genome analyzer IIX platform (Illumina) using TrueSeq Cluster generation kit. At least 20 million sequences of a 36-base single-reads were generated per sample.

Data access
All sequence data are deposited at the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) (accession number SRP029233).

Data analyses
The methods for data analyses are described in detail in the supplementary material.


Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y. and Sun, Y. E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science 329, 444-448.


Supplementary Material

Large hypomethylated domain serves as strong repressive machinery for key developmental genes in vertebrates
Ryohei Nakamura¹, Tatsuya Tsukahara¹, Wei Qu², Kazuki Ichikawa², Takayoshi Otsuka¹, Katsumi Ogoshi⁴, Taro L Saito², Kouji Matsushima⁴, Sumio Sugano³, Shinichi Hashimoto⁴,⁵, Yutaka Suzuki², Shinichi Morishita²,*, and Hiroyuki Takeda¹,*

Figure S1 – S15
Table S10
Supplementary methods
Figure S1. Characteristics of hypomethylation in medaka blastula embryos

(A) Histogram of methylation level at each CpG site for blastula embryos.

(B) Fractions of HMDs overlapped with promoters (defined as regions from 5 kb upstream to 2 kb downstream to the TSSs) and CpG islands.
Figure S2. Characteristics of HMD boundary

The DNA motifs enriched at HMD boundaries. P-value was determined by the MEME program.
Figure S3. Comparison of K27HMD and region called by ChromHMM

(A) Venn diagram showing the number of K27HMD only, ChromHMM called only, and shared locations. Most of the K27HMDs overlapped with regions called by ChromHMM.

(B) Venn diagram showing the overlapped length of genomic regions annotated as K27HMD or ChromHMM.
Figure S4. H3K4me2 enrichment at HMDs

Boxplots showing the enrichment of H3K4me2 ChIP reads for nonK27HMD, K27HMD and randomized regions. In the box plots, the bottom and top of the boxes correspond to the 25th and 75th percentiles and the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile range of the lower and upper quartiles, respectively.
Figure S5. Large K27HMDs mark key transcription factors important for development

(A) Full lists of GO terms significantly enriched for genes marked by large (top) and small (bottom) K27HMDs. PANTHER biological process terms are shown. The x axes values (in logarithmic scale) correspond to the P-values calculated by DAVID tool (Huang et al., 2008).

(B) Boxplots show the gene length of each HMD category. In the boxplots, the bottom and top of the boxes correspond to the 25th and 75th percentiles and the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile range of the lower and upper quartiles, respectively.

(C and D) Genome browser representation of DNA methylation, H3K27me3 and H3K4me2 enrichment
are shown for duplicated genes, *pax6* (B) and *tbx2* (C).

(E and F) in situ hybridization for *pax6* (E), and *tbx2* (F) in somite stage medaka embryos (st. 27). Lateral view (left) and dorsal view (right) are shown. Arrowheads point to specific expressions.
Figure S6. The size of K27HMD correlates with H3K27me3 level

(A) Comparison of the size of K27HMD with mapped ChIP reads per kb for H3K27me3 (left) and H3K4me2 (right). Spearman’s rank correlation coefficient ($r_s$) and p value are shown.

(B) Boxplots show the correlation between the HMD length and mapped ChIP reads per kb for indicated histone modifications and input DNA. In the boxplots, the bottom and top of the boxes correspond to the 25th and 75th percentiles and the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile range of the lower and upper quartiles, respectively.

(C) Comparison of the K27HMD size with H3K27me3 ChIP peak intensity. Spearman’s rank correlation coefficient ($r_s$) and p value are shown.
Figure S7. CTCF peak distribution at HMD in hESCs

Distribution of the CTCF ChIP-seq peaks around boundaries of HMDs larger than 3 kb in hESCs. The number of CTCF ChIP peak centers in 50 bp window were counted.
Figure S8. Large K27HMD genes are highly conserved in protein sequence

Boxplots show percent identity of medaka protein sequences to human orthologues. P-values were calculated using non-paired Wilcoxon tests.
Figure S9. Correlation between low methylated CpG count and H3K27me3 level in hESCs
Comparison of the number of low methylated CpG sites and the highest H3K27me3 ChIP peak intensity inside K27HMD. Spearman’s rank correlation coefficient ($r_s$) and p value are shown.
Figure S10. Large K27HMD is conserved between human and medaka

Size distributions of Class 0 (K27HMD in medaka; green), Class I (nonK27HMD in medaka; orange) and Class II (methylated in medaka; red) human K27HMDs in human ESCs.
Figure S11. Summary of whole genome bisulfite sequencing data

(A) Read coverage at individual positions of bisulfite sequencing.

(B and C) Distributions of methylation rate for CpG cytosine sites (A) and Non-CpG cytosine sites (B).
Figure S12. Active chromatin dependent K27HMD shortening in adult myotome

(A) Relative expression of *zic1* measured by RT-qPCR in dorsal and ventral myotome from *Wt* and *Da*. Expression is normalized to that of the housekeeping gene *ef1a*. Error bars represent s.d. from three biological replicates.

(B) Biological replicate for ChIP-qPCR in Figure 5B. Error bars represent s.d. from three technical replicates.

(C) Genome browser representation of DNA methylation, H3K27me3 and H3K4me2 enrichment in the blastula embryo, adult myotome and liver are shown for large K27HMD covering HoxC cluster. Red dashed box indicates large K27HMD regions. Note that the methylation levels of promoter regions of *hoxc11a, 10a, 3a*, and *8a* remain low in adult myotome (blue dashed boxes).
Figure S13. HMD shortening associates with sustained gene expression in adult tissues

(A) Average DNA methylation around TSSs marked by large K27HMD with elevated DNA methylation in liver.

(B and C) Boxplots show H3K27me3 enrichment (B) and gene expression (C) at hypermethylated and unchanged K27HMDs in adult liver. P-values were calculated using non-paired Wilcoxon tests.

(D) Boxplots show H3K27me3 enrichment for each K27HMD category in blastula embryos.
Figure S14. DNA hypermethylation of nonK27HMD associates with gene silencing

(A) Fraction of nonK27HMDs with DNA hypermethylation in adult myotome and liver

(B) Boxplots show gene expression levels (RPKM) at hypermethylated and unchanged nonK27HMDs for adult myotome and liver. P-values were calculated using non-paired Wilcoxon tests.

(C) Genome browser representation of DNA methylation, H3K27me3 and H3K4me2 enrichment in the blastula embryo, adult myotome and liver are shown for nonK27HMD with hypermethylation (for example, tdrd9; left, rnf212; middle, piwil2; right).
Figure S15. HMD shortening associates with active gene expression also in zebrafish

(A) Boxplots show gene expression levels (RPKM) of genes associate to large (>8 kb) HMDs with unchanged methylation and elevated methylation in adult myotome. P-values were calculated using non-paired Wilcoxon tests.

(B) Genome browser representation of DNA methylation and gene expression in the zebrafish sphere stage embryo and adult myotome are shown for large HMD which undergo shortening (for example, *tbx15*). Relative read coverage normalized by total mapped reads are shown for RNA-seq tracks. Red and blue boxes represent a large HMD in sphere and a shortened HMD in myotome, respectively.
### Table S10. Primers used in this study

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>zic1_promoter_F</td>
<td>CATCAGATGAGCGTGTAGG</td>
</tr>
<tr>
<td>zic1_promoter_R</td>
<td>CTGAGACGACTGAGACGAG</td>
</tr>
<tr>
<td>zic1_negative_F</td>
<td>ACGCTGCATGCATCAAACAAGGC</td>
</tr>
<tr>
<td>zic1_negative_R</td>
<td>TGTCACAACACCCGGGCACAC</td>
</tr>
<tr>
<td>bisulfite_B_F</td>
<td>TGGAAGTTGTATAATAAGTTTTTT</td>
</tr>
<tr>
<td>bisulfite_B_R</td>
<td>AAATATACCACATACTTCACACACCTAC</td>
</tr>
<tr>
<td>bisulfite_C_F</td>
<td>GAGTTTTTGTGGAGTGATGGGGATG</td>
</tr>
<tr>
<td>bisulfite_C_R</td>
<td>AACTTAACCTTTACCTTTATATTTCCCC</td>
</tr>
<tr>
<td>tbx2–1_F</td>
<td>AACGTGCACTGACAGTGAAAC</td>
</tr>
<tr>
<td>tbx2–1_R</td>
<td>TGGGTGAACAAACAGTGGTG</td>
</tr>
<tr>
<td>tbx2–2_F</td>
<td>GTCTTTTCCCCCACAGATG</td>
</tr>
<tr>
<td>tbx2–2_R</td>
<td>CCCAATGACATCTGTCCCTGG</td>
</tr>
<tr>
<td>pax6–1_3race</td>
<td>TGTCAGTCCCGAGGAGCGGAGCCT</td>
</tr>
<tr>
<td>pax6–2_3race</td>
<td>GGTCCAAAGTCCCAGGAAGTGAAAGCA</td>
</tr>
<tr>
<td>zic1_RTqPCR_F</td>
<td>AGCCCTTTCCGTGTCGTTCC</td>
</tr>
<tr>
<td>zic1_RTqPCR_R</td>
<td>CCGACGTGTGGACGTGACATGT</td>
</tr>
<tr>
<td>ef1a_RTqPCR_F</td>
<td>AAGGCTGAGCGTGAGCCTG</td>
</tr>
<tr>
<td>ef1a_RTqPCR_R</td>
<td>CTCACCAAACGCGACAGGCGA</td>
</tr>
</tbody>
</table>
Methods for data analyses

Alignment for ChIP-seq and RNA-seq reads

After removing low-quality reads (those containing five or more undetectable bases in 36mer-length reads), we mapped the remaining reads using Burrows-Wheeler Aligner mapping software (http://bio-bwa.sourceforge.net); no more than three mismatches and no gap were allowed. Only uniquely mapped reads were used for the further analysis. The medaka genome and predicted gene sequences were downloaded from Ensembl database (http://www.ensembl.org).

ChIP-seq data processing and analysis

After the mapping, redundant reads were excluded to avoid potential PCR bias. QuEST software (version 2.4 (Valouev et al., 2008)) was used to detect H3K27me3 peak in blastula embryos. WIG data generated by QuEST were also used for ChromHMM and for the visualization of other modifications. The following QuEST 2.4 settings were used for blastula H3K27me3: KDE (kernel density estimation) bandwidth = 100, ChIP seeding fold enrichment = 10, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 3. For other ChIP data visualization, WIG files were generated using following KDE bandwidth settings: H3K4me1: KDE bandwidth = 100; H3K4me2: KDE bandwidth = 60; H3K4me3: KDE bandwidth = 60; H3K27me3: KDE bandwidth = 100; H3K27Ac: KDE bandwidth = 60.

RNA-seq data processing and analysis

The BAM files of mapped reads were used for calculating the normalized RPKM with SAMMATE software (version 2.6.1 (Xu et al., 2011)).

Identification of HMDs

In this study, medaka genomic regions on scaffolds and ultracontigs were excluded from the analysis. We searched for contiguous regions of low methylated (methylation ratio lower than 0.4) CpG sites. ‘CpG site’ represents a single CpG in the genome. The two close low methylated regions were connected and considered as one domain when they were divided by less than four high methylated CpG sites (methylation ratio higher than 0.4). Regions with more than 9 low methylated CpG sites were defined as hyomethylated domains (HMDs). The boundaries of HMD were defined at the first low methylated CpG site inside the HMD regions. The position of TSS was defined according to the ensemble database (http://www.ensembl.org). Promoter region in this study was defined as regions from 5 kb upstream to 2 kb downstream of the TSS. The HMD overlapped with the promoter region was assigned to the gene, and used to classify genes. When more than two HMD overlapped with promoter region, the closest one to the TSS was selected. CpG island was defined as GC >= 50%, O/E >= 0.6 and length = 200 bp, according to
the previous study (Kasahara et al., 2007).

Identification of K27HMDs
K27HMDs were identified as the presence of H3K27me3 peaks inside the HMD, whereas nonK27HMDs were identified as the absence of H3K27me3 peaks according to QuEST analysis. For figures 2G, 3E and 6D, the number of mapped reads inside HMD of each category was counted and normalized by the length of HMD. For figures 2H, the highest peak signal intensity within the K27HMD calculated by QuEST was compared with low methylated CpG counts (the number of low methylated CpG sites).

Identification of H3K27me3 enriched hypomethylated regions by ChromHMM
We calculated K27HMD regions from methylation levels at individual CpG sites and H3K4me2/H3K27me3 ChIP-seq data. First, for each running window of 200bp in length, we computed the average of \(10(1 – x)\), where \(x\) was the methylation level at each CpG site, and the averages of H3K4me2/H3K27me3 ChIP-seq signal levels calculated by QuEST in the window. Subsequently, we input these epigenetic values into ChromHMM, which used the Poisson distribution to model the input data, and set a real value to 0 if its p-value < 0.0001, and to 1, otherwise. After this binarization, ChromHMM decomposed the input DNA sequence into regions of six chromatin states using the Hidden Markov Model. Finally, we selected such regions that the methylation level was low while H3K27me3 ChIP-seq signal was high, and examined the overlap with K27HMD regions.

Identification of HMD with elevated DNA methylation in adult tissues
The HMDs in which methylation increased were identified as >5 % of low methylated CpGs inside the HMD at blastula became high methylated (> 0.4) in adult myotome (dorsal) or liver. For figure 4C, average CpG methylation ratio in 25 bp window was calculated.

Heatmap generation
For DNA methylation, average methylation in 250 bp window was calculated. The level of windows without CpG site was set to 1. For histone modifications, the numbers of mapped reads in 250 bp window were counted. Heatmap was visualized in log scale using Java TreeView software.

Motif analysis at HMD boundaries
For the identification of motifs around boundaries of large K27HMDs (> 4 kb), we analyzed the sequence of 200 bp (100 bp each for upstream and downstream) using MEME software. Discovered motifs were subjected to TOMTOM software and the CTCF motif was identified. Next we analyzed the genome-wide distribution of CTCF motifs using FIMO software (we set the threshold of p value 1.0E-5) and counted
based on the position from the center of HMD. All softwares used here were on MEME suite (http://meme.nbcr.net/meme/).

**Analysis of nucleosome positioning around CTCF motifs**

For analyzing the nucleosome positioning around CTCF motifs, we utilized the genome-wide nucleosome positioning data sets from the previous report (Sasaki et al., 2009). The nucleosome density around CTCF motifs located near HMD boundary (100 bp upstream to 400 bp downstream from the HMD boundary) was averaged.

**Gene Ontology (GO) analysis**

For the GO analysis, we selected the genes that have human orthologous genes because of the absence of a GO platform in medaka. Among 18,028 medaka ensemble genes (genes on scaffolds and ultracontigs were excluded), 13,301 genes have their single human ortholog. GO analysis was performed by DAVID program (Huang et al., 2009). All human orthologous genes (13,301) were used for background. PANTHER biological processes category was used.

**Sequence conservation analysis**

For the analysis of sequence conservation among vertebrates, the average PhastCons scores profiles around the HMD boundaries were generated with the Conservation Plot tool, part of the Cistrome Analysis pipeline (http://cistrome.dfci.harvard.edu/ap/). The identity of amino acid sequence between medaka and human proteins were downloaded from Ensembl database and compared among groups of DNA methylation state.

**Analysis of human ESCs data**

For the comparative analysis between medaka and human, ChIP-seq, and RNA-seq data sets were downloaded from previous reports (Lister et al., 2009; Lister et al., 2011) (GEO accession no.: GSE16256). Mapping to hg19 reference genome was performed using bowtie program (Langmead et al., 2009). For WGBS, we downloaded processed wig data from Lister et al., 2009, and converted to hg19 using UCSC LiftOver. Low methylated CpG was defined as < 0.6 in human. Other analyses were performed in the same way as medaka data sets. The following QuEST settings were used for H3K27me3 peak detection and visualization; H3K27me3: KDE bandwidth = 100, ChIP seeding fold enrichment = 20, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 3; H3K4me2: KDE bandwidth = 60.

**CTCF ChIP-seq peak distribution around HMD boundaries in human**

CTCF ChIP-seq peak position file was downloaded from the UCSC genome browser (The ENCODE
Project Consortium 2011; GSM733672). The relative position to the HMD boundary was counted.

**Analysis of zebrafish data**
Methylome and RNA-seq data were downloaded from previous reports ((Potok et al., 2013) Accession number: SRP020008). For Bisulfite data, only one end of the paired-end reads was used in this study. Analyses were performed in the same way as medaka data sets.

**Statistical analysis and the data visualization**
The statistical analysis and graph visualization were performed using R software (version 2.14.2). For the visualization of genome-wide data, we integrated the data into UTGB genome browser (Saito et al., 2009).

**Supplemental References**

