



Published in final edited form as:

Nat Genet. 2016 April ; 48(4): 417–426. doi:10.1038/ng.3522.

Active DNA demethylation at enhancers during the vertebrate phylotypic period

Ozren Bogdanovi^{1,2,3}, Arne H Smits⁴, Elisa de la Calle Mustienes², Juan J Tena², Ethan Ford^{1,3}, Ruth Williams⁵, Upeka Senanayake⁵, Matthew D Schultz⁶, Saartje Hontelez⁷, Ila van Kruijsbergen⁷, Teresa Rayon⁸, Felix Gnerlich⁹, Thomas Carell⁹, Gert Jan C Veenstra⁷, Miguel Manzanares⁸, Tatjana Sauka-Spengler⁵, Joseph R Ecker^{6,10}, Michiel Vermeulen⁴, José Luis Gómez-Skarmeta², and Ryan Lister^{1,3}

¹Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Perth, Western Australia, Australia ²Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas, Universidad Pablo de Olavide, Seville, Spain ³Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia ⁴Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, the Netherlands ⁵Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK ⁶Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, California, USA ⁷Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, the Netherlands ⁸Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain ⁹Centre for Integrated Protein Science, Ludwig Maximilians Universität München, Munich, Germany ¹⁰Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, California, USA

Abstract

The vertebrate body plan and organs are shaped during a conserved embryonic phase called the phylotypic stage. However, the mechanisms that guide the epigenome through this transition and their evolutionary conservation remain elusive. Here we report widespread DNA demethylation of enhancers during the phylotypic period in zebrafish, *Xenopus tropicalis* and mouse. These enhancers are linked to developmental genes that display coordinated transcriptional and epigenomic changes in the diverse vertebrates during embryogenesis. Binding of Tet proteins to (hydroxy)methylated DNA and enrichment of 5-hydroxymethylcytosine in these regions

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

Correspondence should be addressed to O.B. (ozren.bogdanovich@uwa.edu.au), M.V. (michiel.vermeulen@science.ru.nl), J.L.G.-S. (jlgomska@upo.es) or R.L. (ryan.lister@uwa.edu.au).

Accession codes. MethylC-seq, TAB-seq, ATAC-seq and RNA-seq data were deposited at the Gene Expression Omnibus, GSE68087; quantitative interaction proteomics data were deposited at ProteomeXchange, PXD001164.

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

Author Contributions: O.B., M.V., J.L.G.-S. and R.L. designed the study. O.B. and E.F. prepared and sequenced MethylC-seq libraries. The data were analyzed by O.B. with the help of R.L., E.F., M.D.S. and J.R.E. Embryo work was performed by O.B., E.d.I.C.M., J.J.T., T.R., M.M. and J.L.G.-S. The zebrafish *sox10* line was prepared by R.W., U.S. and T.S.-S. *Xenopus* ChIP-seq data were generated by S.H., I.v.K. and G.J.C.V. Quantitative interaction proteomics experiments were performed by A.H.S., F.G., T.C. and M.V. Proteomics data were analyzed by A.H.S. and M.V. The manuscript was written by O.B., A.H.S., M.V., J.L.G.-S. and R.L.

Competing Financial Interests: The authors declare no competing financial interests.

implicated active DNA demethylation in this process. Furthermore, loss of function of Tet1, Tet2 and Tet3 in zebrafish reduced chromatin accessibility and increased methylation levels specifically at these enhancers, indicative of DNA methylation being an upstream regulator of phylotypic enhancer function. Overall, our study highlights a regulatory module associated with the most conserved phase of vertebrate embryogenesis and suggests an ancient developmental role for Tet dioxygenases.

Methylation of cytosine residues in genomic DNA is a major, mostly repressive epigenomic modification associated with key biological processes¹. Early vertebrate embryos display large-scale 5-methylcytosine (5mC) dynamics associated with the establishment of totipotency^{2–7}. Interestingly, these 5mC remodeling events occur in a species-specific fashion. Mammalian embryos seem to employ a combination of active and passive mechanisms to remodel their DNA methylomes after fertilization⁸. The active mechanism consists of Tet-dependent oxidation of 5mC that involves the 5-hydroxymethylcytosine (5hmC) intermediate, whereas the passive mechanism is based on 5mC dilution through cell divisions in the absence of methylation maintenance⁹. In contrast, in zebrafish, no Tet activity was detected during pluripotency and 5mC dynamics in the early embryo are reduced to passive reconfiguration of the maternal methylome to match the sperm methylation pattern^{6,7}. Similarly, the timing of zygotic genome activation can vary greatly between species. For example, in mouse, the first major wave of zygotic transcription occurs as early as the 2-cell stage (at 2 d), whereas in *Xenopus tropicalis* and zebrafish the embryonic transcription peaks at 6 h (~4,000 cells) and 4.3 h (64 cells), respectively¹⁰. Another major difference between mammals and anamniotes is the process of implantation and intrauterine embryonic development present in mammals. Notwithstanding these developmental differences, both groups have highly similar morphologies and gene expression patterns during the phylotypic stage, a developmental period associated with body plan and organ formation^{11–13}. However, very little is currently known about the mechanisms and evolutionary conservation of epigenetic patterning during these embryonic stages.

Here we report widespread DNA demethylation of thousands of enhancers associated with conserved regulatory pathways during the phylotypic period in zebrafish, *Xenopus* and mouse. Through whole-genome bisulfite sequencing (WGBS)^{14–16}, quantitative interaction proteomics¹⁷ and loss-of-function approaches, we found that this widespread demethylation event is Tet dependent and required for vertebrate body plan and organ formation. Finally, interrogation of chromatin accessibility (ATAC-seq) and whole-genome methylome profiling of *tet1-tet2-tet3* morphant zebrafish embryos demonstrated an upstream regulatory role for DNA methylation on these conserved genomic elements.

These findings have major implications for the understanding of fundamental processes that guide embryonic development. By unraveling the dependence of key developmental pathways on Tet-dependent demethylation of distal regulatory elements during the vertebrate phylotypic period, we shed light on a previously unknown developmental role for Tet proteins associated with the most conserved phase of vertebrate embryogenesis.

Results

Highly conserved phylotypic epigenome reconfiguration

To select equivalent embryonic stages of vertebrate embryos spanning the phylotypic period, we used as guidelines the interspecies relationships deduced from reciprocal best transcriptome similarity¹² (Fig. 1a). The phylotypic stage in zebrafish and *Xenopus* corresponds to 24 hours post-fertilization (h.p.f.) and stage (st.) 30, respectively, whereas in mouse this stage coincides with embryonic day (E) 9.5 (Fig. 1a, hourglasses). We generated WGBS DNA methylome profiles for four stages of zebrafish, *Xenopus* and mouse embryogenesis corresponding to the blastula (1,000-cell stage⁶, st. 9, E3.5 inner cell mass (ICM)⁸), gastrula (80% epiboly, st. 12.5, E7.5 (ref. 8)), pharyngula (24 h.p.f., st. 30, E9.5) and tailbud/fetus (48 h.p.f., st. 43, E14.5) stages (for details on non-conversion rates and sequencing metrics, see Supplementary Table 1). Overall, both the zebrafish and *Xenopus* genomes were globally highly methylated during these developmental stages, with average 5mC levels of 85% and 91%, respectively (Fig. 1a). Unlike in zebrafish and *Xenopus*, mammalian embryos, including mouse embryos, display extensive alterations in global 5mC levels during early embryonic stages (Fig. 1a), as previously described^{2,5,8,18–20}

To identify genomic regions displaying developmental changes in 5mC levels, we searched the DNA methylome profiles for differentially methylated regions (DMRs; false discovery rate (FDR) = 0.05, minimum change in the fraction of methylated CpG sites (Δ mCG) = 0.2), identifying thousands of localized changes in methylation state that occur during embryonic development in each species²¹ (Fig. 1b, Supplementary Fig. 1a,b and Supplementary Tables 2–20). The identified DMRs displayed a median Δ mCG of 0.33–0.53 (zebrafish), 0.29–0.44 (*Xenopus*) and 0.27–0.53 (mouse) (Supplementary figs. 2–4). In zebrafish embryos, whereas the transition from the 1,000-cell stage (blastula) to the stage with 80% epiboly (gastrula) was characterized by both developmental hyper- and hypomethylation (Fig. 1b and Supplementary Fig. 2a,b), the transitions associated with the phylotypic stage were overwhelmingly characterized by developmental hypomethylation (median Δ mCG = 0.33–0.44; Fig. 1b and Supplementary Fig. 2). A similar trend could be observed in *Xenopus* embryos, where almost no DMRs were identified between the blastula and gastrula stages (Supplementary Fig. 1a) and 5mC remodeling was mostly limited to demethylation (Δ mCG = 0.29–0.44) surrounding the phylotypic stage (Fig. 1b and Supplementary Fig. 3a,b). Similarly, the majority of mouse DMRs found between the late gastrula (E7.5) and tailbud (E14.5) stages were being developmentally demethylated (Δ mCG = 0.38–0.43; Fig. 1b and Supplementary Fig. 4).

To explore whether the DMRs associated with developmental demethylation surrounding the phylotypic stage are implicated in similar processes in the three species examined, we searched for enriched and conserved ontology terms associated with phylo(–)DMR-linked genes²². We identified 63 conserved (shared by all species) and significantly enriched (FDR q < 0.01) terms for early phylotypic DMRs (late gastrula–phylotypic stage demethylation) and 70 such terms for late phylotypic DMRs (phylotypic stage–tailbud/fetus demethylation) (Fig. 1c and Supplementary Tables 21 and 22). The enriched ontologies included terms such as ‘organ development’, ‘pattern specification process’, ‘tissue development’ and

‘anatomical structure development’ (Fig. 1c). Of note, no conserved ontology terms were found for the DMRs identified for the transitions between the other developmental stages we assessed: zebrafish blastula-gastrula hypo- or hypermethylated DMRs did not yield any significant Gene Ontology (GO) categories (FDR $q < 0.05$), whereas almost no DMRs except for the ones associated with phylotypic-stage transitions were identified in *Xenopus*. Early mouse DMRs that became hypomethylated between E3.5 and E7.5 as well as later-stage DMRs that became hypermethylated between E9.5 and E14.5 were enriched in GO categories associated with embryonic development; however, these stages yielded far fewer significant categories than the ones associated with the phylotypic transitions (Supplementary Figs. 5 and 6). Given their similar 5mC dynamics and implication in various developmental processes, henceforth the set of DMRs associated with developmental hypomethylation surrounding the phylotypic stage will be referred to as phylo(–)DMRs.

To address the potential influence of tissue heterogeneity in later-stage embryos and to recapitulate our analysis in a specific cell lineage, we generated a DNA methylome for a neural crest cell population (*sox10*⁺)²³ isolated from 24 h.p.f. zebrafish embryos. This analysis showed a decrease in phylo(–)DMR 5mC levels in *sox10*⁺ cells when compared to blastula (1,000-cell stage) (Fig. 1d). A further loss of 5mC was observed in the adult zebrafish brain, which had 5mC levels comparable to those in *Xenopus* and mouse brains (Fig. 1d). Furthermore, the extent of hypomethylation in phylo(–)DMRs was similar in *sox10*⁺ cells and 24 h.p.f. embryos for both early and late phylo(–)DMRs (Supplementary Fig. 7a). To explore whether these intermediate 5mC levels observed in *sox10*⁺ cells and 24 h.p.f. embryos are a result of different cell populations with either fully methylated or fully unmethylated sequences or of a uniform, partially methylated population, we plotted average mCG levels in individual sequencing reads for phylo(–)DMRs (Supplementary Figs. 7b–d and 8). This analysis indicated that the majority of sequenced reads were either fully unmethylated or fully methylated, thereby supporting the hypothesis of two different methylation states. A large proportion of the reads were either partially or fully unmethylated, indicating that demethylation was likely occurring in the majority of the cells in the embryo. In mouse, however, a substantial proportion of the reads were found in a partially methylated state (Supplementary Fig. 7b), a finding that could be attributed either to demethylation intermediates or the formation of low-methylated regions displaying intermediate 5mC levels, as previously described in mammalian cell lines²⁴. To explore whether phylo(–)DMRs only occur in certain lineages or whether they are found more widely throughout the organism, we analyzed whole-genome DNA methylome profiles corresponding to a number of adult tissues derived from the three embryonic layers (ectoderm, endoderm and mesoderm)²⁵. Average 5mC profiles exhibited hypomethylation at both early and late phylo(–)DMRs across all lineages and in 16 different organs, indicative of widespread phylo(–)DMR usage during organ formation (Fig. 1e).

Overall, our base-resolution DNA methylome profiles suggest a highly conserved process of 5mC reconfiguration that takes place throughout the phylotypic period in diverse vertebrate species, involves developmental hypomethylation and, at least in mouse, affects organs derived from all germ layers.

Phylo(–)DMRs are developmentally activated enhancers

To explore the chromatin configuration and genomic context of phylo(–)DMRs, we used chromatin immunoprecipitation and sequencing (ChIP-seq) data for chromatin marks associated with promoters (trimethylation of histone H3 at lysine 4, H3K4me3), poised enhancers (monomethylation of histone H3 at lysine 4, H3K4me1) and active enhancers (acetylation of histone H3 at lysine 27 (H3K27ac) and p300)^{26–28}. Sorted heat maps of phylo(–)DMRs in zebrafish and *Xenopus* showed a strong developmental enrichment for enhancer but not promoter histone marks (Fig. 2a). Similarly, the mouse phylo(–)DMRs were strongly enriched for H3K4me1 and H3K27ac marks and were enriched to a lesser extent for H3K4me3 marks, indicative of active enhancers^{29–31}. Notably, the other zebrafish DMR groups did not display such a strong enrichment in enhancer chromatin (Supplementary Fig. 9a). However, some of the identified DMRs that became developmentally hypermethylated (DMRs hypomethylated in the 1,000-cell stage and hypermethylated in the epiboly stage and DMRs hypomethylated in the epiboly stage and hypermethylated at 24 h.p.f.) overlapped the promoter regions of genes such as *sostdc1*, *ddx4*, *dazl* and *dnmt3bb.2*, some of which were previously identified as differentially methylated between early embryos and differentiated tissues^{6,7} (Supplementary Fig. 9b). Inspection of the CpG density of phylo(–)DMRs, as measured by the mean CpG levels within and flanking phylo(–)DMRs, indicated similar CpG densities in zebrafish and mouse DMRs, whereas lower CpG density was observed in *Xenopus* DMRs (Fig. 2b). Whereas the CpG densities of phylo(–)DMRs in zebrafish and mouse were similar to that of CpG islands identified through pulldown of unmethylated DNA^{32,33}, *Xenopus* phylo(–)DMRs, and in particular the early *Xenopus* phylo(–)DMRs, displayed considerably lower CpG density than these CpG islands (Fig. 2c and Supplementary Fig. 10a). These results identify phylo(–)DMRs as developmentally activated enhancers of high CpG density in zebrafish and mouse and low and intermediate CpG density in *Xenopus*, and they support previous notions that CpG density alone is not a major driver of regulatory function^{34–36}.

To provide further evidence that phylo(–)DMRs act as developmental enhancers, we intersected the mouse phylo(–)DMR genomic positions with the positions of previously validated enhancers from the VISTA enhancer browser database^{37,38} and obtained 13 intersections corresponding to active enhancers (Supplementary Table 23). Examples of such heart and limb enhancers and their intersection with phylo(–)DMRs are presented in Figure 2d. Intersection of zebrafish phylo(–)DMRs with VISTA enhancers resulted in 36 intersections. These regions displayed a mild degree of developmental demethylation; however, none of them overlapped with statistically significant mouse DMRs (Supplementary Fig. 10b,c and Supplementary Table 23).

Given the high conservation of GO term enrichments (Fig. 1c) and highly similar chromatin configurations of phylo(–)DMRs (Fig. 2a), we postulated that phylo(–)DMR-linked genes should be co-regulated during zebrafish, *Xenopus* and mouse embryogenesis. To investigate this hypothesis, we identified orthologous genes that were linked to phylo(–)DMRs in all the species examined (Supplementary Table 24) and subjected them to pathway enrichment analyses^{39,40}. These analyses showed that orthologous phylo(–)DMR-linked genes were enriched in pathways such as Wnt, Notch and transforming growth factor (TGF)- β ,

implicated in body plan and organ formation (Supplementary Fig. 11). To our knowledge, this is the first indication that these key developmental pathways may be regulated through DNA methylation in multiple vertebrate species. Next, we compared the developmental expression profiles of these orthologous phylo(–)DMR-linked genes in zebrafish, *Xenopus* and mouse at time points corresponding to the blastula (1,000-cell stage, st. 9, blastocyst), late gastrula (bud, st. 12, E8.5) and phylotypic (28 h.p.f., st. 30, E9.5) stages^{41–44}. Hierarchical clustering analysis of scaled RNA sequencing (RNA-seq) data (in transcripts per million, TPM; Fig. 2e) indicated a strong developmental correlation of the orthologous genes, providing further support for phylo(–)DMRs being involved in conserved regulatory networks. Finally, assessment of the evolutionary conservation of zebrafish and mouse phylo(–)DMRs by mapping aggregate sequence conservation scores⁴⁵ demonstrated that phylo(–)DMRs display higher evolutionary conservation than early (blastula or gastrula) or late (adult organ) DMRs (Fig. 2f), consistent with the previous observation of higher evolutionary conservation of putative regulatory regions during similar embryonic stages^{46–48}. Together, these findings support a role for phylotypic enhancer demethylation in the activation and deployment of the pan-vertebrate developmental toolkit necessary for body plan formation and organ specification.

Active demethylation in vertebrate embryos

Several DNA demethylation pathways have been described in zebrafish and *Xenopus laevis*^{49–51}. To obtain better insight into which cellular factors might be implicated in phylo(–)DMR demethylation, we performed a quantitative interaction proteomics screen using 5mC-modified, 5hmC-modified and unmodified DNA oligonucleotide probes¹⁷ and nuclear extracts from zebrafish embryos, to identify developmentally regulated 5mC readers in the dome and 24 h.p.f. stages. We also profiled binding to 5hmC, as this active demethylation intermediate was previously observed in 24 h.p.f. zebrafish embryos^{51,52}. Proteins displaying differential binding to unmethylated and methylated DNA oligonucleotides were identified by an adapted *t* test on label-free quantification (LFQ) intensities¹⁷ (Supplementary Figs. 12 and 13, and Supplementary Tables 25 and 26). We also performed absolute quantification of the nuclear extract proteomes (Supplementary Fig. 12) to directly assess the effects of stage-specific protein abundance on binding enrichments. In total, we identified 98 context-specific 5mC readers, the majority of which displayed stage-specific binding profiles. Of the 45 readers quantified, the differential binding of only 19 could be explained by different protein abundance (Supplementary Fig. 13), thereby demonstrating the dynamics of the 5mC interactome during early vertebrate embryogenesis.

To investigate the evolutionary conservation of 5(h)mC readers in vertebrates, we compared our set of zebrafish readers to the readers previously identified in mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs)¹⁷. In total, 96 proteins for which orthologs are annotated in both species were identified as either 5mC or cytosine readers in zebrafish dome or mouse ESC nuclear extracts. Forty-five of these 96 proteins showed conserved binding (47% $P < 1.1 \times 10^{-45}$, hypergeometric test) (Fig. 3a,b and Supplementary Fig. 6). Readers for cytosine, 5mC and 5hmC in mouse NPCs and 24 h.p.f. zebrafish embryos also displayed substantial overlap: 99 of 249 proteins (40% $P < 5 \times 10^{-37}$, hypergeometric test) for which orthologs are annotated in both species displayed conserved binding (Fig. 3a,b and

Supplementary Fig. 14). Notably, the 5hmC readers were highly enriched for proteins related to DNA repair and DNA demethylation, including Tet proteins and Uhrf2 (Fig. 3b,c), a protein associated with increased Tet processivity in mouse NPCs¹⁷. The majority of proteins involved in DNA repair and demethylation showed a clear trend toward a higher binding affinity for 5mC in the differentiated states and strongest binding to 5hmC (Fig. 3c, bottom row). Together, these results suggest an enrichment of Tet-dependent DNA demethylation pathways occurring during the vertebrate phylotypic period.

The observed enrichment for Tet and DNA repair protein binding to 5mC and 5hmC in zebrafish phylotypic-stage embryos and mouse NPCs suggests that phylo(–)DMRs may be demethylated through a Tet-dependent mechanism. Although the diverse vertebrate species displayed different Tet expression profiles during development, for all species the phylotypic period was characterized by the expression of at least one Tet family member (Fig. 4a). To assess whether phylo(–)DMRs become demethylated through an active mechanism that involves the 5hmC intermediate, we generated base-resolution maps of 5hmC in zebrafish (24 h.p.f.), *Xenopus* (st. 30) and mouse (E9.5) embryos by Tet-assisted bisulfite sequencing (TAB-seq)¹⁶ (for details on non-conversion and TAB-seq metrics, see Supplementary Table 1). Strong and highly localized enrichment of 5hmC was detected in regions marked by active enhancer chromatin and corresponding to phylo(–)DMRs (Fig. 4b). Similarly, mean 5hmC levels in all species were highest in phylo(–)DMRs as compared to other regions that became developmentally hypomethylated in all organisms examined (Fig. 4c). Accordingly, the 5hmC signal in zebrafish and *Xenopus* coincided with developmental demethylation taking place in both early and late phylo(–) DMRs (Fig. 4d and Supplementary Fig. 15). A similar pattern was observed for mouse late phylo(–) DMRs (Fig. 4d and Supplementary Fig. 15), whereas mouse early phylo(–)DMRs were already demethylated by E9.5 (Supplementary Fig. 4) and therefore did not show any localized 5hmC enrichment (Fig. 4d). To explore the developmental correlations between 5hmC and 5mC levels and interrogate whether the quantity of 5hmC is directly linked to phylo(–)DMR demethylation, we plotted the average 5hmC abundance in the phylotypic stage for each CpG dinucleotide found within phylo(–)DMRs against its 5mC level in the subsequent developmental stage (Fig. 4e). We observed an overall negative correlation between these two values, indicating that the quantity of 5hmC signal can serve as a predictor of active demethylation in vertebrate embryos. To further investigate the potential role of active DNA demethylation mechanisms in phylo(–)DMR formation, we used Tet1 ChIP-seq and 5hmC (hMeDIP-seq) data in mouse ESCs^{53,54}. When superimposed over mouse early and late phylo(–)DMRs, both Tet1 and 5hmC displayed a strong enrichment over these regions in mouse ESCs (Supplementary Fig. 16). Furthermore, a positive correlation between the quantity of 5hmC and Tet1 signal, as well as between Tet1 signals from different antibodies, was evident for both early and late phylo(–)DMRs (Supplementary Fig. 16b), thereby supporting the notion of Tet-dependent 5mC-5hmC conversion in phylo(–)DMRs during mammalian embryogenesis. Altogether, these data provide evidence for active DNA demethylation in phylo(–)DMR formation during the vertebrate phylotypic period.

***tet1-tet2-tet3* knockdown reduces phylo(–)DMR accessibility**

To assess the role of Tet proteins in mediating demethylation of phylo(–)DMRs, we inhibited Tet function in zebrafish embryos, first using the morpholino knockdown approach to target Tet3, a major phylotypic 5mC and 5hmC interactor (Fig. 3c). Embryos injected with *tet3* morpholino showed only minor defects, with 23% of the morphants displaying varying degrees of microphthalmia (Supplementary Fig. 17a), as previously reported in a *tet3* morpholino study performed in *Xenopus laevis* embryos⁵⁵. Next, we injected zebrafish embryos with morpholinos targeting all three Tet proteins to generate *tet1-tet2-tet3* morphants⁵⁶. The triple morphants were severely affected, with the majority displaying embryonic lethality, while the embryos that survived gastrulation (23%) displayed short and blended axes, impaired head structures, small eyes and reduced pigmentation (Supplementary Fig. 17a), comparable to the defects recently described in a *tet1-tet2-tet3* zebrafish knockout⁵⁷. We next performed WGBS (for details, see Supplementary Table 1) on morphant embryos ($n = 50$) and compared the 5mC profiles to their counterparts in wildtype embryos. Global mCG levels (in 1-kb non-overlapping windows) were similar in *tet3* and *tet1-tet2-tet3* morphants when compared to wild-type embryos and were highly correlated with mCG levels in the wild-type embryos (Fig. 5a). A similar pattern was observed in CpG islands, key regulatory features of vertebrate genomes associated with gene regulation, identified through CXXC profiling³³. However, comparisons of the *tet1-tet2-tet3* morphants and wild-type embryos showed an increase in the mCG levels of phylo(–)DMRs in the morphants, affecting almost all (93.5%) of these regions. No differences in 5mC increase were observed between early and late phylo(–)DMRs, and no other DMR population displaying developmental demethylation was affected by this perturbation (Supplementary Fig. 17b,c). Also, there was an increase in mCG levels in a subpopulation of CpG islands that corresponded to phylo(–)DMRs and, thus, with sites of aberrant demethylation in the *tet1-tet2-tet3* morphants (Fig. 5a and Supplementary Fig. 17d,e). Taken together, these data highlight an embryonic requirement for Tet proteins and demonstrate abnormalities in 5mC remodeling at key regulatory elements caused by the absence of these proteins.

Next, we wanted to determine whether the loss of Tet proteins and increase in phylo(–)DMR 5mC levels would result in decreased chromatin accessibility of these regulatory elements. We therefore performed ATAC-seq⁵⁸ on two pools of *tet1-tet2-tet3* morphant embryos and their wild-type counterparts (for ATAC-seq metrics, see Supplementary Table 1). Both wild-type and morphant embryos displayed ATAC-seq signal enrichment at phylo(–)DMRs; however, this enrichment was significantly (Wilcoxon test, $P < 0.001$) decreased in *tet1-tet2-tet3* morphant embryos (Fig. 5b,c). Notably, such a change was not observed in a general population of putative distal regulatory elements (PDREs) identified in zebrafish embryos²⁶. Next, we identified phylo(–)DMRs displaying a difference (Fisher's exact test, FDR q 0.05) in ATAC-seq signal between wild-type and morphant embryos. This population consisted of ~35% of the phylo(–)DMRs and was characterized by reduced ATAC-seq signal in >90% of the phylo(–)DMRs, consistent with the average ATAC-seq profiles (Fig. 5d). Finally, to address the impact of Tet1-Tet2-Tet3 loss on embryonic transcription, RNA-seq analysis was performed on two pools of *tet1-tet2-tet3* morphants and their wild-type controls. Of note, unsupervised clustering of the zebrafish embryonic transcriptomes identified a cluster encompassing the *tet1-tet2-tet3* morphant samples, their wild-type

controls and other zebrafish phylotypic-stage samples⁴¹, indicative of morpholino phenotype specificity rather than a developmental delay caused by *tet1-tet2-tet3* knockdown (Supplementary Fig. 18). Differential gene expression analyses⁵⁹ showed bidirectional changes in gene transcript abundance in *tet1-tet2-tet3* morphants ($n = 718$ upregulated genes and $n = 780$ downregulated genes in the morphant) (Fig. 5e and Supplementary Table 27). GO analyses^{39,40} of downregulated genes showed an over-representation of functions associated with transcriptional regulation, similar to those described to be enriched among zebrafish phylo(-)DMR-associated genes (Fig. 5f and Supplementary Fig. 5), whereas no statistically significant gene ontology enrichments were associated with the upregulated group of genes. An example of aberrant phylo(-)DMR demethylation in the *tet1-tet2-tet3* morphant resulting in reduced chromatin accessibility and reduced transcription is depicted in Figure 5g. Taken together, these data link Tet-dependent demethylation with the proper activation of phylotypic-stage enhancers and suggest a regulatory role for 5mC in phylo(-)DMR usage.

Discussion

It is well established that vertebrates remodel their epigenomes during embryogenesis to achieve totipotency^{2–10,60,61}. However, thus far, only very limited insights have been obtained regarding 5mC dynamics and the evolutionary conservation of 5mC patterning in later-stage vertebrate embryos. Here we describe widespread enhancer demethylation coinciding with the phylotypic stage of zebrafish, *Xenopus* and mouse embryogenesis. This demethylation activity is almost exclusively targeted to a subset of embryonic enhancers, phylo(-)DMRs, found in the vicinity of genes that have conserved roles in the establishment of the vertebrate body plan, including key developmental pathways such as Notch-Delta, Wnt and TGF- β . These observations are in line with a recent study that identified distal regulatory elements as targets of 5mC remodeling in zebrafish embryos⁶². Furthermore, we demonstrate that phylo(-)DMRs display low methylation levels in all explored adult organs in mice, indicative of their widespread usage in all embryonic lineages, and that Tet proteins and 5mC-5hmC conversion are required for their demethylation during the phylotypic period in zebrafish. Notwithstanding the implications of active demethylation pathways in this widespread epigenomic reconfiguration event, we do not exclude the possibility that transcription factor binding participates in the demethylation of these genomic regions, as previously described for low-methylated regions in mouse cell cultures²⁴. In fact, a recent study demonstrated Tet3 targeting through transcription factor binding in the mammalian neural lineage⁶³, whereas Tet1 was shown to associate with the transcription factor Tex10 to regulate chromatin conformation at super-enhancers in ESCs⁶⁴.

Mammals display complex embryonic requirements for Tet activity; the *Tet3*^{-/-} knockout mouse shows early embryonic lethality⁶⁵, whereas the *Tet1*^{-/-} (ref. 66) and *Tet2*^{-/-} (refs. 67,68) knockout mice are viable. The double *Tet1*^{-/-}; *Tet2*^{-/-} knockout is perinatal lethal in the majority of embryos; however, a small percentage of *Tet1*^{-/-}; *Tet2*^{-/-} knockout mice can successfully be grown to adulthood⁶⁹. In contrast, anamniotes such as zebrafish and *Xenopus* do not express Tet proteins during early embryonic stages, and only very weak 5hmC signal was detected in pluripotent zebrafish embryos^{6,7}. Nonetheless, here we demonstrate that both anamniotes and mammals employ active demethylation of enhancers

for gene regulation during the phylotypic period, suggestive of an ancient role of the Tet dioxygenases and a pan-vertebrate regulatory logic. Finally, this work provides insights into the roles that 5mC has in embryonic enhancer elements and implicates 5mC as an upstream regulator of phylotypic enhancer function. Indeed, the marked reduction in phylo(–)DMR demethylation and chromatin accessibility upon *tet1-tet2-tet3* knockdown indicates that modulation of the epigenetic state is a critical regulator of the function of these elements. Altogether, our study highlights a highly conserved mechanism used by vertebrates during the specification of the enhancer repertoire needed for body plan and organ formation. This sets the foundation for future studies that will aim to address the precise hierarchical relationships between Tet-dependent demethylation, enhancer activation and transcription factor binding and to determine the embryonic requirements for each of these processes.

URLs

Database for Annotation, Visualization and Integrated Discovery (DAVID), <https://david.ncifcrf.gov/home.jsp>; GREAT: Genomic Regions Enrichment of Annotations Tool, <http://bejerano.stanford.edu/great/public/html>; FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit/index.html; R Project for Statistical Computing, www.r-project.org/; Kallisto, <http://pachterlab.github.io/kallisto/>; methylpy pipeline, <https://bitbucket.org/schultzmatt/methylpy>; Ensembl Biomart, <http://www.ensembl.org/biomart>; ProteomeXchange Consortium, <http://proteomecentral.proteomexchange.org/>; BEDTools suite, <http://bedtools.readthedocs.org/en/latest>.

Online Methods

DNA and RNA isolation

For DNA extraction, zebrafish, *Xenopus* and mouse embryos were lysed in buffer containing 20 mM Tris, pH 8.0, 100 mM NaCl, 15 mM EDTA, 1% SDS and 0.5 mg/ml proteinase K for 3 h at 55 °C. Lysis was followed by two phenol:chloroform:isoamyl alcohol (25:24:1) extractions and subsequent centrifugation (5 min at 17,949g). DNA was then precipitated by adding 0.2 volumes of 4 M ammonium acetate and 3 volumes of 96% ethanol. The reaction was left on ice for a minimum of 30 min. The DNA precipitate was centrifuged for 20 min at 4 °C (17,949g), and the pellet was washed with 500 µl of 70% ethanol and centrifuged for 5 min (17,949g) at room temperature. The pellet was then resuspended in 200 µl of TE buffer, and 1 µl of RNase A (20 µg/µl) was added. The reaction was left to proceed for 30 min at room temperature, after which time the DNA was precipitated with 0.1 volumes of 4 M ammonium acetate and 1 volume of isopropanol on ice for 2 h. The reaction was then centrifuged for 30 min at 4 °C (17,949g), and the pellet was resuspended in TE buffer. RNA extractions were performed with Qiagen RNeasy kits according to the manufacturer's recommendations.

MethylC-seq

For MethylC-seq library generation, genomic DNA was sonicated to an average size of 200 bp using the Covaris sonicator. Sonicated DNA was then purified and end repaired, followed by ligation to methylated Illumina TruSeq sequencing adaptors. Library amplification was performed with KAPA HiFi HotStart Uracil+ DNA polymerase (Kapa Biosystems), using

six cycles of amplification. Single-read MethylC-seq libraries (for details, see Supplementary Table 1) were sequenced on the Illumina HiSeq 1500 platform. The sequenced reads in FASTQ format were mapped to *in silico* bisulfite-converted reference genomes (danRer7, JGI.71 or mm10 for zebrafish, *Xenopus* and mouse, respectively) using the Bowtie alignment algorithm with the following parameters: -e 120 -l 20 -n 0 (ref. 70), as previously reported⁷¹. Published paired-read MethylC-seq data⁶⁻⁸ were mapped with the following parameters: -e 120 -l 20 -n 1 -k 10 -o 4 -I 0 -X 1000. To estimate the bisulfite non-conversion frequency the frequency of all cytosine base calls at reference cytosine positions in the lambda phage genome (unmethylated spike-in control) was normalized by the total number of base calls at reference cytosine positions in the lambda phage genome (Supplementary Table 1).

TAB-seq

TAB-seq library generation was performed with the 5hmC TAB-seq kit (WiseGene, K001) according to the manufacturer's instructions, with minor modifications related to the amount of starting material (zebrafish 24 h.p.f., 400 ng; *Xenopus* st. 30, 300 ng; mouse E9.5, 100 ng). 5-hydroxymethylated pUC19 DNA (WiseGene, S002) was used as the 5hmC standard for the estimation of β -glucose protection of 5hmC from Tet conversion, whereas lambda phage DNA with methylation of all cytosines at CpG sites (WiseGene, S001) was used as the 5mC/cytosine spike-in control. Single-read TAB-seq libraries (for details, see Supplementary Table 1) were sequenced on the Illumina HiSeq 1500 platform. The sequenced reads in FASTQ format were mapped to the *in silico* bisulfite-converted reference genomes (danRer7, JGI.71 or mm10) using the Bowtie alignment algorithm with the following parameters: -e 120 -l 20 -n 1 (refs. 70,71).

ATAC-seq

For ATAC-seq library preparation⁵⁸, ten zebrafish embryos were manually dechorionated and disrupted in 500 μ l of Ginzburg Fish Ringers (55 mM NaCl, 1.8 mM KCl, 1.25 mM NaHCO₃). After washing with cold PBS, 75,000 cells were lysed (lysis buffer: 10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 0.1% Igepal) and incubated for 30 min at 37 °C with TDE1 enzyme. The sample was then purified with the Qiagen MinElute kit, and a PCR reaction was performed with 13 cycles using the Ad1F and Ad2.1R primers and KAPA HiFi HotStart enzyme (Kapa Biosystems). Reads were aligned using the zebrafish danRer7 (zv9) assembly as the reference genome. Duplicated pairs of reads or those separated by more than 2 kb were removed from the analyses. The enzyme cleavage site was determined as the position at -4 (minus strand) or +5 (plus strand) with respect to each read start, and this position was extended by 5 bp in both directions.

Identification of CpG differentially methylated regions

CpG DMRs were identified by the methylpy pipeline^{21,72}. First, a root-mean-square test on each CpG site across all samples was performed. *P* values were simulated using 10,000 permutations. The largest *P* value cutoff was chosen that still satisfied the requirement for an FDR of 0.05. Significantly differentially methylated sites were combined into blocks if they were within 500 bases of one another and had methylation changes in the same direction. Furthermore, blocks that contained fewer than ten differentially methylated sites were

discarded. Only DMRs displaying a minimum change in mCG fraction ($mCG = 0.2$) were used for the analyses, unless otherwise indicated. For zebrafish and *Xenopus*, all the samples (embryonic stages) were compared in the same DMR finding experiment to obtain the methylation status (hyper- or hypomethylated) of the DMRs across all stages. The same was done for mouse embryos at E7.5, E9.5 and E14.5, whereas early mouse DMRs (E3.5 with a globally hypomethylated genome and E7.5) were identified in a separate DMR finding experiment, using the same statistics as described above. The output of the methylpy pipeline is provided as Supplementary Tables 2–5.

DNA methylation and hydroxymethylation analyses

5mC and 5hmC heat maps—Heat maps of 5mC and 5hmC data represent mean methylated CpG levels in 100-bp windows (extended by 3 or 5 kb on each side from the center of the phylo(–)DMR) corrected for the CpG non-conversion rate and the protection rate from CpG hydroxymethylation (TAB-seq). Windows with no coverage were assigned the average value for the adjacent windows. In case of multiple adjacent windows with no coverage, the average genomic value for methylated or hydroxymethylated CpGs was assigned. Calculation of 5hmC levels was performed using the formula

$$p = (n - r) \frac{1}{(1 - r - s)}$$

where p is the corrected 5hmC value, n is the raw 5hmC value, r is the bisulfite sequencing non-conversion rate and s is the 5hmC non-protection rate.

5mC and 5hmC correlation in phylo(–)DMRs—The correlation of hydroxymethylated and methylated CpGs within phylo(–)DMRs was calculated in the following way: for all samples, only the methylated and hydroxymethylated CpGs that displayed a minimum coverage of 4 were used for the analyses. The values (number of methylated cytosines and total number of sequenced cytosines) originating from both DNA strands were merged into a single value to obtain average 5mC and 5hmC levels for each CpG dinucleotide. The fraction of 5hmC was calculated as the average 5hmC (hmCG/CG) level divided by the average 5mC (mCG/CG) level for each CpG dinucleotide in the phylotypic stage. These values were then plotted against the average 5mC level (mCG/CG) in the tailbud stage (zebrafish, 48 h.p.f; *Xenopus*, st. 43; mouse, E14.5), filtering out positions where the value was 0. The plots were generated in R using the smoothScatter and smooth.spline functions.

Comparisons of 5mC levels in wild-type and Tet morphants—Average CpG methylation levels (mCG/CG) were calculated for non-overlapping genomic windows (1 kb), zebrafish CpG islands and non-methylated islands³³, and phylo(–)DMRs. The scatterplots for CpG islands and genomic windows were plotted using the smoothScatter function in R whereas the levels of methylated CpGs in phylo(–)DMRs were plotted using the filled.contour function in R.

5mC levels of individual sequencing read—For this analysis, only phylo(–)DMRs displaying intermediate DNA methylation levels (0.4–0.6) were used. Sequencing reads

containing 5 CpGs and overlapping with these phylo(–)DMRs were used for the calculation. Per-read 5mC levels were calculated as mCG/CG for each individual sequencing read. Because of the lower CpG density of *Xenopus* phylo(–)DMRs, the same analysis was repeated in *Xenopus* with reads with 3 CpGs.

ATAC-seq data analyses

Nucleosome-free ATAC-seq reads in phylo(–)DMRs were counted using BEDTools version 2.0 (ref. 73). Normalized read densities were obtained by multiplying the number of nucleosome-free reads within a phylo(–)DMR by 10^9 and dividing that number by the total number of nucleosome-free reads in the experiment multiplied by phylo(–)DMR length (in bp). To assess the statistical significance of the difference between ATAC-seq signal distributions in *tet1-tet2-tet3* morphants and wild-type embryos in phylo(–)DMRs, the Wilcoxon test was used. As a control, the same calculation was performed on a collection of PDREs identified in zebrafish embryos through ChIP-seq profiling²⁶. Fisher's exact test (FDR $q = 0.05$, minimum fold change = 2) was used to identify phylo(–)DMRs with statistically significant differential abundance of ATAC-seq signal between *tet1-tet2-tet3* morphants and wild-type embryos.

ChIP-seq and Bio-CAP data analysis

Zebrafish, mouse and *Xenopus* CpG islands and non-methylated islands are part of a previously published data set³³ (GSE43512). Zebrafish CpG islands and non-methylated islands used in this study correspond to the embryonic 24 h.p.f. Bio-CAP sample (GSM1064697), whereas *Xenopus* CpG islands and non-methylated islands correspond to the embryonic st. 11–12 Bio-CAP sample (GSM1064693). For mouse, a merged collection of testes (GSM1064678), liver (GSM1064679) and ESC (GSM1064680) CpG islands and non-methylated islands was used. Zebrafish ChIP-seq, H3K27ac, H3K4me3 and H3K4me1 reads²⁶ were obtained from the GEO database (GSE32483). Mouse embryonic H3K4me1, H3K27ac and H3K4me3 mapped reads corresponding to fetal (E14.5) brain, liver, limb and heart²⁷ were obtained from GSE29184, merged and analyzed as one embryonic sample. *Xenopus* ChIP-seq data²⁸ were obtained from GSE67974. Tet1 ChIP-seq and hMeDIP mapped data^{53,54} used in this study correspond to the following identifiers: GSM611192, GSM611194, GSM611199 and GSM659799. Heat maps of ChIP-seq data represent sequencing depth–normalized read density for 100-bp windows extending to 5 kb on each side from the phylo(–)DMR center.

RNA-seq library preparation and data analysis

RNA libraries were prepared using the TruSeq Stranded mRNA Library Prep kit (Illumina) from 400 ng of total extracted RNA according to the manufacturer's protocol. For all libraries, ten cycles of PCR amplification were used. All RNA-seq data, including previously published zebrafish (GSE32900)⁴¹, mouse (GSM1502476 and GSM1545168)^{43,44} and *Xenopus* (GSE43652)⁴² data, were mapped using Kallisto⁷⁴ with default settings. Before mapping the mouse RNA-seq data, the reads were trimmed to 51 bp using the FASTX trimmer from the FASTX-toolkit to normalize for read length differences among the different studies. The reference transcriptomes were generated using canonical transcript isoforms as identified by Ensembl. TPM values were used to assess transcript

abundance. Differential gene expression analysis was performed using the DESeq Bioconductor package⁵⁹. Only genes with FDR $q < 0.05$ were considered significant.

Gene ontology analyses

Gene ontology analyses were performed using the GREAT tool (v3.0.0)²² or the DAVID tool^{39,40}, as indicated. For *Xenopus* GREAT analyses, we calculated the GREAT regions as described before²² using 'basal plus extension' association rule settings. Briefly, each gene was assigned a basal regulatory domain of a minimum distance upstream and downstream of the transcription start site (TSS; regardless of other nearby genes, 5 kb upstream and 1 kb downstream). The gene regulatory domain was extended in both directions to the nearest gene's basal domain but no more than the maximum extension in one direction. These GREAT regions were then intersected using BEDTools version 2.0 (ref. 73) with *Xenopus* phylo(-)DMRs to obtain phylo(-)DMR-associated genes. The TSSs of these identified genes (Ensembl gene models, human reference genome hg19) were uploaded to GREAT to calculate GO enrichments. For multiple-species comparisons, only GO terms displaying minimum FDR $q < 0.01$ for both binomial and hypergeometric tests were used for the analyses.

Expression analysis of orthologous genes

The TPM values for zebrafish, mouse and *Xenopus* phylo(-)DMR-linked genes were calculated using the Kallisto software⁷⁴ as described above and assigned to the orthologous genes identified through the Ensembl orthology tool⁷⁵. The highest expression value for each gene in each species was assigned a value of 1, and the TPM values for the other stages were scaled accordingly. These scaled expression values were clustered using the Ward method and Spearman correlation distance.

Evolutionary conservation of phylo(-)DMRs

Mean evolutionary conservation (phastCons) scores⁴⁵ were calculated for early (1,000-cell (ref. 6)–80% epiboly) DMRs, phylo(-)DMRs (combined 80% epiboly–24 h.p.f. developmental demethylation and 24–48 h.p.f. developmental demethylation) and late (adult muscle–adult brain)⁷ zebrafish (danRer7) DMRs using BEDTools version 2.0 (ref. 73). For mouse, vertebrate phastCons scores (mm10) were calculated for early (E3.5–E7.5; ref. 8) DMRs, phylo(-)DMRs (combined E7.5–E9.5 developmental demethylation and E9.5–E14.5 developmental demethylation) and late (adult liver–adult brain)²⁵ DMRs. DMRs with no phastCons score assigned were excluded from the analyses. The Kruskal-Wallis test (Dunn's post test) was applied to test for the statistical significance of differences in phastCons score distributions.

Morpholino knockdown of Tet proteins

The morpholinos specific for each Tet protein are described in ref. 56. We also designed a specific splice-junction morpholino for the zebrafish Tet3 protein (Supplementary Table 28). Either 12 ng of this morpholino or 3 ng each of a combination of all four morpholinos was injected into zebrafish embryos at the one-cell stage. The observed phenotypes were documented at different time points using a stereoscope (SZX16-DP71, Olympus).

Nuclear extracts from zebrafish embryos

Zebrafish embryos at the dome ($n = 100,000$) and 24 h.p.f. ($n = 25,000$) stages were collected and dechorionated using pronase (Sigma). Embryos were resuspended in cold 250-STMDPS buffer (250 mM sucrose, 50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, protease inhibitors)⁷⁶. A minimal 1:10 ratio of tissue to lysis buffer was ensured during every homogenization. Embryonic tissue was homogenized using a Dounce homogenizer (20 times with loose pestle, 20 times with tight pestle). The homogenate was centrifuged for 15 min at 800g at 4 °C. The cytoplasmic fraction (supernatant) was saved, and cold 250-STMDPS buffer was added to the nuclear pellet to repeat the homogenization and centrifugation steps. The pellet was then resuspended in EXT buffer (20 mM Tris-HCl, pH 7.9, 420 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 10% glycerol, protease inhibitors) and incubated for 30 min at 4 °C while rotating. After passing the extract ten times through a 12-gauge needle, the extract was spun down at 4 °C and 17,949g for 15 min, to remove any precipitate. Batches of nuclear extract were snap frozen in liquid nitrogen and kept at -80 °C. Before proceeding with DNA pulldowns, the extracts were thawed and combined into a single tube, following centrifugation for 20 min in a chilled (4 °C) tabletop centrifuge.

DNA pulldowns and mass spectrometry

For DNA methylation pulldowns, 450 µg of protein per pulldown was used. Probes were immobilized on streptavidin Sepharose beads (GE Healthcare). Protein extracts were then added to the immobilized probes, and, following incubations and washes, bound proteins were digested with trypsin¹⁷. Tryptic peptides were desalted and concentrated using stage-tips⁷⁷ and were applied to nanoLC (Proxeon) coupled online to an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). Four-hour gradients (5–80% acetonitrile) were applied, and the top 15 tandem mass spectrometry spectra were recorded.

Proteomics data analysis

Raw data were analyzed using MaxQuant version 1.3.0.5 with default settings and the options label-free quantification and match between runs enabled⁷⁶. The UniProt *Danio rerio* database was used as the reference proteome. The resulting proteingroups.txt table was filtered for contaminants and reverse hits. The LFQ intensities obtained were \log_2 transformed, and proteins were filtered to have at least three valid values in one group (cytosine, 5mC or 5hmC). For the resulting proteins, missing values were semirandomly imputed from a normal distribution (width = 0.3 and shift = 1.8), on the basis of the assumption that these proteins were under or close to the detection limit. To identify significant interactors, an adapted two-tailed t test was performed (Perseus software), which corrects for multiple testing by applying a permutation-based FDR. Volcano plots were generated in R, in which the LFQ ratio ((h)mC/C) is plotted against the calculated FDR ($-\log_{10}$). The FDR and s0 significance threshold values used are depicted in these volcano plots (Supplementary Fig. 3). The significant (h)mC readers obtained were clustered on the basis of their enrichment ratios in R. iBAQ was performed as described^{17,78}. In brief, 3.3 µg of UPS2 standard (Sigma) was spiked into 10 µg of nuclear extract. Filter-aided sample preparation (FASP) was performed⁷⁹, and the peptides were applied to liquid chromatography and tandem mass spectrometry. Linear regression was performed on the

exact known amounts of the UPS2 standard protein and its measured iBAQ intensities, followed by extrapolation of the absolute protein amounts for the zebrafish proteins. In parallel, 100 µg of extract was digested using FASP, and peptides were fractionated using strong anion exchange (SAX) into five fractions, resulting in a deep proteome. Proteins quantified in the single FASP sample were matched with the iBAQ intensities measured for the deep proteome, which were used for linear regression and extrapolation of absolute quantifications for all proteins in the measured proteome.

Isolation of zebrafish *sox10*⁺ cells

Neural crest cells were isolated from 24 h.p.f. zebrafish transgenic embryos, expressing mCherry under the control of the *sox10* regulatory locus (TgBAC(*Sox10*:Cherry)), using FACS. Embryos were dissociated to a single-cell suspension using collagenase (20 mg/ml) and trypsin (0.05%) solution at 30 °C for 12 min with three intermittent homogenization steps. The reaction was stopped with Hank's blocking solution (1× HBSS Ca, Mg and phenol red free, 0.25% BSA, 10 mM HEPES, pH 8.0). Cells were collected by centrifugation at 500g for 10 min, resuspended in Hank's solution and passed through a cell strainer to remove cell aggregates. Single cells, concentrated by another centrifugation step, were resuspended in ~500 µl of Hank's solution and processed by FACS. Genomic DNA for further analysis was isolated from collected mCherry-positive *sox10*-expressing neural crest cells using the PureLink Genomic DNA Mini kit (K182002, Life Technologies).

Animal procedures

All animal experiments were conducted following the guidelines established and approved by the local governments and the Institutional Animal Care and Use Committee, (Universidad Pablo de Olavide, Spain) always in accordance with best practices outlined by the European Union.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank D. Secco and A. De Mendoza for critical reading of the manuscript. Spanish and Andalusian government grants BFU2013-41322-P and BIO-396 to J.L.G.-S. supported this work. R.L. was supported by an Australian Research Council Future Fellowship (FT120100862) and a Sylvia and Charles Viertel Senior Medical Research Fellowship, and work in the laboratory of R.L. was funded by the Australian Research Council, National Health and Medical Research Council, and the Raine Medical Research Foundation. O.B. is supported by an Australian Research Council Discovery Early Career Researcher Award (DECRA; DE140101962). The laboratory of M.V. is supported by grants from the Netherlands Organisation for Scientific Research (NWO-VIDI; 864.09.003) and Cancer Genomics Netherlands, a European Research Council starting grant (309384) and the European Union Framework Programme 7 Network of Excellence EpiGeneSys. J.R.E. was supported by the Gordon and Betty Moore Foundation (GBMF3034) and is an Investigator of the Howard Hughes Medical Institute. Work in the laboratory of M.M. is funded by grants from the Ministerio de Economía y Competitividad (BFU2011-23083), Comunidad Autónoma de Madrid (CELLDD-CM), and by the Pro-CNIC Foundation. This work has been supported by a grant from the US National Institutes of Health (National Institute of Child Health and Human Development, grant R01HD069344) to G.J.C.V.

References

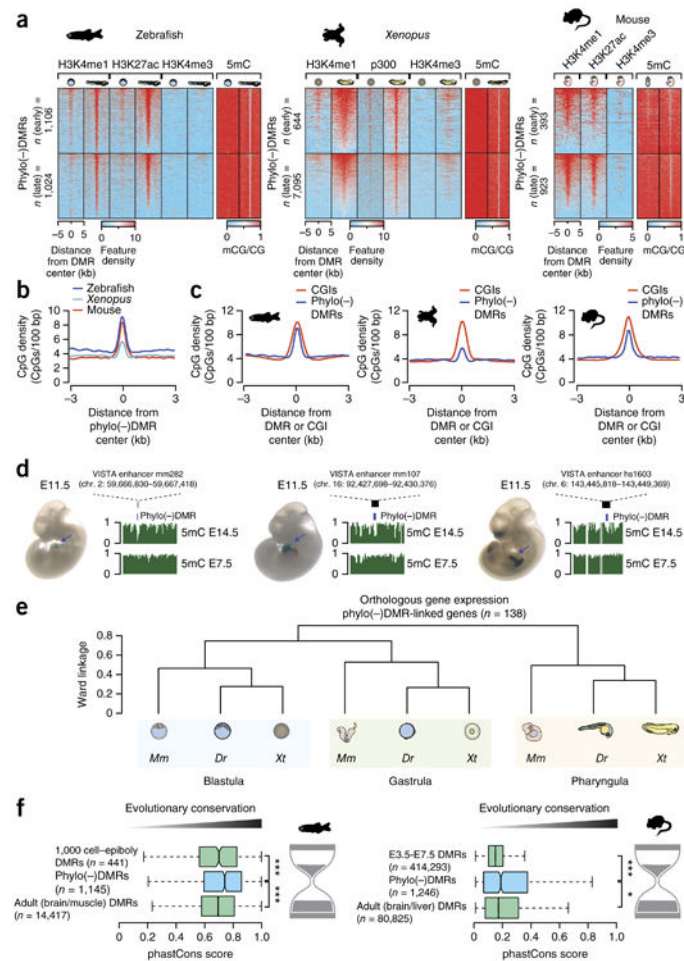
1. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002; 16:6–21. [PubMed: 11782440]
2. Santos F, Hendrich B, Reik W, Dean W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol.* 2002; 241:172–182. [PubMed: 11784103]
3. Borgel J, et al. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet.* 2010; 42:1093–1100. [PubMed: 21057502]
4. Andersen IS, Reiner AH, Aanes H, Aleström P, Collas P. Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biol.* 2012; 13:R65. [PubMed: 22830626]
5. Smith ZD, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature.* 2012; 484:339–344. [PubMed: 22456710]
6. Jiang L, et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell.* 2013; 153:773–784. [PubMed: 23663777]
7. Potok ME, Nix DA, Parnell TJ, Cairns BR. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell.* 2013; 153:759–772. [PubMed: 23663776]
8. Wang L, et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell.* 2014; 157:979–991. [PubMed: 24813617]
9. Seisenberger S, et al. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Phil Trans R Soc Lond B.* 2013; 368:20110330. [PubMed: 23166394]
10. Lee MT, Bonneau AR, Giraldez AJ. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol.* 2014; 30:581–613. [PubMed: 25150012]
11. Domazet-Lošo T, Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 2010; 468:815–818. [PubMed: 21150997]
12. Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun.* 2011; 2:248. [PubMed: 21427719]
13. Kalinka AT, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature.* 2010; 468:811–814. [PubMed: 21150996]
14. Cokus SJ, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature.* 2008; 452:215–219. [PubMed: 18278030]
15. Lister R, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell.* 2008; 133:523–536. [PubMed: 18423832]
16. Yu M, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012; 149:1368–1380. [PubMed: 22608086]
17. Spruijt CG, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell.* 2013; 152:1146–1159. [PubMed: 23434322]
18. Oswald J, et al. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol.* 2000; 10:475–478. [PubMed: 10801417]
19. Smith ZD, et al. DNA methylation dynamics of the human preimplantation embryo. *Nature.* 2014; 511:611–615. [PubMed: 25079558]
20. Guo H, et al. The DNA methylation landscape of human early embryos. *Nature.* 2014; 511:606–610. [PubMed: 25079557]
21. Lister R, et al. Global epigenomic reconfiguration during mammalian brain development. *Science.* 2013; 341:1237905. [PubMed: 23828890]
22. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
23. Dutton KA, et al. Zebrafish colourless encodes *sox10* and specifies non-ectomesenchymal neural crest fates. *Development.* 2001; 128:4113–4125. [PubMed: 11684650]
24. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature.* 2011; 480:490–495. [PubMed: 22170606]

25. Hon GC, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet.* 2013; 45:1198–1206. [PubMed: 23995138]
26. Bogdanovic O, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res.* 2012; 22:2043–2053. [PubMed: 22593555]
27. Shen Y, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature.* 2012; 488:116–120. [PubMed: 22763441]
28. Hontelez S, et al. Embryonic transcription is controlled by maternally defined chromatin state. *Nat Commun.* 2015; 6:10148. [PubMed: 26679111]
29. Creighton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA.* 2010; 107:21931–21936. [PubMed: 21106759]
30. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 2011; 470:279–283. [PubMed: 21160473]
31. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007; 39:311–318. [PubMed: 17277777]
32. Illingworth RS, et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 2010; 6:e1001134. [PubMed: 20885785]
33. Long HK, et al. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife.* 2013; 2:e00348. [PubMed: 23467541]
34. Cohen NM, Kenigsberg E, Tanay A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell.* 2011; 145:773–786. [PubMed: 21620139]
35. Krebs AR, Dessus-Babus S, Burger L, Schübeler D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife.* 2014; 3:e04094. [PubMed: 25259795]
36. Wachter E, et al. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. *eLife.* 2014; 3:e03397. [PubMed: 25259796]
37. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 2007; 35:D88–D92. [PubMed: 17130149]
38. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009; 457:854–858. [PubMed: 19212405]
39. Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1–13. [PubMed: 19033363]
40. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4:44–57. [PubMed: 19131956]
41. Pauli A, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 2012; 22:577–591. [PubMed: 22110045]
42. Paranjpe SS, Jacobi UG, van Heeringen SJ, Veenstra GJ. A genome-wide survey of maternal and embryonic transcripts during *Xenopus tropicalis* development. *BMC Genomics.* 2013; 14:762. [PubMed: 24195446]
43. Auclair G, Guibert S, Bender A, Weber M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* 2014; 15:545. [PubMed: 25476147]
44. Shen L, et al. Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell.* 2014; 15:459–470. [PubMed: 25280220]
45. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005; 15:1034–1050. [PubMed: 16024819]
46. Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. The hourglass and the early conservation models—co-existing patterns of developmental constraints in vertebrates. *PLoS Genet.* 2013; 9:e1003476. [PubMed: 23637639]
47. Nord AS, et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell.* 2013; 155:1521–1531. [PubMed: 24360275]

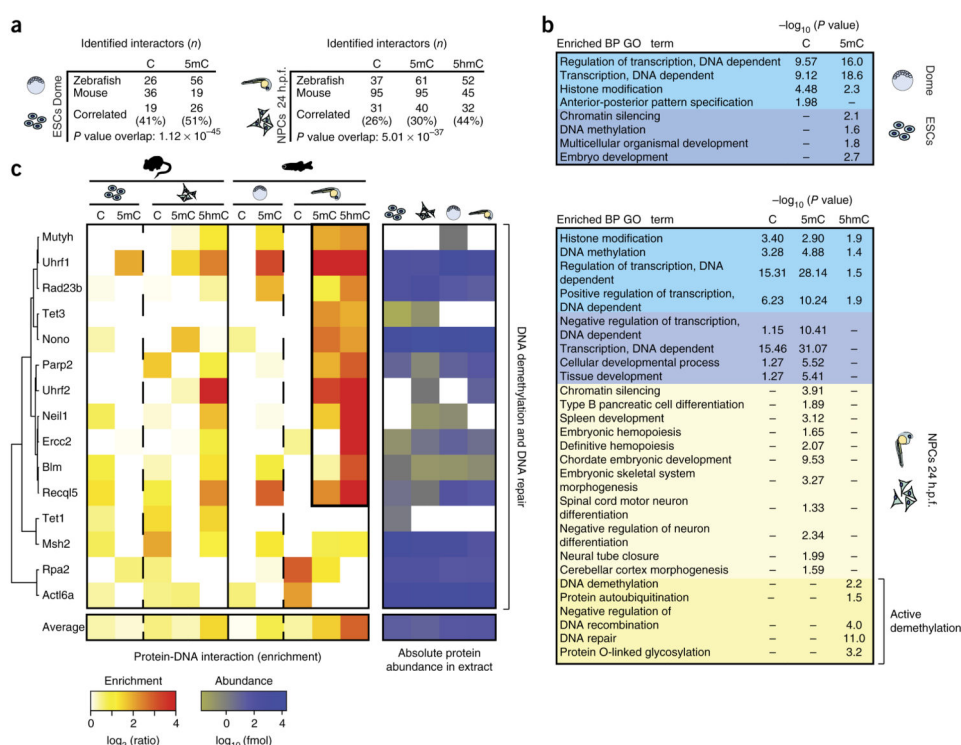
48. Stergachis AB, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell*. 2013; 154:888–903. [PubMed: 23953118]
49. Barreto G, et al. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature*. 2007; 445:671–675. [PubMed: 17268471]
50. Rai K, et al. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell*. 2008; 135:1201–1212. [PubMed: 19109892]
51. Almeida RD, et al. 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. *Epigenetics*. 2012; 7:383–389. [PubMed: 22419071]
52. Kamstra JH, Løken M, Aleström P, Legler J. Dynamics of DNA hydroxymethylation in zebrafish. *Zebrafish*. 2015; 12:230–237. [PubMed: 25751297]
53. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*. 2011; 473:343–348. [PubMed: 21490601]
54. Wu H, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*. 2011; 473:389–393. [PubMed: 21451524]
55. Xu Y, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell*. 2012; 151:1200–1213. [PubMed: 23217707]
56. Ge L, et al. TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes via DNA oxidative demethylation in a zebrafish model. *Mol Cell Biol*. 2014; 34:989–1002. [PubMed: 24396069]
57. Li C, et al. Overlapping requirements for Tet2 and Tet3 in normal development and hematopoietic stem cell emergence. *Cell Rep*. 2015; 12:1133–1143. [PubMed: 26257178]
58. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013; 10:1213–1218. [PubMed: 24097267]
59. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11:R106. [PubMed: 20979621]
60. Vastenhouw NL, et al. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*. 2010; 464:922–926. [PubMed: 20336069]
61. Lindeman LC, et al. Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev Cell*. 2011; 21:993–1004. [PubMed: 22137762]
62. Lee HJ, et al. Developmental enhancers revealed by extensive DNA methylome maps of zebrafish early embryos. *Nat Commun*. 2015; 6:6315. [PubMed: 25697895]
63. Perera A, et al. TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Rep*. 2015; 11:283–294. [PubMed: 25843715]
64. Ding J, et al. Tex10 coordinates epigenetic control of super-enhancer activity in pluripotency and reprogramming. *Cell Stem Cell*. 2015; 16:653–668. [PubMed: 25936917]
65. Gu TP, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*. 2011; 477:606–610. [PubMed: 21892189]
66. Dawlaty MM, et al. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*. 2011; 9:166–175. [PubMed: 21816367]
67. Ko M, et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci USA*. 2011; 108:14566–14571. [PubMed: 21873190]
68. Moran-Crusio K, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011; 20:11–24. [PubMed: 21723200]
69. Dawlaty MM, et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell*. 2013; 24:310–323. [PubMed: 23352810]
70. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
71. Lister R, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*. 2011; 471:68–73. [PubMed: 21289626]
72. Schultz MD, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*. 2015; 523:212–216. [PubMed: 26030523]

73. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842. [PubMed: 20110278]
74. Bray, N., Pimentel, H., Melsted, P., Pachter, L. Near-optimal RNA-Seq quantification. *arXiv*. 2015. <http://arxiv.org/abs/1505.02710>
75. Vilella AJ, et al. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 2009; 19:327–335. [PubMed: 19029536]
76. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008; 26:1367–1372. [PubMed: 19029910]
77. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrixassisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003; 75:663–670. [PubMed: 12585499]
78. Schwanhäusser B, et al. Global quantification of mammalian gene expression control. *Nature*. 2011; 473:337–342. [PubMed: 21593866]
79. Wi niewski JR, Zougman A, Mann M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res*. 2009; 8:5674–5678. [PubMed: 19848406]

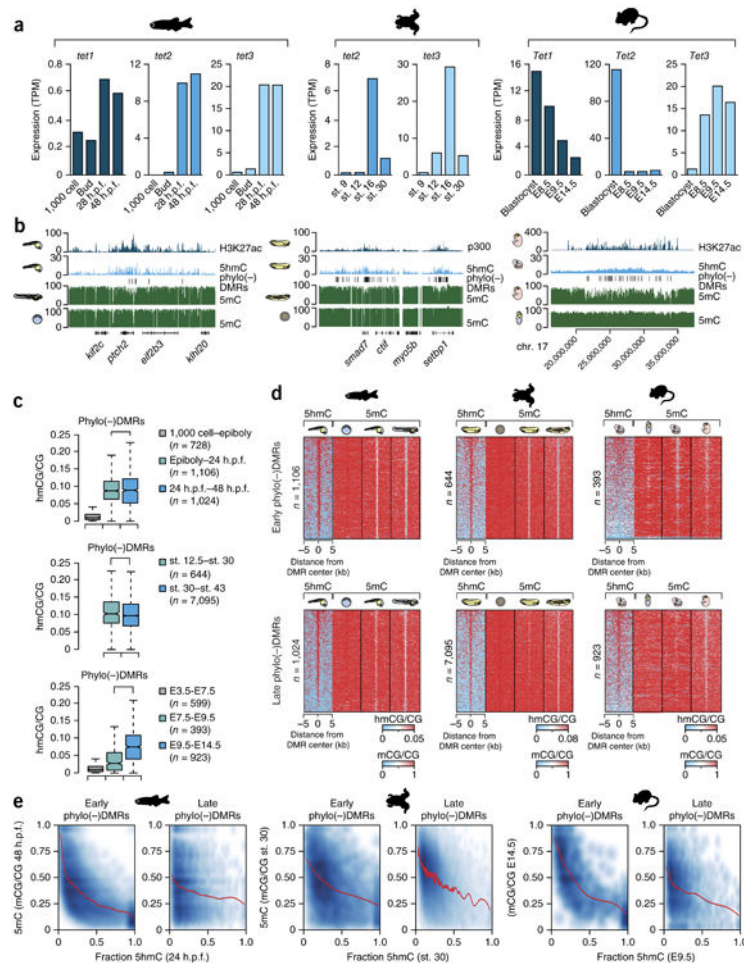
Nat Genet. Author manuscript; available in PMC 2018 April 23.

**Figure 2.**

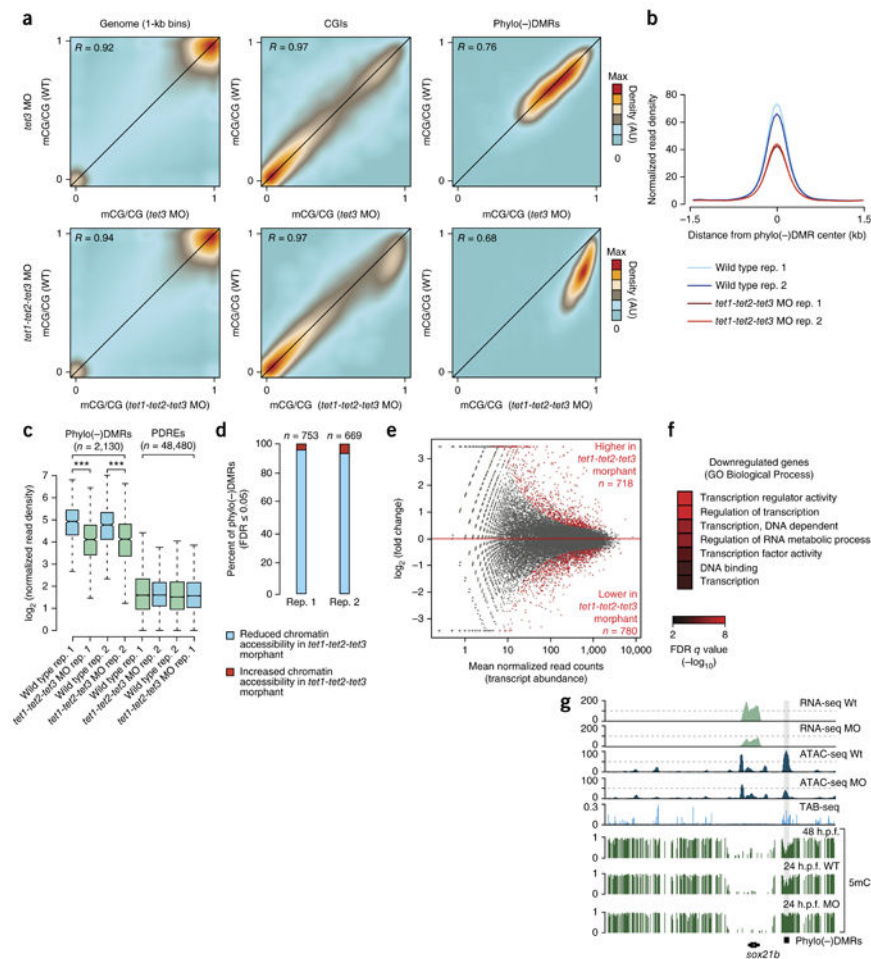
Phylo(-)DMRs are developmentally activated enhancers associated with vertebrate body plan formation. (a) Sorted heat maps of 5mC levels and normalized ChIP-seq read density for H3K4me1, H3K4me3, H3K27ac and p300 binding at phylo(-)DMRs in zebrafish, *Xenopus* and mouse embryos. (b) Mean CpG density (CpGs/100 bp) in zebrafish, *Xenopus* and mouse phylo(-)DMRs. (c) Comparisons of mean CpG density (CpGs/100 bp) for phylo(-)DMRs and CpG islands (CGIs). (d) Genomic overlaps and expression patterns (indicated by arrows) of validated VISTA enhancers associated with mouse phylo(-)DMRs. (e) Hierarchical clustering of transcript abundance (scaled TPMs) of orthologous genes ($n = 138$) linked to phylo(-)DMRs. *Mm*, mouse; *Dr*, zebrafish; *Xt*, *Xenopus*. (f) Evolutionary conservation (aggregate phastCons scores) of DMRs in zebrafish and mouse embryos (Kruskal-Wallis test, Dunn's post test, $*P < 0.05$, $***P < 0.001$). The boxes show the IQR around the median, and the whiskers extend from the minimum value to the maximum value unless the distance to the first or third quartile was more than 1.5 times the IQR.

**Figure 3.**

Active demethylation components bind 5mC and 5hmC during the phylotypic period in vertebrates. **(a)** Total numbers of nuclear proteins found to interact with cytosine, 5mC or 5hmC oligonucleotides. P values (hypergeometric test: zebrafish dome stage–mouse ESCs, $P = 1.12 \times 10^{-45}$; 24 h.p.f. zebrafish embryo–mouse NPCs, $P = 5.01 \times 10^{-37}$) represent the significance of interspecies correlations. **(b)** Enriched GO Biological Process (BP) terms for the conserved cytosine, 5mC or 5hmC readers for pluripotent (dome, ESC) and differentiated (24 h.p.f., NPC) stages. **(c)** Hierarchical correlation-based clustering of the enrichment of DNA demethylation– and DNA repair–linked 5mC and 5hmC readers identified in zebrafish embryos and mouse cell nuclear extracts.

**Figure 4.**

Phylo(-)DMRs are characterized by 5hmC enrichment in vertebrate embryos. (a) Steady-state abundance (TPM) of Tet1, Tet2 and Tet3 transcripts during zebrafish, *Xenopus* and mouse embryogenesis. (b) Genome browser displays demonstrating the co-occurrence of phylo(-)DMRs, 5hmC (TAB-seq) and active enhancer marks (H3K27ac, p300). (c) 5hmC levels in developmentally hypomethylated DMRs. The boxes show the IQR around the median, and the whiskers extend from the minimum value to the maximum value unless the distance to the first or third quartile was more than 1.5 times the IQR. (d) Phylo(-)DMR-centered heat maps of 5hmC and 5mC enrichment. (e) Negative correlation of the fraction of 5hmC (hmCG/CG)/(mCG/CG) at the phylotypic stage and the fraction of 5mC (mCG/CG) at the tailbud stage for individual dinucleotides within phylo(-)DMRs.

**Figure 5.**

Tet proteins are required for phylo(-)DMR demethylation and body plan formation in zebrafish. **(a)** Correlation of 5mC levels in wild-type (WT) and *tet3* and *tet1-tet2-tet3* knockdown embryos across the genome (1-kb non-overlapping windows), in CpG islands and in phylo(-)DMRs. MO, morpholino; AU, arbitrary units. **(b)** Average profiles (normalized read density) of ATAC-seq signal over phylo(-)DMRs in wild-type and *tet1-tet2-tet3* morphant embryos. **(c)** Reduced ATAC-seq signal (Wilcoxon test, *** $P < 0.001$) in phylo(-)DMRs as compared to a general population of PDREs identified in zebrafish embryos. The boxes show the IQR around the median, and the whiskers extend from the minimum value to the maximum value unless the distance to the first or third quartile was more than 1.5 times the IQR. **(d)** Number of regions displaying significantly altered (Fisher's exact test, FDR $q = 0.05$, minimum fold change = 2) chromatin accessibility in *tet1-tet2-tet3* knockdown embryos. **(e)** Statistically significant (Benjamini-Hochberg FDR $q = 0.05$) differentially expressed genes in *tet1-tet2-tet3* morphants. **(f)** GO term enrichments (FDR $q = 0.05$) for genes downregulated in *tet1-tet2-tet3* morphants. **(g)** Genomic example (*sox21b* locus) of reduced chromatin accessibility and reduced gene expression in *tet1-tet2-tet3* morphants when compared to wild-type embryos.