Active DNA demethylation during the vertebrate phylotypic period

Ozren Bogdanović[1,2,3*], Arne H. Smits[4], Juan J. Tena[2] Elisa de la Calle Mustienes[2], Ruth Williams[5], Upeka Senanayake[5], Ethan Ford[1,3], Matthew D. Schultz[6], Saartje Hontelez[7], Ila van Kruijsbergen[7], Teresa Rayon[8], Felix Gnerlich[9], Thomas Carell[9], Gert Jan C. Veenstra[7], Miguel Manzanares[8], Tatjana Sauka-Spengler[5], Joseph R. Ecker[6,10], Michiel Vermeulen[3,*], Jose Luis Gómez-Skarmeta[2,*], Ryan Lister[1,3,*]

[1] ARC Center of Excellence in Plant Energy Biology, School of Chemistry and Biochemistry, Faculty of Science, The University of Western Australia, Perth, WA 6009, Australia

[2] Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide (UPO), Consejo Superior de Investigaciones Científicas (CSIC), 41013 Seville, Spain

[3] The Harry Perkins Institute of Medical Research, Perth, WA 6009, Australia

[4] Radboud University, Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

[5] Weatherall Institute of Molecular Medicine, University of Oxford, UK

[6] Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

[7] Radboud University, Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

[8] Centro Nacional de Investigaciones Cardiovasculares (CNIC), 28029 Madrid, Spain

[9] Center for Integrated Protein Science, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

[10] Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

*Correspondence: ozren.bogdanovich@ uwa.edu.au, m.vermeulen@ncmls.ru.nl, jlgomska@upo.es, ryan.lister@uwa.edu.au

Abstract

The vertebrate body plan and organs are shaped during a highly 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 thousands of enhancers during the phylotypic period in zebrafish, *Xenopus* and mouse. These dynamic enhancers are linked to essential developmental genes that display coordinated transcriptional and epigenomic changes in the diverse vertebrates during embryogenesis. Phylotypic stage-specific binding of Tet proteins to (hydroxy)methylated DNA, and enrichment of hydroxymethylcytosine on these enhancers, implicated active DNA demethylation in this process. Reduced chromatin accessibility and increased methylation levels, specifically on these enhancers, in a zebrafish *tet1/2/3* loss of function system, is indicative of DNA methylation being an upstream regulator of phylotypic enhancer function. Overall, our study reveals a novel regulatory module associated with the most conserved phase of vertebrate embryogenesis and uncovers an ancient developmental role for the Tet dioxygenases.

Methylation of cytosine residues (mC) is a major, mostly repressive epigenomic modification associated with key biological processes {Bird, 2002 #138}. Early vertebrate embryos display notable mC dynamics associated with totipotency establishment {Santos, 2002 #70;Borgel, 2010 #25;Andersen, 2012 #68;Smith, 2012 #30;Jiang, 2013 #22;Potok, 2013 #23}. Interestingly, these mC remodelling events occur in a species-specific fashion. Mammalian embryos employ a combination of active and passive mechanisms to remodel their methylomes after fertilization {Wang, 2014 #21}. The active mechanism consists of Tet dependent oxidation of mC that involves the hydroxymethylcytosine (hmC) intermediate, whereas the passive mechanism is based on mC dilution throughout cell divisions in the absence of methylation maintenance {Seisenberger, 2013 #133}. In zebrafish for example, no Tet activity was detected during pluripotency, and the mC dynamics in the early embryo are reduced to passive reconfiguration of the maternal methylome to match the sperm methylation pattern {Jiang, 2013 #22;Potok, 2013 #23}. However, very little is known about the mechanisms and evolutionary conservation of DNA methylation patterning during later stages of vertebrate embryogenesis. Of particular interest is the vertebrate phylotypic stage, a period during which a developing embryo displays the most morphological and transcriptomic similarities compared to embryos of other vertebrate species, and during which the body plan is established {Domazet-Loso, 2010 #66;Irie, 2011 #20}.

Here we report widespread DNA demethylation of thousands of enhancers associated with highly conserved regulatory pathways during the phylotypic period in zebrafish, *Xenopus* and mouse. Through whole genome bisulfite sequencing (WGBS) {Cokus, 2008 #111;Lister, 2008 #37;Yu, 2012 #67}, quantitative interaction proteomics {Spruijt, 2013 #46}, 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, ATAC-seq and whole genome methylome profiling of *tet1/2/3* morphant 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 vertebrate embryogenesis. By unravelling 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 undescribed developmental role of Tet proteins that was missed in knockout studies in mice due to complex developmental requirements for Tet proteins in mammals {Dawlaty, 2013 #117;Gu, 2011 #59;Moran-Crusio, 2011 #95}. Overall, our study reveals a novel regulatory module associated with the most conserved phase of vertebrate embryogenesis and uncovers an ancient developmental role for the Tet dioxygenases.

RESULTS

Phylotypic hypomethylation

To select equivalent embryonic stages of vertebrate embryos spanning the phylotypic period, we utilized as guidelines the interspecies relationships deduced from reciprocal best transcriptome similarity {Irie, 2011 #20} (Fig. 1a). The phylotypic stage in zebrafish and *Xenopus* correspond to 24 hours post-fertilization (hpf) and stage (st.) 30 respectively, whereas in mouse this stage coincides with embryonic day E9.5 (Fig. 1a – hourglass). We generated WGBS DNA methylome profiles for four stages of zebrafish, *Xenopus* and mouse embryogenesis corresponding to blastula (1K cell {Jiang, 2013 #22} , st.9, E3.5 {Wang, 2014 #21}), gastrula (80% epiboly, st.12, E7.5 {Wang, 2014 #21}), pharyngula (24hpf, st.30, E9.5) and tailbud/fetus (48hpf, st.43, E14.5)

stages, (for details on non-conversion rates and sequencing statistics see Supplementary Table 1). Overall, both zebrafish and *Xenopus* genomes are globally highly methylated during these developmental stages, with average mC levels of 85% and 91%, respectively (Fig. 1a). Unlike zebrafish and *Xenopus*, mammalian embryos including mouse, display significant alterations in mC levels during early embryonic stages (Fig. 1a) as previously described {Oswald, 2000 #72;Santos, 2002 #70;Smith, 2012 #30;Smith, 2014 #28;Guo, 2014 #9;Wang, 2014 #21}.

To identify genomic regions displaying developmental changes in mC, we searched the DNA methylome profiles for differentially methylated regions (DMRs, FDR = 0.05) {Lister, 2013 #6} (Supplementary Tables 2-4 and Supplementary Fig. 1a). While in zebrafish embryos the transition from the 1K cell (blastula) to 80% epiboly (gastrula) stage is characterized by both developmental hyper- and hypo-methylation (Supplementary Fig. 1b), the transitions associated with the phylotypic stage are overwhelmingly characterized by developmental hypomethylation (Fig. 1b and Supplementary Fig.1b). A similar trend can be observed in *Xenopus* embryos, where only a single DMR was identified between blastula and gastrula stages and the mC remodelling appears to be limited to demethylation surrounding the phylotypic stage (Fig. 1b). Similarly, the majority of mouse DMRs found between late gastrula (E7.5) and the phylotypic stage (E9.5) are being developmentally demethylated.

To explore whether the DMRs associated with developmental demethylation surrounding the phylotypic stage are implicated in similar processes in the three examined species, we have employed the GREAT tool {McLean, 2010 #80} to search for enriched and conserved ontology terms. Strikingly, we identified 133 conserved and significantly enriched (*FDR Q*. val. < 0.01) terms for early phylotypic DMRs (late gastrula - phylotypic stage demethylation) in all species, whereas for the late phylotyipc DMRs (phylotypic stage – tailbud/fetus demethylation), that

number was 109 (Fig. 1c and Supplementary Tables 5-6). The enriched ontologies include 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 other developmental stages. Given their similar mC dynamics and implication in various developmental processes, from now forward 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 of later stage embryos, and to recapitulate our results within a specific cell lineage, we generated a WGBS DNA methylome of a pure neural crest cell population (*sox10*+) {Dutton, 2001 #69} isolated from 24hpf zebrafish embryos. As expected, we observe a clear decrease of phylo-DMR mC levels in *sox10*+ cells when compared to blastula (1K cell) embryos and a further loss of methylation in the adult zebrafish brain (Fig. 1d). Similar mC profiles where obtained when comparing embryonic with adult brain mC levels in *Xenopus* and mouse (Fig. 1d) {Lister, 2013 #6}. Next we wanted to explore whether phylo-DMRs affect only certain lineages or whether their usage is a more widespread phenomenom. To that end we took advantage of methylome maps corresponding to a number of adult tissues derived from all three embryonic layers (ectoderm, endoderm and mesoderm) {Hon, 2013 #113}. Average mC profiles revealed strong phylo-DMR demethylation (40 – 50%) 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 profiling reveals a highly conserved process of mC reconfiguration that takes place throughout the phylotypic period in diverse vertebrate species, involves developmental hypomethylation and affects organs derived from all embryonic layers.

Phylo-DMRs are developmentally activated enhancers

To explore the chromatin configuration and genomic context of phylo-DMRs, we utilized ChIP-seq data for promoter (H3K4me3), poised enhancer (H3K4me1) and active enhancer (H3K27ac, p300) chromatin marks {Bogdanovic, 2012 #52;Shen, 2012 #162}. Sorted heatmaps of phylo-DMRs in zebrafish and *Xenopus* demonstrate a strong developmental enrichment for enhancer, but not promoter, histone marks (Fig. 2a). Similarly, the mouse phylo-DMRs are strongly enriched with H3K4me1, H3K27ac but not H3K4me3, indicative of active enhancer chromatin {Creyghton, 2010 #76;Rada-Iglesias, 2011 #15;Heintzman, 2007 #128}. Notably, the early zebrafish DMRs did not display such enrichment in enhancer chromatin (Supplementary Fig.2), however, they coincided with promoters of *vasa*, *dazl* and *ntla* genes, previously identified as differentially methylated between early embryos and differentiated tissues {Jiang, 2013 #22;Potok, 2013 #23} (Supplementary Fig. 2b-d). Inspection of the CpG density of phylo-DMRs by analysing mean CpG levels within and flanking phylo-DMR regions revealed similar CpG densities in zebrafish and mouse DMRs, while lower CpG density was observed in *Xenopus* DMRs (Fig. 2b). While the CpG density of phylo-DMRs in zebrafish and mouse was similar to that of CpG islands identified through unmethylated DNA pulldowns {Illingworth, 2010 #129;Long, 2013 #93}, *Xenopus* phylo-DMRs displayed considerably lower CpG density than those CpG islands (Fig. 2c). These results reveal phylo-DMRs as developmentally activated distal regulatory elements of variable CpG density and confirm previous notions that CpG density alone is not a major driver of regulatory function {Cohen, 2011 #141;Krebs, 2014 #16;Wachter, 2014 #17}.

To provide further proof that phylo-DMRs act as developmental enhancers, we intersected the phylo-DMR genomic positions with previously validated enhancers from the

VISTA enhancer browser database {Visel, 2007 #163;Visel, 2009 #164} and obtained 13 intersections corresponding to enhancers with diverse expression patterns, consistent with phylo-DMR demethylation throughout various lineages (Fig. 1e and Supplementary Table 7). Examples of such heart and limb enhancers and their overlaps with phylo-DMRs are visualized in Fig. 2d.

Given the high conservation of gene ontology enrichments (Fig. 1c) and highly comparable chromatin configuration of phylo-DMRs (Fig. 2a), we postulated that phylo-DMR-linked genes should be co-regulated during zebrafish, *Xenopus* and mouse embryogenesis. Firstly, we identified orthologous genes linked to phylo-DMRs in all examined species (N= 211, Supplementary Table 8) and subjected them to pathway enrichment analyses {Huang da, 2009 #166;Huang da, 2009 #165}. Indeed, we found that orthologues phylo-DMR-linked genes are enriched in pathways such as *Wnt*, *Notch* and *TGF-beta*, implicated in body plan and organ formation (Supplementary Fig. 3 and Supplementary Table 9). To our knowledge this is the first notion of these key developmental pathways being regulated through DNA methylation in multiple vertebrate species. Next we compared the developmental expression profiles of these orthologus phylo-DMR linked genes in *Xenopus*, zebrafish and mouse at three developmental time-points corresponding to blastula (1K cell, st9, blastocyst), late gastrula (bud, st12, E8.5) and the phylotypic stage (28hpf, st30, E9.5) {Pauli, 2012 #82;Paranjpe, 2013 #83;Auclair, 2014 #85;Shen, 2014 #84}. Hierarchical clustering analysis of scaled RNA-seq data (TPMs) (Fig. 2e) revealed a strong developmental correlation of the orthologous genes, with the samples clustering by developmental stage rather than species, further supporting that phylo-DMRs are involved in conserved regulatory networks. Finally, assessment of the evolutionary conservation of zebrafish and mouse phylo DMRs by mapping aggregate conservation scores {Siepel, 2005

#137} revealed that phylo-DMRs display significantly higher evolutionary conservation (Kruskal-Wallis test, $P < 0.01$) than the early (blastula/gastrula) or late (adult organ) DMRs (Fig. 2f). This is consistent with previous studies that described a similar phenomenon through DNaseI hypersensitivity and ChIP-seq profiling {Nord, 2013 #140;Stergachis, 2013 #112;Harmston, 2013 #142}. Together, these findings support a role of 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

In zebrafish and *Xenopus* several DNA demethylation pathways have been described {Barreto, 2007 #132;Rai, 2008 #131;Almeida, 2012 #60}. To obtain a better insight into which cellular factors might be implicated in phylo-DMR demethylation, we performed a quantitative interaction-proteomics screen using mC-modified, hmC-modified and unmodified DNA oligonucleotides {Spruijt, 2013 #46} and nuclear extracts from zebrafish embryos, to identify developmentally regulated mC readers in dome (blastula) and 24hpf (phylotypic stage) stages. We also profiled binding to hydroxymethylcytosine (hmC), as this active demethylation intermediate was previously observed in zebrafish 24hpf embryos {Almeida, 2012 #60}. Proteins displaying differential binding enrichment between unmethylated and methylated DNA oligonucleotides were identified by an adapted *t*-test on label-free quantification (LFQ) intensities {Spruijt, 2013 #46} (Supplementary Fig. 4, 5 and Supplementary Tables 10, 11). We also performed absolute quantification of the nuclear extract proteomes (Supplementary Fig. 4) to directly assess the effects of stage-specific protein abundance on binding enrichments. In total,

we identified 98 context-specific mC readers, the majority of which display stage-specific binding profiles. Of 45 quantified readers, only 19 could be explained by differential protein abundance (Supplementary Fig. 5), thereby demonstrating the dynamics of the mC interactome during early vertebrate embryogenesis.

To investigate the evolutionary conservation of (h)mC readers in vertebrates, we compared our set of zebrafish readers to the readers previously identified in mouse ESCs and NPCs {Spruijt, 2013 #46}. In total, 96 proteins for which orthologues are annotated in both species were identified as either mC or C readers in dome or mESC nuclear extracts. 45 of these 96 proteins show conserved binding (47%, $P < 1.1 \times 10^{-45}$, hypergeometric test) (Fig. 3a, b and Supplementary Fig. 6). Readers for C, mC and hmC in NPCs and 24hpf embryos also display substantial overlap; 99 out of 249 proteins (40%, $P < 5 \times 10^{-37}$, hypergeometric test) for which orthologues are annotated in both species display conserved binding (Fig. 3a,b and Supplementary Fig. 6). Notably, the hmC readers are 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 NPCs {Spruijt, 2013 #46}. All proteins involved in DNA repair/demethylation show a clear trend towards a higher binding affinity to mC in the differentiated states, and strongest binding to hmC (Fig. 3c, average row). Together, this reveals an enrichment of active, Tet-dependent DNA demethylation pathways occurring during the vertebrate phylotypic period.

The observed enrichment in Tet and DNA repair protein binding to mC and hmC in zebrafish phylotypic stage embryos and mouse NPCs suggests that phylo-DMRs may be demethylated through a Tet-dependent mechanism. While the diverse vertebrate species display different Tet expression profiles during development, for all species the phylotypic period is

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 hmC intermediate, zebrafish (24hpf), *Xenopus* (st30) and mouse (E9.5) embryos were profiled by Tet-assisted bisulfite sequencing (TAB-seq){Yu, 2012 #67} (for details on non-conversion and other associated TAB-seq statistics see Supplementary Table 1). Notable enrichment of hmC was detected in regions marked by active enhancer chromatin and corresponding to phylo-DMRs (Fig. 4b). Similarly, average profiles of hmC and mC over phylo-DMRs in fish, frogs and mice demonstrate the enrichment in hmC and depletion of the mC signal occurring during the same embryonic stage (Fig. 4c). Altogether, these base resolution maps reveal strong hmC enrichment on phylo-DMRs and provide further evidence of active DNA demethylation during the vertebrate phylotypic period.

Tet proteins demethylate phylo-DMRs

To assess the effects of inhibiting Tet function in zebrafish embryos, we took advantage of the morpholino (MO) knockdown approach to target the Tet3 protein, a major phylotypic mC/hmC interactor discovered in our quantitative proteomics screen (Fig. 3c). Embryos injected with Tet3 MO showed only minor defects, with 23% of the morphants displaying varying degrees of microphthalmia (Supplementary Fig. 7a) as previously reported in a Tet3 MO study performed in *Xenopus* embryos {Xu, 2012 #89}. To exclude the potential effects of Tet protein functional redundancy, zebrafish embryos were injected with a triple Tet (Tet1/2/3) MO {Ge, 2014 #92}. The triple morphants were severely affected, with the majority displaying embryonic lethality. The embryos that survived gastrulation (23%) displayed short and blended axes, impaired head

structures, small eyes, reduced pigmentation and heart oedemas (Supplementary Fig. 7a). We next performed WGBS (Supplementary Table 1) upon morphant embryos (N = 50) and compared the mC profiles to their wild type counterparts. Global mCG levels (1 kb non-overlapping windows) are similar in both Tet3 and Tet1/2/3 morphants when compared to wild-type embryos, and are 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 profiling {Long, 2013 #93}. However, comparisons between the Tet1/2/3 morphants and wild-type embryos revealed a striking increase in mCG levels of phylo-DMRs in the morphants, affecting almost all (92%) of these regions. Similarly, there was an increase of mCG in a subpopulation of CpG islands that corresponds to phylo-DMRs, and thus with sites of aberrant demethylation in the Tet1/2/3 morphant (Fig. 5a and Supplementary Fig. 7b,c). Taken together, these data reveal an embryonic requirement for Tet proteins and demonstrate abnormalities in mC remodelling at key regulatory elements caused by the absence of these proteins.

Next, we wanted to determine whether the loss of Tet proteins and thus increase in phylo-DMR mC would result in decreased chromatin accessibility of these regulatory elements. We therefore performed ATAC-seq {Buenrostro, 2013 #168} on two pools of *tet1/2/3* MO embryos and their wild type counterparts. Both wild type and morphant embryos display notable ATAC-seq signal enrichment on phylo-DMRs, however that enrichment is significantly (Kruskal-Wallis test, $P < 0.001$) decreased in *tet1/2/3* MO embryos (Fig. 5b,c). Importantly, such a change is not observed on a general population of putative regulatory elements (PDREs) identified in 24hpf embryos through ChIP-seq profiling {Bogdanovic, 2012 #52}. Next, we identified phylo-DMRs displaying a statistically significant change (Fisher's exact test, FDR Q value < 0.05) in ATAC-

seq signal between wild type and morphant embryos. This population consists of ~40% phylo-DMRs and is characterized by reduced ATAC-seq signal in > 90% regions consistent with the average ATAC-seq profiles (Fig. 5d). Finally we wanted to address the impact of Tet1/2/3 loss on embryonic transcription and performed RNA-seq on two pools of *tet1/2/3* morphants and their wild type controls. Of note, unsupervised clustering of zebrafish embryonic transcriptomes revealed a cluster of the *tet1/2/3* MO samples, their wild type 24hpf controls and other zebrafish phylotypic stage samples {Pauli, 2012 #82}, indicative of MO phenotype specificity (Supplementary Fig. 7d). Differential gene expression analyses {Anders, 2010 #169} unravelled bi-directional changes in the transcriptomes of *tet1/2/3* morphants (N = 718 upregulated genes and N = 780 downregulated genes in the morphant) (Fig. 5e and Supplementary Table 12). Gene ontology analyses {Huang da, 2009 #166;Huang da, 2009 #165} of down-regulated genes resulted in an overrepresentation of terms associated with transcriptional regulation similar to those described enriched in phylo-DMR-associated genes (Fig. 5f, Supplementary Tables 5 and 6) whereas no statistically significant enrichments were associated with the upregulated group of genes. An example of aberrant phylo-DMR demethylation in the *tet1/2/3* morphant resulting in reduced chromatin accessibly and reduced transcription is depicted in Fig. 5g. Taken together, our data links *tet1/2/3*-dependent demethylation with the proper activation of phylotypic-stage enhancers and suggests a regulatory role for mC in phylo-DMR usage.

DISCUSSION

It is well established that vertebrates remodel their epigenomes during embryogenesis to achieve totipotency {Santos, 2002 #70;Borgel, 2010 #25;Vastenhouw, 2010 #171;Lindeman, 2011

#173;Andersen, 2012 #68;Smith, 2012 #30;Jiang, 2013 #22;Potok, 2013 #2;Lee, 2014 #172}. However, to date only very limited insights have been obtained regarding mC dynamics and the evolutionary conservation of mC patterning in later stage vertebrate embryos. Here we describe widespread enhancer demethylation that starts before and continues throughout the phylotypic stage of fish, frog 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 play conserved roles in the establishment of the vertebrate body plan, including key developmental pathways such as *Notch/Delta, Wnt* and *TGF-beta*. These observations are in line with a recent study that identified distal regulatory elements as targets of mC remodelling in zebrafish embryos {Lee, 2015 #97}. Furthermore, we demonstrate that this observed wave of enhancer demethylation requires Tet proteins and mC to hmC conversion during the phylotypic period. 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 cultures {Stadler, 2011 #4}. In fact, a recent study demonstrated Tet3 targeting through transcription factor binding in the mammalian neural lineage {Perera, 2015 #130}, whereas Tet1 was shown to associate with the transcription factor Tex10 to regulate chromatin conformation on super-enhancers in ESCs {Ding, 2015 #170}.

Mammals display complex embryonic requirements for Tet activity; the $Tet3^{-/-}$ KO results in early embryonic lethality {Gu, 2011 #59}, whereas the $Tet1^{-/-}$ {Dawlaty, 2011 #101} and $Tet2^{-/-}$ {Ko, 2011 #96;Moran-Crusio, 2011 #95} KOs are viable. The double Tet1/Tet2 KO is lethal in the majority of embryos, however, a small percentage of Tet1/Tet2 KO mice can successfully be grown to adulthood {Dawlaty, 2013 #117}. On the other hand, anamniotes such

as fish and frogs do not express Tet proteins during early embryonic stages and no hmC signal was detected in pluripotent zebrafish embryos {Potok, 2013 #23;Jiang, 2013 #22} . Nonetheless, here we demonstrate that both anamniotes and mammals employ Tet-dependent demethylation of enhancers for gene regulation during the phylotypic period, thereby suggesting an ancient, pan-vertebrate regulatory logic. These findings are supported by the higher evolutionary conservation of phylo-DMRs in fish and mice, when compared to early or late DMRs. Finally, this work provides important insights into the roles that mC plays on embryonic enhancer elements and implicates mC as an upstream regulator of enhancer function. Our study reveals a highly conserved mechanism used by vertebrates during the specification of the enhancer repertoire needed for body plan 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 determine the embryonic requirements for each of these processes.

ONLINE METHODS

Data access

The raw data have been deposited in GEO under the accession number GSE68087: (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ghsxkymsxzkxfut&acc=GSE68087).
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD001164.

MethylC-seq

Genomic DNA from zebrafish and *Xenopus* embryos was obtained as described before {Bogdanovic, 2013 #49}. MethylC-seq library generation was performed as described previously {Lister, 2013 #6}. Library amplification was performed with KAPA HiFi HotStart Uracil+ DNA polymerase (Kapa Biosystems, Woburn, MA), using 6 cycles of amplification. Single-read MethylC-seq libraries (for details see Table S1) were processed and aligned as described previously {Lister, 2011 #33}. For reference genomes used in this study, please see Table S1. Previously published {Jiang, 2013 #22;Potok, 2013 #23;Wang, 2014 #21} paired-read MethylC-seq data was mapped and processed as described previously {Lister, 2011 #33} with the following parameters: -e 120 -l 20 -n 1 -k 10 -o 4 -I 0 -X 1000, except the previous filter that excluded reads containing >3 cytosine bases in the CH context was not applied in this study. To estimate the bisulfite non-conversion frequency, the frequency of all cytosine base-calls at reference cytosine positions in the lambda genome (unmethylated spike in control) was normalized by the total number of base-calls at reference cytosine positions in the lambda genome (Table S1).

TAB-seq

TAB-seq library generation was performed with the 5hmC TAB-seq Kit (WiseGene, USA - Cat# K001) kit as per manufacturer's instructions. 5-hydroxymethylated pUC19 DNA (WiseGene, USA - Cat# S002) was used as the hmC standard for the estimate of β-glucose protection of hmC from Tet conversion, whereas lambda phage DNA with methylation of all cytosines at CpG sites

(WiseGene, USA - Cat# S001) was used as the 5mC/C spike-in control. Calculation of the level of hmC was performed as described previously {Lister, 2013 #6}.

ATAC-seq

ATAC-seq experiments were performed as previously described {Buenrostro, 2013 #149}. Briefly, 10 zebrafish embryos were manually dechorionated and disrupted in 500 µl 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 µM Tris pH7.4, 10 µM NaCl, 3 µM MgCl2, 0.1% IGEPAL) and incubated for 30 min at 37ºC with the TDE1 enzyme. The sample was then purified with Qiagen Minelute kit, and a PCR was performed with 13 cycles using Ad1F and Ad2.1R primers and KAPA HiFi hotstart enzyme (Kapa Biosystems). Reads were aligned using zebrafish danRer7 assembly as the reference genome. Duplicated pairs or those ones separated by more than 2Kb were removed. The enzyme cleavage site was determined as the position -4 (minus strand) or +5 (plus strand) from each read start, and this position was extended 5 bp in both directions.

Identification of CG Differentially Methylated Regions (CG-DMRs)

CG-DMRs were identified as described previously {Lister, 2013 #6} with the following differences: the *P* values were simulated using 5000 permutations. The largest *P* value cut-off was chosen that still satisfied the 0.05 FDR requirement. Significant 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 10 differentially methylated sites were discarded. Finally, CG-DMR blocks were filtered based on a requirement for a minimum number of samples to all show the same significant differential methylation patterns.

ChIP-seq and Bio-CAP data analysis

Zebrafish, mouse and *Xenopus* CGIs/NMIs are part of a previously published dataset {Long, 2013 #93}. The raw reads corresponding to the zebrafish 24hpf post fertilization Bio-CAP sample (SRX/SRX217/SRX217160) were mapped to the zebrafish genome (Zv9/danRer7) using Bowtie 1.0.0 {Langmead, 2009 #103}, allowing up to two mismatches in the seed and retaining only uniquely mapped reads. Zebrafish ChIP-seq, H3K27ac, H3K4me3 and H3K27me3 reads were obtained from a previously published study {Bogdanovic, 2012 #52} (GSE32483). Mouse embryonic H3K4me1/H3K27ac/H3K4me3 mapped reads corresponding to mouse fetal brain, liver, limb and heart were obtained from a previously published study {Shen, 2012 #73}(GEO data GSE29184). The mapped reads were pulled together and analysed as one embryonic sample. The *Xenopus tropicalis* data was obtained from Hontelez *et al* (manuscript in preparation). The data is associated with the following GEO identifier: GSE67974 http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ilypiqmmlfidzol&acc=GSE67974).

RNA-seq data analysis

Data corresponding to zebrafish (GSE32900) {Pauli, 2012 #82}, mouse (GSE60334, GSM1502476, GSE47966) {Auclair, 2014 #85;Shen, 2014 #84;Lister, 2013 #6} and *Xenopus* {Paranjpe, 2013 #83} embryos were mapped using Kallisto (http://pachterlab.github.io/kallisto/) with default settings. Prior to mapping the mouse RNA-seq data, the reads were trimmed to 51 bp using the FASTX trimmer from the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) to normalize for read length differences among different studies. The reference transcriptomes were obtained from ENSEMBL and only canonical isoforms, as identified by ENSEMBL, were used. The Kallisto TPM (Transcripts Per Million) values were used to asses transcript abundance. Differential gene expression was performed as described before using the DEseq package {Anders, 2010 #169}. Only genes with an *FDR Q* value < 0.05 were considered significant.

Gene ontology analyses

Gene ontology analyses were performed using the GREAT tool {McLean, 2010 #80} (http://bejerano.stanford.edu/great/public/html/) or DAVID tool {Huang da, 2009 #165;Huang da, 2009 #166}, as indicated in the main text. For *Xenopus* GREAT analyses, we have calculated the GREAT regions as described before {McLean, 2010 #80} 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 TSS (regardless of other nearby genes, 5kb upstream, 1kb 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. The

XTEV *Xenopus* gene models (with human gene IDs) {Paranjpe, 2013 #135} were used for that purpose.

Expression analysis of orthologous genes

The genes corresponding to phylo-DMRs were identified using GREAT {McLean, 2010 #80}. The TPM values for zebrafish, mouse and *Xenopus* were calculated as described above and assigned to the orthologous genes identified through the ENSEMBL orthology tool {Vilella, 2009 #107}. The highest expression value for each gene within each species was assigned a value of "1" and the other TPM values were scaled accordingly. Such scaled expression values were clustered using the Ward method and Pearson correlation distance.

Morpholino knockdown of Tet proteins

The morpholinos (MOs) specific for each Tet protein are described in {Ge, 2014 #92}. We also designed a specific splice junction MO for the zebrafish Tet3 protein. (5'-TGCATGTCCACAGTAACTTACCACA-3′). Either 12 ng of this MO or 3 ng of a combination of all four MOs were injected in 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 dome (N = 100, 000) and 24 hpf (N = 25, 000) stages were collected and dechorionated as described before {Bogdanovic, 2013 #49}. Nuclear extracts were prepared using a previously described protocol for mass-spectrometry based proteomics {Cox, 2008 #86} with modifications. A minimal 1:10 ratio of tissue:lysis buffer was ensured during every homogenization. The embryonic tissue was homogenized using a Dounce homogenizer (20 X loose pestle, 20 X tight pestle). Batches of nuclear extract were snap frozen in liquid nitrogen and kept at -80°C. Before proceeding with DNA-pulldowns, the extracts were defrosted and combined into a single tube, following a 20 min centrifugation in a chilled (4°C) table top centrifuge.

DNA pull-downs and Mass spectrometry

DNA pull-downs were performed as described previously {Spruijt, 2013 #46}, except that streptavidin sepharose beads were used for the affinity purifications (GE Healthcare). Following incubations and washes, bound proteins were digested with trypsin. Tryptic peptides were desalted and concentrated using stage-tips {Rappsilber, 2003 #108} and applied to nanoLC (Proxeon) coupled online to an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific). 4 hour gradients (5% - 80% acetonitrile) were applied and the top 15 MS/MS spectra were recorded.

Proteomics data analysis

Raw data were analyzed using MaxQuant version 1.3.0.5 using default settings and the options label-free quantification and match between runs enabled {Cox, 2008 #86}. The uniprot *Danio rerio* database was used as reference proteome. The resulting 'proteingroups.txt' table was filtered for contaminants and reverse hits. Obtained label-free quantification (LFQ) intensities were log2 transformed and the proteins were filtered to have at least three valid values in one group (C, mC or hmC). For the resulting proteins, missing values were semi-random imputed by a normal distribution (width = 0.3 and shift = 1.8), based on the assumption that these proteins were under/close to the detection limit. To identify significant interactors, an adapted two-tailed t-test was performed (Persues software), which corrects for multiple testing by applying a permutation-based false discovery rate (FDR). Volcano plots were made in R, in which the LFQ ratio ((h)mC/C) is plotted against the calculated FDR (-log10). The FDR and s0 significance-threshold values used are depicted in these volcano plots (Supplementary Fig. 3). The obtained significant (h)mC readers were clustered on their enrichments ratios in R. iBAQ was performed as described {Schwanhausser, 2011 #109;Spruijt, 2013 #46}. In brief, 3.3 µg UPS2 standard (Sigma) was spiked in 10 µg nuclear extract. Filter-aided sample preparation (FASP) was performed {Wisniewski, 2009 #110} and the peptides were applied to LC-MS/MS. Linear regression was performed on the exact known amounts of the UPS2 standard proteins and their measured iBAQ intensities, followed by extrapolation of the absolute protein amounts for the zebrafish proteins. In parallel, 100 µg extract was digested using FASP and peptides were fractionated using strong anion exchange (SAX) in five fractions, which resulted in a deep-proteome. Proteins quantified in the single FASP sample were matched with the iBAQ intensities measured in the deep-proteome, which were used for linear regression and extrapolation of absolute quantification for all proteins in the measured proteome.

Isolation of zebrafish *sox10*+ cells

Neural crest cells were isolated from 24hpf zebrafish transgenic embryos, expressing mCherry under the control of *sox10* regulatory locus (TgBAC(Sox10:Cherry)), using Fluorescence Activated Cell Sorting (FACS). Embryos were dissociated to a single cell suspension using collagenase (20mg/ml)/trypsin (0.05%) solution at 30°C for 12 mins with three intermittent homogenization steps. Reaction was stopped with Hank's blocking solution (1X HBSS Ca-, Mg-, phenol red-free, 0.25% BSA, 10mM Hepes pH8), cells collected by centrifugation at 500g for 10 mins, re-suspended in Hanks solution and passed through a cell strainer to remove the cell aggregates. Single cells, concentrated by another centrifugation were resuspended in ~500ul 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 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, always in accordance with best practices outlined by the European Union.

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 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 Vermeulen lab is supported by a grant from the Netherlands Organisation for Scientific Research (NWO-VIDI) (864.09.003). J.R.E was supported by the Gordon and Betty Moore Foundation (GMBF3034) 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 Economia y Competitividad (BFU2011-23083), Comunidad Autónoma de Madrid (CELLDD-CM), and by the Pro-CNIC Foundation. This work has been supported by a grant of the US National Institutes of Health (NICHD, grant R01HD069344) to GJCV.

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 analysed by O.B. with the help of R.L., E.F., M.S and J.R.E. Embryo work was performed by O.B., E. d.l. C. M., J.J.T, T.R., M.M., and J.L.G-S. Zebrafish *sox10*+ line was prepared by R.W., U.S., and T. S-S. *Xenopus* ChIP-seq data was generated by S.H., I. v. K., and G.J.C.V. Quantitative interaction proteomics experiments were performed by A.S., F.G., T.C., and M.V. Proteomics data were analysed by A.S. and M.V. The study was written by O.B, A.S., M.V, J.L.G-S and R.L.

FIGURE LEGENDS

Figure 1 DNA methylome dynamics during vertebrate embryogenesis and the phylotypic stage. **(a)** Global DNA methylation (mC) levels at four stages of *Xenopus tropicalis* embryogenesis, zebrafish post-ZGA embryogenesis, and mouse embryogenesis. **(b)** Number, directionality and developmental stage of differentially methylated regions (DMRs, FDR = 0.05) identified in zebrafish, *Xenopus* and mouse embryos. The regions of developmental demethylation are denoted as phylo-DMRs. **(c)** Conserved gene ontology (Biological Process) terms associated with phylo-DMRs in zebrafish *Xenopus* and mouse. **(d)** Progressive demethylation of phylo-DMRs in embryonic tissues and adult brains. **(e)** mC levels of phylo-DMRs in adult mouse tissues.

Figure 2 Phylo-DMRs correspond to evolutionarily conserved developmentally activated enhancers associated with vertebrate body plan formation. **(a)** Sorted heatmaps of mC and normalized ChIP-seq read density for H3K4me1, H3K4me3, H3K27ac, p300 at phylo-DMRs in zebrafish, *Xenopus* and mouse embryos. **(b)** Mean CpG density (CpGs / 100bp) in zebrafish, *Xenopus* and mouse phylo-DMRs. **(c)** Comparisons of mean CpG density (CpGs / 100bp) in phylo-DMRs and CpG islands. **(d)** Genomic overlaps and expression patterns of validated VISTA enhancers associated with phylo-DMRs. **(e)** Hierarchical clustering of transcript abundance (scaled TPMs) of orthologous genes (N = 211) linked to phylo-DMRs. (**f)** Evolutionary conservation (aggregate PhastCons scores) of DMRs in zebrafish and mouse embryos (Kruskal – Wallis test, $P < 0.01$).

Figure 3 Active demethylation components bind mC/hmC during the phylotypic period in vertebrates. **(a)** Total numbers of nuclear proteins found to interact with C, mC or hmC oligos. *P*-values (hypergeometric test) represent the significance of interspecies correlations. (**b)** Enriched GO terms in the conserved C, mC or hmC readers for pluripotent (dome, ESCs) and differentiated stages (NPCs, 24hpf). **(c)** Hierarchical correlation-based clustering of the enrichment of DNA demethylation and DNA repair linked mC / hmC readers identified in zebrafish embryos and mouse cell nuclear extracts.

Figure 4 Phylo-DMRs are characterized by hmC enrichment in vertebrate embryos**. (a)** Steady state abundance (TPMs) of Tet1/2/3 transcripts during zebrafish, *Xenopus* and mouse embryogenesis. (**b)** Genome browser display demonstrating the co-occurrence of phylo-DMRs, hydroxymethylation (hmC, TAB-seq) and active enhancer marks. **(c)** Average profiles of hmC and mC levels over phylo-DMRs extended to -3kb/3kb (zebrafish, *Xenopu*s) or -10kb/10kb (mouse).

Figure 5 Tet proteins are required for phylo-DMR demethylation and body plan formation in vertebrates. **(a)** Correlation of mC levels in wild type and Tet3 and Tet1/2/3 knockdown embryos genome wide (1 kb non-overlapping windows), in CpG islands (CGIs) and in phylo-DMRs. **(b)** Average profiles (normalized read density) of ATAC-seq signal over phylo-DMRs in wild type and *tet1/2/3* morphant embryos. **(c)** Reduced ATAC-seq signal (Kruskal Wallis test, $P < 0.01$) in phylo-DMRs when compared to a general population of PDREs identified in 24hpf zebrafish embryos. **(d)** Number of regions displaying significantly altered (Fisher's exact test, *FDR Q* < 0.05) chromatin accessibility in *tet1/2/3* mutants.

Supplementary Figure 1 DMR size, distribution and dynamics**. (a)** Size distribution of DMRs in zebrafish, *Xenopu*s and mouse. **(b)** Numbers and directionality of identified DMRs in zebrafish, *Xenopus* and mouse

Supplementary Figure 2 Chromatin state and examples of zebrafish DMRs. **(a)** Sorted heatmaps of normalized ChIP-seq read density for H3K4me1, H3K4me3, H3K27ac histone marks at phylo-DMRs and early zebrafish (1K cell vs 80% epiboly) DMRs. **(b)** Genome browser examples of zebrafish promoter DMRs.

Supplementary Figure 3 KEGG pathway enrichment of phylo-DMR-linked genes. Gene ontology (GO) enrichments (KEGG pathway) of orthologous genes linked to zebrafish, *Xenopus* and mouse phylo-DMRs.

Supplementary Figure 4 Stage-specific mC/hmC/C readers and differential protein expression analyses. **(a – c)** The ratio between the measured label-free quantification (LFQ) intensities of the mC or hmC bait over the C bait (x-axis) is plotted against the –log false discovery rate (FDR) (y-axis) as determined by an adapted two-tailed t-test. The used FDR and s0 (the weight of the ratio in the calculation of the significance-threshold) are depicted in the lower-left corner of the plots. The black dots represent the significant proteins for the unmodified (left side of the plot) and modified (right side of the plot) baits. (**d)** The linear regression curves for the UPS2 standard (amount versus iBAQ intensity) and for the FASP iBAQ (amount versus SAX iBAQ intensity) in both the dome and 24hpf nuclear extract (top panel). **(e)** Differential expression analysis of the quantified proteins (4960 in at least one sample) yield enriched GO terms for dome and 24hpf specific proteins.

Supplementary Figure 5 Dynamic c/mC/hmC readers during zebrafish embryogenesis. **(a – b)** Schematic representation of quantitative interaction proteomics and nuclear extract quantitation approaches used in this study. **(c)** Hierarchical correlation-based clustering of the enrichment of mC readers in dome and 24hpf zebrafish extracts and hmC readers in 24hpf zebrafish extracts. Colors indicate the extent of enrichment; white to red indicates no enrichment to high enrichment. Absolute abundance of the readers is depicted in the last two columns, in which yellow to blue indicates sub-femtomole to picomoles per 100 μg extract. Components of active DNA demethylation/DNA repair pathways are indicated by dotted lines.

Supplementary Figure 6 Evolutionary conservation of developmental c/mC/hmC/C readers in zebrafish and mouse. Conserved C and mC readers in mESC and zebrafish dome embryos (left panel) and conserved C, mC and hmC readers in mouse NPC and zebrafish 24hpf embryos (right panel). Colors indicate the degree of binding conservation between the two organisms; purple – significant; blue just below the significance threshold; green – enrichment; brown - family members of this protein are found to be significantly enriched; salmon - protein bears a domain that is significantly enriched in readers for this modification in both organisms. The names are zebrafish gene names, however the homologue mouse gene name is used in cases with no informative zebrafish name available (italic).

Supplementary Figure 7 Tet1/2/3 depletion in zebrafish embryos. **(a)** Morpholino (MO) Tet3 and Tet1/2/3 knockdown phenotypes in zebrafish. **(b)** Sorted heatmap of CGIs (Kdm2b CxxC BIOcap signal) over zebrafish phylo-DMRs. (**c)** Genome browser displays of zebrafish phylo-DMR examples associated with key developmental regulators (sox21b and neurog1). **(d)**

Hierarchical clustering of gene expression profiles (TPMs) of zebrafish *tet1/2/3* and wild type embryos.

Supplementary Table 1 Overview of (hydroxy)methylome data used in this study. Organism, library type, developmental stage (tissue), sequencing statistics, reference genome and data-related references.

Supplementary Tables 2 – 4 Genomic positions and directionality of phylo-DMRs. Genomic positions and directionality of statistically significant (*FDR* = 0.05) phylo-DMRs in zebrafish (Supplementary Table 2), *Xenopus* (Supplementary Table 3) and mouse (Supplementary Table 4).

Supplementary Table 5 - 6 Conserved gene ontology (biological Process) enrichments (*FDR* = 0.01) of early (gastrula – phylotypic stage, Supplementary Table 5) and late (phylotypic stage – tailbud/fetus, Supplementary Table 6) phylo-DMRs.

Supplementary Table 7 Orthologous genes associated with phylo-DMRs in zebrafish, *Xenopus* and mouse. Mouse gene IDs corresponding to orthologous genes (*N* = 211) linked to phylo-DMRs in all the three species.

Supplementary Table 8 Genomic positions of VISTA enhancers overlapping phylo-DMRs. Genomic positions (mouse mm10 genome reference) of previously validated VISTA enhancers overlapping with phylo-DMRs in mouse.

Supplementary Table 9 KEGG pathway enrichments of orthologous phylo-DMR linked genes. Only statistically significant enrichments (Benjamini Hochberg FDR < 0.05) are displayed.

Supplementary Tables 10 – 11 Label-free quantification (LFQ) intensities for zebrafish dome and 24hpf samples. Proteins displaying significantly different enrichment between unmethylated and methylated DNA oligonucleotides in zebrafish dome (Supplementary table 9) and zebrafish 24hpf (Supplementary Table 10) samples were identified by an adapted t-test on label-free quantification (LFQ) intensities.

Supplementary Table 12 Differentially expressed genes (*FDR* < 0.05) between zebrafish *tet1/2/3* morphant and wild type embryos.

Figure 1

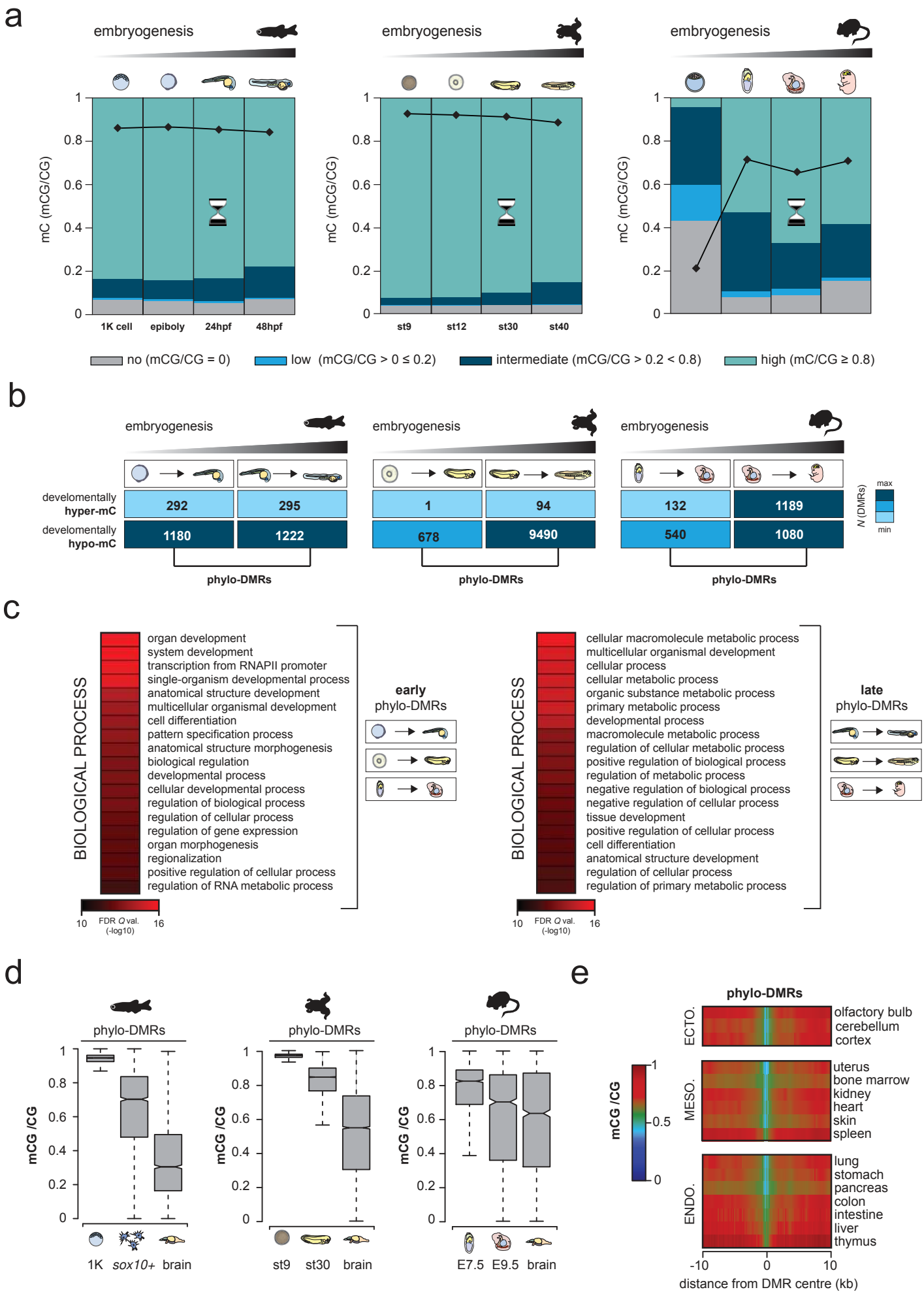

a
embryogenesis
mC (mCG/CG)
1K cell
epiboly
24hpf
48hpf
st9
st12
st30
st40
no (mCG/CG = 0)
low (mCG/CG > 0 ≤ 0.2)
intermediate (mCG/CG > 0.2 < 0.8)
high (mC/CG ≥ 0.8)
b
embryogenesis
develomentally hyper-mC
develomentally hypo-mC
292
295
1180
1222
1
94
678
9490
132
1189
540
1080
phylo-DMRs
N (DMRs)
max
min
c
BIOLOGICAL PROCESS
organ development
system development
transcription from RNAPII promoter
single-organism developmental process
anatomical structure development
multicellular organismal development
cell differentiation
pattern specification process
anatomical structure morphogenesis
biological regulation
developmental process
cellular developmental process
regulation of biological process
regulation of cellular process
regulation of gene expression
organ morphogenesis
regionalization
positive regulation of cellular process
regulation of RNA metabolic process
early phylo-DMRs
FDR Q val. (-log10)
cellular macromolecule metabolic process
multicellular organismal development
cellular process
cellular metabolic process
organic substance metabolic process
primary metabolic process
developmental process
macromolecule metabolic process
regulation of cellular metabolic process
positive regulation of biological process
regulation of metabolic process
negative regulation of biological process
negative regulation of cellular process
tissue development
positive regulation of cellular process
cell differentiation
anatomical structure development
regulation of cellular process
regulation of primary metabolic process
late phylo-DMRs
d
phylo-DMRs
mCG /CG
1K sox10+ brain
st9 st30 brain
E7.5 E9.5 brain
e
phylo-DMRs
ECTO.
olfactory bulb
cerebellum
cortex
MESO.
uterus
bone marrow
kidney
heart
skin
spleen
ENDO.
lung
stomach
pancreas
colon
intestine
liver
thymus
mCG /CG
distance from DMR centre (kb)

Figure 2

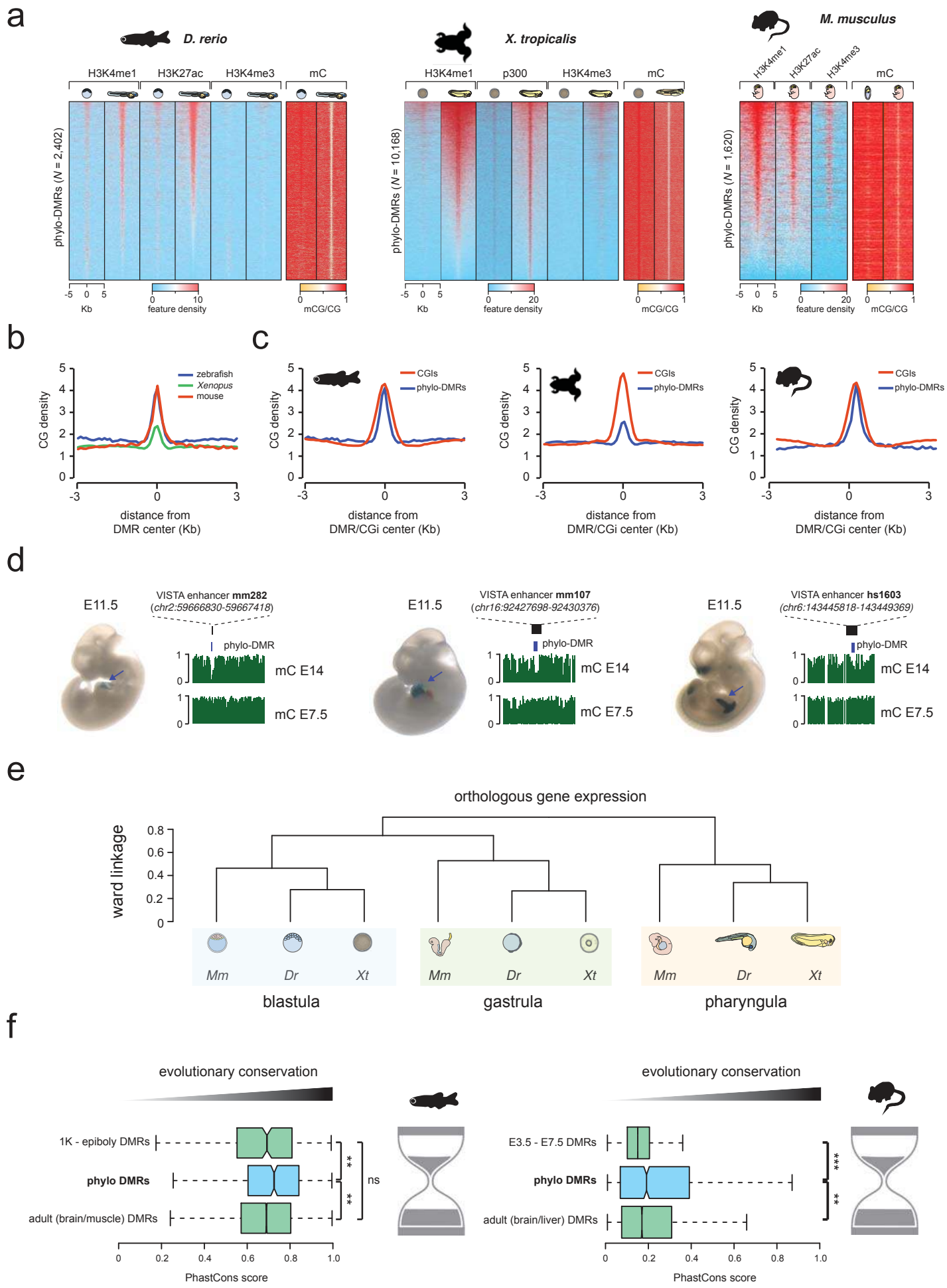

a
D. rerio
X. tropicalis
M. musculus
H3K4me1
H3K27ac
H3K4me3
p300
mC
phylo-DMRs (N = 2,402)
phylo-DMRs (N = 10,168)
phylo-DMRs (N = 1,620)
Kb
feature density
mCG/CG
b
c
zebrafish
Xenopus
mouse
CGIs
phylo-DMRs
CG density
distance from DMR center (Kb)
distance from DMR/CGi center (Kb)
d
E11.5
VISTA enhancer mm282 (chr2:59666830-59667418)
VISTA enhancer mm107 (chr16:92427698-92430376)
VISTA enhancer hs1603 (chr6:143445818-143449369)
phylo-DMR
mC E14
mC E7.5
e
orthologous gene expression
ward linkage
Mm
Dr
Xt
blastula
gastrula
pharyngula
f
evolutionary conservation
1K - epiboly DMRs
phylo DMRs
adult (brain/muscle) DMRs
adult (brain/liver) DMRs
E3.5 - E7.5 DMRs
ns
PhastCons score

Figure 3

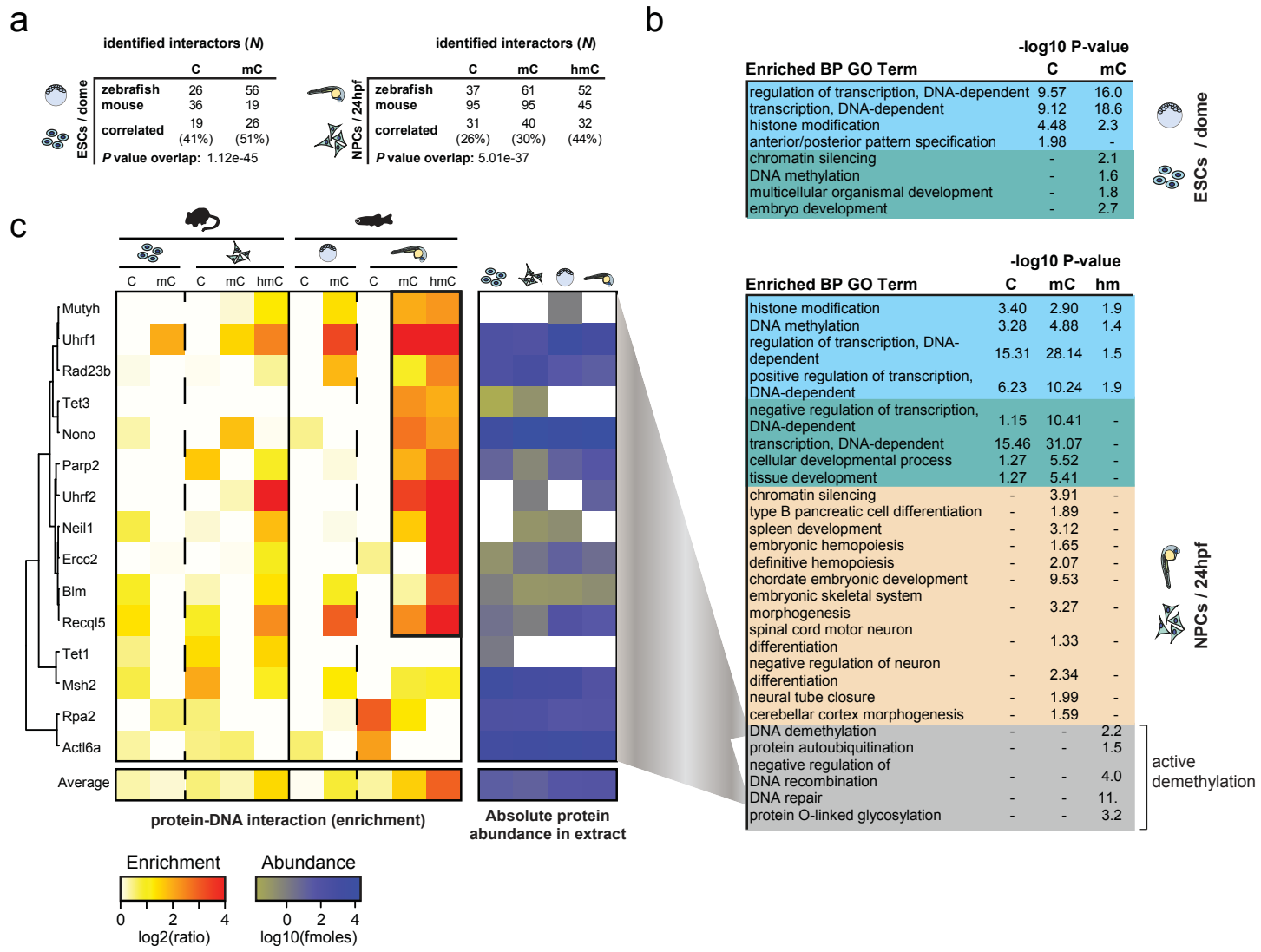
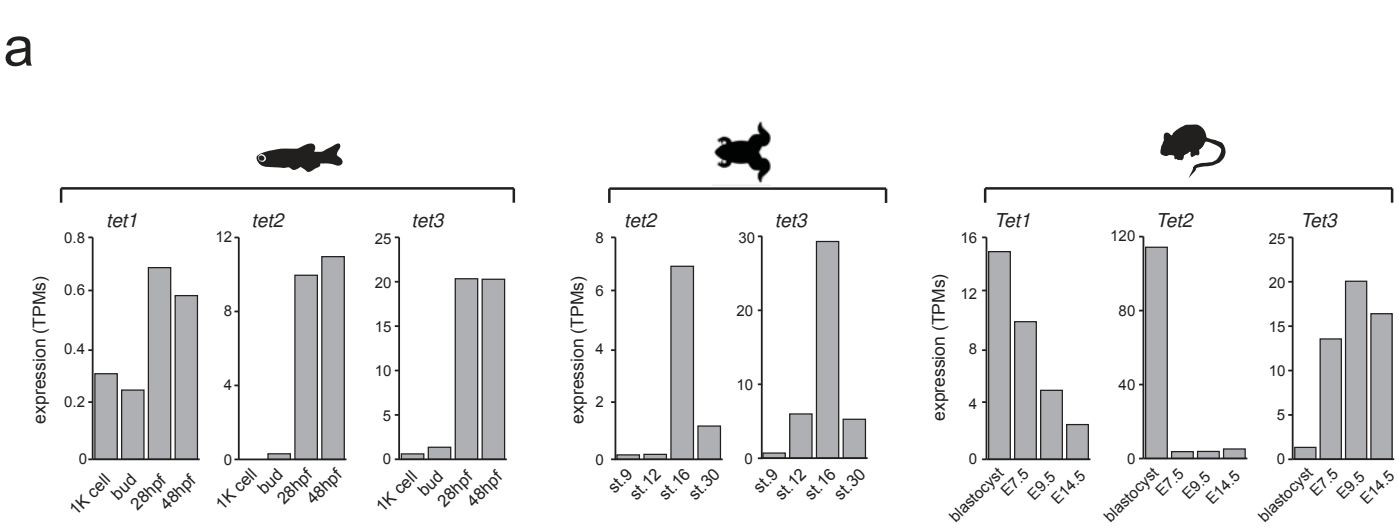
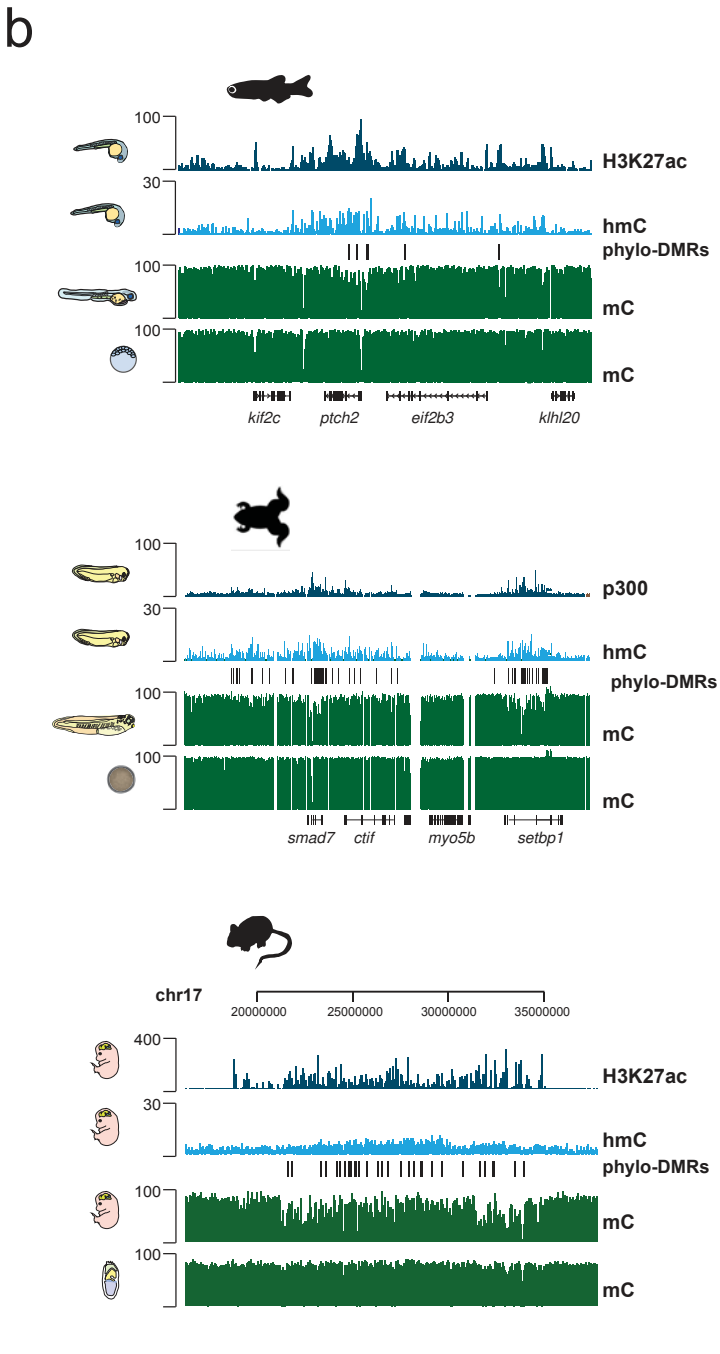

a
identified interactors (N)
ESCs / dome
C mC
zebrafish 26 56
mouse 36 19
correlated 19 (41%) 26 (51%)
P value overlap: 1.12e-45
identified interactors (N)
NPCs / 24hpf
C mC hmC
zebrafish 37 61 52
mouse 95 95 45
correlated 31 (26%) 40 (30%) 32 (44%)
P value overlap: 5.01e-37
b
-log10 P-value
Enriched BP GO Term C mC
regulation of transcription, DNA-dependent 9.57 16.0
transcription, DNA-dependent 9.12 18.6
histone modification 4.48 2.3
anterior/posterior pattern specification 1.98 -
chromatin silencing - 2.1
DNA methylation - 1.6
multicellular organismal development - 1.8
embryo development - 2.7
ESCs / dome
-log10 P-value
Enriched BP GO Term C mC hm
histone modification 3.40 2.90 1.9
DNA methylation 3.28 4.88 1.4
regulation of transcription, DNA-dependent 15.31 28.14 1.5
positive regulation of transcription, DNA-dependent 6.23 10.24 1.9
negative regulation of transcription, DNA-dependent 1.15 10.41 -
transcription, DNA-dependent 15.46 31.07 -
cellular developmental process 1.27 5.52 -
tissue development 1.27 5.41 -
chromatin silencing - 3.91 -
type B pancreatic cell differentiation - 1.89 -
spleen development - 3.12 -
embryonic hemopoiesis - 1.65 -
definitive hemopoiesis - 2.07 -
chordate embryonic development - 9.53 -
embryonic skeletal system morphogenesis - 3.27 -
spinal cord motor neuron differentiation - 1.33 -
negative regulation of neuron differentiation - 2.34 -
neural tube closure - 1.99 -
cerebellar cortex morphogenesis - 1.59 -
DNA demethylation - - 2.2
protein autoubiquitination - - 1.5
negative regulation of DNA recombination - - 4.0
DNA repair - - 11.
protein O-linked glycosylation - - 3.2
NPCs / 24hpf
active demethylation
c
C mC C mC hmC C mC C mC hmC
Mutyh
Uhrf1
Rad23b
Tet3
Nono
Parp2
Uhrf2
Neil1
Ercc2
Blm
Recql5
Tet1
Msh2
Rpa2
Actl6a
Average
protein-DNA interaction (enrichment)
Absolute protein abundance in extract
Enrichment
0 2 4
log2(ratio)
Abundance
0 2 4
log10(fmoles)

Figure 4


a
tet1
tet2
tet3
expression (TPMs)
1K cell
bud
28hpf
48hpf
st.9
st.12
st.16
st.30
Tet1
Tet2
Tet3
blastocyst
E7.5
E9.5
E14.5


b
H3K27ac
hmC
phylo-DMRs
mC
kif2c
ptch2
eif2b3
klhl20
p300
smad7
ctif
myo5b
setbp1
chr17
20000000
25000000
30000000
35000000

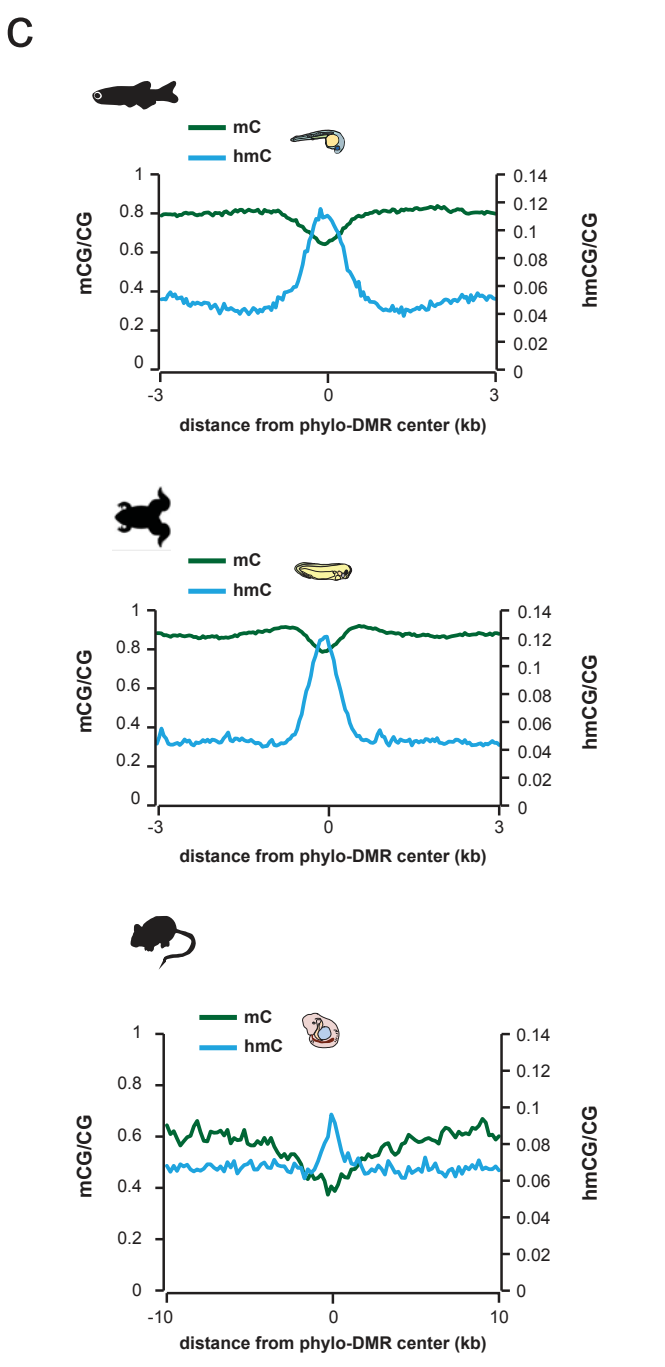
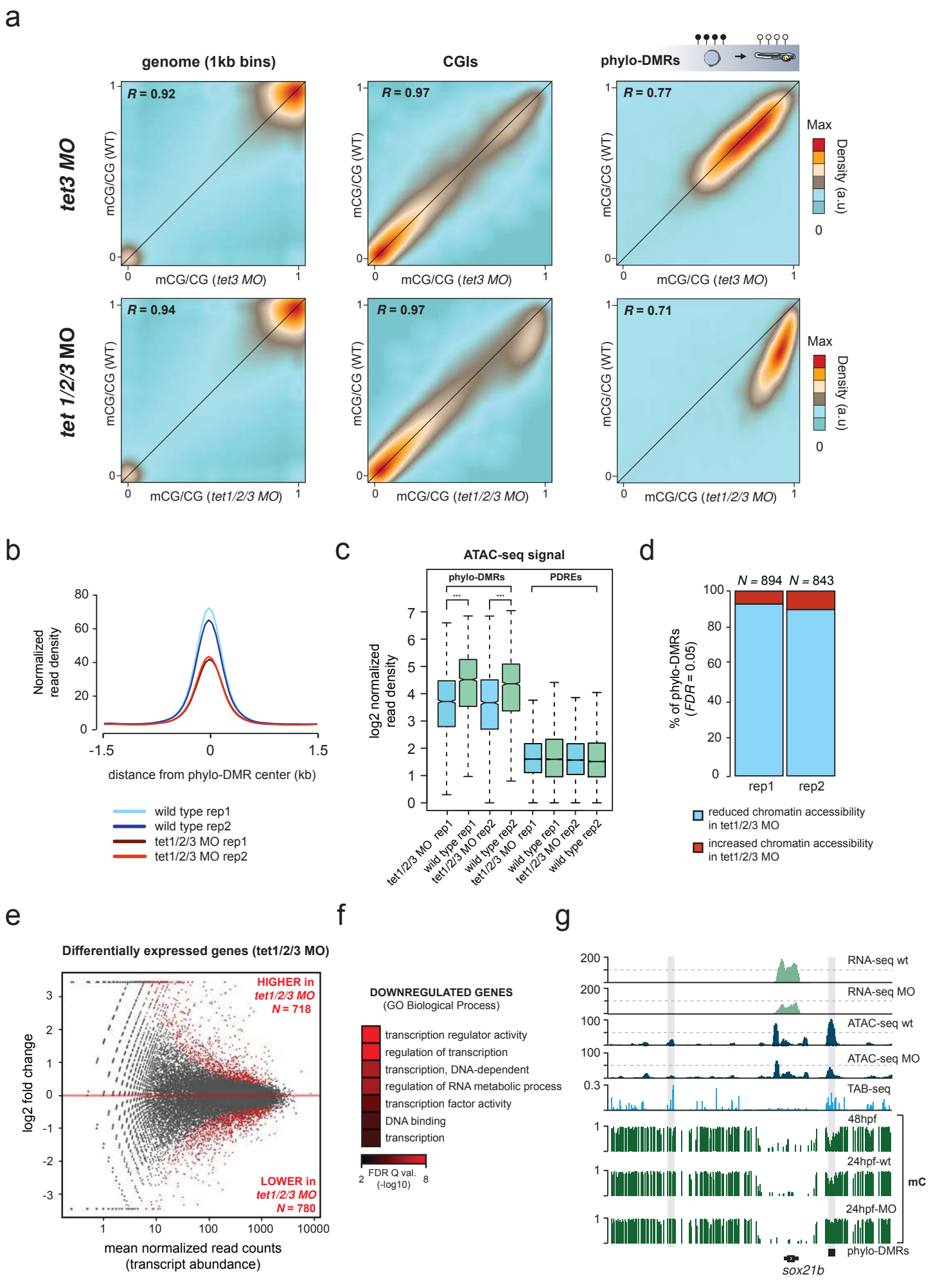

c
mC
hmC
mCG/CG
hmCG/CG
distance from phylo-DMR center (kb)

Figure 5


a
genome (1kb bins)
CGIs
phylo-DMRs
tet3 MO
tet 1/2/3 MO
R = 0.92
R = 0.97
R = 0.77
R = 0.94
R = 0.97
R = 0.71
mCG/CG (WT)
mCG/CG (tet3 MO)
mCG/CG (tet1/2/3 MO)
Max
Density (a.u)
0
b
Normalized read density
distance from phylo-DMR center (kb)
wild type rep1
wild type rep2
tet1/2/3 MO rep1
tet1/2/3 MO rep2
c
ATAC-seq signal
phylo-DMRs
PDREs
log2 normalized read density
d
N = 894
N = 843
% of phylo-DMRs (FDR = 0.05)
rep1
rep2
reduced chromatin accessibility in tet1/2/3 MO
increased chromatin accessibility in tet1/2/3 MO
e
Differentially expressed genes (tet1/2/3 MO)
HIGHER in tet1/2/3 MO N = 718
LOWER in tet1/2/3 MO N = 780
log2 fold change
mean normalized read counts (transcript abundance)
f
DOWNREGULATED GENES
(GO Biological Process)
transcription regulator activity
regulation of transcription
transcription, DNA-dependent
regulation of RNA metabolic process
transcription factor activity
DNA binding
transcription
FDR Q val. (-log10)
g
RNA-seq wt
RNA-seq MO
ATAC-seq wt
ATAC-seq MO
TAB-seq
48hpf
24hpf-wt
24hpf-MO
mC
phylo-DMRs
sox21b