

Supplementary Materials for
**Active DNA demethylation of developmental cis-regulatory regions predates
vertebrate origins**

Ksenia Skvortsova *et al.*

Corresponding author: Ksenia Skvortsova, k.skvortsova@garvan.org.au;
Ozren Bogdanovic, o.bogdanovic@gmail.com

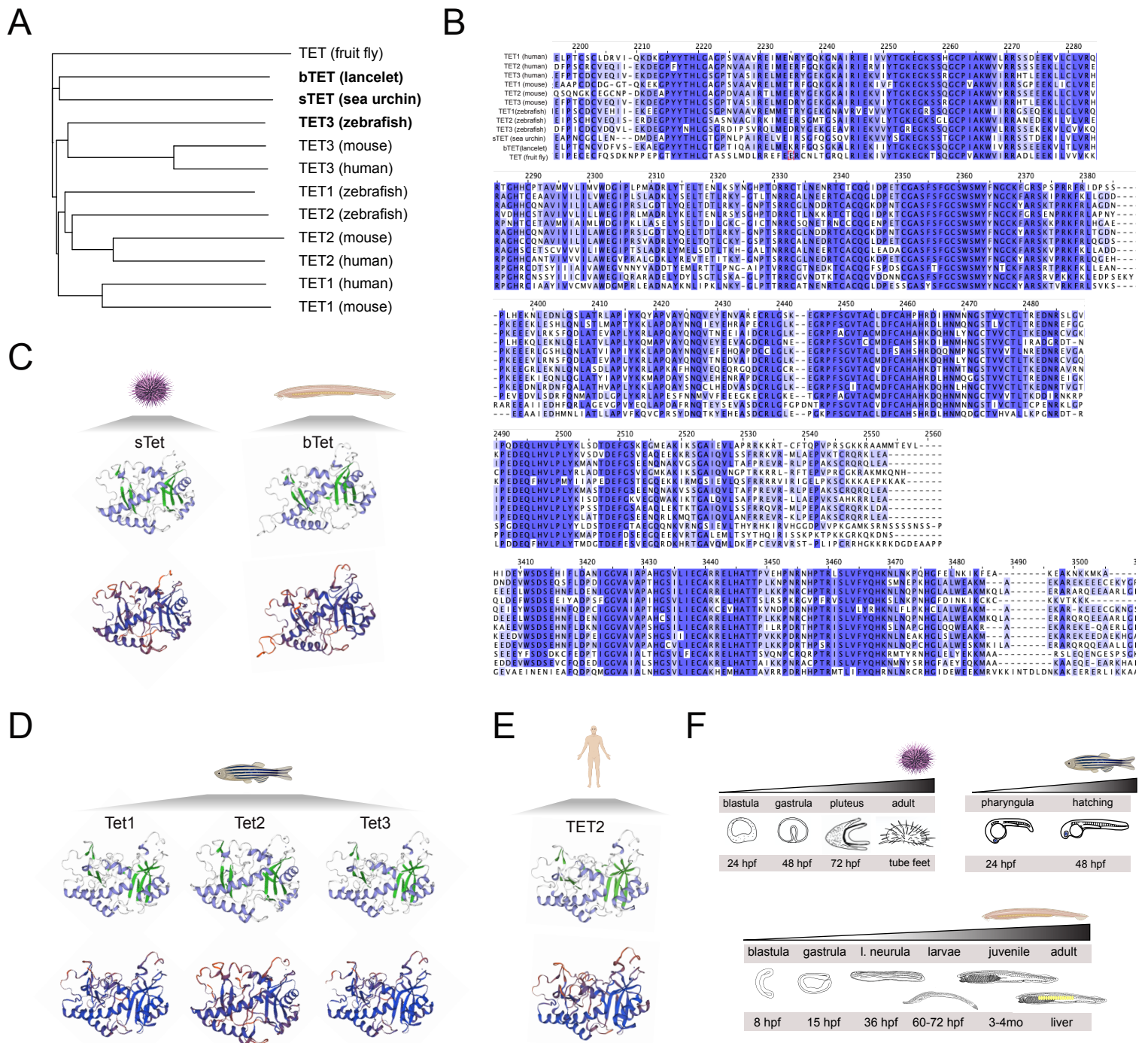
Sci. Adv. **8**, eabn2258 (2022)
DOI: 10.1126/sciadv.abn2258

The PDF file includes:

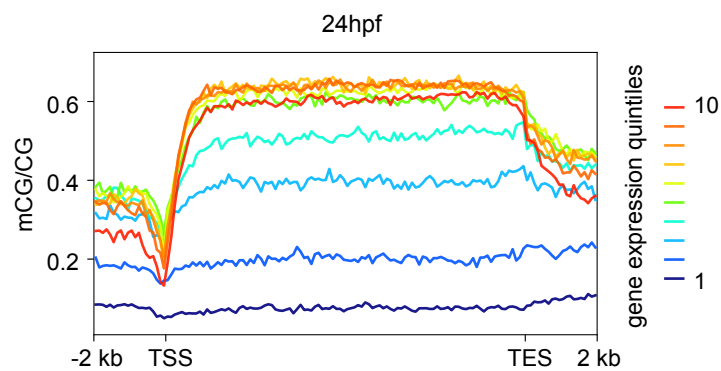
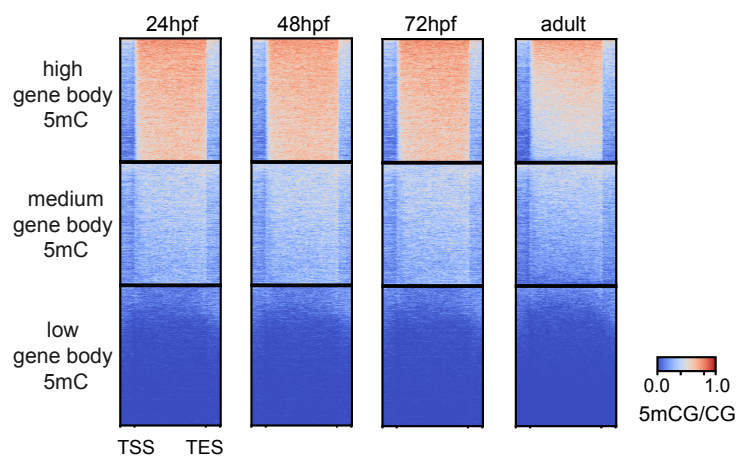
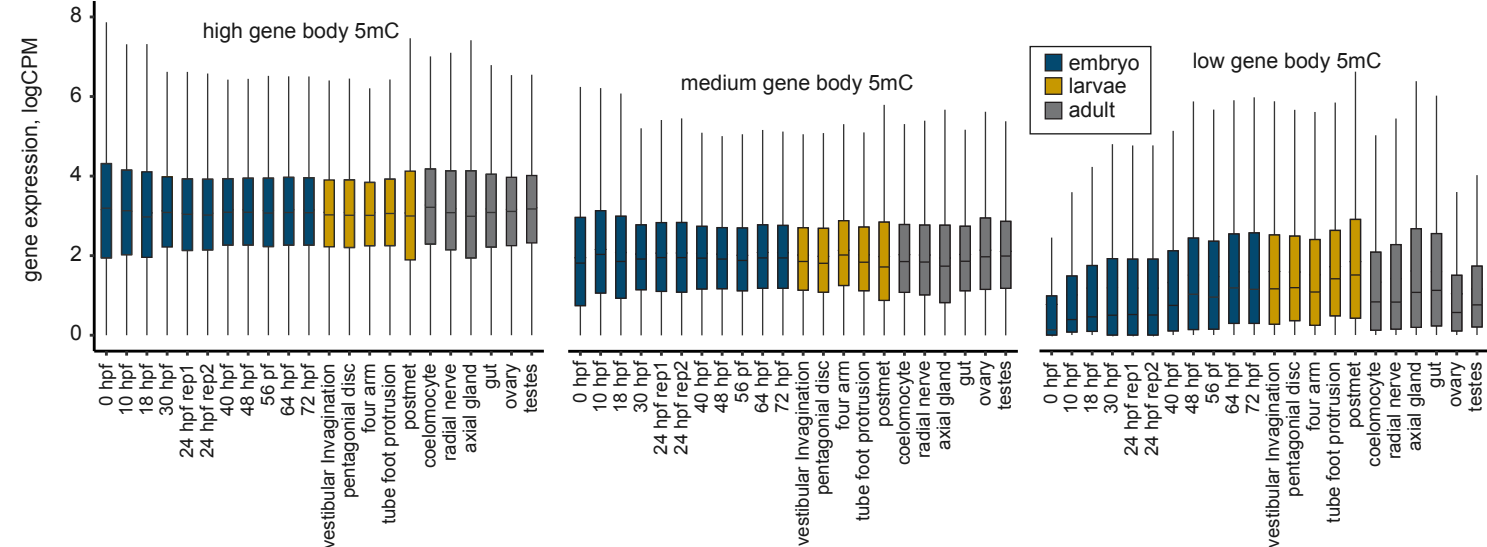
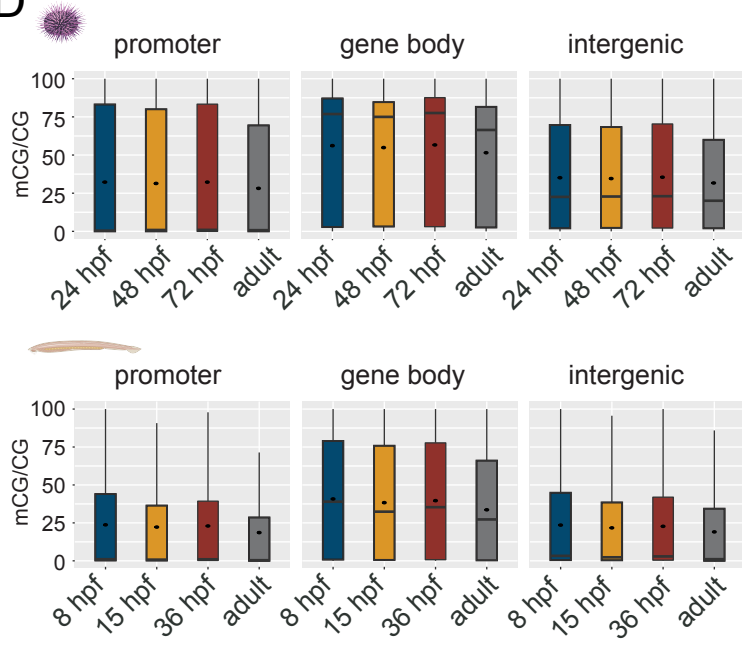
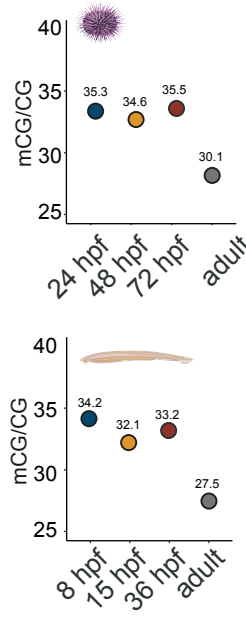
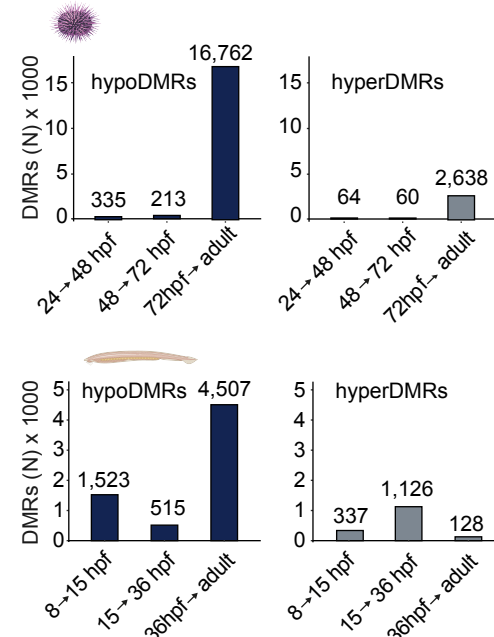
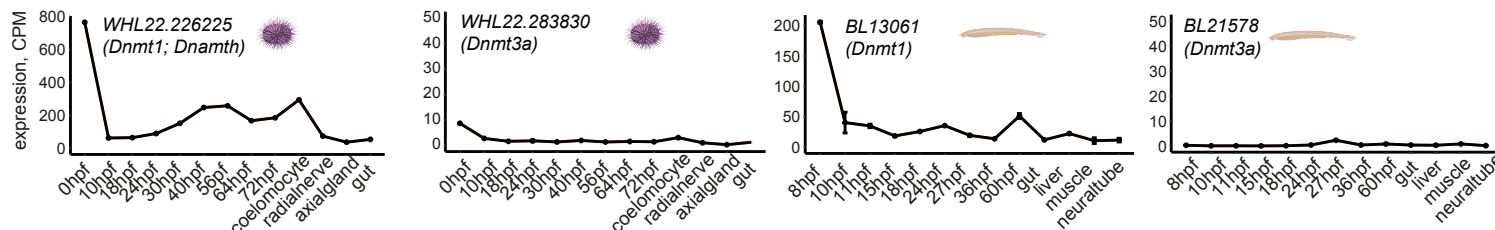
Figs. S1 to S8
Legends for tables S1 to S13

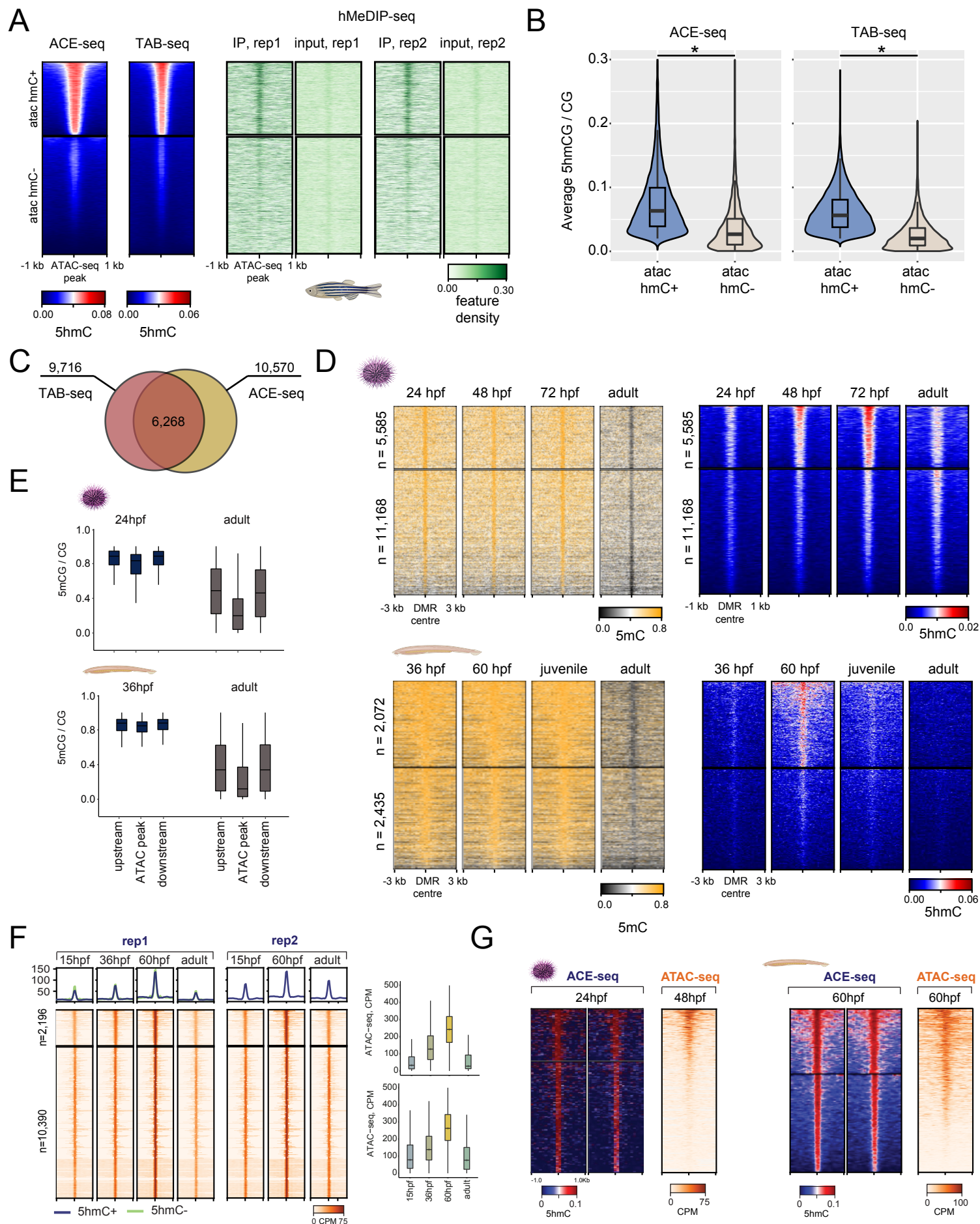
Other Supplementary Material for this manuscript includes the following:

Tables S1 to S13



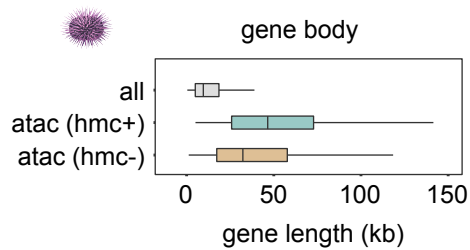
Supplementary Figure 1. Conservation of TET protein sequence and catalytic domain structure. (A) Cladogram depicting the phylogenetic relationships between fruit fly, sea urchin, lancelet, zebrafish, mouse, and human TET proteins. **(B)** Multiple sequence alignment of human, mouse and zebrafish TET1, TET2, TET3 as well as sea urchin, lancelet and fruit fly TET DSBH domains. DSBH domains harbour a large low-complexity insert, which is not shown. The colour of each amino acid indicates percentage identity (PID) with darker blue depicting higher PID and lighter blue - lower PID. **(C-E)** 3D models of methylcytosine dioxygenase domains of sea urchin sTet, lancelet bTet, zebrafish Tet1, Tet2, Tet3, and human TET2, performed using SWISS-MODEL. Top, 3D-model coloured by secondary structure: α -helix is depicted in blue and β -sheet in green. Bottom: 3D-model coloured by “Confidence”, QMEANDisCo local quality score depicting the expected similarity of each residue of the model to the native structure. High confidence is shown in blue, low confidence is shown in red. **(F)** Developmental stages of sea urchin, lancelet, and zebrafish embryonic stages and adult tissues used in the study.

A**B****C****D****E****G****F**

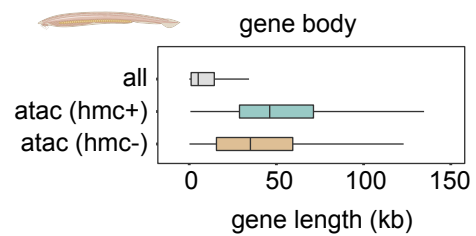


Supplementary Figure 3. Comparison of 5hmC profiling approaches and 5mC/5hmC dynamics at DMRs. (A) 5hmC enrichment as identified by ACE-seq, TAB-seq and hMeDIP-seq plotted over zebrafish ATAC-seq regions identified at 24hpf (Bogdanovic et al, 2016). K-means clustering (k=2) was performed to identify ATAC-seq regions with and without 5hmC. **(B)** Average 5hmC levels within ATAC hmC+ and ATAC hmC- regions. Wilcoxon rank-sum test was performed to compare average 5hmC at ATAC hmC+ and ATAC hmC- regions. **(C)** Overlap between ATAC hmC+ regions identified by ACE-seq (av. 5hmC > 5%) and TAB-seq (av. 5hmC > 5%). **(D)** 5mC and 5hmC dynamics (ACE-seq) at embryo-adult DMRs in sea urchin and lancelet. K-means clustering of 5hmC signal. **(E)** Quantification of developmental 5mC loss at ATAC-seq peaks corresponding to cluster 1 (Fig. 3E, F), and surrounding regions (500 bp up- and downstream). **(F)** Developmental profiles of open chromatin (ATAC-seq) during lancelet development (cluster 1 = hmC +, cluster 2 = 5hmC -). **(G)** 5hmC and ATAC-seq signal at ACE-seq peaks in sea urchin and lancelet embryos.

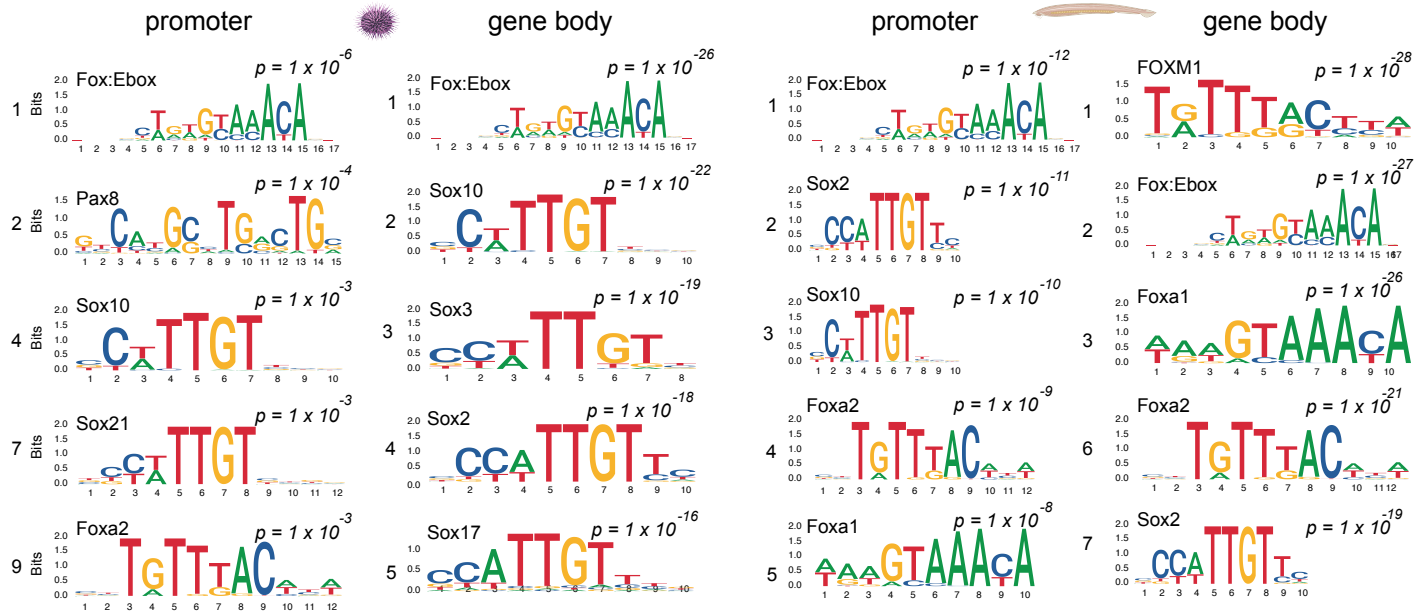
A



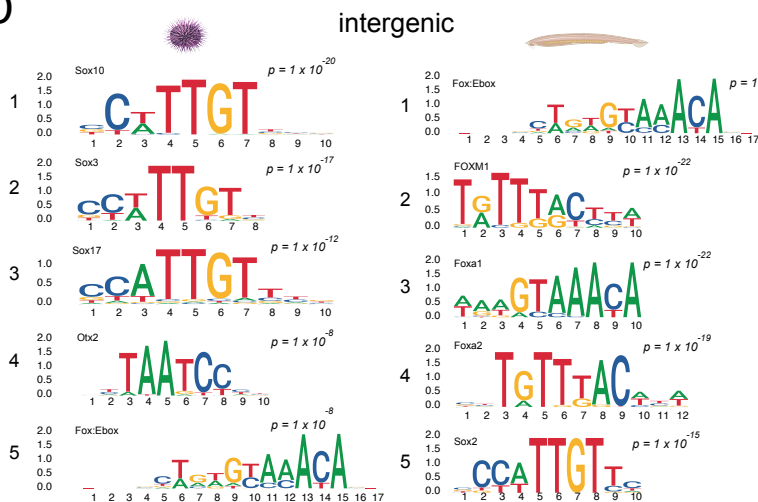
B



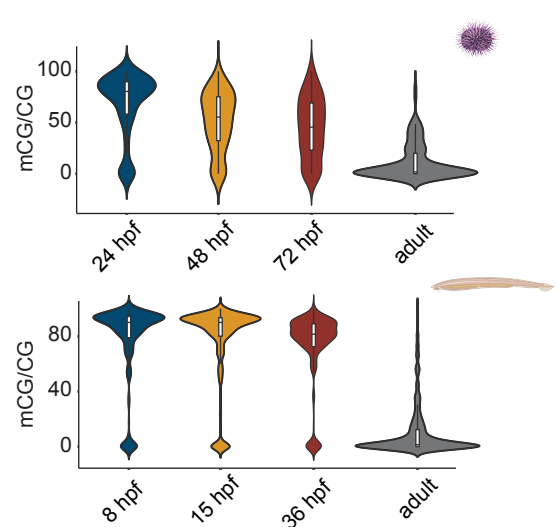
C



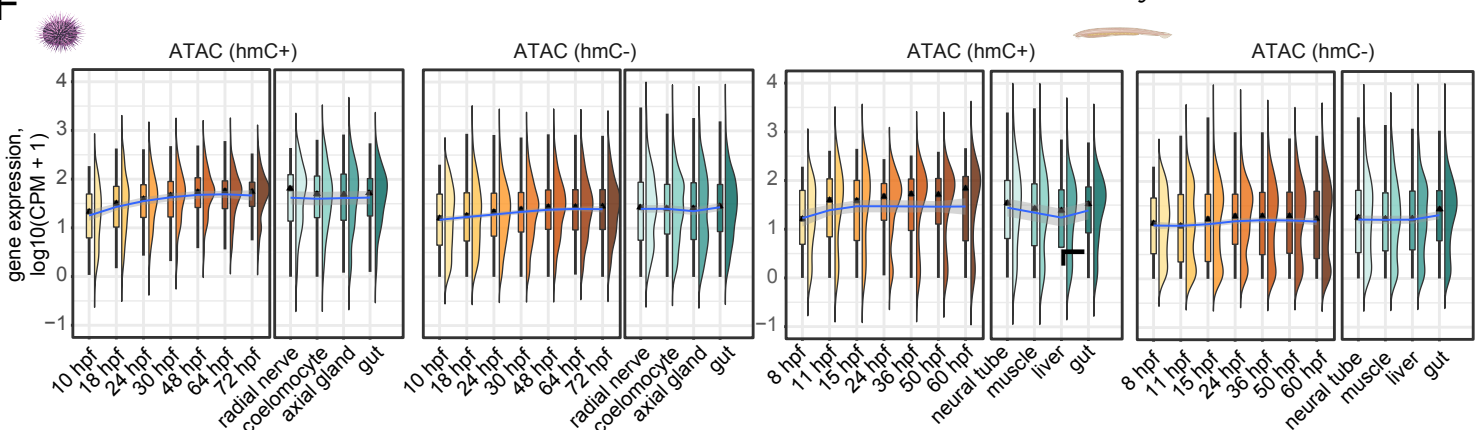
D



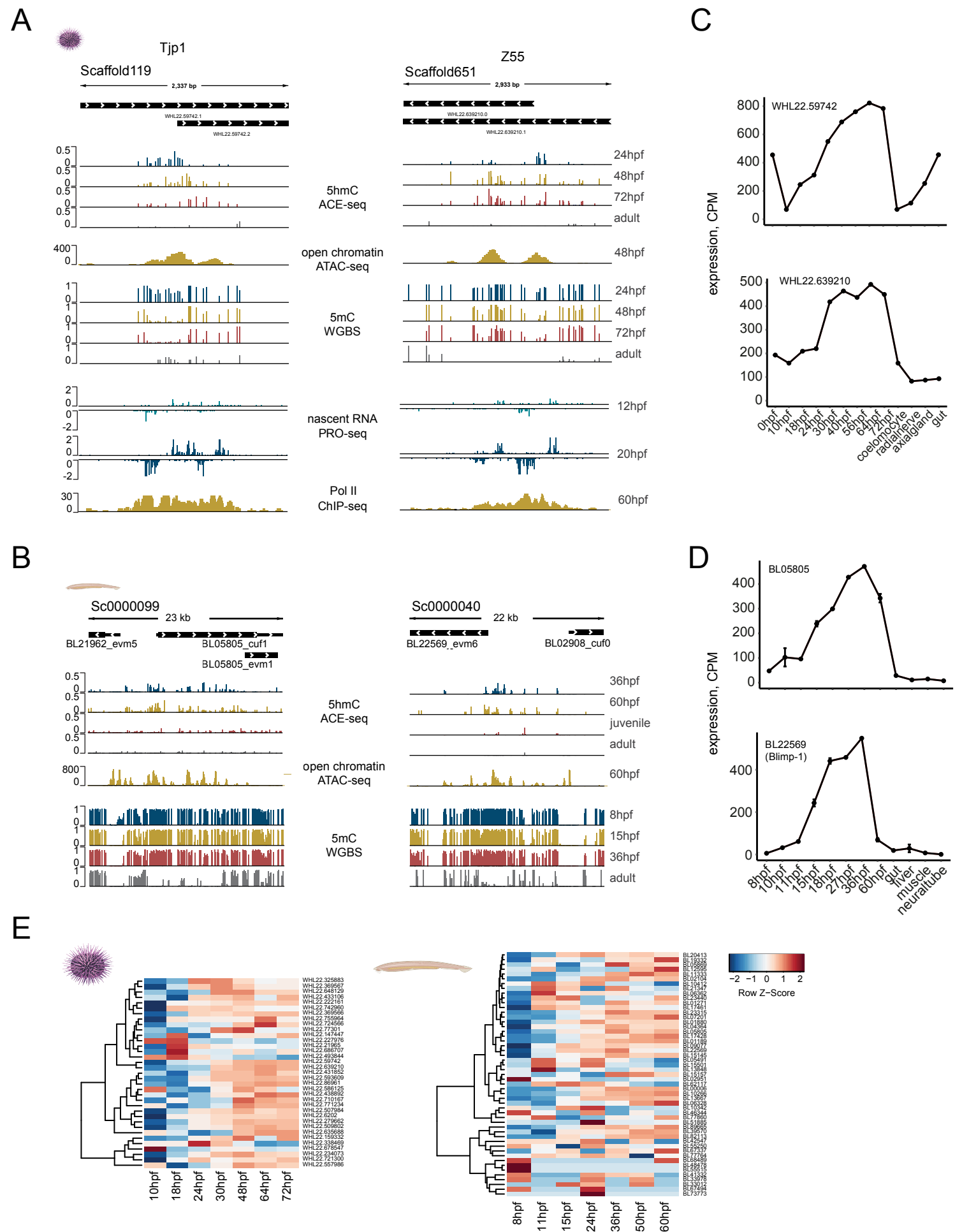
E



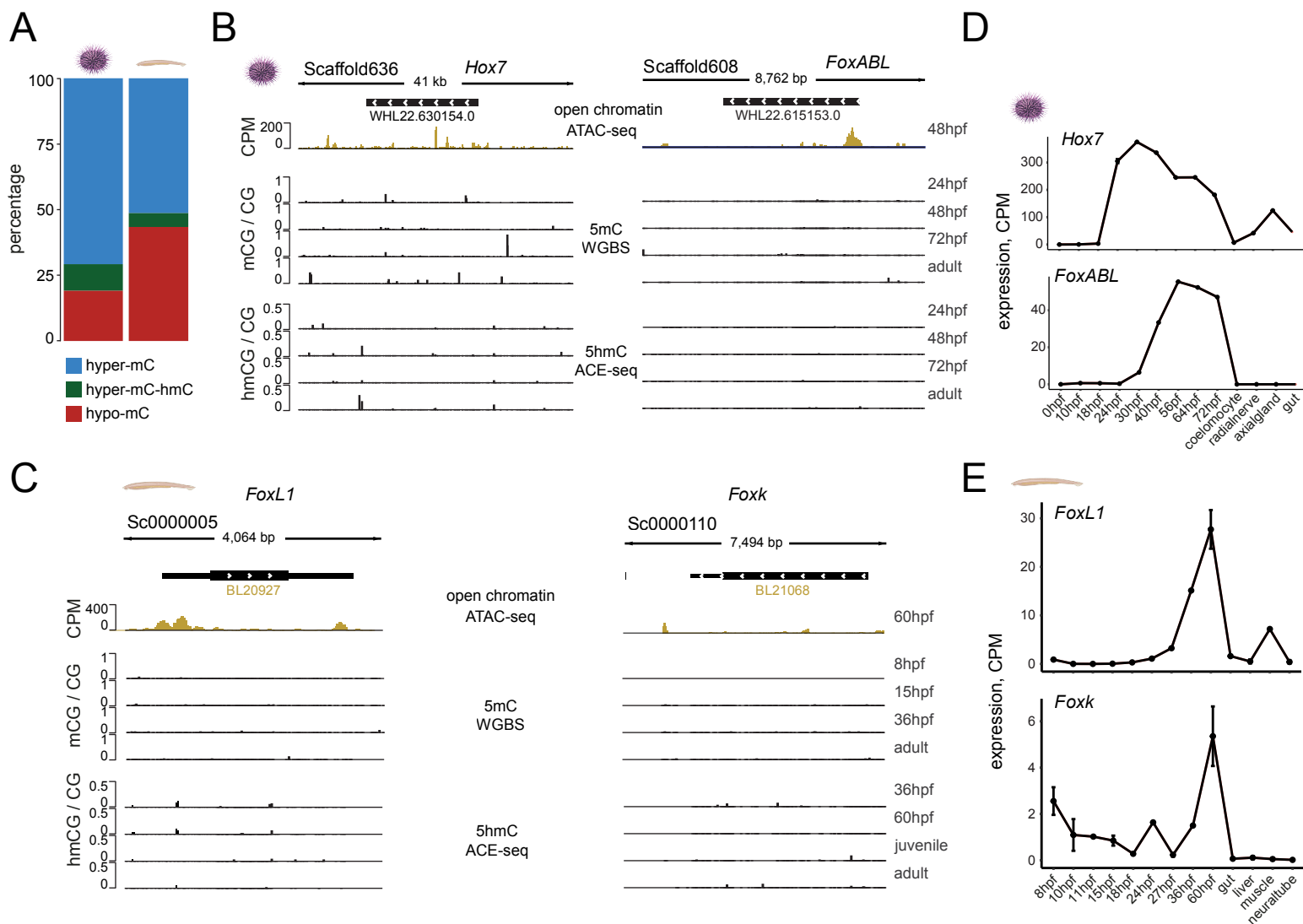
F



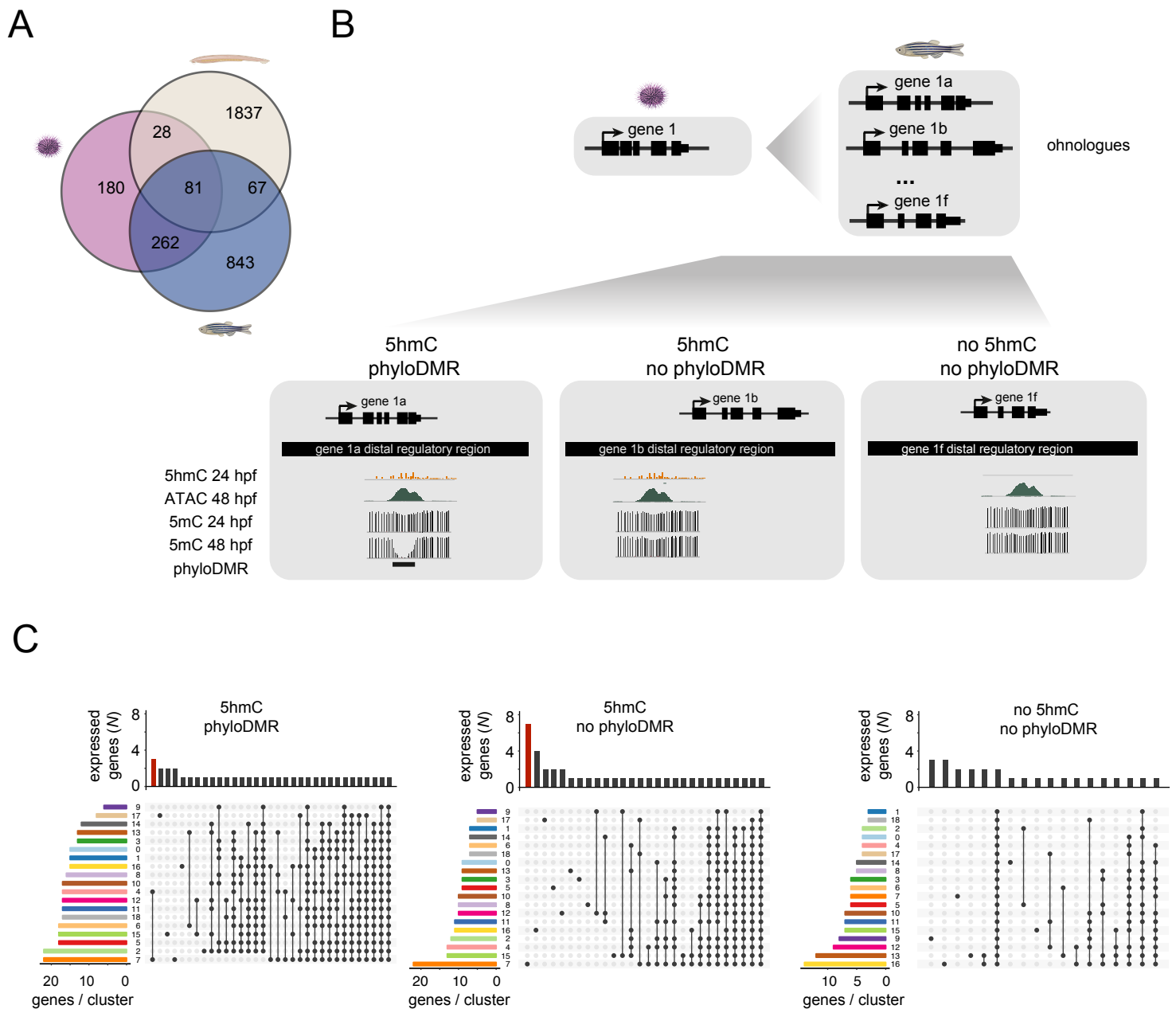
Supplementary Figure 4. Regulatory features and gene expression profiles of 5hmC-linked genes. (A-B) Length distributions of sea urchin (A) and lancelet (B) genes harbouring 5hmC-marked ATAC-seq peaks (gene body ATAC hmc+) and non-5hmC ATAC-seq peaks (gene body ATAC hmc-) within their gene bodies. **(C)** HOMER motif enrichment analysis of ATAC hmc+ regions associated with either gene promoters or gene bodies. **(D)** HOMER motif enrichment analysis of ATAC hmc+ regions associated with intergenic regions. **(E)** Dynamics of 5mC at 5hmC-marked gene promoters during sea urchin and lancelet development. **(F)** Expression profiles of gene body ATAC hmc+ genes as compared to gene body ATAC hmc- genes in sea urchin and lancelet embryos and adult tissues.



Supplementary Figure 5. DNA methylation, hydroxymethylation and gene expression dynamics at the 5hmC-marked promoters of developmental genes during sea urchin and lancelet embryonic development. (A, B) IGV browser tracks depicting 5hmC enrichment at gene promoters of developmental genes, coinciding with loss of 5mC and increased expression in sea urchin (**A**) and lancelet (**B**) genomes. Chromatin accessibility (ATAC-seq), 5mC (WGBS) and 5hmC (ACE-seq) tracks are shown in both species. In addition, nascent transcription (PRO-seq) and PolII enrichment (ChIP-seq) are shown for sea urchin. (**C, D**) Expression dynamics of 5hmC-regulated genes during sea urchin (**C**) and lancelet (**D**) development. (**E**) Gene expression changes (row Z-score) of sea urchin and lancelet genes displaying 5hmC / ATAC-seq peaks overlapping their promoters.

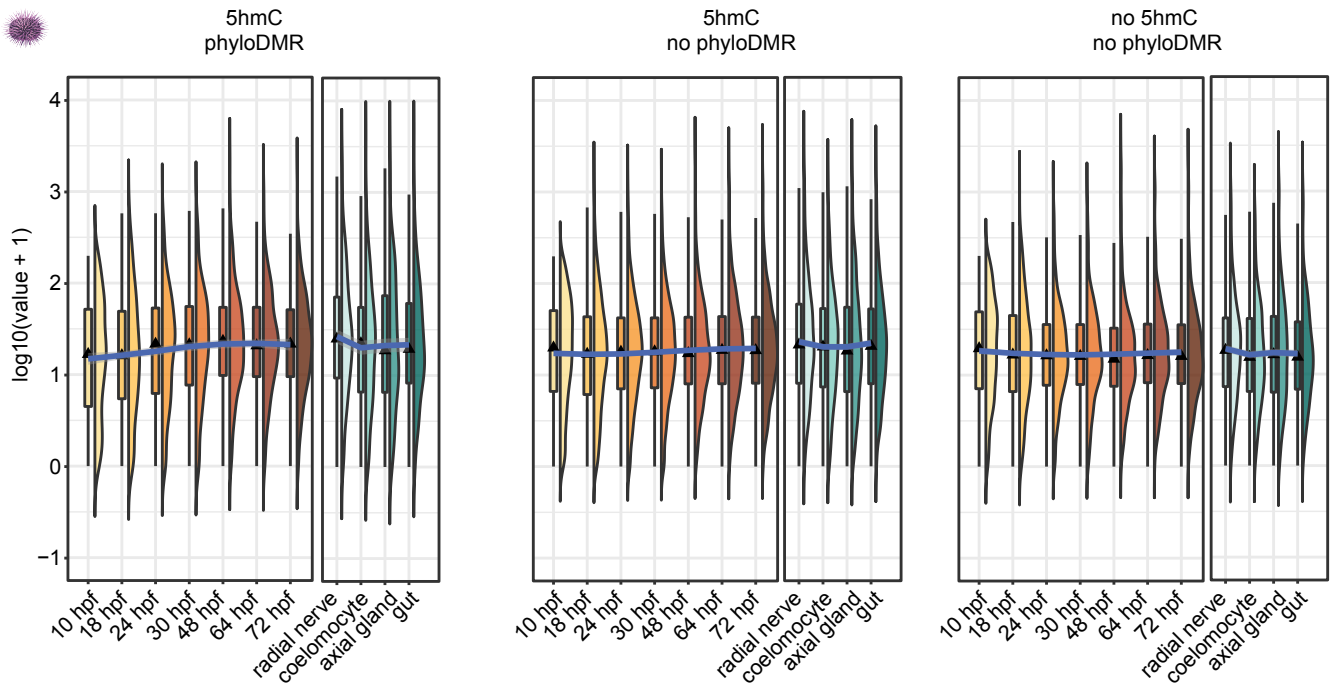


Supplementary Figure 6. DNA methylation, hydroxymethylation and gene expression dynamics at developmental genes during sea urchin and lancelet embryonic development. (A) Percentages of developmental genes in sea urchin (n=2,029), and lancelet (n=2,940) with: hypomethylated gene bodies (hypo-mC), methylated gene bodies with hmC-marked ATAC-seq peaks (hyper-mC-hmC), and methylated gene bodies without hmC-marked ATAC-seq peaks (hyper-mC). (B, C) IGV browser tracks depicting key developmental genes residing within hypomethylated valleys in sea urchin (B) and lancelet (C) genomes. Chromatin accessibility (ATAC-seq), DNA methylation (WGBS) and DNA hydroxymethylation (ACE-seq) tracks are shown. (D) *Hox7* and *FoxABL* and (E) *FoxL1* and *Foxk* gene expression dynamics of during embryonic development.

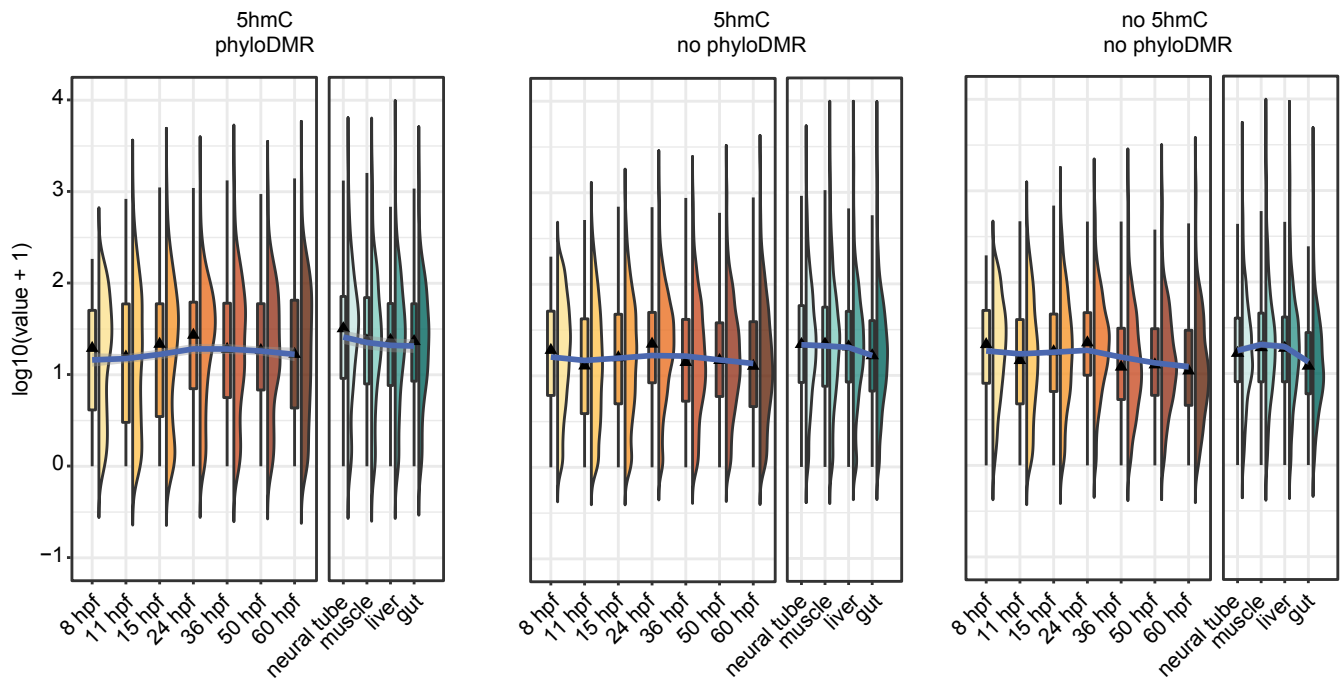


Supplementary Figure 7. Conservation of developmental gene regulatory logic of 5hmC-marked genes in deuterostomes. (A) Triple Venn diagram showing an overlap between ATAC hmC+ genes in sea urchin, lancelet and zebrafish. **(B)** Schematic depicting sea urchin genes associated with developmental 5hmC and their orthologous zebrafish genes, retained after three rounds of whole genome duplication (2R/3R-ohnologues). Zebrafish gene 2R/3R-ohnologues were separated into three groups: genes associated with: (i) 5hmC-marked ATAC-seq peaks overlapping phylo-DMRs (5hmC/phyloDMR), (ii) 5hmC-marked ATAC-seq peaks not overlapping phylo-DMRs (5hmC/no phyloDMR) and (iii) non-5hmC ATAC-seq peaks not overlapping phylo-DMRs (no 5hmC/no phyloDMR) in their distal regulatory domains. Distal regulatory domains were defined using GREAT gene regulatory domain definition (McLean et al, 2010). **(C)** Upset plots showing the number of genes expressed exclusively in a particular cell cluster or an intersection of clusters. A gene was considered to be expressed in a given cell cluster if it was expressed in minimum 25% of cells in the cluster. Horizontal bars depict the number of genes expressed in each cluster. Neuronal genes are marked in red.

A



B



Supplementary Figure 8. Developmental expression of sea urchin and lancelet genes whose orthologs in zebrafish do and do not harbour 5hmC-marked phyloDMRs. (A, B) Expression dynamics of sea urchin and lancelet genes, whose zebrafish orthologs harbour 5hmC-marked ATAC-seq peaks overlapping phylo-DMRs (5hmC/phylo-DMR), or 5hmC-marked ATAC-seq peaks not overlapping phylo-DMRs (5hmC/no phylo-DMR) or non-5hmC ATAC-seq peaks not overlapping phylo-DMRs (no 5hmC/no phylo-DMR) in embryos and adult tissues.

Supplementary Tables 1-13

Table 1. MethylC-seq, ACE-seq, and hMeDIP-seq samples and sequencing metrics

Table 2. Sea Urchin 24hpf vs 48hpf hypoDMRs

Table 3. Sea Urchin 48hpf vs 72hpf hypoDMRs

Table 4. Sea Urchin 72hpf vs adult hypoDMRs

Table 5. Lancelet 8hpf vs 15hpf hypoDMRs

Table 6. Lancelet 15hpf vs 36hpf hypoDMRs

Table 7. Lancelet 36hpf vs adult_hypoDMRs

Table 8. Sea Urchin ATAC (hmC+) peak promoter motifs

Table 9. Sea Urchin ATAC (hmC+) gene body motifs

Table 10. Sea Urchin ATAC (hmC+) intergenic motifs

Table 11. Lancelet ATAC (hmC+) peak promoter motifs

Table 12. Lancelet ATAC (hmC+) peak gene body motifs

Table 13. Lancelet ATAC (hmC+) peak intergenic motifs