Epigenetic machinery is functionally conserved in cephalopods

Filippo Macchi[1], Eric Edsinger[2] and Kirsten C. Sadler[1*]

[1]Program in Biology, New York University Abu Dhabi, Abu Dhabi, P.O. Box 129188, United Arab Emirates

[2]Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 Torrey Pines Road, La Jolla, CA 92037, United States

*** Corresponding author:**
Kirsten C. Sadler
ke27@nyu.edu

Keywords: *Octopus bimaculoides*, cephalopods, epigenetics, DNA methylation, histone methylation, DNMT1, UHRF1

Abstract

Epigenetic regulatory mechanisms are divergent across the animal kingdom, yet little is known about the epigenome in non-model organisms. Unique features of cephalopods make them attractive for investigating behavioral, sensory, developmental and regenerative processes, but using molecular approaches in such studies is hindered by the lack of knowledge about genome organization and gene regulation in these animals. We combined bioinformatic and molecular analysis of *Octopus bimaculoides* to identify gene expression signatures for 12 adult tissues and a hatchling, and investigate the presence and pattern of DNA methylation and histone methylation marks across tissues. This revealed a dynamic gene expression profile encoding several epigenetic regulators, including DNA methylation maintenance factors that were highly conserved and functional in cephalopods, as shown by detection of 5-methyl-cytosine in multiple tissues of octopus, squid and bobtail squid. WGBS of octopus brain and RRBS from a hatchling revealed that less than 10% of CpGs are methylated, highlighting a non-random distribution in the genome of all tissues, with enrichment in the bodies of a subset of 14,000 genes and absence from transposons. Each DNA methylation pattern encompassed genes with distinct functions and, strikingly, many of these genes showed similar expression levels across tissues. In contrast to the static pattern of DNA methylation, the histone marks H3K27me3, H3K9me3 and H3K4me3 were detected at different levels in diverse cephalopod tissues. This suggests the methylome and histone code cooperate to regulate tissue specific gene expression in a way that may be unique to cephalopods.

Introduction

Epigenetic modifications to histones and DNA are conserved features that regulate tissue-specific profiles of gene expression and repress transposable elements (TEs) [1]. There is great diversity across the animal kingdom in how the epigenome accomplishes these complex and important functions, yet mechanistic studies on epigenetic patterning and function have largely focused on a few model organisms [2]. While this has been extremely fruitful in elucidating mechanisms of epigenome patterning, regulation and function, they do not provide a comprehensive understanding of how epigenetic marks are patterned or how they regulate genes, TEs or genome structure across the evolutionary tree.

Examining DNA methylation in diverse animal species provides an illustrative example of how incorporating organismal diversity into epigenetic studies expands the epigenetic lexicon. In vertebrates, the vast majority of CpGs are methylated, with an asymmetric distribution throughout the genome, including enrichment in intergenic regions, exclusion from CpG rich promoters, and a varied pattern on gene bodies [3]. Heavy decoration of repetitive sequences with 5-methyl cytosine (5mC) reflects its important function in suppressing the expression of CpG dense transposons [4]. In contrast, the initial studies on canonical invertebrate model organisms - *Caenorhabditis elegans* and *Drosophila melanogaster* – showed there was a complete lack of DNA methylation [5-7]. The result of this is that detailed functional analysis of DNA methylation in animals has been largely carried in vertebrate model organisms and human samples, creating a large gap in understanding of how the DNA methylome is patterned or what its function is in other animals.

The advance of sequencing technologies has massively expanded the diversity of organisms with fully sequenced genomes, allowing examination of DNA methylation patterns in animals from across the evolutionary tree [3, 8-11]. This revealed that most invertebrates have either no or low levels of CpG methylation. Moreover, the pattern of 5mC distribution is very different in animals with low levels of methylation compared to vertebrate genomes, with methylated CpGs enriched in gene bodies and heavily methylated genes tend to be expressed at higher levels than unmethylated genes [3, 8, 9, 12-16]. Exceptions to these patterns abound; for example, the methylome in a sponge is highly similar to vertebrates [17, 18], the sea squirt *Ciona intestinalis* shows a high level of gene body methylation with an intermediate methylation pattern on repeats [9, 10, 19] and there are dynamic changes in the DNA methylation pattern during oyster development [15, 20], compared to the largely static pattern across tissues and developmental stages in vertebrates.

Unique features of the *Octopus bimaculoides* methylome was highlighted by a recent study using whole genome bisulfite sequencing (WGBS) to analyze brain methylation patterns in diverse animal species [16]. In most animals, TE abundance correlates with genome size, and since DNA methylation serves to keep transposons in check in vertebrates and plants [12], it has been hypothesized that DNA methylation is a key to suppressing TE expression in organisms with large genomes and high transposon burden [21]. This is not the case in *O. bimaculoides,* where half the genome is populated by repetitive elements similarly to vertebrates [22]. Interestingly, the level of DNA methylation is not proportional to the transposon load in this genome [16]. This raises the important question of whether the DNA methylation pattern observed in octopus is conserved in other cephalopods and warrants investigation of the function of DNA methylation in these animals.

In species which have DNA methylation, the pattern of CpG methylation needs to be copied to the daughter strand during DNA replication. This is carried out by the DNA methyltransferase DNMT1 and by UHRF1, which recognizes hemi-methylated DNA generated during DNA replication and recruits DNMT1 [23-28]. UHRF1 also functions as a reader of the histone code, including the canonical heterochromatin mark, trimethylated histone H3 lysine 9 (H3K9me3) [29-34]. The DNMT1-UHRF1-complex represents the core DNA methylation machinery and contributes to establishing heterochromatin domains. Recent studies reported high conservation of DNMTs, UHRF1 and TET proteins across diverse phylogenetic groups of animals including annelid worms [18], sponges [17, 18], mollusks [14, 35], and other invertebrates [3, 16]. This indicates that although 5mC is not ubiquitous throughout the animal kingdom, in those cases where it is present, it plays diverse, albeit incompletely understood, roles.

Cephalopods represent an emerging model system with multiple studies utilizing squid, bobtail squid and octopus for uncovering novel mechanisms of RNA editing, highly complex behavioral regulation, remarkable regenerative capacity, and genome evolution [36-44]. Recent advances in embryo cultivation, standardized aquaculture protocols [45], and genome editing [46] have advanced the utility of cephalopods as new model organisms [47]. Transcriptomic profiling of mollusk embryos [48], brain [49, 50] and multiple tissues of *O. bimaculoides* adults [22] have revealed that key genes having a role in development and neurological processes are highly conserved in these animals. However, a comprehensive analysis of tissue-specific gene expression profiles in cephalopods has not been reported. Moreover, despite significant strides in cephalopod research and few studies reporting the presence of DNA methylation in some species of octopus [16, 51, 52], there is virtually

nothing known about the epigenetic marks that contribute to the regulation of tissue-specific gene expression profiles, transposon suppression or genome organization in cephalopods.

We address this using transcriptomic and methylome datasets and biochemical approaches to analyze *O. bimaculoides* tissue specific gene expression, methylation patterning and histone modifications. We show that DNA methylation and histone modifications are present in multiple tissues of three cephalopod species. Methylome analysis in octopus shows that while methylated CpGs account for less than 10% of all CpGs in the genome and are virtually absent from repetitive DNA and transposons. Methylated CpGs are clustered on the gene bodies of a distinct set of genes which are highly expressed across tissues. This shows that CpG methylation and histone methylation are prominent features of the cephalopod epigenome and suggests that the pattern of DNA methylation is set by characteristics of the genome that are maintained across cell types.

Materials and Methods

Animal husbandry and sample collection. Animals were maintained in a circulating natural sea water aquaculture facility at the Marine Resources Center at the Marine Biological Laboratory and all experiments were performed according to the current policy for the use of cephalopods at Marine Biology Laboratories (MBL, https://www.mbl.edu/policies/files/2018/07/J1.10-CEPH-Policies-Procedures-Jan-1-31-2020-fillable-form.pdf). In brief, adult male octopus *O. bimaculoides*, adult male bobtail squid *Euprymna berryi* and adult squid *Doryteuthis pealeii* (unknown sex) were anesthetized in 3% ethanol in natural sea water for 10 minutes and tissues were dissected, flash frozen in liquid nitrogen, and stored at -80°C. One whole *O. bimaculoides* hatchling at developmental stage of 30 days post-fertilization (30 dpf hatchling) was anesthetized as above and flash frozen.

RNA and DNA extraction. Frozen tissues were ground using a mortar and pestle cooled with liquid nitrogen and placed in dry ice. 15 mg of tissue powder was used to extract either RNA or DNA. RNA was extracted using Trizol (Invitrogen) following the manufacturer's instructions with some modifications. Briefly, during precipitation in isopropanol, 10 µg of Glycoblue (Thermo Fisher Scientific) was added and precipitation was performed overnight at -20°C followed by 1 hour centrifuge at 12000g at 4°C. RNA was resuspended in water and used in the following procedures. Genomic DNA (gDNA) was extracted by using a DNA extraction buffer (10 mM Tris-HCl pH9, 10 mM EDTA, 200 mM NaCl, 0.5% DSD, 200 µg/ml proteinase K) and over-night incubation at 65°C, followed by RNAase treatment with 2 mg/ml PureLink™ RNase A (Invitrogen) for 2 h at 37°C. Then, 0.25 v/v of 5 M potassium

acetate (CH_3CO_2K) was added and the sample centrifuged at 12,000 x g at room temperature to precipitate proteins. 1:1 v/v of isopropanol was added to the supernatant and incubated at -20°C overnight and DNA precipitated by centrifuge at 12,000 x g at room temperature for 15 minutes. DNA was resuspended in water and quantified by Qubit dsDNA Broad Range kit.

Slot blot. Slot blot was performed using 1.5 ng of gDNA that was denatured in 400 mM NaOH/10 mM EDTA and blotted onto nitrocellulose membrane (BioRad) in duplicate for dsDNA and 5mC DNA using a slot blot apparatus (BioRad). Membranes were incubated 1 hour at 80°C, blocked with 5% bovine serum albumin (BSA) in TBST (37 mM NaCl, 20 mM Tris pH 7.5, 0.1% Tween 20), and incubated overnight at 4°C in either anti-dsDNA (Abcam, 1:5,000 in 2% BSA in TBST) or anti-5-methyl-cytosine (5mC – Aviva Biosystem clone 33D3, 1:3,000 in 2% BSA in TBST). Membranes were washed in TBST and probed with anti-mouse HRP secondary antibody (Promega; 1:2,000 in 5% BSA in TBST) for 1 hour at room temperature followed by development in ECL (Thermo Fisher Scientific) or Clarity ECL (BioRad). ChemiDoc (BioRad) was used to detect and quantify the chemiluminescent signal. Gel Analyzer (http://www.gelanalyzer.com) was used to perform quantitative densitometric analysis of the signals and ratio between 5mC and dsDNA was plotted for each sample using GraphPad Prism.

Protein extraction and Western blotting. Frozen tissues were ground using a mortar and pestle cooled with liquid nitrogen and placed in dry ice. 15 mg of tissue powder was used to extract proteins in lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1% v/v NP-40, 10% v/v glycerol, 2mM EDTA). Protein extraction was performed using a probe sonicator (2 sec pulse, 2 sec pause for 5 min at amplitude 30%) and lysates were cleared by centrifuging at 11,000 × g for 15 min at 4°C and quantified using Qubit reagent (Invitrogen). For preparation of samples for SDS PAGE, 4x Laemmli buffer (BioRad) was added to protein extracts, incubated at 95 °C for 5 minutes and 15 µg of proteins were loaded onto 12.5% denaturating gels, electrophoresed, transferred onto PVDF membranes (BioRad), blocked with 5% w/v powdered milk in TBST buffer (20 mM Tris-HCl, 150 mM NaCl, 0.1% v/v Tween 20, pH 8.0) for 1 hour at room temperature, and incubated overnight at 4°C with anti-H3 (SantaCruz, sc-10809, Rabbit polyclonal, 1:5,000), anti-H3K4me3 (Abcam, ab8580, Rabbit polyclonal, 1:1,000), anti-H3K9me3 (Active Motif, 39161, Rabbit polyclonal, 1:1,000) or anti-H3K27me3 (Active Motif, 61017, Mouse monoclonal, 1:1,000) diluted in blocking buffer. After washing with TBST and incubation for 1 hour with anti-Rabbit IgG HRP Conjugate (Promega, 1:2,000) or anti-Mouse IgG HRP Conjugate (Promega, 1:2,000) diluted in blocking buffer followed by washing in TBST, membranes were visualized using Pierce™ ECL Western

Blotting Substrate (Thermo Fisher Scientific) or Clarity ECL substrate (BioRad) on the BioRad ChemiDoc. Immunoblot bands were quantified by densitometry using GelAnalyzer (http://www.gelanalyzer.com).

RNA-seq. Total RNA extracted as described above was treated by DNAse I for 30 minutes at 37°C followed by RNA purification (RapidOut DNA Removal Kit – Thermo Fisher Scientific). RiboZero was used to remove ribosomal RNA and the remaining sample was used for library preparation according to manufacturer's instructions (Illumina) from 250 ng of RNA. Libraries were sequenced on NextSeq550 (Illumina) to obtain 150 bp paired-end reads. Raw FASTQ sequenced reads where first assessed for quality using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads where then passed through Trimmomatic v0.36 [53] for quality trimming and adapter sequence removal, with the parameters (ILLUMINACLIP: trimmomatic_adapter.fa:2:30:10 TRAILING:3 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36). The dataset from hatchling was also processed with Fastp [54] in order to remove poly-G tails and Novaseq/Nextseq specific artifacts. Following the Fastp quality trimming, the reads were assessed again using FastQC. Quality trimmed reads were used to produce psuedoalignments using Kallisto [55], and the Kallisto quantification was assessed with the --bias flag using the reference *O. bimaculoides* genome (PRJNA270931) and its corresponding annotation. The resulting transcripts per kilobase per million (TPMs) from the pseudo-counts were used for further downstream analysis. Files are available on GEO (accession number: GSE188925).

Trinotate transcriptome annotation. The quality trimmed reads were aligned to the *O. bimaculoides* genome (PRJNA270931) using HISAT2 [56] with the default parameters and additionally by providing the –dta flag. The resulting SAM alignments were then converted to BAM format and coordinate sorted using SAMtools v1.3.1 [57]. The sorted alignments were processed through Stringtie v1.3.0 [58] for transcriptome quantification to produce a GTF file per sample. The GTFs were then combined using STRINGTIE merge to produce one merged GTF representing the transcriptome for the genome. Finally, Qualimap [59] v2.2.2 was used to generate RNAseq specific QC metrics per sample. Following the transcriptome quantification steps above, the Trinotate [60] pipeline was used to annotate the transcriptome, in addition to the existing reference annotation. The Trinotate steps as detailed in the software's user manual were followed. Briefly, after transcriptome preparation, BlastP longest ORFs using the longest_orfs protein sequences against the Uniprot database and Pfam domain search using longest ORFs against the Pfam database were integrated into coding region selection using TransDecoder.Predict. In addition, the transcriptome FASTA file was BlastX against the Uniprot database and domain scanned

using HMMscan to generate gene to transcript mappings using transIDmapper.pl with the output exported to an SQLite database.

Reduced Representation Bisulfite Sequencing (RRBS). RRBS was performed on gDNA extracted from the same 30 dpf hatchling of *O. bimaculoides* used in RNA-seq. Briefly, 1,000 ng of genomic DNA was digested with 200 U of MspI (New England Biolabs) for 24 hours at 37°C. Digested DNA was used for preparing the library as previously described [61], with the exception that adaptors used for multiplexing were purchased separately (Next Multiplex Methylated Adaptors - New England Biolabs). Libraries were size-selected by dual-step purification with Ampure XP Magnetic Beads (Beckman Coulter, Agencourt) to specifically select a range of fragments from 175 to 670 bp, as previously described [62]. Bisulfite conversion was performed with Lightning Methylation Kit (ZYMO Research) by following the manufacturer's instructions. Libraries were amplified using KAPA HiFi HotStart Uracil+ Taq polymerase (Roche) and purified with Ampure XP Magnetic Beads (Beckman Coulter, Agencourt) before sequencing. Libraries were sequenced using the Illumina Nextseq550 sequencer. Quality control was undertaken using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Reads were quality trimmed using Trimmomatic [53] to remove low quality reads and adapters. Reads passing quality control were aligned to the reference genome (assembly PRJNA270931, available here: https://groups.oist.jp/molgenu/octopus-genome) using default parameters in Bismark [63], which adopts Bowtie2 as the aligner [64] and calls cytosine methylation at the same time. Fastq files are available on BioSample (accession number: SAMN23139394). RRBS metrics describing lengths of reads, numbers of sequenced and mapped reads and identified citosines were extracted using 'bismark2report' function in Bismark (Table S1A), while conversion rates (99.20%) and coverage stats were extracted using 'processBismarkAln' and 'getCoverageStats' function in methylKit (Table S1B) [65].

Bioinformatic analysis. RNA-seq data was analyzed as described above and visualized in RStudio (R version 4.0). Heatmaps for transcriptomic profiling were performed by using R package 'pheatmap'. Clusters were calculated on the hierarchical clustering of dendrogram (Euclidean distance) based on the normalized expression profile (row z-score across different tissues). The number of clusters was determined based on the optimal performance to discriminate the different tissues used for the analysis. For Gene Ontology (GO), GO terms were downloaded from Ensemble Metazoa (BioMart database). GO enrichment analysis was conducted using the GO hypergeometric over-representation test in the 'ClusterProfiler' package in R. An adjusted p-value < 0.05 was treated as significant for all analyses. Unique stable transcript IDs were annotated with GO terms from BioMart

database and divided by Molecular Function, Biological Process, and Cellular Component terms. Putative members of epigenetic machinery represented in the heatmaps were identified using Trinotate as described above and reported with corresponding human gene symbols. For RRBS analysis on hatchling samples, CpG methylation levels were extracted from Bismarck aligned file with the R package 'methylKit' [65]. CpGs covered at least 10 times (and with minimum phred quality score = 20) were included in the analysis. Whole genome bisulfite sequencing (WGBS) data from the Supra and Sub Esophageal brain regions were obtained from public available GEO datasets (GSE141609) [66]. Specifically, Supra (GSM4209498) and Sub (GSM4209499) esophageal brain Bisulfite-seq files (CGmap files) were downloaded and used as source of CpGs methylation analysis. To compare WGBS data with RRBS, CGmap files were processed as follow: CGmap files were filtered for methylation on CpGs context only (based on column 4 representing the context (CG/CHG/CHH)); strand direction information was obtained converting C into + and G into – (based on column 2 representing the nucleotide on Watson (+) strand); percentage of methylation was obtained multiplying by 100 the methylation-level (from 0 to 1; in 0 to 100 values) (based on column 6 methylation-level). CpGs with methylation levels below 20% were treated as unmethylated and above 80% were considered as methylated. Genomic element annotation and metaplots of CpGs were performed with R package 'genomation'. Repetitive Elements (RE) were identified using the Repeat Masker annotation on the reference genome (assembly PRJNA270931, available here: https://groups.oist.jp/molgenu/octopus-genome), and manually curated to group the classes of transposons (DNA, LTR, LINE, SINE) and non-transposons (Satellite, Simple Repeats, Other RE). Since sequences represented (transcripts or RE) are of unequal length, metaplots were divided into 30 or 15 equal bins (based on the average length of the sequences analyzed). Lines in metaplots represent the winsorized mean values (1–99 percentile) for each bin, and blue shades represent dispersion of 95 percent confidence interval for the mean. Heatmaps of DNA methylation pattern were performed with R package 'EnrichedHeatmap" on the full-length transcripts and 2,000 bp upstream and downstream. DNA methylation average of gene bodies for each transcript was divided into 4 clusters based on kmeans from R package 'stats'. Number of clusters was optimized at 4 based on the ability to discriminate between the distinct patterns of DNA methylation. Center of kmeans from Pattern 1 to 4 are respectively 71.06, 49.11, 28.59, 1.31. Subsequent analyses on methylation pattern were performed using R package 'ggplot' for violin plots, 'pheatmap' for expression profiling, and 'ClusterProfiler' for GO enrichment analysis. All code used for this study is available on Github (https://github.com/SadlerEdepli-NYUAD).

Phylogenomic Pipeline. Human sequences for genes encoding DNA methylation and histone modification factors were collected from Uniprot (October 5, 2021) to generate reference gene sets (RGS) per family (Table S2). Project databases, Metazoa50 and Metazoa19 (a subsample of Metazoa50), representing 50 and 19 animal and unicellular outgroup genomes, respectively, were generated per genome. Genomes were sourced from Ensembl, NCBI RefSeq, or other public databases (Table S3), and species and data set selection was based on representation of major clades and phyla and BUSCO-based evaluations of genome quality (BUSCO; Metazoa10) [67]. Genome gene models were filtered to retain single longest protein isoform per coding gene locus. Fasta file headers were standardized per genome based on NCBI Taxonomy (January 2021) details per species and on common names as provided in Wikipedia (October 2021) (Table S2). Blastp databases were generated per genome using NCBI Blast+ (version 2.6.0) Makeblastdb. Metazoa50 genomes and RGS sequences were domain annotated using Interproscan Pfam (version 5.48-83.0) [68].

Human reference gene sets were blasted against Metazoa19 and Metazoa50 Blast databases per gene family or superfamily using NCBI Blast+ Blastp (e-5 cutoff). All hits were then blasted against the human genome and top hits to any human RGS sequence retained to form the candidate gene set (CGS), which were combined with RGS sequences to form a final gene set. Sequences were aligned in MAFFT (linsi command; version 7.487)[69], alignments were trimmed in ClipKIT (super gappy; version 1.1.3) [70], and maximum likelihood trees were built in IQTREE2 (with ModelFinder; version 2.1.2) [71]. Alignments and trees were visually assessed in Geneious (version 2021.1.1) and in FigTree (version 1.4.4 and iTOL (version 6) [72], respectively. Sequences residing on UHRF1, DNMT1, SETD1B, SETDB1, EZH2, KAT2A, and HDAC8 branches within a larger superfamily tree were collected and then aligned, trimmed, and a tree built for just the family to improve branching structure and support relative to the superfamily tree. Octopus sequences came from the NCBI RefSeq genome and were mapped to the Ensembl genome based on top Blastp hit to match sequence identifier to those used in transcriptome analyses. Trees were color annotated in FigTree and rooted on sponge or a unicellular outgroup or else left unrooted.

Protein alignment. Protein sequences were downloaded from the UniProt database (https://www.uniprot.org) as FASTA formatted files and alignments were performed using ClustalOmega with multiple sequence alignment program [73]. Output alignment files generated in a ClustalW format with character counts were reformatted and colored based on amino acid residue identity using MView (https://www.ebi.ac.uk/Tools/msa/mview/).

Protein sequences of interest were processed with phmmer [74] and queried against HMM target databases using profile hidden Markov models. Sequences matches were calculated by multiple factors, grouped by Pfam domains and homology sequence probability were represented by Bit score.

Swiss-Expasy 3D Model. To build protein homology models we used the Automated Mode (https://swissmodel.expasy.org/docs/help) of the SWISS-MODEL server homology modeling pipeline [75]. In brief, homology modeling proceeds through four main steps: (i) alignment of target sequence and identification of structural templates by BLAST and HHblits; (ii) alignment and sorting of target–template structures based on Global Model Quality Estimation (GMQE) and Quaternary Structure Quality Estimation (QSQE); (iii) model-building relying on ProMod3 [76] and OpenStructure comparative modelling engine [77]; and (iv) model quality evaluation using GMQE estimation score, and another composite estimator QMEAN [78]. SWISS-MODEL Structure comparison tool was used to perform super-positioning of newly computed 3D models of *O. bimaculoides* proteins and published structures of mouse DNMT1 (PDB ID 4da4.1) and UHRF1 (PDB ID 2zke.1).

Statistical analysis. All experiments were carried out on at least 3 biological replicates, except where noted. For slot blot analysis, technical replicates were also included. The number of replicates for each experiment is indicated in each figure. Methods to evaluate statistical significance include unpaired parametric one-way ANOVA test adjusted with Tukey's multiple comparisons test or unpaired non-parametric Kruskal-Wallis test adjusted with Dunn's multiple comparisons test. Tests used are indicated in each graph. All plots were generated in GraphPad Prism 9 or RStudio (R version 4.0). Statistical analysis was performed in GraphPad Prism 9.

Results

Tissue specific gene clusters identified in octopus

We extended previously published RNA-seq datasets obtained from 11 different adult tissues (Figure 1A; [22]) by generating new RNA-seq datasets from the first left (L1) arm of an adult male octopus at distal, medial and proximal locations (between 1.5 and 5.5 cm from the arm tip, Figure S1A-B), and one 30 day old hatchling (Figure S1B-C). We combined these datasets and found that of 40,327 transcripts annotated in the *O. bimaculoides* genome, 38,232 were expressed in at least one sample (TPM > 0). Trinotate [79] was used to better annotate the *O. bimaculoides* genome and assigned putative gene names and

putative UniProt entry names to 23,879 transcripts (59%) identified in the datasets analyzed (Table S4).

Optimized hierarchical clustering of the 38,232 expressed transcripts generated 13 clusters that define specific transcriptomic profiles across these tissues (Figure 1B and Table S5). Some clusters were dominated by transcripts that were nearly exclusively expressed in a single tissue, such as Cluster 6 (testes), whereas other clusters were defined by transcripts expressed in multiple tissues that are functionally related, such as Cluster 5 (tissues with neuronal functions). Interestingly, the distal arm segment was characterized by a distinct set of transcripts in Cluster 10 that were not highly enriched in any other tissue (Figure 1B), potentially reflecting the sensory and regenerative capacity of this structure [36, 41, 80].

Gene Ontology (GO) analysis for Molecular Function (Figure 1C), Biological Process (Figure S2A) and Cellular Component (Figure S2B) revealed that each cluster was enriched for transcripts encoding proteins with shared functions. For example, microtubule binding, motor activity, and cilium movement terms which are prominent for sperm activity characterized genes expressed in testes (i.e. Cluster 6; Figure 1C and S2A). G-protein coupled receptor activity, GTP binding, glutamate receptor activity, and signaling pathways and others critical for neuronal function and were enriched in Clusters 2 and 5, (brain and nervous tissues; Figure 1C and S2A). Cluster 4, which was enriched in transcripts expressed in the arm and hatchling, included terms involved in motor activity, actin filament binding, calcium-ion transmembrane transport, and myosin complex (Figure 1C and S2A-B), reflecting contractile activities of muscles. These findings make predictions about genes regulating cell identity and key processes in *O. bimaculoides* tissues and suggest that complex regulatory mechanisms, potentially based in the epigenome, regulate these gene expression profiles.

DNA methylation machinery is conserved and differentially expressed in octopus

We characterized DNMT and UHRF gene family evolution in animals using a phylogenomic pipeline we designed. Human DNMT1, TRDMT1 (DNMT2), DNMT3A, DNMT3B, and DNMT3L and human UHRF1 and UHRF2 were compared against Metazoa19 and Metazoa50 genomes. We identified DNMT1 (Ensembl Transcript ID: Ocbimv22034501m; NCBI gene symbol: LOC106877272; Figure 2A and S3A) and UHRF1 (Ensembl Transcript ID: Ocbimv22021185m; NCBI gene symbol: LOC106874972; Figure 2B and S3B) homologs in *O. bimaculoides* and across a range of animal species. Both DNMT1 and UHRF1 are absent in the three unicellular outgroups examined (choanoflagellate, *Mongosiga brevicollis*, the flilasterean, *Capsaspora owczarzaki* and the ichthyosporean, *Sphaeroforma artica*), but highly conserved in most animals and were likely present in the last common ancestor, with

a number of independent losses in diverse lineages. DNMT1 was lost in 11 out of 47 Metazoa50 animal species, including the fruit fly *Drosophila melanogaster* (but was present in two other arthropods), and nematodes. UHRF losses were detected for 12 out of 47 Metazoa50 animal species, which match losses of DNMT1 except for the sea squirt, *Ciona savignyi*, which is predicted to have low level of DNA methylation [81]. Several species or clades also exhibited expansions to two or three copies for both genes (Figure 2A-B and S3A-B).

The DNMT1 transcript (hereafter termed DNMT1_OCTBM) encodes a protein that retains the domain structure (Figure 2C) and key amino acid residues (Figure 2D) that are necessary for its methyltransferase function. HMMER analysis of DNMT1_OCTBM to identify sequence homology in the animal reference protein database showed that the C-5 cytosine methyl-transferase domain (Pfam ID: DNA_methylase), which carries out the catalytic function of DNMT1, had the highest homology score and overall, DNMT1_OCTBM had over 50% of identity with vertebrate DNMT1s (Figure S4A-B). We conclude that DNMT1_OCTBM has the necessary features to function as a DNA methyltransferase and to interact with UHRF1.

UHRF1 and UHRF2 family members in mammals, but there is only 1 family member present in mollusks (Figure S3B and [18]). The protein encoded by Ocbimv22021185m (hereafter termed UHRF1_OCTBM) retains features of the mammalian protein, including tandem Tudor and PHD domains that read the histone code and the SRA domain, which binds hemimethylated DNA and facilitates DNMT1 access to CpGs (Figure 2E) [23-26, 30, 31, 82, 83]. The SRA domain of UHRF1_OCTBM has highest homology compared to vertebrate homologs (Figure S4C-D), with key residues involved in hemi-methylated DNA recognition and base flipping [24, 26] completely conserved (Figure 2F). There was much lower homology between UHRF1_OCTBM and UHRF2, and the 5mC binding pocket (residue D474E), the base flipping motif (HVAG thumb loop) and the CpG recognition site (NKRT finger loop) [25] were not conserved (Figure 2F). Another octopus transcript (Ocbimv22020196m; termed YDG-OCTBM), encoded a shorter protein containing a domain similar to the SRA but had low homology to vertebrate UHRF1 or UHRF2 (Figure S4E-F) and lacked key residues for CpG recognition found in both UHRF1 and UHRF2 (Figure S4G). This is further supported by Swiss-Expasy 3D modeling [75] which reveal highly conserved 3D structures for DNMT1_OCTBM, with a near identical structures to the mouse protein (Figure 2G, S5A-B), including DNA double helix correctly located, the 5mC exposed near the C1194 (C1229 in mouse [84, 85]) and S-Adenosyl-L-methionine properly positioned (Figure 3G, S5A-B). Furthermore, the SRA domain of UHRF1_OCTBM aligned to the mouse

domain, with the loops necessary to bind DNA, to recognize CpGs (R612 and R496 in mouse [24, 26]) and to mediate base-flipping (V567 and V451 in mouse) all properly positioned (Figure 2H, S5C-D). This indicates that Ocbimv22034501m encodes the closest DNMT1 homolog and Ocbimv22021185m encodes the UHRF1 ortholog in *O. bimaculoides* and that the encoded proteins retain functional capacity to interact as a complex, for UHRF1 to recognize hemi-methylated DNA and for DNMT1 to methylated CpGs.

UHRF1 and DNMT1 function as a complex, and are typically co-expressed in cells that are actively proliferating. We found this to be the case in octopus tissues, with highest levels of both genes expressed in skin, suckers and arms (Figure 2I). This is consistent with the idea that these cells experience high turnover are expected to have proliferating cells [80]. Interestingly, we found that most genes encoding factors that read (methyl binding proteins; MBD), write (DNMTs) and erase (ten-eleven translocation; TET) DNA methylation in octopus were expressed at low levels in most tissues except for arms, suckers and skin (Figure 2I, Table S6). The relevance of this pattern is not yet known.

DNA methylation is enriched in the bodies of a subset of genes in *O. bimaculoides*

Reports of the presence of DNA methylation in octopus [16, 51, 52] and other mollusks [13, 15, 20] indicate that the DNA methylation machinery is functional in these animals. To test this directly, we used slot blot analysis to detect bulk 5mC levels on genomic DNA (gDNA) extracted from arm tip, brain (optic lobe) and gills of 3 different *O. bimaculoides* adults and a hatchling. We found relatively equivalent levels in all tissues, albeit at less than half of what was detected in mouse liver or zebrafish larvae (Figure 3A).

We next examined the genome-wide distribution of methylated and unmethylated CpGs by performing RRBS on gDNA isolated from a 30 dpf hatchling and by analyzing previously published WGBS datasets from 2 regions of the adult brain (Supra Esophageal - Supra E; and Sub Esophageal - Sub E) [16]. There are 80,660,576 CpGs in the *O. bimaculoides* genome, considering both DNA strands. RRBS profiling of the hatchling genome covered 2.98% of these (2,403,266 CpGs) and WGBS of the Supra E and Sub E covered 80.39% (64,845,686 CpGs) and 58.33% (47,053,504 CpGs), respectively (Figure S6A). In all samples, the vast majority of CpGs had methylation levels categorized as not methylated (<20% methylation) and CpGs categorized as methylated (>80% methylated) was low, ranging from 3.63% in hatchlings and 7.53% in the Sub E brain (Figure S6A-B). Approximately 7% of CpGs showed an intermediate pattern of methylation (between 20-80%) which could reflect tissue heterogeneity or a stochastic pattern of methylation on these CpGs. The methylation pattern (Figure S6B) is in marked contrast to the bimodal pattern

found in vertebrates and some species of sponge [17, 18], where majority of CpGs are methylated and the remaining are unmethylated. These findings indicate that the octopus methylome resembles other invertebrates [3, 8, 12] including another mollusk, the Japanese oyster *Crassostrea gigas* [3, 14], but is distinct from invertebrate model organisms which lack methylation entirely.

DNA methylation serves to repress transposon expression in vertebrates, and therefore the bulk of methylated CpGs are found in intergenic regions, with very little tissue-specific variation [3, 86, 87]. To investigate if there was a tissue specific CpG methylation pattern in octopus, we identified 1,425,557 CpGs that were commonly detected in the 2 brain and hatchling methylome datasets and then plotted the methylation levels on each CpG in the two brain samples and overlaid the methylation levels from the hatchling (Figure 3B). This showed a linear correlation between brain samples and that nearly all CpGs that were either unmethylated or highly methylated in brain had the same pattern in hatchling (Figure 3B), indicating little tissue-specific variation.

The scaffold nature of the *O. bimaculoides* genome [22] makes it difficult to accurately approximate the proportion that is intergenic compared to genic, but clearly the vast majority is intergenic, which is also where over 85% of CpGs reside (Figure 3C). These intergenic CpGs are comparatively depleted of methylation and there is a strong enrichment of methylated CpGs (>80%) detected in introns and exons (Figure 3C). Moreover, while there was high CpG density across the length of TEs, methylation levels were consistently low (Figure 3D), regardless of the TE class (Figure S6C). In contrast, DNA methylation was comparatively higher in gene bodies, with the average methylation level across gene bodies did not exceed 50% in any tissue (Figure 3E). Although we cannot exclude that our finding of low methylation levels across TEs could be attributed to their poor annotation, our data indicate that DNA methylation is not a major epigenetic mechanism of TE repression in octopus. We speculate that in cephalopods, like in other species with low DNA methylation levels, gene bodies are the main target of methylation [12].

The 5mC is a binary mark, so that each residue in individual cells can be either methylated or unmethylated. Therefore, we reasoned that the intermediate methylation pattern observed across all gene bodies represented heterogeneity in the CpG methylation pattern of genes. We tested this in the Supra E brain sample since it was the sample with the highest amount of covered CpGs (Figure S6A). Using k-means clustering of genes based on CpG methylation levels averaged for each transcript identified 4 distinct Methylation Patterns (Figure 4A-B): Pattern 1 had the fewest number of genes (5,482) and was characterized by

having high CpG methylation across the entire gene body and promoter; this was distinguished from genes in Pattern 2 (8,972), which had high methylation across the gene body and downstream of the transcription termination site (TTS) but lacked methylation at the transcript start site (TSS) and promoter. Genes in Pattern 3 (6,545) were characterized by having unmethylated CpGs in the first third of the gene body but higher methylation in the 3' end of the gene and around the TTS. Pattern 4 was the inverse of Pattern 1, as it had the largest number of genes (17,586) and virtually no methylated CpGs (Figure 4A and Table S7). Strikingly, these Methylation Patterns were the same in the Sub E and hatchling samples (Figure 4B, Figure S7A-B), indicating that the pattern of DNA methylation on gene is set by a cellular or genomic feature that is consistent across tissues.

We determine the relationship between gene body methylation and expression by plotting the expression (log2(TPM+1)) of all the transcripts in each Methylation Pattern. The highly methylated genes (Patterns 1-2) were expressed at the highest levels, while unmethylated genes (Pattern 4) were expressed at the lowest levels in the Supra E samples (Figure 4C). To examine the direct correlation between methylation level and expression, transcripts in the Supra E sample were grouped into percentiles based on expression levels (log2(TPM+1)) and then calculated the mean methylation level for all genes in each percentile (Figure 4D). This revealed a direct correlation between DNA methylation and expression for genes at the mid-range of expression levels (1.5 to 4.5 log2(TPM+1)) but there was low or no correlation between methylation levels and expression in genes that had low to no DNA methylation (Figure 4D). This correlation was also found in Sub E and hatchling samples (Figure S7C). Strikingly, a subset of genes in Patterns 1 and 2 were expressed at high levels in all tissues, whereas genes in Pattern 4 were either silenced or expressed at more variable levels across tissues. This suggests that hypermethylation of gene bodies serves to promote expression of genes that are sustained at high levels in all tissues and that hypomethylated genes are silenced.

GO analysis showed that specific gene functions were enriched in each Methylation Pattern. Trafficking, signaling, translation and metabolism characterized genes in Patterns 1-2 and the unmethylated genes in Pattern 4 participated in metabolism, processes that occur at the cell surface, or are related to neuronal functions (Figure 4F). This indicates that the methylation pattern on gene bodies, or lack thereof, defined genes that play roles in similar biological processes. It also suggests that the genes in Pattern 1 may be involved in processes that are required by all cells, such as protein secretion, metabolism and translation.

Studies in other invertebrates suggested a correlation between gene body methylation, gene length, and age [14]. Transcripts in each *O. bimaculoides* Methylation Pattern had a large range of average lengths (from TSS to TTS), with the highest variability in fully methylated transcripts (Pattern 1), with slightly shorter genes in Pattern 4 and slightly longer, on average transcripts in Pattern 2 and 3 (Figure S7D). In summary, this analysis shows that there is a similar pattern of DNA methylation on genes across tissues that positively correlates with gene expression, and genes with distinct functions are marked by similar methylation patterns.

Conserved epigenetic machinery elicits a dynamic pattern of histone modifications in octopus tissues

H3K9me3 and H3K27me3 are highly conserved marks of repressed chromatin, while H3K4me3 marks actively transcribed genes and histone acetylation serves to open chromatin. We investigated the conservation of some well-studied enzymes that write these marks. There was a many-to-one human-octopus ortholog of the H3K4me3 histone methyl transferase (SETD1B in humans; Figure 5A) and a many-to-one human-octopus ortholog of the H3K9me3 histone methyl transferase (SETDB1 in humans; Figure 5B); both showing high conservation across metazoans. A many-to-one human-octopus ortholog of the H3K27me3 methyltransferases (human EZH2) is present in octopus and conserved across species (Figure S8A). Representative examples from the histone acetylation modifying enzymes KAT2A and HDAC8 revealed a many-to-one human-octopus ortholog of KAT2A and a one-to-one human-octopus ortholog of HDAC8 in octopus and conserved in other animals (Figure S8B-C).

We next examined the expression profile of these genes and others categorized by Trinotate identification (Table S6) as regulators of histone methylation (Figure 5C and 9A) and acetylation (Figure 9B). A tissue-specific expression profile was identified for most of these genes, with the testes and neuronal tissue expressing the highest levels of the H3K4 and H3K9 methyltransferases as well as high levels of the H3K4 demethylases (Figure 5C, Table S6), and genes that erase H3K27me2/3 were higher in the arms and hatchlings. Interestingly, expression analysis of all histone methyltransferase clearly show that specific subsets of HMTs characterize a particular tissue (Figure S9A). For instance, several of the enzymes that remove methylation from H3K4 (KDM5A/B) are enriched in testes, retina, brain and suckers, while skin and suckers are enriched for the demethylases that act on H3K9 or H3K36 (i.e. KDM4A/C; Figure 5C).

We determine if these enzymes were functionally active in *O. bimaculoides* using Western blotting for H3K27me3, H3K9me3 and H3K4me3 in the arm, brain, gills, and hatchling. This showed a distinct patterns in each tissue: H3K4me3 was detected at high levels in brain and gills and lower levels in the arm, while H3K9me3 was higher in brain compared to gills and arm (Figure 5D). H3K27me3 showed comparable level in brain and gills and lower in arm, albeit overall lower level in all samples compared to H3K4me3 and H3K9me3. The low levels of all histone modifications investigated in hatchling could reflect the low level of expression of all histone methyltransferases in this sample. Although the complex signature of gene expression makes it difficult to predict which histone modification marks prevail in each tissue, these data suggest histone code plays a prominent role in regulating tissue identity.

DNA methylation and histone marks are present in squid and bobtail squid

To determine if the epigenetic modifications identified in *O. bimaculoides* were conserved in other cephalopods, we analyzed tissues from *D. pealeii* (longfin inshore squid) and *E. berryi* (bobtail squid) for 5mC levels by slot blot H3K27me3, H3K9me3 and H3K4me3 by Western blot. 5mC levels in *D. pealeii* and *E. berryi* tissues were similar to those found in *O. bimaculoides* and markedly lower than levels in mouse or zebrafish (Figure 6A-B). Histone marks were detected in arm tissue from both squid and bobtail squid (Figure 6C). Interestingly, H3K27me3 and H3K4me3 showed comparable levels to the mouse liver, while H3K9me3 was high in squid and bobtail squid arms compared to mouse tissue (Figure 6C-D). Therefore, the mechanisms that regulate the 5mC and histone methylation are conserved in cephalopods.

Discussion

The role of epigenetics in cell identity, plasticity, development, and disease is one of the most investigated aspects of biology that is heavily bias towards vertebrate models. This study identifies new features of the epigenome in cephalopods including DNA and histone methylation. We show that DNMT1 and UHRF1 are evolutionary conserved and identify critical residues necessary for their distinct functions identical in human and octopus homologs. Interestingly, in most species, DNMT1 and UHRF1 evolved together with DNA methylation, so that the presence of these genes predicts the presence of CpG methylation [3, 16-18].

We demonstrate functionality of these factors and those that methylate histone tails across multiple tissues in multiple cephalopod tissues and species. Less than 10% of all CpGs are methylated in either developing or adult tissues from *O. bimaculoides.* This is in stark

contrast to vertebrates, where over 80% of CpGs are methylated and to popular invertebrate model organisms, which lack DNA methylation entirely. In octopus, methylated CpGs are clustered in the gene bodies of a subset of transcripts and are absent from transposons and other repetitive sequences. Importantly, this pattern is largely static across developing and adult tissues and correlates with genes that have similar levels of expression across many tissues. In contrast, histone methylation levels are different across tissues in octopus, squid, and bobtail squid. This suggests the intriguing idea that genes which are not highly tissue specific are maintained by DNA methylation whereby those that are dynamically expressed are regulated by a histone code or other epigenetic features.

Evolution, by convergent or stochastic independent events, has selected for variations in the canonical patterns of methylation [3, 12], illustrated by the high level of DNA methylation across the genome of the sponge *A. queenslandica* [17], the clustering of DNA methylation on repetitive elements in the centipede *S. maritima* [88] and the intermediate pattern of methylation in *C. intestinalis*, where it is high on gene bodies and intermediate on repetitive elements [9, 10, 19]. The methylome in *O. bimaculoides* appears be more typical, as it resembles the gene body methylation pattern found in other mollusks [13, 14, 20, 35] and many invertebrates [3, 8, 9, 13, 17]. While gene body methylation is positively correlated with genes expression in organisms ranging from sponges [17, 18] to oysters [13, 20] to mammals [9, 12], it is not clear how this functions. Importantly, we found CpG levels and the pattern of DNA methylation to be nearly identical across adult and developing tissues in *O. bimaculoides* and to be at the same level in all tissues examined in other cephalopods. Interestingly, our finding that the most methylated and the most unmethylated genes in octopus had similar expression patterns across tissues suggests that CpG methylation has the same function in different cell types and argues against this epigenetic modification as a player in the dynamic regulation of gene expression.

In vertebrates, DNA methylation represses transposons [4]. We [89-91] discovered that zebrafish who have lost DNA methylation have multiple severe phenotypes including embryonic lethality, apoptosis, cell cycle defects and innate immune activation [62, 92]. These findings are concurrent with studies in oysters [93] and an annelid worm [18] where blocking DNA methylation causes developmental defects and prevents regeneration. The phenotypes in zebrafish which lack DNA methylation are in part, attributed to activation of TEs [62, 92]. However, while TEs represents about half of the octopus [44], there is a stark contrast between high levels of DNA methylation on TEs in vertebrates and virtual absence of TE methylation in cephalopods and most other invertebrates. This raises the important questions of: what represses TEs in these species and, if DNA methylation is not regulating

TEs in octopus, what is it's function? Answering these questions will require experiments where the DNA methylation machinery can be manipulated in these animals.

In contrast to the consistency of overall level and pattern of DNA methylation across octopus tissues, we found that histone modification levels differ, suggesting a specific histone code regulates the tissue-specific transcriptome. The report of H3K4, H3K9 and H3K36 methylation changes during oyster development [94] suggests that these are dynamic and important regulators of mollusk development. An interesting study profiling H3K4me1/me2/me3, H3K36me3 and H3K27Ac in anemone showed that the epigenetic landscape was similar to that found in bilaterians, with conserved regulation for enhancers in this species [95]. Moreover, the finding that the histone pattern in organisms with diverse DNA methylomes recapitulates the vertebrate pattern – such as the planarian *Schmidtea mediterranea,* which lacks DNA methylator complexes and DNA methylation, as well the highly methylated sponge *A. queenslandica* [96, 97] – suggests that many organisms decouple these different marks in patterning the epigenetic landscape. Future studies integrating histone and DNA methylation profiling with transcriptomics in cephalopod tissues can address how these patterns are integrated.

This work uncovers previously unknown features of the octopus epigenome, which can expand our understanding of epigenetic functionality beyond the few species that are used as a paradigm for knowledge in this field. Studying non-model organisms opens new challenges. For instance, annotation of epigenetic modifiers is based on functions that homologous proteins have in mammals and a thorough functional investigation for each homologous protein should be performed to unequivocally determine their activity in cephalopods. Unfortunately, size and temperate life cycle features of *O. bimaculoides* limits its possibilities for genetic manipulation. Squid, where genetic engineering has been demonstrated [46], and other cephalopod species which are actively being developed as genetic models [47] can serve as alternatives. Identifying the epigenetic patterns of octopus is a first important step in deciphering the how these patterns function to regulate the extraordinary features of these animals will be important future steps.

Acknowledgments

We gratefully acknowledge the staff in the Marine Resources Center at the Marine Biological Laboratory and in NYUAD Genomics Core Technology Platform, especially Nizar Drou for assistance with annotating the octopus genome. Patrice Delaney and Catherine Palmer provided assistance with the sample collection, Anjana Ramdas Nair edited the manuscript, and all the members of the Sadler lab provided insightful discussion. We especially thank Kirsten Peramba for assistance in octopus care and sample collection and Jan Hsiao and Lola Deng for assistance in building the phylogenomic pipeline. Funding was from the National Science Foundation, NIH R21 (MH119646), HFSP Research Program Grant (RGP0060/2017), Vetlesen Foundation Fellowship (to EE) and the NYUAD Faculty Research Fund (to KCS).

Author Contributions

The project was conceived and designed by KCS and FM, samples were collected by EE and KCS, analysis was carried out by FM and EE; FM, EE and KCS wrote the paper.

Figure legends

Figure 1. Expression profiling identifies tissue specific gene-clusters with distinct functional annotation.

A. Schematic representation of *O. bimaculoides* anatomy highlighting the tissues analyzed in this study. **B**. Heatmap of the expression profile of 13 different tissues (including 12 different tissues derived from adult animals and one whole-body hatchling). Rows are divided into 13 clusters calculated on the hierarchical clustering of dendrogram (Euclidean distance). **C.** Gene Ontology (GO) of genes in each cluster. Dot size represents gene ratio between observed and expected transcripts in each GO category (with padj < 0.05), dot colors represent adjusted p-value and numbers inside dots indicate number of transcripts for each GO term in each cluster.

Figure 2. DNA methylation machinery is conserved and differentially expressed across different tissues

Phylogenetic trees of **A.** DNMT1 and **B.** UHRF1 in a representative subset of 19 metazoan and outgroup species. **C**. DNMT1 domain structure in *H. sapiens* and *O. bimaculoides*. Numbers indicate amino acid residues for each species. **D**. Alignment of the C-terminal Catalytic Domain (CTD) of *O. bimaculoides* to *M. musculus*, *H. sapiens* and *D. rerio* shows that the major residue needed for DNMT1 catalytic function is highly conserved among the species. Residue functionality was assigned based on the mouse DMNT1 ortholog. **E.** Domain structure of UHRF1 in *H. sapiens* and *O. bimaculoides*. **F.** Alignment of the *O. bimaculoides* SRA domain to *M. musculus*, *H. sapiens* and *D. rerio* shows that all major residues needed for UHRF1 functionality are conserved among the species. Residue functionality is based on the mouse UHRF1 ortholog [24, 26]. Alignment to UHRF2 shows no conservation of the critical residues between the SRA domain in *O. bimaculoides* and UHRF2 in *H. sapiens* and *M. musculus*. **G.** Structural superposition of the 3D structure of BAH1, BAH2, and CTD domains in *M. musculus* (grey) with the 3D model of the same domains in *O. bimaculoides* (red). DNA is represented in brown and critical residues for 5mC deposition are highlighted in red. **H.** Structural superposition of the 3D structure of the SRA domain in *M. musculus* (grey) with the 3D model of the same domain in *O. bimaculoides* (green). DNA is represented in brown and residues critical for CpGs recognition and 5mC base flipping are highlighted in red. **I**. Expression profiles of DNA methylation machinery. Gene names are extracted from Trinotate (Table S6).

Figure 3. DNA methylation is enriched in a subset of gene bodies in *O. bimaculoides*.

A. Slot blot of gDNA extracted from the distal arm tip (1.5 cm of right arm 2), brain (optical lobe), and gills of one representative *O. bimaculoides* adult animal, and whole 30 day od hatchling (same biological sample used for RNA-seq and RRBS). gDNA extracted from one representative male mouse adult liver and one representative pool of whole 5 dpf zebrafish larvae was blotted on the same membrane for comparison. The 5mC signal was normalized to double stranded DNA (dsDNA) and then each sample was normalized to levels zebrafish larvae. Each dot represents one biological replicate. p-value were calculated by unpaired parametric one-way ANOVA test adjusted with Tukey's multiple comparisons test. Adjusted p-value are indicated as *** < 0.001, ** < 0.01. Comparisons other than with zebrafish or mouse resulted in not significant adjusted p-values > 0.05 (not indicated in the graph). **B.** Scatter plot of DNA methylation levels of common CpGs across Supra Esophageal (Supra E) and Sub Esophageal (Sub E) brain. Dot color represents DNA methylation levels in hatchling and size of the dots indicates scaled proportion of CpGs represented by each dot. **C.** Genomic annotation of CpGs contained in *O. bimaculoides* genome, and of those covered by WGBS in brain and RRBS in hatchling. CpGs were divided into methylated (> 80%) and not methylated (< 20%) CpGs and relative genomic annotation was performed. **D.** Metaplot represents DNA methylation levels of transposable elements in Supra E and Sub E brain and hatchling. CpG density of octopus genome is represented for the same region. Each region is divided in 15 bins. **E.** Metaplot represents the DNA methylation levels in Supra E and Sub E brain and hatchling for full-length transcripts. CpG density of octopus genome has been represented for the same region. Each transcript has been divided in 30 bins from Transcription Star Site (TSS) to Transcription Termination Site (TTS).

Figure 4. DNA methylation on gene bodies correlates with gene expression

A. Heatmap of DNA methylation levels detected in Supra E brain samples across full-length transcripts and 2000 bp upstream and downstream of start site. All transcripts have been divided in 4 clusters by k-means based on the DNA methylation average across each transcript. **B.** Line plot represents average of DNA methylation of Supra E and Sub E brain and hatchling for each cluster defined in the heatmap on Supra E sample. **C.** Violin plot displays the distribution of transcript expression values (as log2(TPM+1)) for each methylation pattern in Supra E brain samples. p-values were calculated by unpaired non-parametric Kruskal-Wallis test adjusted with Dunn's multiple comparisons test. **** indicates adjusted p-value < 0.0001, and n.s. indicates not significant adjusted p-values > 0.9999. **D.** Overall DNA methylation of transcripts (TSS to TTS of each transcript) in Supra E brain is regressed against transcripts expression (log2(TPM+1)). Transcripts were grouped by percentile of expression values and each dot represents the average value of DNA methylation for each percentile. **E.** The expression profile (as log2(TPM+1)) of transcripts

categorized in each methylation pattern have been represented across the 13 different tissues. Column clustering is supervised according to the sample order established in Fig 1B. Row clustering is calculated on the hierarchical clustering of dendrogram (Euclidean distance). **F.** Gene Ontology (GO Biological Process) annotation of transcripts in each methylation pattern. Dot size represents gene ratio between observed and expected transcripts in each GO category (with padj < 0.05), dot colors represent adjusted p-value and numbers inside the dots indicate number of transcripts for each GO term.

Figure 5. Conserved epigenetic machinery elicits a dynamic pattern of histone modifications in different octopus tissues

A. Phylogenetic tree of SETD1B, responsible for H3K4me3 deposition, in a representative subset of 19 metazoan and outgroup species. **B.** Phylogenetic tree of SETDB1, responsible for H3K9me3 deposition, in a representative subset of 19 metazoan and outgroup species. **C**. Heatmap of the main histone methylation factors, methyltransferase (writers) and de-methylases (erasers). **D.** Western blot of histone H3 and relative modification (H3K4me3, H3K9me3, H3K27me3) performed on arm (1.5 cm of Right Arm 2), brain (optical lobe), gills and hatchling of *O. bimaculoides*, in comparison with quiescent mouse adult liver. Quantification of each sample measured by western blot was normalized to histone H3 and each histone mark was normalized to the relative mark in mouse liver as control. Each dot represents one biological replicate.

Figure 6. DNA methylation and histone marks are present in squid, and bobtail squid.

A. Slot blot detecting 5mC on gDNA extracted from arm (2-3 cm of Right Arm 2), brain (optical lobe), gills of one representative animal of *D. pealeii* and *E. berryi*, the analogous tissues of *O. bimaculoides* adults and a 30 dpf hatchling, one representative male mouse adult liver and one representative pool of 5 dpf larvae of zebrafish. **B.** Quantification of 5-mC measured by slot blot was normalized to double stranded DNA (dsDNA) and each sample was normalized to zebrafish larvae as control. Each dot represents one biological replicate. **C.** Western blot of histone H3 and relative modification (H3K4me3, H3K9me3, H3K27me3) performed on an arm of *D. pealeii* and *E. berryi*, in comparison with quiescent mouse adult liver. **D.** Quantification of each sample measured by Western blot was normalized to histone H3 and each histone mark was normalized to the relative mark in mouse liver as control. Each dot represents one biological replicate.

Supplemental Figure legends

Supplemental Figure 1. Representative images of *O. bimaculoides.*

A. Picture of anesthetized adult male. **B.** Representative image of *O. bimaculoides* arm for collection of distal, medial and proximal samples. **C.** A clutch of 30 dpf hatchlings and **D.** the hatchling used for DNA, RNA and protein extraction.

Supplemental Figure 2. Distinct biological processes and cellular components of differentially expressed genes octopus tissues.

GO analysis of **A.** Biological Process and **B.** Cellular Component of each gene cluster identified in Figure 1.

Supplemental Figure 3. Extended phylogenetic analysis shows high conservation of DNMT1 and UHRF1.

A. Phylogenetic tree of DNMT1 in a representative subset of 50 metazoan and outgroup species. **B.** Phylogenetic tree of UHRF1 in a representative subset of 50 metazoan and outgroup species.

Supplemental Figure 4. *O. bimaculoides* genome encodes conserved features of DNMT1 and UHRF1 but not UHRF2.

A. Similarity for each domain of *O. bimaculoides* DNMT1 determined by HMMER. Bit score indicates homology score. **B**. Table shows percentage of sequence identity calculated by ClustalOmega for the *O. bimaculoides* DNMT1 protein compared to human, mouse, and zebrafish in a multiple sequence alignment. **C.** Table shows similarity for each domain (with a PFAM ID) of *O. bimaculoides* UHRF1 when compared to the proteome database using HMMER. Bit score indicates homology score. **D.** Table shows percentage of sequence identity calculated by ClustalOmega for the *O. bimaculoides* UHRF1 protein compared to human, mouse, and zebrafish in a multiple sequence alignment. **E.** Similarity for each domain of the *O. bimaculoides* YDG_OCTBM protein determined using HMMER. Bit score indicates homology score. **F.** Percent identity calculated by ClustalOmega for the *O. bimaculoides* YDG_OCTBM protein compared to human, mouse, and zebrafish in a multiple sequence alignment. **G.** Alignment of the SRA domain in *O. bimaculoides* YDG_OCTBM to human, mouse, and zebrafish shows that all the major residues needed for the correct functionality of UHRF1 are not conserved in YDG_OCTBM of *O.bimaculoides*. Alignment to UHRF2 shows no conservation of critical residues between YDG_OCTBM in *O. bimaculoides* and UHRF2 in human, mouse. Residues functionality is based on mouse orthologs

Supplemental Figure 5. Structural modelling of DNMT1 and UHRF1.

A. 3D structure of the BAH1, BAH2 and CTD domains of DNMT1 in *M. musculus*. **B.** 3D model of BAH1, BAH2 and CTD domains of DNMT1_OCTBM in *O. bimaculoides*. **C.** 3D structure of SRA domain of UHRF1 in *M. musculus*. **D.** 3D model of SRA domain of UHRF1_OCTBM in *O. bimaculoides*.

Supplemental Figure 6. Pattern of DNA methylation identified by WGBS and RRBS in octopus tissues.

A. Table describing the number and relative percentage of CpGs covered in the *O. bimaculoides* genome by each technique and sample analyzed. CpGs were classified based on the percentage of methylation as methylated (> 80%) or not methylated (< 20%). **B.** Scaled violin plot of CpGs identified by WGBS and RRBS in Supra E and Sub E brain and in one whole-body 30 dpf hatchling. Numbers on the lines indicate the percent of CpGs that are detected as >80% methylated. **C.** Box plot describing the percentage of methylation of CpGs contained in Repetitive Elements (RE) divided by class in the 30 dpf hatchling. Box-and-whisker plots have a center line at the median, lower and upper hinges correspond to first and third quartiles, and whiskers extend from hinges to largest or smallest values no further than 1.5 × IQR (inter-quartile range), while data beyond the end of the whiskers are outlying points that are plotted individually.

Supplemental Figure 7. Relationship between DNA methylation pattern gene expression, and gene length in octopus.

A. Heatmap of DNA methylation of all full-length transcripts and 2000 bp upstream and downstream as detected in the Sub E brain WGBS data. Clustering and rank order is dictated by Supra E brain samples shown in Figure 4A. **B.** Heatmap of DNA methylation in hatchling full-length transcripts and 2000 bp upstream and downstream. Clustering and rank order is dictated by Supra E brain samples shown in Figure 4A. **C.** Overall DNA methylation of transcripts (TSS to TTS of each transcript) in Sub E brain and hatchling is regressed against transcripts expression (log2(TPM+1)). Transcripts were grouped by percentile of expression values and each dot represents the average value of DNA methylation for each percentile. **D.** Violin plot displays the distribution of transcript length (as log10(width)) in each methylation pattern. p-values were calculated by unpaired non-parametric Kruskal-Wallis test adjusted with Dunn's multiple comparisons test. **** indicates p-value adjusted < 0.0001.

Supplemental Figure 8. Supplemental phylogenetic trees.

A. Phylogenetic tree of EZH2, responsible of H3K27me3 deposition, in a representative subset of 19 metazoan and outgroup species. **B.** Phylogenetic tree of KAT2A, mainly responsible of H3K9ac deposition, in a representative subset of 19 metazoan and outgroup species. **C.** Phylogenetic tree of HADC8, mainly responsible of H3K9ac removal, in a representative subset of 19 metazoan and outgroup species.

Supplemental Figure 9. Histone methyltransferases, acetyltransferases and de-acetylases have a tissue specific expression pattern in octopus

A. Heatmap of the extended panel of histone methyltransferases. **B**. Heatmap of the main histone acetylation factors, acetyltransferase (writers) and de- acetylases (erasers).

Supplemental Table legends

Supplemental Table 1. Statistical metrics of RRBS. **A.** Statistic values of alignment performed in Bismark. **B.** Read coverage statistics per base analyzed by methylKit.

Supplemental Table 2. Reference Gene Set (RGS) sources and details of proteins utilized in phylogenomic pipeline.

Supplemental Table 3. Genome sources and details of organisms utilized in phylogenomic pipeline.

Supplemental Table 4. Trinotate table containing the putative gene names of the transcriptome of *O. bimaculoides*.

Supplemental Table 5. Table of Ensembl transcript IDs contained in each Cluster.

Supplemental Table 6. Table of Ensembl transcript IDs contained in each heatmap of epigenetic factors. The mapping across different identifiers is also reported for each transcript.

Supplemental Table 7. Table of Ensembl transcript IDs contained in each Pattern.

References

1. Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet *17*, 487-500.
2. Stricker, S.H., Koferle, A., and Beck, S. (2017). From profiles to function in epigenomics. Nat Rev Genet *18*, 51-66.
3. de Mendoza, A., Lister, R., and Bogdanovic, O. (2019). Evolution of DNA Methylome Diversity in Eukaryotes. J Mol Biol.
4. Deniz, O., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. Nature reviews. Genetics *20*, 417-431.
5. Rae, P.M., and Steele, R.E. (1979). Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of Drosophila and the rDNA of other insects. Nucleic Acids Res *6*, 2987-2995.
6. Bird, A.P., and Taggart, M.H. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res *8*, 1485-1497.
7. Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.
8. Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.
9. Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.
10. Keller, T.E., Han, P., and Yi, S.V. (2016). Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary. Mol Biol Evol *33*, 1019-1028.
11. Long, H.K., King, H.W., Patient, R.K., Odom, D.T., and Klose, R.J. (2016). Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. Nucleic Acids Res *44*, 6693-6706.
12. Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr Biol *20*, R780-785.
13. Gavery, M.R., and Roberts, S.B. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (Crassostrea gigas). BMC Genomics *11*, 483.
14. Wang, X., Li, Q., Lian, J., Li, L., Jin, L., Cai, H., Xu, F., Qi, H., Zhang, L., Wu, F., et al. (2014). Genome-wide and single-base resolution DNA methylomes of the Pacific oyster Crassostrea gigas provide insight into the evolution of invertebrate CpG methylation. BMC Genomics *15*, 1119.
15. Riviere, G., He, Y., Tecchio, S., Crowell, E., Gras, M., Sourdaine, P., Guo, X., and Favrel, P. (2017). Dynamics of DNA methylomes underlie oyster development. PLoS Genet *13*, e1006807.
16. de Mendoza, A., Poppe, D., Buckberry, S., Pflueger, J., Albertin, C.B., Daish, T., Bertrand, S., de la Calle-Mustienes, E., Gomez-Skarmeta, J.L., Nery, J.R., et al. (2021). The emergence of the brain non-CpG methylation system in vertebrates. Nat Ecol Evol *5*, 369-378.
17. de Mendoza, A., Hatleberg, W.L., Pang, K., Leininger, S., Bogdanovic, O., Pflueger, J., Buckberry, S., Technau, U., Hejnol, A., Adamska, M., et al. (2019). Convergent evolution of a vertebrate-like methylome in a marine sponge. Nat Ecol Evol *3*, 1464-1473.

18. Planques, A., Kerner, P., Ferry, L., Grunau, C., Gazave, E., and Vervoort, M. (2021). DNA methylation atlas and machinery in the developing and regenerating annelid Platynereis dumerilii. BMC Biol *19*, 148.
19. Suzuki, M.M., Kerr, A.R., De Sousa, D., and Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res *17*, 625-631.
20. Song, K., Li, L., and Zhang, G. (2017). The association between DNA methylation and exon expression in the Pacific oyster Crassostrea gigas. PLoS One *12*, e0185224.
21. Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.
22. Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and Rokhsar, D.S. (2015). The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature *524*, 220-224.
23. Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M.J., and Zhou, M.M. (2008). Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. J Biol Chem *283*, 34490-34494.
24. Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E., and Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature *455*, 826-829.
25. Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature *455*, 822-825.
26. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818-821.
27. Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.
28. Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.
29. Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., Xue, S., Duan, S., Allali-Hassani, A., Zuo, X., et al. (2011). Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem *286*, 24300-24311.
30. Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., et al. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell *43*, 275-284.
31. Xie, S., Jakoncic, J., and Qian, C. (2012). UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. J Mol Biol *415*, 318-328.
32. Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H., and Xu, Y. (2013). Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. J Biol Chem *288*, 1329-1339.
33. Fang, J., Cheng, J., Wang, J., Zhang, Q., Liu, M., Gong, R., Wang, P., Zhang, X., Feng, Y., Lan, W., et al. (2016). Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. Nat Commun *7*, 11197.

34. Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F., and Leonhardt, H. (2009). The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res.
35. Manner, L., Schell, T., Provataris, P., Haase, M., and Greve, C. (2021). Inference of DNA methylation patterns in molluscs. Philos Trans R Soc Lond B Biol Sci *376*, 20200166.
36. Sumbre, G., Gutfreund, Y., Fiorito, G., Flash, T., and Hochner, B. (2001). Control of octopus arm extension by a peripheral motor program. Science *293*, 1845-1848.
37. Ogura, A., Ikeo, K., and Gojobori, T. (2004). Comparative analysis of gene expression for convergent evolution of camera eye between octopus and human. Genome Res *14*, 1555-1561.
38. Nesher, N., Levy, G., Grasso, F.W., and Hochner, B. (2014). Self-recognition mechanism between skin and suckers prevents octopus arms from interfering with each other. Curr Biol *24*, 1271-1275.
39. Imperadore, P., Shah, S.B., Makarenkova, H.P., and Fiorito, G. (2017). Nerve degeneration and regeneration in the cephalopod mollusc Octopus vulgaris: the case of the pallial nerve. Sci Rep *7*, 46564.
40. Edsinger, E., and Dolen, G. (2018). A Conserved Role for Serotonergic Neurotransmission in Mediating Social Behavior in Octopus. Curr Biol *28*, 3136-3142 e3134.
41. van Giesen, L., Kilian, P.B., Allard, C.A.H., and Bellono, N.W. (2020). Molecular Basis of Chemotactile Sensation in Octopus. Cell *183*, 594-604 e514.
42. Lange, M.M. (1920). On the regeneration and finer structure of the arms of the cephalopods. Journal of Experimental Zoology *31*, 1-57.
43. Liscovitch-Brauer, N., Alon, S., Porath, H.T., Elstein, B., Unger, R., Ziv, T., Admon, A., Levanon, E.Y., Rosenthal, J.J.C., and Eisenberg, E. (2017). Trade-off between Transcriptome Plasticity and Genome Evolution in Cephalopods. Cell *169*, 191-202 e111.
44. Belcaid, M., Casaburi, G., McAnulty, S.J., Schmidbaur, H., Suria, A.M., Moriano-Gutierrez, S., Pankey, M.S., Oakley, T.H., Kremer, N., Koch, E.J., et al. (2019). Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc Natl Acad Sci U S A *116*, 3030-3035.
45. Vidal, E.A., Villanueva, R., Andrade, J.P., Gleadall, I.G., Iglesias, J., Koueta, N., Rosas, C., Segawa, S., Grasse, B., Franco-Santos, R.M., et al. (2014). Cephalopod culture: current status of main biological models and research priorities. Adv Mar Biol *67*, 1-98.
46. Crawford, K., Diaz Quiroz, J.F., Koenig, K.M., Ahuja, N., Albertin, C.B., and Rosenthal, J.J.C. (2020). Highly Efficient Knockout of a Squid Pigmentation Gene. Curr Biol *30*, 3484-3490 e3484.
47. Albertin, C.B., and Simakov, O. (2020). Cephalopod Biology: At the Intersection Between Genomic and Organismal Novelties. Annu Rev Anim Biosci *8*, 71-90.
48. De Oliveira, A.L., Wollesen, T., Kristof, A., Scherholz, M., Redl, E., Todt, C., Bleidorn, C., and Wanninger, A. (2016). Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. BMC Genomics *17*, 905.
49. Sadamoto, H., Takahashi, H., Okada, T., Kenmoku, H., Toyota, M., and Asakawa, Y. (2012). De novo sequencing and transcriptome analysis of the central nervous system of mollusc Lymnaea stagnalis by deep RNA sequencing. PLoS One *7*, e42546.
50. Maselli, V., Polese, G., Al-Soudy, A.S., Buglione, M., and Di Cosmo, A. (2020). Cognitive Stimulation Induces Differential Gene Expression in Octopus vulgaris: The Key Role of Protocadherins. Biology (Basel) *9*.

51. Garcia-Fernandez, P., Garcia-Souto, D., Almansa, E., Moran, P., and Gestal, C. (2017). Epigenetic DNA Methylation Mediating Octopus vulgaris Early Development: Effect of Essential Fatty Acids Enriched Diet. Front Physiol *8*, 292.
52. Diaz-Freije, E., Gestal, C., Castellanos-Martinez, S., and Moran, P. (2014). The role of DNA methylation on Octopus vulgaris development and their perspectives. Front Physiol *5*, 62.
53. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.
54. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884-i890.
55. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol *34*, 525-527.
56. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357-360.
57. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
58. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature protocols *11*, 1650-1667.
59. Garcia-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Gotz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics *28*, 2678-2679.
60. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.
61. Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., and Melnick, A.M. (2015). Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. J Vis Exp, e52246.
62. Magnani, E., Macchi, F., Madakashira, B.P., Zhang, C., Alaydaroos, F., and Sadler, K.C. (2021). uhrf1 and dnmt1 Loss Induces an Immune Response in Zebrafish Livers Due to Viral Mimicry by Transposable Elements. Front Immunol *12*, 627926.
63. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.
64. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.
65. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol *13*, R87.
66. Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., et al. (2014). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res *24*, 1613-1623.
67. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol *1962*, 227-245.
68. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236-1240.

69. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol *30*, 772-780.
70. Steenwyk, J.L., Buida, T.J., 3rd, Li, Y., Shen, X.X., and Rokas, A. (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol *18*, e3001007.
71. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol *37*, 1530-1534.
72. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res *49*, W293-W296.
73. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res *47*, W636-W641.
74. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res *46*, W200-W204.
75. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res *46*, W296-W303.
76. Studer, G., Tauriello, G., Bienert, S., Biasini, M., Johner, N., and Schwede, T. (2021). ProMod3-A versatile homology modelling toolbox. PLoS Comput Biol *17*, e1008667.
77. Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johner, N., Schenk, A.D., Philippsen, A., and Schwede, T. (2013). OpenStructure: an integrated software framework for computational structural biology. Acta Crystallogr D Biol Crystallogr *69*, 701-709.
78. Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, R., Haas, J., and Schwede, T. (2020). QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics *36*, 1765-1771.
79. Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G., et al. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Rep *18*, 762-776.
80. Nodl, M.T., Fossati, S.M., Domingues, P., Sanchez, F.J., and Zullo, L. (2015). The making of an octopus arm. Evodevo *6*, 19.
81. Aliaga, B., Bulla, I., Mouahid, G., Duval, D., and Grunau, C. (2019). Universality of the DNA methylation codes in Eucaryotes. Sci Rep *9*, 173.
82. Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol.
83. Rothbart, S.B., Dickson, B.M., Ong, M.S., Krajewski, K., Houliston, S., Kireev, D.B., Arrowsmith, C.H., and Strahl, B.D. (2013). Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev *27*, 1288-1298.
84. Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res *35*, 4301-4312.

85. Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science *335*, 709-712.
86. Zhang, C., Hoshida, Y., and Sadler, K.C. (2016). Comparative Epigenomic Profiling of the DNA Methylome in Mouse and Zebrafish Uncovers High Interspecies Divergence. Front Genet *7*, 110.
87. Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477-481.
88. de Mendoza, A., Pflueger, J., and Lister, R. (2019). Capture of a functionally active methyl-CpG binding domain by an arthropod retrotransposon family. Genome Res *29*, 1277-1286.
89. Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell *162*, 974-986.
90. Beck, M.A., Fischer, H., Grabner, L.M., Groffics, T., Winter, M., Tangermann, S., Meischel, T., Zaussinger-Haas, B., Wagner, P., Fischer, C., et al. (2021). DNA hypomethylation leads to cGAS-induced autoinflammation in the epidermis. EMBO J, e108234.
91. Sadler, K.C., Krahn, K.N., Gaur, N.A., and Ukomadu, C. (2007). Liver growth in the embryo and during liver regeneration in zebrafish requires the cell cycle regulator, uhrf1. Proc Natl Acad Sci U S A *104*, 1570-1575.
92. Chernyavskaya, Y., Mudbhary, R., Zhang, C., Tokarz, D., Jacob, V., Gopinath, S., Sun, X., Wang, S., Magnani, E., Madakashira, B.P., et al. (2017). Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. Development *144*, 2925-2939.
93. Riviere, G., Wu, G.C., Fellous, A., Goux, D., Sourdaine, P., and Favrel, P. (2013). DNA methylation is crucial for the early development in the Oyster C. gigas. Mar Biotechnol (NY) *15*, 739-753.
94. Fellous, A., Lefranc, L., Jouaux, A., Goux, D., Favrel, P., and Riviere, G. (2019). Histone Methylation Participates in Gene Expression Control during the Early Development of the Pacific Oyster Crassostrea gigas. Genes (Basel) *10*.
95. Schwaiger, M., Schonauer, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res *24*, 639-650.
96. Gaiti, F., Jindrich, K., Fernandez-Valverde, S.L., Roper, K.E., Degnan, B.M., and Tanurdzic, M. (2017). Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. Elife *6*.
97. Dattani, A., Kao, D., Mihaylova, Y., Abnave, P., Hughes, S., Lai, A., Sahu, S., and Aboobaker, A.A. (2018). Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells. Genome Res *28*, 1543-1554.

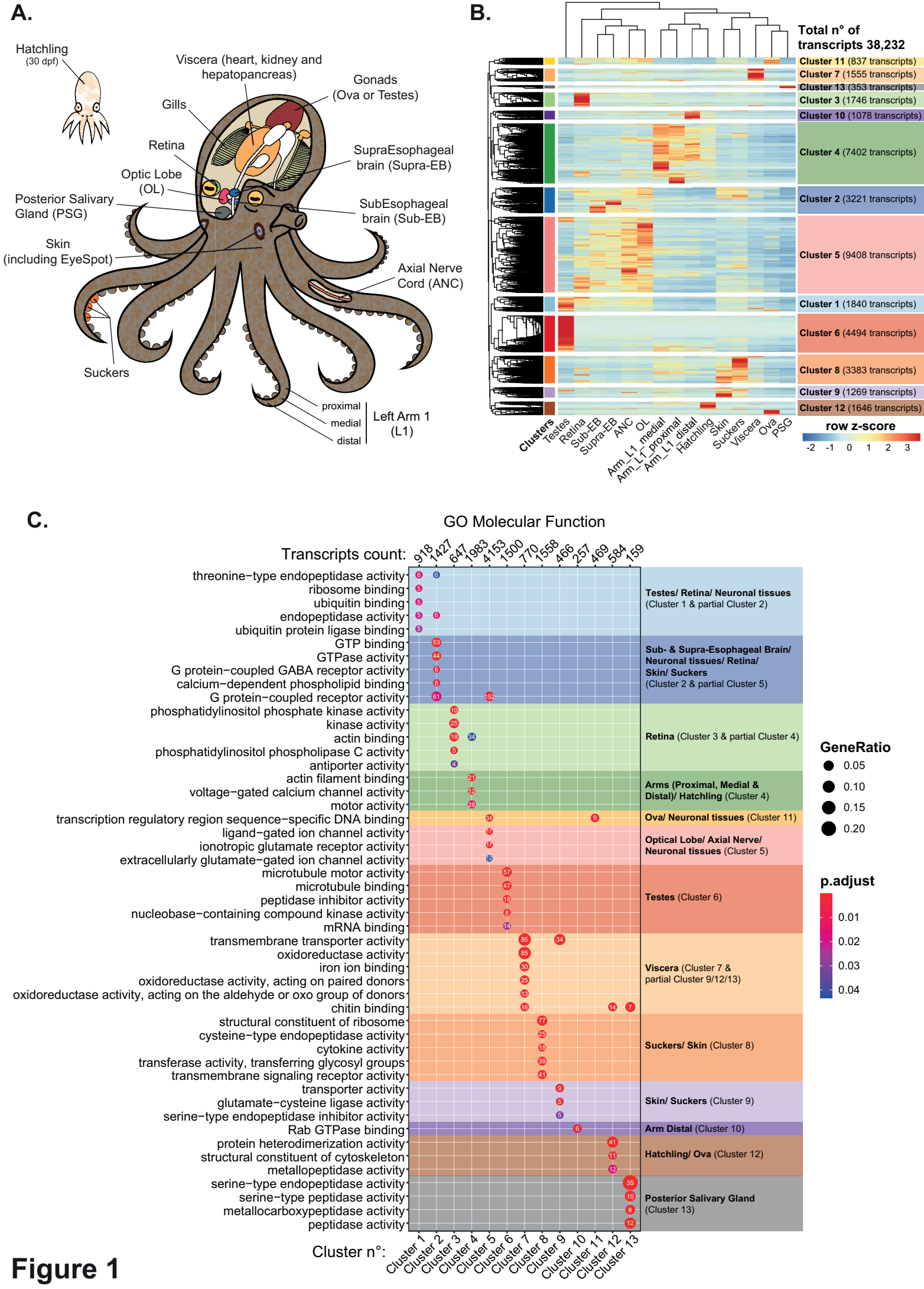
A.
Hatchling
(30 dpf)
Viscera (heart, kidney and hepatopancreas)
Gonads
(Ova or Testes)
Gills
Retina
SupraEsophageal brain (Supra-EB)
Optic Lobe
(OL)
Posterior Salivary Gland (PSG)
SubEsophageal brain (Sub-EB)
Skin
(including EyeSpot)
Axial Nerve Cord (ANC)
Suckers
proximal
medial
distal
Left Arm 1
(L1)
B.
Total n° of transcripts 38,232
Cluster 11 (837 transcripts)
Cluster 7 (1555 transcripts)
Cluster 13 (353 transcripts)
Cluster 3 (1746 transcripts)
Cluster 10 (1078 transcripts)
Cluster 4 (7402 transcripts)
Cluster 2 (3221 transcripts)
Cluster 5 (9408 transcripts)
Cluster 1 (1840 transcripts)
Cluster 6 (4494 transcripts)
Cluster 8 (3383 transcripts)
Cluster 9 (1269 transcripts)
Cluster 12 (1646 transcripts)
Clusters
Testes
Retina
Sub-EB
Supra-EB
ANC
OL
Arm_L1_medial
Arm_L1_proximal
Arm_L1_distal
Hatchling
Skin
Suckers
Viscera
Ova
PSG
row z-score
-2 -1 0 1 2 3
C.
GO Molecular Function
Transcripts count:
918 1427 647 1983 4153 1500 770 1558 466 257 469 584 159
threonine−type endopeptidase activity
ribosome binding
ubiquitin binding
endopeptidase activity
ubiquitin protein ligase binding
GTP binding
GTPase activity
G protein−coupled GABA receptor activity
calcium−dependent phospholipid binding
G protein−coupled receptor activity
phosphatidylinositol phosphate kinase activity
kinase activity
actin binding
phosphatidylinositol phospholipase C activity
antiporter activity
actin filament binding
voltage−gated calcium channel activity
motor activity
transcription regulatory region sequence−specific DNA binding
ligand−gated ion channel activity
ionotropic glutamate receptor activity
extracellularly glutamate−gated ion channel activity
microtubule motor activity
microtubule binding
peptidase inhibitor activity
nucleobase−containing compound kinase activity
mRNA binding
transmembrane transporter activity
oxidoreductase activity
iron ion binding
oxidoreductase activity, acting on paired donors
oxidoreductase activity, acting on the aldehyde or oxo group of donors
chitin binding
structural constituent of ribosome
cysteine−type endopeptidase activity
cytokine activity
transferase activity, transferring glycosyl groups
transmembrane signaling receptor activity
transporter activity
glutamate−cysteine ligase activity
serine−type endopeptidase inhibitor activity
Rab GTPase binding
protein heterodimerization activity
structural constituent of cytoskeleton
metallopeptidase activity
serine−type endopeptidase activity
serine−type peptidase activity
metallocarboxypeptidase activity
peptidase activity
Testes/ Retina/ Neuronal tissues
(Cluster 1 & partial Cluster 2)
Sub- & Supra-Esophageal Brain/ Neuronal tissues/ Retina/ Skin/ Suckers
(Cluster 2 & partial Cluster 5)
Retina (Cluster 3 & partial Cluster 4)
Arms (Proximal, Medial & Distal)/ Hatchling (Cluster 4)
Ova/ Neuronal tissues (Cluster 11)
Optical Lobe/ Axial Nerve/ Neuronal tissues (Cluster 5)
Testes (Cluster 6)
Viscera (Cluster 7 & partial Cluster 9/12/13)
Suckers/ Skin (Cluster 8)
Skin/ Suckers (Cluster 9)
Arm Distal (Cluster 10)
Hatchling/ Ova (Cluster 12)
Posterior Salivary Gland
(Cluster 13)
GeneRatio
0.05
0.10
0.15
0.20
p.adjust
0.01
0.02
0.03
0.04
Cluster n°:
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5 Cluster 6 Cluster 7 Cluster 8 Cluster 9 Cluster 10 Cluster 11 Cluster 12 Cluster 13

Figure 1

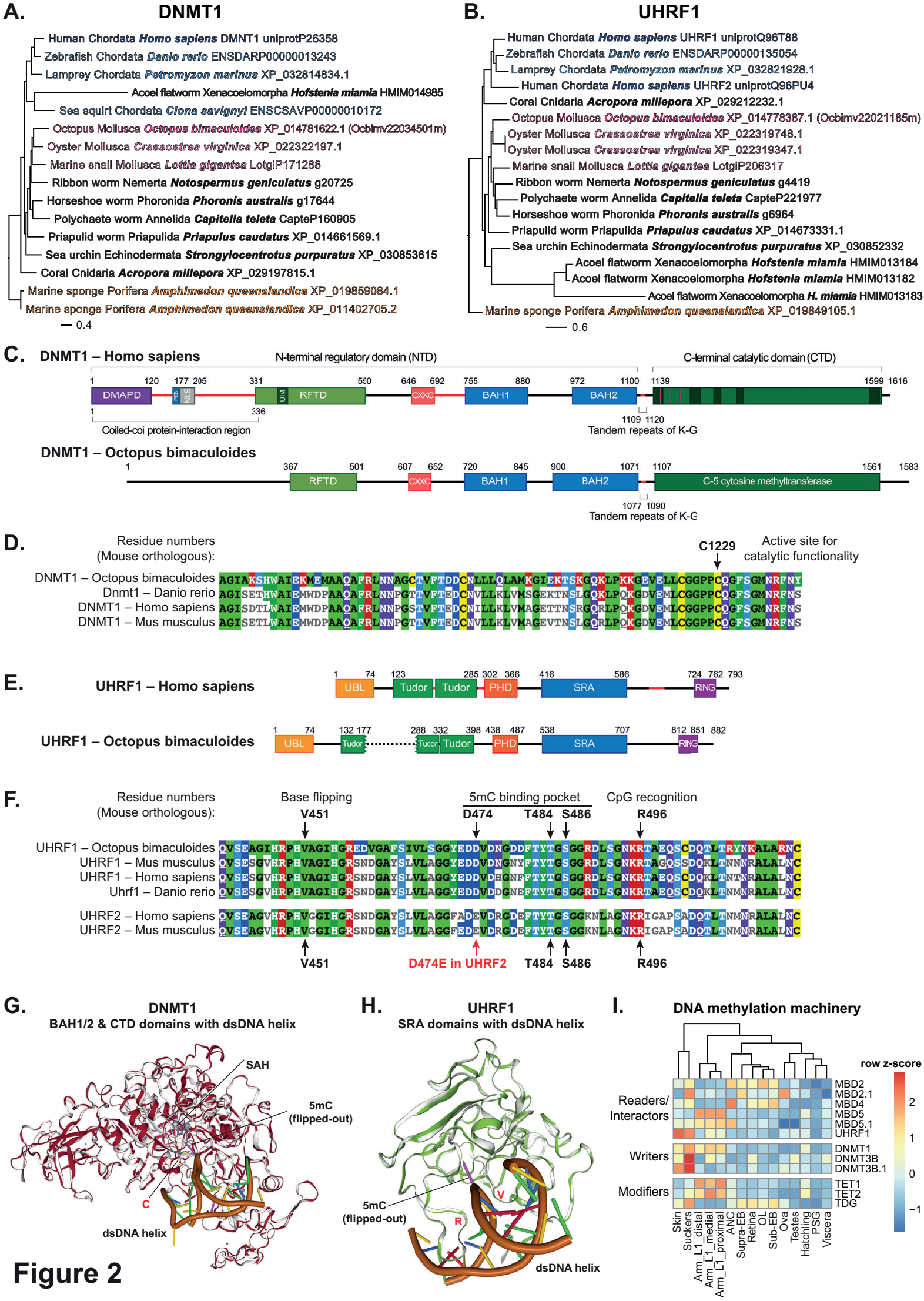
A.
DNMT1
Human Chordata Homo sapiens DNMT1 uniprotP26358
Zebrafish Chordata Danio rerio ENSDARP00000013243
Lamprey Chordata Petromyzon marinus XP_032814834.1
Acoel flatworm Xenacoelomorpha Hofstenia miamia HMIM014985
Sea squirt Chordata Ciona savignyi ENSCSAVP00000010172
Octopus Mollusca Octopus bimaculoides XP_014781622.1 (Ocbimv22034501m)
Oyster Mollusca Crassostrea virginica XP_022322197.1
Marine snail Mollusca Lottia gigantea LotgiP171288
Ribbon worm Nemerta Notospermus geniculatus g20725
Horseshoe worm Phoronida Phoronis australis g17644
Polychaete worm Annelida Capitella teleta CapteP160905
Priapulid worm Priapulida Priapulus caudatus XP_014661569.1
Sea urchin Echinodermata Strongylocentrotus purpuratus XP_030853615
Coral Cnidaria Acropora millepora XP_029197815.1
Marine sponge Porifera Amphimedon queensland ica XP_019859084.1
Marine sponge Porifera Amphimedon queensland ica XP_011402705.2
0.4
B.
UHRF1
Human Chordata Homo sapiens UHRF1 uniprotQ96T88
Zebrafish Chordata Danio rerio ENSDARP00000135054
Lamprey Chordata Petromyzon marinus XP_032821928.1
Human Chordata Homo sapiens UHRF2 uniprotQ96PU4
Coral Cnidaria Acropora millepora XP_029212232.1
Octopus Mollusca Octopus bimaculoides XP_014778387.1 (Ocbimv22021185m)
Oyster Mollusca Crassostrea virginica XP_022319748.1
Oyster Mollusca Crassostrea virginica XP_022319347.1
Marine snail Mollusca Lottia gigantea LotgiP206317
Ribbon worm Nemerta Notospermus geniculatus g4419
Polychaete worm Annelida Capitella teleta CapteP221977
Horseshoe worm Phoronida Phoronis australis g6964
Priapulid worm Priapulida Priapulus caudatus XP_014673331.1
Sea urchin Echinodermata Strongylocentrotus purpuratus XP_030852332
Acoel flatworm Xenacoelomorpha Hofstenia miamia HMIM013184
Acoel flatworm Xenacoelomorpha Hofstenia miamia HMIM013182
Acoel flatworm Xenacoelomorpha H. miamia HMIM013183
Marine sponge Porifera Amphimedon queensland ica XP_019849105.1
0.6
C.
DNMT1 – Homo sapiens
N-terminal regulatory domain (NTD)
C-terminal catalytic domain (CTD)
DMAPD
PBD
NLS
UIM
RFTD
CXXC
BAH1
BAH2
Coiled-coi protein-interaction region
Tandem repeats of K-G
DNMT1 – Octopus bimaculoides
RFTD
CXXC
BAH1
BAH2
C-5 cytosine methyltransferase
Tandem repeats of K-G
D.
Residue numbers (Mouse orthologous):
C1229
Active site for catalytic functionality
DNMT1 – Octopus bimaculoides
Dnmt1 – Danio rerio
DNMT1 – Homo sapiens
DNMT1 – Mus musculus
E.
UHRF1 – Homo sapiens
UBL
Tudor
Tudor
PHD
SRA
RING
UHRF1 – Octopus bimaculoides
UBL
Tudor
Tudor
Tudor
PHD
SRA
RING
F.
Residue numbers (Mouse orthologous):
Base flipping
5mC binding pocket
CpG recognition
V451
D474
T484
S486
R496
UHRF1 – Octopus bimaculoides
UHRF1 – Mus musculus
UHRF1 – Homo sapiens
Uhrf1 – Danio rerio
UHRF2 – Homo sapiens
UHRF2 – Mus musculus
V451
D474E in UHRF2
T484
S486
R496
G.
DNMT1
BAH1/2 & CTD domains with dsDNA helix
SAH
5mC (flipped-out)
C
dsDNA helix
H.
UHRF1
SRA domains with dsDNA helix
5mC (flipped-out)
V
R
dsDNA helix
I.
DNA methylation machinery
row z-score
Readers/ Interactors
MBD2
MBD2.1
MBD4
MBD5
MBD5.1
UHRF1
Writers
DNMT1
DNMT3B
DNMT3B.1
Modifiers
TET1
TET2
TDG
Skin
Suckers
Arm_L1_distal
Arm_L1_medial
Arm_L1_proximal
ANC
Supra-EB
Retina
OL
Sub-EB
Ova
Testes
Hatchling
PSG
Viscera

Figure 2

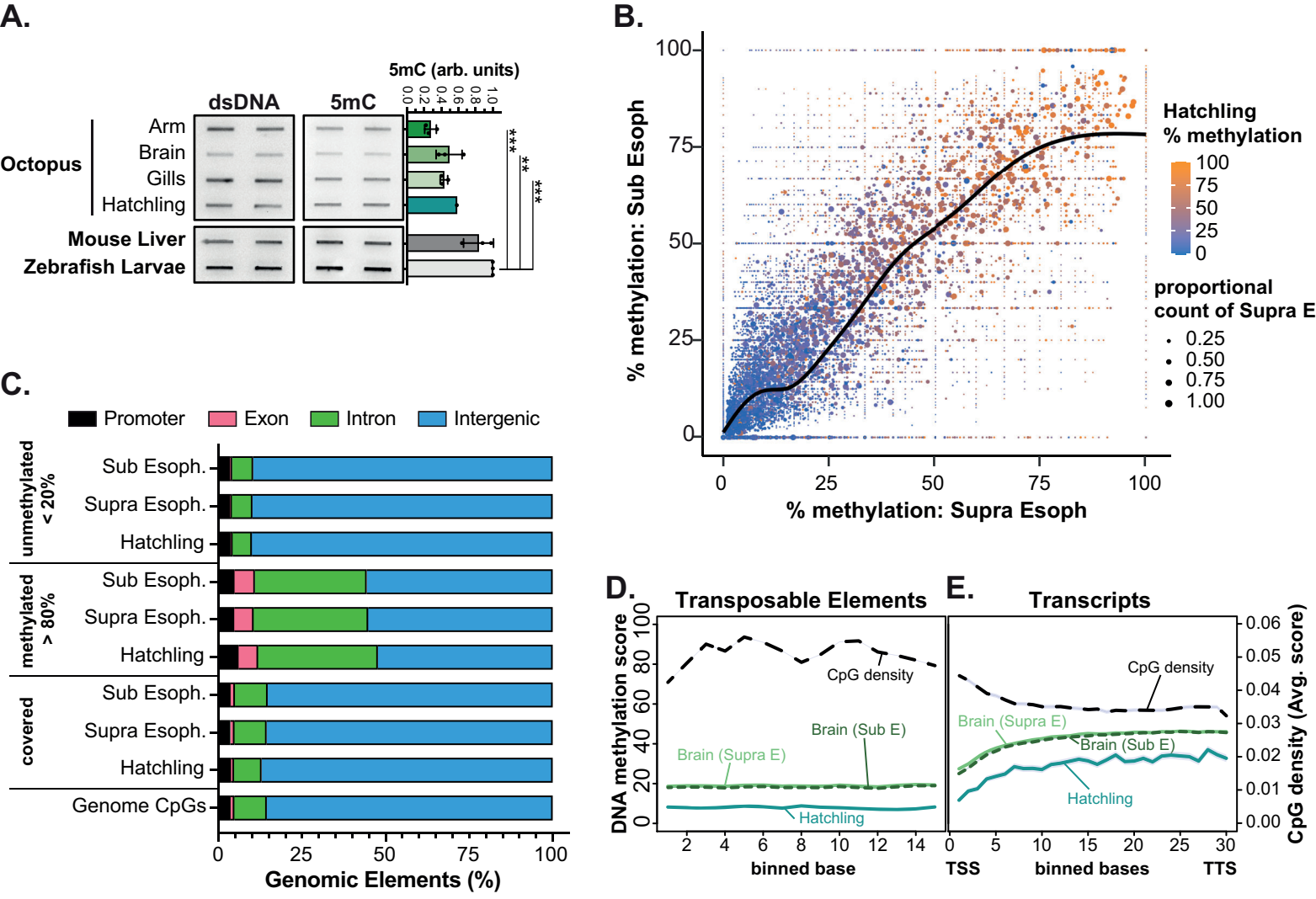
A.
5mC (arb. units)
dsDNA
5mC
Octopus
Arm
Brain
Gills
Hatchling
Mouse Liver
Zebrafish Larvae
B.
% methylation: Sub Esoph
% methylation: Supra Esoph
Hatchling % methylation
proportional count of Supra E
C.
Promoter
Exon
Intron
Intergenic
unmethylated < 20%
methylated > 80%
covered
Sub Esoph.
Supra Esoph.
Hatchling
Genome CpGs
Genomic Elements (%)
D.
Transposable Elements
DNA methylation score
CpG density
Brain (Sub E)
Brain (Supra E)
Hatchling
binned base
E.
Transcripts
CpG density (Avg. score)
TSS
binned bases
TTS

Figure 3

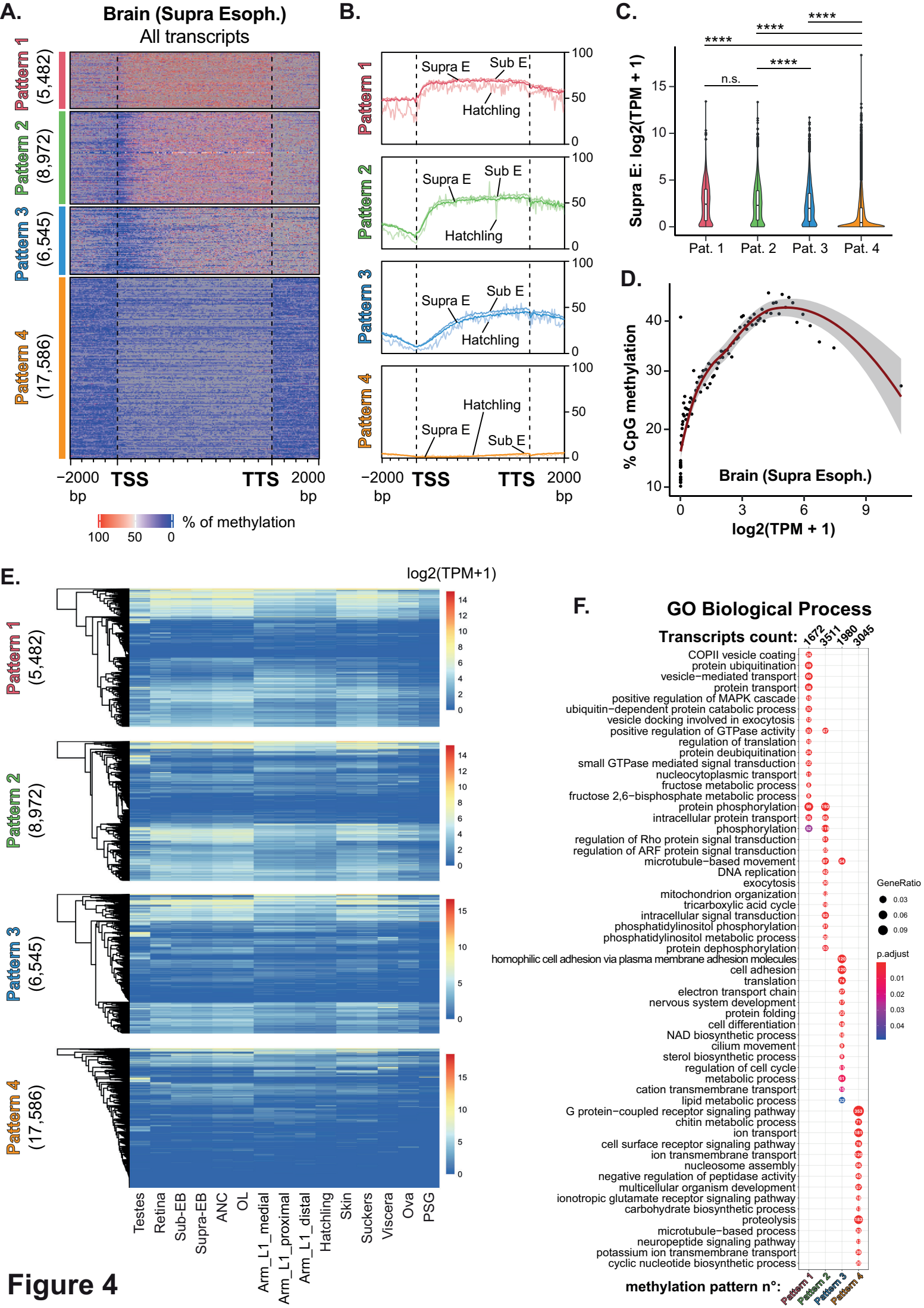
A.
Brain (Supra Esoph.)
All transcripts
Pattern 1 (5,482)
Pattern 2 (8,972)
Pattern 3 (6,545)
Pattern 4 (17,586)
−2000 bp
TSS
TTS
2000 bp
% of methylation
100
50
0
B.
Supra E
Sub E
Hatchling
C.
Supra E: log2(TPM + 1)
n.s.
****
Pat. 1
Pat. 2
Pat. 3
Pat. 4
D.
% CpG methylation
Brain (Supra Esoph.)
log2(TPM + 1)
E.
log2(TPM+1)
Testes
Retina
Sub-EB
Supra-EB
ANC
OL
Arm_L1_medial
Arm_L1_proximal
Arm_L1_distal
Hatchling
Skin
Suckers
Viscera
Ova
PSG
F.
GO Biological Process
Transcripts count:
1672
3511
1980
3045
COPII vesicle coating
protein ubiquitination
vesicle−mediated transport
protein transport
positive regulation of MAPK cascade
ubiquitin−dependent protein catabolic process
vesicle docking involved in exocytosis
positive regulation of GTPase activity
regulation of translation
protein deubiquitination
small GTPase mediated signal transduction
nucleocytoplasmic transport
fructose metabolic process
fructose 2,6−bisphosphate metabolic process
protein phosphorylation
intracellular protein transport
phosphorylation
regulation of Rho protein signal transduction
regulation of ARF protein signal transduction
microtubule−based movement
DNA replication
exocytosis
mitochondrion organization
tricarboxylic acid cycle
intracellular signal transduction
phosphatidylinositol phosphorylation
phosphatidylinositol metabolic process
protein dephosphorylation
homophilic cell adhesion via plasma membrane adhesion molecules
cell adhesion
translation
electron transport chain
nervous system development
protein folding
cell differentiation
NAD biosynthetic process
cilium movement
sterol biosynthetic process
regulation of cell cycle
metabolic process
cation transmembrane transport
lipid metabolic process
G protein−coupled receptor signaling pathway
chitin metabolic process
ion transport
cell surface receptor signaling pathway
ion transmembrane transport
nucleosome assembly
negative regulation of peptidase activity
multicellular organism development
ionotropic glutamate receptor signaling pathway
carbohydrate biosynthetic process
proteolysis
microtubule−based process
neuropeptide signaling pathway
potassium ion transmembrane transport
cyclic nucleotide biosynthetic process
GeneRatio
0.03
0.06
0.09
p.adjust
0.01
0.02
0.03
0.04
methylation pattern n°:
Pattern 1
Pattern 2
Pattern 3
Pattern 4

Figure 4

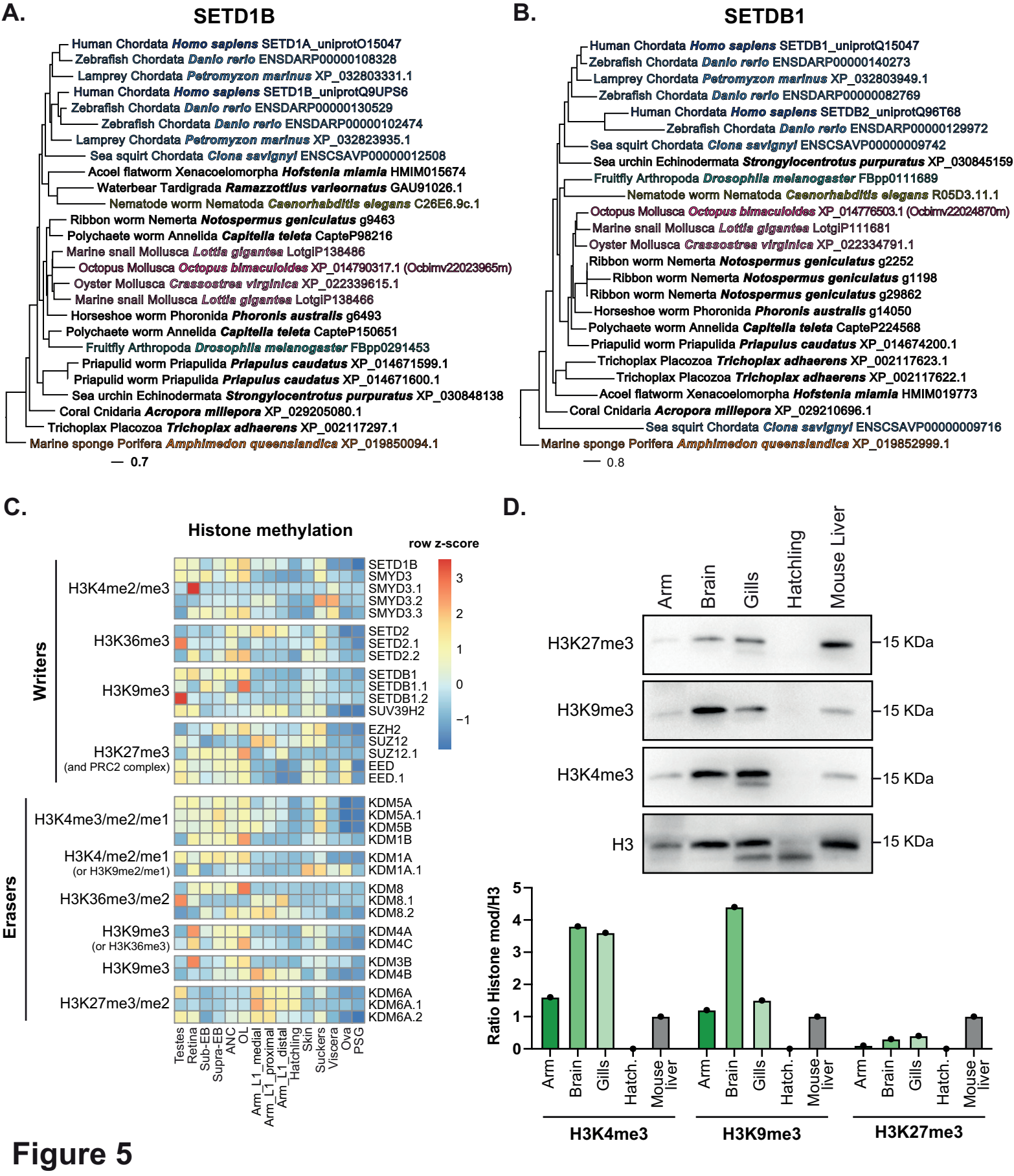
A.
SETD1B
Human Chordata Homo sapiens SETD1A_uniprotO15047
Zebrafish Chordata Danio rerio ENSDARP00000108328
Lamprey Chordata Petromyzon marinus XP_032803331.1
Human Chordata Homo sapiens SETD1B_uniprotQ9UPS6
Zebrafish Chordata Danio rerio ENSDARP00000130529
Zebrafish Chordata Danio rerio ENSDARP00000102474
Lamprey Chordata Petromyzon marinus XP_032823935.1
Sea squirt Chordata Ciona savignyi ENSCSAVP00000012508
Acoel flatworm Xenacoelomorpha Hofstenia miamia HMIM015674
Waterbear Tardigrada Ramazzottius varieornatus GAU91026.1
Nematode worm Nematoda Caenorhabditis elegans C26E6.9c.1
Ribbon worm Nemerta Notospermus geniculatus g9463
Polychaete worm Annelida Capitella teleta CapteP98216
Marine snail Mollusca Lottia gigantea LotgiP138486
Octopus Mollusca Octopus bimaculoides XP_014790317.1 (Ocbimv22023965m)
Oyster Mollusca Crassostrea virginica XP_022339615.1
Marine snail Mollusca Lottia gigantea LotgiP138466
Horseshoe worm Phoronida Phoronis australis g6493
Polychaete worm Annelida Capitella teleta CapteP150651
Fruitfly Arthropoda Drosophila melanogaster FBpp0291453
Priapulid worm Priapulida Priapulus caudatus XP_014671599.1
Priapulid worm Priapulida Priapulus caudatus XP_014671600.1
Sea urchin Echinodermata Strongylocentrotus purpuratus XP_030848138
Coral Cnidaria Acropora millepora XP_029205080.1
Trichoplax Placozoa Trichoplax adhaerens XP_002117297.1
Marine sponge Porifera Amphimedon queenslandica XP_019850094.1
0.7
B.
SETDB1
Human Chordata Homo sapiens SETDB1_uniprotQ15047
Zebrafish Chordata Danio rerio ENSDARP00000140273
Lamprey Chordata Petromyzon marinus XP_032803949.1
Zebrafish Chordata Danio rerio ENSDARP00000082769
Human Chordata Homo sapiens SETDB2_uniprotQ96T68
Zebrafish Chordata Danio rerio ENSDARP00000129972
Sea squirt Chordata Ciona savignyi ENSCSAVP00000009742
Sea urchin Echinodermata Strongylocentrotus purpuratus XP_030845159
Fruitfly Arthropoda Drosophila melanogaster FBpp0111689
Nematode worm Nematoda Caenorhabditis elegans R05D3.11.1
Octopus Mollusca Octopus bimaculoides XP_014776503.1 (Ocbimv22024870m)
Marine snail Mollusca Lottia gigantea LotgiP111681
Oyster Mollusca Crassostrea virginica XP_022334791.1
Ribbon worm Nemerta Notospermus geniculatus g2252
Ribbon worm Nemerta Notospermus geniculatus g1198
Ribbon worm Nemerta Notospermus geniculatus g29862
Horseshoe worm Phoronida Phoronis australis g14050
Polychaete worm Annelida Capitella teleta CapteP224568
Priapulid worm Priapulida Priapulus caudatus XP_014674200.1
Trichoplax Placozoa Trichoplax adhaerens XP_002117623.1
Trichoplax Placozoa Trichoplax adhaerens XP_002117622.1
Acoel flatworm Xenacoelomorpha Hofstenia miamia HMIM019773
Coral Cnidaria Acropora millepora XP_029210696.1
Sea squirt Chordata Ciona savignyi ENSCSAVP00000009716
Marine sponge Porifera Amphimedon queenslandica XP_019852999.1
0.8
C.
Histone methylation
row z-score
3
2
1
0
−1
Writers
H3K4me2/me3
SETD1B
SMYD3
SMYD3.1
SMYD3.2
SMYD3.3
H3K36me3
SETD2
SETD2.1
SETD2.2
H3K9me3
SETDB1
SETDB1.1
SETDB1.2
SUV39H2
H3K27me3
(and PRC2 complex)
EZH2
SUZ12
SUZ12.1
EED
EED.1
Erasers
H3K4me3/me2/me1
KDM5A
KDM5A.1
KDM5B
KDM1B
H3K4/me2/me1
(or H3K9me2/me1)
KDM1A
KDM1A.1
H3K36me3/me2
KDM8
KDM8.1
KDM8.2
H3K9me3
(or H3K36me3)
KDM4A
KDM4C
H3K9me3
KDM3B
KDM4B
H3K27me3/me2
KDM6A
KDM6A.1
KDM6A.2
Testes
Retina
Sub-EB
Supra-EB
ANC
OL
Arm_L1_medial
Arm_L1_proximal
Arm_L1_distal
Hatchling
Skin
Suckers
Viscera
Ova
PSG
D.
Arm
Brain
Gills
Hatchling
Mouse Liver
H3K27me3
H3K9me3
H3K4me3
H3
15 KDa
Ratio Histone mod/H3
5
4
3
2
1
0
Arm
Brain
Gills
Hatch.
Mouse liver
H3K4me3
H3K9me3
H3K27me3

Figure 5

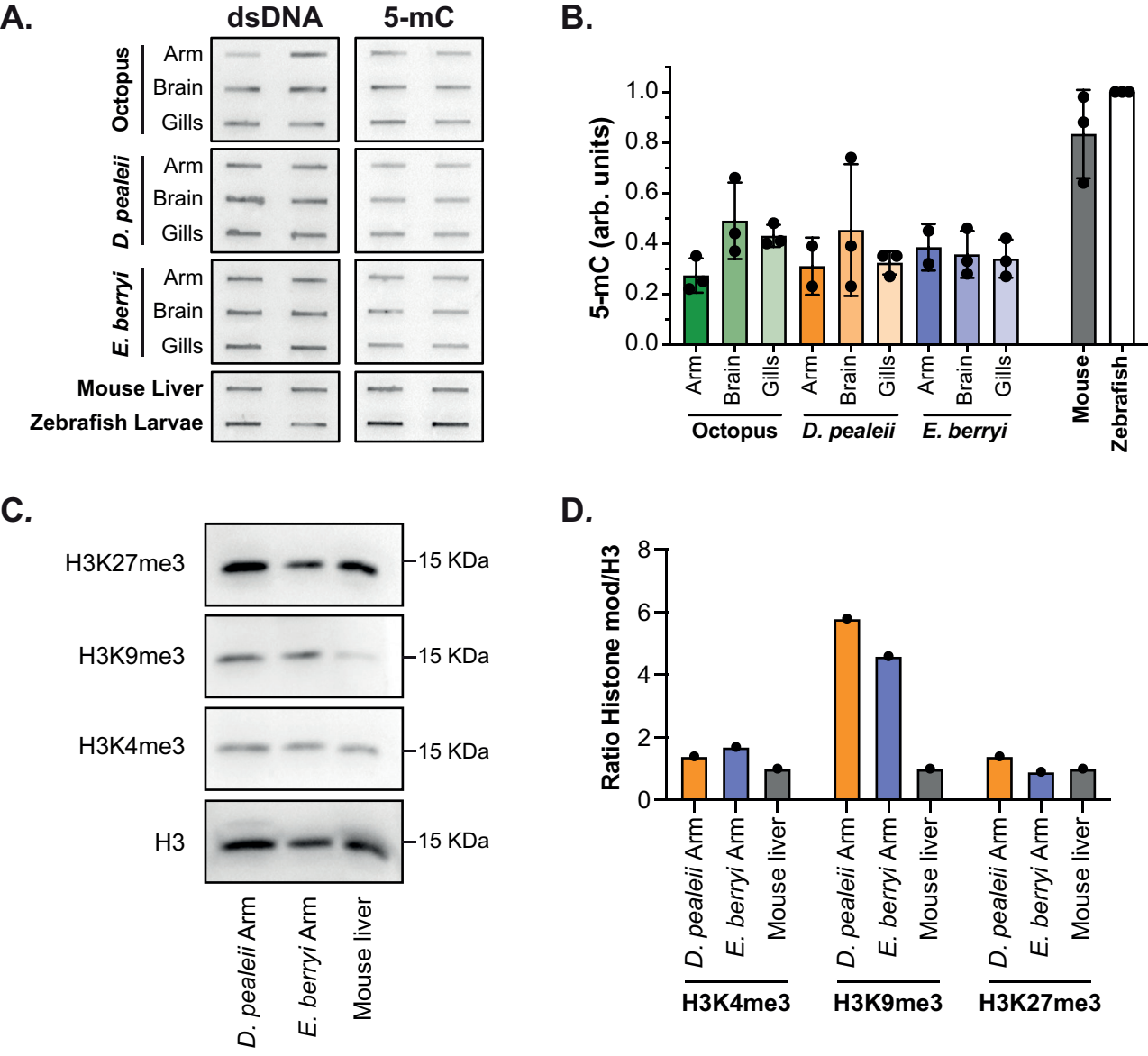
A.
dsDNA
5-mC
Octopus
Arm
Brain
Gills
D. pealeii
Arm
Brain
Gills
E. berryi
Arm
Brain
Gills
Mouse Liver
Zebrafish Larvae
B.
5-mC (arb. units)
1.0
0.8
0.6
0.4
0.2
0.0
Arm
Brain
Gills
Arm
Brain
Gills
Arm
Brain
Gills
Mouse
Zebrafish
Octopus
D. pealeii
E. berryi
C.
H3K27me3
H3K9me3
H3K4me3
H3
15 KDa
15 KDa
15 KDa
15 KDa
D. pealeii Arm
E. berryi Arm
Mouse liver
D.
Ratio Histone mod/H3
8
6
4
2
0
D. pealeii Arm
E. berryi Arm
Mouse liver
D. pealeii Arm
E. berryi Arm
Mouse liver
D. pealeii Arm
E. berryi Arm
Mouse liver
H3K4me3
H3K9me3
H3K27me3

Figure 6