



RESEARCH ARTICLE

The elephant shark methylome reveals conservation of epigenetic regulation across jawed vertebrates [version 1; referees: 2 approved]

Julian R. Peat , Oscar Ortega-Recalde , Olga Kardailsky, Timothy A. Hore 

Department of Anatomy, University of Otago, Dunedin, 9016, New Zealand

v1 First published: 20 Apr 2017, 6:526 (doi: [10.12688/f1000research.11281.1](https://doi.org/10.12688/f1000research.11281.1))
 Latest published: 20 Apr 2017, 6:526 (doi: [10.12688/f1000research.11281.1](https://doi.org/10.12688/f1000research.11281.1))

Abstract

Background: Methylation of CG dinucleotides constitutes a critical system of epigenetic memory in bony vertebrates, where it modulates gene expression and suppresses transposon activity. The genomes of studied vertebrates are pervasively hypermethylated, with the exception of regulatory elements such as transcription start sites (TSSs), where the presence of methylation is associated with gene silencing. This system is not found in the sparsely methylated genomes of invertebrates, and establishing how it arose during early vertebrate evolution is impeded by a paucity of epigenetic data from basal vertebrates.



Methods: We perform whole-genome bisulfite sequencing to generate the first genome-wide methylation profiles of a cartilaginous fish, the elephant shark *Callorhynchus milii*. Employing these to determine the elephant shark methylome structure and its relationship with expression, we compare this with higher vertebrates and an invertebrate chordate using published methylation and transcriptome data.


Results: Like higher vertebrates, the majority of elephant shark CG sites are highly methylated, and methylation is abundant across the genome rather than patterned in the mosaic configuration of invertebrates. This global hypermethylation includes transposable elements and the bodies of genes at all expression levels. Significantly, we document an inverse relationship between TSS methylation and expression in the elephant shark, supporting the presence of the repressive regulatory architecture shared by higher vertebrates.

Conclusions: Our demonstration that methylation patterns in a cartilaginous fish are characteristic of higher vertebrates imply the conservation of this epigenetic modification system across jawed vertebrates separated by 465 million years of evolution. In addition, these findings position the elephant shark as a valuable model to explore the evolutionary history and function of vertebrate methylation.

Open Peer Review

Referee Status:  

	Invited Referees	
	1	2
version 1 published 20 Apr 2017	 report	 report

- Matthew M. Hindle**, University of Edinburgh UK
- Arthur Georges** , University of Canberra Australia, **Changwei Shao**, Chinese Academy of Fishery Sciences China

Discuss this article

Comments (0)

Corresponding author: Timothy A. Hore (tim.hore@otago.ac.nz)

How to cite this article: Peat JR, Ortega-Recalde O, Kardailsky O and Hore TA. **The elephant shark methylome reveals conservation of epigenetic regulation across jawed vertebrates [version 1; referees: 2 approved]** *F1000Research* 2017, 6:526 (doi: [10.12688/f1000research.11281.1](https://doi.org/10.12688/f1000research.11281.1))

Copyright: © 2017 Peat JR *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This research was funded by a University of Otago Research Committee Grant (111899.01.R.LA)
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 20 Apr 2017, 6:526 (doi: [10.12688/f1000research.11281.1](https://doi.org/10.12688/f1000research.11281.1))

Introduction

The methylation of DNA at cytosine bases constitutes an epigenetic regulatory system that is essential for the development of bony vertebrates¹⁻³. Of particular significance is the modification of CG dinucleotides, whose symmetry allows methylation signals in this context to be perpetuated by maintenance methyltransferases following DNA replication⁴. CG methylation and the epigenetic memory encoded by it thus form a stable but flexible storage system for molecular information.

The methylomes of studied vertebrates – including bony fish, amphibians and mammals – exhibit similar global patterns in which the majority of CG sites are methylated in somatic tissues⁵⁻⁹. Regulatory elements such as promoters and enhancers are an important exception to this pervasive methylation landscape, particularly when associated with short CG-rich regions termed CpG islands. At the transcription start site (TSS), the presence of methylation is associated with transcriptional silencing, an effect achieved through the inhibition of transcription factor binding and the action of proteins that recognise methylated DNA and induce an inaccessible chromatin configuration^{10,11}. The inverse relationship of TSS methylation with gene expression has been documented across a wide range of vertebrate taxa^{5,8,12-16}, indicating an evolutionarily important function. The molecular machinery that invokes an inactive state in response to methylation signals also appears to be conserved¹⁰. Differences in methylation at regulatory regions are linked to the definition of cell fate during developmental progression and the stable maintenance of this identity in differentiated tissues^{5,17-19}. Indeed, widespread erasure of methylation marks in the cells of humans and mice plays a prominent role in the reprogramming of fate specification in both natural and experimental systems^{17,20,21}.

High levels of methylation outside the TSS of genes also serve an important function in vertebrate genomes. A substantial fraction of vertebrate genomes is composed of repetitive transposable elements (TEs), whose activity must be repressed to safeguard genome integrity^{22,23}. These elements are ubiquitously methylated in vertebrate somatic tissues^{8,9,16,24}, and experiments performed in mammalian model systems has shown this to be critical for their transcriptional repression²⁵. Hypermethylation of gene bodies is also a conserved feature of vertebrate genomes, and – unlike methylation at the TSS – this is compatible with active transcription in all species profiled to date^{5,8,12,14-16,26-28}. Although the relationship of intragenic methylation with gene expression levels is complex and appears to vary across taxa and even cell-type^{5,8,12-16,26,28,29}, it has been shown to suppress spurious transcription³⁰ and regulate exon splicing^{31,32} in mammalian systems.

The distribution and regulatory functions of methylation in vertebrates are unique amongst the metazoa, but the evolution of this system is poorly understood. In striking contrast to the pervasive hypermethylation that characterises vertebrates, invertebrate genomes are sparsely methylated and certain species such as the nematode *Caenorhabditis elegans* and fruit fly *Drosophila melanogaster* are apparently devoid of cytosine methylation^{14,33-36}. Where present, the predominant pattern is a mosaic configuration, in which unmethylated regions are interspersed with hypermethylated sequences, the latter preferentially located in gene bodies

and in loose positive association with transcription^{14,33-35,37}. Significantly, invertebrates lack the inverse relationship between TSS methylation and expression that constitutes a key regulatory mechanism in vertebrates, and the low levels of methylation do not appear to act as a control against TE activity in their genomes^{14,35,38-41}.

Methylation in *Ciona intestinalis*, a sea squirt belonging to the subphylum tunicata, the chordate lineage most closely related to vertebrates⁴² (Figure 1), typifies the invertebrate mosaic pattern^{14,33,35}. The methylation system present in higher vertebrates can thus be inferred to have evolved at some point after the divergence of tunicates from vertebrate progenitors (~680 Mya⁴³) and before the radiation of bony fish and tetrapods (~430 Mya⁴³; Figure 1). Understanding the timing of this progression at greater resolution and the factors that stimulated its development is hindered by the absence of methylation data from basal vertebrate classes.

Here, we use whole-genome bisulfite sequencing to generate the first methylation profiles of a cartilaginous fish, the elephant shark *Callorhynchus milii*. Through detailed comparison with published methylation and expression datasets, we demonstrate that the elephant shark methylome is characteristic of vertebrates in its global hypermethylation – including at TEs and gene bodies – and, crucially, association with transcriptional silencing at the TSS. These findings indicate conservation of a complex methylation system across jawed vertebrates separated by 465 million years of evolution, and identify the elephant shark as an important model to examine the origins and function of methylation in vertebrates.

Methods

Tissue source and DNA extraction

Elephant shark tissue samples were sourced as by-product of deceased animals harvested from commercial fishing in the Otago coastal region. As such, no animal ethics permission was applicable in this circumstance. No animal experimentation or manipulation was undertaken as defined by the Animal Welfare Act (2009, New Zealand), or according to guidelines issued by the New Zealand National Animal Ethics Advisory Committee (NAEAC, Occasional Paper No 2, 2009, ISBN 978-0-478-33858-4).

DNA was purified using a modified magnetic bead approach⁴⁴. Briefly, cells were first homogenised in “GITC” lysis buffer (4 M Guanidine thiocyanate, Sigma G6639; 50 mM Tris, Thermo 15568-025; 20 mM EDTA; Thermo 15575-020; 2% Sarkosyl, Sigma L9150-50G; 0.1% Antifoam, Sigma A8311-50ML), and this lysate mixture was then combined with TE-diluted Sera-Mag Magnetic SpeedBeads (GE Healthcare, GEHE45152105050250) and isopropanol in a volumetric ratio of 2:3:4, respectively. Following capture with a neodymium magnet, beads were washed once with isopropanol, twice with 70% ethanol and resuspended in filter-sterile milliQ water.

Preparation of WGBS-seq libraries

WGBS-seq was undertaken using a post-bisulfite adapter tagging (PBAT) method adapted from Peat *et al.*, 2014⁴⁵. Briefly,

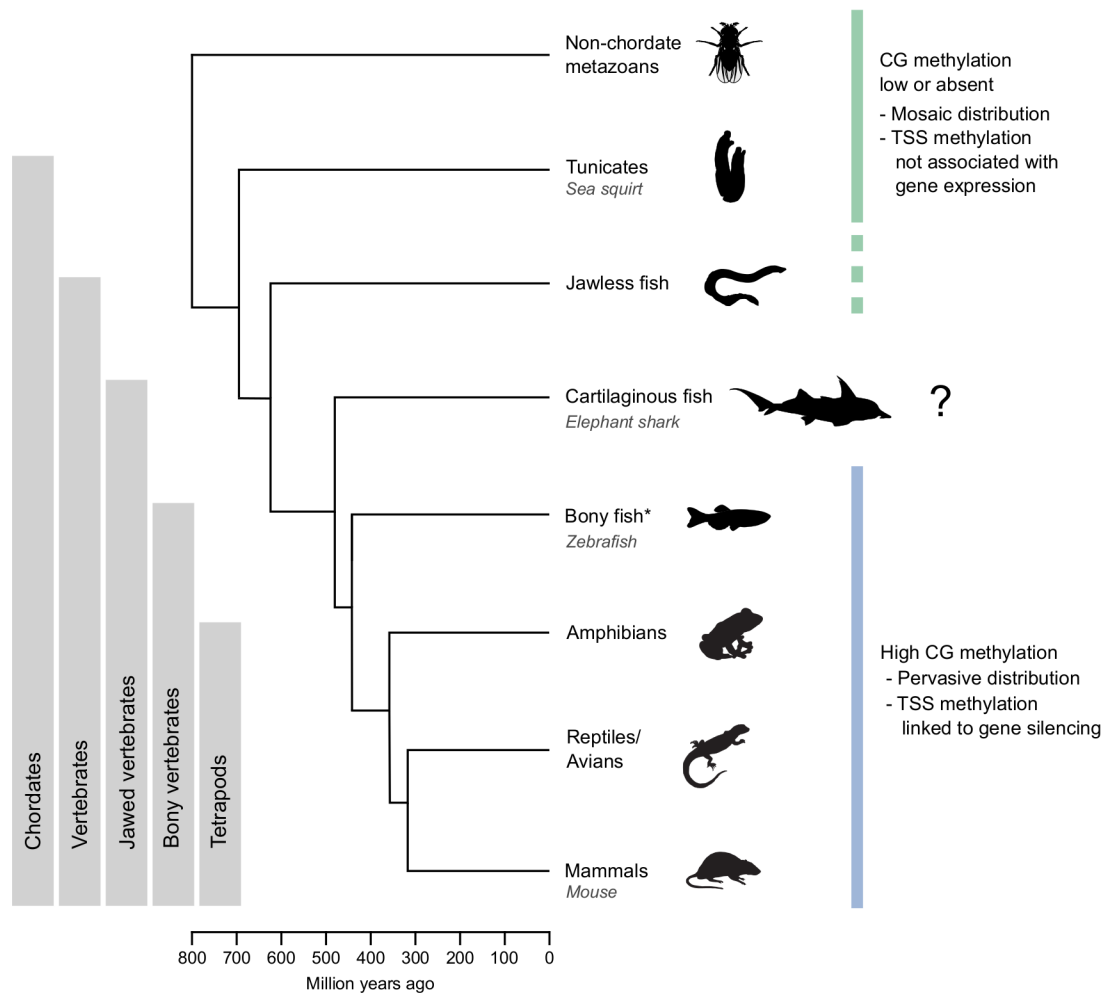


Figure 1. Phylogenetic tree showing major vertebrate groups, invertebrate outgroups and defining characteristics of their methylomes. The genomes of higher vertebrates are pervasively hypermethylated, with the exception of regulatory elements such as transcription start sites (TSSs), where the presence of methylation is associated with gene silencing (blue line). In contrast, invertebrate genomes are generally sparsely methylated in a mosaic pattern, and lack the inverse relationship between TSS methylation and expression that characterises vertebrates (green line). Certain invertebrate species appear to lack methylation altogether. Due to a paucity of data from basal vertebrate species, the evolutionary history of the CG methylation system present in higher vertebrates is unclear. Preprint methylation data from the sea lamprey *Petromyzon marinus* is not indicated here (see discussion). The names of organisms examined in this study are noted underneath the appropriate class. * The lobe-finned fish (sarcopterygii), as well as the cephalochordata (a basal chordate taxon), have been omitted for clarity. The following terms have been treated as equivalent: jawless fish and cyclostomata, jawed vertebrates and gnathostomata, cartilaginous fish and chondrichthyes, bony vertebrates and euteleostomi. Median divergence times from the TimeTree database⁴³ were used to construct the tree.

50–100 ng of purified DNA was subjected to bisulfite conversion using the Imprint DNA modification kit (Sigma, MOD50). Converted DNA underwent first strand synthesis with a biotin-labelled adapter sequence possessing seven random nucleotides at its 3' end (BioP5N7, biotin- ACACTCTTCCCTACAC-GACGCTCTCCGATCTNNNNNNN). The product of first strand synthesis was captured using streptavidin-coated Dynabeads (Thermo, 11205D) and magnetic immobilisation. Double-stranded DNA was created using the immobilized first-strand as a template and an additional adapter that also possesses seven random nucleotides at its 3' end (P7N7, GTGACTGGAGTTCAGACGT-GTGCTCTCCGATCTNNNNNNN). Unique molecular barcodes

and sequences necessary for binding to Illumina flow-cells were added to libraries by PCR using 1X HiFi HotStart Uracil+ Mix (KAPA, KK2801 and 10 μ M indexed Truseq-type oligos), with thermal cycling as follows: 12 \times (94°C, 80 sec; 65°C, 30 sec; 72°C, 30 sec).

For deep sequencing, libraries were sequenced with a single-end 100bp protocol on a HiSeq 2500 instrument (Illumina) using rapid run mode. For low-coverage sequencing of additional samples, libraries were sequenced on a MiSeq instrument (Illumina) until the desired depth (at least 15,000 mapped CG calls) was attained.

Detailed sequencing results are provided in [Table S1](#).

Bioinformatic processing of WGBS-seq dataset

Mapped CG methylated calls for mouse liver⁵ were downloaded from [GEO](#) (accession GSE42836, sample GSM1051157) and analysed directly. For zebrafish muscle⁸ (SRA study SRP020008, run SRR800081) and sea squirt muscle¹⁴ (GEO accession GSE19824, sample GSM497251), raw sequencing data was downloaded and processed along with elephant shark WGBS-seq data generated in this study as follows.

Trimming was performed to remove both poor-quality calls and adapters sequences using [TrimGalore](#) (v0.4.0, default parameters). For the elephant shark data, 10bp were also removed from the 5' end of reads to account for sequence biases associated with PBAT library construction.

Trimmed reads were aligned using [Bismark](#)⁴⁶ (v0.14.3, default parameters) with the `--pbat` option specified for elephant shark data. The following genome assemblies were used for alignment: zebrafish, GRCz10; elephant shark, 6.1.3; sea squirt, KH. For sea squirt and elephant shark, alignment was only performed against scaffolds larger than 277kb to avoid gene annotation issues and assembly artefacts. The deep-sequenced elephant shark data generated in this study was additionally mapped to the mitochondrial genome.

Bismark mapping reports were used to determine global methylation levels for low-coverage elephant shark data. All other datasets were deduplicated and CG methylation calls extracted using [Bismark](#) (`--comprehensive` and `--merge_non_CG` options specified).

The number of mapped cytosine calls for sequencing performed in this study are provided in [Table S1](#). The frequency of non-CG methylation indicates the maximum rate of non-conversion during the bisulfite treatment step; by this measure, all libraries had a bisulfite conversion efficiency of at least 98.9%.

Bootstrap sampling to determine margin of error in low-coverage WGBS-seq

In order to determine the number of CG methylation calls required to accurately predict genome-wide methylation levels, bootstrap sampling of reads from the deep-sequenced male elephant shark dataset was performed to generate regular intervals of CG calls from approximately 100 to 30,000. These reads were trimmed, mapped and methylation quantified as described above, and following 1000 iterations, the proportion of data falling within the 0.5-99.5 percentiles was calculated to generate a 99% confidence interval. An asymptotic model described by the equation $y = 2.208/\sqrt{x}$ was used to fit a curve to the data. At our minimum sequencing depth of 15,000 CG calls, bootstrap sampling predicts a margin of error (99% confidence interval) of approximately ± 1.8 methylation percentage points.

Analysis of deep-sequenced WGBS-seq datasets

CG methylation calls were imported into the [SeqMonk](#) program (v1.37.1) for analysis. For elephant shark and sea squirt, custom

[SeqMonk](#) genomes were built using GFF annotation files downloaded from [NCBI](#) and [Ensembl](#), respectively.

To analyse methylation at the level of individual CG dinucleotides, we generated an annotation track of each CG site using [Bowtie](#) v1.1.2⁴⁷. A minimum of five methylation calls was required for inclusion of a CG site in analyses.

For mouse, zebrafish and elephant shark, precompiled annotation tracks of repetitive elements generated using the [RepeatMasker](#) program were downloaded from UCSC. For sea squirt, we generated these annotations by running the [RepeatMasker](#) program (v4.0.6) on the KH assembly with the `-s` option and specifying *Ciona intestinalis* as the species. The various classes of transposable elements were extracted from these annotation files and where indicated, merged for analysis. A minimum of five calls was applied as a threshold for inclusion when quantifying individual elements.

To examine methylation profiles across genes or TEs and neighbouring sequences, methylation was quantified at individual CGs and the mean plotted across a size-standardised gene or TE as well as 10kb upstream and downstream regions, using the quantitation trend plot function. Figures were produced using [Prism](#) (GraphPad, v7), with smoothing applied to flanking regions by averaging 100 neighbours.

Transcription start sites were defined as 200bp centred on the first nucleotide of an annotated mRNA, and a minimum of five methylation calls was applied as a threshold for inclusion in analyses. For analysis of gene bodies, 2kb running windows were quantified (with a minimum of 50 methylation calls applied for inclusion) within annotated mRNAs, excluding 1kb at the 5' end, and the mean was reported for each mRNA.

Violin plots and histograms were drawn using the [ggplot2](#) package⁴⁸ in [R](#).

Bioinformatic processing of RNA-seq datasets

We downloaded raw sequencing data from previous studies as follows; sea squirt muscle¹⁴, GEO accession GSE19824, sample GSM497252; elephant shark liver⁴⁹, SRA study SRP013772, run SRR513760; zebrafish muscle⁸, SRA study SRP020008, run SRR800045; mouse liver (ENCODE Consortium^{50,51}), GEO accession GSE78583, sample GSM2072415.

Trimming was performed to remove both poor-quality calls and adapters sequences using [TrimGalore](#) (v0.4.0, default parameters). In addition, 12bp were removed from the 5' end of sea squirt reads and 10bp from the 5' end of both elephant shark and mouse reads to avoid sequence biases.

Trimmed reads were aligned to the reference genomes described above with [HISAT2](#)⁵² (v2.0.5) using single-end or paired-end mode, as appropriate. Known splice sites were specified from a file built from GTF annotation files downloaded from [Ensembl](#) (release 87) using the [HISAT2](#) python script. No GTF file was

available for elephant shark, so a GFF annotation file downloaded from NCBI was first converted to GTF format using the gffread program (<https://github.com/gperrea/gffread>).

Analysis of RNA-seq datasets

Alignments from HISAT2 were imported into the SeqMonk program, specifying a minimum mapping quality of 60 to select only uniquely aligned reads.

The RNA-seq quantitation pipeline was used to generate raw read counts across the exons of nuclear protein-coding genes with a correction for any DNA contamination. Counts were corrected by transcript length and genes were divided into quintiles according to expression level.

Results

Genome-wide methylation profiles of the elephant shark, *Callorhynchus milii*

To generate genome-wide methylation profiles, we extracted DNA from the liver tissue of one female and one male adult elephant shark and performed whole-genome bisulfite sequencing (WGBS-seq). Detailed sequencing results are provided in Table S1.

As described in the somatic tissues of other vertebrates, we found that methylation is much more prevalent in nuclear DNA at CG dinucleotides (69 – 71.6%) than in non-CG context (0.8 – 1%) or mitochondrial DNA (1.6 – 2.5%; Figure 2A). Low-coverage WGBS-seq demonstrated similar global methylation levels in three additional individuals for liver, and in spleen and pancreas samples (Figure 2B). While we observed a small trend for lower methylation in female samples (Figure 2B;

female mean 66.4%, male mean 68.6%), this was not significant according to a *t*-test ($p=0.2308$) and within the margin of error expected at this sequencing depth (Figure S1).

We proceeded with further analysis of CG methylation in deep-sequenced liver datasets as an example of the elephant shark somatic methylome, and combined male and female samples to enhance sequencing coverage.

The elephant shark genome is pervasively methylated

Existing data indicate that methylation patterns differ markedly between vertebrates and invertebrates. In order to delineate the characteristics of these disparate systems and establish their relationship to the elephant shark methylome, we reanalysed published WGBS-seq data from two vertebrates, mouse (*Mus musculus*)⁵ and zebrafish (*Danio rerio*)⁸, as well as an invertebrate from the closest chordate outgroup, the sea squirt *Ciona intestinalis*¹⁴ (Table 1A).

As expected from analysis of global levels, examination of methylation at individual CG dinucleotides in the elephant shark showed that the majority of sites are highly methylated ($\geq 80\%$), and fewer than one tenth are unmethylated (Figure 3A). Both this pattern and the global methylation level are comparable to mouse and zebrafish (Figure 3A–B). In contrast, mean methylation in the invertebrate sea squirt is only 22.9%, and over two thirds of CG sites are unmethylated.

A further striking distinction is evident when the genome is profiled in 2kb running windows. High methylation levels are pervasive in the elephant shark genome (Figure 3C–D), resembling

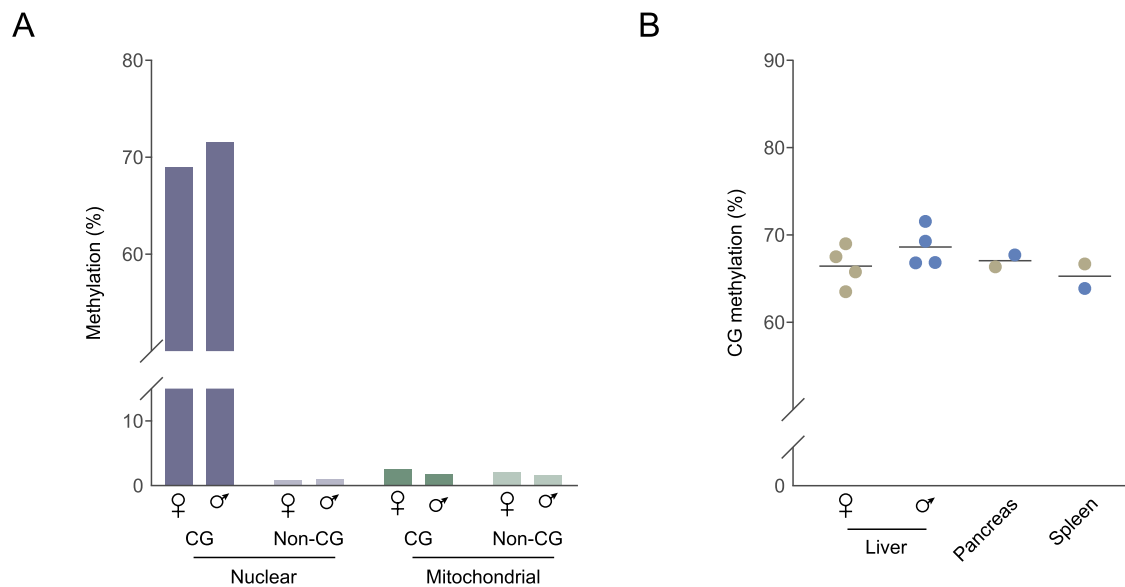


Figure 2. Global methylation levels in elephant shark somatic tissues. A: Global methylation levels of deep-sequenced liver samples in different contexts. 'CG' refers to symmetrical CG dinucleotides; 'Non-CG' indicates all other sequence contexts. **B:** Global CG methylation levels in elephant shark tissues examined by low-coverage sequencing. The horizontal bar indicates the mean; gold dots, female samples; blue dots, male samples. The difference between female and male liver samples is not significant according to a *t*-test, and within the technical margin of error expected at the threshold sequencing depth used (± 1.8 methylation percentage points; Figure S1).

Table 1. Published WGBS-seq and RNA-seq datasets used for comparative analysis in this study. Accession numbers are provided in the methods.

Species	A. WGBS-seq Datasets		B. RNA-seq Datasets	
	Reference	Tissue	Reference	Tissue
<i>Callorhynchus milii</i> Elephant shark	This study	Liver	Venkatesh <i>et al.</i> , 2014 ⁴⁹	Liver
<i>Ciona intestinalis</i> Sea squirt	Zemach <i>et al.</i> , 2010 ¹⁴	Muscle	Zemach <i>et al.</i> , 2010 ¹⁴	Muscle
<i>Danio rerio</i> Zebrafish	Potok <i>et al.</i> , 2013 ⁸	Muscle	Potok <i>et al.</i> , 2013 ⁸	Muscle
<i>Mus musculus</i> Mouse	Hon <i>et al.</i> , 2013 ⁵	Liver	ENCODE Consortium ^{50,51}	Liver

the structure of other vertebrate methylomes. In contrast, the sea squirt methylome is characterised by a bimodal but largely unmethylated distribution (Figure 3C), resulting from a mosaic pattern in which background hypomethylation is punctuated by shorter stretches of methylated sequences (Figure 3D). Interestingly, running windows show a broader distribution of methylation in elephant shark than in mouse or zebrafish (Figure 3C). Whether this is a feature of basal vertebrates generally or of elephant shark specifically will require analysis of methylation patterns in additional cartilaginous fish.

Transposable elements are hypermethylated in the elephant shark

Having established that the global structure of the elephant shark methylome is characteristic of vertebrates, we sought to determine the profile and impact of methylation at specific functional elements.

Transposable elements (TEs) are highly methylated in vertebrate genomes, a feature which is linked to the necessity of repressing their transcription to prevent destabilising transposase activity^{8,9,16,22–24}. The generally low levels of methylation at TEs in invertebrates such as the sea squirt do not appear to regulate their activity^{14,38,39}.

Examination of methylation patterns at TEs and flanking sequences showed that the elephant shark exhibits hypermethylation at the large majority of TEs and a slight increase in mean methylation relative to adjacent regions (Figure 4A–B), conforming to the pattern of other vertebrates. While mean methylation levels of TEs in sea squirt are moderately elevated compared to flanking sequences, the large majority of TEs are hypomethylated. Little variation in methylation was observed between the two predominant TE classes in the elephant shark genome, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs; Figure 4C–D), indicating that – as in other vertebrates^{8,9,16,24} – hypermethylation of TEs is ubiquitous.

Methylation at elephant shark transcription start sites is associated with gene silencing

Silencing of gene expression through the deposition of methylation at transcription start sites (TSSs) constitutes an important regulatory mechanism in vertebrates, but appears to be absent from invertebrates^{5–10,12–16,35,40,41}. To compare the relationship of methylation and transcription in elephant shark with higher vertebrates and the sea squirt, we made use of tissue-matched published RNA-seq datasets^{8,14,49,50} (Table 1B) to classify protein-coding genes into expression quintiles.

Hypomethylation at the TSS of expressed genes constitutes a conspicuous exception to the otherwise pervasively methylated elephant shark genome, matching the higher vertebrates examined (Figure 5A–C). Significantly, we document an inverse relationship between TSS methylation and expression level in the elephant shark (Figure 5A, Figure 5E). A bimodal distribution in which a large proportion of sequences are methylated at low expression levels contrasts with negligible methylation at most TSSs of intermediate and highly expressed genes. The association of TSS methylation with transcriptional silencing is a distinguishing feature of higher vertebrate methylomes^{5,8,12–16} that is recapitulated here for zebrafish and mouse (Figure 5B–C, Figure 5E), and its presence in the elephant shark indicates that methylation at the TSS induces repression in a similar manner. Consistent with reports showing that invertebrates lack this wide variation in TSS methylation as a function of expression level^{14,35,41}, the large majority of sea squirt TSSs are hypomethylated at all expression levels and methylation levels at the TSS are comparable to intergenic sequences (Figure 5D–E).

Interestingly, a larger number of TSSs at highly expressed genes remain methylated in elephant shark compared to mouse and zebrafish. This may suggest that the association of methylation with repression is less absolute than in higher vertebrates, but could also be attributed to poorer TSS annotation in the less intensively-studied and incompletely assembled elephant shark genome.

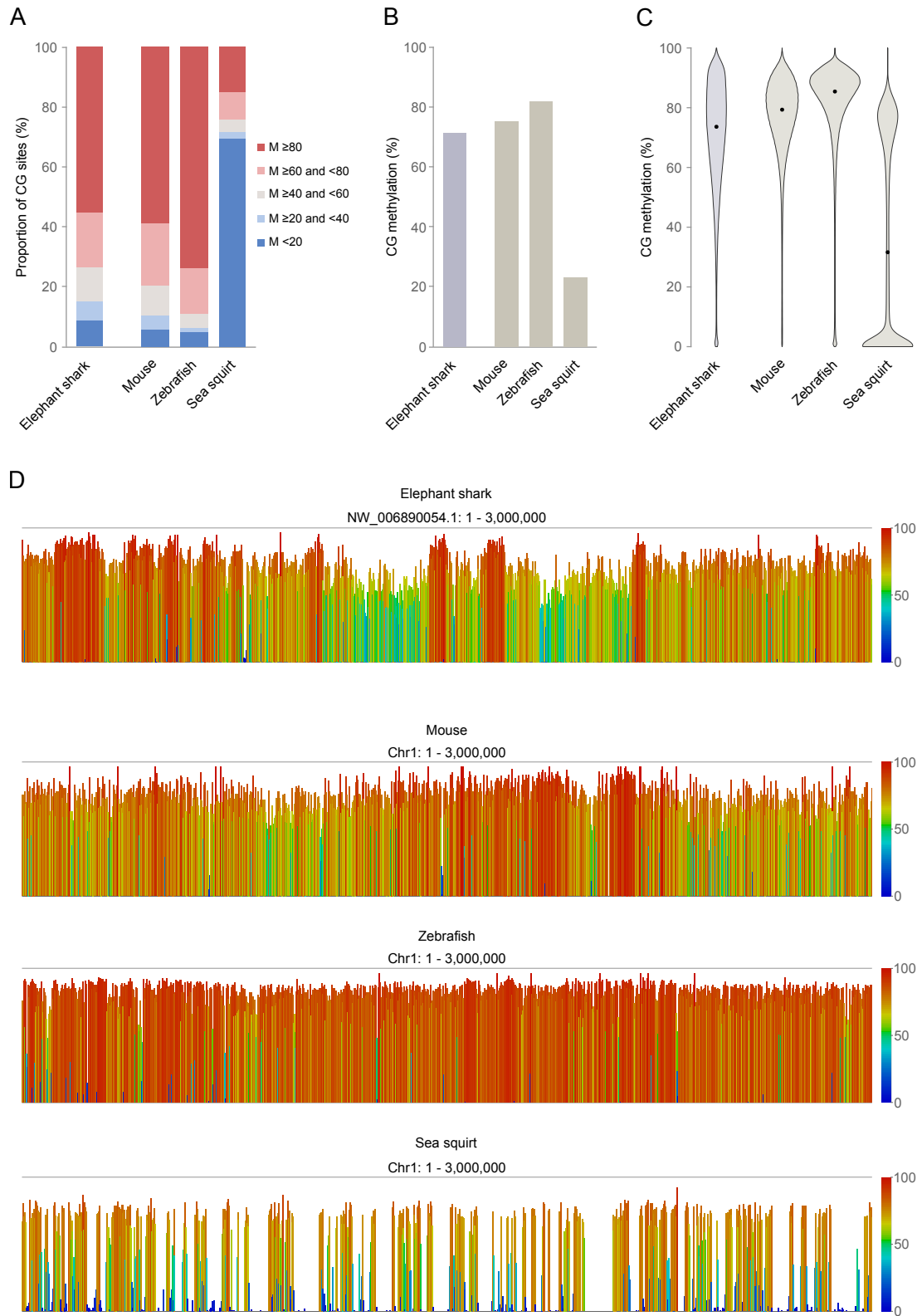


Figure 3. Global structure of the elephant shark methylome. A: Distribution of methylation at individual CG dinucleotides. 'M' denotes percentage CG methylation. **B:** Mean methylation of CG dinucleotides. **C:** Distribution of methylation within 2kb running windows covering the entire genome. Black dots denote the median. **D:** Genome screenshots of methylation quantified in 2kb running windows over the first 3Mb of chromosome 1 in sea squirt, zebrafish and mouse, and of the largest scaffold (NW_006890054.1) in elephant shark. These regions were arbitrarily chosen as an unbiased section of each genome.

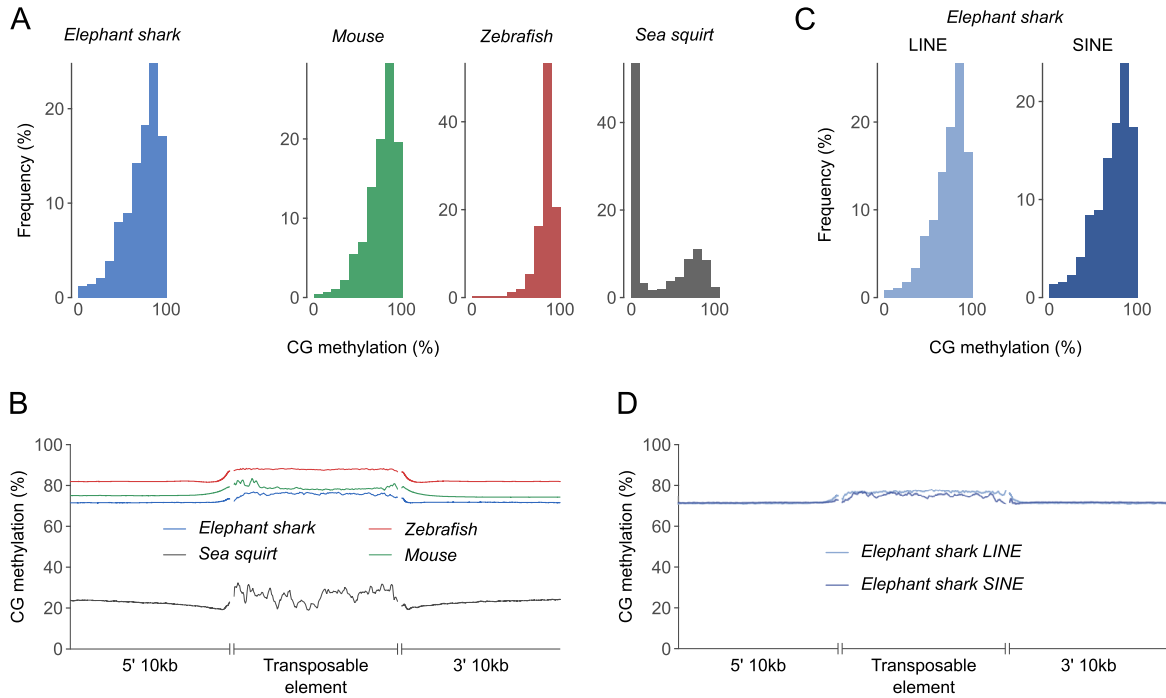


Figure 4. Methylation patterns at transposable elements. A: Distribution of methylation at transposable elements. Mean methylation values are divided into 10 bins. **B:** Mean CG methylation across transposable elements and 10kb flanking regions. Quantification was performed at the level of individual CG dinucleotides. Flanking regions were smoothed by averaging 100 neighbours. **C:** Distribution of methylation at long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) in the elephant shark genome. Mean methylation values are divided into 10 bins. **D:** Methylation at long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) in the elephant shark genome, plotted as in (B).

The methylomes of higher vertebrates and invertebrates also differ within gene bodies. While intragenic methylation in sea squirt forms the bimodal distribution reported in invertebrates^{35,37}, and most silenced genes lack methylation, vertebrate gene bodies are generally hypermethylated at all expression levels (Figure 5F). Intragenic methylation in the elephant shark is characteristic of this vertebrate pattern. In addition, higher expression levels are associated with moderately elevated gene body methylation in elephant shark liver, but not in zebrafish muscle or mouse liver. Given the limited understanding of the role played by intragenic methylation in the regulation of vertebrate gene expression, the functional relevance of this relationship is unclear.

Discussion

Methylation of CG dinucleotides forms a heritable but flexible epigenetic memory that constitutes a critical regulatory system in bony vertebrates, where it is employed in the modulation of gene expression and suppression of transposon element activity. The genomes of studied vertebrates are pervasively hypermethylated, with the exception of regulatory elements such as transcriptional start sites (TSSs), where the presence of methylation is linked to transcriptional silencing^{1–10,12–16,22–25}. These features are not found in the sparsely methylated genomes of invertebrates, including chordates closely related to vertebrates^{14,33–40}, but establishing when this important regulatory system arose and the factors that drove its

development has been impeded by a lack of methylation data from basal vertebrates (Figure 1).

In this study, we employ WGBS-seq to generate the first genome-wide methylation profiles of a cartilaginous fish, the elephant shark *Callorhynchus milii*. Through detailed comparison with published methylation and expression datasets, we demonstrate that the elephant shark methylome is characteristic of higher vertebrates and in clear contrast to the prevailing invertebrate configuration.

We first note that methylation in the elephant shark is primarily located in symmetric CG context, where comparable global methylation levels of approximately 65–70% were found by low-coverage WGBS-seq in the male and female liver, as well as in the spleen and pancreas (Figure 2). The similarity of male and female methylation indicates that, unlike certain bony fish species⁵³, the uncharacterised sex-determination mechanism in the elephant shark is not associated with large differences in global methylation. Examination of liver profiles at higher resolution demonstrated that – like higher vertebrates – the majority of elephant shark CG sites are methylated, and this is ubiquitous throughout the genome rather than concentrated in short stretches in the invertebrate mosaic pattern, typified by the sea squirt (Figure 3). The global hypermethylation of the elephant shark genome includes both major transposon classes, LINEs and SINEs (Figure 4), whose

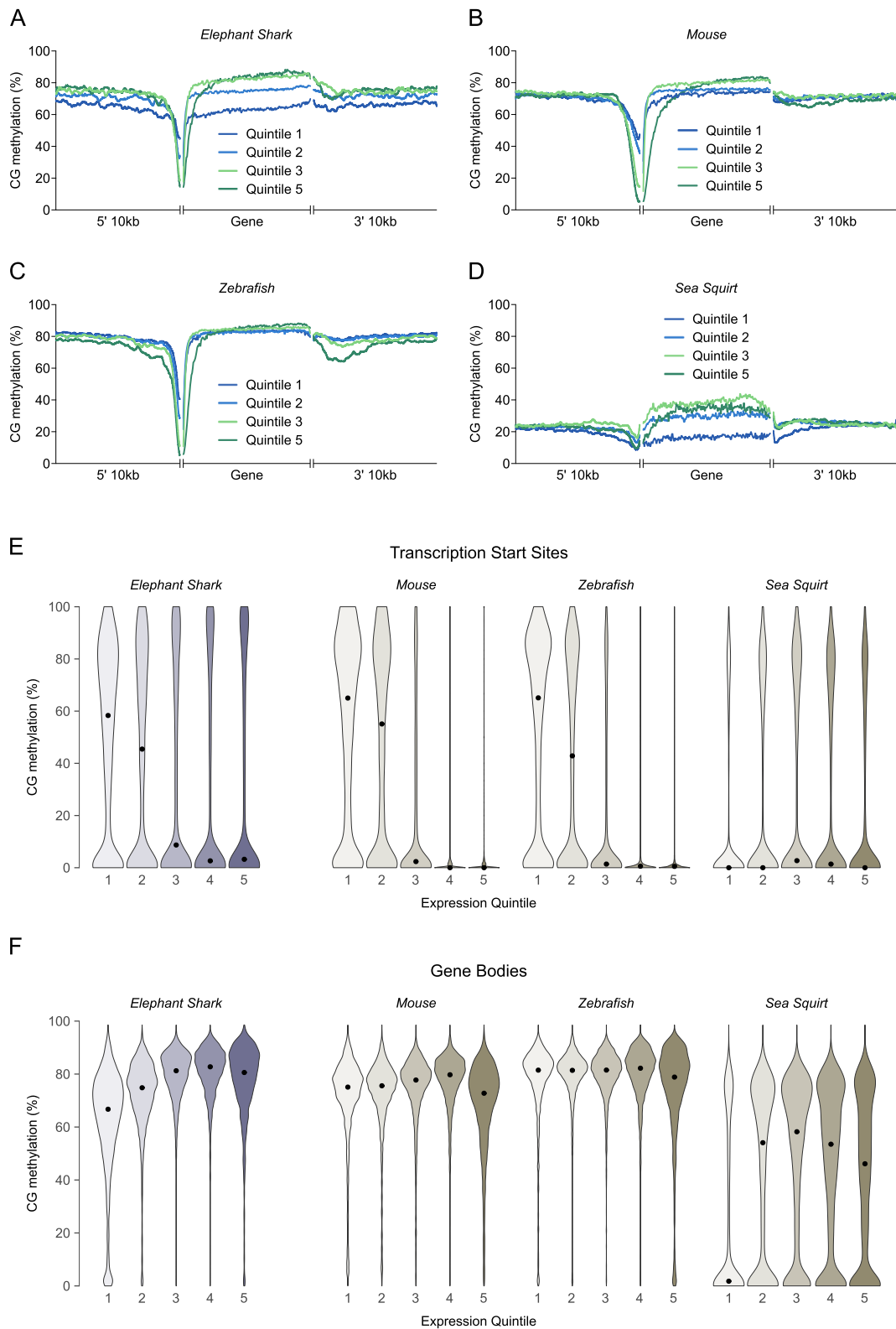


Figure 5. Relationship between methylation and gene expression. **A – D:** Mean CG methylation across genes and 10kb flanking regions, classified into quintiles according to expression level in RNA-seq datasets (5 = highest). Quintile 4 is omitted for clarity. Quantification was performed at the level of individual CG dinucleotides. Flanking regions were smoothed by averaging 100 neighbours. **E – F:** Distribution of methylation at the transcription start site (**E**) and within the body (**F**) of genes classified into quintiles according expression level (5 = highest). Each violin is scaled to the same maximum width (total area is not constant between violins) to demonstrate distributions for each quintile. Black dots denote the median.

transcriptional repression is thought to be an important function of vertebrate methylation systems as a safeguard against destabilising transposition activity.

Crucially, the elephant shark mirrors higher vertebrates in their inverse relationship of methylation with expression at the TSS (Figure 5); most expressed genes are unmethylated while a large proportion of inactive genes are hypermethylated at the TSS. This indicates that TSS methylation represses gene expression in a similar fashion in the elephant shark, and implies that this key regulatory mechanism – which is absent from invertebrates – is present in cartilaginous fish. While the association of TSS methylation with silencing is conserved across the vertebrates examined, we also observe that a greater number of expressed genes are methylated at the TSS in elephant shark than in mouse or zebrafish. It will be important to clarify whether this arises from the poorer annotation of the less intensively studied elephant shark genome, or a meaningful biological difference in the repressive potency of methylation in this system.

The hypermethylation of most gene bodies at all levels of transcription is a feature of higher vertebrate methylomes, that our data show is also shared by the elephant shark (Figure 5). We additionally document an interesting association between higher expression levels and elevated methylation in the elephant shark, a trend which is absent from the higher vertebrate tissues we examined. The relationship between intragenic methylation and expression is complex and appears to vary between vertebrate taxa and even within the tissues of a single species^{5,8,12–16,26,28,29}. Indeed, although a variety of functions for intragenic methylation have been suggested, including suppression of spurious transcription and regulation of exon splicing^{30–32}, their generality is poorly understood, particularly outside mammalian systems. Significant further research will be required to uncover the impact of intragenic methylation in vertebrate genomes and determine the biological relevance of its positive relationship with expression in the elephant shark.

Evolutionary history of the vertebrate methylation system

The observation that methylation patterns in a cartilaginous fish are characteristic of higher vertebrates implies the conservation of a complex methylation system across jawed vertebrates separated by 465 million years of evolution (Figure 1). Of particular note, they support the common presence of a regulatory architecture that links methylation at the TSS to transcriptional repression.

Preprint methylome data from the sea lamprey *Petromyzon marinus*, a basal jawless vertebrate, indicate that this species lacks the genome-wide hypermethylation and functional relationships of higher vertebrates (<https://doi.org/10.1101/033233>). While the data from this study has not yet been released, the authors state that methylation patterns in sea lamprey more closely resemble those of the sea squirt and appear to represent a transitional intermediate. In the context of our findings, this implies that the evolution of the higher vertebrate methylation system was achieved after the emergence of jawed vertebrates (~600 Mya⁴³),

but before the divergence of bony and cartilaginous fish (~465 Mya⁴³; Figure 1). These data further identify cartilaginous fish as the most divergent class to possess a DNA modification system similar to our own, and position the elephant shark as a valuable model to examine the function and evolution of the vertebrate methylation system. As the slowest evolving vertebrate documented⁴⁹, the elephant shark bears the closest resemblance to the most recent common ancestor of all jawed vertebrates, enhancing its appeal in this respect. Moreover, the extensive orthology of its small genome to those of tetrapods⁴⁹ facilitates comparative studies.

Transposon aggressiveness correlates with the degree of sexual outcrossing in the host, and repression of this destabilising activity has been proposed as a major reason for genome-wide hypermethylation in sexually-reproducing organisms such as plants and vertebrates^{14,38,54}. This control mechanism appears to have been discarded as unnecessary in early asexual metazoans, and alternative suppression systems such as the piwi-piRNA pathway were developed in their sexually-reproducing invertebrate descendants^{54,55}. The reason for the apparent reinvention of methylation-based silencing in vertebrates is unclear. Comparison of TE dynamics in the cells of elephant shark and basal chordates offers the opportunity to determine whether the need for additional control mechanisms was a primary driver for genome-wide hypermethylation in jawed vertebrates.

We note that in addition to substantial physiological changes, the emergence of jawed vertebrates was accompanied by major innovations in gene regulatory networks, notably non-coding RNA elements⁴⁹. These advances may have facilitated, or conversely been enabled by, the development of a complex methylation system during the same time period. The role of the whole-genome duplications that occurred in vertebrate progenitors⁵⁶ in the acquisition of components that act downstream of the methylation signal, or as a stimulus for new mechanisms of regulating gene dosage, also merits further investigation.

Methylation of elements that modulate gene expression forms an epigenetic memory that plays an important role in defining and stabilising cell identity in higher vertebrates^{5,17–21}. The reprogramming of this specification in the germline to regenerate full developmental competence after fertilisation, and the pathways employed to achieve this – such as active demethylation by ten-eleven-translocase (TET) enzymes, vary considerably across vertebrates⁵⁷. Examination of these phenomena in the elephant shark will provide insight into the evolutionary history of epigenetic control in the life cycle and its consequences for vertebrate development.

Our findings provide fresh perspective on an important epigenetic modification. The elephant shark methylome delineates the evolutionary extent of the complex methylation system found in higher vertebrates, and sets the scene for comparative studies that will address longstanding questions about the primary purpose of this system and how these functions evolved from the mosaic pattern of invertebrates. It will be particularly pertinent

to understand the development of the mechanism that links TSS methylation to transcriptional repression. Epigenetic studies in the elephant shark also open promising avenues to explore the ways in which methylation is put to use during development and the specification of cell fate, and the conservation of these strategies amongst vertebrates.

Data availability

All raw WGBS-seq data (including low-coverage WGBS-seq data), as well as mapped CG call files for male and female liver deep-sequencing, are deposited in the [GEO database](#) under accession number GSE96683.

Source for published WGBS-seq datasets:

Mapped CG methylated calls for mouse liver⁵ were downloaded from [GEO](#), accession GSE42836, sample GSM1051157.

Raw sequencing data for zebrafish muscle⁸ were downloaded from [SRA](#), study SRP020008, run SRR800081.

Raw sequencing data for sea squirt muscle¹⁴ were downloaded from [GEO](#), accession GSE19824, sample GSM497251.

Source for published RNA-seq datasets:

Raw sequencing data for elephant shark liver⁴⁹ were downloaded from [SRA](#), SRA study SRP013772, run SRR513760.

Raw sequencing data for mouse liver (ENCODE Consortium^{50,51}) were downloaded from [GEO](#), accession GSE78583, sample GSM2072415.

Raw sequencing data for zebrafish muscle⁸ were downloaded from [SRA](#), study SRP020008, run SRR800045.

Raw sequencing data for sea squirt muscle¹⁴ were downloaded from [GEO](#), accession GSE19824, sample GSM497252;

Author contributions

TAH conceived the project and JRP and TAH designed the study. JRP performed data analysis and wrote the manuscript. OK prepared WGBS-seq libraries. OO-R performed bootstrap sampling. TAH supervised the study, assisted with data analysis and contributed to the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

This research was funded by a University of Otago Research Committee Grant (111899.01.R.LA).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors would like to acknowledge Andrew Grey (University of Otago) for kindly providing statistical advice. We are also grateful to Les McNoe and Aaron Jeffs (Otago Genomics) for assistance with high-throughput sequencing, and Felix Krueger and Simon Andrews (Babraham Institute) for bioinformatics advice.

Supplementary material

Table S1: Whole-genome bisulfite sequencing of elephant shark somatic tissues.

The table lists the number of cytosine calls at either symmetric CG dinucleotides ('CG') or in other sequence contexts ('non-CG'), mapped against the elephant shark 6.1.3 genome assembly or mitochondrial DNA. Details of bioinformatic processing are provided in the methods section. For deep-sequenced samples, the number of calls following deduplication are given. The frequency of non-CG methylation indicates the maximum rate of non-conversion during the bisulfite treatment step; by this measure, all libraries had a bisulfite conversion efficiency of at least 98.9%.

[Click here to access the data.](#)

Figure S1: Bootstrap sampling to determine margin of error in low-coverage WGBS-seq.

Empirical prediction of the margin of error (99% confidence interval) associated with low coverage WGBS-seq, as calculated by bootstrap sampling of the deep-sequenced male elephant shark liver dataset. Details of the sampling approach are provided in the methods section. An asymptotic model with the equation $y=2.208/\sqrt{x}$ was used to fit a curve to the data. At our minimum sequencing depth of 15,000 CG calls, bootstrap sampling predicts a margin of error of approximately ± 1.8 methylation percentage points.

[Click here to access the data.](#)

References

1. Li E, Bestor TH, Jaenisch R: **Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.** *Cell.* 1992; 69(6): 915–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Okano M, Bell DW, Haber DA, et al.: **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.** *Cell.* 1999; 99(3): 247–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Tittle RK, Sze R, Ng A, et al.: **Uhrf1 and Dnmt1 are required for development and maintenance of the zebrafish lens.** *Dev Biol.* 2011; 350(1): 50–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Goll MG, Bestor TH: **Eukaryotic cytosine methyltransferases.** *Annu Rev Biochem.* 2005; 74: 481–514.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Hon GC, Rajagopal N, Shen Y, et al.: **Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues.** *Nat Genet.* 2013; 45(10): 1198–206.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Lister R, Pelizzola M, Dowen RH, et al.: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature.* 2009; 462(7271): 315–22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Session AM, Uno Y, Kwon T, et al.: **Genome evolution in the allotetraploid frog *Xenopus laevis*.** *Nature.* 2016; 538(7625): 336–343.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Potok ME, Nix DA, Parnell TJ, et al.: **Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern.** *Cell.* 2013; 153(4): 759–72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Meissner A, Mikkelsen TS, Gu H, et al.: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature.* 2008; 454(7205): 766–70.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Bogdanović O, Veenstra GJ: **DNA methylation and methyl-CpG binding proteins: developmental requirements and function.** *Chromosoma.* 2009; 118(5): 549–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Chen PY, Feng S, Joo JW, et al.: **A comparative analysis of DNA methylation across human embryonic stem cell lines.** *Genome Biol.* 2011; 12(7): R62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Laurent L, Wong E, Li G, et al.: **Dynamic changes in the human methylome during differentiation.** *Genome Res.* 2010; 20(3): 320–331.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. McGaughey DM, Abaan HO, Miller RM, et al.: **Genomics of CpG methylation in developing and developed zebrafish.** *G3 (Bethesda).* 2014; 4(5): 861–869.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Zemach A, McDaniel IE, Silva P, et al.: **Genome-wide evolutionary analysis of eukaryotic DNA methylation.** *Science.* 2010; 328(5980): 916–919.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Laine VN, Gossman TI, Schachtschneider KM, et al.: **Evolutionary signals of selection on cognition from the great tit genome and methylome.** *Nat Commun.* 2016; 7: 10474.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Derks MF, Schachtschneider KM, Madsen O, et al.: **Gene and transposable element methylation in great tit (*Parus major*) brain and blood.** *BMC Genomics.* 2016; 17: 332.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Lee HJ, Hore TA, Reik W: **Reprogramming the methylome: erasing memory and creating diversity.** *Cell Stem Cell.* 2014; 14(6): 710–719.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Ziller MJ, Gu H, Müller F, et al.: **Charting a dynamic DNA methylation landscape of the human genome.** *Nature.* 2013; 500(7463): 477–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Hemberger M, Dean W, Reik W: **Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal.** *Nat Rev Mol Cell Biol.* 2009; 10(8): 526–37.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Bagci H, Fisher AG: **DNA demethylation in pluripotency and reprogramming: the role of Tet proteins and cell division.** *Cell Stem Cell.* 2013; 13(3): 265–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Seisenberger S, Andrews S, Krueger F, et al.: **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.** *Mol Cell.* 2012; 48(6): 849–62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet.* 1997; 13(8): 335–340.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. O'Neill RJ, O'Neill MJ, Graves JA: **Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid.** *Nature.* 1998; 393(6680): 68–72.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Smith ZD, Chan MM, Humm KC, et al.: **DNA methylation dynamics of the human preimplantation embryo.** *Nature.* 2014; 511(7511): 611–615.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development.** *Nat Rev Genet.* 2013; 14(3): 204–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Zhong Z, Du K, Yu Q, et al.: **Divergent DNA Methylation Provides Insights into the Evolution of Duplicate Genes in Zebrafish.** *G3 (Bethesda).* 2016; 6(11): 3581–3591. pii: g3.116.032243.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Bogdanovic O, Long SW, van Heeringen SJ, et al.: **Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis.** *Genome Res.* 2011; 21(8): 1313–1327.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Jjingo D, Conley AB, Yi SV, et al.: **On the presence and role of human gene-body DNA methylation.** *Oncotarget.* 2012; 3(4): 462–474.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Guo JU, Ma DK, Mo H, et al.: **Neuronal activity modifies the DNA methylation landscape in the adult brain.** *Nat Neurosci.* 2011; 14(10): 1345–1351.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Neri F, Rapelli S, Krepelova A, et al.: **Intragenic DNA methylation prevents spurious transcription initiation.** *Nature.* 2017; 543(7643): 72–77.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Maunakea AK, Chepelev I, Cui K, et al.: **Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition.** *Cell Res.* 2013; 23(11): 1256–69.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Shukla S, Kavak E, Gregory M, et al.: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature.* 2011; 479(7371): 74–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Tweedie S, Charlton J, Clark V, et al.: **Methylation of genomes and genes at the invertebrate-vertebrate boundary.** *Mol Cell Biol.* 1997; 17(3): 1469–1475.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Feng S, Cokus SJ, Zhang X, et al.: **Conservation and divergence of methylation patterning in plants and animals.** *Proc Natl Acad Sci U S A.* 2010; 107(19): 8689–8694.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Suzuki MM, Kerr AR, De Sousa D, et al.: **CpG methylation is targeted to transcription units in an invertebrate genome.** *Genome Res.* 2007; 17(5): 625–631.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Raddatz G, Guzzardo PM, Olova N, et al.: **Dnmt2-dependent methylomes lack defined DNA methylation patterns.** *Proc Natl Acad Sci U S A.* 2013; 110(21): 8627–31.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Sarda S, Zeng J, Hunt BG, et al.: **The Evolution of Invertebrate Gene Body Methylation.** *Mol Biol Evol.* 2012; 29(8): 1907–1916.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Simmen MW, Leitgeb S, Charlton J, et al.: **Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome.** *Science.* 1999; 283(5405): 1164–1167.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Holland LZ, Gibson-Brown JJ: **The *Ciona intestinalis* genome: when the constraints are off.** *Bioessays.* 2003; 25(6): 529–532.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Keller TE, Han P, Yi SV: **Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary.** *Mol Biol Evol.* 2016; 33(4): 1019–1028.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Suzuki MM, Yoshinari A, Obara M, et al.: **Identical sets of methylated and nonmethylated genes in *Ciona intestinalis* sperm and muscle cells.** *Epigenetics Chromatin.* 2013; 6(1): 38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Delsuc F, Brinkmann H, Chourrout D: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature.* 2006; 439: 965–968.
[Publisher Full Text](#)
43. Hedges SB, Dudley J, Kumar S: **TimeTree: A public knowledge-base of divergence times among organisms.** *Bioinformatics.* 2006; 22(23): 2971–2972.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. Deangelis MM, Wang DG, Hawkins TL: **Solid-phase reversible immobilization for the isolation of PCR products.** *Nucleic Acids Res.* 1995; 23(22): 4742–4743.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Peat JR, Dean W, Clark SJ, et al.: **Genome-wide Bisulfite Sequencing in Zygotes Identifies Demethylation Targets and Maps the Contribution of TET3 Oxidation.** *Cell Rep.* 2014; 9(6): 1990–2000.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Krueger F, Andrews SR: **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.** *Bioinformatics.* 2011; 27(11): 1571–2.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Langmead B, Trapnell C, Pop M, et al.: **Ultrafast and memory-efficient alignment**

- of short DNA sequences to the human genome. *Genome Biol.* 2009; **10**(3): R25.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Wickham HM: **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2009.
[Reference Source](#)
49. Venkatesh B, Lee AP, Ravi V, *et al.*: **Elephant shark genome provides unique insights into gnathostome evolution**. *Nature*. 2014; **505**(7482): 174–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Sloan CA, Chan ET, Davidson JM, *et al.*: **ENCODE data at the ENCODE portal**. *Nucleic Acids Res.* 2016; **44**(D1): D726–D732.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome**. *Nature*. 2012; **489**(7414): 57–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements**. *Nat Methods*. 2015; **12**(4): 357–360.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Shao C, Li Q, Chen S, *et al.*: **Epigenetic modification and inheritance in sexual reversal of fish**. *Genome Res.* 2014; **24**(4): 604–615.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. Zemach A, Zilberman D: **Evolution of eukaryotic DNA methylation and the pursuit of safer sex**. *Curr Biol.* 2010; **20**(17): R780–R785.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Aravin AA, Hannon GJ, Brennecke J: **The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race**. *Science*. 2007; **318**(5851): 761–764.
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Smith JJ, Kuraku S, Holt C, *et al.*: **Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution**. *Nat Genet.* 2013; **45**(4): 415–21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Bogdanović O, Gómez-Skarmeta JL: **Embryonic DNA methylation: insights from the genomics era**. *Brief Funct Genomics.* 2013; **13**(2): 121–130.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 18 May 2017

doi:10.5256/f1000research.12170.r22066

 **Arthur Georges**¹ , **Changwei Shao**²

¹ Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia

² Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China

In this paper, the authors report use of whole-genome bisulfite sequencing to generate the first genome-wide methylation profiles for a cartilaginous fish, the elephant shark *Callorhynchus milii*. The results are preceded by an excellent introduction to the topic, pointing to the pertinent recent literature. The work is significant because the cartilaginous fish arise from a lineage that branches early from the vertebrate phylogeny and their methylation profiles have not before been examined. The authors are able to compare this methylation profile with that of other vertebrates – zebrafish, mouse and sea squirt -- to show that the cartilaginous fishes exhibit a methylation profile that is characteristic of vertebrates generally, and in contrast to the pattern shown by invertebrates, including the chordate sea squirt. They were also able to report an inverse relationship between TSS methylation and gene expression in the elephant shark, supporting the presence of the repressive regulatory architecture shared by other vertebrates so far examined. This study narrows considerably the evolutionary window in which the widespread methylation pattern characteristic of vertebrates evolved. They demonstrate conservation of a complex methylation system across jawed vertebrates separated by 465 million years of evolution.

The interpretation of the results is sound for the most part, and gives sufficient evidence to support their conclusion of conservation of methylation across jawed vertebrate, which fills the gap in our knowledge on the methylation system of vertebrate compared with invertebrate.

We have concerns about the detail presented on the methylation data to support the conclusions. Information on the basic methylation data is an omission that needs to be rectified, and the reasons to focusing solely on CpG sites for further analyses. What quantity of methylome data was produced, what was the average depth per strand for each sample, and what was the mapping ratio? What was the density of methylated genomic cytosines and how did the detailed distribution pattern of mCs differ with context?

The most interesting point of the paper is that the elephant shark resembled the methylation pattern of other vertebrate lineages based on the characterization of genomic methylation. DNA methylation is catalyzed by the three DNA methyltransferases (DNMT1, 2 and 3). Is it possible to also add a comparative analysis of DNMT to see if the cartilaginous fishes also resemble other vertebrate lineages (multiple copy) in comparison with invertebrates (single copy), including the non-vertebrate chordates?

The authors interpret similarity of male and female methylation to indicate that, unlike some bony fish

species, the uncharacterised sex-determination mechanism in the elephant shark is not associated with large differences in global methylation. We think the authors mean that the outcome of sexual differentiation in the elephant shark does not yield, or is not governed by, large differences in global methylation. Specifically, the reference to sex related differences in methylation in the bony fish species is to work that demonstrates a difference between ovary and testes. As such it refers to differences in tissue specific patterns of methylation, not necessarily related to sex determination itself.

There is a semantic point, and the authors may wish to word it out of the manuscript. It centres on the use of basal taxa and higher vertebrates. The elephant shark studied is extant and so has had as much time to evolve and diverge from the common ancestor as we have. We suggest replacing references to the elephant shark as basal by "the elephant shark has arisen from an early branch of the vertebrate phylogeny". The logic on the implications of the results remains unchanged. Also consider wording out "higher vertebrates" with "other vertebrate lineages".

Apart from those points, the authors might like to address some repetition of points made in the manuscript with a view to removing them and tightening up the manuscript. There appears to be an error in the caption of Figure 5 which states that Quintile 4 is omitted for clarity, but it appears to be included.

We really liked this paper. it is an excellent contribution to our understanding of the evolution of the distinctive global methylation pattern of vertebrates. The work is couched in an introduction to the topic that provides the reader with a good introduction to the topic of global methylation patterns, by which to place the work in context.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 11 May 2017

doi:10.5256/f1000research.12170.r22183

**Matthew M. Hindle**

The Roslin Institute, University of Edinburgh, Edinburgh, UK

Peat *et al.* reveal the first methylome of a cartilaginous fish. A valuable contribution to comparative epigenetics that sheds light on the emergence of DNA methylation as a regulatory mechanism in vertebrate genomes. The lack of any published methylomes between Sea Squirts and Zebrafish emphasises just how important this paper is for understanding the evolution of epigenetics as a mechanism for transcript regulation.

The paper is a great read and an excellent example of how to present genome-wide methylation data in an interesting and compelling format. The main claim that the elephant shark genome has a “higher vertebrate” like methylation profile is convincingly made in Figure 3. I am also persuaded by the higher vertebrate like correlation of CpG methylation and transposable elements (Fig 4B) and TSS (Fig 5 A-D). It is particularly encouraging that they managed to show such a convincing correlation in a non-model species, where genomic annotation is often inaccurate and incomplete. It is tempting to infer that the differences between the elephant shark methylome and those of higher vertebrates indicate an evolutionary intermediate/transitionary stage in the functional importance of TSS methylation to repress transcription (authors indicate there may be “less absolute repression than in higher vertebrates”). However, the authors appear to be very cautious in limiting their claims and acknowledge that the reduced amplitude of the TSS methylation correlation in Fig4B compared to mouse and zebrafish could be due to TSS miss-annotation in a non-model genome. It will be interesting to see if improvements of TSS annotation with CAGE or similar data alter the differences in the TSS methylation ratio. Given the authors carefully qualify their claims on the observed differences of elephant shark methylomes to higher vertebrates, it would be unreasonable to request that they improve on the reference annotation. Similarly, the transcript correlation differences between elephant shark and higher vertebrates in Figure 5 E-F could be the result of incomplete or erroneous transcript models. However, it is very interesting that compared to higher vertebrates there appears to be underrepresentation of CpG methylation for gene bodies in low-level expressed genes and for TSS an slight overrepresentation of CpG methylation in highly expressed genes. Again it is a tantalising finding but difficult to draw conclusions because of the annotation quality differences and the differences in the RNASeq datasets.

It is a shame the WGBS Sea squirt and Zebrafish data is muscle rather than liver (Table 1) but the authors made use of what was available.

I use a similar Babraham pipeline for BS data and all the methods appear to be appropriate.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
