
Figures and figure supplements

Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates

Hannah K Long, et al.

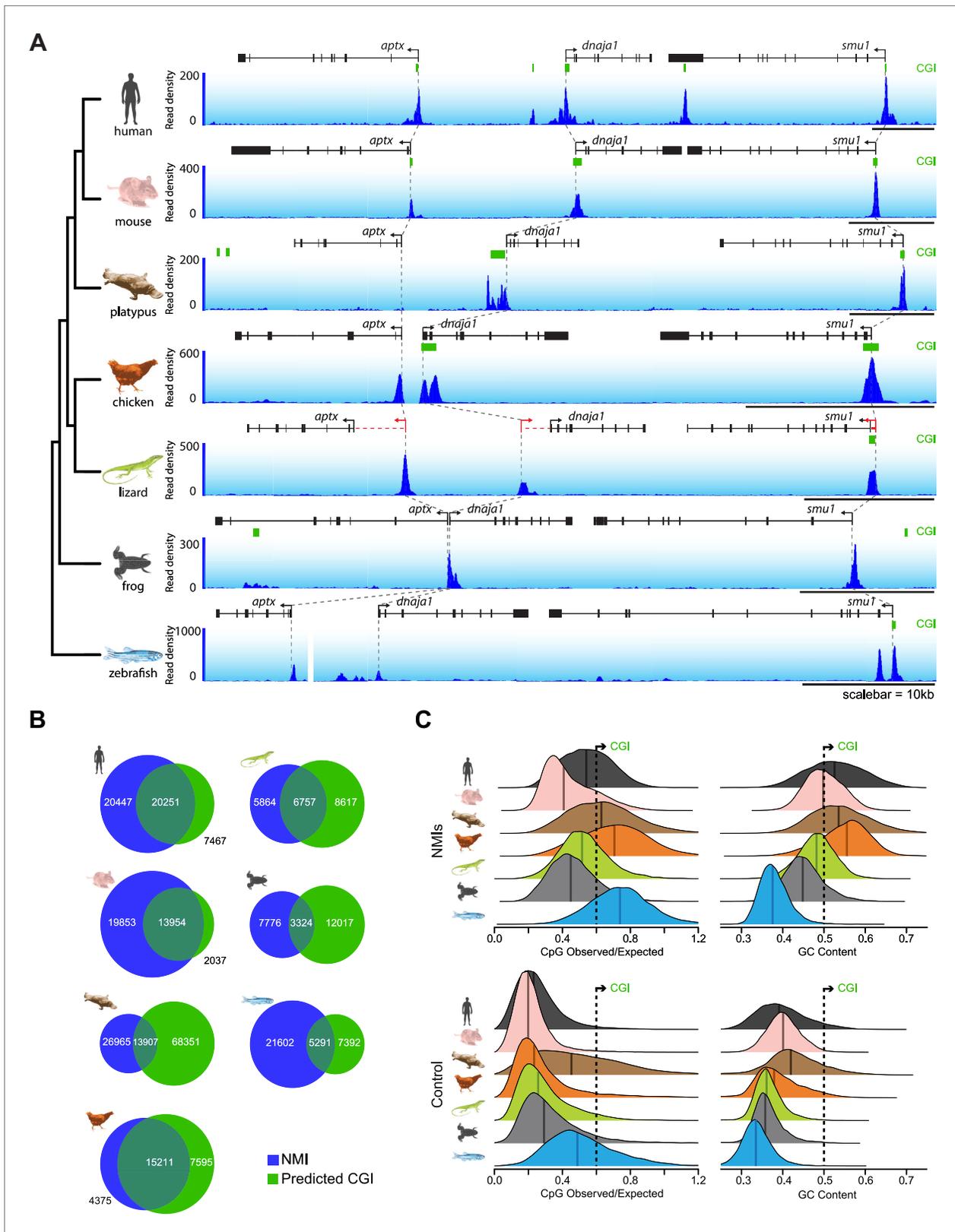


Figure 1. CpG island predictions do not accurately identify non-methylated islands of DNA in vertebrate genomes. **(A)** Non-methylated DNA profiles in testes at a representative syntenic region for seven vertebrate species. Genes are shown in black (improved annotation of gene TSSs using RNA-seq Figure 1. Continued on next page

Figure 1. Continued

data is shown in red), CpG island predictions in green (CGI), and non-methylated DNA profiles are shown in blue. A phylogenetic tree (left) highlights the evolutionary relationship among the seven species. Dashed grey lines highlight the relationship between the gene TSSs across the species. A gap in the zebrafish profile indicates that *aptx* is found at a separate locus from *dnaja1* and *smu1*. **(B)** The genome-wide overlap between CpG islands (green) and non-methylated islands (blue) is depicted as a Venn diagram for each of the species. **(C)** Nucleotide properties of non-methylated islands and control regions are depicted as density plots. CpG observed/expected (left) and GC content (right) are shown for NMI and control regions of the genome. Median values are shown as dark vertical lines. Thresholds for CpG island prediction are indicated (black dashed line).

DOI: [10.7554/eLife.00348.003](https://doi.org/10.7554/eLife.00348.003)

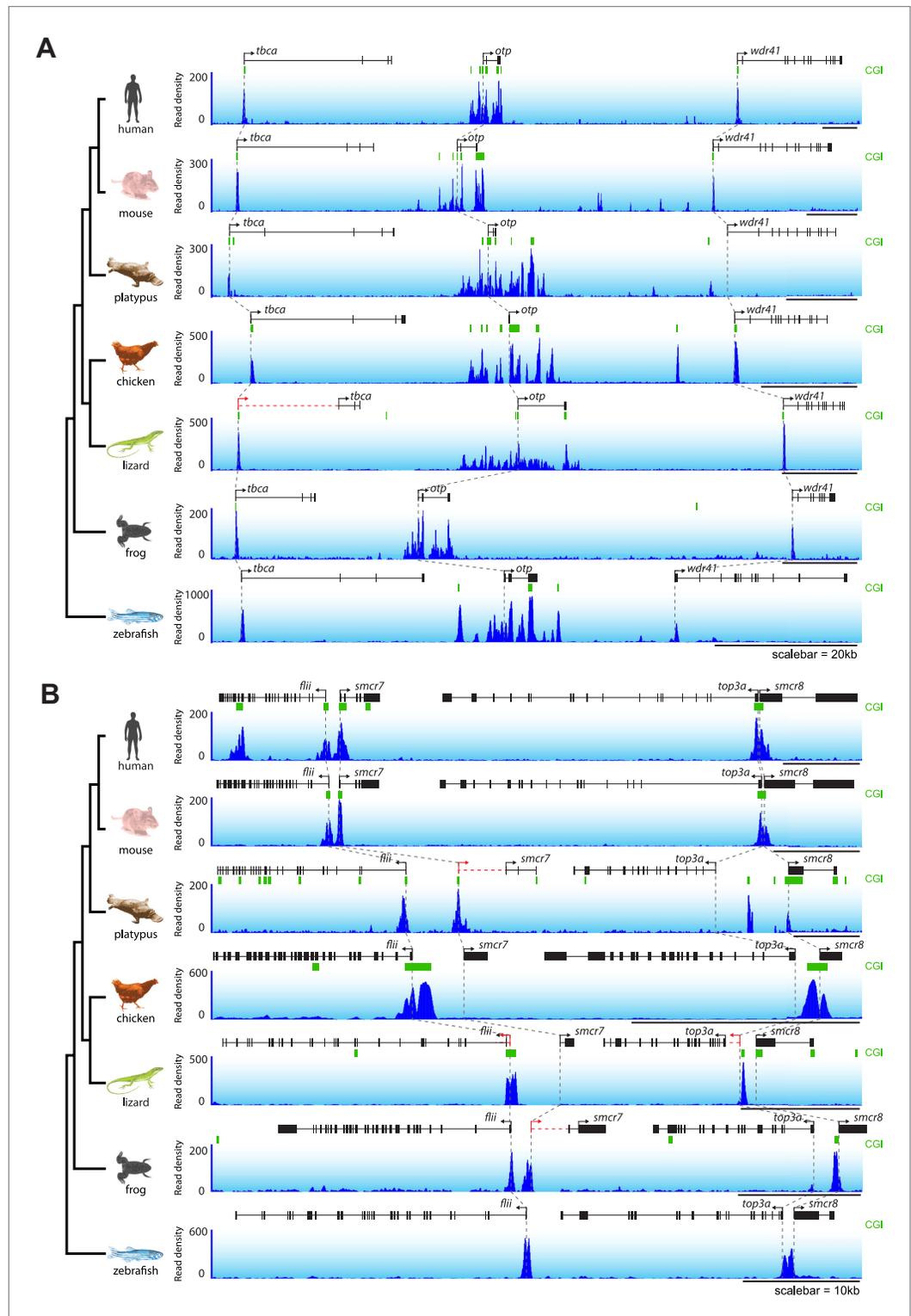


Figure 1—figure supplement 1. NMIs are a conserved feature of vertebrate promoters as illustrated by two syntenic loci. **(A)** and **(B)** Profiles of non-methylated DNA are shown in testes at two representative syntenic regions for seven vertebrate species. Genes are shown in black (improved annotation of gene TSSs using RNA-seq data is shown in red), CpG island predictions in green, and non-methylated DNA profiles are shown in blue. A phylogenetic tree (left) highlights the evolutionary relationship among the seven species and dashed grey lines highlight the relationship between the gene TSSs across the species.

DOI: [10.7554/eLife.00348.004](https://doi.org/10.7554/eLife.00348.004)

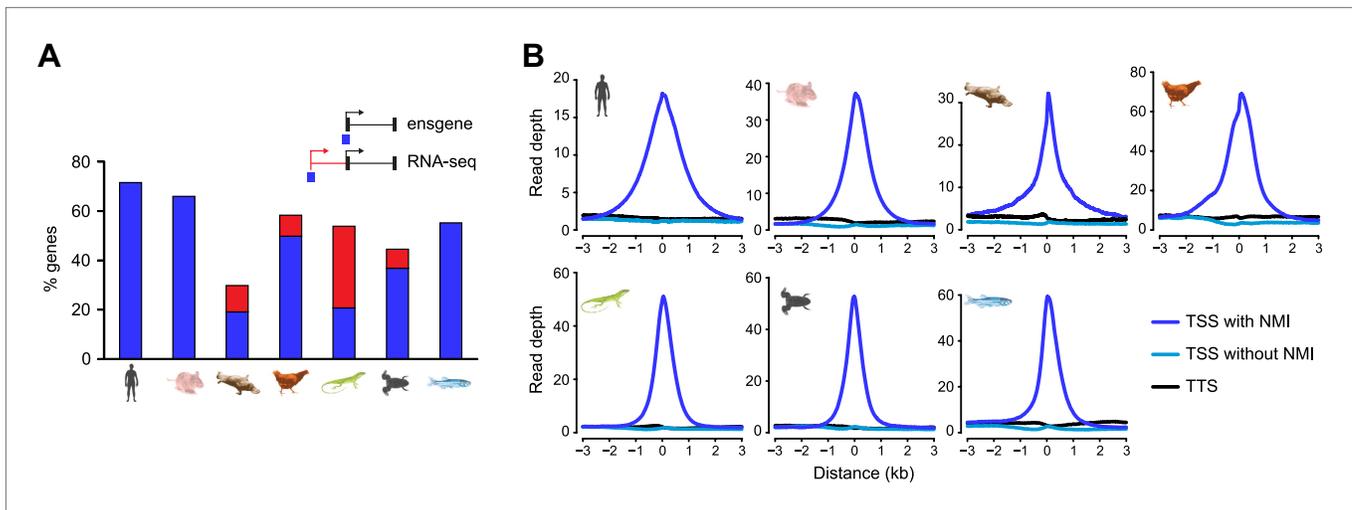


Figure 2. Non-methylated islands are associated with gene promoters in vertebrate genomes. **(A)** A histogram depicting the proportion of protein-coding transcription start sites (TSSs) which are overlapped by an NMI for all seven species. Blue bars indicate overlap with annotated TSSs and red bars indicate overlap with additional TSSs identified using RNA-seq data (platypus, chicken and lizard) or Xtev gene sets (frog). **(B)** Profiles of non-methylated DNA were plotted over a 6-kb window centred on all TSSs with an NMI (dark blue), without an NMI (blue), and for all transcription termination sites (TTS, black). The non-methylated DNA signal peaks at the TSS of gene promoters in all vertebrates.

DOI: 10.7554/eLife.00348.005

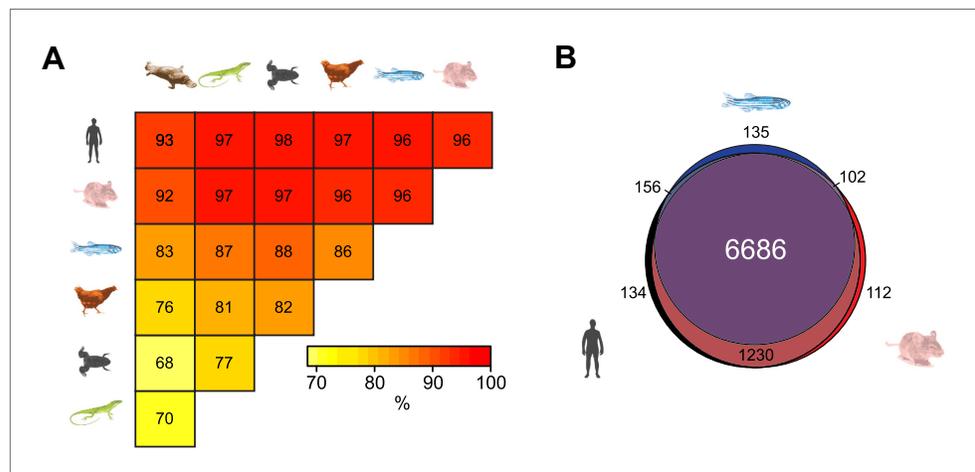


Figure 3. Non-methylated islands are a highly conserved epigenetic feature of vertebrate gene promoters. **(A)** The presence of NMIs at orthologous gene TSSs is preserved as illustrated by a pairwise analysis of NMIs at vertebrate gene orthologues. The percentage of NMIs conserved at orthologous gene TSSs was calculated in a pairwise manner and found to be highly statistically significant for all comparisons across the seven vertebrate species ($p < 10^{-10}$, hypergeometric test). **(B)** A proportional Venn diagram illustrating the three-way comparison of NMI presence at conserved human-mouse-zebrafish gene orthologue TSSs.

DOI: 10.7554/eLife.00348.006

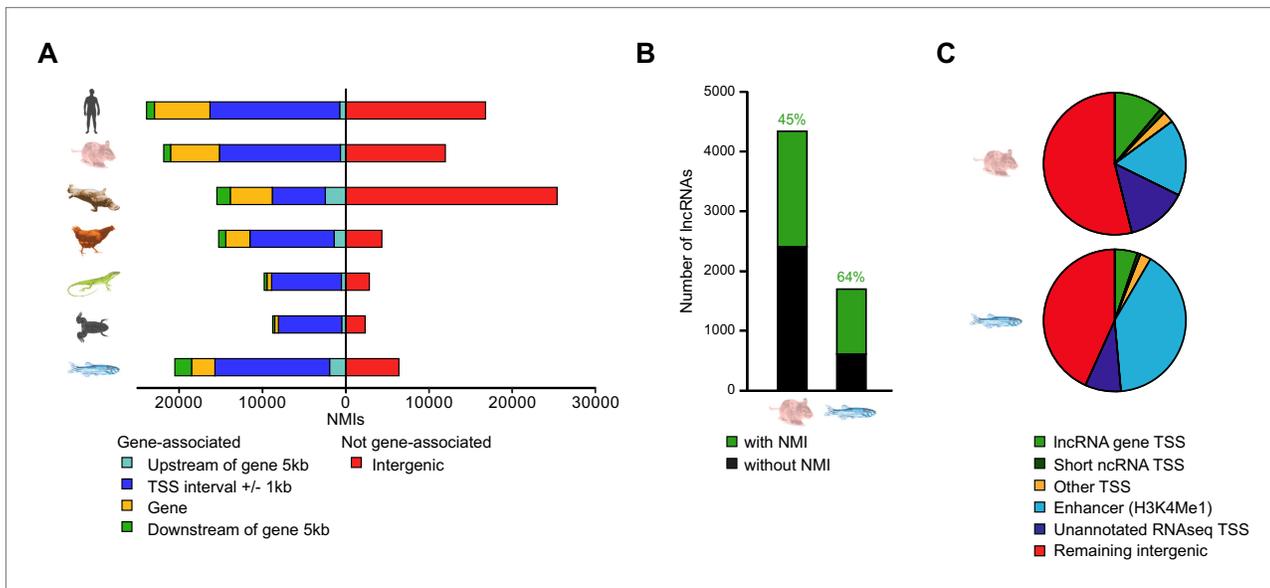


Figure 4. Intergenic NIMs are associated with distal regulatory elements, non-coding RNAs, and unannotated transcripts. **(A)** Most NIMs are associated with known protein-coding genes (left) but a substantial proportion are located within intergenic regions of the genome (right). **(B)** NIMs (green) are found at 45% and 64% of all known long non-coding RNA (lncRNA) TSSs (black) in mouse and zebrafish respectively. **(C)** A pie chart depicting the proportion of intergenic NIMs (>5 kb from a protein-coding gene) associated with different genomic features in mouse embryonic stem (ES) cells and zebrafish 24 hpf embryos. The association was performed hierarchically in the following order: lncRNA TSSs, other non-coding RNA TSSs (miRNAs, rRNAs, snRNAs, or snoRNAs), other TSSs (pseudogenes and processed transcripts), putative enhancer mark H3K4me1 and novel RNA-seq TSSs. This analysis indicates that intergenic NIMs mark novel transcriptional units or regulatory elements.

DOI: [10.7554/eLife.00348.007](https://doi.org/10.7554/eLife.00348.007)

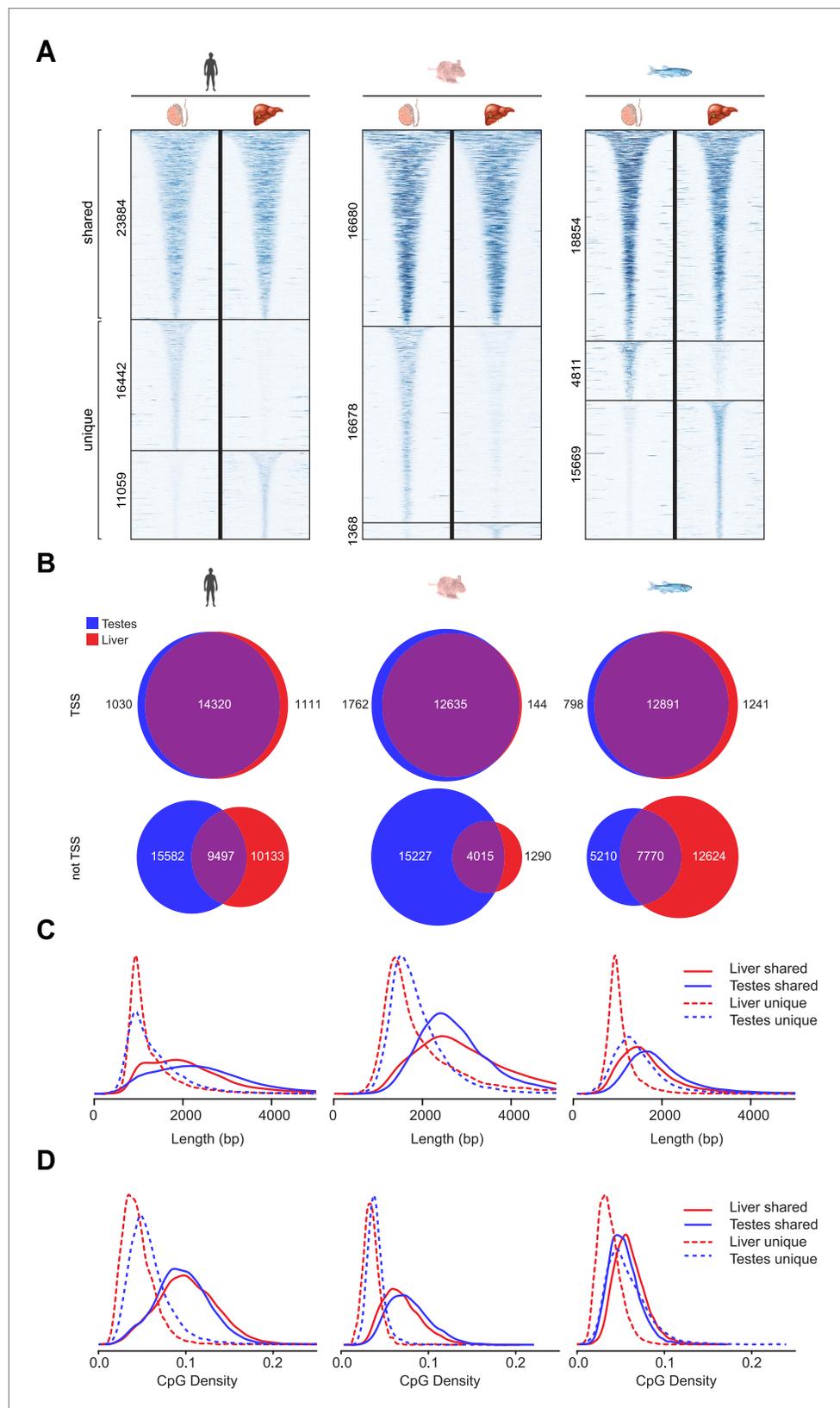


Figure 5. Differential methylation of a subset of NMs. **(A)** All vertebrate genomes have a subset of NMs that are subject to differential methylation as illustrated by a heat map of non-methylated DNA signal from testes and liver *Figure 5. Continued on next page*

Figure 5. Continued

in human, mouse and zebrafish. In each case NMIs are ranked according to length and clustered as shared (upper) or unique (lower) between the two tissues. A 5-kb window centred at the NMI is shown and read density is indicated by colour intensity. **(B)** The overlap of NMIs identified in liver and testes is depicted by Venn diagrams for NMIs associated with protein-coding TSSs (upper) and for NMIs away from TSSs (lower). NMIs at TSSs are generally non-methylated in both tissues whereas differentially methylated NMIs tend to be found away from TSSs. **(C)** NMI length distribution plots for shared (Shared NMIs, solid line) or unique (Unique NMIs, dashed line) NMIs from testes (blue) or liver (red). Shared NMIs tend to be longer than tissue-specific unique NMIs. **(D)** CpG density distribution plots for shared (solid line) or unique (dashed line) NMIs from testes (blue) or liver (red). Shared NMIs tend to have higher CpG density than unique NMIs.

DOI: [10.7554/eLife.00348.008](https://doi.org/10.7554/eLife.00348.008)

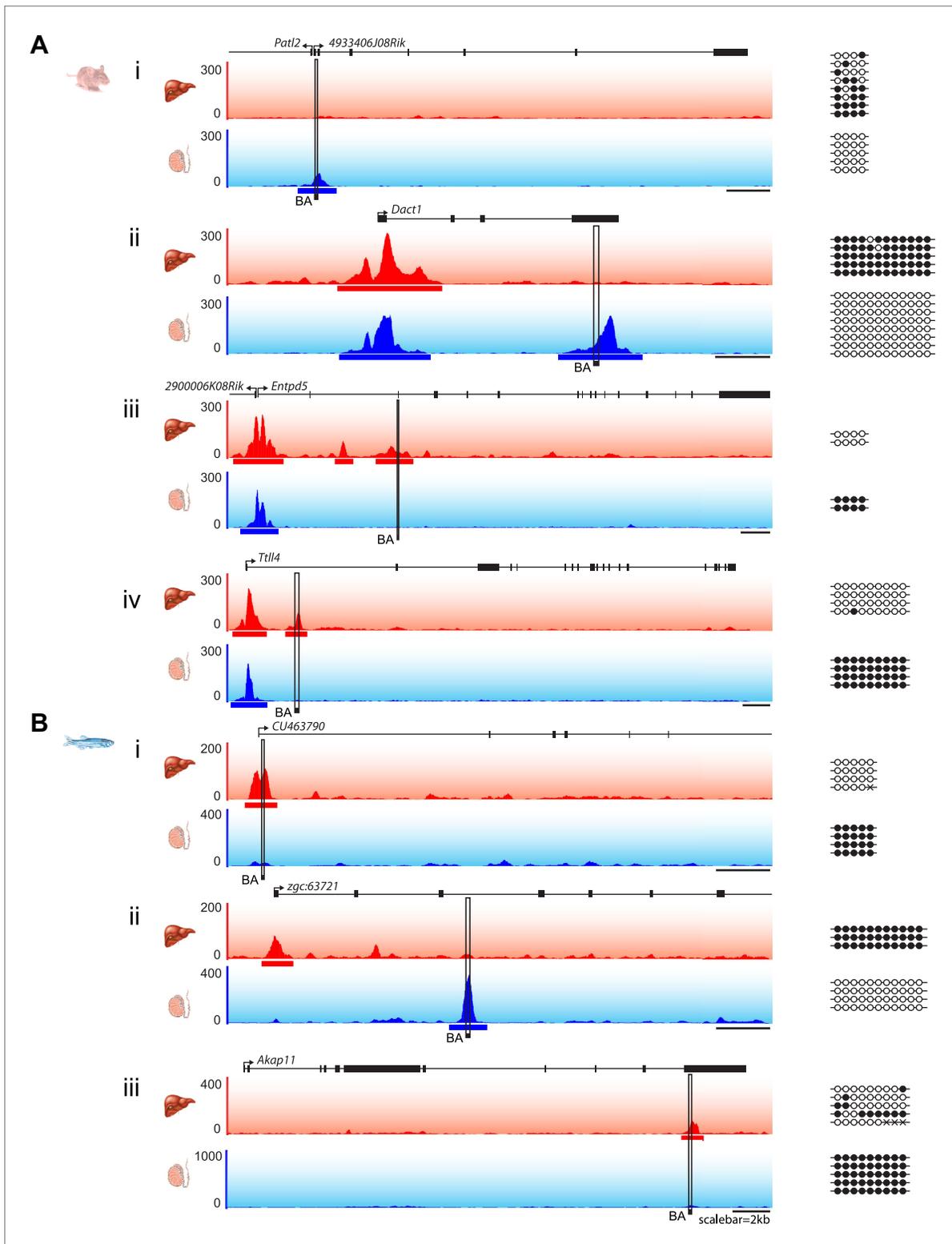


Figure 5—figure supplement 1. Validation of differentially methylated NMs between liver and testes in mouse and zebrafish by bisulfite sequencing. **(A, i–iv)** Mouse NMs unique to liver or testes were analysed by bisulfite sequencing to verify that the regions were indeed differentially methylated. Traces of non-methylated DNA are depicted for differentially methylated regions in mouse liver (red) and testes (blue) with NMs depicted as bars under the traces. The y-axis depicts read density. Methylation status of the unique NMs was confirmed using the indicated bisulfite PCR amplicon (BA, black rectangle). Empty and filled circles represent non-methylated and methylated CpG dinucleotides, respectively. **(B, (i–iii))** Zebrafish NMs unique to liver or testes were validated by bisulfite sequencing as in **(A)**.

DOI: [10.7554/eLife.00348.009](https://doi.org/10.7554/eLife.00348.009)

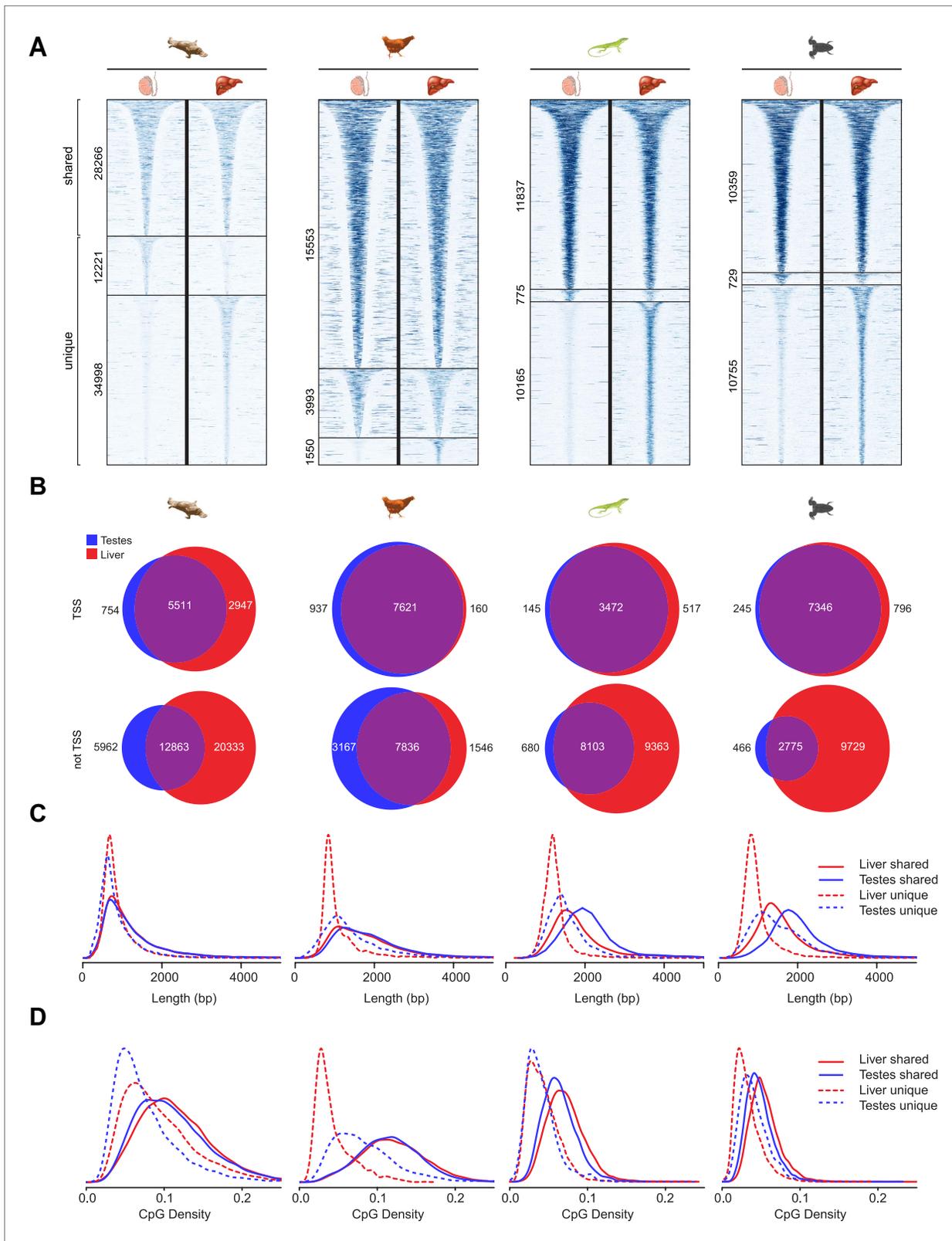


Figure 5—figure supplement 2. Differential methylation of NMIs in platypus, chicken, lizard and frog and length distributions of NMIs from all seven vertebrates. **(A)** A heat map of non-methylated DNA signal from testes and liver in platypus, chicken, lizard and frog. In each case NMIs are ranked according to length and clustered as shared (upper) or unique (lower) between the two tissues. A 5-kb window centred at the NMI is shown and read *Figure 5—figure supplement 2. Continued on next page*

Figure 5—figure supplement 2. Continued

density is indicated by colour intensity. (B) Venn diagrams demonstrate that shared NMI are found predominantly at protein-coding gene TSSs (upper) and unique NMIs tend to be found away from TSSs (lower). (C) NMI length distribution plots for shared (Shared NMIs, solid line) or unique (Unique NMIs, dashed line) NMIs from testes (blue) or liver (red). Shared NMIs tend to be longer than tissue-specific unique NMIs. (D) CpG density distribution plots for shared (solid line) or unique (dashed line) NMIs from testes (blue) or liver (red). Shared NMIs tend to have higher CpG density than unique NMIs.

DOI: 10.7554/eLife.00348.010

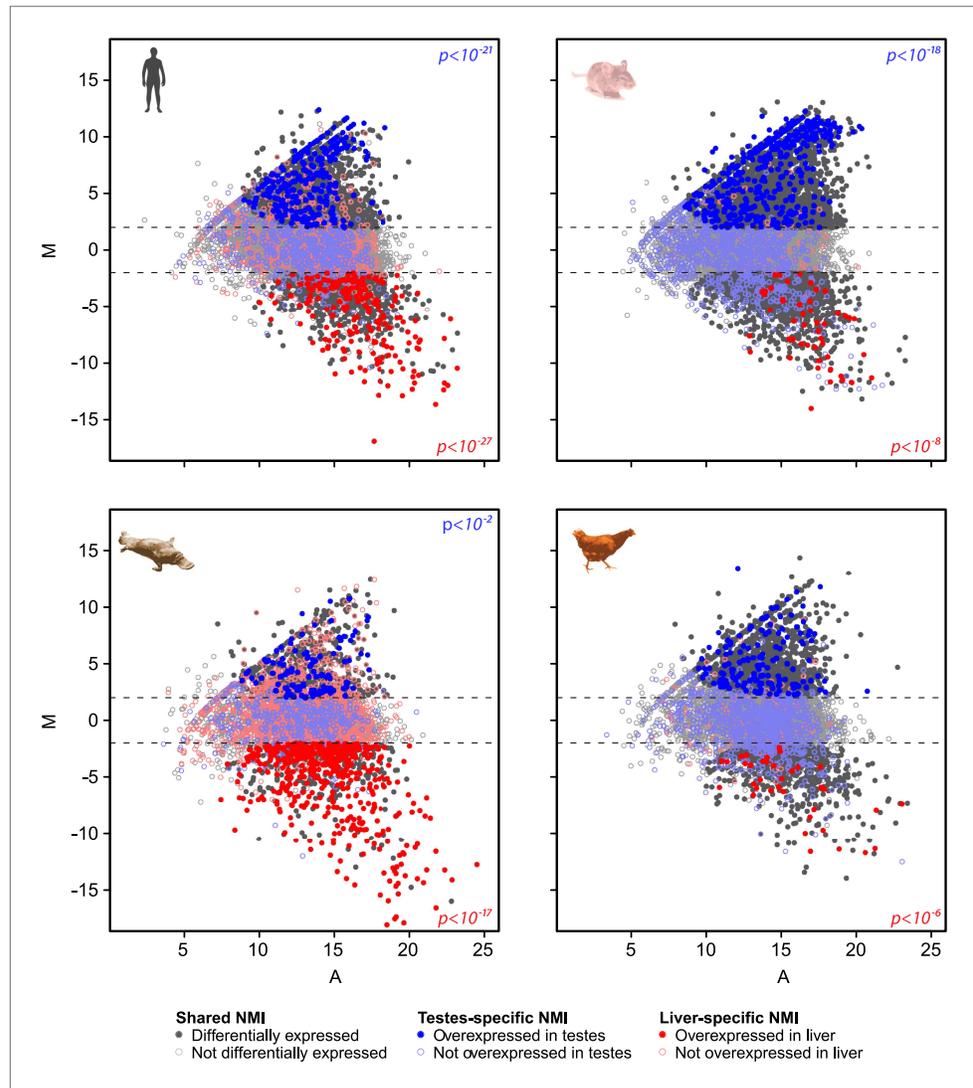


Figure 5—figure supplement 3. Genes with TSS-associated testes or liver specific NMIs are over-represented for increased differential expression in the same tissue. MA plots depicting expression differences for genes with TSS-associated NMIs from liver and testes for human, mouse, platypus and chicken. Genes are coloured according to whether they share an NMI in both liver and testes (grey) or have an NMI only in liver (red) or testes (blue). Genes are further distinguished as being differentially expressed or overexpressed in a tissue-specific manner (dark, filled circle) or not (light, open circles). The log mean expression of the gene from both liver and testes is displayed on the x axis (A) and the log ratio of gene expression is displayed on the y axis (M). The dotted lines indicate a fold change threshold of two. Genes with tissue-specific NMIs were significantly over-represented in the set of genes which had increased differential expression in seven out of eight cases (Fisher's exact test, human testes $p < 10^{-21}$, liver $p < 10^{-27}$; mouse testes $p < 10^{-18}$, liver $p < 10^{-8}$; platypus testes $p < 10^{-2}$, liver $p < 10^{-17}$; chicken liver $p < 10^{-6}$).

DOI: 10.7554/eLife.00348.011

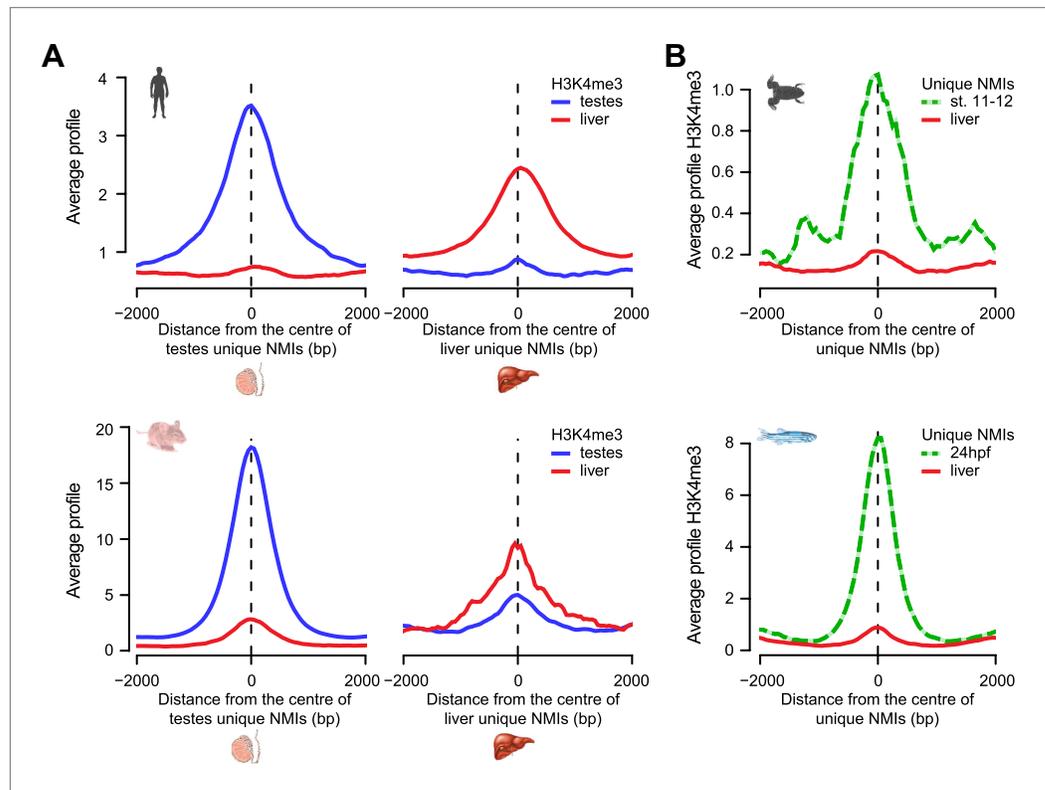


Figure 6. Chromatin modification at NIMs depends on their underlying DNA methylation state. **(A)** H3K4me3 read density from testes (blue) and liver (red) is profiled over testes unique (left) and liver unique (right) NIMs for human (upper) and mouse (lower) and displayed as an average profile. At differentially methylated loci, the histone H3K4me3 modification is found preferentially in the tissue with the non-methylated NIM. **(B)** The H3K4me3 signal (profiled in frog stage 11–12 embryos and zebrafish 24 hpf) is present specifically at unique NIMs from frog stage 11–12 and zebrafish 24 hpf (green) and not at unique NIMs from the liver (red).

DOI: [10.7554/eLife.00348.012](https://doi.org/10.7554/eLife.00348.012)

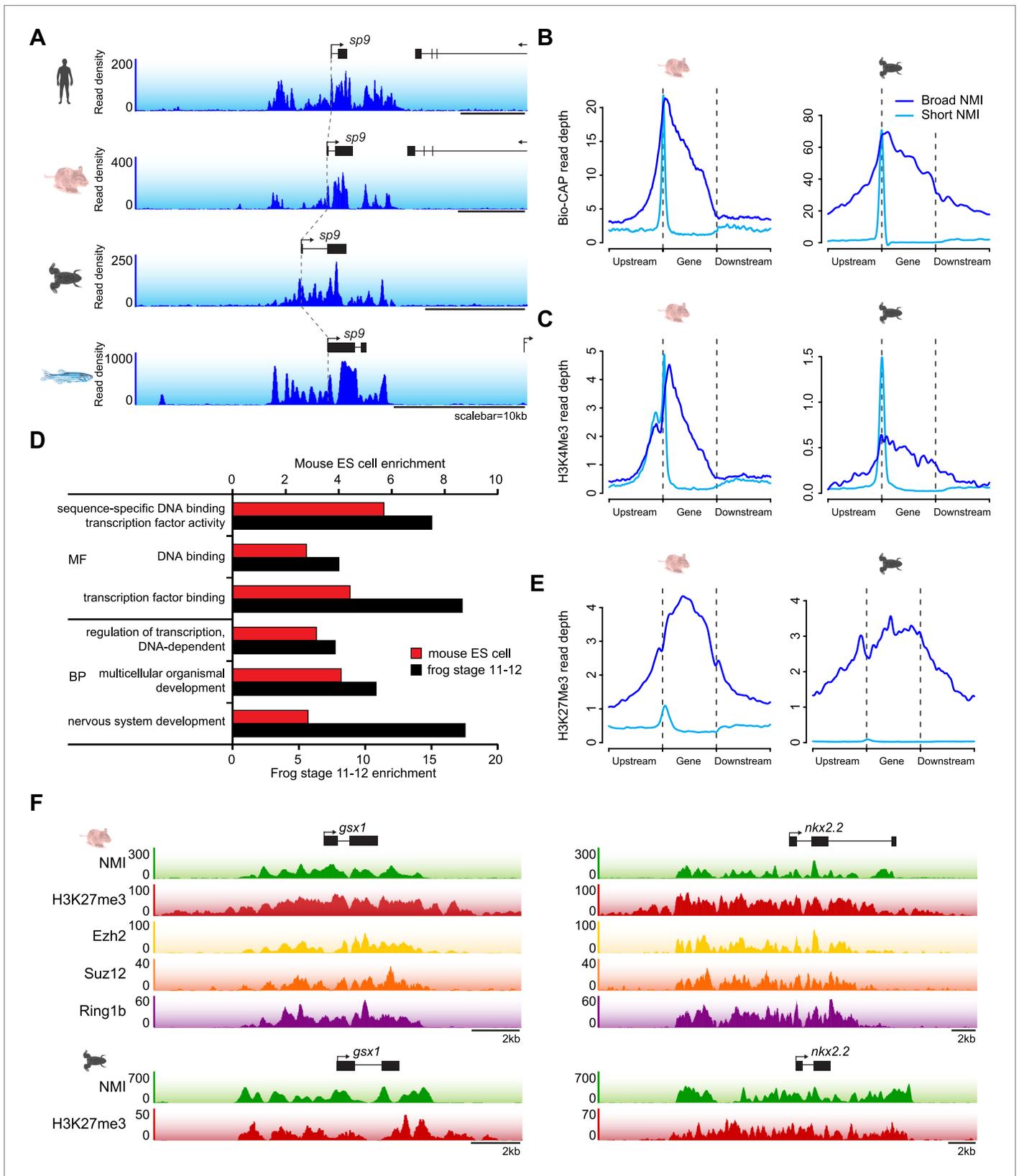


Figure 7. A unique class of broad non-methylated islands encompass polycomb-regulated developmental genes. **(A)** An example of a broad region of non-methylated DNA associated with the *sp9* gene for four representative species (human, mouse, frog and fish). Dashed grey lines highlight the

Figure 7. Continued on next page

Figure 7. Continued

location of the gene TSSs across the four species. (B) Non-methylated DNA profiles are depicted for genes associated with broad NMs (dark blue) and canonical NMs (light blue) in mouse embryonic stem (ES) cells and frog stage 11–12. The profile is scaled to show an averaged gene with one gene length depicted upstream and downstream. (C) H3K4me3 ChIP-seq signal from mouse and frog was plotted as in (B). H3K4me3 profiles reflect the underlying non-methylated DNA profiles. (D) Genes associated with broad NMs were analysed by gene ontology (GO) analysis for mouse ES cell and frog stage 11–12. Broad NMs are found to be significantly enriched for GO term categories associated with sequence-specific DNA binding, transcriptional regulation and development. MF: molecular function; BP: biological process. $p < 10^{-5}$ for all GO terms. (E) H3K27me3 ChIP-seq signal from mouse and frog was plotted for the same gene sets as in (B). The profile is scaled to show an averaged gene with three gene lengths depicted upstream and downstream. As for H3K4me3, H3K27me3 ChIP-seq profiles correspond to the underlying non-methylated DNA profile. (F) A representative example of two broadly non-methylated genes *gsx1* and *nkx2.2* for mouse and frog. In both species, the broad non-methylated regions (green) are associated with the polycomb repressive mark H3K27me3 (red). In addition, in mouse, polycomb repressive complex 2 (*ezh2*, yellow and *suz12*, orange) and polycomb repressive complex 1 (*ring1b*, purple) components are associated with the broad non-methylated regions. The y-axis depicts read density. Genes are depicted above the profiles in black.

DOI: 10.7554/eLife.00348.013

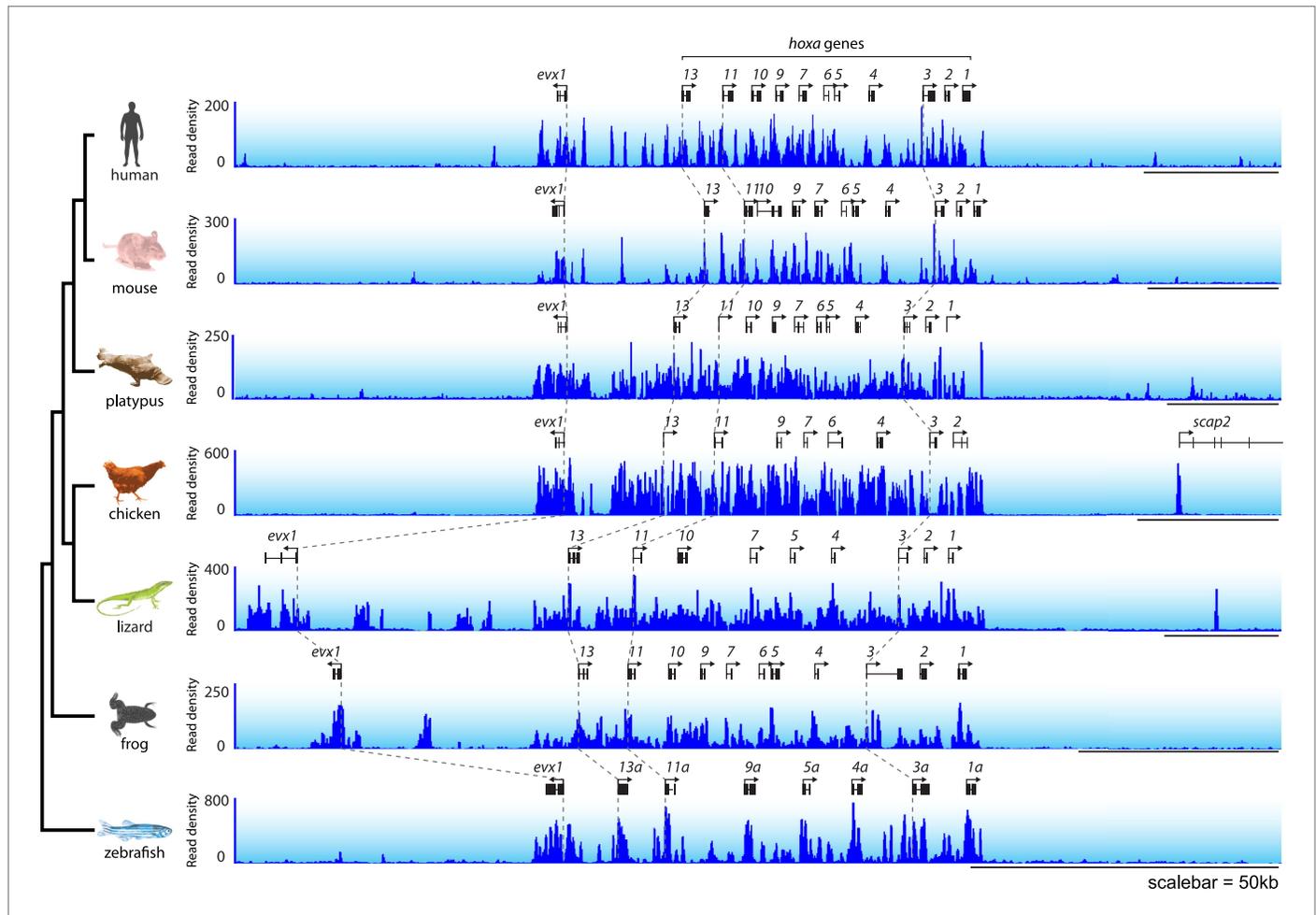


Figure 7—figure supplement 1. *Hox* gene clusters are characterized by broad NMs. The *Hoxa* gene cluster from all seven vertebrate species is associated with broad regions of non-methylated DNA. Genes are shown in black and non-methylated DNA profiles are shown in blue and dashed grey lines highlight the relationship between conserved gene TSSs across the species.

DOI: 10.7554/eLife.00348.014