



Figures and figure supplements

Chromatin accessibility dynamics across *C. elegans* development and ageing

Jürgen Jänes et al

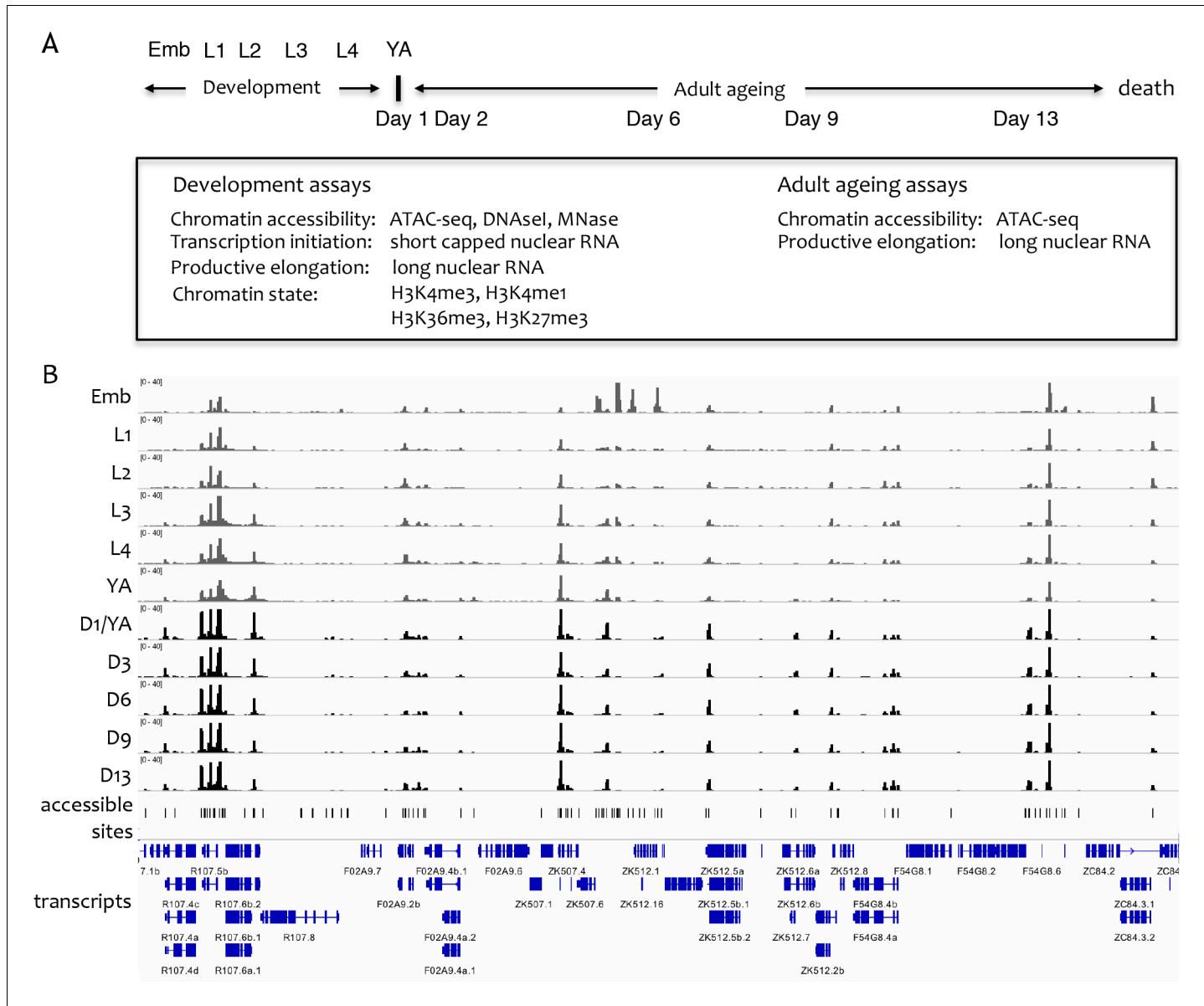


Figure 1. Overview of the project. (A) Overview of genome-wide assays and time points of developmental and ageing samples. For development samples, chromatin accessibility, transcription initiation, productive elongation, and chromatin state were profiled in six stages of wild-type animals (embryos, four larval stages, young adults). For ageing samples, chromatin accessibility and productive transcription elongation were profiled in five time points of sterile adult *glp-1* mutants (Day 1/Young adult, Day 2, Day 6, Day 9, Day 13). (B) Representative screen shot of normalized genome-wide accessibility profiles in the eleven samples (chrIII:9,041,700–9,196,700, 154 kb).

DOI: <https://doi.org/10.7554/eLife.37344.002>

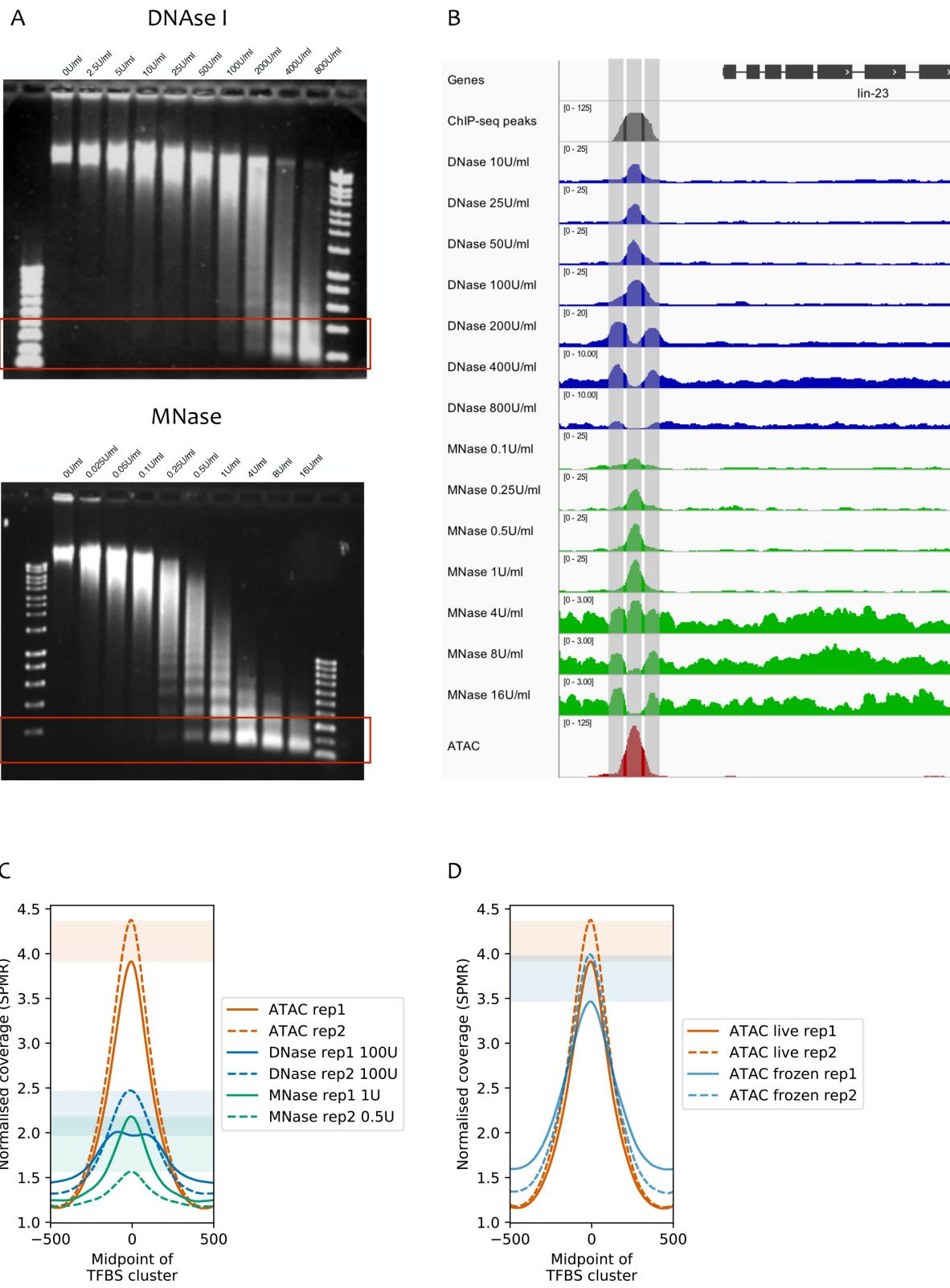


Figure 1—figure supplement 1. Comparison of ATAC-seq to concentration courses of DNase I-seq and MNase-seq. (A) Genomic DNA digested using different concentrations of DNase I (top) or MNase (bottom). Red rectangles highlight approximate size ranges subjected to paired-end Illumina sequencing. Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

sequencing. (B) SPMR-normalized coverage of a DNase I concentration series (blue tracks), MNase concentration series (green tracks), and ATAC-seq (red track) at the *lin-23* locus (chrII:6,369,650–6,373,750, 4.1 kb). The modENCODE/modERN ChIP-seq peak call pileup (grey track) shows a TF binding region upstream of the gene. Different concentrations of nuclease show different types of signal. Low concentrations of DNase I and MNase produce a peak in the middle of the TF-binding region, at the expected NDR (middle vertical bar). At higher concentrations, both enzymes show a peak at the -1 and $+1$ nucleosomes (left and right vertical bars). ATAC-seq has a single large peak centered in the middle of the TF-binding region. (C) Mean normalized coverage at transcription factor binding sites defined by clustering modENCODE/modERN peak calls ($n = 36,389$; Materials and methods) in ATAC-seq, DNase-seq, and MNase-seq (the latter two are shown at concentrations with the highest accessibility enrichment). ATAC-seq shows higher signal than DNase-seq or MNase-seq. Shaded rectangles show the range of signal between assay replicates at the midpoint of the TFBS cluster. (D) Normalized read coverage of ATAC-seq prepared from nuclei harvested from live (red), or frozen (blue) embryos. Shaded rectangles are used as in (C).

DOI: <https://doi.org/10.7554/eLife.37344.003>

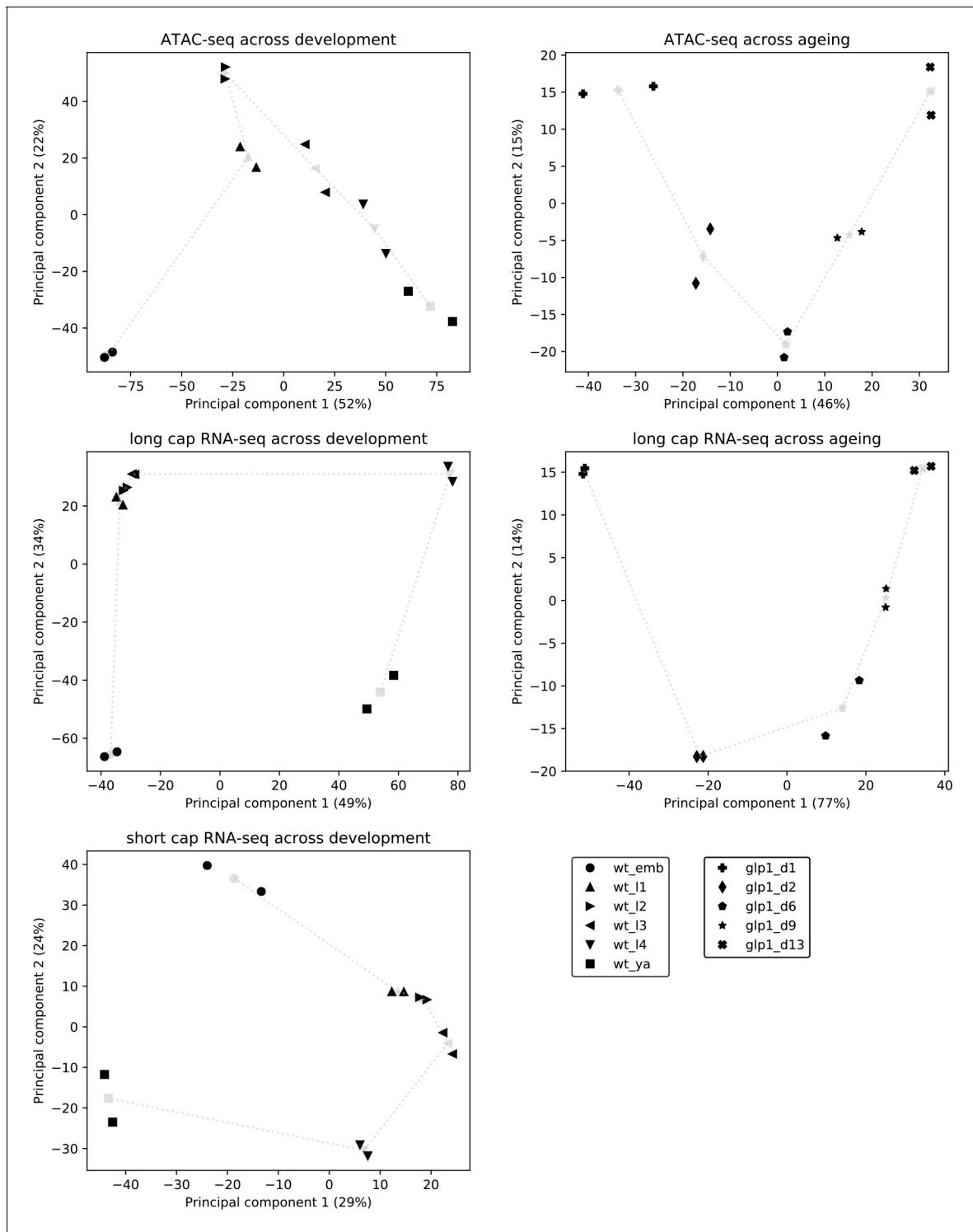


Figure 1—figure supplement 2. Reproducibility and broad relatedness of ATAC-seq and RNA-seq data. Reproducibility and broad relatedness of the different samples and time points in ATAC-seq (top), long cap RNA-seq (middle), and short cap RNA-seq (bottom). We applied PCA to peak accessibility at promoters (ATAC-seq), read counts at annotated genes (long cap RNA-seq), and 5' end read counts at promoters (short cap RNA-seq). Black markers show different biological samples (two per time point) whereas gray markers show centroids, calculated by averaging the two samples collected at the same time point. Gray dotted lines show the time progression across development or ageing.

DOI: <https://doi.org/10.7554/eLife.37344.004>

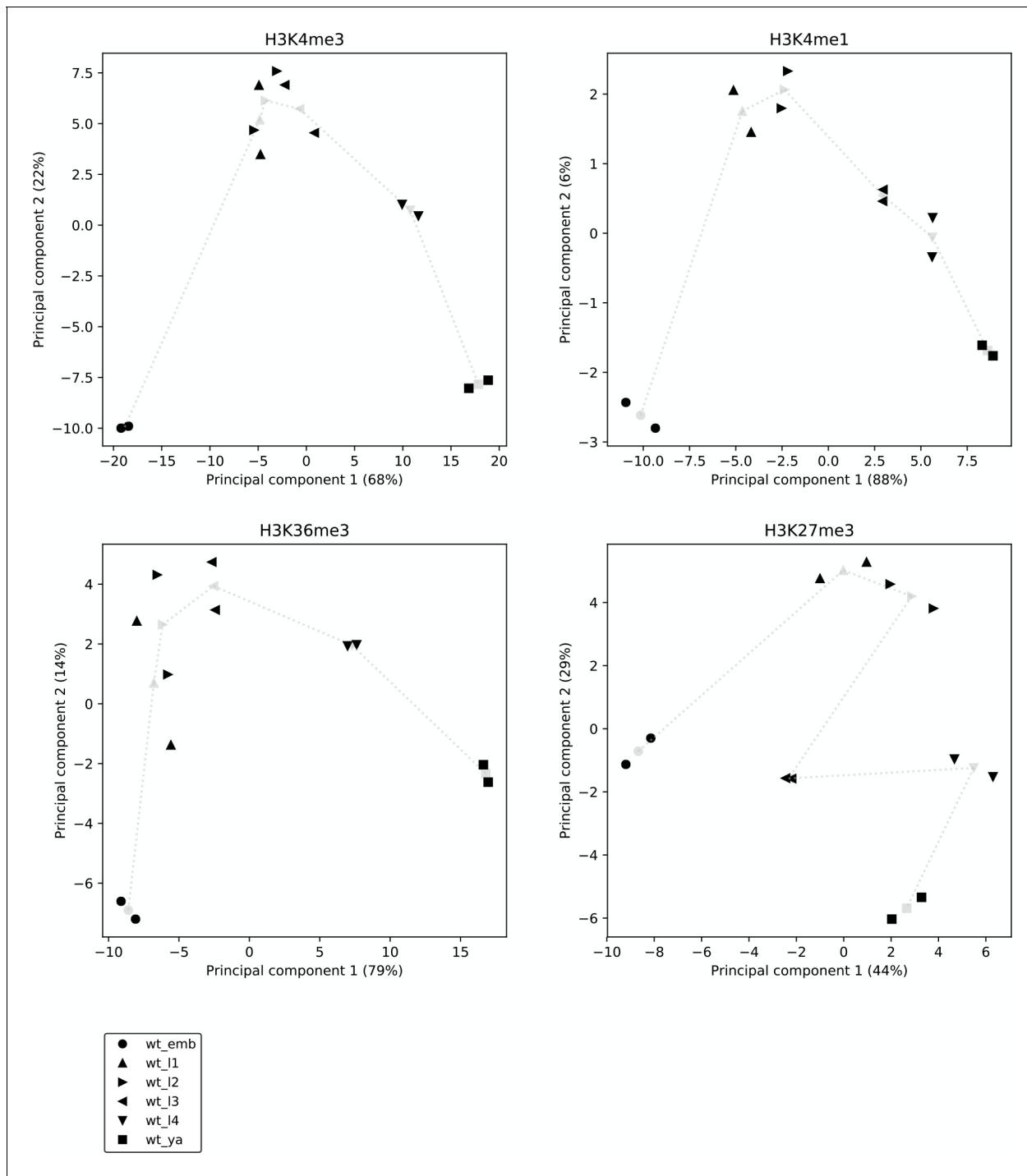


Figure 1—figure supplement 3. Reproducibility and broad relatedness of the histone modification data. Reproducibility and broad relatedness of the different samples and time points in H3K4me3 (top left), H3K4me1 (top right), H3K36me3 (bottom left), and H3K27me3 ChIP-seq (bottom right). We applied PCA to genic regions, from the most upstream promoter to the annotated 3' end, considering genes with at least one promoter. Black markers show different biological samples (two per time point), whereas gray markers show centroids, calculated by averaging the two samples collected at the same time point. Gray dotted lines show the time progression across development.

DOI: <https://doi.org/10.7554/eLife.37344.005>

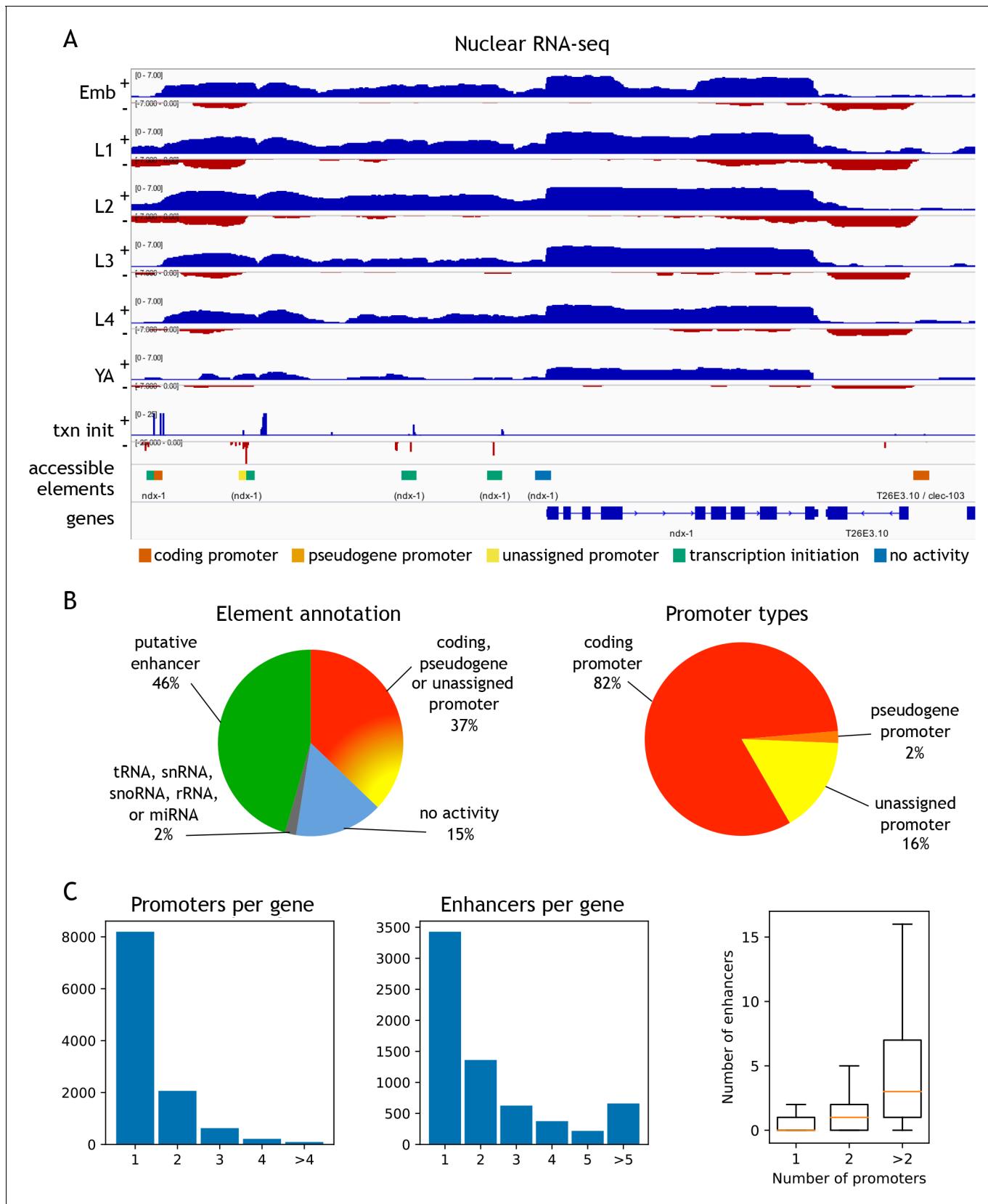


Figure 2. Annotation of accessible elements. (A) Top, strand-specific nuclear RNA in each developmental stage monitors transcription elongation; plus strand, blue; minus strand, red. Below is transcription initiation signal, accessible elements (colored by annotation), and gene models (chr12:12,675,000–Figure 2 continued on next page

Figure 2 continued

12,683,400, 8.4 kb). The left side of each element is colored by the reverse strand annotation whereas the right side of an element is colored by the forward strand annotation (color key at bottom). (B) Left, distribution of accessible sites in four categories: promoters (one or both strands), putative enhancers, no activity, or overlapping a tRNA, snRNA, snoRNA, rRNA, or miRNA. Right, distribution of different types of promoter annotations. (C) Left, distribution of the number of promoters and enhancers per gene; right, boxplot shows that genes with more promoters also have more enhancers.

DOI: <https://doi.org/10.7554/eLife.37344.007>

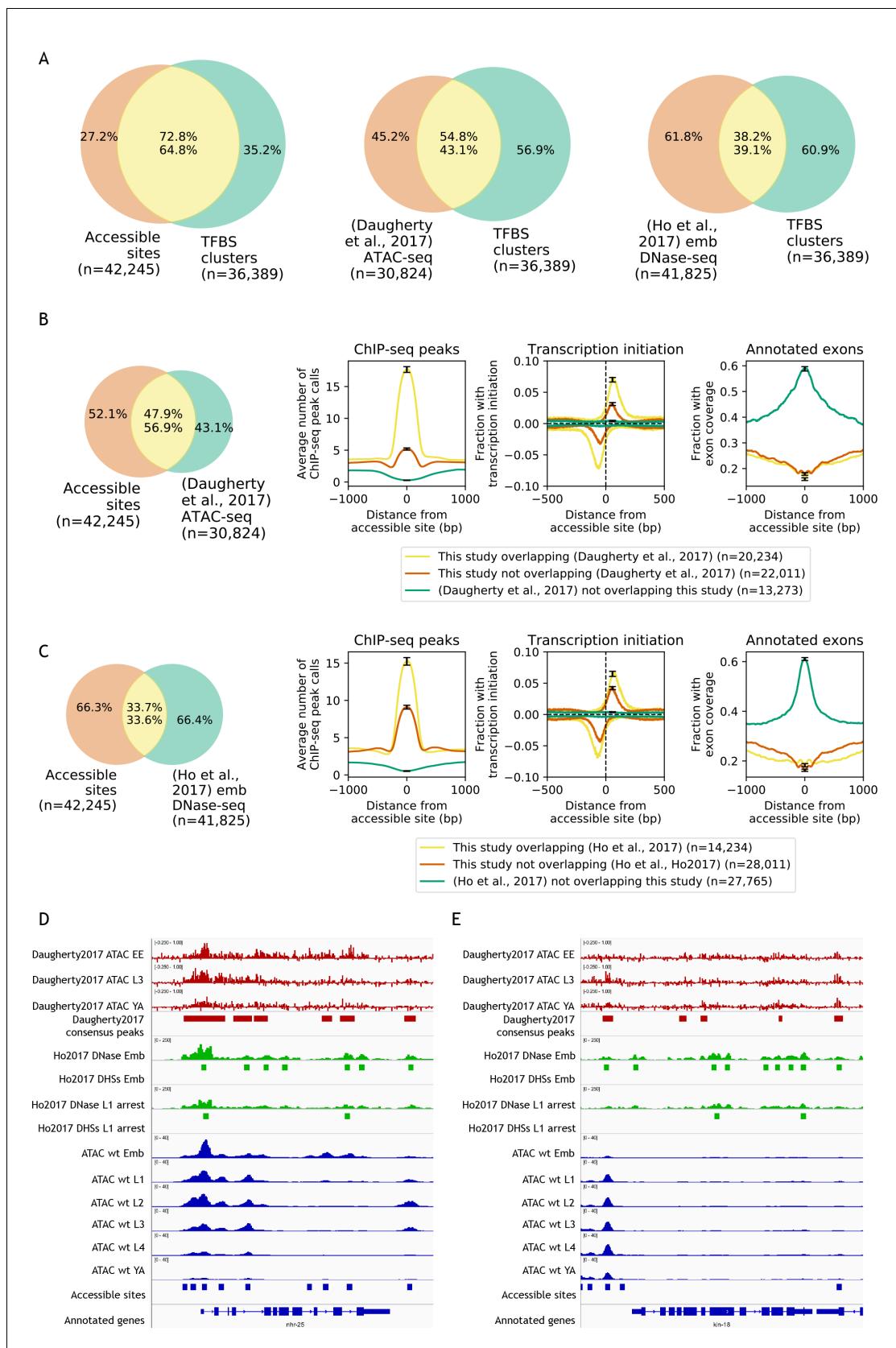


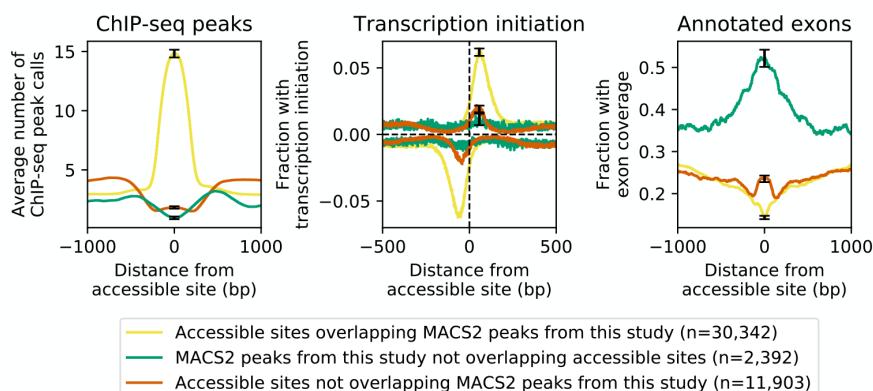
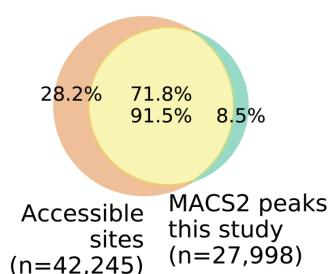
Figure 2—figure supplement 1. Comparisons to previous accessibility maps. (A) Venn diagrams showing the overlap of transcription factor binding sites defined by clustering modENCODE/modERN peak calls ($n = 36,389$; Materials and methods) to accessible sites from this study and two previous studies. (B) and (C) Comparison of ChIP-seq peak distribution, transcription initiation, and annotated exons around accessible sites for overlapping and non-overlapping regions across three studies. (D) and (E) Genomic tracks for *rnf25* and *kin-18* showing Daugherty2017 ATAC EE, L3, YA, consensus peaks; Ho2017 DNase Emb, DHSs Emb, L1 arrest, DHSs L1 arrest; ATAC wt Emb, L1, L2, L3, L4, YA; Accessible sites; Annotated genes.

Figure 2—figure supplement 1 continued

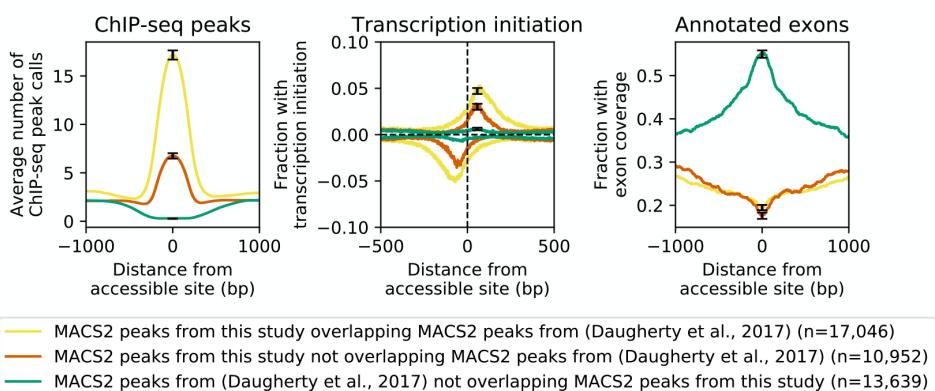
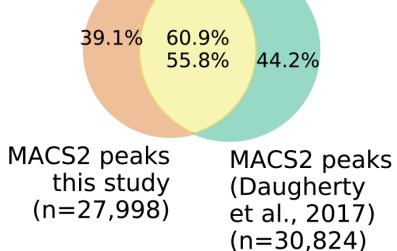
studies (**Daugherty et al., 2017; Ho et al., 2017**). (B) Comparison of accessible sites defined in this study to accessible sites defined in **Daugherty et al. (2017)**. (C) Comparison of accessible sites defined in this study to accessible sites defined in **Ho et al. (2017)**. (B,C) Leftmost plot shows overlaps between accessible sites. Rightmost three plots compare regions found in both studies or unique to only one study, for mean profile of modENCODE/modERN peak call pileup, fraction of sites with transcription initiation signal (negative values are reverse strand signals), and fraction overlapping an exon. (D,E) IGV screenshots (D: *nhr-25*, chrX:13,007,000–13,015,000, 8 kb; E: *kin-18*, chrIII:6,117,000–6,125,500, 8.5 kb) of stage-specific accessibility profiles and peak calls from **Daugherty et al. (2017)** (top, red), **(Ho et al., 2017)** (middle, green), and this study (bottom, blue).

DOI: <https://doi.org/10.7554/eLife.37344.008>

A



B



C

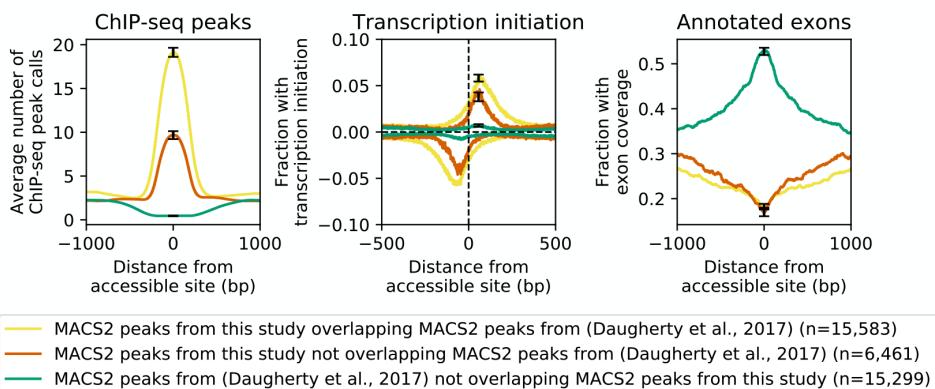
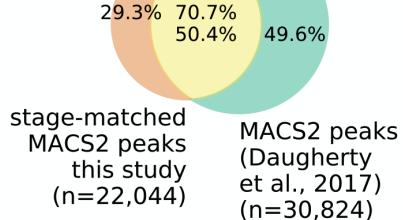


Figure 2—figure supplement 2. Effect of differences in peak calling methods on the types of identified accessible sites. As done by **Daugherty et al. (2017)**, MACS2 was used to call peaks on the ATAC-seq data reported here using MACS2 parameters –shift –75 –gsize ce -q 5e-2 –nomodel –extsize 150 –bdg –keep-dup all –call-summits –SPMR. Peaks from biological replicates of the same stage were combined by intersecting peak calls from the two biological replicates, and then peaks from each stage were combined by taking their union. This identified 27,998 peaks used for this analysis. (A) Comparison between accessible regions identified by the focal enrichment peak calling method used in this study ($n = 42,245$) to those defined using MACS2 ($n = 27,998$). (B) Comparison of ATAC-seq MACS2 peak calls from our data to ATAC-seq MACS2 peak calls from **(Daugherty et al., 2017)**. (C) Comparison of stage-matched ATAC-seq MACS2 peak calls from our data to ATAC-seq MACS2 peak calls from **Daugherty et al. (2017)**. (A–C) Leftmost plots show overlaps between accessible sites. Rightmost three plots compare regions found in both sets of peak calls or unique to only one. Figure 2—figure supplement 2 continued on next page

Figure 2—figure supplement 2 continued

set, for mean profile of modENCODE/modERN peak call pileup, fraction of sites with transcription initiation signal (negative values are reverse strand signals), and fraction overlapping an exon.

DOI: <https://doi.org/10.7554/eLife.37344.009>

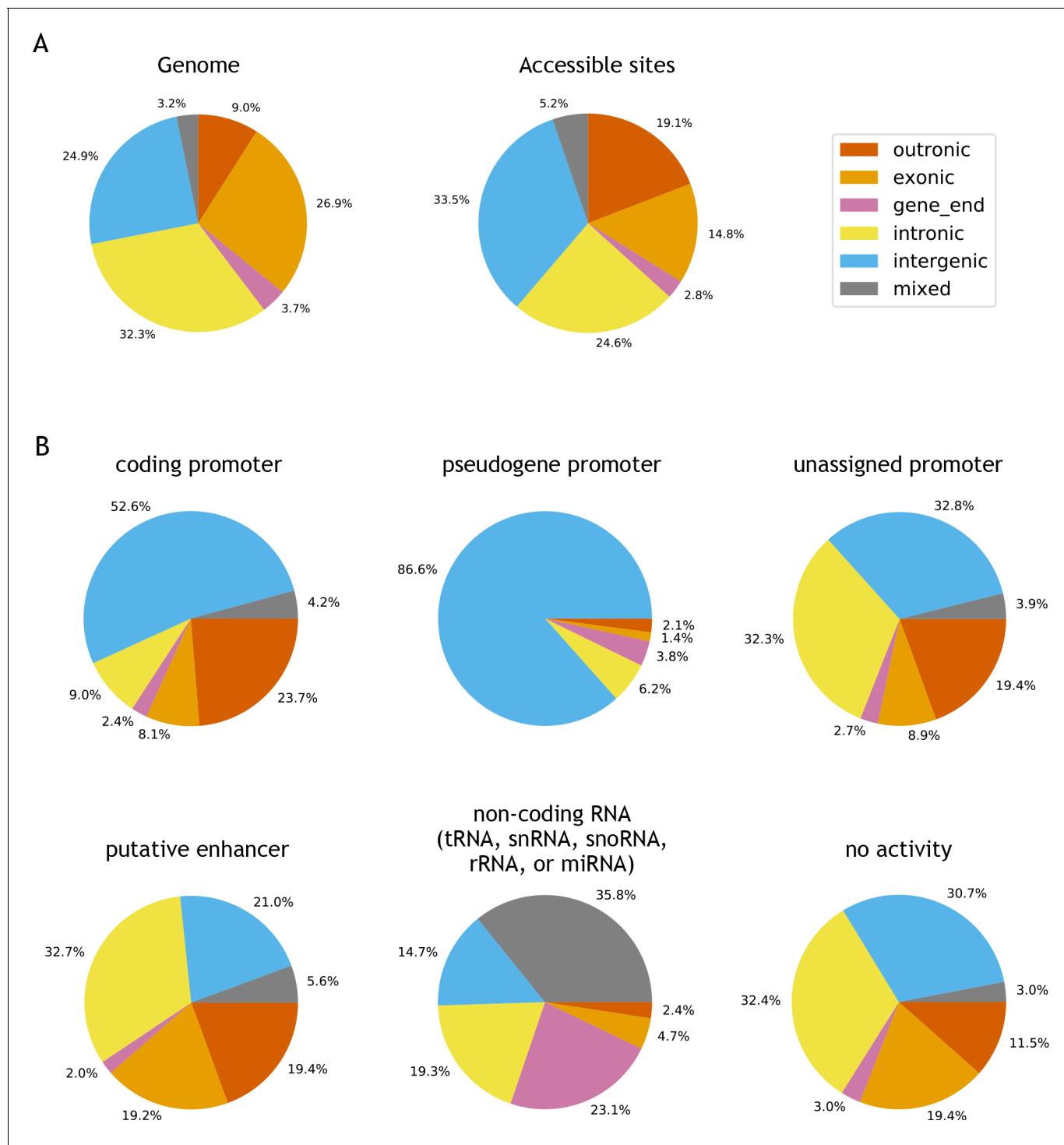


Figure 2—figure supplement 3. Genomic locations of accessible sites. (A) Left: distribution of bases in the *C. elegans* genome, partitioned into outronic, exonic, intronic, intergenic or mixed, based on the regulatory annotation. Right: distribution of genomic region type at accessible sites. (B) Distribution of genomic region at specific types of accessible sites.

DOI: <https://doi.org/10.7554/eLife.37344.010>

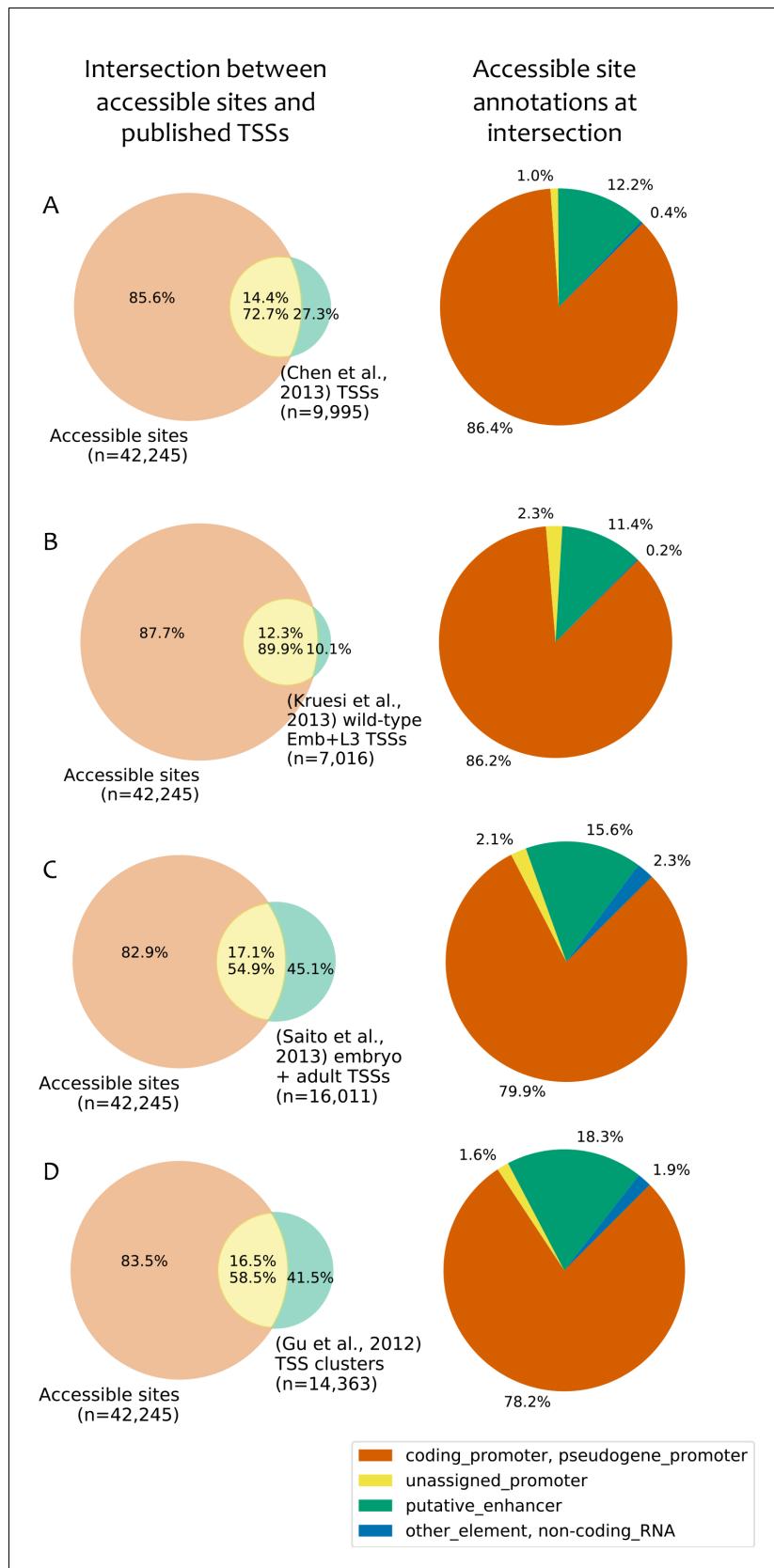


Figure 2—figure supplement 4. Comparison to published TSS maps. (A–D) Left: overlap between accessible sites and TSS annotations from (A) (Chen et al., 2013); (B) (Kruesi et al., 2013); (C) (Saito et al., 2013); (D) (Gu et al., 2012). Right: distribution of annotations at the intersection. Figure 2—figure supplement 4 continued on next page

Figure 2—figure supplement 4 continued

al., 2012). Right: accessible site annotations of elements that overlap a TSS in the indicated study. TSSs were considered to overlap an accessible site if they were located within 150 bp of peak accessibility. For **Gu et al. (2012)**, TSSs were clustered using a single-linkage approach using a distance threshold of 50 bp, and the overlaps are based on those clusters.

DOI: <https://doi.org/10.7554/eLife.37344.011>

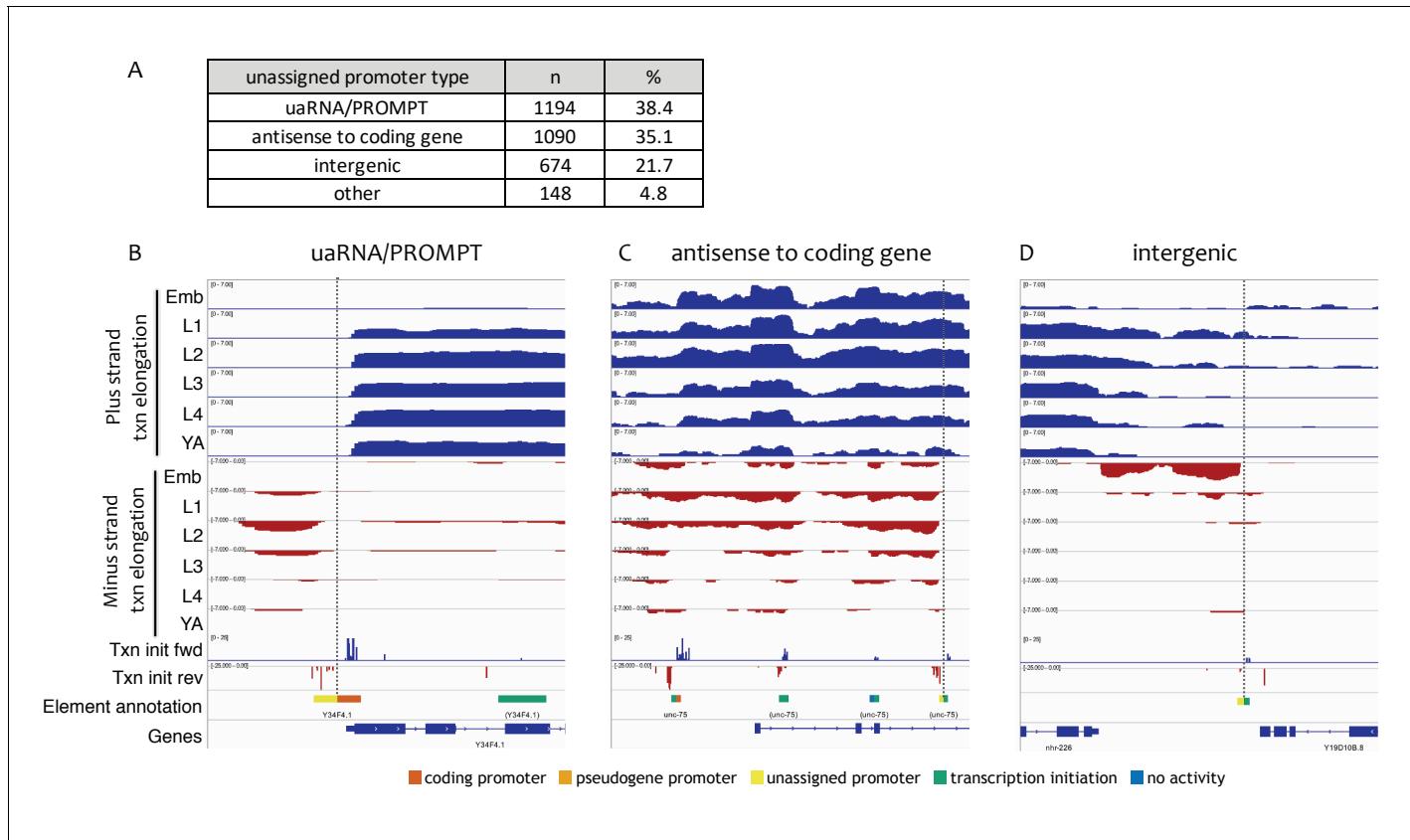


Figure 2—figure supplement 5. Types of unassigned promoters. (A) Types and numbers of unassigned promoters. (B–D) Examples of transcription patterns at unassigned promoters. Shown are forward and reverse strand nuclear RNA-seq signals to indicate genomic regions with transcription elongation, forward and reverse strand transcription initiation signal (pooled across stages), and accessible elements colored with left halves indicating reverse strand annotation and right halves indicating forward strand annotation. Vertical dotted lines highlight unassigned promoters. (B) uaRNA/PROMPT (chrIII:1,020,500–1,021,700, 1.2 kb), (C) antisense to coding gene (chrI:11,590,000–11,596,000, 6 kb), (D) intergenic (chrV:2,296,000–2,300,500, 4.5 kb).

DOI: <https://doi.org/10.7554/eLife.37344.012>

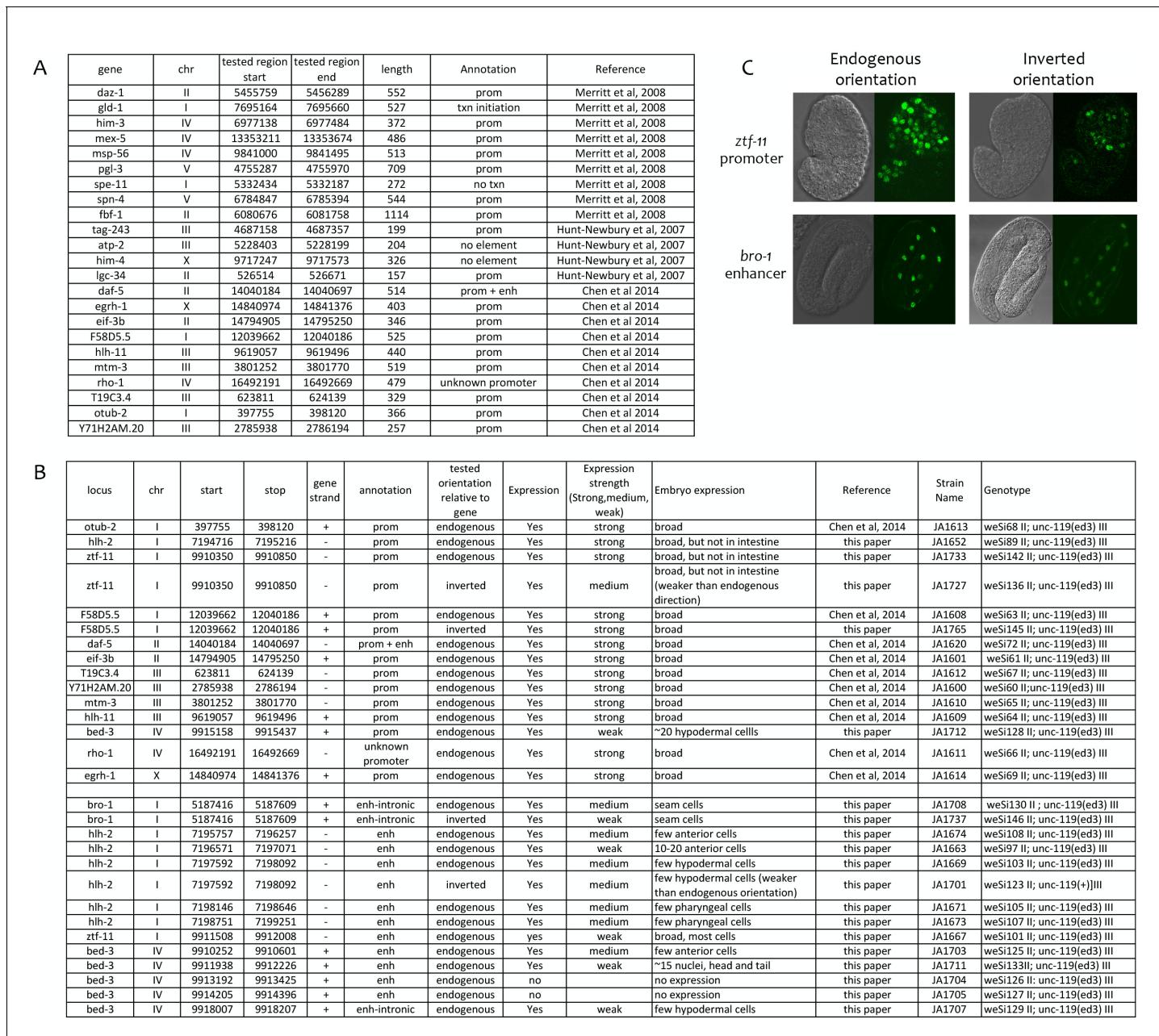


Figure 2—figure supplement 6. Transgenic tests of annotated promoters and enhancers for promoter activity. (A) Comparison of annotations to 23 elements previously shown to function as promoters in transgenic assays ([Merritt et al., 2008](#); [Hunt-Newbury et al., 2007](#); [Chen et al., 2014](#)). (B) Indicated elements were fused to *his-58::gfp* (see Materials and methods) and the resulting transgenic strains tested for GFP expression in embryos. Elements were cloned in the endogenous orientation relative to their associated gene or in inverted orientation, as indicated. In expression strength column, 'strong' and 'medium' indicate high and low level of GFP visible in live embryos; 'weak' indicates expression only visible by immunofluorescence. (C) Examples of transgene expression. Shown is expression driven by the ztf-11 promoter and the bro-1 enhancer in both orientations; DIC image on left, HIS-58-GFP on right.

DOI: <https://doi.org/10.7554/eLife.37344.013>

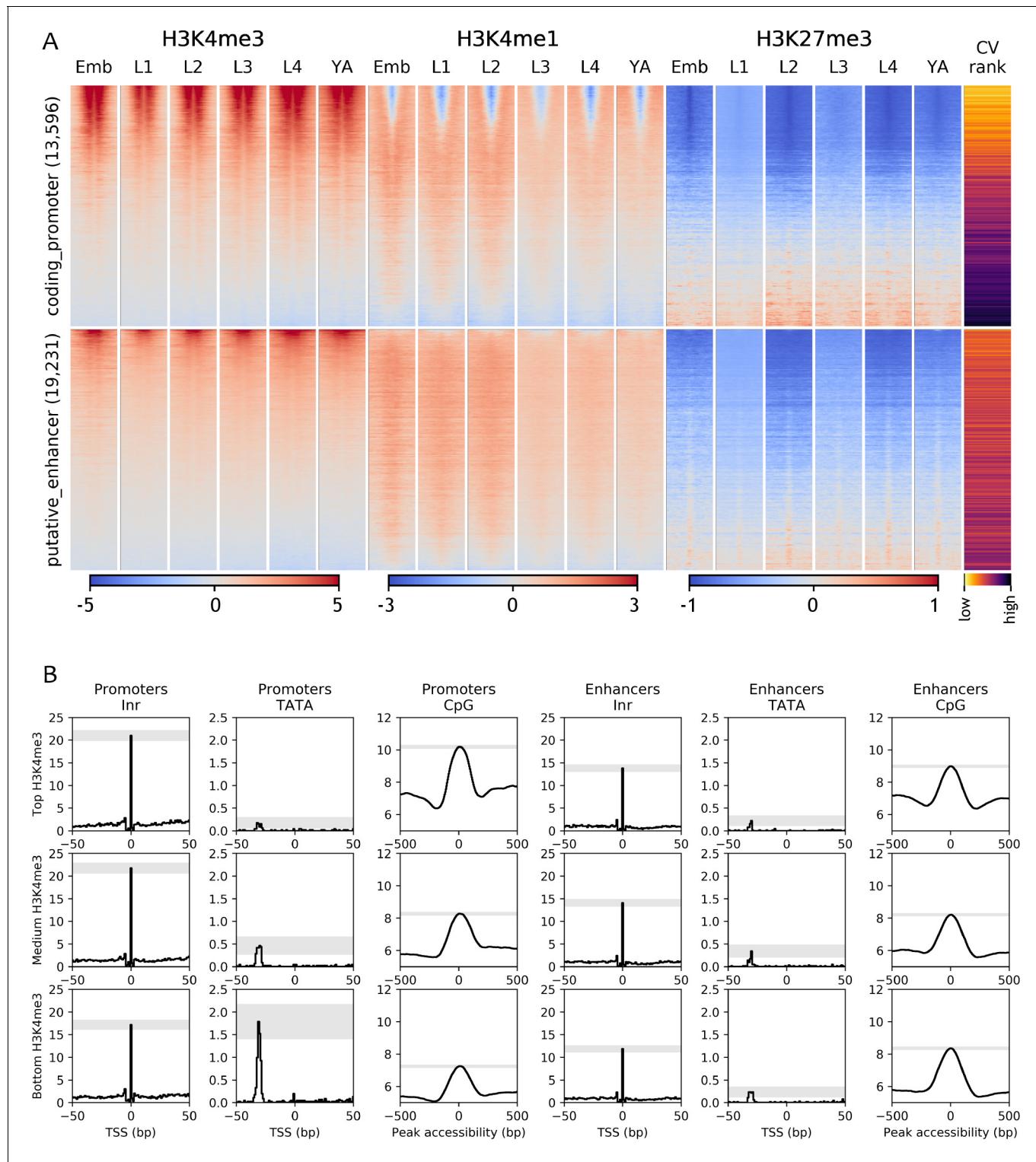


Figure 3. Chromatin state and sequence features of promoters and enhancers. (A) Heatmaps of indicated histone modifications and CV values at coding promoters (top), and enhancers (bottom), aligned at element midpoints. Elements are ranked by mean H3K4me3 levels. Low CV values indicate broad expression across development and cell types and high CV values indicate regulated expression. Promoters of genes with low CV values have high H3K4me3 levels. (B) Distribution of initiator Inr motif, TATA motif, and CpG content at coding promoters and enhancers, separated by H3K4me3 level (top, middle, and bottom thirds). Grey-shaded regions represent 95% confidence intervals of the sample mean at the genomic position with the highest signal.

DOI: <https://doi.org/10.7554/eLife.37344.015>

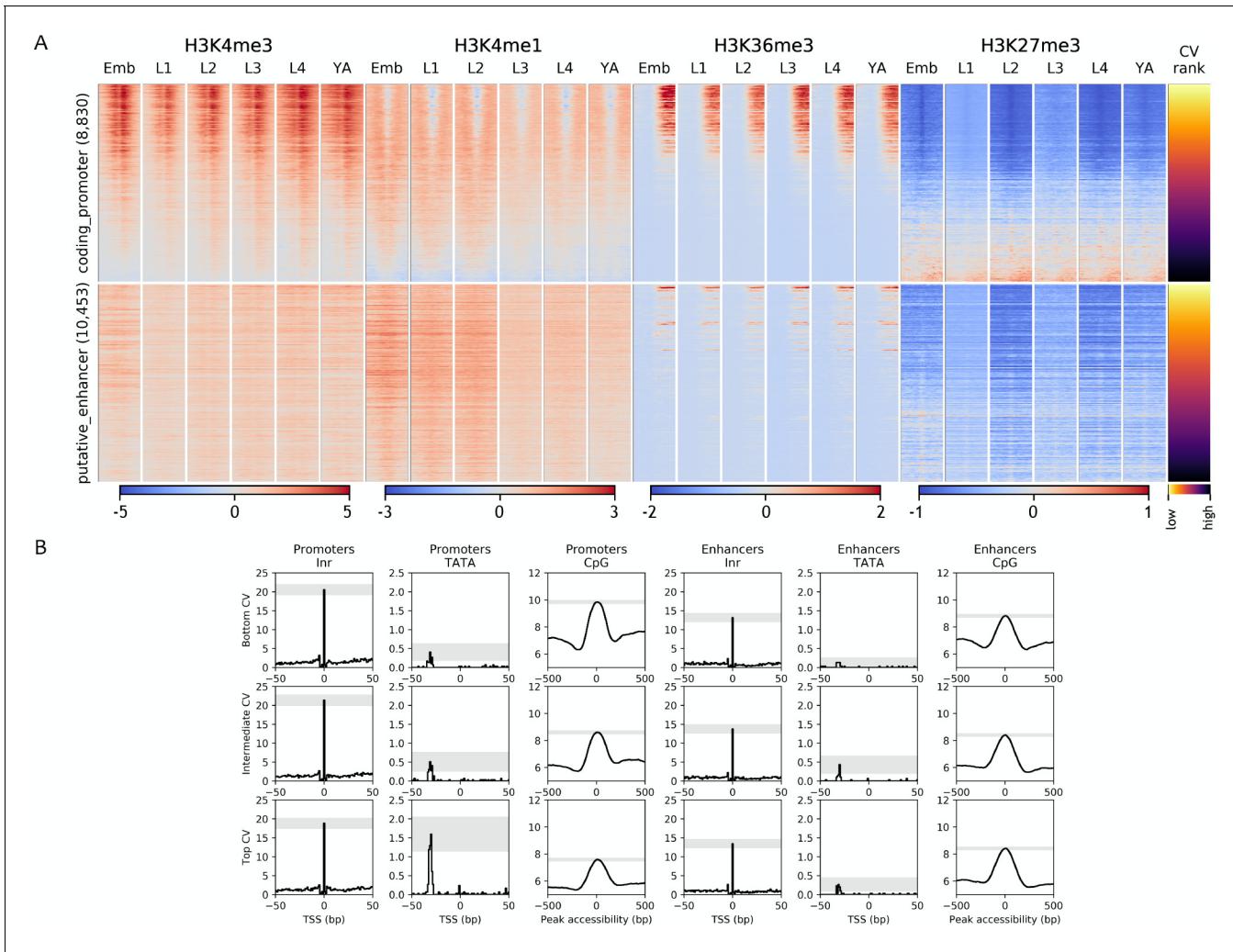


Figure 3—figure supplement 1. Chromatin state and sequence features of promoters and enhancers sorted by CV value. (A) Heatmaps of indicated histone modifications and CV values at coding promoters (top), and enhancers (bottom). Elements are ranked by CV value. Low CV values indicate broad expression across development and cell types and high CV values indicate regulated expression. H3K4me3, H3K4me1, and H3K27me3 signals are aligned at element midpoints; H3K36me3 is aligned at the start of the associated gene annotation. (B) Distribution of initiator Inr motif, TATA motif, and CpG content at coding promoters and enhancers, separated by CV value (top (high CV), middle, and bottom (low CV)). Grey-shaded regions represent 95% confidence intervals of the sample mean at the genomic position with the highest signal.

DOI: <https://doi.org/10.7554/eLife.37344.016>

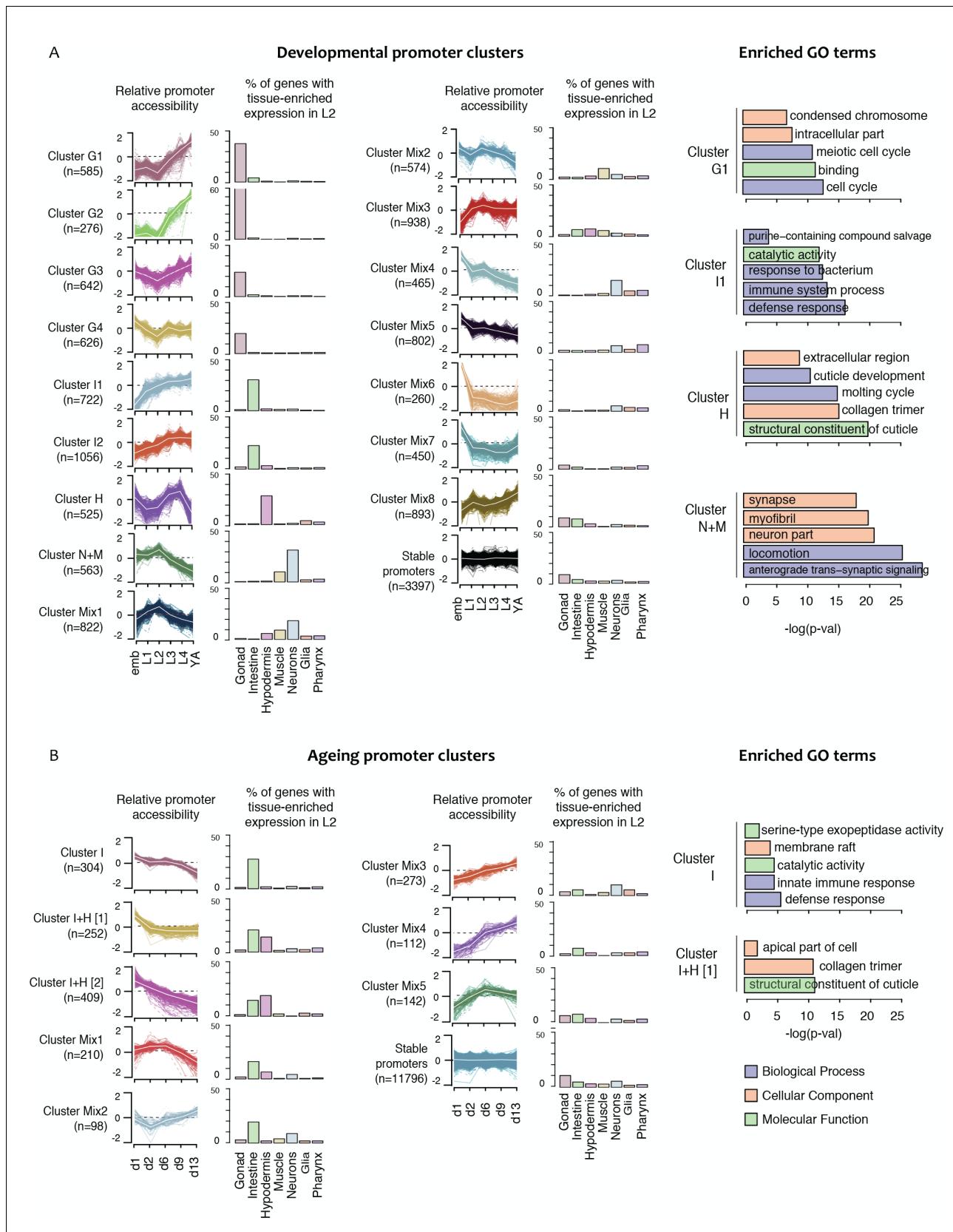


Figure 4. Shared dynamics of promoter accessibility in development and ageing. Clusters of promoters with shared relative accessibility patterns across (A) development or (B) ageing. Relative promoter accessibility is \log_2 of the depth-normalized ATAC-seq coverage at a given time point divided by the

Figure 4 continued

mean ATAC-seq coverage across the time series (see Materials and methods). The percentage of associated genes that have enriched expression in the indicated tissues was determined from single-cell L2 larval RNA-seq data (**Cao et al., 2017**; see Materials and methods). Right hand panels show examples of GO terms enriched in genes associated with development or ageing clusters.

DOI: <https://doi.org/10.7554/eLife.37344.017>

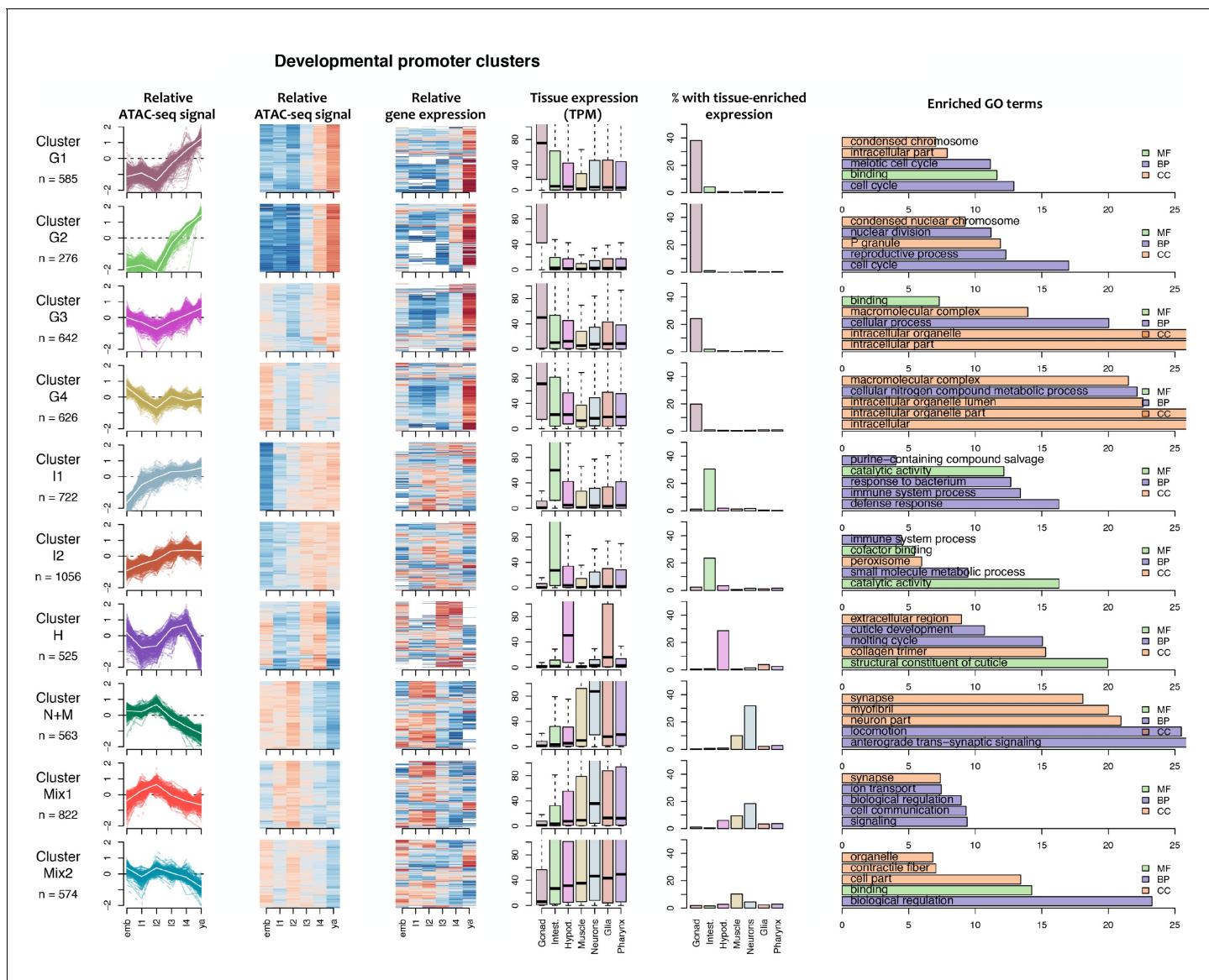


Figure 4—figure supplement 1. Characteristics of developmental promoter clusters (continued in **Figure 4—figure supplement 2**). Characteristics of promoter clusters with shared accessibility patterns across development. Relative promoter ATAC-seq coverage is shown across the time series as a graph (each line representing a promoter) and a heatmap (each row representing a promoter). Values are scaled from -2 (dark blue) to 0 (white) to +2 (dark red). Heatmap showing relative expression of the associated genes across the time series, using the same color scale. Tissue expression box plots show TPMs of clustered genes in individual tissues (data from Cao et al., 2017). Percentage of genes with tissue-enriched expression shows the percentage of genes within the cluster with enriched expression in the indicated tissues. Enriched GO terms show the top five enriched GO terms obtained for each cluster from the corresponding list of genes using gProfiler. MF = Molecular Function, CC = Cellular Component, BP = Biological Process.

DOI: <https://doi.org/10.7554/eLife.37344.018>

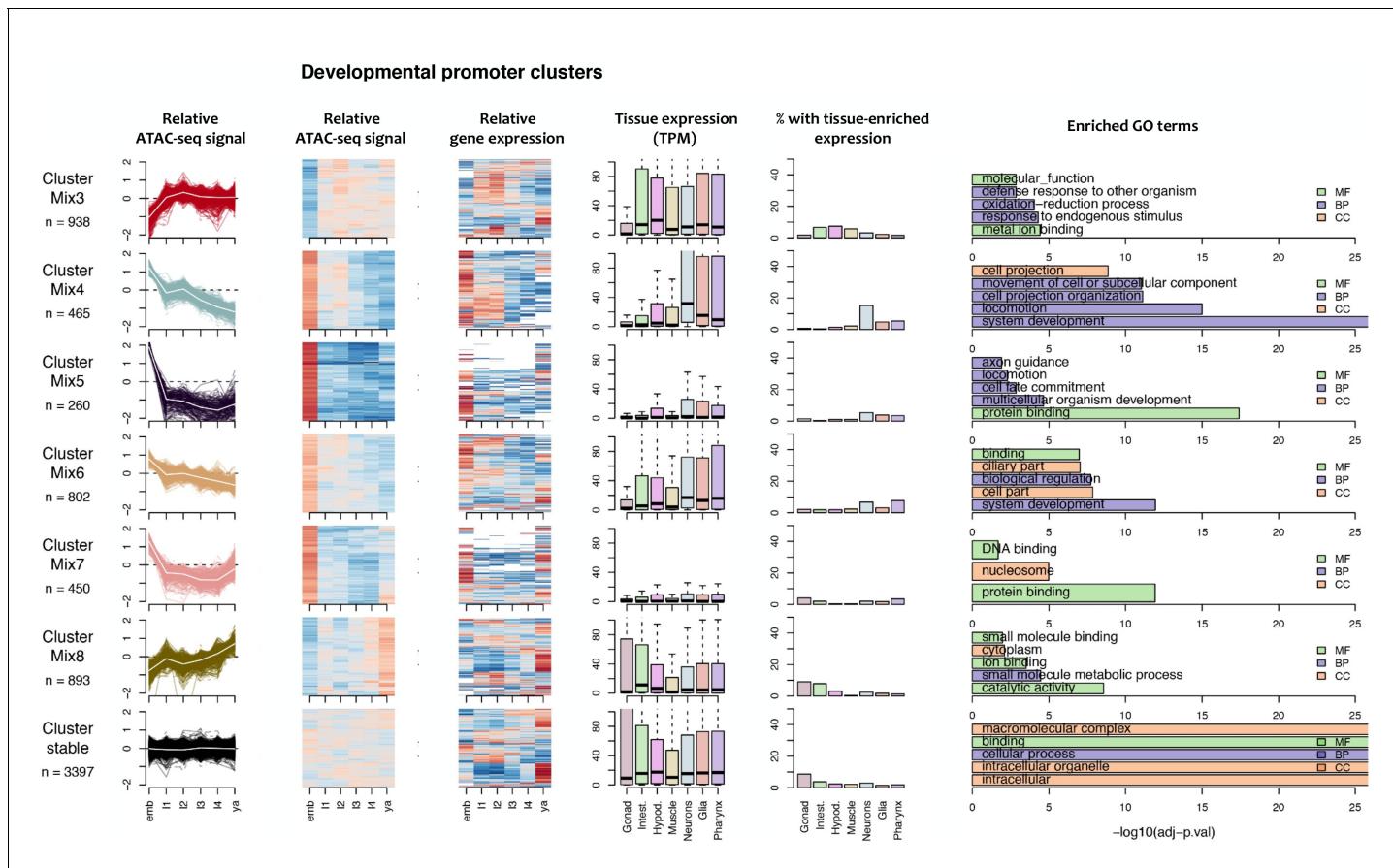


Figure 4—figure supplement 2. Characteristics of developmental promoter clusters (continued from **Figure 4—figure supplement 1**). Characteristics of promoter clusters with shared accessibility patterns across development. Relative promoter ATAC-seq coverage is shown across the time series as a graph (each line representing a promoter) and a heatmap (each row representing a promoter). Values are scaled from -2 (dark blue) to 0 (white) to +2 (dark red). Heatmap showing relative expression of the associated genes across the time series, using the same color scale. Tissue expression box plots show TPMs of clustered genes in individual tissues (data from **Cao et al., 2017**). Percentage of genes with tissue-enriched expression shows the percentage of genes within the cluster with enriched expression in the indicated tissues. Enriched GO terms show the top five enriched GO terms obtained for each cluster from the corresponding list of genes using gProfiler. MF = Molecular Function, CC = Cellular Component, BP = Biological Process.

DOI: <https://doi.org/10.7554/eLife.37344.019>

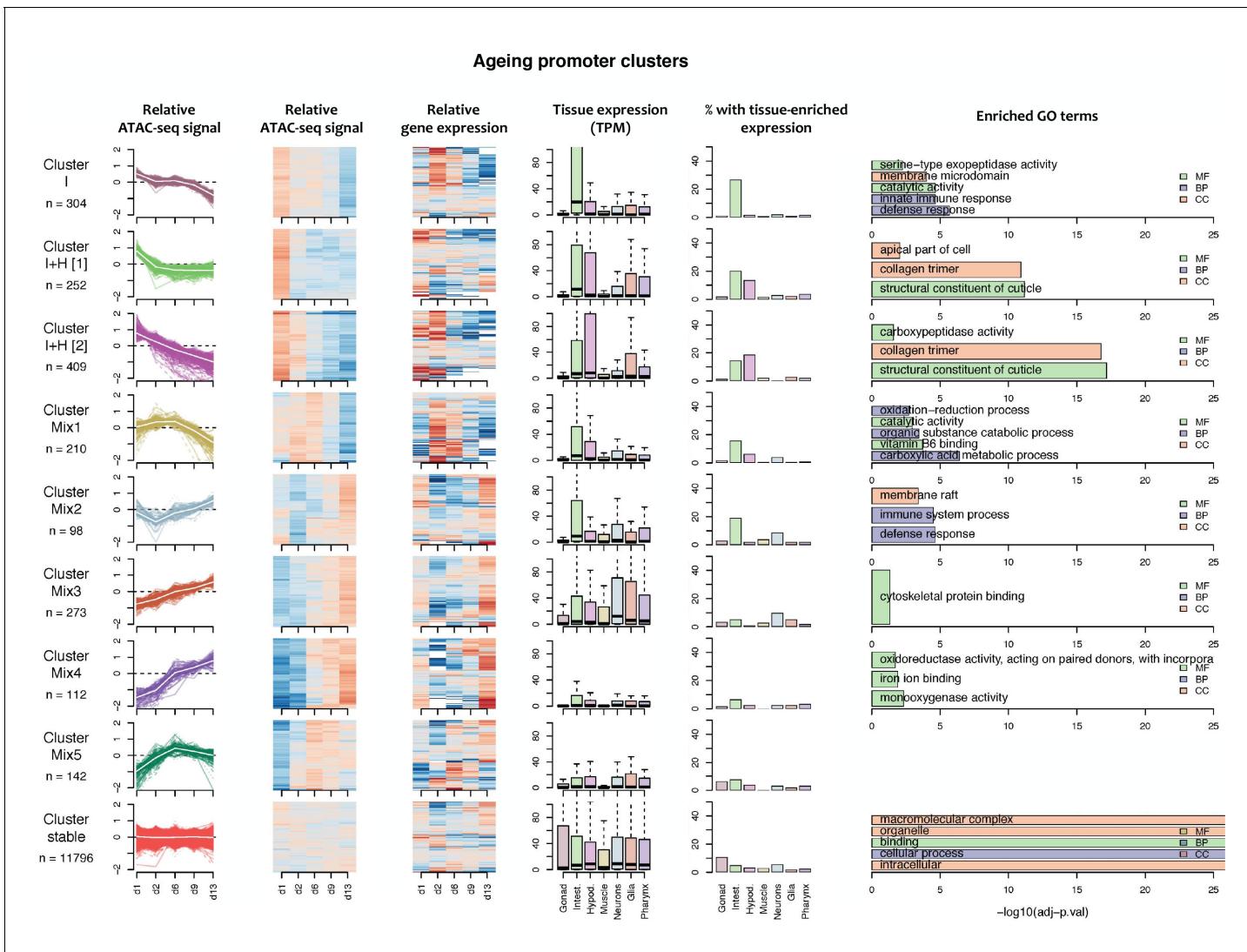
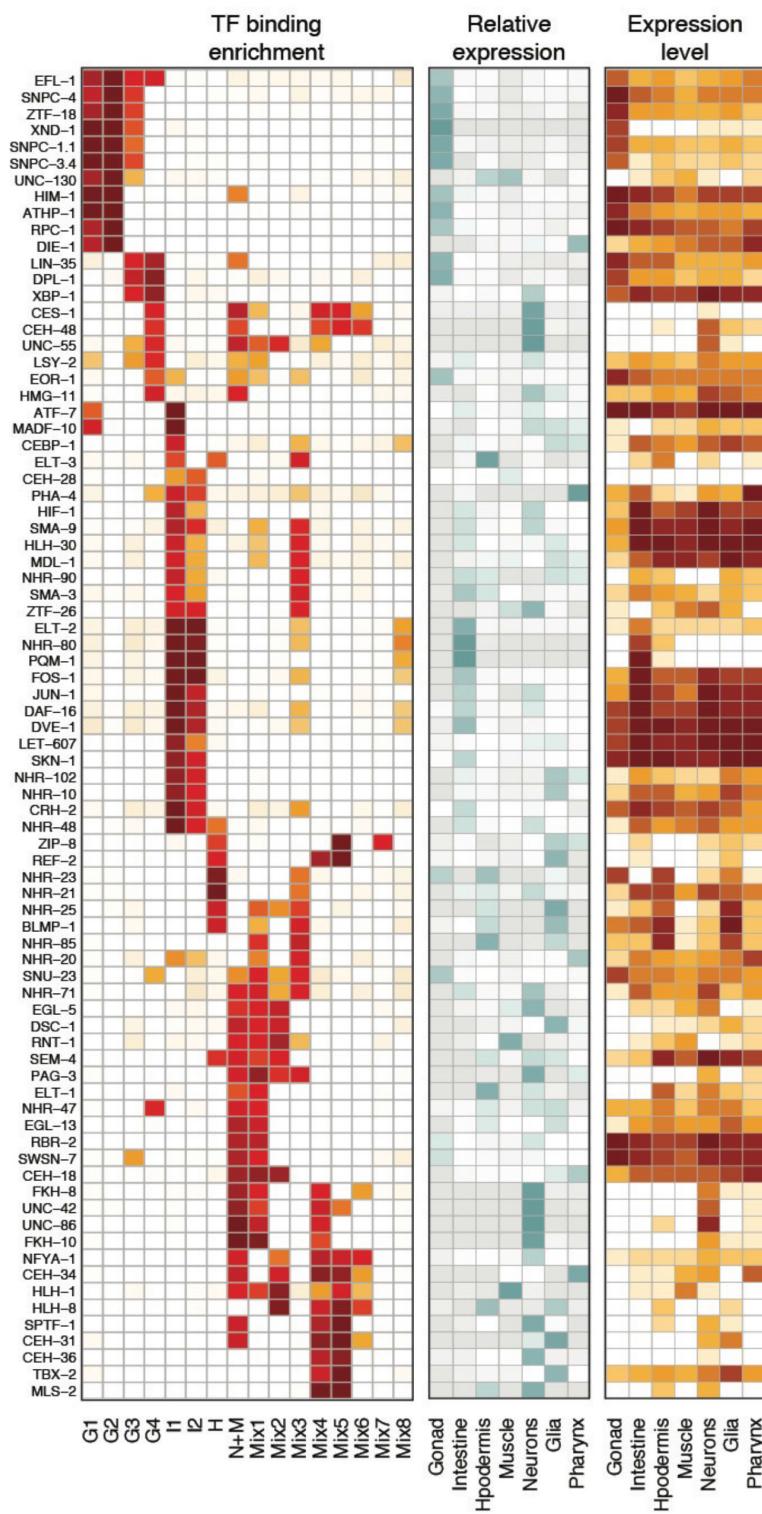


Figure 4—figure supplement 3. Characteristics of ageing promoter clusters. Characteristics of promoter clusters with shared accessibility patterns across ageing. Relative promoter ATAC-seq coverage is shown across the time series as a graph (each line representing a promoter) and a heatmap (each row representing a promoter). Values are scaled from -2 (dark blue) to 0 (white) to +2 (dark red). Heatmap showing relative expression of the associated genes across the time series, using the same color scale. Tissue expression box plots show TPMs of clustered genes in individual tissues (data from Cao et al., 2017). Percentage of genes with tissue-enriched expression shows the percentage of genes within the cluster with enriched expression in the indicated tissues. Enriched GO terms show the top five enriched GO terms obtained for each cluster from the corresponding list of genes using gProfiler. MF = Molecular Function, CC = Cellular Component, BP = Biological Process. Note that Cluster Mix5 does not have any enriched GO terms.

DOI: <https://doi.org/10.7554/eLife.37344.020>

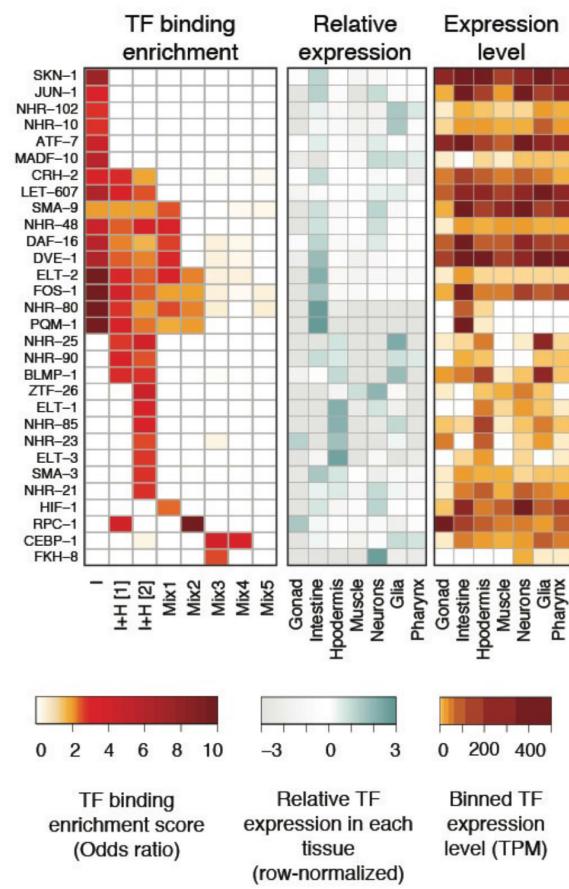
A

Developmental promoter clusters



B

Ageing promoter clusters



TF binding enrichment score (Odds ratio)

Relative TF expression in each tissue (row-normalized)

Binned TF expression level (TPM)

Figure 5. Transcription factor binding enrichment in developmental and ageing promoter clusters. Transcription factor (TF) binding enrichments in developmental (A) or ageing (B) promoter clusters from **Figure 4**. TF-binding data are from modENCODE/modERN (Araya et al., 2014; Kudron et al., 2014). See also **Figure 5 continued on next page**.

Figure 5 continued

al., 2018); peaks in HOT regions were excluded (see Materials and methods). Only TFs enriched more than twofold in at least one cluster are shown, and only enrichments with a $p < 0.01$ (Fisher's exact test) are shown. Plots show TF binding enrichment odds ratio (left), expression of the TF in each tissue relative to its expression across all tissues ($\log_2(\text{TF tissue TPM}/\text{mean of the TF's TPMs across all tissues})$), middle), and the decile of expression of the TF in each tissue (right; TPMs < 1 are not taken into account when calculating TPMs deciles). Expression data are from Cao et al. (2017). Legends for Figure Supplements.

DOI: <https://doi.org/10.7554/eLife.37344.022>