Chromatin accessibility is dynamically regulated across *C. elegans* development and ageing

Jürgen Jänes[1,^], Yan Dong[1,^], Michael Schoof[1,§], Jacques Serizay[1,§], Alex Appert[1], Chiara Cerrato[1], Carson Woodbury[1], Ron Chen[1,3], Carolina Gemma[1,4], Ni Huang[1], Djem Kissiov[1,5], Przemysław Stempor[1], Annette Steward[1], Eva Zeiser[1], Sasha Sauer[2], and Julie Ahringer[1,*]

[^]equal contribution, [§]equal contribution

*author for correspondence (j.ahringer@gurdon.cam.ac.uk)

[1]The Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge UK

[2]Max Delbrück Center for Molecular Medicine (BIMSB and BIH), Robert-Rössle-Strasse 10, 13092 Berlin, Germany; Max Planck Institute for Molecular Genetics, Otto-Warburg Laboratories, Ihnestrasse 63-73, 14195 Berlin, Germany

[3]current address: School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

[4]current address: Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital Campus, London, UK.

[5]current address: University of California, Berkeley, 142 Life Sciences Addition # 3200, Berkeley, CA 94720-3200, USA

Abstract

An essential step for understanding the transcriptional circuits that control development and physiology is the global identification and characterization of regulatory elements. Here we present the first map of regulatory elements across the development and ageing of an animal, identifying 42,245 elements accessible in at least one *C. elegans* stage. Based on nuclear transcription profiles, we define 15,714 protein-coding promoters and 19,231 putative enhancers, and find that both types of element can drive orientation-independent transcription. Additionally, hundreds of promoters produce transcripts antisense to protein coding genes, suggesting involvement in a widespread regulatory mechanism. We find that the accessibility of most elements is regulated during development and/or ageing and that patterns of accessibility change are linked to specific developmental or physiological processes. The map and characterization of regulatory elements across *C. elegans* life provides a platform for understanding how transcription controls development and ageing.

Introduction

The genome encodes the information for organismal life. Because the deployment of genomic information depends in large part on regulatory elements such as promoters and enhancers, their identification and characterization is essential for understanding genome function and its regulation.

Regulatory elements are typically depleted for nucleosomes, which facilitates their identification using sensitivity to digestion by nucleases such as DNase I or Tn5 transposase, termed DNA accessibility (Sabo et al. 2006; Crawford et al. 2006; Buenrostro et al. 2013). In different organisms, large repertoires of regulatory elements have been determined by profiling DNA accessibility genome-wide in different cell types and developmental stages (Thomas et al. 2011; Kharchenko et al. 2011; Thurman et al. 2012; Yue et al. 2014; Roadmap Epigenomics Consortium et al. 2015; Daugherty et al. 2017; Ho et al. 2017). However, no study has yet investigated regulatory element usage across the life of an animal, from the embryo to the end of life. Such information is important, because different transcriptional programs operate in different periods of life and ageing. *C. elegans* is ideal for addressing this question, as it has a simple anatomy, well defined cell types, and short development and lifespan. A map of regulatory elements and their temporal dynamics would facilitate understanding of the genetic control of organismal life.

Active regulatory elements have previously been shown to have different transcriptional outputs and chromatin modifications (Andersson 2015; Kim and Shiekhattar 2015). Transcription is initiated at both promoters and enhancers, with most elements having divergent initiation events from two independent sites (Core et al. 2008;

Kim et al. 2010; De Santa et al. 2010; Koch et al. 2011; Chen et al. 2013). However, promoters and enhancers differ in the production of stable transcripts. At protein-coding promoters, productive transcription elongation produces a stable transcript, whereas enhancers and the upstream divergent initiation from promoters generally produce short, aborted, unstable transcripts (Core et al. 2014; Andersson et al. 2014; Rennie et al. 2017).

Promoters and enhancers have also been shown to be differently enriched for specific patterns of histone modifications. In particular, promoters often have high levels of H3K4me3 and low levels of H3K4me1, whereas enhancers tend to have the opposite pattern of higher H3K4me1 and lower H3K4me3 (Heintzman et al. 2007, 2009). However, in human and Drosophila cell lines it was observed that H3K4me3 and H3K4me1 levels correlate with levels of transcription at regulatory elements, rather than whether the element acts as a promoter or an enhancer (Core et al. 2014; Henriques et al. 2018; Rennie et al. 2018). Further, analyses of genes that are highly regulated in development showed that their promoters lacked chromatin marks associated with activity (including H3K4me3), even when the associated genes are actively transcribed (Zhang et al. 2014; Pérez-Lluch et al. 2015). Therefore, stable elongating transcription, rather than histone modification patterns, appears to be the defining feature that distinguishes active promoters from active enhancers (reviewed in Andersson 2015; Andersson et al. 2015; Kim and Shiekhattar 2015; Henriques et al. 2018; Rennie et al. 2018).

Regulatory elements have not been systematically mapped and annotated in *C. elegans*. Promoter identification has been hampered because the 5' ends of ~70% of protein-coding transcripts are trans-spliced to a 22nt leader sequence (Allen et al. 2011).

Because the region from the transcription initiation site to the trans-splice site (the "outron") is removed and degraded, the 5' end of the mature mRNA does not mark the transcription start site. To overcome this difficulty, previous studies identified transcription start sites for some genes through profiling transcription initiation and elongation in nuclear RNA or by inhibiting *trans*-splicing at a subset of stages (Gu et al. 2012; Chen et al. 2013; Kruesi et al. 2013; Saito et al. 2013). In addition, two recent studies used ATAC-seq or DNAse I hypersensitivity to map regions of accessible chromatin in some developmental stages, and predicted element function by proximity to first exons or chromatin state (Daugherty et al. 2017; Ho et al. 2017).

Towards building a comprehensive map of regulatory elements and their use during the life of an animal, here we used multiple assays to systematically identify and annotate accessible chromatin in the six *C. elegans* developmental stages and at five time points of adult ageing. Strikingly, most elements undergo a significant change in accessibility during development and/or ageing. Clustering the patterns of accessibility changes in promoters reveals groups that act in shared processes. This map makes a major step towards defining regulatory element use during *C. elegans* life.

Results and Discussion

Defining and annotating regions of accessible DNA

To define and characterize regulatory elements across *C. elegans* life, we collected biological replicate samples from a developmental time course and an ageing time course (Figure 1A). The developmental time course consisted of wild-type samples from each of the six developmental stages (embryos, four larval stages, and young adults). For the ageing time course, we used *glp-1(e2144ts)* mutants to prevent progeny production, as they lack germ cells at the restrictive temperature. Five adult ageing time points were collected, starting from the young adult stage (day 1) and ending at day 13, just before the major wave of death.

Figure 1A outlines the datasets generated. For all developmental and ageing time points, we used ATAC-seq to identify accessible regions of DNA. We also sequenced strand-specific nuclear RNA (>200nt long) to determine regions of transcriptional elongation, because previous work demonstrated that this approach could capture outron signal linking promoters to annotated exons (Chen et al. 2013; Kruesi et al. 2013; Saito et al. 2013). For the development time course, we additionally sequenced short (<100nt) capped nuclear RNA to profile transcription initiation, profiled four histone modifications to characterize chromatin state (H3K4me3, H3K4me1, H3K36me3, and H3K27me3), and performed a DNase I concentration course to investigate the relative accessibility of elements. Micrococcal nuclease (MNase) data were also collected for the embryo stage. As previously noted by others, we found that the ATAC-seq accessibility signal is similar to that observed using a low concentration DNase I or MNase, and that

the ATAC-seq data has the highest signal to noise ratio (Buenrostro et al. 2013; Figure 1 - figure supplement 1A).

To define sites that are accessible in at least one developmental or ageing stage, focal peaks of significant ATAC-seq enrichment were identified across all developmental and ageing samples, yielding 42,245 individual elements (Figure 1B, Figure 1 - source data 1; see Methods for details). Of these, 72.8% overlap a transcription factor binding site (TFBS) mapped by the modENCODE or modERN projects (Araya et al. 2014; Kudron et al. 2017), supporting their potential regulatory functions (Figure 2 - figure supplement 1A).

Two recent studies reported accessible regions in *C. elegans* identified using DNase I hypersensitivity or ATAC-seq (Ho et al. 2017; Daugherty et al. 2017). The 42,245 accessible elements defined here overlap 33.7% of (Ho et al. 2017) DNase I hypersensitive sites and 47.9% of (Daugherty et al. 2017) ATAC-seq peaks (Figure 2 - figure supplement 1B,C). Examining the non-overlapping sites from pairwise comparisons, it appears that differences in peak calling methods account for some of the differences. Accessible regions determined here required a focal peak of enrichment whereas the other studies found both focal sites and broad regions with increased signal. Consistent with these differences in methods, sites reported in the two studies but not identified here are enriched for exonic chromatin, depleted for both TFBS and transcription initiation sites, and often found in broad regions of increased accessibility across transcriptionally active gene bodies (Figure 2 - figure supplement 1B-E).

To functionally classify elements, we annotated each of the 42,245 elements for transcription initiation and transcription elongation signals on both strands (Figure 2A;

Figure 2 - source data 1; see Methods for details). Overall, 37.1% of elements had evidence of promoter activity, indicated by an increase in transcription elongation signal in at least one stage and one direction, and 82.3% of elements had transcription initiation signal. These patterns of nuclear transcription were used together with gene and ncRNA annotations to functionally separate the accessible elements into six classes: protein-coding promoter, pseudogene promoter, unknown promoter, putative enhancer, ncRNA (tRNA, snoRNA, rRNA, or miRNA), or other (Figure 2B). Elements were defined as promoters where there was a significant increase in transcription elongation signal originating at the element. Promoters were assigned to protein-coding or pseudogenes if there was continuous signal extending from the element to an annotated first exon (covering the outron). Promoters were annotated as “unknown” if transcription elongation signal was not linked to an annotated gene. Elements were annotated as putative enhancers where there was transcription initiation signal but no significant transcription elongation signal (hereafter referred to as “enhancers”). Elements were assigned to the ncRNA class if they overlapped an annotated tRNA, snoRNA, rRNA, or miRNA. Finally, elements with no transcriptional activity were annotated as “other”. Overall, accessible sites are enriched for being located within outrons or intergenic regions (Figure 2 - figure supplement 2A).

Of the 42,245 elements, 13,596 were defined as protein-coding promoters: 11,478 elements are unidirectional promoters and 2,118 are divergent promoters that drive expression of two oppositely oriented protein-coding genes (Figure 2 - source data 1). In total, promoters were defined for 11,196 protein-coding genes, with 3,000 genes having >1 promoter (Figure 2C). The protein-coding promoter annotations show good overlap with four sets of TSSs previously defined based on mapping transcription (Chen

et al. 2013; Kruesi et al. 2013; Saito et al. 2013; Gu et al. 2012; 76.8%–85.1%; Figure 2 - figure supplement 5). A further 19,231 elements were defined as enhancers. Enhancers were assigned to a gene if they are located within the region from its most upstream promoter to its gene end, with the rest left unassigned; 6,668 genes have at least one associated enhancer, and 3,240 genes have >1 enhancer (Figure 2C).

The locations of unknown promoters suggest different potential functions. A large fraction (34.9%) generate antisense transcripts within the body of a protein coding gene, suggesting a possible role in regulating expression of the associated gene (Figure 2 - figure supplement 4). Another large group (41.1%) produce antisense transcripts from an element that is a protein coding promoter in the sense direction, a pattern seen in many mammalian promoters (Figure 2 - figure supplement 4; Preker et al. 2008; Flynn et al. 2011; Sigova et al. 2013). Most of the rest (19.1%) are intergenic and may define promoters for unannotated transcripts.

Patterns of histone marks at promoters and enhancers

Promoters and enhancers show general differences in patterns of histone modifications, such as higher levels of H3K4me3 at promoters or H3K4me1 at enhancers, and chromatin states are frequently used to define elements as promoters or enhancers (Heintzman et al. 2007; Ernst and Kellis 2010; Ernst et al. 2011; Kharchenko et al. 2011; Hoffman et al. 2013; Daugherty et al. 2017). However it has been shown that H3K4me3 levels correlate with transcriptional activity rather than with function (Pekowska et al. 2011; Core et al. 2014; Andersson et al. 2014; Henriques et al. 2018; Rennie et al.

2018), suggesting that defining regulatory elements solely based on chromatin state is likely to lead to incorrect annotations.

To further investigate the relationship between chromatin marking and element function, we mapped four histone modifications at each developmental stage (H3K4me3, H3K4me1, H3K27me3, H3K36me3) and examined their patterns around coding promoters and enhancers. As expected, many coding promoters had high levels of H3K4me3 and were depleted for H3K4me1 (Figure 3A). Moreover, enhancers had generally low levels of H3K4me3 and higher levels of H3K4me1 than promoters (Figure 3A). However many elements did not have these patterns. For example, about 50% of coding promoters have a high level of H3K4me1 and no or low H3K4me3 marking (Figure 3A).

To investigate the nature of these patterns, we examined coefficients of variation of gene expression (CV; Gerstein et al. 2014) of the associated genes. Genes with broad stable expression across cell types and development, such as housekeeping genes, have low gene expression variation and hence a low CV value. In contrast, genes with regulated expression, such as those expressed only in particular stages or cell types have a high CV value. We found a strong inverse correlation between a gene's CV value and its promoter H3K4me3 level (-0.64, Spearman's rank correlation; Figure 3; Figure 3 - figure supplement 1A). Furthermore, promoters with low or no H3K4me3 marking are enriched for H3K27me3 (Figure 3; Figure 3 - figure supplement 1A), which is associated with regulated gene expression (Tittel-Elmer et al. 2010; Pérez-Lluch et al. 2015; Evans et al. 2016). These results support the view that H3K4me3 marking may be a specific feature of promoters with broad stable activity, consistent with the finding that active promoters of regulated genes lack H3K4me3 (Pérez-Lluch et al. 2015). The

profiling here was done in whole animals, which may have precluded detecting modifications occurring in a small number of nuclei. Nevertheless, the results indicate that chromatin state alone is not a reliable metric for element annotation. Histone modification patterns at many promoters resemble those at enhancers, and vice versa.

Promoters and enhancers also share sequence features. Both are enriched for initiator INR elements, although enhancers have a slightly lower INR frequency (Figure 3B). Promoters and enhancers show a similar level of enrichment for CpG dinucleotides, with the exception of promoters with high H3K4me3 and low CV values (broadly expressed genes), which have higher CpG content than other elements (Figure 3B and Figure 3 - figure supplement 1B). As expected from other studies, promoters also differ from enhancers by the presence of TATA motifs, which occur predominantly in genes with low H3K4me3 and high CV values (i.e., with regulated expression; Figure 3B and Figure 3 - figure supplement 1B).

Promoters and enhancers can drive gene expression in an orientation independent manner

To validate the promoter annotations, we compared them with studies where small regions of DNA had been defined as promoters using transgenic assays. These comprised ten regions defined based on transcription initiation signal (Chen et al. 2014), nine regions defined based on proximity to a germ line gene (Merritt et al. 2008), and four defined by proximity to the first exon of a muscle expressed gene (Hunt-Newbury et al. 2007). Of these 23 regions, 21 overlap an element in our set of accessible sites, 19 of which are annotated as protein coding promoters (Figure 2 - Figure supplement 5A). One of the remaining two is annotated as an enhancer and the other overlaps an

accessible element for which no transcriptional signal was detected. We further directly tested three elements annotated as promoters (for *hlh-2, ztf-11* and *bed-3* genes), and found that all three drove robust expression of a histone-GFP reporter (Figure 2 - figure supplement 5A). Overall, there is good concordance between promoter annotation and promoter activity.

Most of the elements annotated as protein-coding promoters are flanked by bidirectional transcription initiation signal (74.0%), similar to the pattern seen in mammals. Most (82.6%) are unidirectional promoters, producing a protein-coding transcript in one direction, but no stable transcript from the upstream initiation site. To test whether such upstream antisense initiation sites could function as promoters, we inverted the orientation of two of the unidirectional promoters found to be active (*ztf-11* and F58D5.5). If the lack of *in vivo* transcription elongation was a property of the element or initiation site itself, the GFP fusion should not be expressed. However, we observed that the two inverted unidirectional promoters both drove GFP expression. The expression patterns generated were similar in both orientations, although one was weaker when inverted (Figure 2 - figure supplements 6B,C). These results suggest that signals for productive elongation occur downstream of the transcription initiation site.

Similar to the upstream antisense transcription initiation observed at promoters, enhancers also show transcription initiation signals but generally do not produce stable transcripts (Core et al. 2014; Andersson et al. 2014). Previous studies have reported that some enhancers can function as promoters in transgenic assays and also at endogenous loci (Kowalczyk et al. 2012; Leung et al. 2015; Nguyen et al. 2016; van Arensbergen et al. 2016; Mikhaylichenko et al. 2018). To assess the potential promoter activities of *C. elegans* enhancers, we directly fused 12 putative enhancers that had

transcription initiation signal in embryos to a histone-GFP reporter gene and assessed transgenic strains for embryo expression. Two of the tested enhancers are located in introns, and one of these, from the *bro-1* gene, has been previously validated as an enhancer (Brabin et al. 2011); most of the others are associated with the *hlh-2* or *ztf-11* genes. We found that 10 of 12 tested regions drove reporter expression in embryos, including the two intronic enhancers (Figure 2 - figure supplement 5B,C). Whereas the *hlh-2* and *ztf-11* promoters drove strong, broad expression, the associated enhancers were active in a smaller number of cells and expression levels were overall lower (Figure 2 - figure supplement 6B,C). We also tested two enhancers in inverted orientation and found that both showed similar activity in both orientations, as observed for the two tested promoters (Figure 2 - figure supplement 5B,C). The percentage of enhancers that functioned as active promoters is higher than that observed in a cell based assay (Nguyen et al. 2016), possibly because all cell types are tested in an intact animal. Episomal based assays have also been reported to underestimate activity (Inoue et al. 2016).

Extensive regulation of chromatin accessibility in development

We observed that chromatin accessibility is highly dynamic across development, with most elements showing a significant change within the developmental time course (71%, >=2-fold change, FDR<0.01; Figure 4 - source data 1; see Methods). To investigate how accessibility relates to gene expression, we focused on the 13,596 elements annotated as protein-coding promoters. Of these, 10,199 displayed significant changes in accessibility in development, with the remaining 3397 promoters classified as having stable accessibility.

We defined 16 clusters of promoters with shared accessibility dynamics in development (Figure 4A and Figure 4 - figure supplement 1; Figure 4 - source data 1; Methods). Within clusters, we observed that promoter accessibility and nuclear RNA levels are usually correlated (mean r=0.47 (sd=0.11) across all clusters), indicating that accessibility is a good metric of promoter activity and overall gene expression (Figure 4 - figure supplement 1).

The shared patterns of developmental accessibility suggest that promoters within each cluster may control genes involved in common processes. To explore this possibility, we took advantage of recent single cell profiling data obtained from L2 larvae, which provides gene expression measurements for different tissues (Cao et al. 2017). We find that half of the developmental promoter clusters are enriched for genes with tissue biased expression (Figure 4A and Figure 4 - figure supplement 1). Based on these patterns of enrichment, we defined four gonad clusters (G1-G4), two intestine clusters (I1, I2), one hypodermal cluster (H) and one cluster enriched for neural and muscle genes (N+M) (Figure 4A and Figure 4 - figure supplement 1). Genes associated with the remaining eight clusters (Mix1–8) are generally expressed in multiple tissues, but predominantly in the soma (Figure 4A and Figure 4 - figure supplement 1). As expected, genes linked to the stable promoters are widely expressed (Figure 4A and Figure 4 - figure supplement 1). Interestingly, clusters associated with the same tissues-specific profiles can exhibit similar variations in accessibility but with different amplitude. For instance, gonad clusters G1 and G2 both show a sharp increase in accessibility at the L3 stage, however the increase is 1.5-fold larger in G2 than in G1. The gonad clusters are

generally characterized by an increase of promoter accessibility starting in L3 when germ cell number strongly increases.

To further investigate promoter clusters sharing accessibility dynamics, we performed Gene Ontology analyses on the associated genes. As expected, we found that clusters containing genes enriched for expression in a particular tissue are also associated with GO terms related to that tissue (Figure 4C and Figure 4 - figure supplement 1). For instance, cluster H contains genes highly expressed in hypodermis and GO terms linked to cuticle development. Of note, the four accessibility clusters enriched for expression in germ line are associated with GO terms for different sets of germ line functions (Figure 4 - figure supplement 1). Similarly, the two intestinal clusters also identify genes with different types of intestinal function. Furthermore, accessibility dynamics can reflect the temporal function of the associated promoters. For instance, cluster Mix4 has GO terms indicative of neuronal development and highest accessibility in the embryo, when many neurons develop. These results suggest that promoter clusters contain genes acting in a shared process and having a similar mode of regulation.

To identify potential transcriptional regulators, we asked whether the binding of particular transcription factors is enriched in any promoter clusters, using TF binding data from the modENCODE and modERN projects (Boyle et al. 2014; Kudron et al. 2017). TFs with enriched binding were found for each cluster, and the expression of such TFs was generally enriched in the expected tissue (Figure 5A ; Methods). For example, ELT-2, an intestine specific GATA protein (Fukushige et al. 1998), has enriched binding at promoters of intestinal clusters 1 and 2. Similarly, hypodermal

transcription factors BLMP-1 (Horn et al. 2014), NHR-25 (Gissendanner and Sluder 2000) and ELT-3 (Gilleard et al. 1999) are enriched in the hypodermal promoter cluster, and germ line XND-1 transcription factor (Wagner et al. 2010) is enriched in the germ line clusters of promoters. Following this approach, we identified novel tissue-specific associations for uncharacterized transcription factors, such as ZTF-18 and ATHP-1 with promoters of germ line clusters and CRH-2 with the intestinal clusters (Figure 5A). Taken together, the results suggest that promoters with shared accessibility patterns have shared cell-type specific activity, and they highlight potential regulators that are candidates for future studies.

Analysis of Ageing clusters

We next focused on chromatin accessibility changes during ageing. In contrast to the development time course, the accessibility of most promoters is stable during ageing, with only 13% (n=1,800) of promoters showing changes (Figure 4 - source data 1). Interestingly, 75% of these also had regulated accessibility in development.

As for the development time course, we clustered accessibility changes in ageing. We identified eight clusters of promoters with similar accessibility changes across ageing and annotated them based on tissue biases in gene expression (Figure 4B; Figure 4 - source data 1). This defined one intestinal cluster (I), two clusters enriched for intestine or hypodermal biased expression (I+H) and 5 mixed clusters. Several mixed clusters show weak gene expression enrichments, such as intestine expression in Mix1-2 and neural expression in Mix3 (Figure 4B). As observed for the

development clusters, enriched GO terms were consistent with gene expression biases (Figure 4B, Figure 4 - figure supplement 2).

We then evaluated the enrichment of transcription factors at each ageing promoter cluster. The binding of DAF-16/FoxO, a master regulator of ageing (Lin et al. 2001), is associated with four ageing promoter clusters (Figure 5B). Consistent with a prominent role in the intestine (Figure 4B; Kaplan and Baugh 2016), promoter clusters enriched for DAF-16 binding are enriched for intestinal genes (Figure 4B). The binding enrichment patterns of five other TFs implicated in ageing (DVE-1, NHR-80, ELT-2, FOS-1 and PQM-1 (Uno et al. 2013; Folick et al. 2015; Goudeau et al. 2011; Mann et al. 2016; Tian et al. 2016; Mao et al. 2016; Tepper et al. 2013) are similar to DAF-16 (Figure 5B). These TFs and DAF-16 are also enriched in developmental intestine promoter clusters (Figure 5A), supporting cooperation in development and ageing. A group of hypodermal TFs including BLMP-1, ELT-1 and ELT-3 are found enriched at promoters in the two I + H ageing clusters (Figure 5B). Finally, CEPB-1 binding is enriched in clusters Mix3 and Mix4, which are characterized by a continuous increase of promoter accessibility across ageing. This suggests a potential role of CEBP-1 in activating a subset of genes during ageing, as it is the case for its homologue CEBP-β in mouse (Sandhir and Berman 2010).

Conclusion

For the first time, we systematically map regulatory elements across the lifespan of an animal. We identified 42,245 accessible sites in *C. elegans* chromatin and functionally annotated them based on transcription patterns at the accessible site. This avoided the

problems of histone-mark based approaches for defining element function (Core et al. 2014; Henriques et al. 2018; Rennie et al. 2018). Our map identified promoters active across development and ageing, but we did not find promoters for every gene. Classes that would have been missed are those for genes expressed only in males or dauer larvae (which we did not profile) and genes not active under laboratory conditions. In addition, whole-animal profiling would miss promoters active in only a small number of cells. In the future, assaying accessible chromatin and nuclear transcription in specific cell types should identify many of these missed elements.

We found that accessibility of most elements changes during the life of the worm, supporting a key role played by chromatin structure. Despite the map being based on bulk profiling in whole animals, we find that regulatory elements with shared accessibility dynamics often share patterns of tissue-specific expression, GO annotation, and TF binding. The promoters within clusters are therefore excellent starting points for studies of cell- and process-specific gene expression. In summary, our identification of regulatory elements across *C. elegans* life together with an initial characterisation of their properties provides a key resource that will enable future studies of transcriptional regulation in development and ageing.

Methods

Collection of developmental time series samples

Wild-type N2 were grown at 20°C in liquid culture to the adult stage using standard S-basal medium with HB101 bacteria, animals bleached to obtain embryos, and the embryos hatched without food in M9 buffer for 24 hrs at 20°C to obtain synchronized starved L1 larvae. L1 larvae were grown in a further liquid culture at 20°C to the desired stage, then collected, washed in M9, floated on sucrose, washed again in M9, then frozen into "popcorn" by dripping embryo or worm slurry into liquid nitrogen. Popcorn were stored at -80°C until use. Times of growth were L1 (4 hrs), L2 (20 hrs), L3 (30 hrs), L4 (45 hrs), young adults (60 hrs). Mixed populations of embryos were collected by bleaching cultures of synchronized one day old adults.

Collection of ageing time series samples

glp-1(e2144) were raised at 15°C on standard NGM plates seeded with OP50 bacteria. Embryos were obtained by bleaching gravid adults and then approximately 6000 placed at 25°C on 150mm 2% NGM plates seeded with a 30X concentrated overnight culture of OP50. For harvest, worms were washed 3X in M9 and then worm slurry was frozen into popcorn by dripping into liquid nitrogen and stored at -80°C. Harvest times after embryo plating were D1/YA (53 hrs), D2 (71 hrs), D6 (167 hrs), D9 (239 hrs), D13 (335 hrs).

Nuclear isolation and ATAC-seq

Frozen embryos or worms (1–3 frozen popcorns) were broken by grinding in a mortar and pestle or smashing using a Biopulverizer, then the frozen powder was thawed in 10 ml Egg buffer (25 mM HEPES pH 7.3, 118 mM NaCl, 48 mM KCl, 2 mM CaCl2, 2 mM MgCl2). Ground worms were pelleted by spinning at 1500 g for 2 minutes, then resuspended in 10ml working Buffer A (0.3M sucrose, 10 mM Tris pH 7.5, 10 mM MgCl2, 1mM DTT, 0.5 mM spermidine 0.15 mM spermine, protease inhibitors (Roche complete, EDTA free) containing 0.025% IGEPAL CA-630. The sample was dounced 10X in a 14ml stainless steel tissue grinder (VWR), then the sample spun 100g for 6 min to pellet large fragments. The supernatant was kept and the pellet resuspended in a further 10 ml Buffer A, then dounced for 25 strokes. This was spun 100g for 6 min to pellet debris and the supernatants, which contain the nuclei, were pooled, spun again at 100g for 6 min to pellet debris, and transferred to a new tube. Nuclei were counted using a hemocytometer. One million nuclei were transferred to a 1.5 ml tube and spun 2000g for 10min to pellet. ATAC-seq was performed essentially as in (Buenrostro et al. 2013). The supernatant was removed, the nuclei resuspended in 47.5 ul of tagmentation buffer, incubated for 30 minutes at 37°C with 2.5 ul Tn5 enzyme (Illumina Nextera kit), and then tagmented DNA purified using a MinElute column (Qiagen) and converted into a library using the Nextera kit protocol. Typically, libraries were amplified using 12–16 PCR cycles. ATAC-seq was performed on two biological replicates for each developmental stage and each ageing time point.

DNAse I and MNase mapping

Replicate concentration courses of DNase I were performed for each stage as follows. Twenty million nuclei were digested in Roche DNAse I buffer for 10 minutes at 25C using 2.5, 5, 10, 25, 50, 100, 200, and 800 units/ml DNase I (Roche), then EDTA was added to stop the reactions. Embryo micrococcal nuclease (MNase) digestion concentration courses for embryos were made by digesting nuclei with 0.025, 0.05, 0.1, 0.25, 0.5, 1, 4, 8, or 16 units/ml MNase in 10mM Tris pH 7.5, 10mM MgCl2, 4mM CaCl2 for 10 minutes at 37C. Reactions were stopped by the additon of EDTA. Following digestions, total DNA was isolated from the nuclei following proteinase K and RNase A digestion, then large fragments removed by binding to Agencourt AMPure XP beads (0.5 volumes). Small double cut fragments < 300 bp were isolated either using a Pippen prep gel (protocol 1) or using Agencourt AMPure XP beads (protocol 2). DNA was converted into sequencing libraries using the Illumina Truseq kit or a homemade equivalent.

Transcription initiation and nuclear RNA profiling

Nuclei were isolated and then chromatin associated RNA (development series) or nuclear RNA (ageing series) was isolated. Chromatin associated RNA was isolated as in (Pandya-Jones and Black 2009), resuspending washed nuclei in Trizol for RNA extraction. To isolate nuclear RNA, nuclei were directly mixed with Trizol. Following purification, RNA was separated into fractions of 17–200nt and >200nt using Zymo clean and concentrate columns. To profile transcription elongation ("long cap RNA-seq") in the nucleus, stranded libraries were prepared from the >200nt RNA fraction using the NEB Next Ultra Directional RNA Library Prep Kit (#E7420S). Libraries were made from two

biological replicates for each developmental stage and each ageing time point. To profile transcription initiation ("short cap RNA-seq"), stranded libraries were prepared from the 17–200nt RNA fraction. Non-capped RNA was degraded by first converting uncapped RNAs into 5'-monophosphorylated RNAs using RNA polyphosphatase (Epibio), then treating with 5' Terminator nuclease (Epibio). The RNA was treated with calf intestinal phosphatase to remove 5' phosphates from undegraded RNA, and decapped using Tobacco Acid Pyrophosphatase (Epicentre), Cap-Clip Acid Pyrophosphatase (CellScript, for one L2 and one L3 replicate) or Decapping Pyrophosphohydrolase, (Dpph tebu-bio, for one L3 replicate) and then converted into sequencing libraries using the Illumina TruSeq Small RNA Preparation Kit kit. Libraries were size selected to be 145–225 bp long on a 6% acrylamide gel, giving inserts of 20–100 bp long. Libraries were made from two biological replicates for each developmental stage. During the course of this work, the TAP enzyme stopped being available; the Cap-Clip and Dpph enzymes perform less well than TAP. One L3 and one YA replicate was made using a slightly different protocol. Embryo short cap RNA-seq data from (Chen et al. 2013) was also included in the analyses (GSE42819).

ChIP-seq

Balls of frozen embryos or worms were ground to a powder using a mortar and pestle or a Retch Mixer Mill to break animals into pieces. Frozen powder was thawed into 1% formaldehyde in PBS, incubated 10 minutes, then quenched with 0.125M glycine. Fixed tissue was washed 2X with PBS plus protease inhibitors, once in FA buffer, then resuspended in 1ml FA buffer per 1 mL of ground worm powder and the extract

sonicated to an average size of 200 base pairs with a Diagenode Bioruptor or Bioruptor Pico for 25 pulses of 30 seconds followed by 30 seconds pause. For ChIP, 500ug protein extract was incubated 2ug antibody in FA buffer with protease inhibitors overnight at 4°C, then incubated with magnetic beads conjugated to secondary antibodies for 2hrs at 4°C. Magnetic beads bound to immunoprecipitate were washed at room temperature twice in FA+protease inhibitors (50mM Hepes pH7.5, 1mM EDTA, 1% TritonX-100, 0.1% sodium deoxycholate, and 150mM NaCl), then once each in FA with 0.5M NaCl, FA with 1M NaCl, 0.25M LiCl (containing 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris pH8) and finally twice with TE pH8. Immunoprecipitated DNA was then eluted at twice with 1%SDS, 250mM NaCl, 10mM Tris pH8, 1mM EDTA at 65°C. Eluted DNA was treated with RNase for 30min at 37C and crosslinks reversed by overnight incubation at 65°C with 200ug/ml proteinaseK, and the DNA purified using a Qiagen column. Following ChIP and DNA purification, libraries were prepared using the Illumina TruSeq kit. Fragments in the 250–350 base pair range were selected using Agencourt AMPure XP beads. Two biological replicate ChIPs were conducted for each histone modification at each developmental time point (Embryo, L1, L2, L3, L4, YA). Antibodies used were: anti-H3K4me3 (Abcam ab8580, GR273043-4), anti-H3K4me1 (Abcam ab8895, GR149140-2), anti-H3K36me3 (Abcam ab9050, lot GR288636-2), and anti-H327me3 (Wako 309-95259).

Data processing

The WBcel215/ce10 (WS220) version of the *C. elegans* genome was used throughout the study. Reads were aligned using bwa-backtrack (H. Li and Durbin 2009) in

single-end (ATAC-seq, short cap RNA-seq, ChIP-seq) or paired-end mode (ATAC-seq - developmental only, DNase-seq, MNase-seq, long cap RNA-seq). Low-quality (q < 10), mitochondrial and modENCODE-blacklisted (Boyle et al., 2014) reads were discarded at this point.

For ATAC-seq, normalised genome-wide accessibility profiles from single-end reads were then calculated with MACS2 (Zhang et al. 2008) using the parameters --format BAM --bdg --SPMR --gsize ce --nolambda --nomodel --extsize 150 --shift -75 --keep-dup all. Developmental ATAC-seq was also processed in paired-end mode (ATAC-seq libraries of ageing samples were single-end). We did not observe major differences between accessible sites identified from paired-end, and single-end profiles, and therefore use single-end profiles throughout the study for consistency.

Short cap and long cap data was processed essentially as in (Chen et al. 2013). Following alignment, and filtering, transcription initiation was represented using strand-specific coverage of 5' ends of short cap reads. Transcription elongation was represented as strand-specific coverage of long cap reads, with regions between read pairs filled in. For browsing, transcription elongation signal was normalised between samples by sizeFactors calculated from gene-level read counts using DESeq2 (Love, Huber, and Anders 2014). Normalised (linear) coverage signal was then further log-transformed with $log_2(normalised\_coverage + 1)$.

ChIP-seq data was processed as in (Chen et al. 2014). After alignment and filtering, the BEADS algorithm was used to generate normalised ChIP-seq coverage tracks (Cheung et al. 2011).

For downstream analyses - aggregate plots, heatmaps, screen shots - stage-specific tracks were obtained by averaging normalised signal across two biological replicates. Manipulations of genome-wide signal were performed using bedtools (Quinlan and Hall 2010), UCSC utilities (Kent et al. 2010), and wiggleTools (Zerbino et al. 2014). Computationally intensive steps were managed and parallelised using snakemake (Köster and Rahmann 2012). Genome-wide data was visualised using the Integrative Genomics Viewer (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

Identification of accessible sites

Accessible sites were identified as follows. We first identified concave regions (regions with negative smoothed second derivative) from ATAC-seq coverage averaged across all stages and replicates. This approach is extremely sensitive, identifying a large number (>200,000) of peak-like regions. We then scored all peaks in each sample using the magnitude of the sample-specific smoothed second derivative. We used IDR(Q. Li et al. 2011) on the scores to assess stage-specific signal levels and biological reproducibility, setting a conservative cutoff at 0.001. Final peaks boundaries were set to peak accessibility extended by 75bp on both sides. We found that calling peaks using paired end or single end data were highly similar, but some regions were captured better by one or the other. Developmental ATAC-seq datasets were sequenced paired-end and ageing datasets single-end. Peaks were therefore called separately using developmental paired-end data, developmental single-end data extended to 150bp and shifted 75bp upstream, and ageing (single-end only) data, and then merged. This was achieved by successively including peaks from the three sets if they did not overlap a peak already

identified in an earlier set. Figure 1 - source data 1 gives peak calls and ATAC peak heights at each stage.

Annotation of regulatory elements

Patterns of nuclear transcription were used to annotate elements. At each stage, separately on both strands, we assessed 1) initiating and elongating transcription at the site, 2) continuity of transcription from the site to the closest downstream gene, and 3) positioning of nearby exons (on the matching strand). WormBase WS260 genome annotations - with coordinates backlifted to WBcel215/ce10 (WS220) - were used throughout this study.

To assess for transcription elongation at an accessible site, we counted 5' ends of long cap reads upstream (-250:-75), and downstream (+75:+250) of peak accessibility. We then used two approaches to identify sites with a local increase in transcription elongation. First, we used DESeq2 to test for an increase in downstream vs upstream counts ("jump" method). Statistical significance was called at log2FoldChange > 1.5, and adjusted p-value < 0.1 (one-sided test). To capture additional regions with weak signal ("incr" method), we accepted sites with 0 reads upstream, at least one read in both biological replicates downstream, and 3 reads total when summed across both biological replicates.

To assess transcription initiation, we pooled short cap across all six wild-type stages, and included two additional embryo replicates from (Chen et al. 2013). The pooled signal was filtered for reproducibility by only keeping signal at base pairs with

non-zero transcription initiation in at least two replicates. We then required the presence of at least one base pair with reproducible signal within 125bp of peak accessibility to designate an accessible site as having transcription initiation. For every site, we also defined a representative transcription initiation mode as the position with maximum short cap signal within 125bp of peak accessibility. For sites without reproducible short cap signal, we used an extrapolated, "best-guess" position at 60bp downstream of peak accessibility.

We annotated accessible sites as coding_promoter or pseudogene_promoter if they fulfilled the following four criteria. 1) The accessible site had transcription initiation, and passed at least one of the elongation tests (jump or incr), or passed both elongation tests (jump and incr). 2) Transcription initiation mode at the accessible site was either upstream of the closest first exon, or, in the presence of a UTR, up to 250bp downstream within the UTR. (The closest first exon was chosen based on the distance between the 5' end of the first exon and peak accessibility at the accessible site, allowing the 5' end of the exon to be up to 250bp upstream or anywhere downstream of peak accessibility). 3) The region from peak accessibility to the closest first exon did not contain the 5' end of a non-first exon. 4) Distal sites (peak accessibility >250bp from the closest first exon) were additionally required to (a) have continuous long cap coverage from 250bp downstream of peak accessibility to the closest first exon, and (b) be further than 250bp away from any non-first exon.

We then further attempted to assign a single, lower-confidence promoter to genes that were not assigned a promoter so far. For every gene without promoter assignments, we re-examined sites that fulfilled criteria (2-4), and were either intergenic,

or within 250bp of the closest first exon. We then annotated the site with the largest jump test log2FoldChange as the promoter, if it was also larger than 1.

Next, sites within 250bp of the 5' end of an annotated tRNA, snoRNA, miRNA, snRNA or rRNA were annotated as non-coding_RNA. Intergenic sites more than 250bp away from annotated exons that had initiating transcription, and passed the jump test were annotated as unknown_promoter. All remaining sites were annotated as transcription_initiation or no_transcription based on whether they had transcription initiation.

Elements were then annotated on each strand based on aggregating transcription patterns across stages by determining the "highest" annotation using the ranking of: coding_promoter, pseudogene_promoter, non-coding_RNA, unknown_promoter, transcription_initiation, no_transcription. Element type was then defined using the following ranking: coding_promoter on either strand => coding_promoter; pseudogene_promoter on either strand => pseudogene_promoter; non-coding_RNA on either strand => non-coding_RNA; unknown_promoter on either strand => unknown_promoter; transcription_initiation on either strand => putative_enhancer; all remaining sites => other_element. Figure 2 - source data 1 gives annotation information.

Motif analyses

Inr and TATA consensus sequences were obtained from (Sloutskin et al. 2015), and mapped with zero mismatches using homer (Heinz et al. 2010).

Clustering of promoter accessibility

Regulatory elements with regulated accessibility were determined as follows. All elements (n=42,245) were tested for a difference in ATAC-seq coverage between any developmental time point or between any ageing time point using DESeq2 (Love, Huber, and Anders 2014). Sites with >= 2 absolute fold change and adjusted p-value < 0.01 were defined as "regulated" (n=30,032 in development and 6,590 in ageing), and regulated promoters (n=10,199 in development and 1,800 in ageing) were used in clustering analyses.

For clustering analyses, depth-normalised ATAC-seq coverage of each promoter was calculated at each time point in development or ageing. The relative accessibility of a regulated promoter was calculated at each time point in development or ageing by applying the following formula:

$$log_2\left(ATACseq\ coverage_{time\ point\ i} + 1\right) - log_2(mean\ ATACseq\ coverage\ across\ time\ points + 1) \quad .$$

For each promoter, its mean ATAC-seq coverage across time points was calculated separately for developmental or ageing time course. Clustering was performed using *k*-medoids, as implemented in the pam() method of the cluster R package (Maechler et al. 2017). Different numbers of clusters were tested for clustering of regulatory elements in developmental and ageing datasets and the ones with the best homogeneity of normalized changes in ATAC-seq signals within each cluster was chosen. We manually merged two ageing clusters showing comparable accessibility and tissue-specific genes enrichment (resulting in the cluster I+H [2]). Clusters labels were determined based on

which tissues showed a strong enrichment for tissue-enriched genes within each cluster (> 3.5-fold increase in the proportion of tissue-enriched genes between each tissue).

To compare accessibility and gene expression, FPM-normalised gene-level read counts were calculated using DESeq2, and then averaged across biological replicates. For visualisation, relative expression levels were calculated using the approach described above for relative promoter accessibility (see formula above), with FPM values instead of ATAC-seq coverage values.

Tissue-specific enrichment analyses used single-cell RNA-seq data from (Cao et al. 2017). Genes were considered enriched in a given tissue if they had a fold-change >= 3 between the first and the second tissues with the highest expression and an adjusted p-value < 0.01. GO enrichments were evaluated using the R package gProfileR (Reimand et al. 2016). Significant enrichment was set at an adjusted p-value of 0.05, and hierarchically redundant terms were automatically removed by gProfileR.

Cluster-specific transcription factors binding enrichment

Optimal IDR-thresholded transcription factors ChIP-seq peaks datasets were downloaded from ENCODE portal. When multiple datasets were available for a given transcription factor, peaks from all of the datasets were combined and merged. ChIP-seq profiles were manually inspected and datasets showing poor-quality tracks were discarded, resulting in a set of 152 transcription factors.

To define TFBS clusters (Figure 1 - figure supplement 1C, Figure 2 - figure supplement 1), we extended the TF peak calls to 200bp on either side of the summit, and clustered overlapping peak calls using a single-linkage approach.

Prior to analysis of TF peak enrichment at annotated promoters (Figure 5), any promoter overlapping with more than 10 transcription factors peaks was considered as “hot” and removed from the initial set of 13,596 annotated promoters, resulting in 8,351 to be assessed by enrichment analysis. Only transcription factors with more than 200 peaks overlapping “non-hot” promoters were kept, to remove potentially weak TF peaks datasets constituted only of residual “hot” peaks. Following this stringent filtering, 62 transcription factors could be assayed for binding enrichment. A transcription factor peak was assigned to a promoter if any portion of the 400bp region centered at the peak summit intersected the promoter. Transcription factor binding enrichment in each cluster was estimated using the odds ratio and enrichments with an associated p-value < 0.01 (Fisher’s exact test) were kept. Transcription factors which did not show enrichment higher than 2 in any cluster were discarded. Figure 5 summarizes the transcription factor binding enrichment in each cluster during development or ageing. Relative tissue expression profiles of each transcription factor at the L2 stage (data from Cao et al. 2017) was calculated in each tissue by taking the log2 of its expression (TPM) in the tissue divided by its mean expression across all tissues. A pseudo-value of 0.1 was first added to all the TPM values before calculation of the relative levels of expression.

Construction of transgenic lines

Transgene constructs were made using three-site Gateway cloning (Invitrogen) as in (Chen et al. 2014). Site 1 has the regulatory element sequence to be tested, site 2 has a synthetic outron (OU141; Conrad et al. 1995) fused to *his-58* (plasmid pJA357), and site 3 has gfp-tbb-2 3'UTR (pJA256; Zeiser et al. 2011) in the MosSCI compatible vector pCFJ150, which targets Mos site Mos1(ttTi5605); MosSCI lines were generated as described (Frøkjaer-Jensen et al. 2008).

Data access

ATAC-seq, ChIP-seq, DNase/MNase-seq, long/short cap RNA-seq data from this study have been deposited in the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE114494.

Acknowledgements

We thank C. Bradshaw for bioinformatics support, K. Harnish for sequencing, and F. Carelli, C. Gal, and A. Frapporti for comments on the manuscript. The work was supported by Wellcome Trust Senior Research Fellowships to JA (054523 and 101863), a Wellcome Trust PhD fellowship to JJ (097679), a Sir Robert Edwards Scholarship from Churchill College, an English Speaking Union Graduate Scholarship, and funding from the Cambridge Trust to MS, a Medical Research Council DTP studentship to JS, and a Thouron award to CW. This study was also supported by the European Sequencing and

Genotyping Infrastructure (funded by the EC, FP7/2007-2013) under Grant Agreement 26205 (ESGI) as part of the transnational access program. We thank Drs. Hans Lehrach and Marie-Laure Yaspo for generous support of the ESGI project, Dr. Marc Sultan for setting up sequencing technology platforms, and Mathias Linser and the rest of the sequencing team of the Department of Vertebrate Genomics at the Max Planck Institute for Molecular Genetics for technical assistance. We also acknowledge core support from the Wellcome Trust (092096) and Cancer Research UK (C6946/A14492).

References

Allen MA, Hillier LW, Waterston RH, Blumenthal T. 2011. A global analysis of C. elegans trans-splicing. *Genome Res* **21**: 255–264.

Andersson R. 2015. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays* **37**: 314–323.

Andersson R, Refsing Andersen P, Valen E, Core LJ, Bornholdt J, Boyd M, Heick Jensen T, Sandelin A. 2014. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* **5**: 5336.

Andersson R, Sandelin A, Danko CG. 2015. A unified architecture of transcriptional regulatory elements. *Trends Genet*. http://dx.doi.org/10.1016/j.tig.2015.05.007.

Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier LW, Janette J, Jiang L, et al. 2014. Comparative analysis of regulatory information and circuits across distant species. *Nature* **512**: 453–456.

Brabin C, Appleford PJ, Woollard A. 2011. The Caenorhabditis elegans GATA factor ELT-1 works through the cell proliferation regulator BRO-1 and the Fusogen EFF-1 to maintain the seam stem-like fate. *PLoS Genet* **7**: e1002200.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**: 661–667.

Chen RA-J, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LW, Jeffers TE, Ahringer J. 2013. The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. *Genome Res* **23**: 1339–1347.

Chen RA-J, Stempor P, Down TA, Zeiser E, Feuer SK, Ahringer J. 2014. Extreme HOT regions are CpG-dense promoters in C. elegans and humans. *Genome Res* **24**: 1138–1146.

Conrad R, Lea K, Blumenthal T. 1995. SL1 trans-splicing specified by AU-rich synthetic RNA inserted at the 5' end of Caenorhabditis elegans pre-mRNA. *RNA* **1**: 164–170.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread

pausing and divergent initiation at human promoters. *Science* **322**: 1845–1848.

Crawford GE, Davis S, Scacheri PC, Renaud G, Halawi MJ, Erdos MR, Green R, Meltzer PS, Wolfsberg TG, Collins FS. 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat Methods* **3**: 503–509.

Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. 2017. Chromatin accessibility dynamics reveal novel functional enhancers inC. elegans. *Genome Res* **27**: 2096–2107.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**: e1000384.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**: 817–825.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–49.

Evans KJ, Huang N, Stempor P, Chesney MA, Down TA, Ahringer J. 2016. Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. *Proc Natl Acad Sci U S A*. http://dx.doi.org/10.1073/pnas.1608162113.

Flynn RA, Almada AE, Zamudio JR, Sharp PA. 2011. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A* **108**: 10460–10465.

Folick A, Oakley HD, Yu Y, Armstrong EH, Kumari M, Sanor L, Moore DD, Ortlund EA, Zechner R, Wang MC. 2015. Aging. Lysosomal signaling molecules regulate longevity in Caenorhabditis elegans. *Science* **347**: 83–86.

Frøkjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, Grunnet M, Jorgensen EM. 2008. Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet* **40**: 1375–1383.

Fukushige T, Hawkins MG, McGhee JD. 1998. The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. *Dev Biol* **198**: 286–302.

Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, et al. 2014. Comparative analysis of the transcriptome across distant species. *Nature* **512**: 445–448.

Gilleard JS, Shafi Y, Barry JD, McGhee JD. 1999. ELT-3: A Caenorhabditis elegans GATA factor expressed in the embryonic epidermis during morphogenesis. *Dev Biol* **208**: 265–280.

Gissendanner CR, Sluder AE. 2000. nhr-25, the Caenorhabditis elegans ortholog of ftz-f1, is required for epidermal and somatic gonad development. *Dev Biol* **221**: 259–272.

Goudeau J, Bellemin S, Toselli-Mollereau E, Shamalnasab M, Chen Y, Aguilaniu H. 2011. Fatty acid desaturation links germ cell loss to longevity through NHR-80/HNF4 in C. elegans. *PLoS Biol* **9**: e1000599.

Gu W, Lee H-C, Chaves D, Youngman EM, Pazour GJ, Conte D Jr, Mello CC. 2012. CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. *Cell* **151**: 1488–1500.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.

Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC, Adelman K. 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev*. http://dx.doi.org/10.1101/gad.309351.117.

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, et al. 2013. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41**: 827–841.

Ho MCW, Quintero-Cadena P, Sternberg PW. 2017. Genome-wide discovery of active regulatory elements and transcription factor footprints in Caenorhabditis elegans using DNase-seq. *Genome Res*. http://dx.doi.org/10.1101/gr.223735.117.

Horn M, Geisen C, Cermak L, Becker B, Nakamura S, Klein C, Pagano M, Antebi A. 2014. DRE-1/FBXO11-dependent degradation of BLMP-1/BLIMP-1 governs C. elegans developmental timing and maturation. *Dev Cell* **28**: 697–710.

Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, et al. 2007. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. *PLoS Biol* **5**: e237.

Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, Ahituv N, Shendure J. 2016. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res*. http://dx.doi.org/10.1101/gr.212092.116.

Kaplan REW, Baugh LR. 2016. L1 arrest, daf-16/FoxO and nonautonomous control of post-embryonic development. *Worm* **5**: e1175196.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480–485.

Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187.

Kim T-K, Shiekhattar R. 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162**: 948–959.

Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18**: 956–963.

Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D, et al. 2012. Intragenic enhancers act as alternative promoters. *Mol Cell* **45**: 447–458.

Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. 2013. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. *Elife* **2**: e00808.

Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, Kirkey M, Hammonds AS, Gersch J, Ammouri H, et al. 2017. The modERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors. *Genetics*. http://dx.doi.org/10.1534/genetics.117.300657.

Leung D, Jung I, Rajagopal N, Schmitt A, Selvaraj S, Lee AY, Yen C-A, Lin S, Lin Y, Qiu Y, et al. 2015. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**: 350–354.

Lin K, Hsin H, Libina N, Kenyon C. 2001. Regulation of the Caenorhabditis elegans longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nat Genet* **28**: 139–145.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2017. cluster: Cluster Analysis Basics and Extensions.

Mann FG, Van Nostrand EL, Friedland AE, Liu X, Kim SK. 2016. Deactivation of the GATA Transcription Factor ELT-2 Is a Major Driver of Normal Aging in C. elegans. *PLoS Genet* **12**: e1005956.

Mao XR, Kaufman DM, Crowder CM. 2016. Nicotinamide mononucleotide adenylyltransferase promotes hypoxic survival by activating the mitochondrial unfolded protein response. *Cell Death Dis* **7**: e2113.

Merritt C, Rasoloson D, Ko D, Seydoux G. 2008. 3′ UTRs Are the Primary Regulators of

Gene Expression in the C. elegans Germline. *Curr Biol* **18**: 1476–1482.

Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM. 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32**: 42–57.

Nguyen TA, Jones RD, Snavely AR, Pfenning AR, Kirchner R, Hemberg M, Gray JM. 2016. High-throughput functional comparison of promoter and enhancer activities. *Genome Res*. http://dx.doi.org/10.1101/gr.204834.116.

Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**: 1896–1908.

Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau J-C, Ferrier P, Spicuglia S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J* **30**: 4198–4210.

Pérez-Lluch S, Blanco E, Tilgner H, Curado J, Ruiz-Romero M, Corominas M, Guigó R. 2015. Absence of canonical marks of active chromatin in developmentally regulated genes. *Nat Genet* **47**: 1158–1167.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851–1854.

Rennie S, Dalby M, Lloret-Llinares M, Bakoulis S, Dalager Vaagensø C, Heick Jensen T, Andersson R. 2018. Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter-encoded enhancer activities. *Nucleic Acids Res*. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky244/4962481 (Accessed April 12, 2018).

Rennie S, Dalby M, Lloret-Llinares M, Bakoulis S, Vaagenso CD, Jensen TH, Andersson R. 2017. Transcription start site analysis reveals widespread divergent transcription in D. melanogaster and core promoter encoded enhancer activities. *bioRxiv* 221952. https://www.biorxiv.org/content/early/2017/11/18/221952 (Accessed November 20, 2017).

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* **3**: 511–518.

Saito TL, Hashimoto S-I, Gu SG, Morton JJ, Stadler M, Blumenthal T, Fire A, Morishita S. 2013. The transcription start site landscape of C. elegans. *Genome Res* **23**: 1348–1361.

Sandhir R, Berman NEJ. 2010. Age-dependent response of CCAAT/enhancer binding proteins following traumatic brain injury in mice. *Neurochem Int* **56**: 188–193.

Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC, et al. 2013. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A* **110**: 2876–2881.

Tepper RG, Ashraf J, Kaletsky R, Kleemann G, Murphy CT, Bussemaker HJ. 2013. PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. *Cell* **154**: 676–690.

Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, et al. 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biol* **12**: R43.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.

Tian Y, Garcia G, Bian Q, Steffen KK, Joe L, Wolff S, Meyer BJ, Dillin A. 2016. Mitochondrial Stress Induces Chromatin Reorganization to Promote Longevity and UPR(mt). *Cell* **165**: 1197–1208.

Tittel-Elmer M, Bucher E, Broger L, Mathieu O, Paszkowski J, Vaillant I. 2010. Stress-induced activation of heterochromatic transcription. *PLoS Genet* **6**: e1001175.

Uno M, Honjoh S, Matsuda M, Hoshikawa H, Kishimoto S, Yamamoto T, Ebisuya M, Yamamoto T, Matsumoto K, Nishida E. 2013. A fasting-responsive signaling pathway that extends life span in C. elegans. *Cell Rep* **3**: 79–91.

van Arensbergen J, FitzPatrick VD, de Haas M, Pagie L, Sluimer J, Bussemaker HJ, van Steensel B. 2016. Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol*. http://dx.doi.org/10.1038/nbt.3754.

Wagner CR, Kuervers L, Baillie DL, Yanowitz JL. 2010. xnd-1 regulates the global recombination landscape in Caenorhabditis elegans. *Nature* **467**: 839–843.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355–364.

Zeiser E, Frøkjær-Jensen C, Jorgensen E, Ahringer J. 2011. MosSCI and gateway compatible plasmid toolkit for constitutive and inducible expression of transgenes in the C. elegans germline. *PLoS One* **6**: e20082.

Zhang H, Gao L, Anandhakumar J, Gross DS. 2014. Uncoupling transcription from covalent histone modification. *PLoS Genet* **10**: e1004202.

Figure legends

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, Day 2, Day 6, Day 9, Day 13). (B) Representative screen shot of normalised genome-wide accessibility profiles in the eleven samples.

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, element annotation coloured as in (B) right, and gene model. (B) Accessible elements by annotation class. (C) Left, distribution of the number of promoters and enhancers per gene; right, boxplot shows that genes with more promoters also have more enhancers.

Figure 3. Heatmaps of HMs/factors and properties of enhancers and promoters.

(A) Heatmaps of indicated histone modifications and CV values at coding promoters (top), and enhancers (bottom). Elements are ranked by mean H3K4me3 levels. CV values are correlated with 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).

Figure 4. Shared dynamics of promoter accessibility in development and ageing.

(A,B) Clusters of promoters showing shared relative accessibility patterns across (A) development or (B) ageing. Relative accessibility at each time point was defined by the log2 of the depth-normalized ATAC-seq coverage at each time divided by the mean ATAC-seq coverage across the time series (see Methods). For each cluster, the percentage of associated genes with tissue-enriched expression determined from single-cell L2 larval RNA-seq data (Cao et al. 2017) for each tissue is also shown. (C,D) Examples of GO terms enriched in (C) developmental or (D) ageing clusters.

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. 2017); peaks in HOT regions were excluded (see Methods). Only TFs enriched more than 2-fold 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, relative tissue expression (log2(tissue TPM/mean TPM across all tissues)), and tissue expression level TPM. Expression data are from (Cao et al. 2017).

Legends for Figure Supplements

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. (B) SPMR-normalised coverage of a DNase I concentration series (blue tracks), MNase concentration series (green tracks), and ATAC-seq (red track) at the *lin-23* locus. 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 normalised coverage at transcription factor binding sites defined by clustering modENCODE/modERN peak calls (n=36,389; 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. (D) Normalised read coverage of ATAC-seq prepared from nuclei harvested from live (red), or frozen (blue) embryos.

Figure 2 - figure supplement 1. Comparisons to existing accessibility maps.

(A) Venn diagrams showing the overlap of transcription factor binding sites defined by clustering modENCODE/modERN peak calls (n=36,389; Methods) to accessible sites from this study and two previous studies ((Daugherty et al. 2017) and (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; remaining plots compare regions found in only one study or both studies. Plots show 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 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).

Figure 2 - figure supplement 2. 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.

Figure 2 - figure supplement 3. 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: 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 150bp of peak accessibility. For (Gu et al., 2012), TSSs were clustered using a single-linkage approach using a distance threshold of 50bp, and the overlaps are based on those clusters.

Figure 2 - figure supplement 4. Types of unknown promoters.

(A) Position and orientation of sites annotated as unknown_promoter on the forward or reverse strands relative to gene annotations (n=3,026). (B–D) Examples of transcription patterns at unknown promoters (B) coding_promoter_antisense, (C) genic_region_antisense, (D) intergenic.

Figure 2 - figure supplement 5. Tests of promoter activity of annotated promoters and enhancers

(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 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.

Figure 3 - figure supplement 1. Histone marks and motif enrichments sorted by CV value.

(A) As Figure 3A, except that accessible sites were ranked based on CV values, and the heatmap additionally includes H3K36me3, aligned at the start of the associated gene annotation. (B) As Figure 3B, except that the three groups of promoters and enhancers

were defined based on CV values. Bottom CV represents stable expression and top CV regulated expression.

Figure 4 - figure supplement 1. Characteristics of developmental clusters.

Clusters of promoters showing shared accessibility patterns across development. The first column of plots shows promoter relative ATAC coverage across the time series as described in Methods. The second column shows the same information displayed as a heatmap, each row representing a promoter. Values are contained within a color scale from -2 (dark blue) to 0 (white) to +2 (dark red). The third column shows relative gene expression across the same time series as described in Methods, each row representing a gene. The same color scale is used here. The boxplots column represents expression (in TPM) of clustered genes in individual tissues (data from Cao et al 2017). The barplots column represents the percentage of genes within the cluster enriched in each tissue as described in Methods. Finally, the horizontal bar plots column shows the top 5 enriched GO terms obtained for each cluster from the corresponding list of genes using gProfiler as described in Methods. MF=Molecular Function, CC=Cellular Component, BP=Biological Process.

Figure 4 - figure supplement 2. Characteristics of ageing clusters.

Similar to Figure 4 **-** figure supplement 1 for ageing clusters.

Source data

Figure 1 - source data 1. Accessible sites identified using ATAC-seq

- **chrom, start, end** location of the accessible site in bed-style coordinates (ce10)
- **atac_%stage_height** maximum SPMR-normalised ATAC-seq signal at the peak in %stage (one of *wt_emb*, *wt_l1*, *wt_l2*, *wt_l3*, *wt_l4*, *wt_ya*, *glp1_d1*, *glp1_d2*, *glp1_d6*, *glp1_d9*, *glp1_d13*)
- **atac_source** source of the ATAC-seq peak call (see Methods)
    - *atac_wt_pe* wt (developmental) ATAC-seq treated as paired-end
    - *atac_wt_se* wt (developmental) ATAC-seq treated as single-end
    - *atac_glp1_se* glp-1 (aging) ATAC-seq, single-end only

Figure 2 - source data 1. Regulatory annotation of accessible sites

- **chrom, start, end** location of the accessible site in bed-style coordinates (ce10)
- **annot** final regulatory element type, obtained by combining strand-specific transcription patterns (see Methods)
- **annot_%strand** annotation of the strand-specific transcription patterns at the site (%strand is either *fwd* or *rev*)
- **promoter_gene_id_%strand**, **promoter_locus_id_%strand**, **promoter_gene_biotype_%strand** WormBase gene id, locus id, biotype for sites annotated as coding_promoter, pseudogene_promoter or non-coding_RNA on %strand
- **associated_gene_id**, **associated_locus_id** WormBase gene id, locus id of genes whose gene body or outron region overlaps the site. These are defined for

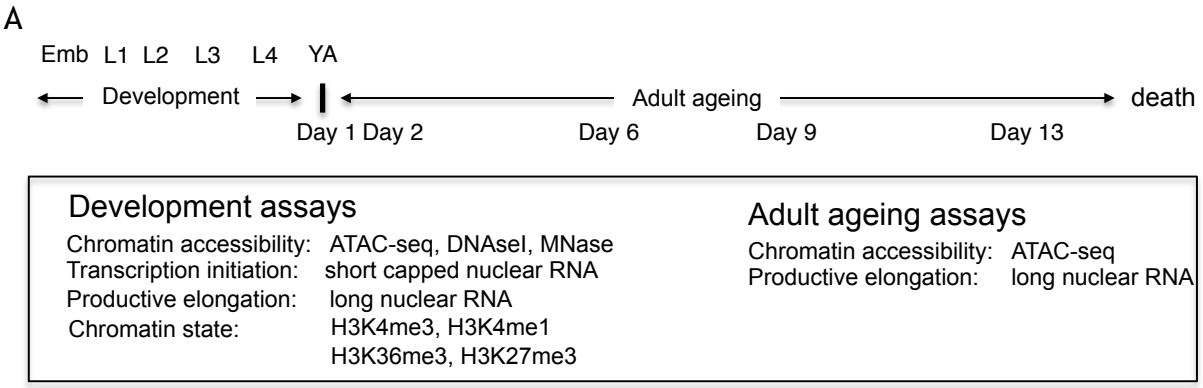
for sites annotated as unknown_promoter, putative_enhancer or other_element. If a site overlaps multiple genes, all overlaps are reported, separated by commas.

- **tss_%strand** representative transcription initiation mode on %strand (Methods)
- **scap_%strand_passed** True or False based on whether the site has reproducible transcription initiation (Methods)
- **lcap_%stage_%strand_passed_jump** True or False based on whether the site passed the jump test for elongating transcription (Methods, %stage is one of *wt_emb*, *wt_l1*, *wt_l2*, *wt_l3*, *wt_l4*, *wt_ya*, *glp1_d1*, *glp1_d2*, *glp1_d6*, *glp1_d9*, *glp1_d13*)
- **lcap_%stage_%strand_passed_incr** True or False based on whether the site passed the incr test for elongating transcription (Methods)

Figure 4 - source data 1. Promoter accessibility clusters in development and ageing

- **chrom, start, end** location of the accessible site in bed-style coordinates (ce10)
- **devel_is_dynamic** True or False based on whether the site shows differential accessibility between any two developmental stages
- **ageing_is_dynamic** True or False based on whether the site shows differential accessibility between any two ageing time points
- **devel_prom_cluster_label** assigned developmental accessibility promoter cluster
- **ageing_prom_cluster_label** assigned ageing accessibility promoter cluster

Figure 1



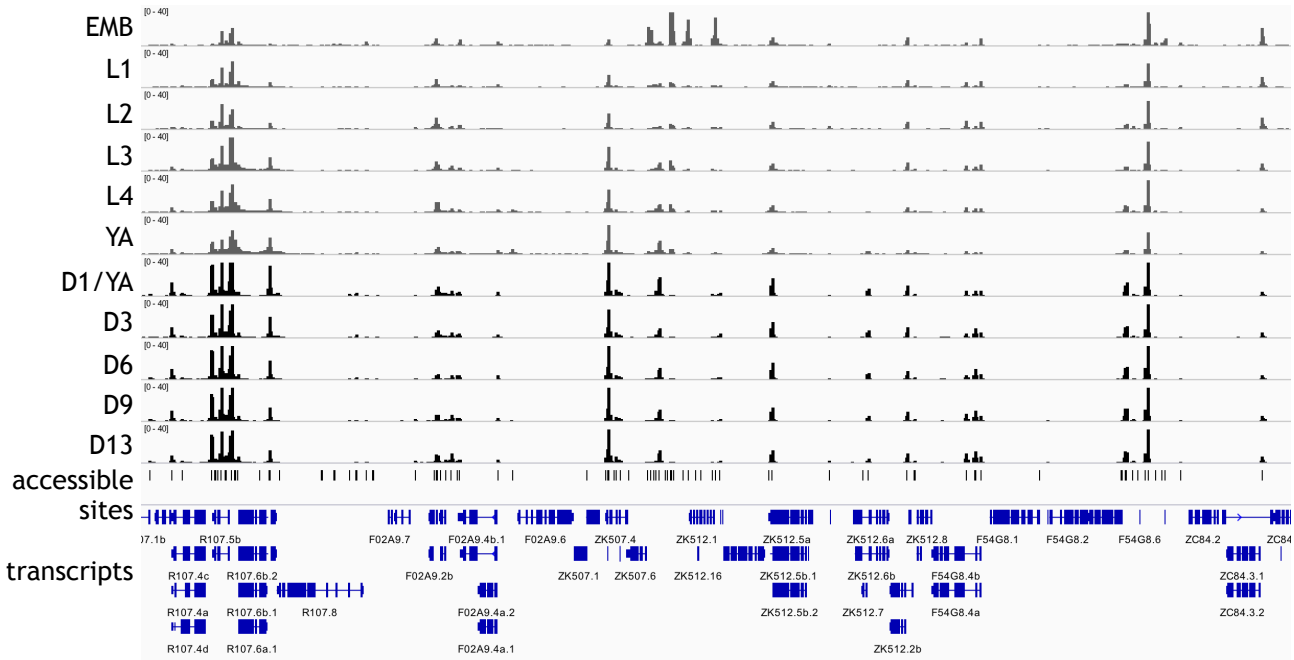
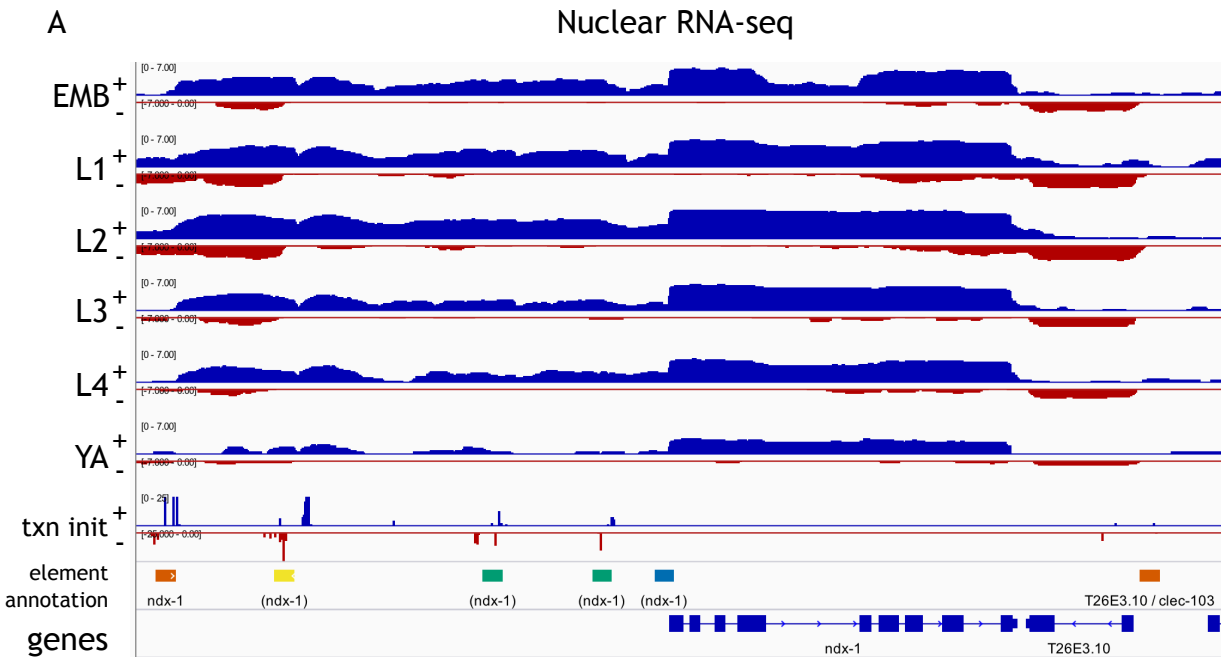
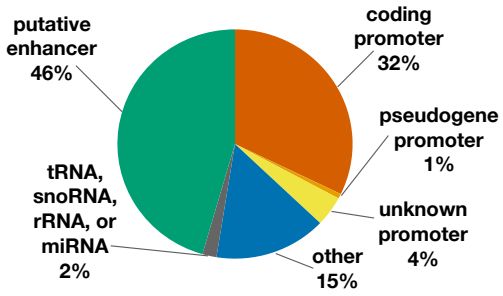


Figure 2



B

all	42,245
coding promoter	13,596
pseudogene promoter	291
unknown promoter	1,791
putative enhancer	19,231
tRNA, snoRNA, rRNA, or miRNA	824
other	6,512



C

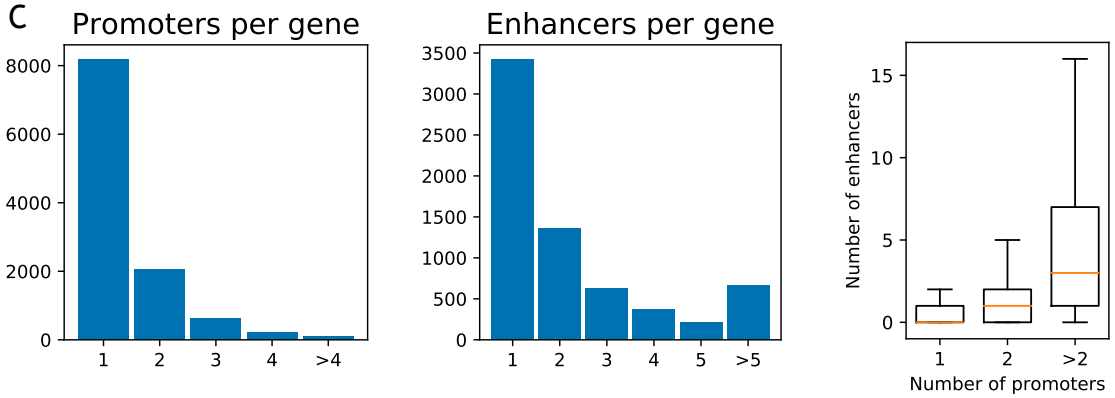
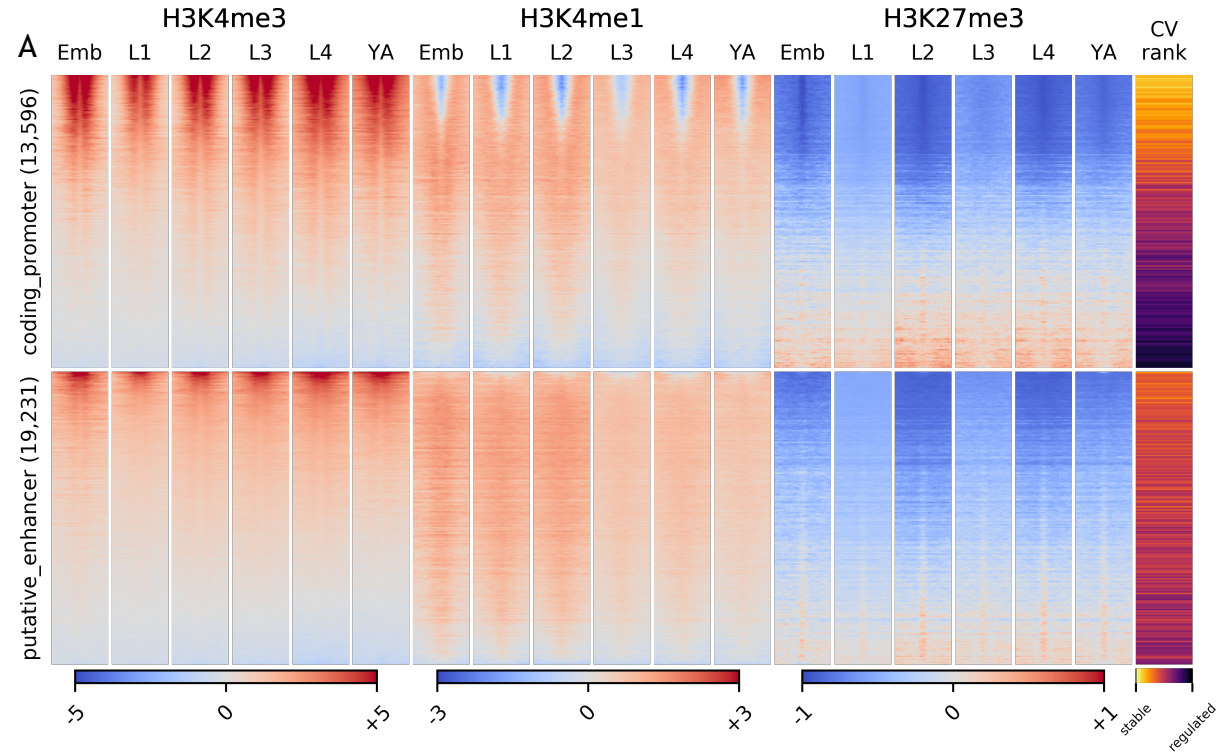


Figure 3



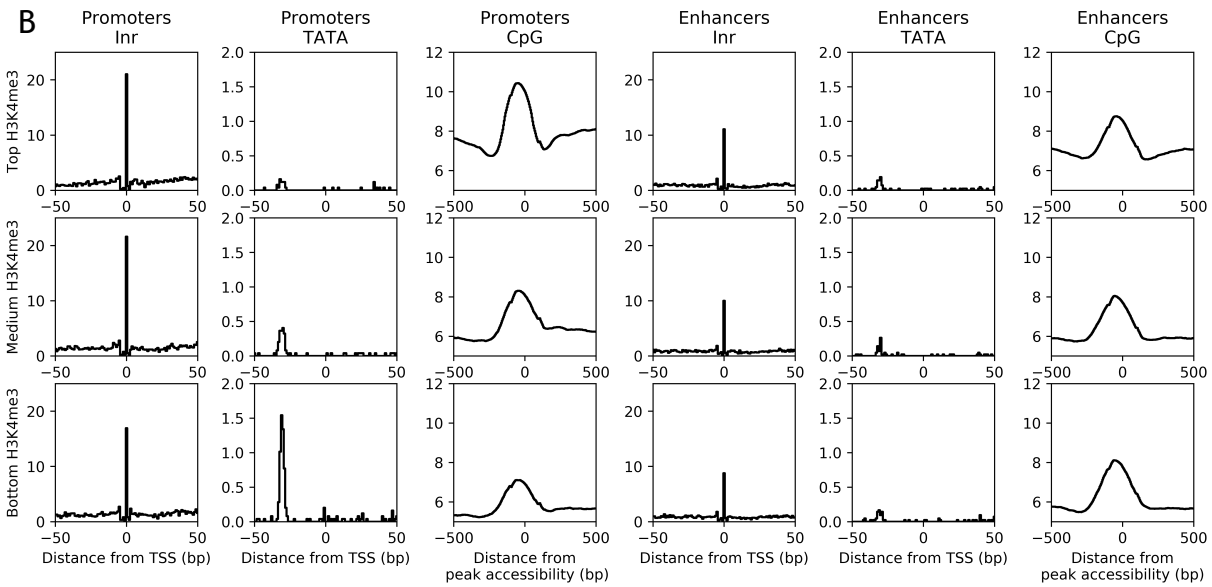


Figure 4

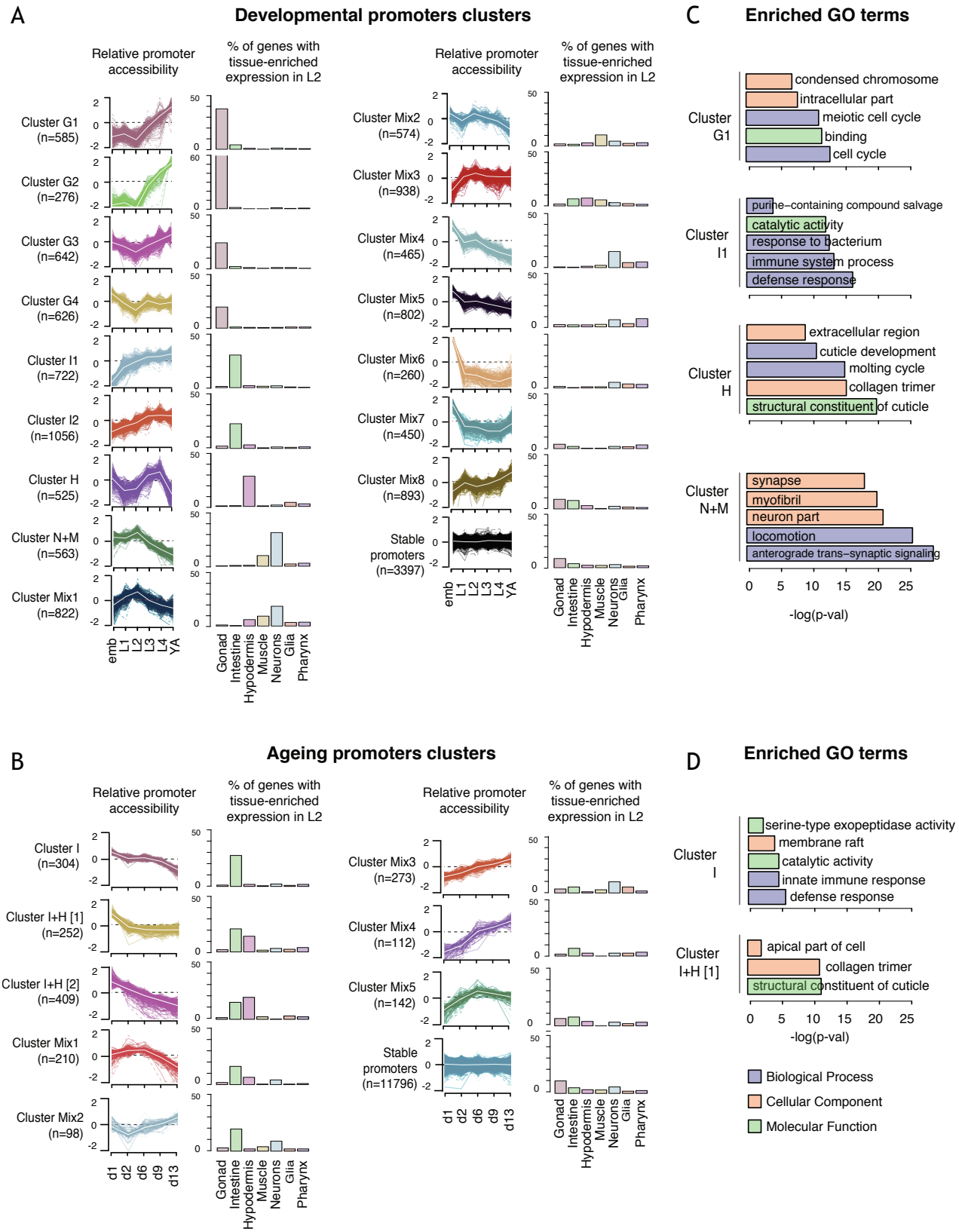


Figure 5

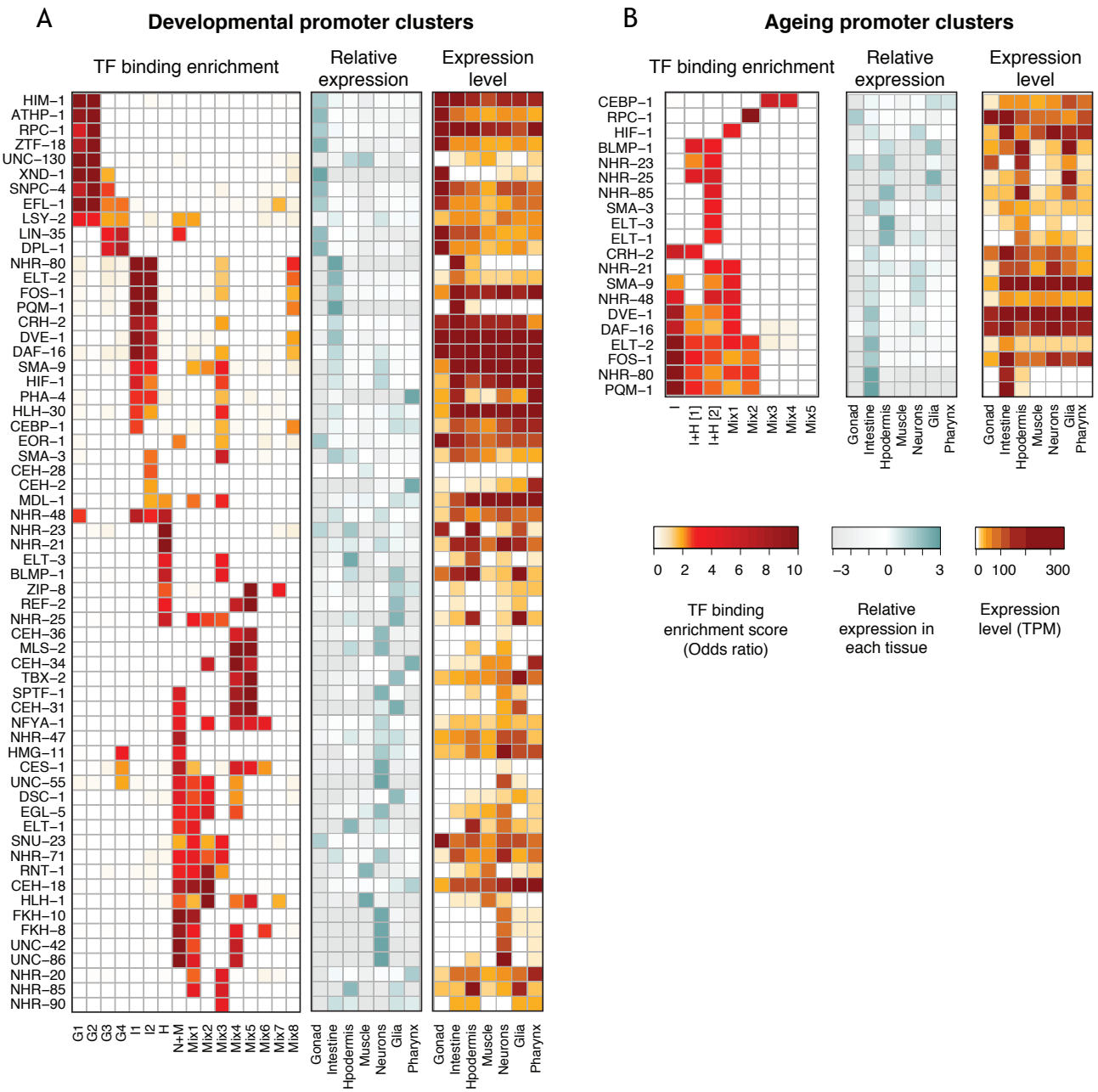


Figure 1—figure supplement 1

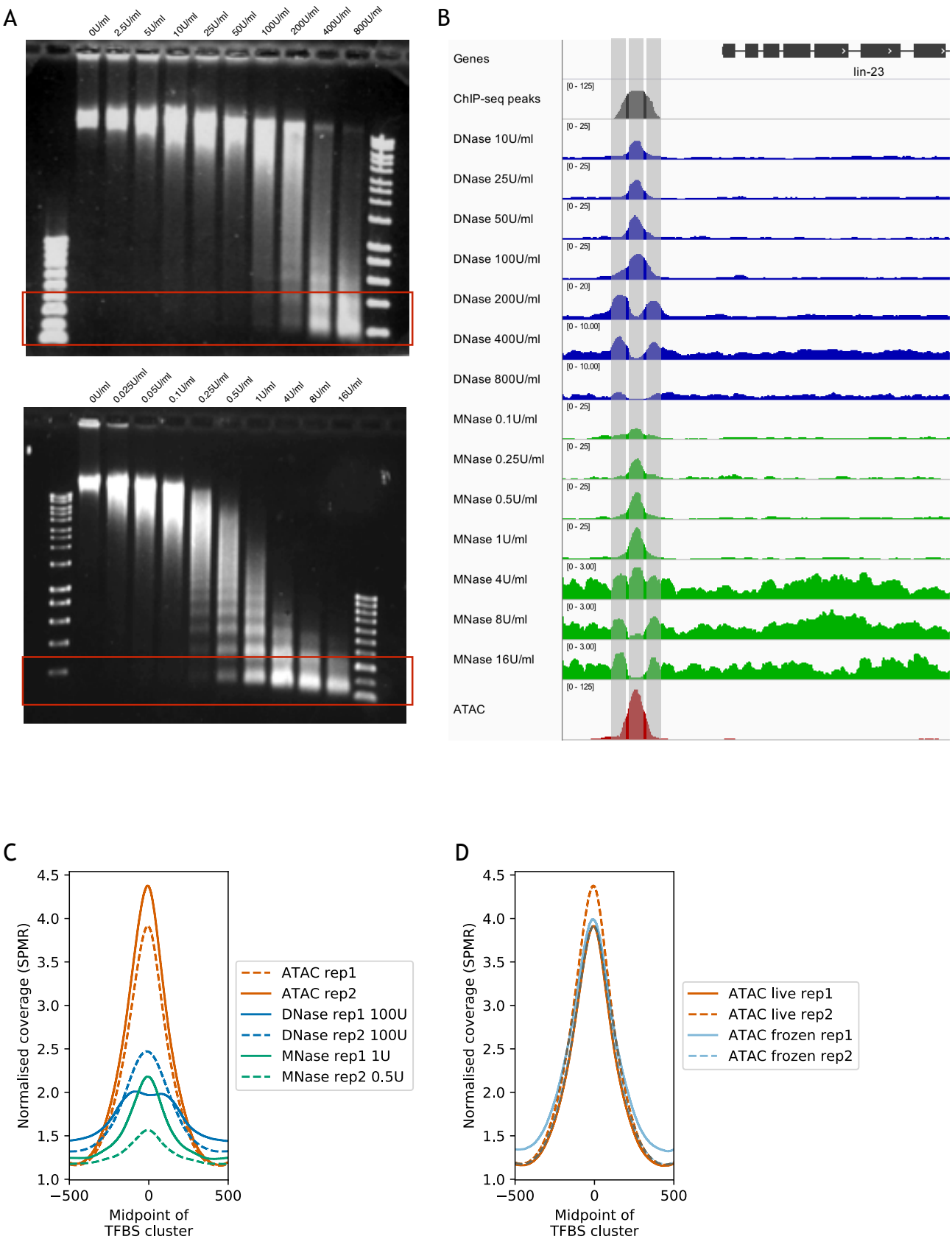


Figure 2—figure supplement 1

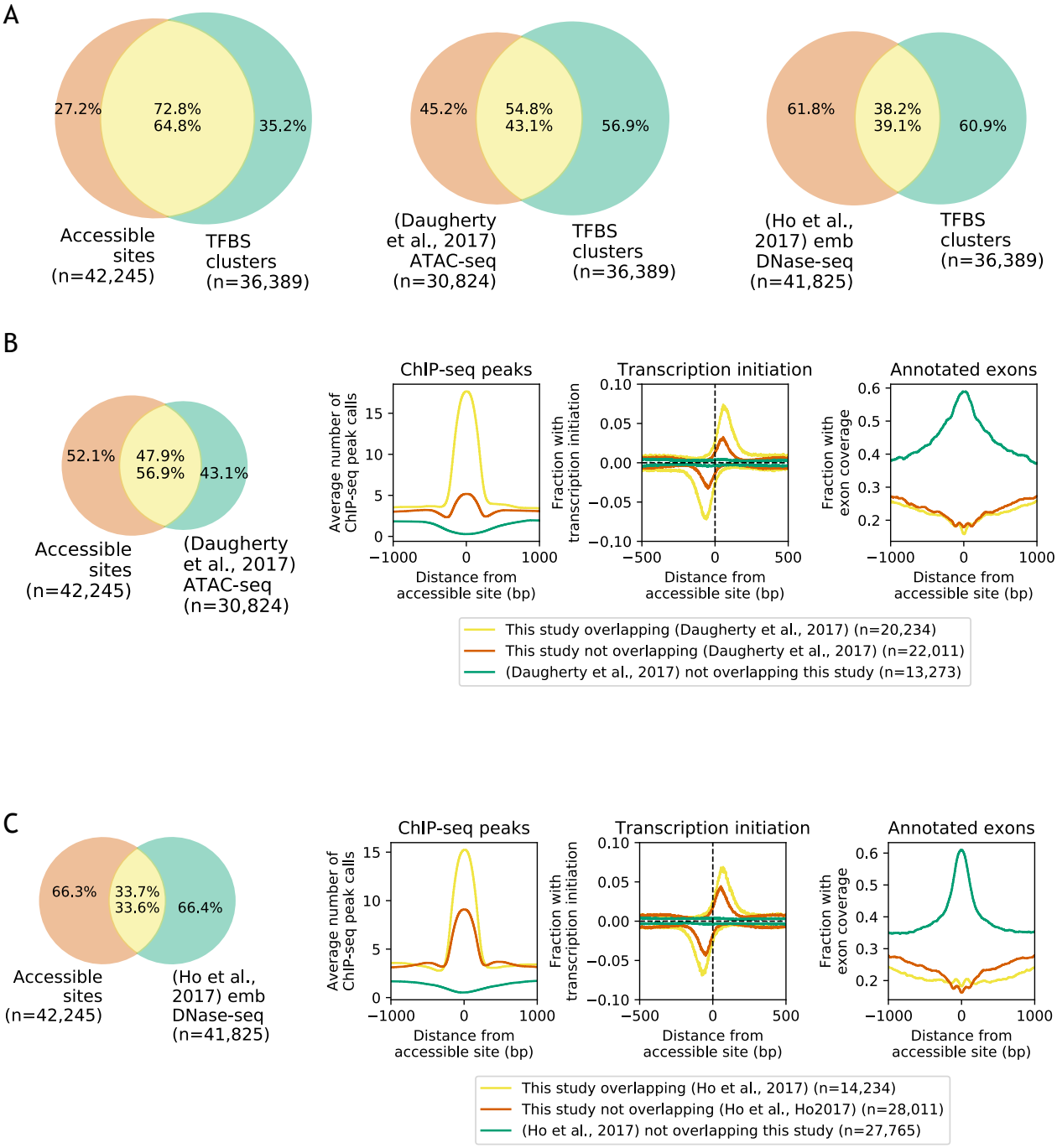


Figure 2—figure supplement 1, continued

D

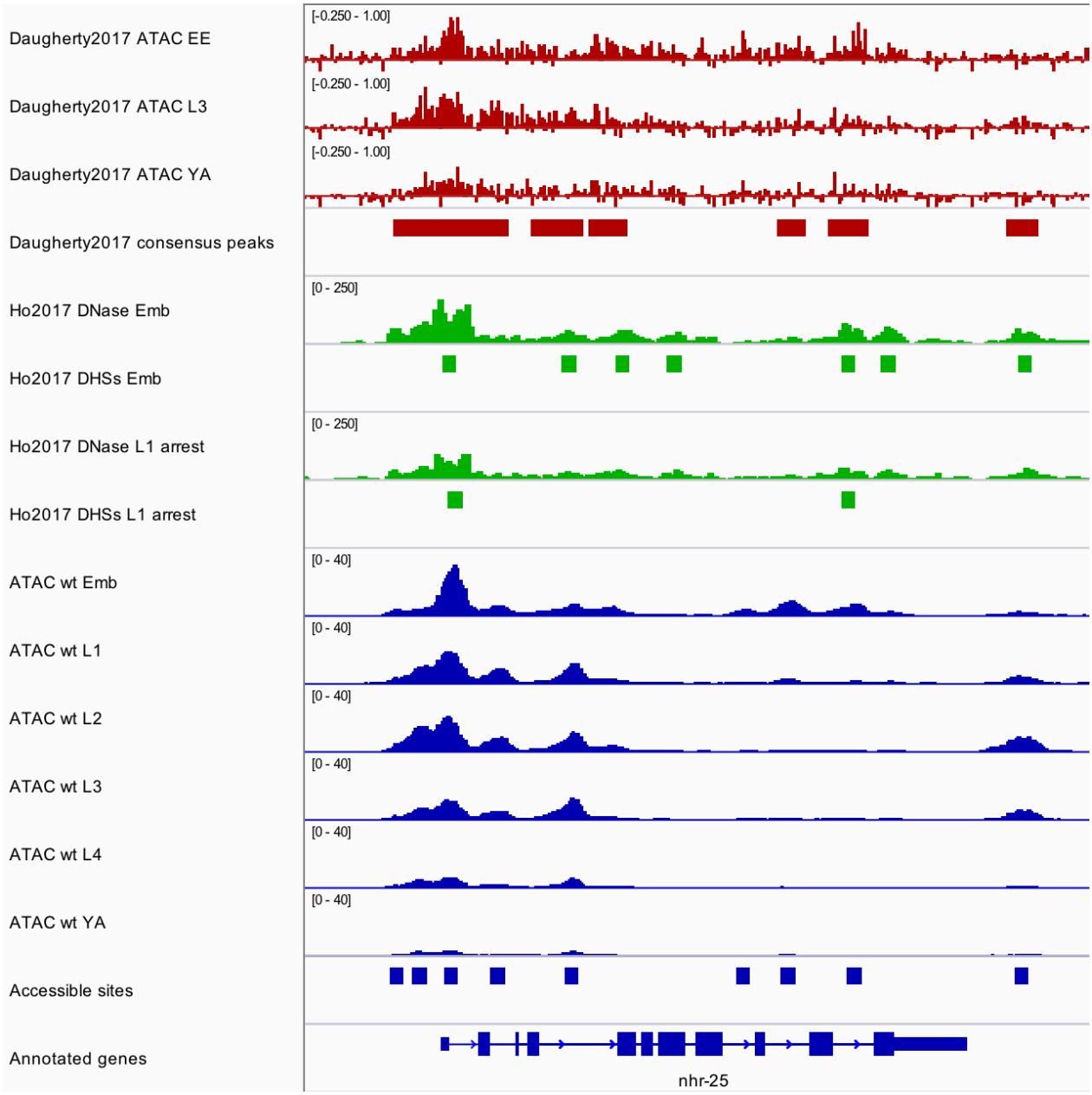


Figure 2—figure supplement 1, continued

E

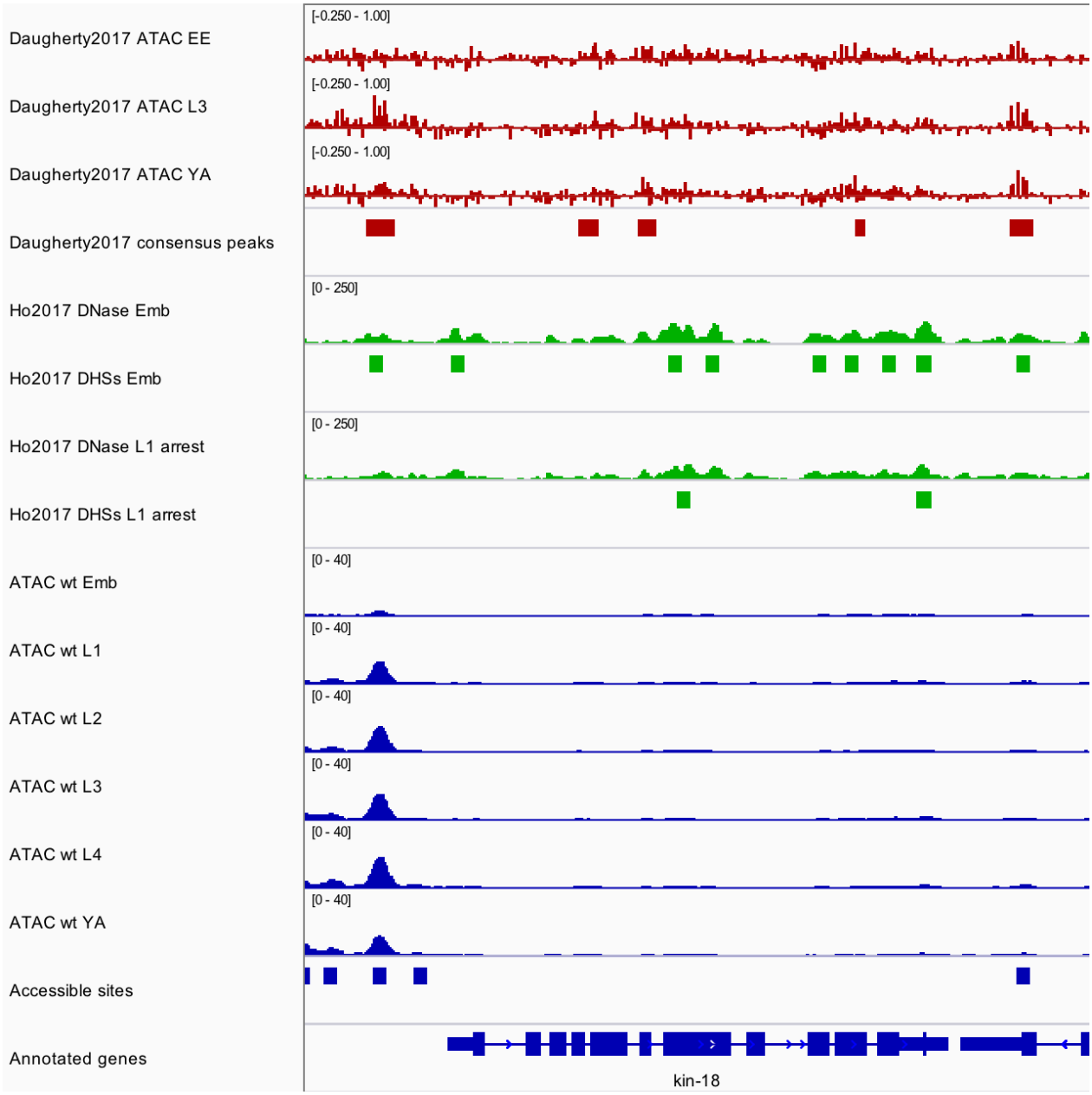


Figure 2—figure supplement 2

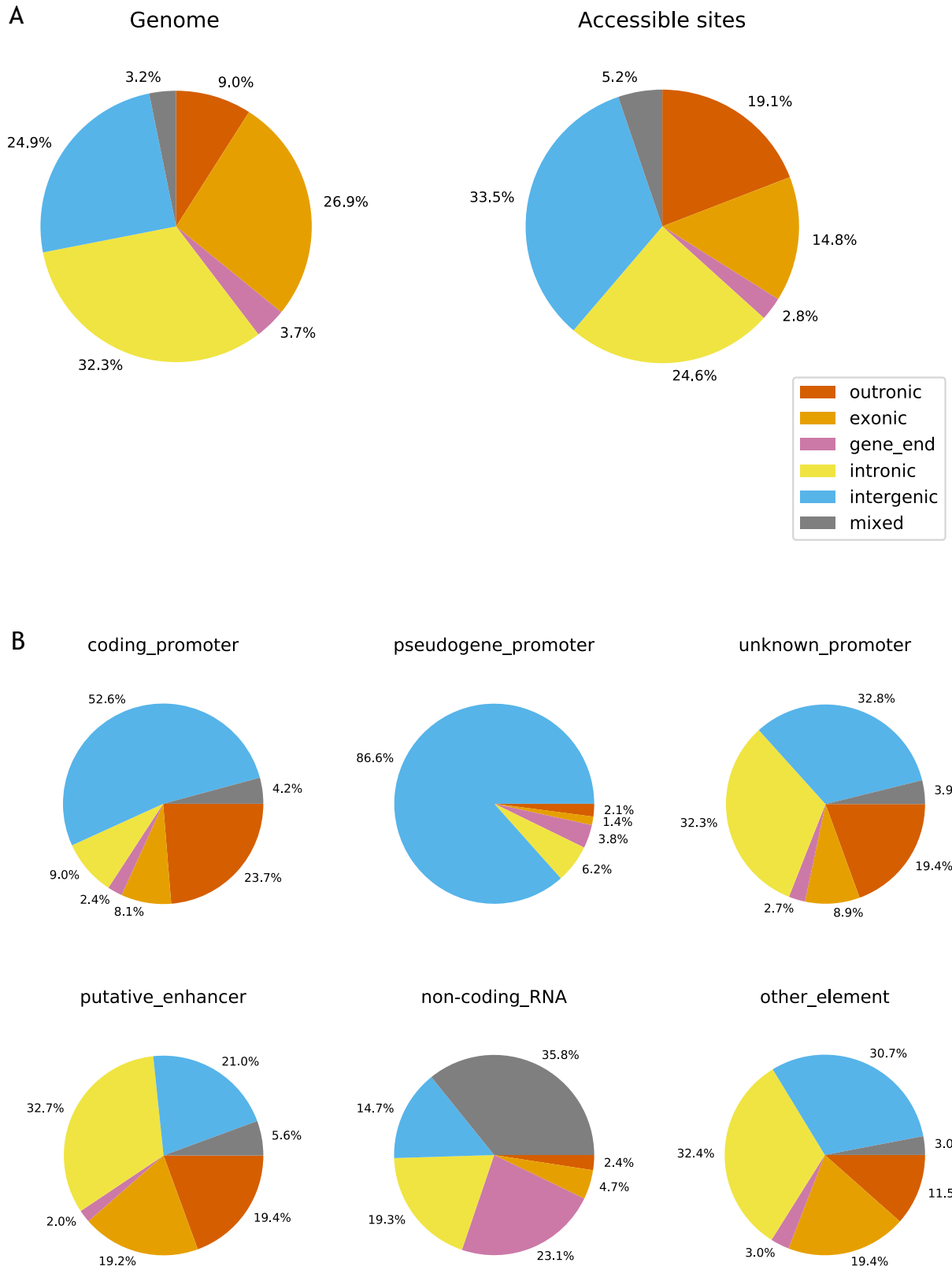


Figure 2—figure supplement 3

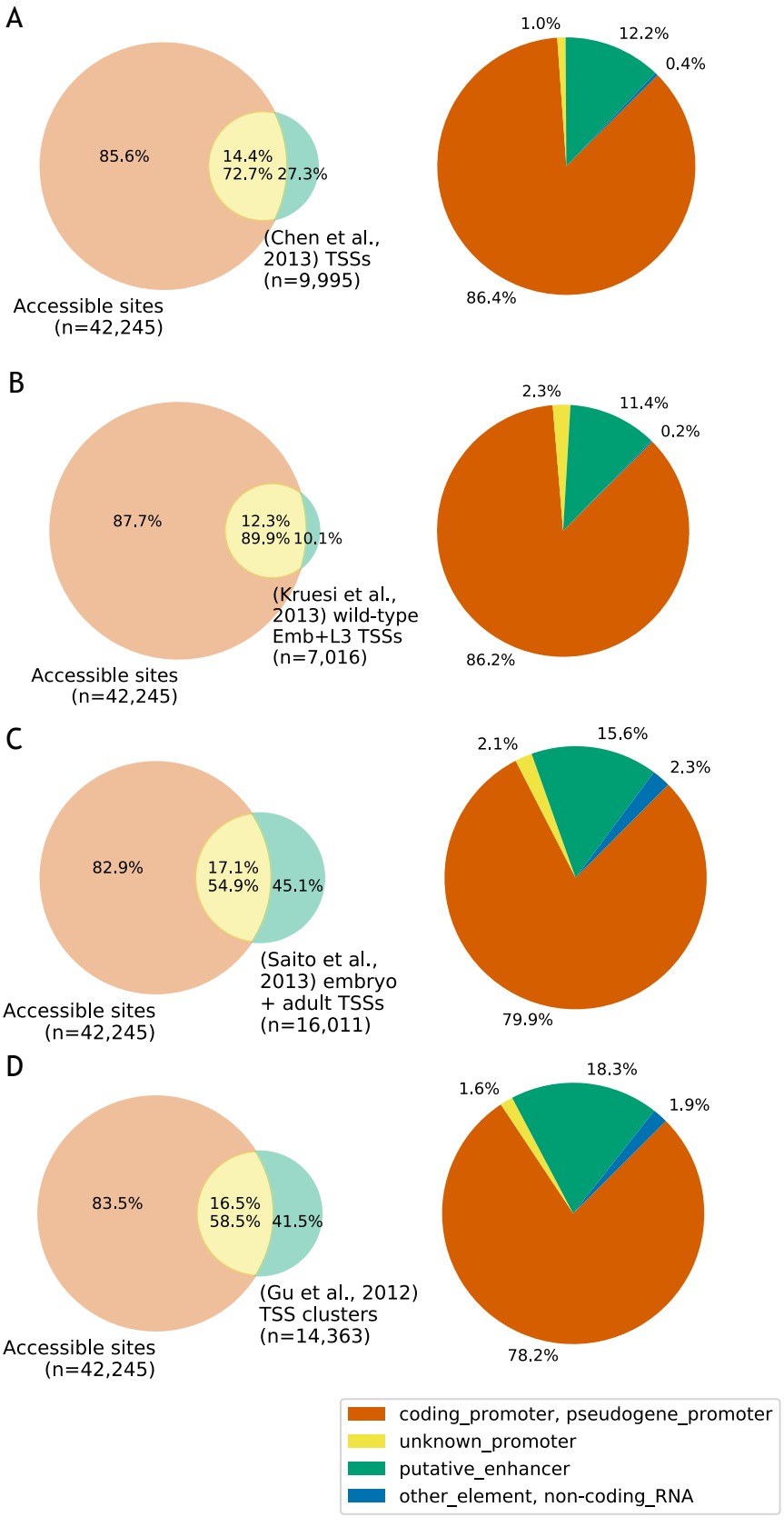
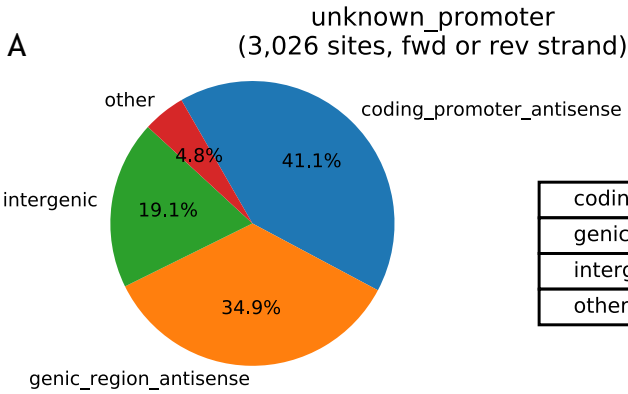


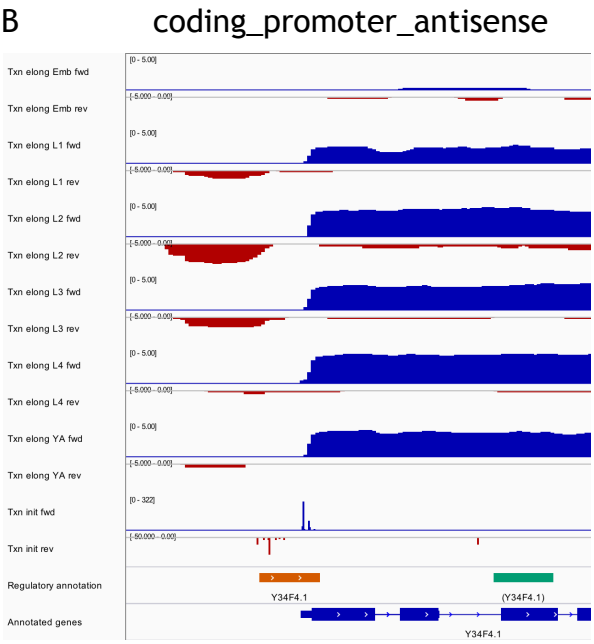
Figure 2—figure supplement 4

A

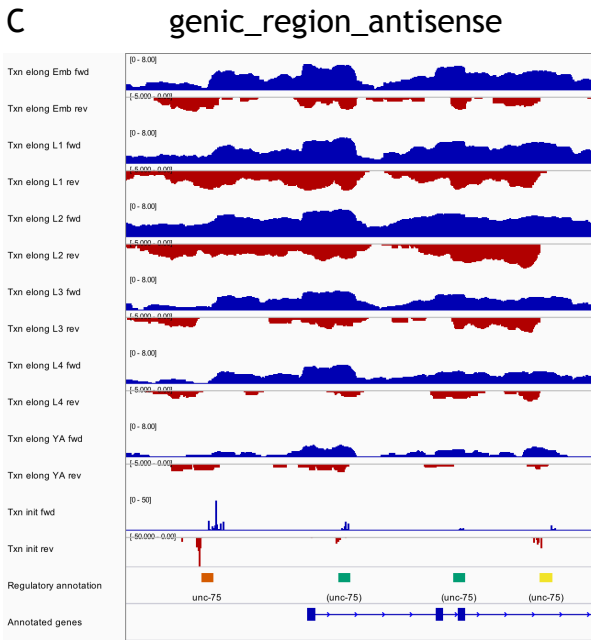


	unknown_promoter_type
coding_promoter_antisense	1244
genic_region_antisense	1057
intergenic	579
other	146

B



C



D

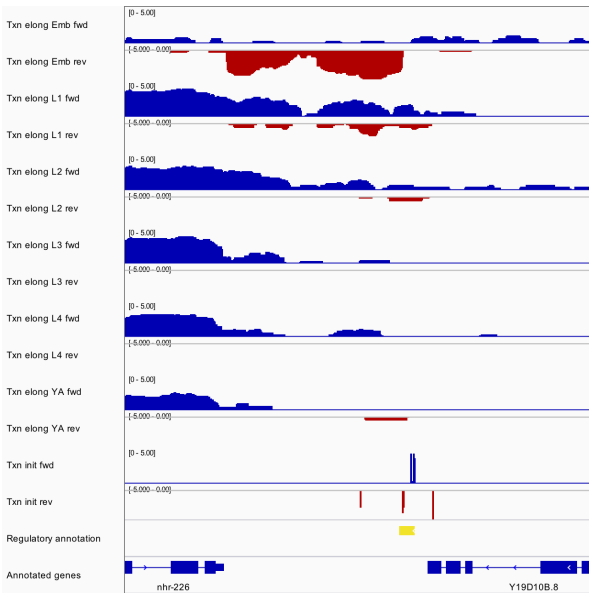
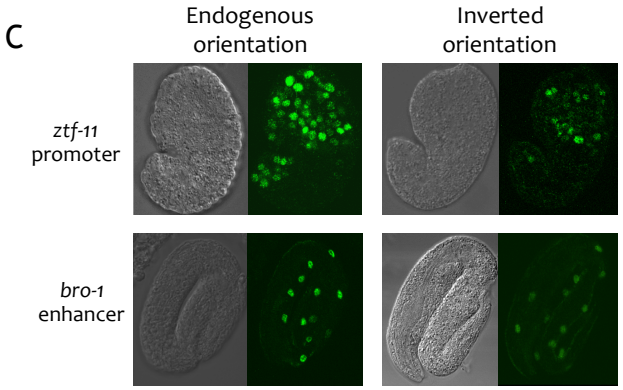


Figure 2—figure supplement 5

A

gene	chr	tested region start	tested region end	length	Annotation	Reference
daz-1	II	5455759	5456289	552	prom	Merritt et al, 2008
gld-1	I	7695164	7695660	527	txn initiation	Merritt et al, 2008
him-3	IV	6977138	6977484	372	prom	Merritt et al, 2008
mex-5	IV	13353211	13353674	486	prom	Merritt et al, 2008
msp-56	IV	9841000	9841495	513	prom	Merritt et al, 2008
pgl-3	V	4755287	4755970	709	prom	Merritt et al, 2008
spe-11	I	5332434	5332187	272	no txn	Merritt et al, 2008
spn-4	V	6784847	6785394	544	prom	Merritt et al, 2008
fbf-1	II	6080676	6081758	1114	prom	Merritt et al, 2008
tag-243	III	4687158	4687357	199	prom	Hunt-Newbury et al, 2007
atp-2	III	5228403	5228199	204	no element	Hunt-Newbury et al, 2007
him-4	X	9717247	9717573	326	no element	Hunt-Newbury et al, 2007
lgc-34	II	526514	526671	157	prom	Hunt-Newbury et al, 2007
daf-5	II	14040184	14040697	514	prom + enh	Chen et al 2014
egrh-1	X	14840974	14841376	403	prom	Chen et al 2014
eif-3b	II	14794905	14795250	346	prom	Chen et al 2014
F58D5.5	I	12039662	12040186	525	prom	Chen et al 2014
hlh-11	III	9619057	9619496	440	prom	Chen et al 2014
mtm-3	III	3801252	3801770	519	prom	Chen et al 2014
rho-1	IV	16492191	16492669	479	unknown promoter	Chen et al 2014
T19C3.4	III	623811	624139	329	prom	Chen et al 2014
otub-2	I	397755	398120	366	prom	Chen et al 2014
Y71H2AM.20	III	2785938	2786194	257	prom	Chen et al 2014

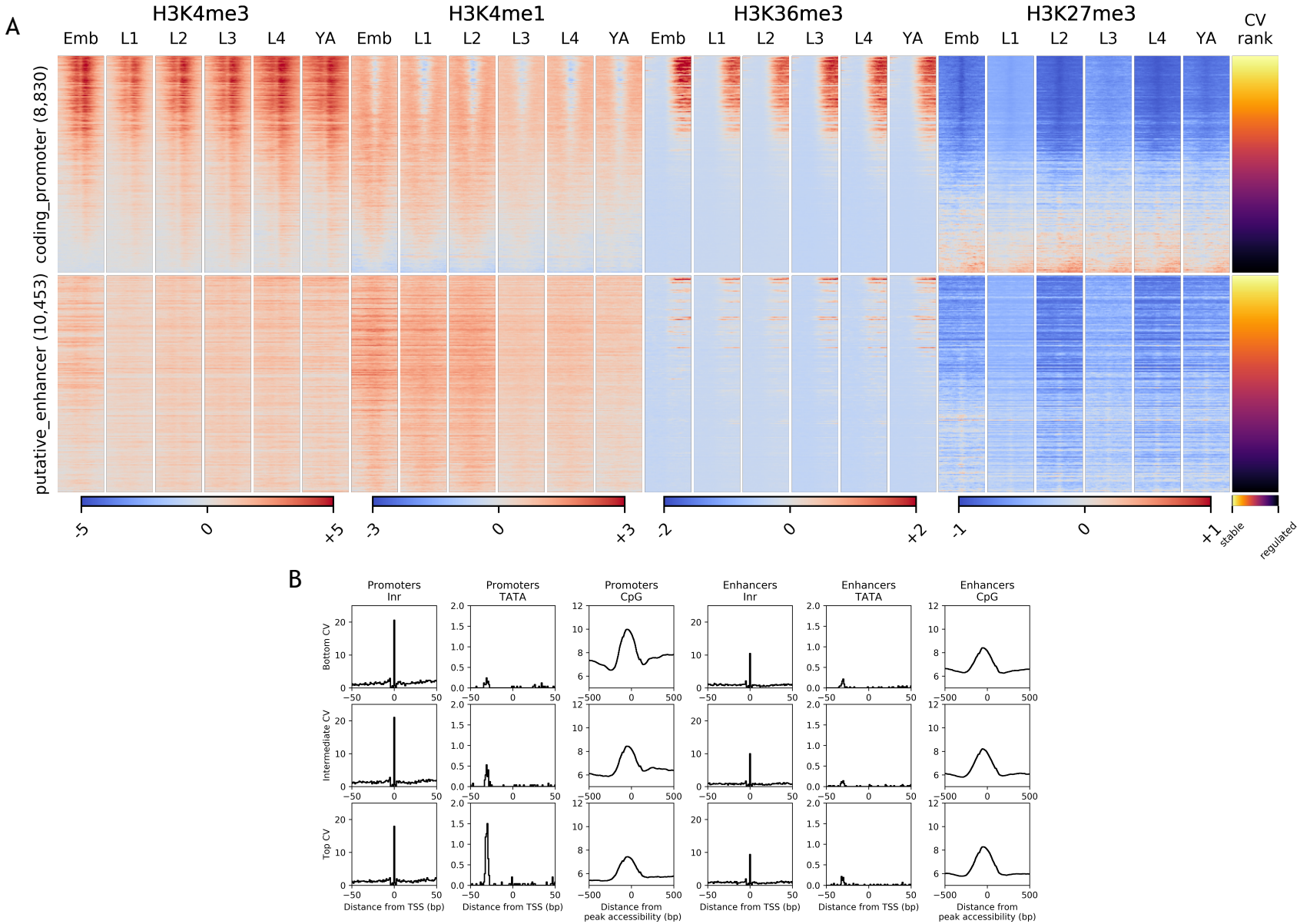
C



B

locus	chr	start	stop	gene strand	annotation	tested orientation relative to gene	Expression	Expression strength (Strong,medium, weak)	Embryo expression	Reference	Strain Name	Genotype
otub-2	I	397755	398120	+	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1613	weSi68 II; unc-119(ed3) III
hlh-2	I	7194716	7195216	-	prom	endogenous	Yes	strong	broad, but not in intestine	this paper	JA1652	weSi89 II; unc-119(ed3) III
ztf-11	I	9910350	9910850	-	prom	endogenous	Yes	strong	broad, but not in intestine	this paper	JA1733	weSi142 II; unc-119(ed3) III
ztf-11	I	9910350	9910850	-	prom	inverted	Yes	medium	broad, but not in intestine (weaker than endogenous direction)	this paper	JA1727	weSi136 II; unc-119(ed3) III
F58D5.5	I	12039662	12040186	+	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1608	weSi63 II; unc-119(ed3) III
F58D5.5	I	12039662	12040186	+	prom	inverted	Yes	strong	broad	this paper	JA1765	weSi145 II; unc-119(ed3) III
daf-5	II	14040184	14040697	-	prom + enh	endogenous	Yes	strong	broad	Chen et al, 2014	JA1620	weSi72 II; unc-119(ed3) III
eif-3b	II	14794905	14795250	+	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1601	weSi61 II; unc-119(ed3) III
T19C3.4	III	623811	624139	-	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1612	weSi67 II; unc-119(ed3) III
Y71H2AM.20	III	2785938	2786194	-	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1600	weSi60 II;unc-119(ed3) III
mtm-3	III	3801252	3801770	-	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1610	weSi65 II; unc-119(ed3) III
hlh-11	III	9619057	9619496	+	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1609	weSi64 II; unc-119(ed3) III
bed-3	IV	9915158	9915437	+	prom	endogenous	Yes	weak	~20 hypodermal cellls	this paper	JA1712	weSi128 II; unc-119(ed3) III
rho-1	IV	16492191	16492669	-	unknown promoter	endogenous	Yes	strong	broad	Chen et al, 2014	JA1611	weSi66 II; unc-119(ed3) III
egrh-1	X	14840974	14841376	+	prom	endogenous	Yes	strong	broad	Chen et al, 2014	JA1614	weSi69 II; unc-119(ed3) III
bro-1	I	5187416	5187609	+	enh-intronic	endogenous	Yes	medium	seam cells	this paper	JA1708	weSi130 II ; unc-119(ed3) III
bro-1	I	5187416	5187609	+	enh-intronic	inverted	Yes	weak	seam cells	this paper	JA1737	weSi146 II; unc-119(ed3) III
hlh-2	I	7195757	7196257	-	enh	endogenous	Yes	medium	few anterior cells	this paper	JA1674	weSi108 II; unc-119(ed3) III
hlh-2	I	7196571	7197071	-	enh	endogenous	Yes	weak	10-20 anterior cells	this paper	JA1663	weSi97 II; unc-119(ed3) III
hlh-2	I	7197592	7198092	-	enh	endogenous	Yes	medium	few hypodermal cells	this paper	JA1669	weSi103 II; unc-119(ed3) III
hlh-2	I	7197592	7198092	-	enh	inverted	Yes	medium	few hypodermal cells (weaker than endogenous orientation)	this paper	JA1701	weSi123 II; unc-119(+)]III
hlh-2	I	7198146	7198646	-	enh	endogenous	Yes	medium	few pharyngeal cells	this paper	JA1671	weSi105 II; unc-119(ed3) III
hlh-2	I	7198751	7199251	-	enh	endogenous	Yes	medium	few pharyngeal cells	this paper	JA1673	weSi107 II; unc-119(ed3) III
ztf-11	I	9911508	9912008	-	enh	endogenous	yes	weak	broad, most cells	this paper	JA1667	weSi101 II; unc-119(ed3) III
bed-3	IV	9910252	9910601	+	enh	endogenous	Yes	medium	few anterior cells	this paper	JA1703	weSi125 II; unc-119(ed3) III
bed-3	IV	9911938	9912226	+	enh	endogenous	Yes	weak	~15 nuclei, head and tail	this paper	JA1711	weSi133II; unc-119(ed3) III
bed-3	IV	9913192	9913425	+	enh	endogenous	no		no expression	this paper	JA1704	weSi126 II: unc-119(ed3) III
bed-3	IV	9914205	9914396	+	enh	endogenous	no		no expression	this paper	JA1705	weSi127 II; unc-119(ed3) III
bed-3	IV	9918007	9918207	+	enh-intronic	endogenous	Yes	weak	few hypodermal cells	this paper	JA1707	weSi129 II; unc-119(ed3) III

Figure 3—figure supplement 1



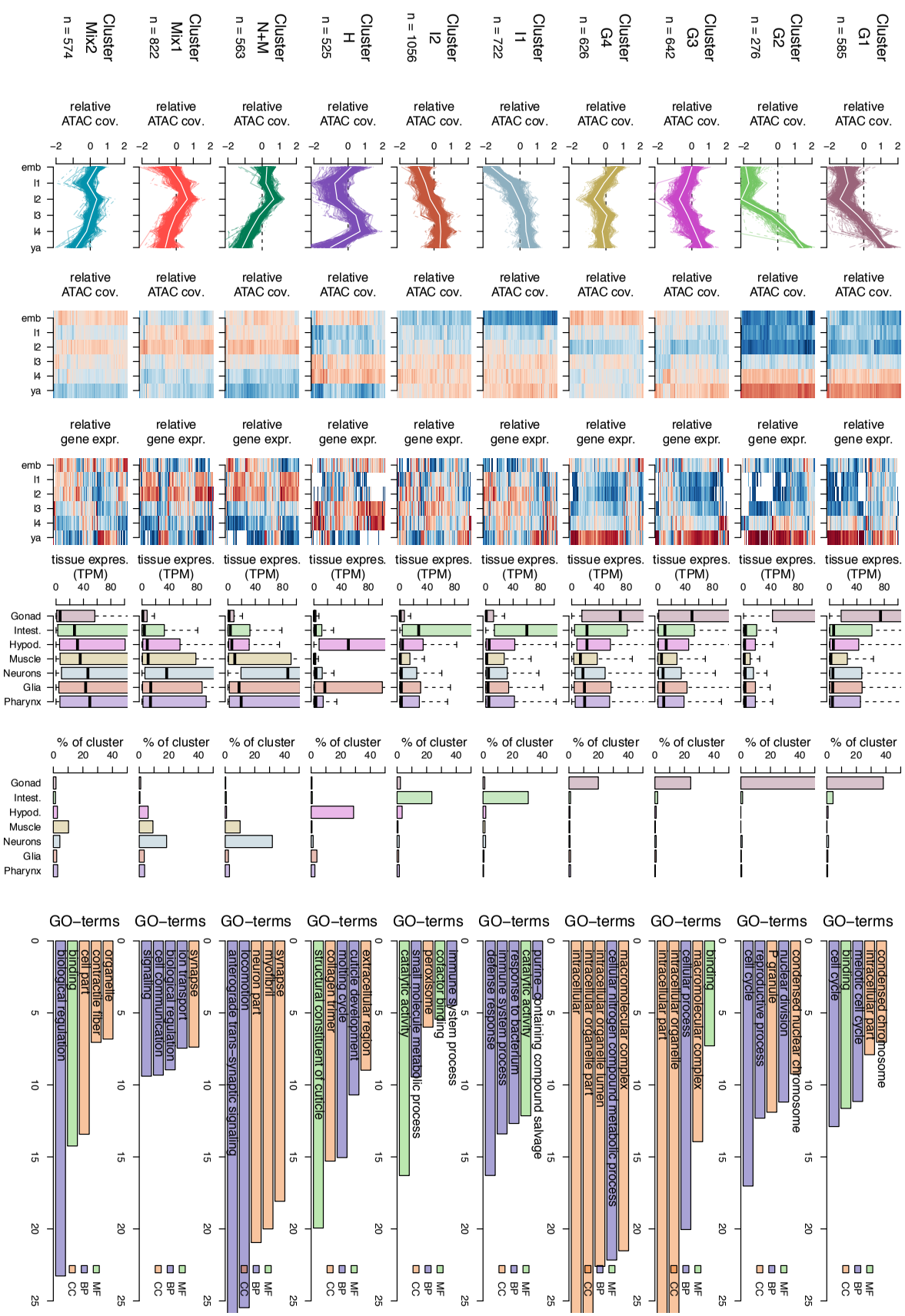


Figure 4—figure supplement 1

Figure 4—figure supplement 1, continued

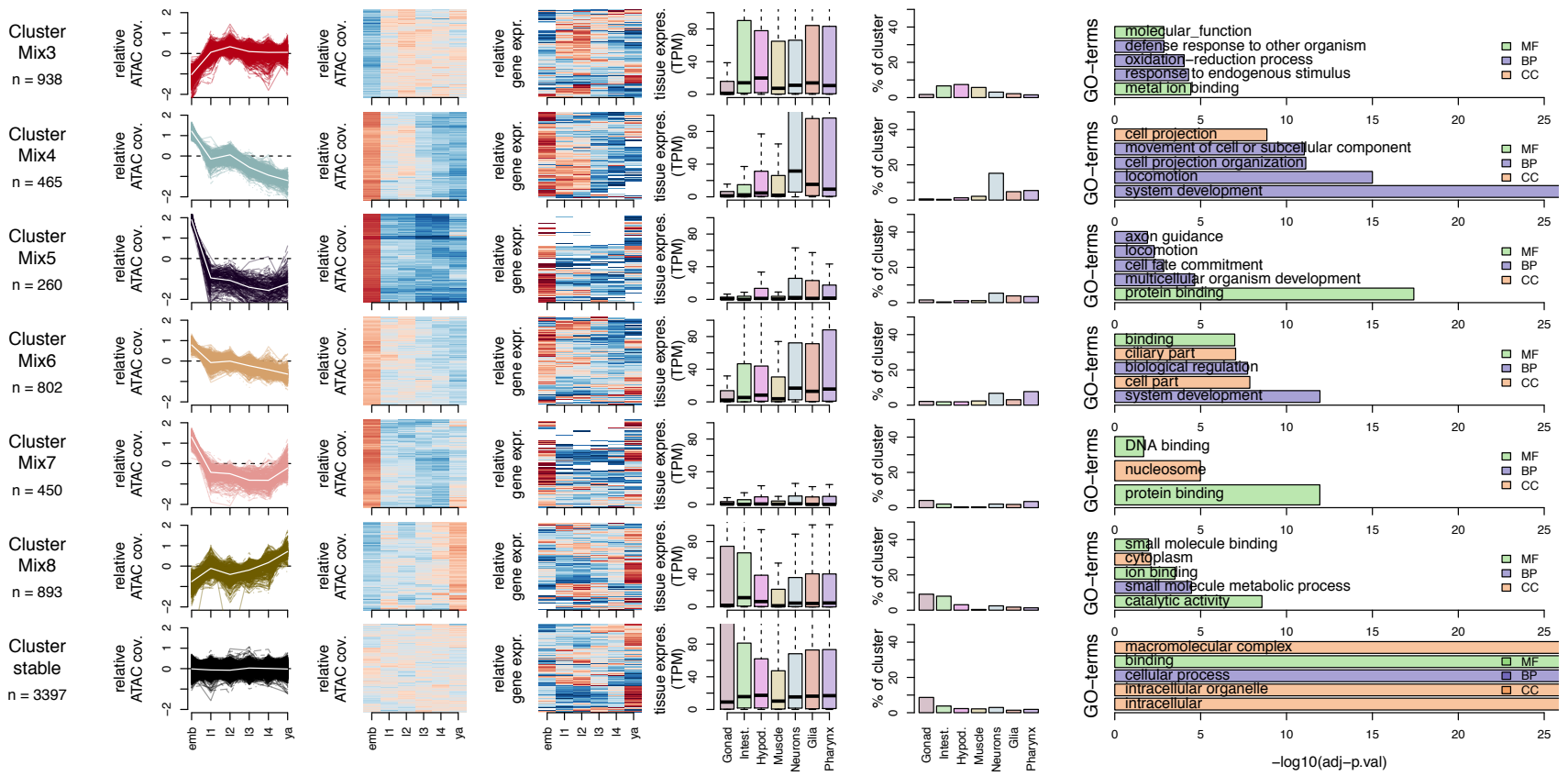


Figure 4—figure supplement 2