Tissue-specific transcription factor target identification in the *Caenorhabditis elegans* epidermis using targeted DamID

Dimitris Katsanos[1] and Michalis Barkoulas[1]

[1]Department of Life Sciences, Imperial College, London SW7 2AZ, United Kingdom.

Corresponding Author: m.barkoulas@imperial.ac.uk

Running title: LIN-22 and NHR-25 targets in the *C. elegans* epidermis

Summary

Transcription factors are key orchestrators of development in multicellular animals and display complex patterns of expression, as well as tissue-specific binding to targets. However, our ability to map transcription factor-target interactions in specific tissues of intact animals remains limited. We introduce here targeted DamID (TaDa) as a method to identify transcription factor targets with tissue-specific resolution in *C. elegans*. We focus on the epidermis as a paradigm and demonstrate that TaDa circumvents problems with Dam-associated toxicity and allows reproducible identification of putative targets. Using a combination of TaDa and single-molecule FISH (smFISH), we refine the positions of LIN-22 and NHR-25 within the epidermal gene network. We reveal direct links between these two factors and the cell differentiation programme, as well as the Wnt signalling pathway. Our results illustrate how TaDa and smFISH can be used to dissect the architecture of tissue-specific gene regulatory networks.

Keywords: DamID, transcription factors, LIN-22, NHR-25, seam cells, epidermis, stem cells, *C. elegans,* tissue-specific

Highlights:

- TaDa circumvents Dam-associated toxicity by keeping levels of Dam expression low.
- TaDa allows the recovery of tissue-specific methylation profiles representing TF binding.
- Methylation signal is enriched in regulatory regions of the genome.
- LIN-22 and NHR-25 targets reveal a link to cell differentiation and Wnt signalling.

Introduction

Development is a tightly regulated process guiding the reproducible formation of complex multicellular organisms starting from single cells. This remarkable transformation requires genetic information to be decoded in a manner that allows cell type diversity to emerge through spatiotemporal control of gene expression. In this process, transcription factors are key players as they participate in gene regulatory networks that receive inputs from internal cues and the external environment to govern robust developmental decisions across systems (Davidson, 2010; Oliveri and Davidson, 2007). Therefore, elucidating the architecture of developmental gene networks by identifying the participating transcription factors, as well as their biologically relevant targets, is fundamental for increasing our mechanistic understanding of how specialised cells and tissues develop.

An important cell type in the *C. elegans* epidermis is the seam cells, which follow a stem cell-like pattern of symmetric and asymmetric divisions throughout larval development to produce most epidermal nuclei (Chisholm and Hsiao, 2012; Sulston and Horvitz, 1977). Proliferative symmetric divisions occur at the second larval stage and increase the stem cell pool, whereas asymmetric cell divisions re-iterate at each larval stage to generate differentiated hypodermal or neuronal cells, while maintaining the seam cell number through self-renewal (Sulston and Horvitz, 1977). As in other stem cell contexts, Wnt signalling plays a key role in seam cell maintenance and the regulation of seam cell division patterns (Gleason and Eisenmann, 2010). In the canonical Wnt pathway, β-catenin is targeted for degradation in the absence of Wnt ligand binding to receptors. Upon Wnt receptor activation, β-catenin is stabilised, enters to the nucleus and, along with the transcription factor POP-1/T-cell factor/lymphoid enhancer factor (TCF/LEF), activates Wnt target genes (Sawa and Korswagen, 2013). The Wnt/β-catenin asymmetry (WβA) is a modified version of the Wnt pathway adapted for the purposes of asymmetric cell division, where asymmetric distribution of Wnt pathway components polarises the mother cell and leads to asymmetric inheritance of the potential for Wnt pathway activation in the two daughter cells (Lam and Phillips, 2017).

Besides Wnt signalling, a number of conserved transcription factors have also been identified influencing cell fate decisions in the epidermis. Due to the highly tractable nature of *C. elegans,* this highlights the potential value of this model to reveal fundamental mechanisms regulating the balance between cell proliferation and differentiation (Joshi et al., 2010) . For example, RNT-1 is the *C. elegans* homolog of Runx and has been shown to act together with its binding partner BRO-1/ CBFβ to promote proliferative/symmetric seam cell divisions (van der Horst et al., 2019; Kagoshima et al., 2007). Mutations in Runx genes and CBFβ are known to cause various leukaemias in humans (Cameron and Neil, 2004), and mutations or overexpression in *C. elegans* leads to seam cell loss or tissue hyperplasia respectively (Nimmo et al., 2005). Another example is the Hairy/Enhancer of split (*Hes*)-related bHLH transcription factor LIN-22, which acts to suppress neurogenesis in seam cells (Wrischnik and Kenyon, 1997) and maintain correct division asymmetry by antagonising the Wnt signalling pathway (Katsanos et al., 2017). The GATA family of transcription factors is linked to various types of cancer in humans (Zheng and Blobel, 2010), and related factors in *C. elegans*, such as ELT-1 and EGL-18, are thought to specify seam cell fate (Gorrepati et al., 2013; Koh and Rothman, 2001; Smith et al., 2005). Finally, transcription factors of the nuclear hormone receptor family have also been reported to be play a role in epidermal development (Antebi, 2006; Miyabayashi et al., 1999). A key example is NHR-25, which has been shown to act in the establishment of cell contacts, while it acts more broadly in vulval development, molting and neurogenesis in the T lineage (Chen et al., 2004; Gissendanner and Sluder, 2000; Hajduskova et al., 2009; Hayes et al., 2006; Shao et al., 2013; Šilhánková et al., 2005).

Despite the increase in our knowledge of individual TFs playing a role in epidermal development, we still have a limited understanding of the underlying gene network and how this drives the decision between cell proliferation and differentiation. This is largely due to the fact that the majority of the known interactions are not direct and remain at the genetic level. Identification of transcription factor targets has been predominantly pursued by ChIP

experiments in *C. elegans* (Araya et al., 2014; Celniker et al., 2009; Kudron et al., 2018), with only one example using DamID-chip to identify targets of DAF-16 (Schuster et al., 2010). Furthermore, an additional challenge in multicellular organisms is to characterise DNA-protein interactions with tissue-specific resolution. Targeted DamID (TaDa) allows tissue-specific target identification (Southall et al., 2013) and has been successfully employed in *Drosophila* to dissect mechanisms of neuronal fate determination (Sen et al., 2019; Vissers et al., 2018), as well as in mammalian stem cell lines to characterise the binding profile of pluripotency factors (Cheetham et al., 2018). TaDa offers tissue-specificity while overcoming the major limitation of conventional DamID, which relates saturated methylation due to high levels of Dam expression (van Steensel and Henikoff, 2000; van Steensel et al., 2001). This obstacle is circumvented in TaDa with the introduction of a primary ORF of *mCherry* followed by 2 STOP codons and a frame-shift preceding the Dam-fusion. Due to the universal property of eukaryotic ribosomes to reinitiate translation after a STOP codon at a reduced frequency (Kozak, 1987, 2001), this transgene design results in low levels of expression of the Dam-TF fusion in the tissue of interest (Figure 1A). We introduce here TaDa in *C. elegans* and demonstrate that this method is powerful in revealing TF targets within a tissue of interest. We employ TaDa to predict and study targets of LIN-22 and NHR-25 to improve our understanding of the gene network underlying epidermal cell fate decisions.

Results

TaDa circumvents Dam-associated toxicity and allows the recovery of TF-specific methylation profiles

To investigate whether TaDa can be used to identify TF targets in *C. elegans* tissues, we focused on the epidermis. We chose to profile two TFs, the bHLH *Hes*-related factor LIN-22 and the nuclear hormone receptor NHR-25, which are both key players in epidermal development (Chen et al., 2004; Gissendanner and Sluder, 2000; Wrischnik and Kenyon, 1997). We chose LIN-22 because our previous work implicated this factor in Wnt-dependent seam cell patterning, although its direct targets remain unknown (Katsanos et al., 2017;

Wrischnik and Kenyon, 1997). On the other hand, NHR-25 targets had been studied before at the whole animal level using ChIP-seq, which we reasoned that it would facilitate comparisons across different methodologies (Araya et al., 2014; Shao et al., 2013). We assembled TaDa transgenes under the *wrt-2* promoter, which drives expression primarily in the seam cells. We included a *C. elegans* optimised *mCherry* as the primary ORF, followed by 2 STOP codons and an extra nucleotide for frame-shift before the TF-Dam (*lin-22:dam* and *nhr-25:dam)* or control (*NLS-GFP:dam)* fusions (Figure 1B). We also designed a versatile vector for use by the community that contains all these key features in a backbone that is compatible for single-copy transgenesis via universal MosSCI (Frokjaer-Jensen et al., 2014). This vector contains an *attR4-attL1* Gateway cloning site to allow insertion of any promoter of choice, as well as restrictions sites to allow the assembly of N- or C- terminal fusions of the protein of interest with Dam (Figure S1A).

Induced ubiquitous Dam expression by a heat-shock promoter has been shown to produce saturated methylation in *C. elegans* (Schuster et al., 2010; van Steensel and Henikoff, 2000), but the effects of tissue-specific expression of Dam-fusions on methylation levels and toxicity had not been previously investigated. To this end, we constructed *lin-22:dam* and *NLS-GFP:dam* fusions with and without *mCherry* as the primary ORF (Figure 1B), and inserted these into the genome as single-copy transgenes. We found that transgenic lines lacking *mCherry* showed developmental defects, such as a dumpy and exploding vulva phenotype, as well as significantly reduced brood size compared to animals containing the *mCherry* primary ORF, which appeared comparable to wild-type controls (Figure 1C, D). These aberrant phenotypes were likely due to high levels of Dam expression as they were observed both in *lin-22:dam* and *NLS-GFP:dam* fusions without the primary ORF, and correlated with increased amount of DNA methylation (Figure S1B). We conclude that the TaDa transgene configuration reduces toxicity in *C. elegans*, as previously reported in other systems as well (Southall et al., 2013; van Steensel and Henikoff, 2000).

We first tested the single-copy transgenics harbouring TaDa transgenes for expression, methylation capacity and functionality of the fusion. The expression of *mCherry* was used as proxy to confirm the spatial localisation of the Dam-fusions. Microscopy at the L4 stage revealed low *mCherry* expression in the seam cells, as expected for single copy transgenes (Figure S2A). The presence of distinct smears of amplification products from recovered methylated DNA indicated the methylation capacity of all fusions (Figure S2B, C). TF-Dam fusions were tested for functionality in a multi-copy transgenesis assay. We reasoned that despite the low level of expression of the TF-Dam fusions, if they have retained their functionality, they would still be sufficient to rescue loss-of-function mutations or induce ectopic phenotypes when multiple copies are present. Indeed, *lin-22(icb38)* mutants show an increase in the number of postdeirid (PDE) neurons, which was suppressed in transgenics carrying the *lin-22:dam* construct as an extrachromosomal array (*icbEx54)* (Figure S2D). Similarly, the *nhr-25:dam* fusion was sufficient to increase seam cell number as a multi-copy array (*icbEx184*) (Figure S2E).

We sequenced amplicons for both TFs at two developmental stages (L2 and L4) and processed the data using the damidseq pipeline (Marshall and Brand, 2015). We generated normalised aligned read count maps to evaluate replicate reproducibility and assess the level of correlation between samples. High level of reproducibility was observed for all fusions between biological replicates (Figure S3A, B, C) with higher Pearson correlation coefficients for comparisons within a given *TF:dam* fusion. Importantly, control samples were found to cluster separately from TF-Dam samples (Figure S3A, B, C). Separate clustering of the *lin-22:dam* and *nhr-25:dam* samples was found by Pearson correlation and principal component analysis (Figure S3C, D, E). These results highlight that TF-Dam fusions produce specific and distinct methylation patterns compared to the non-targeted control fusions.

TaDa signal enrichment profiles map to regulatory regions of the genome

To investigate the profile of signal enrichment across the genome, we compared *TF:dam* and *NLS-GFP:dam* sample pairs and calculated the normalised $\log_2$(*TF:dam/NLS-GFP:dam*)

ratio scores per GATC fragment based on normalised aligned read count maps (Figure S3F). To address whether the observed signal enrichment is likely to reflect TF binding, we focused on putative targets of LIN-22 and NHR-25, based on previous genetic analysis or ChIP-seq experiments respectively. We observed significant enrichment in the *lin-22:dam* profiles near the *frizzled* receptor *lin-17*, the Hox gene *mab-5* and the *C. elegans achaete-scute* homologue *lin-32*, all reported to be genetic interactors of *lin-22* (Katsanos et al., 2017; Wrischnik and Kenyon, 1997) (Figure 2A). With regard to the *nhr-25:dam* profiles, significant enrichment was found in regions near the target genes *idh-1* and *rpl-3,* as well as the *nhr-25* locus itself, supporting previous evidence for self-regulation (Shao et al., 2013) (Figure 2B).

Qualitative inspection of the enrichment profiles suggested a preference for enrichment in intergenic regions (Figure 2A, B), which are likely to harbour *cis*-regulatory sequences for TFs to bind. Indeed, aggregate genome-wide signal profiles showed increased average enrichment scores in upstream sequences directly proximal to the transcriptional start site (TSS) of protein coding genes (Figure 2C). Statistical processing of the signal profiles was used to identify peaks that may harbour TF target sites. In particular, 1965 and 1972 peaks were identified for *lin-22:dam*, while 2044 and 2169 peaks were found for *nhr-25:dam,* in both cases at the L2 and L4 stage respectively (Table S1). Hierarchical clustering of the localisation and score of those peaks that lie between 5 kb upstream and 1kb downstream of genes, was consistent with a preference for signal enrichment in upstream to TSS regions (Figure 2D). The localisation of the peaks in relation to genes was further dissected by assigning peaks to their single closest gene. For all profiles, the majority of the peaks (>94%) were assigned to genes, with the largest proportion of peaks (between 41.2% and 47.3%) found to be upstream to or overlapping the TSS of their assigned gene (Figure 2E). From the peaks classified as upstream of genes, approximately half (46% - 54%) localised within the first 2 kb upstream. A significant proportion of peaks were found to be within genes, but around a quarter of those (24.3% - 28.5%) were exclusively residing within introns, which are known to contain regulatory elements in *C. elegans* (Fuxman Bass et al., 2014) (Figure 2E).

In stark contrast, a very small percentage of peaks, (4.2% - 9%) were exclusively within exons, despite exons having a median size that is double compared to that of introns (Spieth et al., 2014).

To consolidate the link between the identified peaks and regulatory regions of the genome, we studied the overlap between our data and open chromatin signal tracks from ATAC-seq experiments performed on whole-animal L2 and L4 *C. elegans* (Jänes et al., 2018). TaDa peaks for both TF-Dam fusions showed increased average chromatin openness compared to neighbouring regions (Figure 3A), with the majority of the peaks showing direct overlap with regions of increased chromatin accessibility (Figure 3B). The genome-wide overlaps were highly significant ($p << 0.0001$ by Monte Carlo simulations) with only a small fraction (11% - 27.6%) of the total identified peaks for each TF and developmental stage not overlapping with an open chromatin element (Figure 3C). Instead, the majority of peaks were overlapping either coding promoters (35.2% - 46.7%) or putative enhancers (48% - 61.2%) (Figure 3C). Taken together, these findings demonstrate the intergenic localisation of peak enrichment at the genome-wide level, and highlight the regulatory potential of the identified regions as putative TF binding sites.

Comparisons of peak localisation profiles across methods and transcription factors

To evaluate the TF target identification capacity of TaDa against other methods, we compared data obtained using the *nhr-25:dam* fusion and two published ChIP-seq datasets from L1 and L2 staged animals from Shao *et al*., 2013 (L1) and Araya *et al*., 2014 (L2). Initial qualitative assessment of the signal tracks around the above-mentioned targets (*nhr-25, idh-1*) showed good agreement, as well as revealed enrichment near epidermal genes, such as the TF *egl-18* (Figure 4A).

On a genome-wide level, aggregate signal of the *nhr-25:dam* L2 and L4 over all ChIP-seq L1 and L2 peaks exhibited strong overlapping localisation (Figure 4B), further supporting the similarity of peak profiles across methods. More specifically, both the L2 and L4 *nhr-*

25:dam peaks showed significant localisation overlaps across the genome with the 683 peaks of the NHR-25 ChIP-seq L1 (based on reanalysis of published data) and the 5980 peaks of the ChIP-seq L2 datasets (Figure 4C), as assessed by Monte Carlo simulations. We found a substantial overlap with the L2 ChIP-seq dataset, with approximately 37% (726 peaks) and 38.5% (835 peaks) of the total L2 and L4 TaDa peaks overlapping respectively. Out of the 726 *nhr-25:dam* L2 peaks overlapping the ChIP-seq L2 peaks, 543 had overlaps occurring with the highest enriched GATC fragment of the peak. In contrast to ChIP-seq L2, only 7.3% and 7.2% of TaDa peaks L2 and L4 were overlapping with ChIP-seq L1 peaks, which is likely to result from higher level of noise in this particular L1 dataset.

We also examined the overlap between the TaDa peak profiles identified here for the two TFs. Interestingly, *lin-22:dam* L2 and L4 signal showed an increased average preference to map onto both L2 and L4 TaDa NHR-25 peaks in comparison to adjacent regions, and vice versa for the *nhr-25:dam* signal profiles (Figure 4D). Since LIN-22 and NHR-25 both participate in epidermal development, it is conceivable that this similarity may reflect proximity or overlap of TF binding sites. Alternatively, the shared peaks may reflect genomic sites where promiscuous binding of multiple transcription factors occurs. Such regions, usually referred to as High-Occupancy Target (HOT) regions have been previously determined in ChIP-seq experiments (Araya et al., 2014). TaDa peaks showed significant ($p << 0.0001$ by Monte Carlo simulations), but small (8% - 19%) overlap with L2 and L4 stage HOT regions, compared to a 34% of overlap between NHR-25 ChIP-seq L2 peaks and HOT regions. Furthermore, out of the 2167 HOT regions found in L2 (Araya et al., 2014) only 292 (13%) were overlapping TaDa NHR-25 L2 peaks, in stark contrast to the 1807 (83%) that were overlapping with NHR-25 ChIP-seq L2 peaks.

To further investigate the overlaps between the TaDa profiles for the two TFs, all the pairwise comparisons of overlapping peaks were made. At the genome wide level all the studied overlaps were found to be non-random by Monte Carlo simulations. Around 60% of peaks of each profile overlapped across the two stages for the same TF, whereas across

TFs overlapping peaks were <33% of the total (Figure 4E left). Overlaps and their significance were re-calculated for restricted promoter regions (defined here as 5 kb upstream to 500 bp downstream of TSS), to avoid inflating *p*-values by interrogating the entire genome even though TF binding sites are expected to localise on promoters. Again, we found promoter overlaps to be highly significant (Figure 4E right) and represented the majority of the genome-wide overlaps (67% - 72%). In addition, less than a third (24.3% - 32.5%) of the across-TF peak overlaps in promoters occurred in HOT regions. Overall, these results strongly support that the observed overlaps are likely due to genuine TF binding largely to the same promoters.

To compare the binding profile of LIN-22 and NHR-25 to that of other TFs, we acquired peak profiles for various TFs from ChIP-seq experiments conducted at L2 from the modERN database (Kudron et al., 2018) and investigated the global peak localisation pattern they exhibit. We included key TFs known to act specifically in the epidermis, such as ELT-1 and ELT-3, along with others that play broader roles in development including neurogenesis. Profile-wide comparisons of peak localisation between factors, based on overlap and proximity statistics, can provide a measure of co-association and statistically rank factors based on the similarity of their binding (Chikina and Troyanskaya, 2012). Previous calculations of co-association matrices have shown that TFs clustering together often regulate the same targets or act in the same tissue (Araya et al., 2014; Kudron et al., 2018). We found that our NHR-25 and LIN-22 TaDa datasets together with strictly epidermal factors clustered separately (Figure 4F). Interestingly, ChIP-seq and TaDa profiles for NHR-25 did not cluster together (Figure 4F). Since NHR-25 is also expressed and acts in non-epidermal cells, which is likely to be captured in ChIP-seq but not in the TaDa profiles, this may lead to distinct positioning of the NHR-25 binding profile based on whole animal versus tissue-specific putative targets (Figure 4F, G). These results highlight the value of TaDa in tissue-specific studies of TF binding and suggest that LIN-22 and NHR-25 share to some extent targets in epidermal development.

LIN-22 and NHR-25 DNA-binding motif identification by TaDa

We used the TaDa datasets to explore whether we could identify any motif associated to DNA binding. For NHR-25, a DNA binding motif (5'-TGACCTTG-3') has been previously reported by ChIP-seq and confirmed by functional studies (Araya et al., 2014; Barkoulas et al., 2016). This motif was used here as a reference to determine the filtering strategy required to allow its identification starting from TaDa peaks. Peak intensity, size and conservation are commonly used parameters for peak filtering (Araya et al., 2014; Southall et al., 2014). Peak intensity reflects TaDa enrichment scores, size is restricted either by filtering the length of peaks or by isolating GATC fragments of highest enrichment, and conservation is assessed based on PhastCons7way scores (Spieth et al., 2005). We found that selection of the peaks showing the highest enrichment had the highest impact on recovering the NHR-25 binding motif.

Sequences restricted to the overlap between L2 and L4 peaks for each factor were used to increase the probability they reflect real binding sites. Out of these, only the top 200 with the highest averaged enrichment score were used for the analysis. The identified *de novo* motif for NHR-25 was consistent with the one previously reported (Figure S4A left), and when run against a database of known motifs (Fornes et al., 2020) showed significant similarity to those of the *nhr-25* human orthologue NR5A1 (p=3.04×10^{-6}) and mouse orthologue *Nr5a2* (p=2.35×10^{-5}) (Figure S4B left). The same analysis was carried out for LIN-22 and identified a motif that matches an E-box sequence (5'-CANNTG-3') (Figure S4A right). This is similar to reported motifs for the human orthologue HES1 (Lichtenberg et al., 2018). Comparison to known motifs showed significant similarity to that of the human HEY1 (Hes-related Family bHLH transcription factor with YRPW Motif 1) factor (p=3.29×10^{-3}), and HLH-1/ *MyoD* (p=2.45×10^{-4}), which *Hes1* is known to antagonise for sites and binding partners (Sasai et al., 1992) (Figure S4B right).

To assess if the identified motifs broadly represent preference for TF binding as determined in TaDa signal profiles, aggregate signal was mapped onto motif sites. For the

NHR-25 motif, the *nhr-25:dam* L2 and L4 signal showed increased average preference for regions that include the motif as opposed to neighbouring sequences, while the *lin-22:dam* signal did not show any enrichment (Figure S4C). The reverse relationship was observed when the signal profiles were mapped onto the LIN-22 motif (Figure S4D). Some limited enrichment was observed for *nhr-25:dam* L2 data over the LIN-22 motif, which may reflect its more “noisy” nature (Figure S4A) and thus higher frequency in the genome (1.5 times more frequent than the NHR-25 motif).

TaDa-identified LIN-22 and NHR-25 targets relate to their known biological functions

To identify putative target genes that LIN-22 and NHR-25 are likely to regulate, significant peaks were assigned to neighbouring genes. This assignment resulted in a set of 2809 genes for LIN-22 at L2 and 2833 genes at L4, and 3552 genes for NHR-25 at L2 and 3724 genes at L4 (Figure 5A, B and Table S2). The majority of the identified genes were shared between the two stages for each factor, with >63% of genes in any dataset being shared between L2 and L4 (Figure 5A, B). LIN-22 controls aspects of epidermal development by suppressing neurogenesis and influencing ray formation in males (Wrischnik and Kenyon, 1997). Related functions were reflected in GO-term analysis for the identified target gene sets at L2 and L4 with significantly enriched terms like neurogenesis, neuron development, regulation of neuron differentiation and male anatomical structure morphogenesis (Figure 5C-E). In addition, the L2 dataset was found to be enriched for genes related to the Wnt signalling pathway (KEGG adjusted *p*-values of 3.28×10^{-5}). In a similar manner, both the L2 and L4 datasets of putative NHR-25 targets were enriched for multiple GO terms related to NHR-25 biological functions. For example, terms for structural constituents of the cuticle and molting cycle were found to be amongst the most significantly enriched (Figure 5F-H), matching the NHR-25 role in larval molting and new cuticle formation (Chen et al., 2004; Chisholm and Hsiao, 2012; Gissendanner and Sluder, 2000; Hayes et al., 2006). Neuron-related GO terms were also significantly enriched in both stages (Figure 5F-H) and may

correspond to target genes involved in neuronal development from seam cell descendants (Hajduskova et al., 2009).

Comparison between LIN-22 and NHR-25 target genes and between targets identified via TaDa and other methods

We sought to consolidate that the genes we identified with TaDa are likely to be true targets of LIN-22 and NHR-25. To this end, we first compared the NHR-25 TaDa-identified targets with NHR-25 ChIP-seq datasets and found very significant overlap for all pairwise or higher order dataset intersections (Figure S5A). For example, the overlap with the ChIP-seq L2 dataset containing 7438 genes was large containing 62% of L2 and 61.6% of the L4 TaDa genes. Some of the non-overlapping genes from ChIP-seq may reflect targets of NHR-25 outside the epidermis that are not captured by TaDa. Consistent with this hypothesis, tissue enrichment analysis for genes exclusively found by ChIP-seq showed enrichment for neuron and reproduction-related tissues instead of the epidermis (Figure S5B, C).

As there are currently no ChIP-seq datasets available for LIN-22, we used a ChIP-seq dataset for HES1, the human homologue of *lin-22* (Encode accession ENCSR109ODF). Intersections showed statistically significant overlaps for both stages used in TaDa (approximately 23.4% of overlapping genes from both the L2 and L4 datasets) (Figure S5D). In addition, GO term analysis for the genes found in the overlap showed enrichment for the same GO terms previously identified in the TaDa gene sets alone, indicating possible conservation in regulatory interactions of LIN-22/HES1 across species (Figure S5D). The LIN-22 TaDa target genes were also intersected with a list of 52 *in silico* predicted downstream genetic interactors of *lin-22* (Zhong and Sternberg, 2006), and significant overlaps were identified. These included *lin-32* and *mab-5,* but also other genes known to participate in seam cell development like *unc-62* (Hughes et al., 2013), *nmy-2* (Ding and Woollard, 2017) *and rnt-1*, for which there was no prior evidence for a link to *lin-22* (Figure S5B and Table S2). Taken together, TaDa-identified genes are likely to be genuine targets of LIN-22 and NHR-25 in the epidermis.

We finally assessed the overlap between the identified targets for the two TFs. Interestingly, significant overlap was found in all pairwise comparisons across factors with 37%- 46.4% of putative LIN-22 target genes overlapping NHR-25 datasets and 29.3%-35.3% of putative NHR-25 target genes overlapping LIN-22 datasets (Figure S5E). The intersection between all TaDa datasets contained 663 genes, which showed enrichment for GO terms related to neurogenesis and development (Figure S5E). These results indicate that functions of the two TFs during epidermal development are likely to be executed by a combination of distinct and shared target genes.

LIN-22 and NHR-25 targets suggest a link to cell differentiation

We then focused on specific examples of predicted targets identified by signal enrichment in TaDa to investigate in more detail. With regard to putative targets of LIN-22, we focused on *rnt-1* and *cki-1*, which showed significant signal enrichment in their promoter regions (Figure 6A). First, *rnt-1* is known to promote symmetric proliferative seam cell divisions (Nimmo et al., 2005). We found that in *lin-22(icb49)* null mutants, *rnt-1* expression is significantly increased in V1-V4 lineages, indicating that LIN-22 is likely to act as a repressor of *rnt-1* (Figure 6B, C). Comparison of *cki-1* expression between wild-type and *lin-22(icb49)* mutants revealed a significant reduction in transcript levels in V1-V4 lineages (Figure 6D, E). This finding suggests that LIN-22 may activate *cki-1* expression to promote the differentiation programme.

We have previously shown that *lin-17* expression expands to the anterior end in *lin-22* loss-of-function animals (Katsanos et al., 2017). Here, we identified with TaDa two sites of LIN-22 signal enrichment over two GATC fragments on the *lin-17* promoter, one at 598-856 bp and another one at 3456-3713 bp upstream of the *lin-17* ATG (Figure 6A, F). Therefore, we decided to investigate whether these regulatory elements are likely to be required for this interaction. Interestingly, a Vista analysis for conservation (Frazer et al., 2004) by comparing the *lin-17* promoter sequences between *C. elegans* and related *Caenorhabditis* species revealed two conserved regions (termed CRE1 and CRE2), which precisely overlapped with

the most enriched GATC fragments (Figure 6F). To test whether these *cis*-regulatory elements are likely to recruit LIN-22 to supress *lin-17* expression, we constructed transcriptional reporters containing the CRE1 or CRE2 sequence fused to a minimal core promoter driving the expression of histone bound GFP. Multi-copy arrays were created for each element (*icbEx177* for CRE1 and *icbEx180* for CRE2) and introduced into the *lin-22(icb49)* mutant background. Both reporters were sufficient to drive expression in the posterior epidermis in wild-type animals. Furthermore, this expression expanded to more anterior epidermis in the *lin-22(icb49)* mutant background (Figure 6G, H, I). Taken together, these data indicate that LIN-22 is likely to bind to these regulatory elements on the *lin-17* promoter to supress *lin-17* expression.

With regard to putative targets of NHR-25, we focused on *egl-18* and *elt-1,* both GATA transcription factors that promote seam cell fate (Figure 7A). In particular, *egl-18* is a target of the Wnt signalling pathway that specifies and maintains seam cell fate (Gorrepati et al., 2013; Koh and Rothman, 2001). Expansion of *egl-18* expression to seam cell daughters that normally adopt the hypodermal differentiation programme has been shown to correlate with ectopic maintenance of the seam cell fate (Gorrepati et al., 2013; Katsanos et al., 2017). Furthermore, *elt-1* is thought to be the master epidermal fate regulator in *C. elegans* and is known to regulate *nhr-25* in the embryo (Chisholm and Hsiao, 2012; Gilleard and Mcghee, 2001). To test whether these genes are likely to be targets of NHR-25, we carried out smFISH quantifications in control and *nhr-25* RNAi treated animals at the L3 division stage. The efficacy of the RNAi treatment was confirmed by scoring for an increase in terminal seam cell number (Figure 7B). In *nhr-25* RNAi treated animals, *egl-18* was found to be expressed in anterior daughters (Vn.papa and Vn.pppa) of dividing V1-V4 seam cells, whereas it was confined to posterior daughters (Vn.papp and Vn.pppp) in control animals (Figure 7C). Quantification of the signal showed an overall increase in *egl-18* expression in V1-V4 daughters (Figure 7D), indicating that NHR-25 is likely to be a suppressor of *egl-18.* In a similar manner, *elt-1* expression was found to be increased in anterior daughters of the V1-

V4 lineages in *nhr-25* RNAi treated animals (Figure 7E, F). These findings suggest that NHR-25 promotes the differentiation programme by directly suppressing key seam cell fate-promoting factors.

Discussion

TaDa allows transcription factor target identification in *C. elegans* tissues

In this study, we introduce targeted DamID as a method to identify transcription factor targets in a tissue-specific manner in *C. elegans*. Tissue-specific or constitutive levels of expression of Dam-fusions have been known to lead to toxicity, which we also confirmed here in the case of the *C. elegans* epidermis (Schuster et al., 2010; Southall et al., 2013; van Steensel and Henikoff, 2000). The TaDa transgene configuration overcomes this obstacle by minimising the levels of Dam expression in the tissue of interest. We focus on two transcription factors, namely LIN-22 and NHR-25, to establish this method and begin dissecting the epidermal gene network. We show that TaDa can be used in early larval stages and dividing cell populations, such as the seam cells at the L2 stage, while still leading to robust identification of putative targets. This is particularly interesting because hemi-methylated DNA around GATC sites post cell division is a less efficient substrate for DpnI cleavage during the wet lab protocol, and as such it is thought to be more challenging to recover sufficient amount of methylated DNA for the analysis.

Currently, ChIP-seq is the most commonly used method to identify TF targets, with many datasets available based on large-scale projects (Celniker et al., 2009; Kudron et al., 2018). We present here evidence that TaDa is comparable to ChIP-seq in identifying target genes, while it offers two key advantages. First, TaDa allows the identification of TF targets in a specific tissue of interest without cell isolation, as opposed to all tissues where the TF is expressed (Aughey and Southall, 2016; Aughey et al., 2019). Second, TaDa requires substantially less material than ChIP-seq for the analysis. We found approximately 2000 animals to be sufficient to acquire enough methylated DNA to produce 21.5M reads on

average from the seam cells, which represent 3% of the total cells of *C. elegans*. This is also consistent with observations in mammalian cell lines, where 333x more cells were used to identify targets for the same TF by ChIP-seq than TaDa (Cheetham et al., 2018).

Genome-wide assessment of TaDa peaks confirmed enrichment in intergenic regions for both LIN-22 and NHR-25. The most increased average signal was found proximally upstream to the TSS and within 2 kb upstream of genes, which is consistent with ChIP-seq experiments (Araya et al., 2014). In addition, peaks localising within genes showed substantial preference for introns compared to exons, as previously described both in DamID and ChIP-seq experiments in *C. elegans* (Fuxman Bass et al., 2014; Schuster et al., 2010). Introns are known to carry enhancers in *C. elegans* (Fuxman Bass et al., 2014), so this preference is likely to reflect binding to such elements. Further assessment of the regulatory potential of the regions were TaDa peaks localised showed that at least 72% of total peaks were overlapping with open chromatin elements. Coding promoters and putative enhancers were largely found to overlap with TaDa peaks for both TFs. Therefore, the overall localisation profiles of TaDa peaks strongly support that TaDa signal represents genuine TF binding sites. Even though TaDa peaks are broader than those obtained in ChIP-seq peaks, we demonstrate that TaDa data can be used to identify precise TF binding motifs. In addition, we experimentally validate putative binding sites for LIN-22 on two GATC fragments on the *lin-17* promoter showing the highest local score. This highlights the potential to extract focal peak information from TaDa experiments and predict where TF binding events may occur.

When direct comparison across methods could be made, the majority of TaDa putative target genes (>61%) were also identified by stage-matched ChIP-seq. Global pairwise co-association analysis of peak localisation patterns further supported that LIN-22 and NHR-25 TaDa profiles showed the highest similarity with each other, clustering together with other epidermal factor datasets and separately to whole animal NHR-25 ChIP-seq results. The NHR-25 ChIP-seq L2 dataset was approximately three times larger in terms of peaks compared to TaDa L2 (2044 peaks in TaDa compared to 5980 in ChIP-seq). Although

tissue specificity is likely to be an important contributor to this difference, there are other alternative explanations. Both ChIP-seq and DamID methods have inherent biases due to the protocol and reagents used, for example the quality of the Dam fusions (Ramialison et al., 2017), issues with antibody affinity or PCR biases (Meyer and Liu, 2014). A key limitation in all DamID-based methods is the dependence on availability of GATC sites, which could hinder detection of some targets. However, the average length of GATC fragments in *C. elegans* is approximately 367 bp and the median 204 bp, so depletion of GATC sites is unlikely to be pervasive enough to significantly reduce detection of targets. The detection of bound targets in ChIP happens after chemical crosslinking *in vitro*, which has been hypothesised to produce some biases, perhaps including some of the high-occupancy target (HOT) regions determined by ChIP-seq experiments to be bound by multiple transcription factors (Teytelman et al., 2013). Comparisons of genomic locations presented here indicated that only 13% of L2 HOT regions were represented in NHR-25 TaDa, in contrast to 83% in ChIP-seq, highlighting that TaDa peak profiles are less inclusive of HOT regions.

TaDa and smFISH provide a powerful combination to dissect gene regulatory networks with high spatial resolution

Our study refines the position of *lin-22* and *nhr-25* in the epidermal gene network. Based on our findings and previous literature (Brabin et al., 2011; Cassata et al., 2005; Chisholm and Hsiao, 2012; van der Horst et al., 2019; Hughes et al., 2013; Katsanos et al., 2017; Koh and Rothman, 2001; Thompson et al., 2016), we propose here an updated epidermal gene network underlying cell fate decisions (Figure 7G). In the case of *nhr-25,* we identify the seam cell fate regulators *egl-18* and *elt-1* as putative direct targets. So far, *nhr-25* was thought to act on seam cell patterning by regulating seam cell shape and the re-establishment of cell-to-cell contacts post division (Chen et al., 2004; Gissendanner and Sluder, 2000; Šilhánková et al., 2005). Our findings reveal a more prominent role for *nhr-25* in the specification of the hypodermal cell fate through direct suppression of seam cell specifying factors.

We previously proposed that LIN-22 influences seam cell divisions by antagonising Wnt signalling through regulation of the expression of the *frizzled* receptor *lin-17* (Katsanos et al., 2017). LIN-22 is also known to supress neurogenesis in V1-V4 seam cells through suppression of the Hox gene *mab-5,* as well as the pro-neuronal factor *lin-32* (Wrischnik and Kenyon, 1997). However, evidence for a direct interaction was lacking in all these cases. We demonstrate here that LIN-22 is likely to directly bind *mab-5* and *lin-32* upstream sequences in the seam cells. We also demonstrated that LIN-22 may supress *lin-17* by binding to two conserved elements on its promoter. Interestingly, LIN-22 was found to target components at various levels of the Wnt signalling cascade, such as Wnt secretion *(mom-1, mig-14*), the destruction complex (*mig-5, gsk3*), β-catenin (*bar-1*), and Wnt receptors (*mom-5, lin-17*) (Table S2). Although *Hes*-related factors are commonly associated with Notch signalling in other systems (Kageyama et al., 2007), these findings link LIN-22 to the regulation of Wnt signalling in *C. elegans*.

In light of our TaDa findings, LIN-22 appears to have a role in mediating correct hypodermal differentiation in the epidermis. LIN-22 was found to supress *rnt-1,* the Runx homologue of *C. elegans,* which in complex with BRO-1 promote seam cell fate and symmetric divisions by suppressing *pop-1* (van der Horst et al., 2019). This interaction is consistent with the observation of ectopic symmetric divisions in *lin-22* mutants (Katsanos et al., 2017). LIN-22 was also found to activate the cell cycle inhibitor *cki-1*. This gene is expressed in the seam cells and *cki-1* RNAi increases seam cell number (Buck et al., 2009; Hong et al., 1998). Since *cki-1* drives G1 arrest allowing differentiation to progress (Hong et al., 1998; Matus et al., 2015), reduction of *cki-1* expression may contribute to the supernumerary seam cells observed in *lin-22* mutants. Canonical *Hes* factors are thought to act as suppressors (Kageyama et al., 2007), for example the mammalian homologue of *cki-1*, $p27^{Kip1}$ has been shown to be directly supressed by Hes1 in mice (Murata et al., 2005). In contrast to canonical HES factors, LIN-22 lacks a Groucho interacting domain (Schlager et

al., 2006), so it may act as a repressor or activator depending on co-factors or competition with other transcription factors for binding.

TaDa offers the potential to dissect in the future complex TF behaviours related to tissue-specific target binding. This includes the expanded family of HLH factors in *C. elegans* known to bind different targets depending on their dimerising partner that is often tissue-specifically expressed (Grove et al., 2009). We demonstrate here that TaDa predictions can be consolidated at single cell resolution using smFISH in appropriate mutant or silenced backgrounds. This experimental combination can add quantitative strengths to the described network connections, thereby facilitating mathematical modelling of developmental gene networks (Piano et al., 2006). The revised epidermal network we present here represents a framework for future experiments to build upon and validate newly predicted interactions. Given the conserved nature of some of the participating factors, this network can inform about interactions that underlie stem cell behaviour and the developmental logic of robust cell fate patterning in other stem cell contexts.

Acknowledgments

We thank Tony Southall, Colin McClure and Gabriel Aughey for discussions on setting up the method. We thank members of the Barkoulas lab and David Matus for comments on the manuscript. Some *C. elegans* strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). We acknowledge the support from the European Commission [ROBUSTNET- 639485]. DK was the recipient of an Imperial College President's PhD scholarship.

Author contributions

DK carried out all experiments and data analysis. MB supervised the work. DK and MB wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figure Legends:

Figure 1: Design of TaDa constructs to find LIN-22 and NHR-25 targets prevents toxicity.

(A) Rationale of TaDa for analysing tissue or cell type-specific protein-DNA interactions. The bicistronic design of TaDa with a primary ORF of *mCherry* followed by two STOP codons and a frame-shift permits low levels of tissue-specific expression of a fusion between the protein of interest (POI) and Dam. This is thought to prevent toxicity and extensive non-specific methylation. **(B)** Illustration of the key features of single-copy transgenes used in this study for LIN-22 and NHR-25 target identification by TaDa (left). Transgenes used to assess the requirement of an *mCherry* primary ORF to prevent toxicity and high methylation levels are also shown (right). **(C)** Representative brightfield images of adult WT and transgenic animals with *lin-22:dam* or *NLS-GFP:dam* fusions on plates showing the exploding vulva phenotype in the absence of the primary *mCherry* ORF. This phenotype was found in 54.5% of the *lin-22:dam* and 75% of the *NLS-GFP:dam* animals. White arrows point to internal tissue outside the body. **(D)** Quantification of brood size in the above strains. Note defect in brood size in animals carrying *lin-22:dam* and *NLS-GFP:dam* fusions (n=15). Scale bars in C are 100 µm. In Derror bars indicate standard error of the mean and black stars indicate statistically significant differences in the mean with a one-way ANOVA/Dunnett's test(* $p<0.05$, **** $p<0.0001$).

Figure 2: LIN-22 and NHR-25 signal is enriched in upstream gene regulatory regions. **(A-B)** Examples of signal profiles in regions with statistically significant enrichment peaks. For *lin-22:dam* (A), enrichment is shown in promoters of the putative LIN-22 targets *lin-17*, *mab-5* and *lin-32,* while for *nhr-25:dam* (B), enrichment is shown around *nhr-25* and in the promoters of previously confirmed targets *idh-1* and *rpl-3*. The Y-axes represent $\log_2$*(TF:dam/NLS-GFP:dam)* scores. Scale bar length is 2 kb as indicated. **(C)** Aggregation

plots for all indicated *TF:dam* fusions in both developmental stages, showing average enrichment scores in 10 bp bins for regions of equal length across all of the specified genomic features indicated on the X-axis. Strong enrichment preference is seen upstream to gene regions. Plots show 5 kb upstream of the TSS of genes to 1 kb downstream of the TES, with gene bodies pushed into a 2 kb pseudo-length. Y-axes are z-scores for the plotted sequence length and shaded areas represent 95% confidence intervals. **(D)** Heatmaps representing the hierarchically clustered localisation and enrichment score of all statistically significant peaks (FDR<0.05) within 5 kb upstream and 1 kb downstream of a gene. **(E)** Proportions of peaks residing in various genomic locations relative to the gene they are assigned to. To calculate proportions, peaks were assigned to the single closest gene, if any, when their centre coordinate localised within 6 kb upstream and 1 kb downstream of the TSS and TES respectively. Ratios above pie charts show the number of assigned peaks to the total number of significant peaks found. The proportions of peaks within genes with exclusive intron or exon localisation are indicated under the pie charts.

Figure 3: LIN-22 and NHR-25 peaks largely overlap with open chromatin sequences.

(A) Aggregation plots of open chromatin signal from ATAC-seq L2 and L4 data (Jänes et al., 2018) over LIN-22 and NHR-25 TaDa L2 and L4 peaks respectively, indicate increased chromatin openness at the sites of peaks. Y-axes are z-scores for the plotted sequence length and shaded areas represent 95% confidence intervals. **(B)** Heatmaps of open chromatin signal centred on TaDa TF peaks show that the majority of peaks overlap with open chromatin regions. The key shows previously calculated ATAC-seq genome-wide normalised coverage scores (Jänes et al., 2018). In A and B, ±5 kb around the peak centres is shown. **(C)** Proportions of total peaks for each indicated TF at each stage overlapping with different categories of regulatory annotated open chromatin elements. Note that one TF peak may overlap with more than one open chromatin element.

Figure 4: Analysis of peak localisation profiles shows agreement with ChIP-seq data and overlap between the two transcription factors.

(A) Representative snapshots showing good agreement between *nhr-25:dam* TaDa data and two ChIP-seq signal profiles from L1 (Shao et al., 2013) and L2 (Araya et al., 2014) staged animals over *nhr-25*, *idh-1* and *egl-18*. Signal enrichment shown here forms statistically significant peaks. The Y-axes for TaDa samples represent $\log_2$*(nhr-25:dam/NLS-GFP:dam)* scores while for ChIP-seq the fragment pile-up per million reads score. Scale bar is 2 kb as indicated **(B)** Aggregation plots showing enrichment of *nhr-25:dam* L2 and L4 signal over centre of NHR-25 ChIP-seq L1 (left) and L2 (right) peaks. **(C)** Graph indicating the number of peaks from *nhr-25:dam* L2 and L4 samples that overlap with the L1 or with the L2 NHR-25 ChIP-seq peaks datasets. **(D)** Aggregation plots of *lin-22:dam* L2 and L4 signal over NHR-25 L2 and L4 TaDa peaks (top) and *nhr-25:dam* L2 and L4 signal over LIN-22 L2 and L4 TaDa peaks. All show increased signal for the peak location. **(E)** Graph showing the numbers of peaks from each colour-coded sample that overlap with the peaks of the sample indicated on the X-axis. The graph is separated in two parts representing numbers across the whole genome or restricted to promoter regions 5 kb upstream to 500 bp downstream of the TSS. **(F)** Heatmap of symmetrized co-association values, hierarchically clustered, for all pairwise comparisons between peak localisation patterns of the indicated TFs from L2 animals. All peak profiles for the compared TFs (except of the TaDa produced here) derive from ChIP-seq experiments (Kudron et al., 2018). Note the separate clustering of TADA with epidermal factors. **(G)** Heatmap of tissue-specific expression levels from (Cao et al., 2017) in the gonad, neurons, pharynx, body wall, muscle, intestine, epidermis and glia for the TFs tested for peak pattern co-association. The heatmap is scaled per gene. In (B) and (D), ±5 kb around the peak centres have been plotted and the Y-axes represent z-scores for the plotted sequence length and shaded areas represent 95% confidence intervals. In (C) and (E), the exact number of peaks with overlaps is written above the bars and *p*-values from Monte Carlo simulations for the statistical significance of overlaps is printed inside the bar.

Figure 5: Enriched GO terms for LIN-22 and NHR-25 putative targets relate to their known biological functions.

(A-B) Venn diagrams for putative target intersections between genes identified at L2 and L4 for the *lin-22:dam* (A) and *nhr-25:dam* (B) fusions (significance of intersection size with a hypergeometric distribution test $p << 0.0001$ for both). **(C-H)** Plots of selected significantly enriched terms from GO term analysis for the genes in the L2 (C-F), L4 (E-H) or intersection datasets (D-G). For both TFs at all stages, relevant terms are recovered related to neurogenesis and development for *lin-22:dam* or molting for *nhr-25:dam*.

Figure 6: LIN-22 activates *cki-1* and represses *rnt-1 and lin-17.*

(A) *lin-22:dam* signal enrichment profiles forming significant peaks around genes identified as putative LIN-22 targets. Y-axes are $\log_2$*(lin-22:dam/NLS-GFP:dam)* scores. **(B)** Representative *cki-1* smFISH images from WT and *lin-22(icb49)* animals at the late L1 stage. **(C)** Quantification of *cki-1* mRNA spots in H0-V4 seam cells, showing a significant reduction of expression in V1-V4 cells ($50 \leq n \leq 126$). **(D)** Representative *rnt-1* smFISH images from WT and *lin-22(icb49)* animals at the late L1 stage. **(E)** Quantification of *rnt-1* mRNA spots in H0-V4 seam cells, showing a significant increase in expression in V1-V4 cells ($65 \leq n \leq 167$). **(F)** rVista analysis of the *lin-17* promoter identified two conserved, putative regulatory elements (*CRE1* and *CRE2*) between the *Caenorhabditis* species indicated and *C. elegans*. The position of the elements is shown on the *C. elegans* sequence as a reference. They are both overlapping GATC fragments with local maximum enrichment scores. **(G-H)** Representative brightfield and fluorescence images of L4 transgenic animals carrying transcriptional reporters for the *lin-17 CRE1* (*icbEx177* transgene) (G) and *CRE2* (*icbEx180* transgene) (H) sequences fused to *GFP:H2B* in WT and the *lin-22(icb49)* mutant background. Note that expression is restricted to few posterior cells in WT and expands to the anterior epidermis in mutants. White arrowheads indicate epidermal GFP expression. *lin-22(icb49)* mutants also carry a *dat-1p::GFP* reporter labelling dopaminergic neurons, which is distinguishable by the visible axons. **(I)** Quantification of the proportion of transgenic animals with epidermal expression anterior to the vulva in WT (n=35 for *CRE1* and n=24 for *CRE2*) and *lin-22(icb49)* mutant animals (n=47 for *CRE1* and n=20 for *CRE2*) carrying the

above reporters. Error bars indicate the standard error of the proportion. Black stars show statistically significant differences of the proportion with a Fisher's exact test, *** $p<0.001$. In B and D, seam cells are labelled with *SCMp:GFP* and black spots correspond to investigated mRNAs. Scale bars are 2 kb in A, 10 µm in B and D, 500 bp in F, 100 µm in G and H.

Figure 7: NHR-25 represses *egl-18* and *elt-1.*

(A) *nhr-25:dam* signal enrichment profiles in regions with significant peaks around genes identified as putative NHR-25 targets. **(B)** Seam cell number scoring in control and *nhr-25* RNAi treated animals based on *SCMp::GFP* marker expression. Animals were grown on the same plates used for smFISH to assess treatment efficacy ($n \geq 36$). **(C)** Representative *egl-18* smFISH images from control and *nhr-25* RNAi treated animals during the L3 division. **(D)** Quantification of *egl-18* mRNA spots in the V1-V4 L3 division daughter cells**,** showing a significant increase in expression in all cells ($60 \leq n \leq 88$). **(E)** Representative *elt-1* smFISH images from control and *nhr-25* RNAi treated animals during the L3 division. **(F)** Quantification of *elt-1* mRNA spots in the V1-V4 L3 division daughter cells showing a significant increase in expression in the anterior daughters (Vn.papa and Vn.pppa lineages) ($82 \leq n \leq 116$). **(G)** A combined network of epidermal cell fate interactions based on findings from this study and previous published research. Solid lines correspond to direct regulatory interactions found by TaDa, with arrowheads indicating activation and flatheads repression. Black lines indicate literature-described interactions, magenta lines LIN-22 targets and blue lines NHR-25 targets found and validated in this study. Blue and magenta dashed lines indicate new interactions predicted by TaDa. Black dashed lines indicate genetic interactions that is yet unknown whether they are direct or not. The *elt-1* activation of *bro-1* has been shown to be direct (Brabin et al., 2011). This is an updated version of the networks presented in (Chisholm and Hsiao, 2012; Koh and Rothman, 2001; Thompson et al., 2016), including data also from (Brabin et al., 2011; Cassata et al., 2005; van der Horst et al., 2019; Katsanos et al., 2017; Nimmo et al., 2005) and new links identified in this study.

STAR Methods

C. elegans maintenance and genetics

The *C. elegans* strains used in this study were maintained on standard Nematode Growth Medium (NGM) plates and were grown monoxenically on a lawn of *E. coli* OP50 at 20 °C. For TaDa experiments, strains were grown for at least two generations on a lawn of a *dam*$^-$ */dcm*$^-$ *E. coli* mutant of the K12 strain (New England biolabs, C2925). The laboratory reference N2 strain was used as the reference strain in this study. A complete list of strains used in this study is available in Table S3.

Molecular cloning

To construct a TaDa backbone plasmid for epidermal expression of TFs fused upstream of Dam, the pCFJ151 universal MosSCI vector (Frokjaer-Jensen et al., 2014) was digested with BcuI/BspTI enzymes. The promoter of *wrt-2* was amplified from N2 lysate with oligos PB16 and PB7, the *C. elegans* optimised *mCherry* was amplified from the pAA64 (Barkoulas et al., 2016) plasmid using oligos PB8 and PB17, the *Dam* sequence was amplified from the pUAST attB LT3 Dam plasmid (gift by Tony Southall) using oligos PB18 and PB13 and the *unc-54 3' UTR* was amplified from N2 lysate using oligos PB14 and PB15. All 4 fragments and the digested backbone were inserted in a multi-fragment Gibson assembly reaction to produce the pPB7(*wrt-2p::mCherry::Dam::unc-54 3'UTR* + *cb-unc-119*) plasmid. The XmaJI site between *mCherry* and *Dam* was digested to linearise the vector and allow the in-frame to *Dam* insertion of the *nhr-25* coding sequence amplified using oligos PB19 and PB20 from N2 cDNA to produce the pPB10(*wrt-2p::mCherry::nhr-25:Dam::unc-54 3'UTR* + *cb-unc-119*) plasmid. To construct the seam cell driven *lin-22:Dam* fusion and the *NLS-GFP:Dam* control for the TF TaDa experiments, the *lin-22* gene was amplified with oligos DK11 and DK12 from fosmid WRM0627dG07 while *NLS-GFP* was amplified from plasmid pPD93_65 (Fire Lab vector Kit) using oligos DK15 and DK16. Both amplicons were inserted upstream and in-frame with *Dam* by Gibson assembly in an XmaJI linearised pPB7 vector like above. The

resulting plasmids produced were pDK4(*wrt-2p::mCherry::lin-22:Dam::unc-54 3'UTR + cb-unc-119*) and pDK8(*wrt-2p::mCherry::nhr-25:Dam::unc-54 3'UTR + cb-unc-119*). To test the importance of the *mCherry* primary ORF to the viability of animals and methylation levels, versions of the *lin-22:Dam* and *NLS-GFP:Dam* TaDa constructs without *mCherry* were produced. In more detail, the pPB7 plasmid was digested with BcuI/MunI, the 4085 bp and 6141 bp fragments were excised, extracted and kept. *lin-22* was amplified from pDK4 using DK102 and DK11. The 2 digestion fragments, *the lin-22* amplicon and the repair oligo DK103 were all inserted into a Gibson reaction to produce pDK49(*wrt-2p::lin-22:Dam::unc-54 3'UTR + cb-unc-119*) which was then digested with BcuI/XmaJI to remove *lin-22* and insert via Gibson assembly a DK108 and DK15 amplified fragment of NLS-GFP from pDK8 to generate pDK50(*wrt-2p::NLS-GFP:dam::unc-54 3'UTR + cb-unc-119*).

For ease of future applications, a versatile TaDa vector called pDK7 was constructed. Briefly, the *att* recombination cassette of pDest R4-R3 (Invitrogen) was amplified including the *attR4* site, *ccdb* and *CamR* genes but excluding the *attR3* site using the oligos DK17 and DK18 including half of the *attL1* site sequence on the 3' DK18 primer. *mCherry* was amplified from pPB7 using oligos DK19 and DK20 carrying the other half of the *attL1* site on the 5' of DK19. *dam* was amplified from pPB7 with oligos DK21 and DK22. The *unc-54 3'UTR* was amplified from pPB7 with oligos DK23 and DK24. All 4 fragments were inserted in a Gibson assembly reaction along with BcuI/BspTI doubly digested pCFJ151 vector to generate pDK7(*attR4-L1::mCherry::Dam-myc::unc-54 3'UTR + cb-unc-119*).

For the *lin-17 CRE1* and *CRE2* transcriptional reporters the oligos DK115 and DK116 along with DK118 and DK119 were used to amplify each of the respective regions from N2 lysate. The *Δpes-10* core promoter was amplified from L3135 using either the *CRE1* or the *CRE2* compatible forward primers DK117 and DK120 along with the DK107 reverse and was cloned along with the respective CRE amplicon in a NheI/XmaJI digested pDK16 to create pDK59(*lin-17CRE1:: Δpes-10::GFPo-H2B::unc-54 3'UTR + cb-unc-119*) and pDK60(*lin-*

17CRE2:: Δpes-10::GFPo-H2B::unc-54 3'UTR + cb-unc-119). A complete list of the oligos used in this study is presented in Table S4.

Single copy transgenesis

Single-copy transgenic lines were produced using the *Mos1*-mediated single-copy insertion (MosSCI) method (Frokjaer-Jensen et al., 2014). Briefly, ~30 day-one adult animals of the EG6699 strain with a *Mos1* transposon insertion on chromosome II (*ttTi5605* locus) showing the uncoordinated (Unc) phenotype were injected for each transgene insertion. All the MosSCI injection mixes comprised of 50 ng/µl of a universal MosSCI vector carrying the transgene of interest flanked by the *ttTi5605* left and right recombination arms along with plasmids harbouring the *Mos1* transposase (pCFJ601 at 50 ng/µl), a heat-shock inducible *peel-1* toxin (pMA122 at 10 ng/µl) and co-injection markers (pGH8 at 5 ng/µl, *myo-2::dsRed* at 2.5 ng/µl, *myo-3::mCherry* at 5 ng/µl). Post-injection, animals were kept at 25 °C until plates were completely starved. The heat-shock treatment that follows was performed at 34 °C for 3.5 hours, after which the plates were allowed to recover for 3 hours at room temperature before “reverse chunking” was performed, where NGM chunks from the lawn of a new plate were placed on top of the starved, treated plate with the OP50 lawn facing upwards (O’Connell, 2010). The next day the top of lawns were screened for *unc-119(-)* rescued animals with absence of co-injection markers, which were transferred on different NGM plates per injected P0. Homozygous lines were confirmed molecularly for the presence of single-copy transgene insertions using oligos NM3880 and NM3884. A complete list of the transgenes produced for this study along with injection mix make-up information is available in Table S5.

Microscopy

For seam cell imaging, live animals were mounted on fresh 2% agarose pads containing 100 µM NaN_3 for immobilisation on glass slides. The slides were then imaged using either an AxioScope A1 (Zeiss) upright epifluorescence microscope with a Light Emitting Diode (LED)

light source fitted with a RETIGA R6™ camera (Q IMAGING) controlled via the Ocular software (Q IMAGING) or on an inverted Ti-eclipse fully motorised epifluorescence microscope (Nikon) with a metal halide light source fitted with an iKon M DU-934, 1024 x 1024 CCD-17291 camera (Andor) controlled via the NIS-Elements software (Nikon). Scoring of the terminal seam cell or postdeirid neuron number phenotype was performed in late-L4 or early adult animals, carrying the *SCMp::GFP* or *dat-1p::GFP* markers. The lateral side most proximal to the objective was counted for every animal.

To perform smFISH, large populations of animals were synchronised by bleaching and were subsequently grown at 20 °C for 18 hours to reach the late L1 and 35 hours for the L3 asymmetric division stage. Animals were fixed in 4% formaldehyde (Sigma-Aldrich) in 1x PBS (Ambion) for 45 min on a vertical Stuart™ Rotating disk (Cole-Palmer). They were washed with 1.5 ml 1x PBS twice before being stored at 4 °C in 70% Ethanol for at least 24 hours for permeabilisation. Hybridisation was performed in 100 µl of buffer (100 mg/ml dextran sulphate (Sigma-Aldrich), 10% formamide in 2x SSC) at 30 °C for 16 hours with1 µl of a diluted in water probe added. The probe dilution was between 1:5 and 1:50 of a custom-made mixture of 21-48 Cy5-labelled oligonucleotides (Biomers) targeting the gene of interest. A complete list of probes and their dilution used in this study along with their sequences is available in Table S4. Animals were washed in a solution of 10% formamide, 2x SSC, stained with DAPI and resuspended for imaging in 100 µl GLOX buffer (0.4% glucose, 10 mMTris-HCl in 2x SSC) supplemented with 1 µl of 3.7 mg/ml glucose oxidase (Sigma-Aldrich) and 1 µl of 5 mg/ml Catalase (Sigma-Aldrich). Imaging was performed using the Nikon set-up described above using the seam cells closest to the objective lens as homing coordinates to acquire 17 Z-stack slices with a step of 0.8 µm for each of the DAPI, Cy5 and GFP channels (Semrock). Acquisition was performed using a 100x oil immersion objective with exposure set at 100 ms for DAPI at 1/32 reduced light intensity, 3 s for Cy5 and 300 ms for GFP. Analysis was performed using a custom MATLAB® (MathWorks) pipeline (Barkoulas et al., 2013). In brief, selected animal DAPI and GFP images were used to

annotate seam cells and draw regions of interest (ROIs) around the nuclei for at least 5 slices within which smFISH spots would be counted. An animal specific threshold for spot detection was set by manually sampling spots.

TaDa wet lab protocol

Strains for TaDa experiments were transferred onto *dam⁻/dcm⁻* plates by spot bleaching and two biological replicates in separate plate were established per strain that were processed simultaneously. For each strain and replicate, nine 55 mm plates fully populated by gravid adults were used for large-scale egg preparation with isolated embryos seeded in new *dam⁻ /dcm⁻* plates that were incubated at 20 °C. Half of the resulting synchronised populations were collected in a 15 ml centrifuge tube after 24 hours at the L2 stage and the other half after 48 hours at the L4 stage using 2 ml M9 buffer per plate. All collected populations underwent extensive washing by centrifuging animals at 1200 g for 3 min, removing the supernatant and washing the pellet (~100 µl) with 10 ml M9 at least 5 times. Animal pellets were frozen at -20 °C before gDNA extraction.

For gDNA extraction pellets were lysed using 750 µl of Cell Lysis Solution (QIAGEN) containing 100 µg/ml of Proteinase K on a heat-block at 55 °C shaking at 500 RPM for 16 hours overnight. Lysates were treated with 4 µl of 100 mg/ml RNase A (QIAGEN) at 37 °C shaking at 500 RPM for 3 hours. In turn, 250 µl of Protein Precipitation Solution (QIAGEN) was added to each sample and were incubated on ice for 5 min followed by vigorous vortexing for at least 30 sec, another 5 min incubation on ice and a centrifugation at 6000 g for 10 min at 4 °C. The supernatant for each sample was treated with isopropanol, washed with 70% ethanol and the pellet was air-dried for 1 hour. DNA pellets were hydrated with 55 µl of UltraPure™ distilled water and were left for 48 hours at 4 °C for the DNA pellet to dissolve.

For isolation and amplification of the GATC methylated gDNA fragments the protocol followed here is an adapted version of the one presented in (Marshall et al., 2016) for TaDa

in *Drosophil*a with minimal alterations. Of the extracted gDNA samples above, a total amount of 5 µg was transferred into a new 1.5 ml tube and was brought to 43 µl with the addition of UltraPure™ Distilled Water. For those samples where 5 µg were not available, 43 µl of the original sample were transferred. To each of those tubes 5 µl of 10x CutSmart Buffer (New England Biolabs) and 2 µl of DpnI restriction enzyme (New England Biolabs) were added and mixed by gentle flicking to prevent shearing of gDNA. The samples were incubated at 37 °C for 16 hours overnight and were in turn cleaned-up using the QIAquick PCR Purification kit (QIAGEN) and eluted with 40 µl of 50 °C water. 30 µl of each clean digestion product were split equally in two 0.2 ml PCR tubes (15 µl in each) and 4 µl of *Adaptor ligation buffer* (5x T4 DNA Ligase Buffer (New England Biolabs) and 10 µM of the dsAdR adaptor along with 1 µl (400 U) T4 DNA Ligase (New England Biolabs) were added in each. The samples were then incubated at 16 °C for 2 hours, followed by 10 min at 65 °C in a thermocycler. The double-stranded adaptor dsAdR was initially prepared by mixing equal volumes of 100 µM of the single stranded oligos AdRT and AdRb in a 1.5ml tube, immersing in a boiling-hot water-bath and letting it to cool down to room temperature to allow for gradual annealing. Following the adaptor ligation each sample was mixed with 20 µl of a 2x DpnII Digestion Buffer and 10 units (1 µl) of DpnII restriction enzyme (New England Biolabs) mastermix and were incubated for 3 hours at 37 °C. At this stage for each original sample two 40 µl digestion reaction products were available. For methylated DNA amplification by PCR each one of these products was mixed with 118 µl of *DamID PCR buffer* and 2 µl (10 units) of MyTaq™ DNA polymerase (Bioline) and were aliquoted at 40 µl in 4 different 0.2 ml PCR tubes. The *DamID PCR buffer* consisted of 1.36x MyTaq™ Buffer (Bioline) and 1.06 µM of the DamID PCR primer (Adr_PCR) that anneals on the adaptor sequence. In total for each original gDNA sample 8 PCR reactions were performed using the following cycling program:

Single cycle of steps 1-4: 72 °C for 10 min, 94 °C for 30 sec, °C for 5 min, 72 °C for 15 min

3 Cycles of steps 5-7: 94 °C for 30 sec, 65 °C for 1 min, 72 °C for 10 min

21 Cycles of steps 8-10: 94 °C for 30 sec, 65 °C for 1 min, 72 °C for 2 min

Final extension step: 72°C 5 min, Slow cool down to room temperature and storing at 10 °C.

The number of cycles for steps 8-10 was increased from 17 described in (Marshall et al., 2016) to 21, as is more commonly used in conventional DamID experiments in *C. elegans* (Askjaer et al., 2014). The 8 reactions per sample were pooled and were cleaned-up using QIAquick PCR Purification kit (QIAGEN). To remove the adaptor sequences from the resulting PCR products up to 2.5 µg of product was transferred to a new 1.5 ml tube and was digested with AlwI restriction enzyme (New England Biolabs) at 37 °C for 16 hours. The products were cleaned-up again using the QIAquick PCR Purification kit. These final products were sent to GENEWIZ® for library preparation and Next Generation Sequencing (NGS) using the Illumina® HiSeq platform.

Calculation of TaDa signal profiles for LIN-22 and NHR-25 binding

FASTQ files representing single-end reads for each sample and replicate, were initially assessed using the fastq-stats perl script (available at https://github.com/owenjm/damid_misc/blob/master/fastq-stats) for uncut adaptors, primer dimer and internal GATC content as a post-sequencing quality control step for the wet lab executed protocol. The sequencing reads were mapped on the *C. elegans* genome, sequence alignment read-count maps were generated and normalised $\log_2$(*TF:dam/NLS-GFP:dam*) ratio scores were calculated per GATC fragment of the genome using the perl script damidseq_pipeline v1.4.5 (Marshall and Brand, 2015) (available at https://github.com/owenjm/damidseq_pipeline). The pipeline was used calling Bowtie 2 v2.3.4 (Langmead and Salzberg, 2012) for alignment to *C. elegans* bowtie indices from genome assembly WBcel235 (available from illumina® iGenomes page), Samtools v1.9 (Li et al., 2009) for alignment manipulations and a GATC fragment file with the coordinates of all GATC fragments across the *C. elegans* genome in gff format, built from a WBcel235 FASTA file (available at https://www.ensembl.org/Caenorhabditis_elegans/Info/Index) using the gatc.track.marker.pl script (available at https://github.com/owenjm/damidseq_pipeline). The number of usable reads that have been acquired by the experiments presented here and map only once to the genome varied from over 6 million up

to ~30 million reads per sample with the genome coverage being between ~9 and 45 times. From the 2 replicates per TF-Dam fusion and control-Dam fusion all pairwise genome-wide $\log_2$(*TF:dam/NLS-GFP:dam*) calculations were performed and averaged into a single signal profile of $\log_2$(*TF:dam/NLS-GFP:dam*) enrichment scores per GATC of the genome for each TF at each developmental stage. Genomic coordinate files produced and used throughout this study were converted between formats (BED, GFF, Bedgraph, BigWIg, Wig, GTF) using Excel, the Convert between GTrack/BED/WIG/bedGraph/GFF/FASTA files tool of the Galaxy powered GSuite Hyperbrowser (elixir) (at https://hyperbrowser.uio.no/hb/!mode=advanced) and the UCSC browser binaries bedGraphToBigWig, BigWigToBedGraph, bigWigToWig. BED and GFF signal and feature track files were visualised using the SignalMap NimbleGen software (Roche).

Peak-calling and gene-assignment

Identification of statistically significant enriched peaks across the genome for each TF and developmental stage was performed using the perl script find_peaks (available at https://github.com/owenjm/find_peaks) with an FDR<0.05 and default settings with the averaged $\log_2$(*TF:dam/NLS-GFP:dam*) signal profiles as input. The output is a list of genomic interval coordinates for statistically significant peaks with a peak enrichment score and an FDR value. Significant peaks were assigned to genes using UROPA (Kondili et al., 2017) as a web tool (available at http://loosolab.mpi-bn.mpg.de/UROPA_GUI/) with Caenorhabditis_elegans.WBcel235.99.gtf (from http://www.ensembl.org/Caenorhabditis_elegans/Info/Index) as the genome annotation file. Peaks were assigned to genes on any strand when their centre coordinate was positioned up to 6 kb upstream of a gene start site or 1 kb downstream of the end site. The location of the peak relative to the gene was assigned based on the full length of the peak and the strand of the gene. To avoid discarding valid regulatory relationships and due to the compactness of the *C. elegans* genome no prioritisation was set for peaks when multiple assignments were available and all are reported.

For the NHR-25 L1 ChIP-seq dataset (Shao et al., 2013) used in this study for peak localisation comparisons, raw sequencing data (GEO number: GSE44710) were processed into significant peak profiles using MACS2 on Galaxy (https://usegalaxy.eu/). NHR-25 and input bedgraph signal tracks were inserted into the MACS2 bdgcmp tool (default: Poisson *p*value algorithm) to deduct noise and identify NHR-25 specific signal. The output was then used with the MACS2 bdgpeakcall tool (cutoff 1.0, min-length 200 and max-gap 30) to generate genome-wide profiles of significant peaks. Profiles representing different replicates were merged using bedtools merge with averaged heights for peaks that overlap.

Pearson's correlation and principal component analysis

Correlation between samples and reproducibility of replicates was assessed using the deeptools3 (Ramírez et al., 2016) multiBamSummary (--binSize 300) and plotCorrelation (--corMethod pearson, --whatToPlot heatmap, --skipZeros, --removeOutliers) tools on Galaxy (https://usegalaxy.eu/). Principal component analysis was performed using the deeptools3 plotPCA tool on the multiBamSummary-calculated read count density summary matrices.

Aggregation plots and heatmaps of signal localisation

Signal localisation preference around given genomic features presented as aggregation plots or heatmaps were generated using the SeqPlots GUI application (Stempor and Ahringer, 2016) with specific settings mentioned individually for each presented result. Aggregation plots represent signal averages for 10 bp bins in regions of varying but specified length around positional features of the genome. For genes all of their start and end coordinates, based on the largest transcript and used here as the transcriptional start site (TSS) and transcriptional end site (TES) of genes, are anchored to two positions of the X-axis and their genic sequence is pushed or stretched to a pseudo-length of usually 2 kb. For other features the midpoint coordinate is used to align all to the same position on the X-axis which then extents upstream and/or downstream of that region. For each position around the feature an average is calculated across all the features to generate the aggregation plot line

with a shaded area representing the 95% confidence interval. When z-scores are presented on the Y-axes, those have been calculated as deviations from the mean signal seen across the plotted region. In heatmaps each line represents each occurrence of a genomic feature indicated on the X-axis along with a surrounding region of a given length also indicated on the X-axis. Colourscales indicate the positional enrichment score calculated as averages per 10 bp bins.

Assessment of overlaps between sets of genomic intervals or gene sets

To identify overlapping peaks between samples or other genomic interval or features the bedtools intersect tool was used with settings dependent on the prospected outcome of the processing. To test if sets of genomic coordinates representing various features show statistically significant overlaps across the genome, Monte Carlo simulations have been performed using the python pipeline OLOGRAM, part of the gtftk package (Ferré et al., 2019). *p*-values are calculated based on the occurrence of intersections between intervals and overall length of overlap (in bp) across the genome. For statistical assessment of the level of association between patterns of peak (TaDa or ChIP-seq peaks) localisation across the genome for different TFs, the IntervalStats tool (Chikina and Troyanskaya, 2012) as part of the coloc-stats webserver (https://hyperbrowser.uio.no/coloc-stats/) was used. In brief, for the TFs used in this study ChIP-seq optimal IDR-thresholded peak coordinate files from L2 animals were downloaded from the modERN (Kudron et al., 2018) and modENCODE (Celniker et al., 2009) databases and were combined along with the L2 TaDa TF samples into a Gsuite of genomic tracks on coloc-stats. Each peak file was then used as query against the GSuite of reference sequences to calculate the IntervalStats statistic for co-localisation for all pairwise comparisons of peak coordinates. The values in the resulting comparison matrix representing comparisons between the same two TFs with different directionality (query-reference) were averaged to symmetrise the matrix and calculate the final co-association values that were plotted as a heatmap using the R package heatmap3 and hierarchical clustering. To assess the statistical significance of overlaps between sets of

genes, hypergeometric distribution or Fisher's exact tests were performed either on http://nemates.org/MA/progs/overlap_stats.html or using the R software package *SuperExactTest* (Wang et al., 2015) respectively. For both tests when sets of coding genes are compared the size of the sampling pool was set to 20191, the number of annotated coding genes in the WBcel235.99 release. Representation of overlaps was either in the form of Venn diagrams generated using http://bioinformatics.psb.ugent.be/webtools/Venn/ or in the form of the output of the *SuperExactTest* package.

TF motif identification by TaDa

Identification of motifs from transcription factor TaDa peaks was done here using HOMER (Heinz et al., 2010). Peak interval files were used as input for the findMotifsGenome.pl script using the ce11 genome assembly, masking of the sequences and the option to analyse the size of sequences provided by the interval file (options: ce11 -size given -mask). The logos presented here for motifs identified using homer were generated after converting the homer positional weight matrix into a transfac matrix using the RSAT (Nguyen et al., 2018) Metazoa convert matrix tool (http://rsat.sb-roscoff.fr/convert-matrix_form.cgi) and importing to Weblogo3 (Crooks et al., 2004) (http://weblogo.threeplusone.com/) for logo drawing. Identification of similar known motifs to the *de novo* identified motifs was performed using the Tomtom tool of MEME-suite (Gupta et al., 2007). The default parameters were used and the interrogated motif matrices were compared against the JASPAR core 2018 non-redundant database.

Gene set enrichment analysis

Gene sets identified in this study were assessed for enriched gene ontology (GO) terms or association with tissue specific expression using the worbase.org Enrichment Analysis tool (Angeles-Albores et al., 2016; David Angeles-Albores, Raymond Y. N. Lee) (https://wormbase.org/tools/enrichment/tea/tea.cgi), using a *q*-value threshold of 0.1. For significant GO terms presented here the $-\log_{10}q$ value is plotted. Association of gene sets

with biological pathways was evaluated using the gProfiler gOst tool (https://biit.cs.ut.ee/gprofiler/gost) and a g:SCS calculated significance threshold of 0.05.

RNAi by feeding

Knockdown of *nhr-25* was performed by feeding animals on lawns of bacteria expressing dsRNA targeting *nhr-25* (clone commercially available from Source Bioscience). Bacteria were grown overnight and then seeded directly onto NGM plates containing 1 μM IPTG, 25 μg/ml ampicillin and 6.25 μg/ml tetracycline. 5 L4 animals of the strain of interest were transferred on RNAi plates and were allowed to lay progeny that were observed for gene knockdown phenotypic effects at the stage of interest. Control treatments were performed in parallel, with precisely the same experimental conditions, by feeding animals on lawns of the same strain of HT115 bacteria that do not however express dsRNA targeting a gene.

Statistical analysis

Statistical analysis for comparisons between datasets that is not covered in the above paragraphs was performed using GrapPad prims 7 (www.graphpad.com). To test differences in the mean between seam cell scoring or smFISH counting datasets an unpaired two-tailed t-test was performed when the comparison was between two datasets and a one-way Analysis of Variance (ANOVA) was performed when multiple datasets were compared. One-way ANOVA was followed by a Dunnet's post hoc test when the mean of multiple datasets was compared to that of a control. The significance level used throughout is $p<0.05$.

Data availability

All raw sequence files and processed signal files have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (accession number:XXXX).

Supplementary Tables

Table S1. Genome coordinates of significant TaDa peaks for NHR-25 and LIN-22 at the L2 and L4 stage.

Table S2. Assigned genes per TaDa peak for NHR-25 and LIN-22 at the L2 and L4 stage.

Table S3. Strains used in this study.

Table S4. Primers used in this study (general oligos and smFISH).

Table S5. List of transgenes and injection mixes used in this study.

Supplemental Figure Legends

Figure S1: A versatile *C. elegans* TaDa cloning platform and assessment of *in vivo* methylation by Dam-fusions.

(A) Graphic illustration of the design of a versatile *C. elegans* universal TaDa cloning platform with its key features. From left to right the plasmid contains: universal MosSCI recombination sites (ttTi5605_R and L), an LR attR4-attL1 Gateway® cloning site for promoter insertion, an *mCherry* primary ORF followed by 2 x STOP codons, indicated in red, a frameshift in yellow and a unique XmaJI restriction site followed by *dam*, unique restriction cites for PaeI, PacI, NheI and an in-frame STOP codon prior to an *unc-54 3'UTR*. An unc-119 expression cassette is also included to aid screening of transgenics. **(B)** Example of amplification products from methylated gDNA extracted from strains carrying the *dam* fusions described in this study and a WT dam^{-} strain fed on dam^{-} *E. coli.* Note that all samples except for the *dam*$^{-}$ strain show a pronounced 2 kb to 200 bp smear. Wells for all samples were loaded with the same volume of reaction product. Also note the marked increase in methylation observed in the absence of *mCherry* as a primary ORF. Strains grown on *dam*$^{+}$ OP50 show extensive amplification due to bacterial GATC-methylated DNA.

Figure S2: Validation of expression, methylation capacity and functionality of TaDa transgenes.

(A) Confirmation of single-copy transgene expression in the *wrt-2* domain (seam cells) using *mCherry* expression (primary ORF). Animals were imaged at the L4 stage and white arrows mark expression in the seam. **(B-C)** Amplification products from methylated gDNA showing 200-2000 bp smears for the LIN-22 (C) and NHR-25 (D) TaDa experiments. Extractions were performed at L2 and L4 stages, with each combination of Dam-fusion and stage being represented by two biological replicates. The same volume of amplification product was loaded for each sample in each gel. **(D)** Overexpression of the *lin-22:dam* construct as a multi-copy transgene (*icbEx54*) in the *lin-22(icb38)* mutant significantly decreases the number of PDE neurons labelled by *dat-1p::GFP* ($n \geq 31$). **(E)** Overexpression of the *nhr-25:dam* fusion as a multi-copy transgene (*icbEx184*) in the *nhr-25(ku217)* mutant significantly increased the mean seam cell number ($n \geq 31$). Scale bars in A are 20 µm. In D and E error bars indicate standard error of the mean and black stars indicate statistically significant differences in the mean with a t-test (* $p<0.05$, **** $p<0.0001$).

Figure S3: Analysis of sequencing results reveals replicate reproducibility and fusion-dependent methylation.

(A-B) Pearson correlation heatmaps based on normalised aligned read count maps for the LIN-22 (A) and NHR-25 (B) TaDa experiments. The correlation coefficient for each pairwise comparison is printed in each cell of the heatmaps. All Dam-fusions show strong replicate reproducibility and high within-fusion correlation, indicated by high correlation coefficients. Note that TF and control samples show low correlation between them and cluster separately. **(C)** Summary heatmap of Pearson correlations between all samples for the TF TaDa performed in this study. Note the high correlation between control samples from different experiments and low correlation with distinct separate clustering between the *lin-22:dam* and *nhr-25:dam* samples. Samples belonging to the LIN-22 or NHR-25 experiment are indicated with purple or blue font respectively. **(D)** Principal component analysis on normalised aligned

read count maps for all samples shows distinct grouping between TF and control fusions. **(E)** Principal component analysis only on *lin-22:dam* and *nhr-25:dam* replicates shows distinct grouping between the two TF samples. Key in the middle is shared between D and E. **(F)** Example of averaged signal enrichment profiles across chromosome I (locations of protein coding genes are indicated as black blocks) for *lin-22:dam* and *nhr-25:dam* fusions in L2 and L4 stages. The Y-axes represent $\log_2$*(TF:dam/NLS-GFP:dam)* scores. Scale bar length is 2 Mb.

Figure S4: NHR-25 and LIN-22 DNA-binding motifs identified using TaDa peaks.

(A) *De novo* identified DNA-binding motifs for NHR-25 (left) and LIN-22 (right) from TaDa peaks. These motifs were found in 41% of the LIN-22 L2, 33% of the LIN-22 L4, 47% of the NHR-25 L2 and 51% of the NHR-25 L4 total TaDa peaks, with the co-occurrence being statistically significant using Monte Carlo simulation (*p*-values << 0.0001 in all cases)**. (B)** Similar known motifs to the TaDa-identified NHR-25 (left) and LIN-22 (right) motifs are shown from available databases. **(C-D)** Aggregation plots of TaDa signal for all TFs and available stages over instances of the NHR-25 TaDa motif (C) and the LIN-22 TaDa motif (D). Both TFs show strong preference for their respective motif compared to the alternative motif. Regions of ±5 kb around the peak centres have been plotted, the y-axes represent mean enrichment scores for the plotted sequences and shaded areas represent 95% confidence intervals.

Figure S5: TaDa identified targets for NHR-25 and LIN-22 show significant overlaps between them and with other published datasets.

(A) Barplot of all possible intersection sizes between: *nhr-25:dam* L2 and L4 identified target genes, NHR-25 ChIP-seq L1 identified target genes (Shao et al., 2013) and NHR-25 ChIP-seq L2 (Araya et al., 2014) peaks assigned to genes using the same method used in this study. Selected enriched GO-terms for gene sets from the large intersections between TaDa and ChIP-seq L2 are shown. **(B-C)** Tissue-enrichment analysis in the gene-set of putative

targets common between NHR-25 L2 ChIP-seq and TaDa (B) or the gene-set exclusively identified by ChIP-seq (C). In both cases, the top 15 most overrepresented tissues are shown. **(D)** Barplot showing the size of all possible intersections between: the identified target genes for *lin-22:dam* L2 and L4, genes predicted to be downstream genetic interactors of *lin-22* (Zhong and Sternberg, 2006) and *C. elegans* orthologues of HES1 (*H. sapiens*) targets based on ChIP-seq (Encode project accession: ENCSR109ODF). Genes in the pairwise intersection between TaDa and predicted interactors are listed in full above the bars. Selected enriched GO terms for gene sets from the intersections between orthologues of HES1 ChIP-seq targets and TaDa targets are shown. The list of orthologues might contain more than one *C. elegans* homologue for each human gene of the original list. **(E)** Circular plot of all possible intersections between TaDa identified target genes for LIN-22 and NHR-25 at L2 and L4. Selected enriched GO terms are shown for the genes common in all datasets. In A and B, the size of each individual gene-set is printed at the bottom right of each graph. In A, B and C the statistical significance of each intersection assessed with a Fisher's exact test is indicated in the key. All observed intersections shown in this figure were statistically significant ($p < 0.004$).

References:

Angeles-Albores, D., N. Lee, R.Y., Chan, J., and Sternberg, P.W. (2016). Tissue enrichment analysis for *C. elegans* genomics. BMC Bioinformatics *17*, 366.

Antebi, A. (2006). Nuclear hormone receptors in *C. elegans*. WormBook 1–13.

Araya, C.L., Kawli, T., Kundaje, A., Jiang, L., Wu, B., Vafeados, D., Terrell, R., Weissdepp, P., Gevirtzman, L., MacE, D., et al. (2014). Regulatory analysis of the *C. elegans* genome with spatiotemporal resolution. Nature *512*, 400–405.

Askjaer, P., Ercan, S., and Meister, P. (2014). Modern techniques for the analysis of chromatin and nuclear organization in *C. elegans*. WormBook 1–35.

Aughey, G.N., and Southall, T.D. (2016). Dam it's good! DamID profiling of protein-DNA

interactions. Wiley Interdiscip. Rev. Dev. Biol. *5*, 25–37.

Aughey, G.N., Cheetham, S.W., and Southall, T.D. (2019). DamID as a versatile tool for understanding gene regulation. Dev. *146*, 1–8.

Barkoulas, M., van Zon, J.S., Milloz, J., van Oudenaarden, A., and Félix, M.A. (2013). Robustness and Epistasis in the *C. elegans* Vulval Signaling Network Revealed by Pathway Dosage Modulation. Dev. Cell *24*, 64–75.

Barkoulas, M., Vargas Velazquez, A.M., Peluffo, A.E., and Félix, M.A. (2016). Evolution of New cis-Regulatory Motifs Required for Cell-Specific Gene Expression in *Caenorhabditis*. PLoS Genet. *12*, 1–23.

Brabin, C., Appleford, P.J., and 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.

Buck, S.H., Chiu, D., and Saito, R.M. (2009). The cyclin-dependent kinase inhibitors, *cki-1* and *cki-2*, act in overlapping but distinct pathways to control cell cycle quiescence during *C. elegans* development. Cell Cycle *8*, 2613–2620.

Cameron, E.R., and Neil, J.C. (2004). The Runx genes: Lineage-specific oncogenes and tumor suppressors. Oncogene *23*, 4308–4314.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science (80-.). *357*, 661–667.

Cassata, G., Shemer, G., Morandi, P., Donhauser, R., Podbilewicz, B., and Baumeister, R. (2005). ceh-16/engrailed patterns the embryonic epidermis of *Caenorhabditis elegans*. Development *132*, 739–749.

Celniker, S.E., Dillon, L.A.L., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the

genome. Nature *459*, 927–930.

Cheetham, S.W., Gruhn, W.H., Ameele, J. Van Den, Krautz, R., and Southall, T.D. (2018). Targeted DamID reveals differential binding of mammalian pluripotency factors. Development.

Chen, Z., Eastburn, D.J., and Han, M. (2004). The *Caenorhabditis elegans* Nuclear Receptor Gene *nhr-25* Regulates Epidermal Cell Development. Mol. Cell. Biol. *24*, 7345–7358.

Chikina, M.D., and Troyanskaya, O.G. (2012). An effective statistical evaluation of chipseq dataset similarity. Bioinformatics *28*, 607–613.

Chisholm, A.D., and Hsiao, T.I. (2012). The *Caenorhabditis elegans* epidermis as a model skin. I: Development, patterning, and growth. Wiley Interdiscip. Rev. Dev. Biol.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188–1190.

David Angeles-Albores, Raymond Y. N. Lee, J.C. and P.W.S. (2017) Phenotype and gene ontology enrichment as guides for disease modeling in C. elegans. Manuscr. Prep.

Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. Nature *468*, 911–920.

Ding, S.S., and Woollard, A. (2017). Non-muscle myosin II is required for correct fate specification in the *Caenorhabditis elegans* seam cell divisions. Sci. Rep. *7*, 3524.

Ferré, Q., Charbonnier, G., Sadouni, N., Lopez, F., Kermezli, Y., Spicuglia, S., Capponi, C., Ghattas, B., and Puthier, D. (2019). OLOGRAM: determining significance of total overlap length between genomic regions sets. Bioinformatics *36*, 1920–1922.

Fornes, O., Castro-Mondragon, J.A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. Nucleic Acids Res. *48*,

D87–D92.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. *32*, 273–279.

Frokjaer-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovski, A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in C. elegans using a modified Mosl transposon. Nat. Methods *11*, 529–534.

Fuxman Bass, J.I., Tamburino, A.M., Mori, A., Beittel, N., Weirauch, M.T., Reece-Hoyes, J.S., and Walhout, A.J.M. (2014). Transcription factor binding to *Caenorhabditis elegans* first introns reveals lack of redundancy with gene promoters. Nucleic Acids Res. *42*, 153–162.

Gilleard, J.S., and Mcghee, J.D. (2001). Activation of Hypodermal Differentiation in the *Caenorhabditis elegans* Embryo by GATA Transcription Factors ELT-1 and ELT-3. Mol. Cell. Biol. *Vol. 21*, 2533–2544.

Gissendanner, C.R., and Sluder, A.E. (2000). *nhr-25*, the *Caenorhabditis elegans* ortholog of ftz-f1, is required for epidermal and somatic gonad development. Dev. Biol. *221*, 259–272.

Gleason, J.E., and Eisenmann, D.M. (2010). Wnt signaling controls the stem cell-like asymmetric division of the epithelial seam cells during *C. elegans* larval development. Dev. Biol.

Gorrepati, L., Thompson, K.W., and Eisenmann, D.M. (2013). *C. elegans* GATA factors EGL-18 and ELT-6 function downstream of Wnt signaling to maintain the progenitor fate during larval asymmetric divisions of the seam cells. Development *140*, 2093–2102.

Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L., and Walhout, A.J.M. (2009). A Multiparameter Network Reveals Extensive Divergence between *C. elegans* bHLH Transcription Factors. Cell *138*, 314–327.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying

similarity between motifs. Genome Biol. *8*.

Hajduskova, M., Jindra, M., Herman, M.A., and Asahina, M. (2009). The nuclear receptor NHR-25 cooperates with the Wnt/β-catenin asymmetry pathway to control differentiation of the T seam cell in *C. elegans*. J. Cell Sci. *122*, 3051–3060.

Hayes, G.D., Frand, A.R., and Ruvkun, G. (2006). The *mir-84* and *let-7* paralogous microRNA genes of *Caenorhabditis elegans* direct the cessation of molting via the conserved nuclear hormone receptors NHR-23 and NHR-25. Development *133*, 4631–4641.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hong, Y., Roy, R., and Ambros, V. (1998). Developmental regulation of a cyclin-dependent kinase inhibitor controls postembryonic cell cycle progression in *Caenorhabditis elegans*. Development *125*, 3585–3597.

van der Horst, S.E.M., Cravo, J., Woollard, A., Teapal, J., and van den Heuvel, S. (2019). *C. elegans* Runx/CBFβ suppresses POP-1 TCF to convert asymmetric to proliferative division of stem cell-like seam cells. Development *146*, 1–14.

Hughes, S., Brabin, C., Appleford, P.J., and Woollard, A. (2013). CEH-20/Pbx and UNC-62/Meis function upstream of *rnt-1*/Runx to regulate asymmetric divisions of the C. elegans stem-like seam cells. Biol. Open *2*, 718–727.

Jänes, J., Dong, Y., Schoof, M., Serizay, J., Appert, A., Cerrato, C., Woodbury, C., Chen, R., Gemma, C., Huang, N., et al. (2018). Chromatin accessibility dynamics across *C. elegans* development and ageing. Elife *7*, 267–272.

Joshi, P.M., Riddle, M.R., Djabrayan, N.J. V, and Rothman, J.H. (2010). *Caenorhabditis elegans* as a model for stem cell biology. Dev. Dyn. *239*, 1539–1554.

Kageyama, R., Ohtsuka, T., and Kobayashi, T. (2007). The Hes gene family: Repressors and oscillators that orchestrate embryogenesis. Development *134*, 1243–1251.

Kagoshima, H., Nimmo, R., Saad, N., Tanaka, J., Miwa, Y., Mitani, S., Kohara, Y., and Woollard, A. (2007). The *C. elegans* CBF homologue BRO-1 interacts with the Runx factor, RNT-1, to promote stem cell proliferation and self-renewal. Development *134*, 3905–3915.

Katsanos, D., Koneru, S.L., Mestek Boukhibar, L., Gritti, N., Ghose, R., Appleford, P.J., Doitsidou, M., Woollard, A., van Zon, J.S., Poole, R.J., et al. (2017). Stochastic loss and gain of symmetric divisions in the *C. elegans* epidermis perturbs robustness of stem cell number. PLoS Biol. *15*, 1–31.

Koh, K., and Rothman, J.H. (2001). ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. Development *128*, 2867–2880.

Kondili, M., Fust, A., Preussner, J., Kuenne, C., Braun, T., and Looso, M. (2017). UROPA: A tool for Universal RObust Peak Annotation. Sci. Rep. *7*, 1–12.

Kozak, M. (1987). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. Mol. Cell. Biol. *7*, 3438–3445.

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. Nucleic Acids Res. *29*, 5226–5232.

Kudron, M.M., Victorsen, A., Gevirtzman, L., Hillier, L.W., Fisher, W.W., Vafeados, D., Kirkey, M., Hammonds, A.S., Gersch, J., Ammouri, H., et al. (2018). The modern resource: genome-wide binding profiles for hundreds of *Drosophila* and *Caenorhabditis elegans* transcription factors. Genetics *208*, 937–949.

Lam, A.K., and Phillips, B.T. (2017). Wnt signaling polarizes *C. elegans* asymmetric cell divisions during development. In Results and Problems in Cell Differentiation, pp. 83–114.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat.

Methods *9*, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lichtenberg, K.H. de, Funa, N., Nakic, N., Ferrer, J., Zhu, Z., Huangfu, D., and Serup, P. (2018). Genome-Wide Identification of HES1 Target Genes Uncover Novel Roles for HES1 in Pancreatic Development. BioRxiv 335869.

Marshall, O.J., and Brand, A.H. (2015). Damidseq-pipeline: An automated pipeline for processing DamID sequencing datasets. Bioinformatics *31*, 3371–3373.

Marshall, O.J., Southall, T.D., Cheetham, S.W., and Brand, A.H. (2016). Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. Nat. Protoc. *11*, 1586–1598.

Matus, D.Q., Lohmer, L.L., Kelley, L.C., Schindler, A.J., Kohrman, A.Q., Barkoulas, M., Zhang, W., Chi, Q., and Sherwood, D.R. (2015). Invasive Cell Fate Requires G1 Cell-Cycle Arrest and Histone Deacetylase-Mediated Changes in Gene Expression. Dev. Cell *35*, 162–174.

Meyer, C.A., and Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. *15*, 709–721.

Miyabayashi, T., Palfreyman, M.T., Sluder, a E., Slack, F., and Sengupta, P. (1999). Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. Dev. Biol. *215*, 314–331.

Murata, K., Hattori, M., Hirai, N., Shinozuka, Y., Hirata, H., Kageyama, R., Sakai, T., and Minato, N. (2005). Hes1 directly controls cell proliferation through the transcriptional repression of p27Kip1. Mol. Cell. Biol. *25*, 4262–4271.

Nguyen, N.T.T., Contreras-Moreira, B., Castro-Mondragon, J.A., Santana-Garcia, W., Ossio,

R., Robles-Espinoza, C.D., Bahin, M., Collombet, S., Vincens, P., Thieffry, D., et al. (2018). RSAT 2018: Regulatory sequence analysis tools 20th anniversary. Nucleic Acids Res. *46*, W209–W214.

Nimmo, R., Antebi, A., and Woollard, A. (2005). *mab-2* encodes RNT-1, a *C. elegans* Runx homologue essential for controlling cell proliferation in a stem cell-like developmental lineage. Development *132*, 5043–5054.

O'Connell, K.F. (2010). Reverse chunking, a simple and effective method for identifying *unc-119(+)* transgenics. Worm Breeder's Gaz. *Volume 18*.

Oliveri, P., and Davidson, E.H. (2007). Built to run, not fail. Science (80-.). *315*, 1510–1511.

Piano, F., Gunsalus, K.C., Hill, D.E., and Vidal, M. (2006). *C. elegans* network biology: a beginning. WormBook 1–20.

Ramialison, M., Waardenberg, A.J., Schonrock, N., Doan, T., de Jong, D., Bouveret, R., and Harvey, R.P. (2017). Analysis of steric effects in DamID profiling of transcription factor target genes. Genomics *109*, 75–82.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Sasai, Y., Kageyama, R., Tagawa, Y., Shigemoto, R., and Nakanishi, S. (1992). Two mammalian helix-loop-helix factors structurally related to *Drosophila hairy* and *Enhancer of split*. Genes Dev. *6*, 2620–2634.

Sawa, H., and Korswagen, H.C. (2013). Wnt signaling in *C. elegans*. WormBook 1–18.

Schlager, B., Röseler, W., Zheng, M., Gutierrez, A., and Sommer, R.J. (2006). HAIRY-like transcription factors and the evolution of the nematode vulva equivalence group. Curr. Biol. *16*, 1386–1394.

Schuster, E., McElwee, J.J., Tullet, J.M.A., Doonan, R., Matthijssens, F., Reece-Hoyes, J.S., Hope, I.A., Vanfleteren, J.R., Thornton, J.M., and Gems, D. (2010). DamID in *C. elegans* reveals longevity-associated targets of DAF-16/FoxO. Mol. Syst. Biol. *6*, 1–6.

Sen, S.Q., Chanchani, S., Southall, T.D., and Doe, C.Q. (2019). Neuroblast-specific chromatin landscapes allows the integration of spatial and temporal cues during *Drosophila* neurogenesis. Elife *8*, 399–404.

Shao, J., He, K., Wang, H., Ho, W.S., Ren, X., An, X., Wong, M.K., Yan, B., Xie, D., Stamatoyannopoulos, J., et al. (2013). Collaborative regulation of development but independent control of metabolism by two epidermis-specific transcription factors in *Caenorhabditis elegans*. J. Biol. Chem. *288*, 33411–33426.

Šilhánková, M., Jindra, M., and Asahina, M. (2005). Nuclear receptor NHR-25 is required for cell-shape dynamics during epidermal differentiation in *Caenorhabditis elegans*. J. Cell Sci. *118*, 223–232.

Smith, J.A., McGarr, P., and Gilleard, J.S. (2005). The *Caenorhabditis elegans* GATA factor *elt-1* is essential for differentiation and maintenance of hypodermal seam cells and for normal locomotion. J. Cell Sci. *118*, 5709–5719.

Southall, T., Gold, K., Egger, B., Davidson, C., Caygill, E., Marshall, O., and Brand, A. (2013). Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: Assaying RNA pol II occupancy in neural stem cells. Dev. Cell.

Southall, T.D., Davidson, C.M., Miller, C., Carr, A., and Brand, A.H. (2014). Dedifferentiation of neurons precedes tumor formation in lola mutants. Dev. Cell *28*, 685–696.

Spieth, J., Hillier, L.W., and Wilson, R.K. (2005). Evolutionarily conserved elements in vertebrate , insect , worm , and yeast genomes. Genome Res. 1034–1050.

Spieth, J., Lawson, D., Davis, P., Williams, G., and Howe, K. (2014). Overview of gene structure in *C. elegans*. WormBook 1–18.

van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat. Biotechnol. *18*, 424–428.

van Steensel, B., Delrow, J., and Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. *27*, 304–308.

Stempor, P., and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res. *1*, 14.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Dev. Biol. *56*, 110–156.

Teytelman, L., Thurtle, D.M., Rine, J., and Van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Proc. Natl. Acad. Sci. U. S. A. *110*, 18602–18607.

Thompson, K.W., Joshi, P., Dymond, J.S., Gorrepati, L., Smith, H.E., Krause, M.W., and Eisenmann, D.M. (2016). The Paired-box protein PAX-3 regulates the choice between lateral and ventral epidermal cell fates in *C. elegans*. Dev. Biol. *412*, 191–207.

Vissers, J.H.A., Froldi, F., Schröder, J., Papenfuss, A.T., Cheng, L.Y., and Harvey, K.F. (2018). The Scalloped and Nerfin-1 transcription factors cooperate to maintain neuronal cell fate. Cell Rep. *25*, 1561-1576.e7.

Wang, M., Zhao, Y., and Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. Sci. Rep. *5*, 1–12.

Wrischnik, L. a, and Kenyon, C.J. (1997). The role of *lin-22, a hairy/enhancer of split* homolog, in patterning the peripheral nervous system of *C. elegans*. Development *124*, 2875–2888.

Zheng, R., and Blobel, G.A. (2010). GATA transcription factors and cancer. Genes and Cancer *1*, 1178–1188.

Zhong, W., and Sternberg, P.W. (2006). Genome-wide prediction of *C. elegans* genetic interactions. Science (80-.). *311*, 1481–1484.

Figure 1

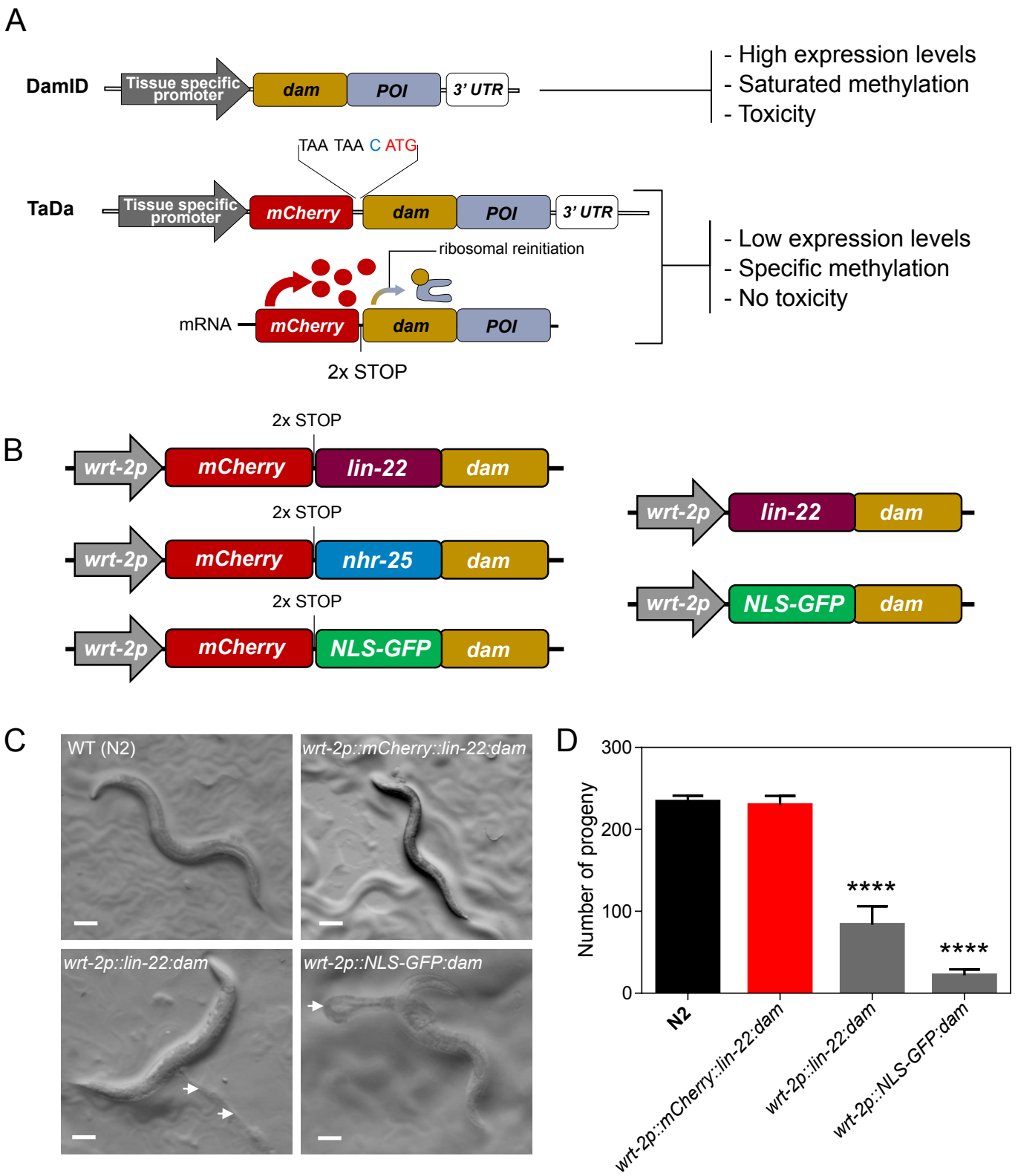

A
DamID
Tissue specific promoter
dam
POI
3' UTR
- High expression levels
- Saturated methylation
- Toxicity
TAA TAA C ATG
TaDa
Tissue specific promoter
mCherry
dam
POI
3' UTR
ribosomal reinitiation
mRNA
mCherry
dam
POI
2x STOP
- Low expression levels
- Specific methylation
- No toxicity
B
2x STOP
wrt-2p
mCherry
lin-22
dam
2x STOP
wrt-2p
mCherry
nhr-25
dam
2x STOP
wrt-2p
mCherry
NLS-GFP
dam
wrt-2p
lin-22
dam
wrt-2p
NLS-GFP
dam
C
WT (N2)
wrt-2p::mCherry::lin-22:dam
wrt-2p::lin-22:dam
wrt-2p::NLS-GFP:dam
D
Number of progeny
300
200
100
0
****
****
N2
wrt-2p::mCherry::lin-22:dam
wrt-2p::lin-22:dam
wrt-2p::NLS-GFP:dam

Figure 2

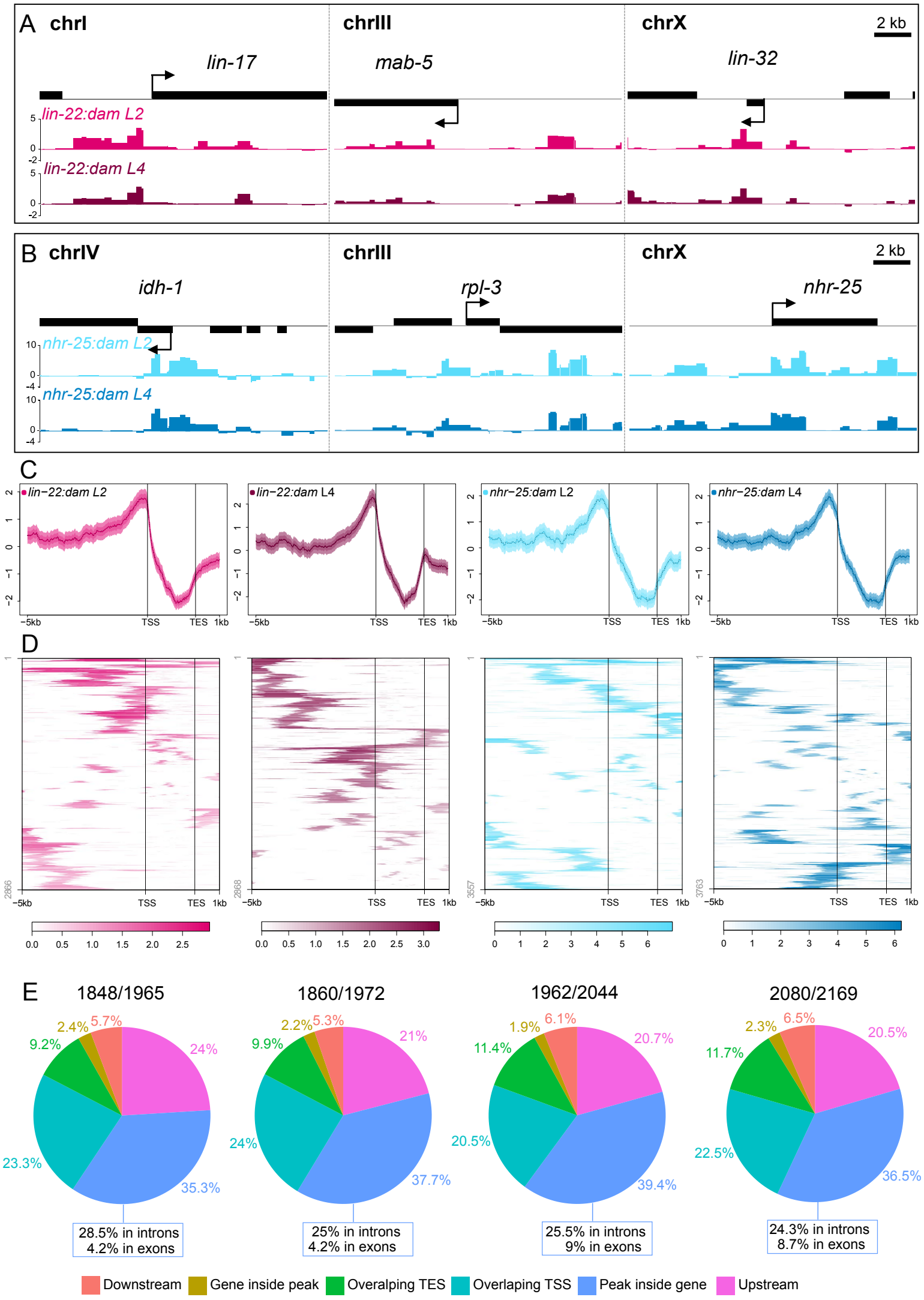
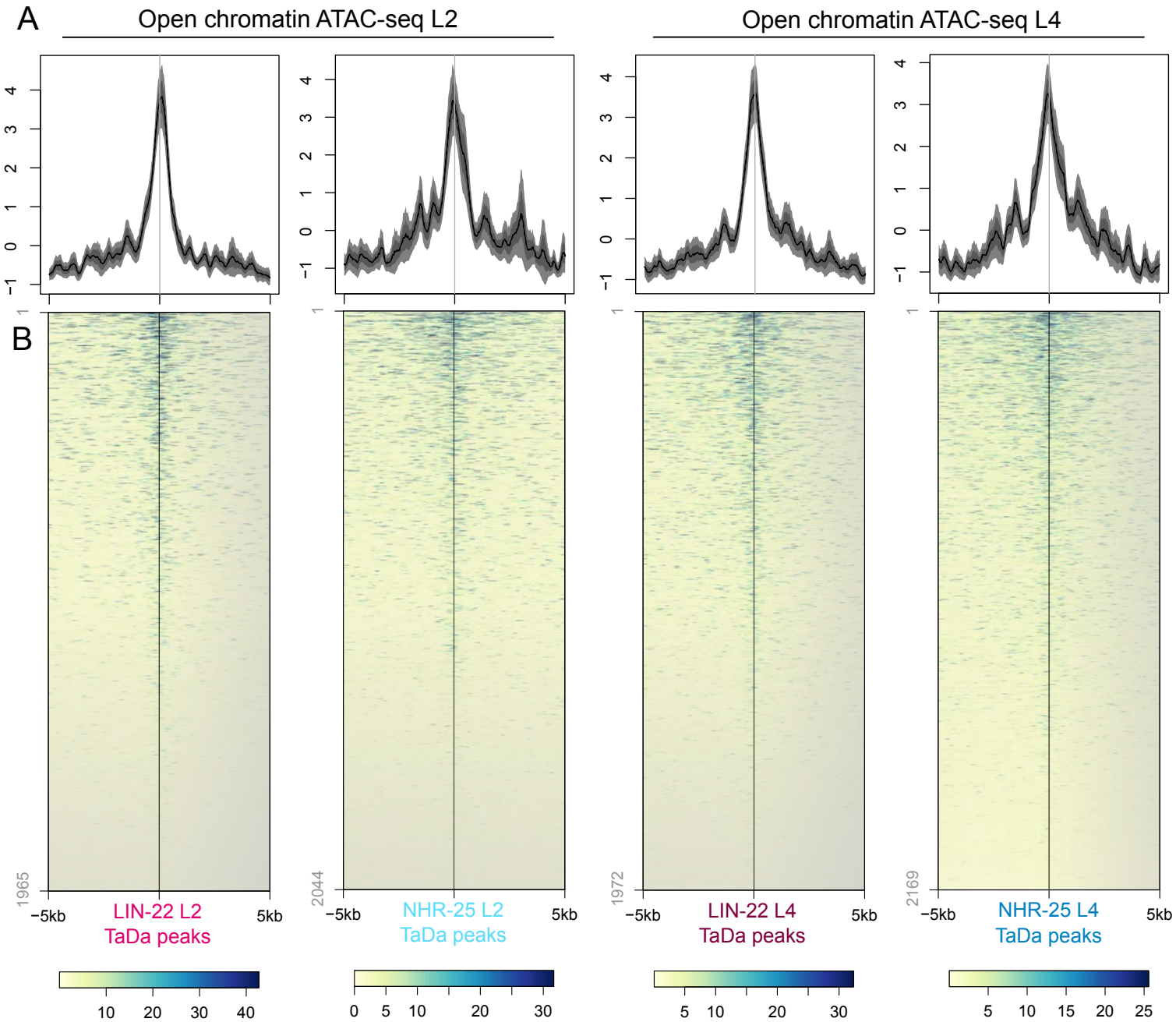

A
chrI
chrIII
chrX
2 kb
lin-17
mab-5
lin-32
lin-22:dam L2
lin-22:dam L4
B
chrIV
chrIII
chrX
2 kb
idh-1
rpl-3
nhr-25
nhr-25:dam L2
nhr-25:dam L4
C
lin−22:dam L2
lin−22:dam L4
nhr−25:dam L2
nhr−25:dam L4
−5kb
TSS
TES
1kb
D
E
1848/1965
1860/1972
1962/2044
2080/2169
28.5% in introns
4.2% in exons
25% in introns
4.2% in exons
25.5% in introns
9% in exons
24.3% in introns
8.7% in exons
Downstream
Gene inside peak
Overalping TES
Overlaping TSS
Peak inside gene
Upstream

Figure 3


A
Open chromatin ATAC-seq L2
Open chromatin ATAC-seq L4
B
1965
2044
1972
2169
−5kb
5kb
LIN-22 L2
TaDa peaks
NHR-25 L2
TaDa peaks
LIN-22 L4
TaDa peaks
NHR-25 L4
TaDa peaks
10 20 30 40
0 5 10 20 30
5 10 20 30
5 10 15 20 25

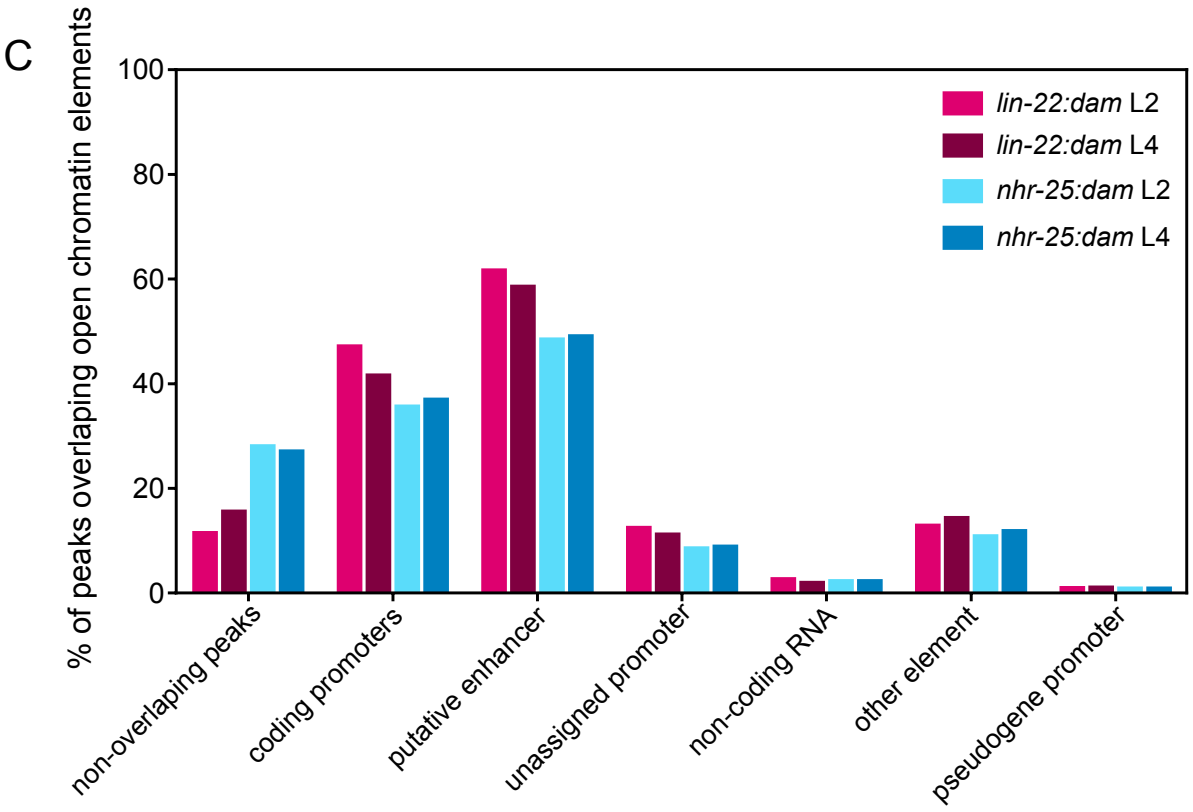
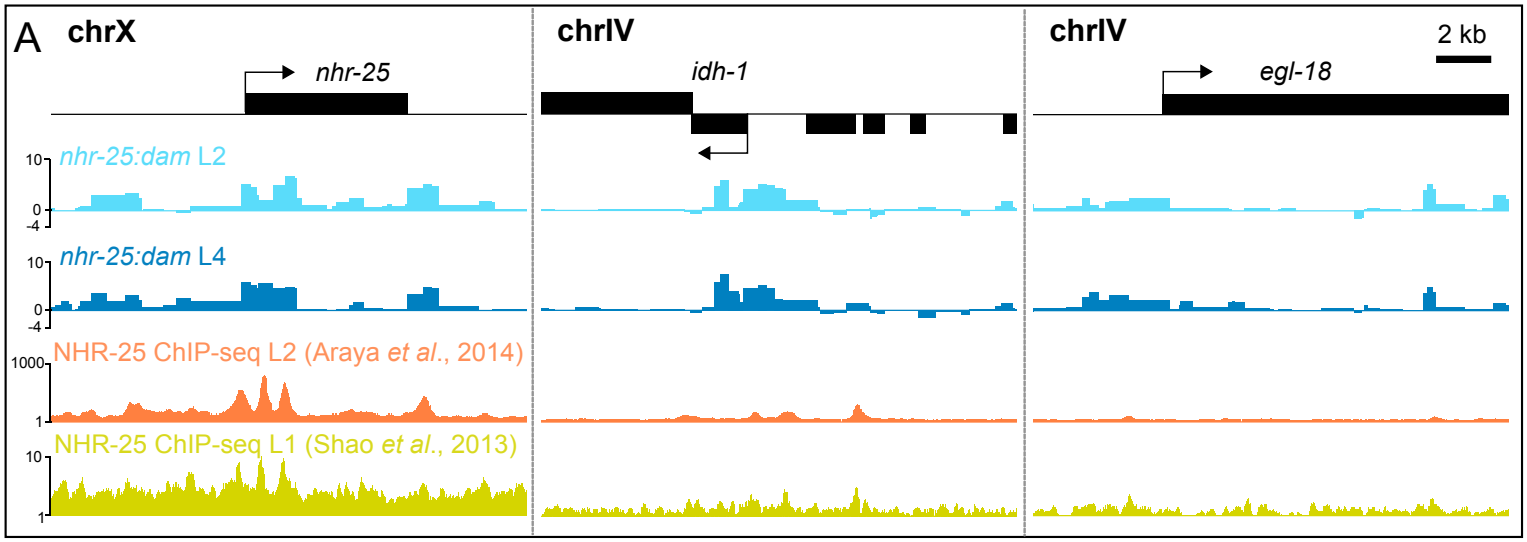
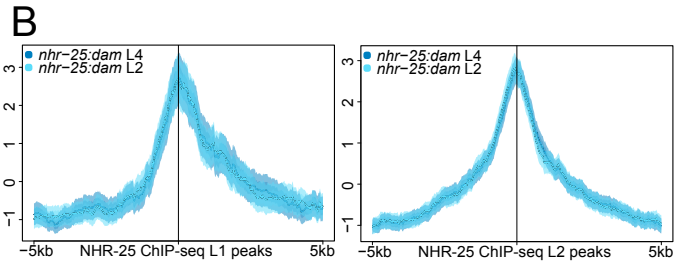
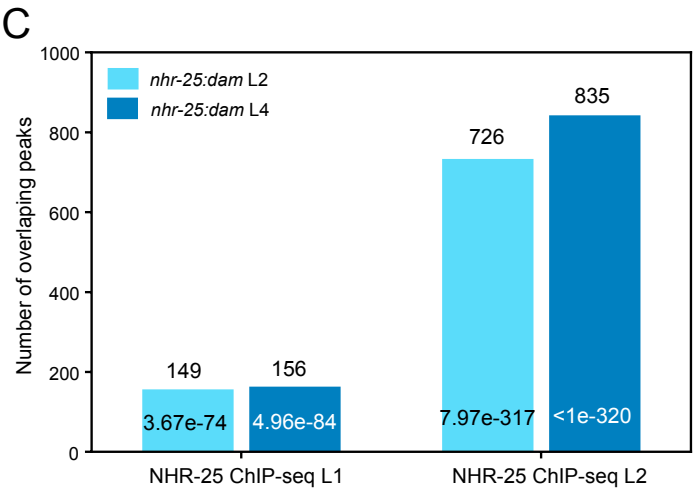
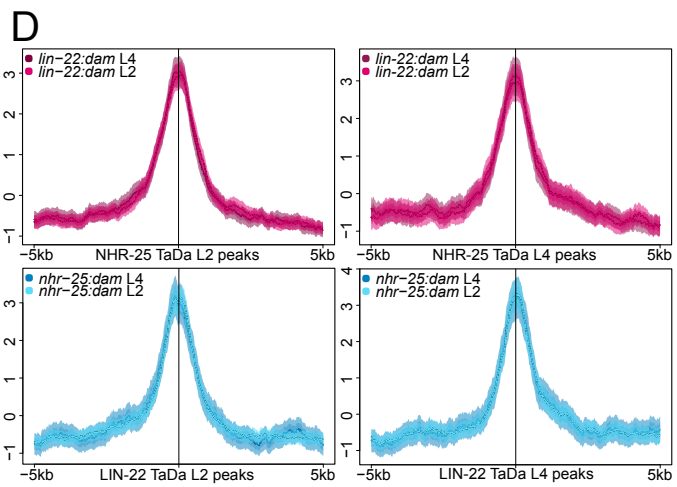

C
% of peaks overlaping open chromatin elements
100
80
60
40
20
0
lin-22:dam L2
lin-22:dam L4
nhr-25:dam L2
nhr-25:dam L4
non-overlaping peaks
coding promoters
putative enhancer
unassigned promoter
non-coding RNA
other element
pseudogene promoter

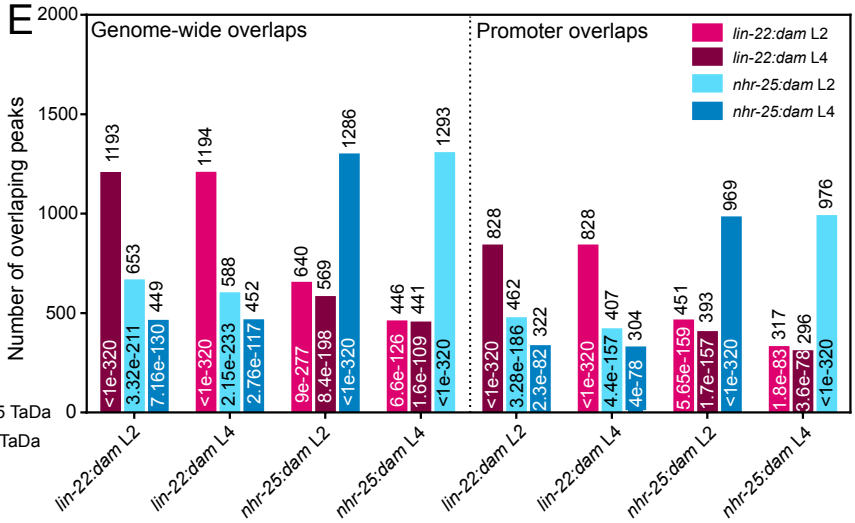
Figure 4

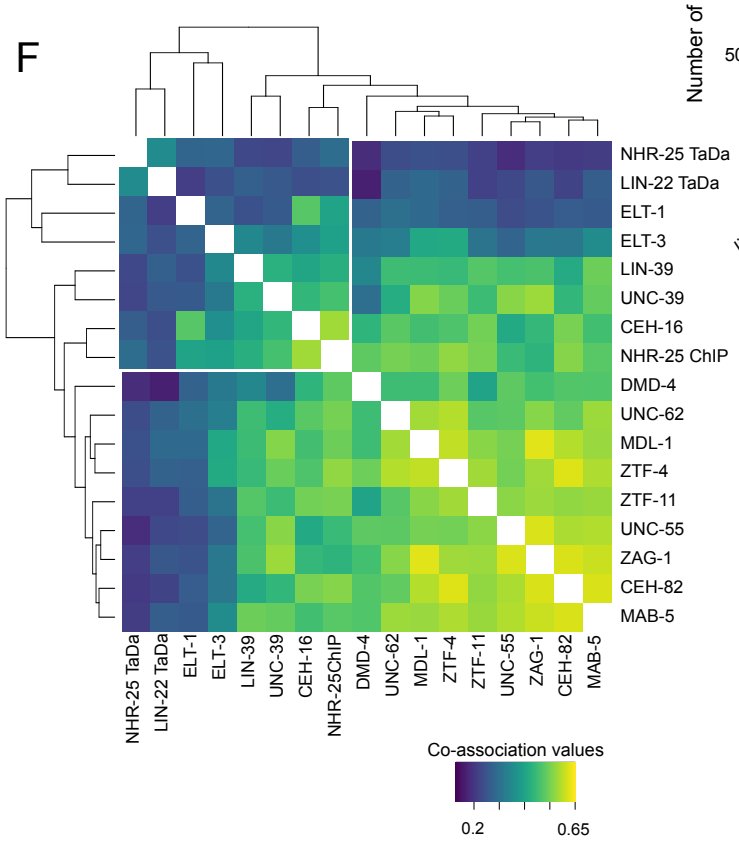

A
chrX
chrIV
chrIV
2 kb
nhr-25
idh-1
egl-18
nhr-25:dam L2
nhr-25:dam L4
NHR-25 ChIP-seq L2 (Araya et al., 2014)
NHR-25 ChIP-seq L1 (Shao et al., 2013)


B
nhr−25:dam L4
nhr−25:dam L2
−5kb
NHR-25 ChIP-seq L1 peaks
5kb
NHR-25 ChIP-seq L2 peaks


C
nhr-25:dam L2
nhr-25:dam L4
Number of overlaping peaks
149
156
726
835
3.67e-74
4.96e-84
7.97e-317
<1e-320
NHR-25 ChIP-seq L1
NHR-25 ChIP-seq L2


D
lin−22:dam L4
lin−22:dam L2
nhr−25:dam L4
nhr−25:dam L2
−5kb
NHR-25 TaDa L2 peaks
NHR-25 TaDa L4 peaks
LIN-22 TaDa L2 peaks
LIN-22 TaDa L4 peaks
5kb


E
Genome-wide overlaps
Promoter overlaps
lin-22:dam L2
lin-22:dam L4
nhr-25:dam L2
nhr-25:dam L4
Number of overlaping peaks
1193
653
449
1194
588
452
640
569
1286
446
441
1293
828
462
322
828
407
304
451
393
969
317
296
976
<1e-320
3.32e-211
7.16e-130
2.15e-233
2.76e-117
9e-277
8.4e-198
6.6e-126
1.6e-109
3.28e-186
2.3e-82
4.4e-157
4e-78
5.65e-159
1.7e-157
1.8e-83
3.6e-78


F
NHR-25 TaDa
LIN-22 TaDa
ELT-1
ELT-3
LIN-39
UNC-39
CEH-16
NHR-25 ChIP
DMD-4
UNC-62
MDL-1
ZTF-4
ZTF-11
UNC-55
ZAG-1
CEH-82
MAB-5
Co-association values
0.2
0.65

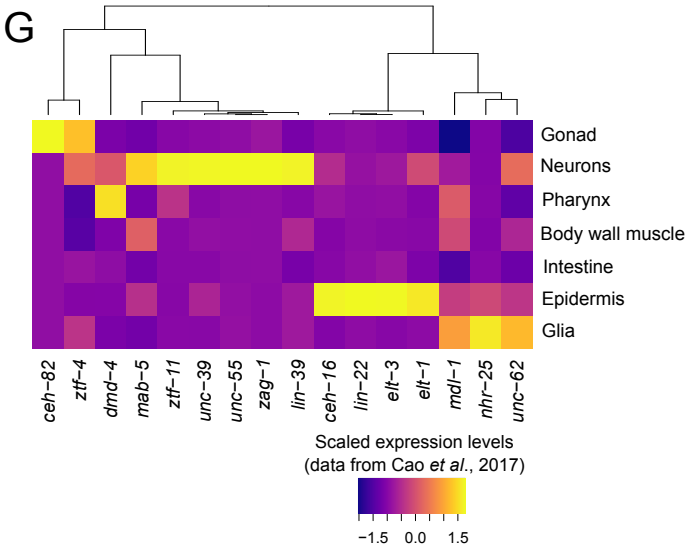

G
Gonad
Neurons
Pharynx
Body wall muscle
Intestine
Epidermis
Glia
ceh−82
ztf−4
dmd−4
mab−5
ztf−11
unc−39
unc−55
zag−1
lin−39
ceh−16
lin−22
elt−3
elt−1
mdl−1
nhr−25
unc−62
Scaled expression levels
(data from Cao et al., 2017)
−1.5
0.0
1.5

Figure 5

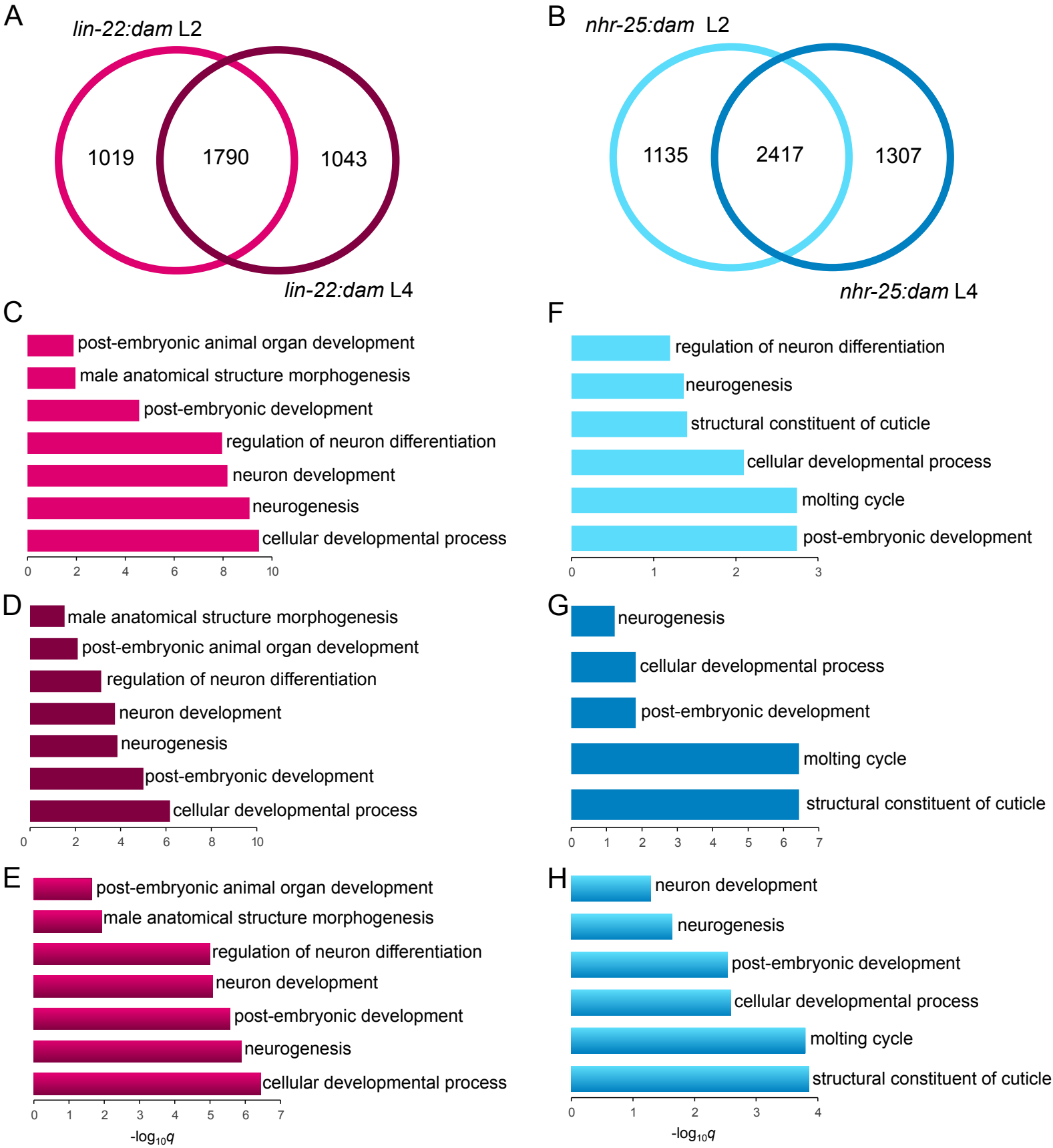

A
lin-22:dam L2
1019
1790
1043
lin-22:dam L4
B
nhr-25:dam L2
1135
2417
1307
nhr-25:dam L4
C
post-embryonic animal organ development
male anatomical structure morphogenesis
post-embryonic development
regulation of neuron differentiation
neuron development
neurogenesis
cellular developmental process
0 2 4 6 8 10
D
male anatomical structure morphogenesis
post-embryonic animal organ development
regulation of neuron differentiation
neuron development
neurogenesis
post-embryonic development
cellular developmental process
0 2 4 6 8 10
E
post-embryonic animal organ development
male anatomical structure morphogenesis
regulation of neuron differentiation
neuron development
post-embryonic development
neurogenesis
cellular developmental process
0 1 2 3 4 5 6 7
-log10q
F
regulation of neuron differentiation
neurogenesis
structural constituent of cuticle
cellular developmental process
molting cycle
post-embryonic development
0 1 2 3
G
neurogenesis
cellular developmental process
post-embryonic development
molting cycle
structural constituent of cuticle
0 1 2 3 4 5 6 7
H
neuron development
neurogenesis
post-embryonic development
cellular developmental process
molting cycle
structural constituent of cuticle
0 1 2 3 4
-log10q

Figure 6

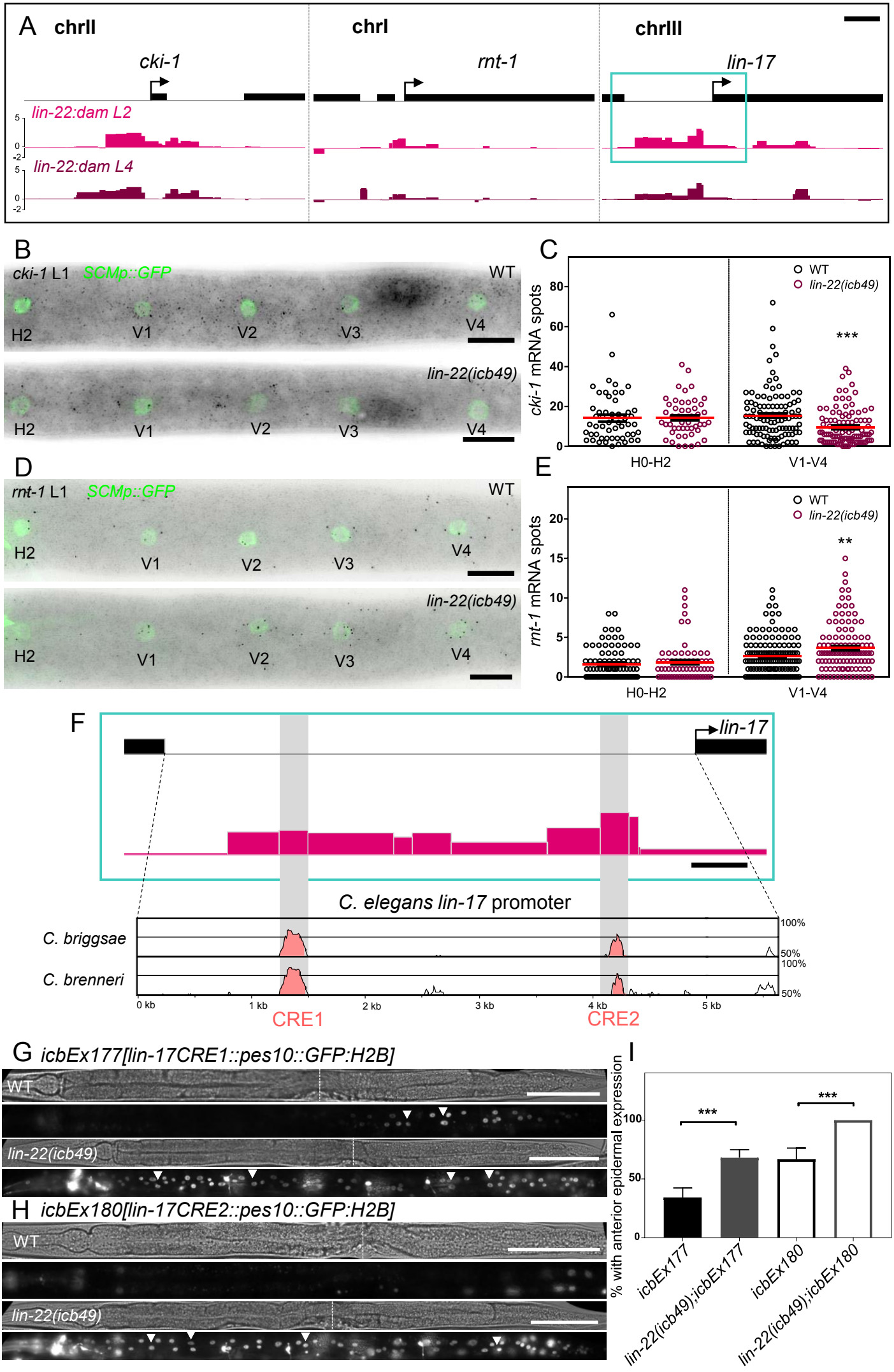
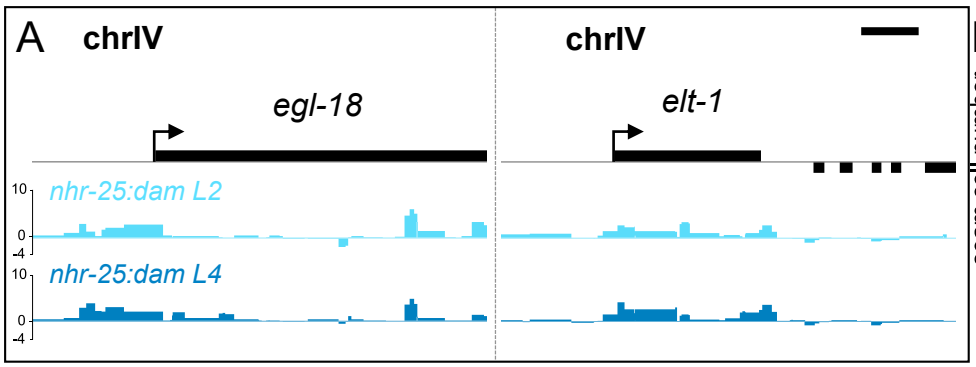
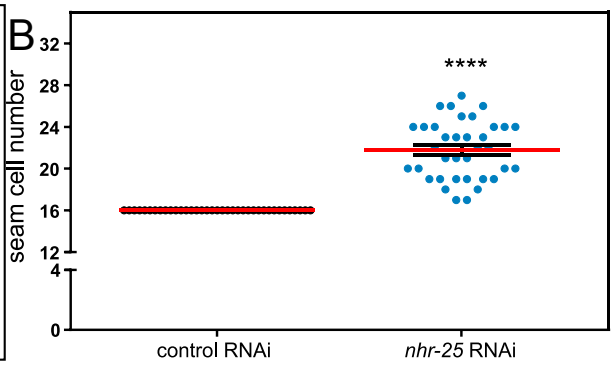
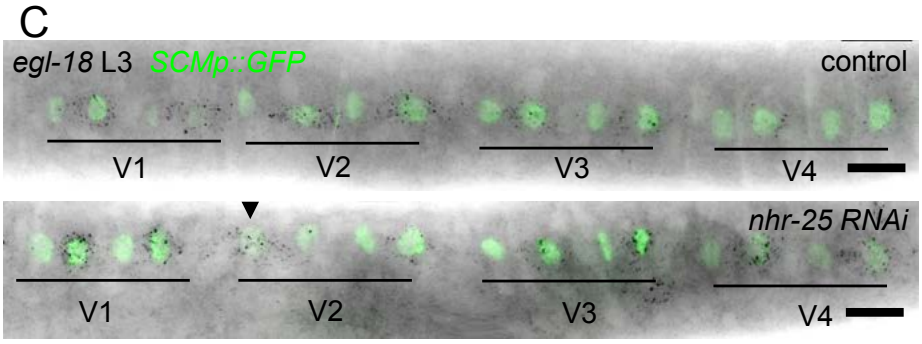
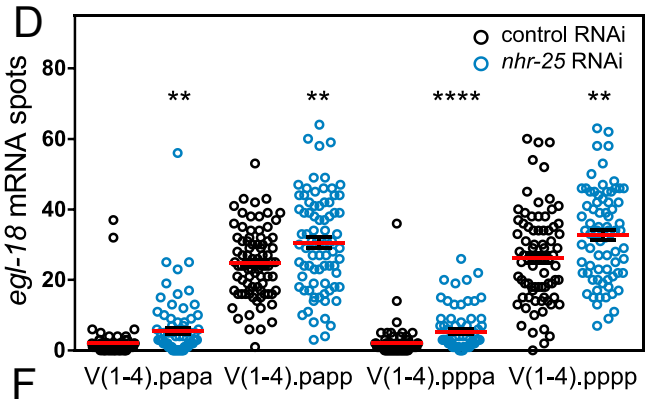

A
chrII
chrI
chrIII
cki-1
rnt-1
lin-17
lin-22:dam L2
lin-22:dam L4
B
cki-1 L1
SCMp::GFP
WT
lin-22(icb49)
H2
V1
V2
V3
V4
C
cki-1 mRNA spots
WT
lin-22(icb49)
***
H0-H2
V1-V4
D
rnt-1 L1
SCMp::GFP
WT
lin-22(icb49)
E
rnt-1 mRNA spots
WT
lin-22(icb49)
**
H0-H2
V1-V4
F
lin-17
C. elegans lin-17 promoter
C. briggsae
C. brenneri
100%
50%
0 kb
1 kb
2 kb
3 kb
4 kb
5 kb
CRE1
CRE2
G icbEx177[lin-17CRE1::pes10::GFP:H2B]
WT
lin-22(icb49)
H icbEx180[lin-17CRE2::pes10::GFP:H2B]
WT
lin-22(icb49)
I
% with anterior epidermal expression
***
***
icbEx177
lin-22(icb49);icbEx177
icbEx180
lin-22(icb49);icbEx180

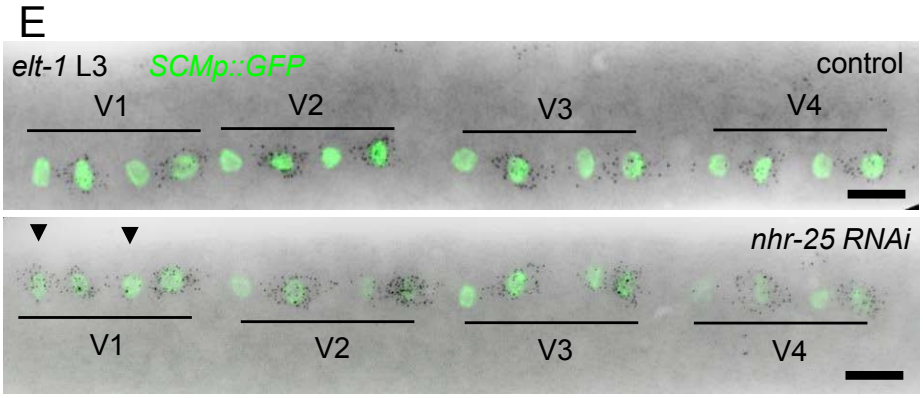
Figure 7

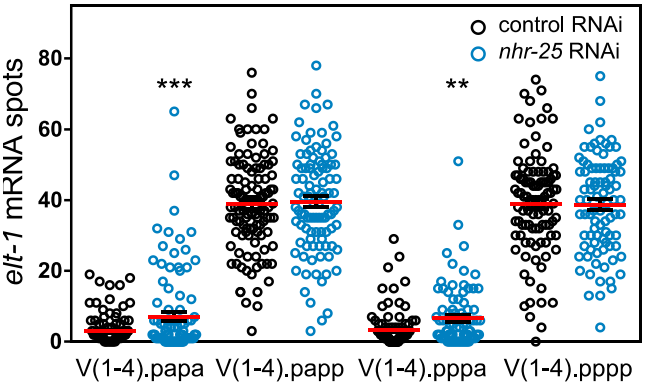

A
chrIV
chrIV
egl-18
elt-1
nhr-25:dam L2
nhr-25:dam L4


B
seam cell number
****
control RNAi
nhr-25 RNAi


C
egl-18 L3 SCMp::GFP
control
V1
V2
V3
V4
nhr-25 RNAi


D
control RNAi
nhr-25 RNAi
egl-18 mRNA spots
**
**
****
**
V(1-4).papa
V(1-4).papp
V(1-4).pppa
V(1-4).pppp


E
elt-1 L3 SCMp::GFP
control
V1
V2
V3
V4
nhr-25 RNAi


F
control RNAi
nhr-25 RNAi
elt-1 mRNA spots
***
**
V(1-4).papa
V(1-4).papp
V(1-4).pppa
V(1-4).pppp

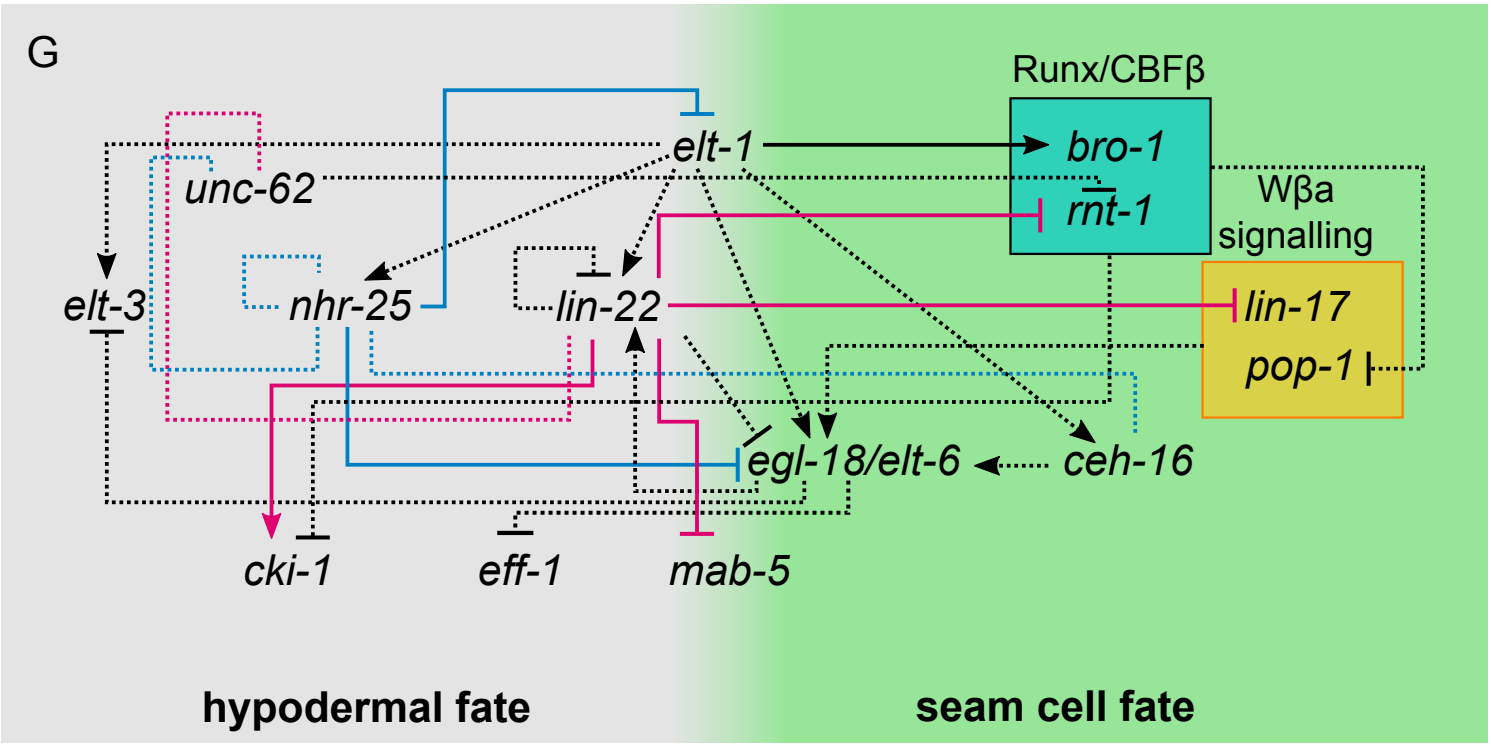
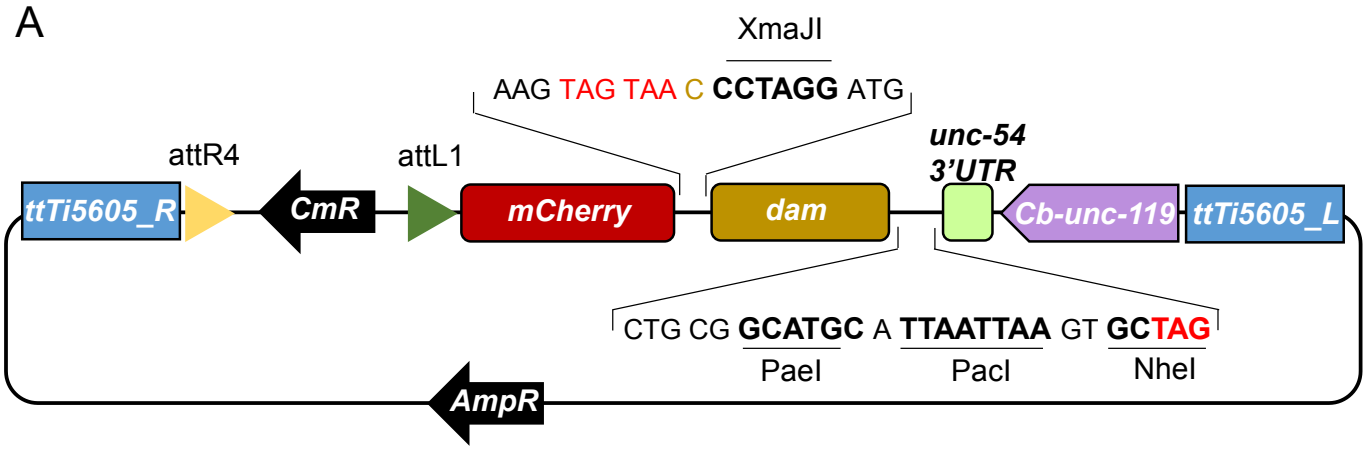

G
Runx/CBFβ
elt-1
bro-1
rnt-1
unc-62
Wβa signalling
elt-3
nhr-25
lin-22
lin-17
pop-1
egl-18/elt-6
ceh-16
cki-1
eff-1
mab-5
hypodermal fate
seam cell fate

Figure S1

A


XmaJI
AAG TAG TAA C CCTAGG ATG
attR4
attL1
unc-54
3'UTR
ttTi5605_R
CmR
mCherry
dam
Cb-unc-119
ttTi5605_L
CTG CG GCATGC A TTAATTAA GT GCTAG
PaeI
PacI
NheI
AmpR

B

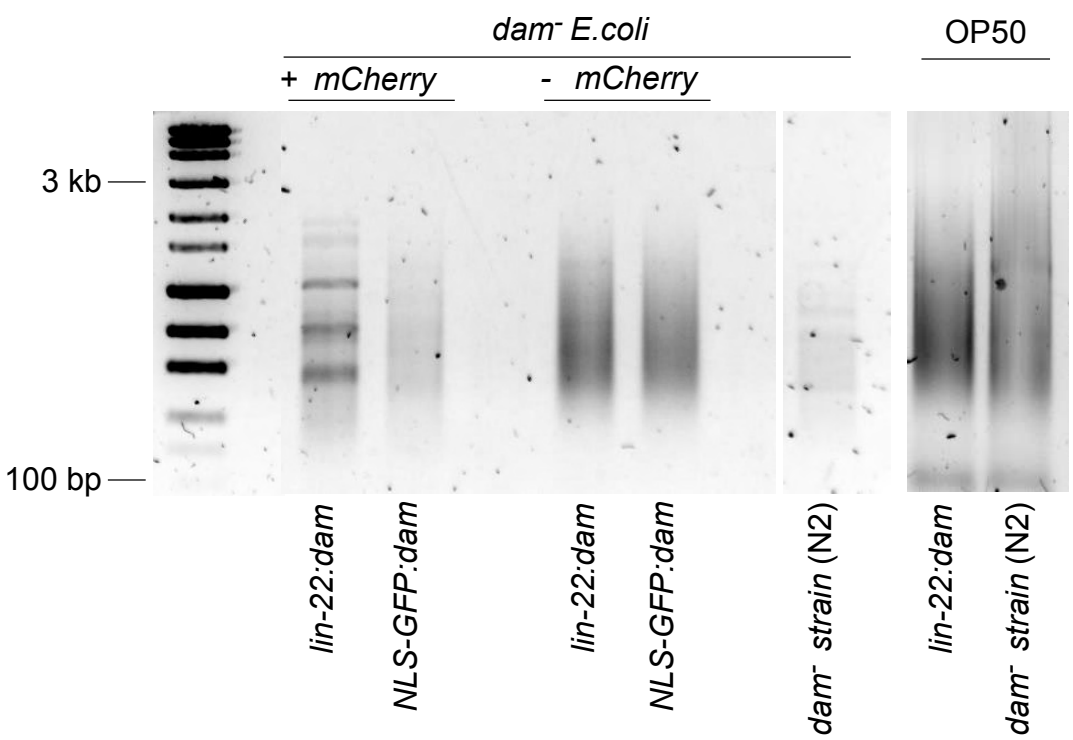

dam⁻ E.coli
OP50
+ mCherry
- mCherry
3 kb
100 bp
lin-22:dam
NLS-GFP:dam
lin-22:dam
NLS-GFP:dam
dam⁻ strain (N2)
lin-22:dam
dam⁻ strain (N2)

Figure S2

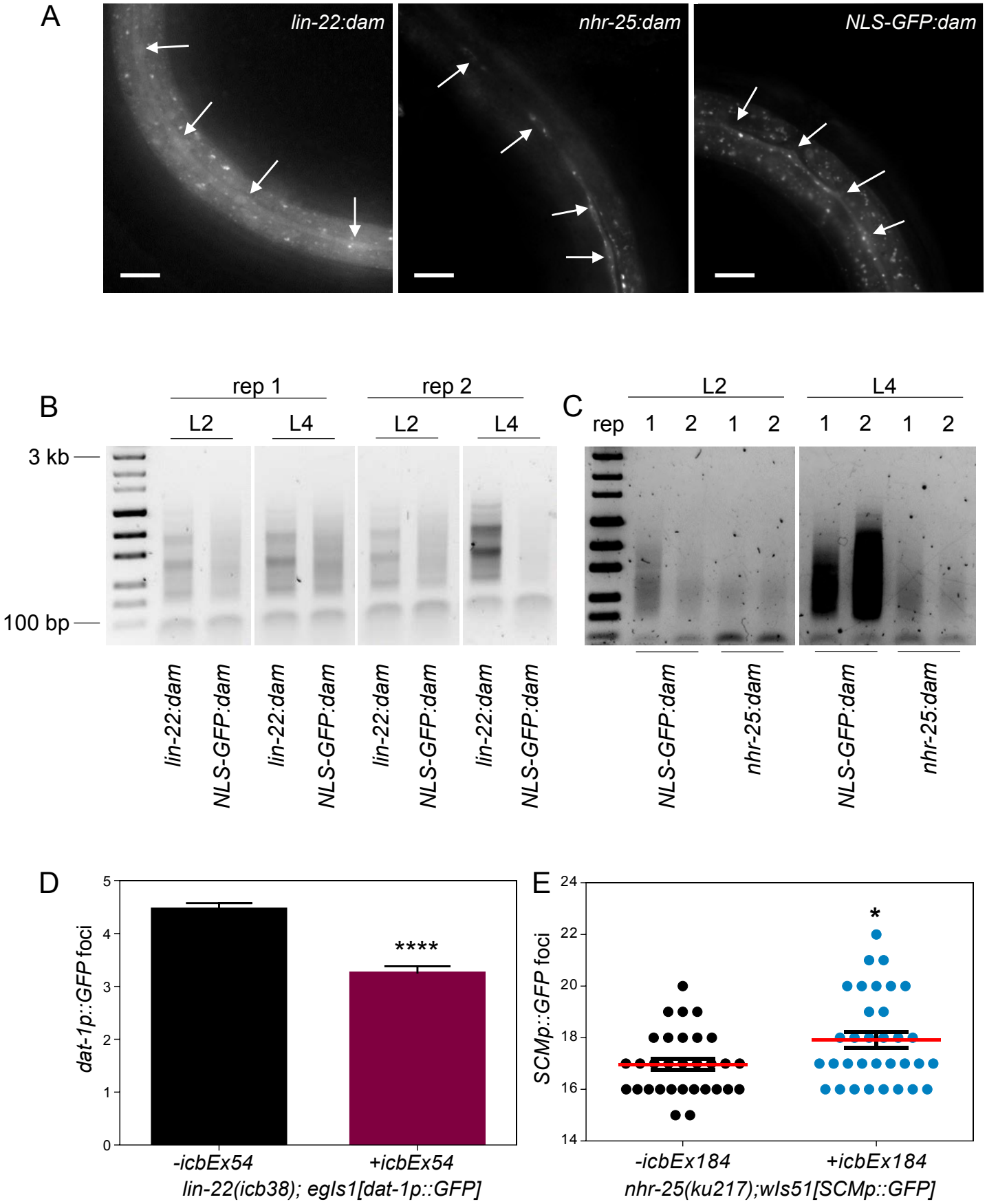

A
lin-22:dam
nhr-25:dam
NLS-GFP:dam
B
rep 1
rep 2
L2
L4
L2
L4
3 kb
100 bp
lin-22:dam
NLS-GFP:dam
lin-22:dam
NLS-GFP:dam
lin-22:dam
NLS-GFP:dam
lin-22:dam
NLS-GFP:dam
C
L2
L4
rep
1 2 1 2
1 2 1 2
NLS-GFP:dam
nhr-25:dam
NLS-GFP:dam
nhr-25:dam
D
dat-1p::GFP foci
0
1
2
3
4
5
****
-icbEx54
+icbEx54
lin-22(icb38); egIs1[dat-1p::GFP]
E
SCMp::GFP foci
14
16
18
20
22
24
*
-icbEx184
+icbEx184
nhr-25(ku217);wIs51[SCMp::GFP]

Figure S3

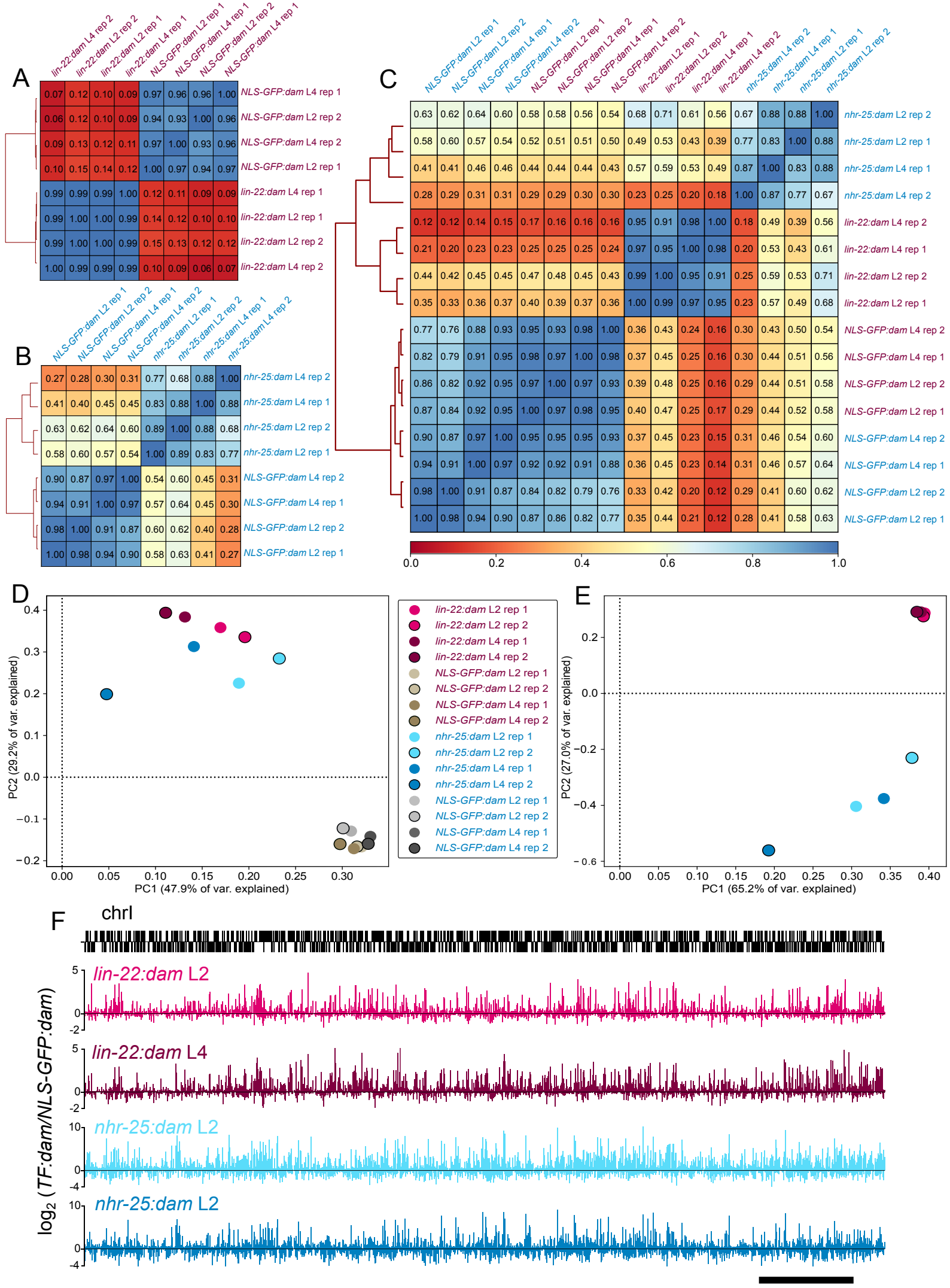
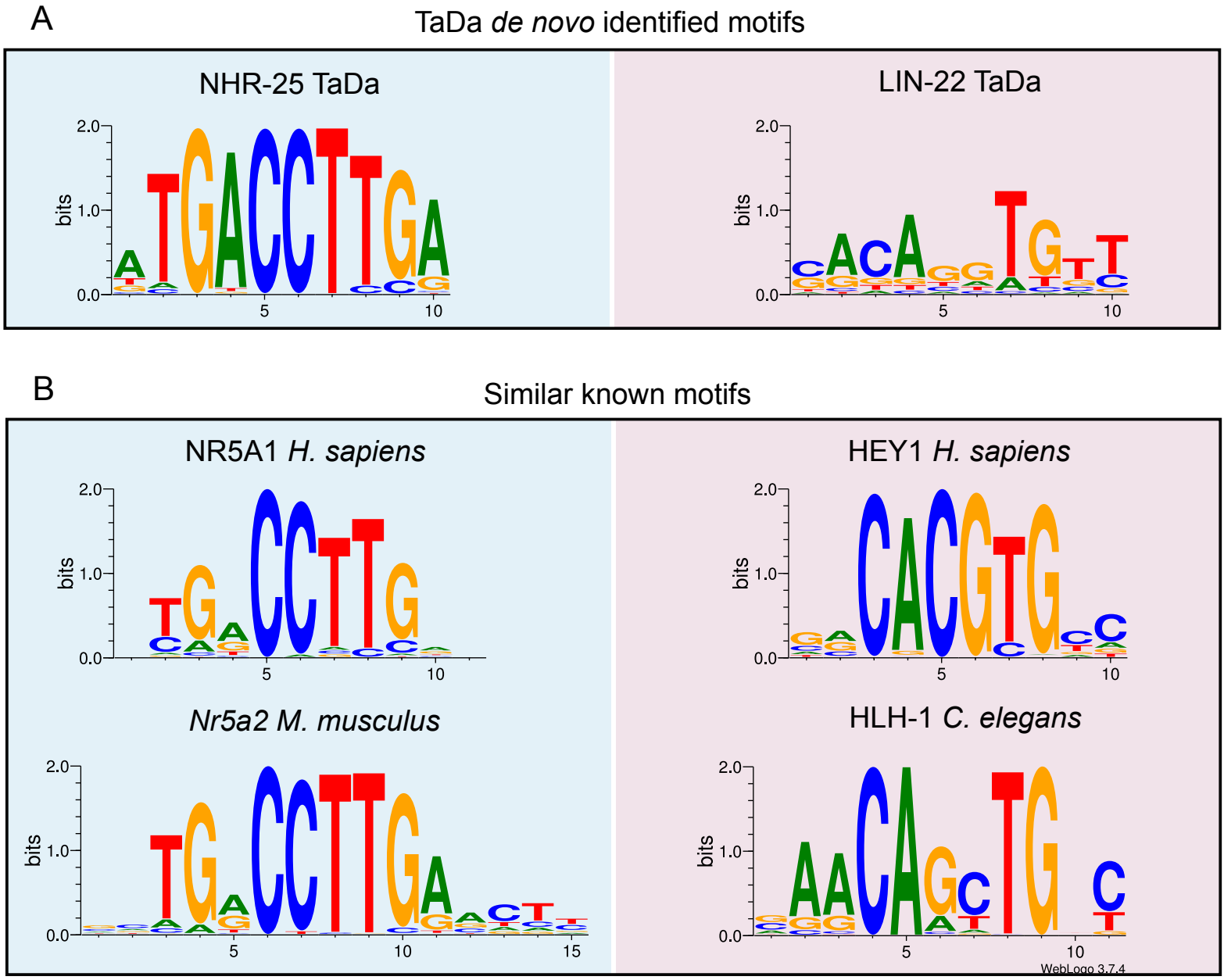

A
lin-22:dam L4 rep 2
lin-22:dam L2 rep 2
lin-22:dam L2 rep 1
lin-22:dam L4 rep 1
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 1 0.07 0.12 0.10 0.09 0.97 0.96 0.96 1.00
NLS-GFP:dam L2 rep 2 0.06 0.12 0.10 0.09 0.94 0.93 1.00 0.96
NLS-GFP:dam L4 rep 2 0.09 0.13 0.12 0.11 0.97 1.00 0.93 0.96
NLS-GFP:dam L2 rep 1 0.10 0.15 0.14 0.12 1.00 0.97 0.94 0.97
lin-22:dam L4 rep 1 0.99 0.99 0.99 1.00 0.12 0.11 0.09 0.09
lin-22:dam L2 rep 1 0.99 1.00 1.00 0.99 0.14 0.12 0.10 0.10
lin-22:dam L2 rep 2 0.99 1.00 1.00 0.99 0.15 0.13 0.12 0.12
lin-22:dam L4 rep 2 1.00 0.99 0.99 0.99 0.10 0.09 0.06 0.07
B
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 2
nhr-25:dam L2 rep 1
nhr-25:dam L2 rep 2
nhr-25:dam L4 rep 1
nhr-25:dam L4 rep 2
nhr-25:dam L4 rep 2 0.27 0.28 0.30 0.31 0.77 0.68 0.88 1.00
nhr-25:dam L4 rep 1 0.41 0.40 0.45 0.45 0.83 0.88 1.00 0.88
nhr-25:dam L2 rep 2 0.63 0.62 0.64 0.60 0.89 1.00 0.88 0.68
nhr-25:dam L2 rep 1 0.58 0.60 0.57 0.54 1.00 0.89 0.83 0.77
NLS-GFP:dam L4 rep 2 0.90 0.87 0.97 1.00 0.54 0.60 0.45 0.31
NLS-GFP:dam L4 rep 1 0.94 0.91 1.00 0.97 0.57 0.64 0.45 0.30
NLS-GFP:dam L2 rep 2 0.98 1.00 0.91 0.87 0.60 0.62 0.40 0.28
NLS-GFP:dam L2 rep 1 1.00 0.98 0.94 0.90 0.58 0.63 0.41 0.27
C
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 2
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 2
lin-22:dam L2 rep 1
lin-22:dam L2 rep 2
lin-22:dam L4 rep 1
lin-22:dam L4 rep 2
nhr-25:dam L4 rep 2
nhr-25:dam L4 rep 1
nhr-25:dam L2 rep 1
nhr-25:dam L2 rep 2
nhr-25:dam L2 rep 2 0.63 0.62 0.64 0.60 0.58 0.58 0.56 0.54 0.68 0.71 0.61 0.56 0.67 0.88 0.88 1.00
nhr-25:dam L2 rep 1 0.58 0.60 0.57 0.54 0.52 0.51 0.51 0.50 0.49 0.53 0.43 0.39 0.77 0.83 1.00 0.88
nhr-25:dam L4 rep 1 0.41 0.41 0.46 0.46 0.44 0.44 0.44 0.43 0.57 0.59 0.53 0.49 0.87 1.00 0.83 0.88
nhr-25:dam L4 rep 2 0.28 0.29 0.31 0.31 0.29 0.29 0.30 0.30 0.23 0.25 0.20 0.18 1.00 0.87 0.77 0.67
lin-22:dam L4 rep 2 0.12 0.12 0.14 0.15 0.17 0.16 0.16 0.16 0.95 0.91 0.98 1.00 0.18 0.49 0.39 0.56
lin-22:dam L4 rep 1 0.21 0.20 0.23 0.23 0.25 0.25 0.25 0.24 0.97 0.95 1.00 0.98 0.20 0.53 0.43 0.61
lin-22:dam L2 rep 2 0.44 0.42 0.45 0.45 0.47 0.48 0.45 0.43 0.99 1.00 0.95 0.91 0.25 0.59 0.53 0.71
lin-22:dam L2 rep 1 0.35 0.33 0.36 0.37 0.40 0.39 0.37 0.36 1.00 0.99 0.97 0.95 0.23 0.57 0.49 0.68
NLS-GFP:dam L4 rep 2 0.77 0.76 0.88 0.93 0.95 0.93 0.98 1.00 0.36 0.43 0.24 0.16 0.30 0.43 0.50 0.54
NLS-GFP:dam L4 rep 1 0.82 0.79 0.91 0.95 0.98 0.97 1.00 0.98 0.37 0.45 0.25 0.16 0.30 0.44 0.51 0.56
NLS-GFP:dam L2 rep 2 0.86 0.82 0.92 0.95 0.97 1.00 0.97 0.93 0.39 0.48 0.25 0.16 0.29 0.44 0.51 0.58
NLS-GFP:dam L2 rep 1 0.87 0.84 0.92 0.95 1.00 0.97 0.98 0.95 0.40 0.47 0.25 0.17 0.29 0.44 0.52 0.58
NLS-GFP:dam L4 rep 2 0.90 0.87 0.97 1.00 0.95 0.95 0.95 0.93 0.37 0.45 0.23 0.15 0.31 0.46 0.54 0.60
NLS-GFP:dam L4 rep 1 0.94 0.91 1.00 0.97 0.92 0.92 0.91 0.88 0.36 0.45 0.23 0.14 0.31 0.46 0.57 0.64
NLS-GFP:dam L2 rep 2 0.98 1.00 0.91 0.87 0.84 0.82 0.79 0.76 0.33 0.42 0.20 0.12 0.29 0.41 0.60 0.62
NLS-GFP:dam L2 rep 1 1.00 0.98 0.94 0.90 0.87 0.86 0.82 0.77 0.35 0.44 0.21 0.12 0.28 0.41 0.58 0.63
0.0 0.2 0.4 0.6 0.8 1.0
D
PC2 (29.2% of var. explained)
PC1 (47.9% of var. explained)
lin-22:dam L2 rep 1
lin-22:dam L2 rep 2
lin-22:dam L4 rep 1
lin-22:dam L4 rep 2
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 2
nhr-25:dam L2 rep 1
nhr-25:dam L2 rep 2
nhr-25:dam L4 rep 1
nhr-25:dam L4 rep 2
NLS-GFP:dam L2 rep 1
NLS-GFP:dam L2 rep 2
NLS-GFP:dam L4 rep 1
NLS-GFP:dam L4 rep 2
E
PC2 (27.0% of var. explained)
PC1 (65.2% of var. explained)
F
chrI
lin-22:dam L2
lin-22:dam L4
nhr-25:dam L2
nhr-25:dam L2
log2 (TF:dam/NLS-GFP:dam)

Figure S4


A
TaDa *de novo* identified motifs
NHR-25 TaDa
LIN-22 TaDa
bits
B
Similar known motifs
NR5A1 *H. sapiens*
HEY1 *H. sapiens*
Nr5a2 M. musculus
HLH-1 *C. elegans*
WebLogo 3.7.4

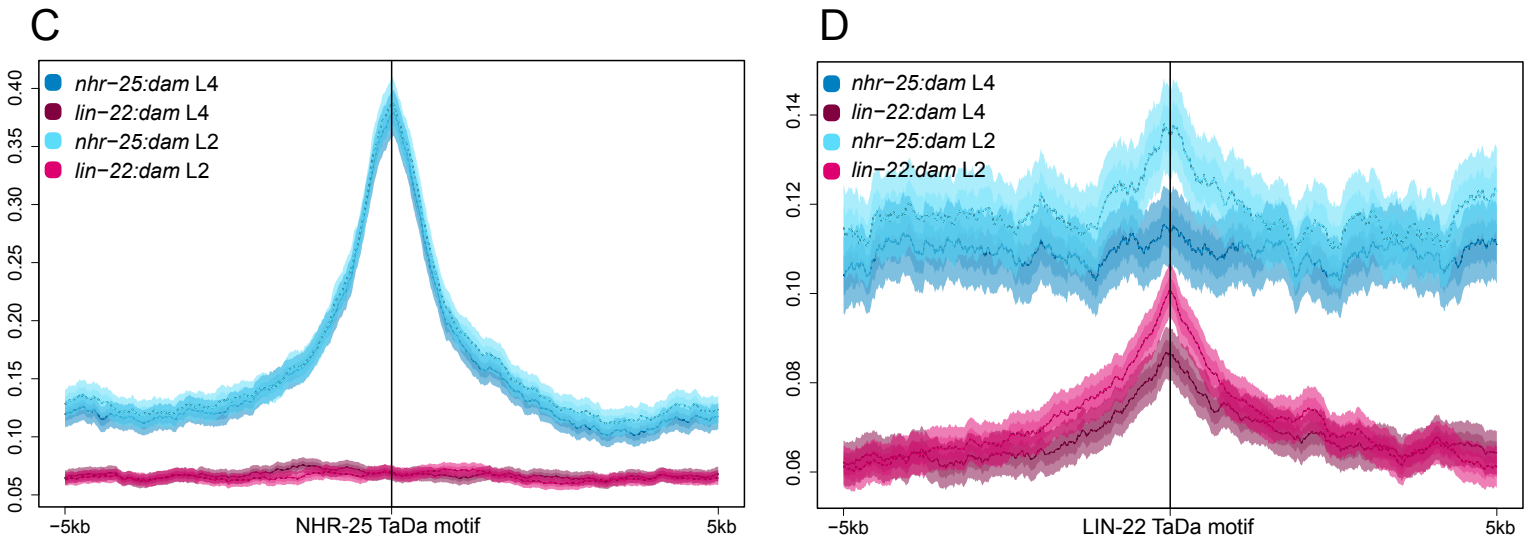
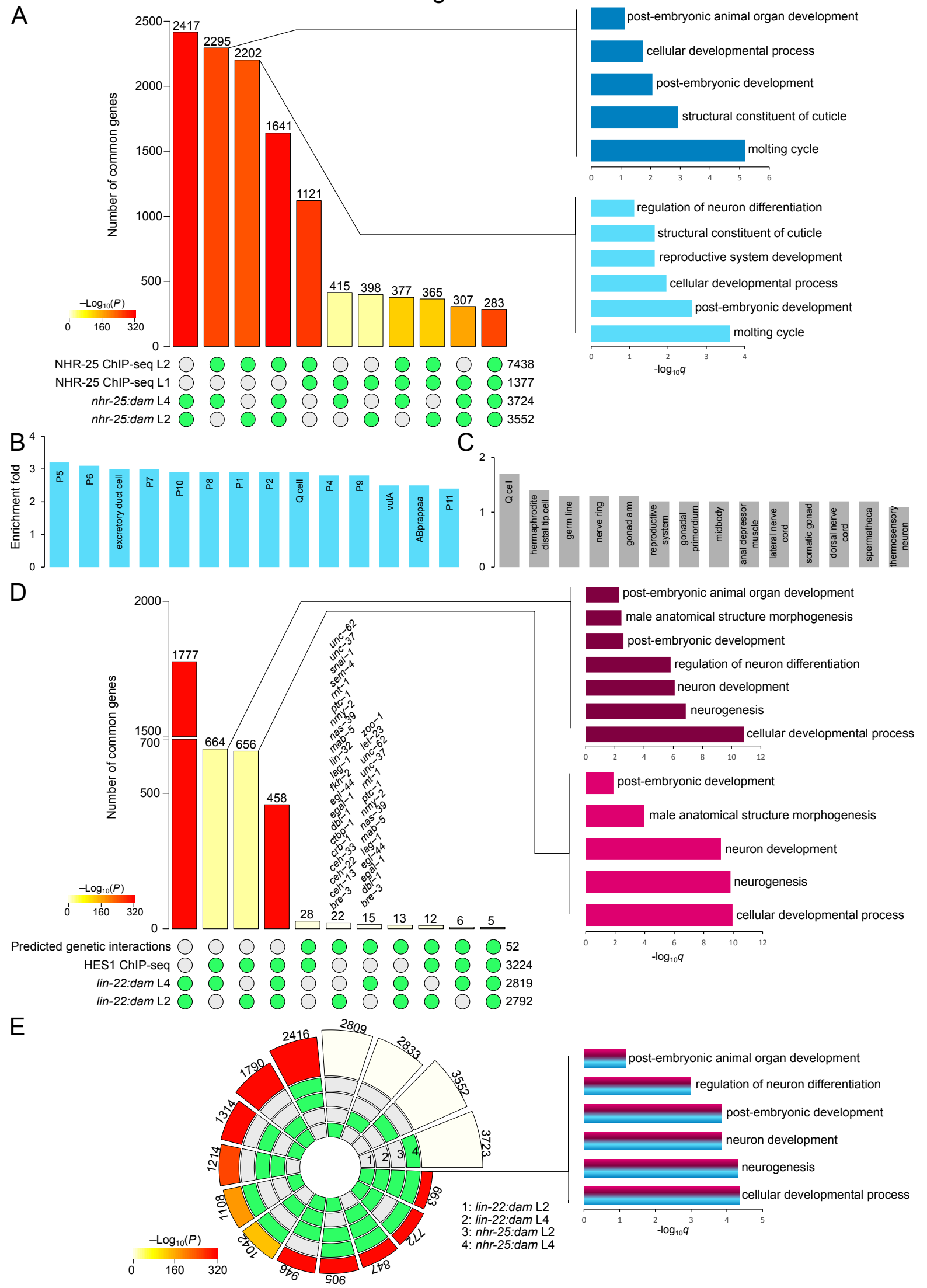

C
D
nhr−25:dam L4
lin−22:dam L4
nhr−25:dam L2
lin−22:dam L2
−5kb
NHR-25 TaDa motif
5kb
LIN-22 TaDa motif

Figure S5


A
Number of common genes
2417
2295
2202
1641
1121
415
398
377
365
307
283
–Log10(P)
0
160
320
NHR-25 ChIP-seq L2
NHR-25 ChIP-seq L1
nhr-25:dam L4
nhr-25:dam L2
7438
1377
3724
3552
post-embryonic animal organ development
cellular developmental process
post-embryonic development
structural constituent of cuticle
molting cycle
regulation of neuron differentiation
structural constituent of cuticle
reproductive system development
cellular developmental process
post-embryonic development
molting cycle
-log10q
B
Enrichment fold
P5
P6
excretory duct cell
P7
P10
P8
P1
P2
Q cell
P4
P9
vulA
ABprappaa
P11
C
Q cell
hermaphrodite distal tip cell
germ line
nerve ring
gonad arm
reproductive system
gonadal primordium
midbody
anal depressor muscle
lateral nerve cord
somatic gonad
dorsal nerve cord
spermatheca
thermosensory neuron
D
Number of common genes
1777
664
656
458
28
22
15
13
12
6
5
unc-62
unc-37
snai-1
sem-4
rnt-1
ptc-1
nmy-2
nas-39
mab-5
lin-32
lag-1
fkh-2
egl-44
egal-1
dbl-1
ctbp-1
crb-1
ceh-33
ceh-22
ceh-13
bre-3
zoo-1
lef-23
Predicted genetic interactions
HES1 ChIP-seq
lin-22:dam L4
lin-22:dam L2
52
3224
2819
2792
post-embryonic animal organ development
male anatomical structure morphogenesis
post-embryonic development
regulation of neuron differentiation
neuron development
neurogenesis
cellular developmental process
post-embryonic development
male anatomical structure morphogenesis
neuron development
neurogenesis
cellular developmental process
-log10q
E
2416
2809
2833
3552
3723
663
772
847
905
946
1042
1108
1214
1314
1790
1: lin-22:dam L2
2: lin-22:dam L4
3: nhr-25:dam L2
4: nhr-25:dam L4
post-embryonic animal organ development
regulation of neuron differentiation
post-embryonic development
neuron development
neurogenesis
cellular developmental process
-log10q