Targeted DamID in *C. elegans* reveals a role for LIN-22 and NHR-25 in epidermal cell differentiation

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: TaDa identifies LIN-22 and NHR-25 targets

Abstract

Transcription factors are key players in gene networks controlling cell fate specification during development. In multicellular organisms, they can display complex patterns of expression and binding to their targets, which necessitates tissue-specific characterisation of transcription factor-target interactions. Here, we focus on *C. elegans* seam cell development, which is used as a model of robust epidermal stem cell patterning. Despite our knowledge of multiple transcription factors playing a role in epidermal development, the composition of the gene network underlying cell fate patterning remains largely unknown. We introduce Targeted DamID (TaDa) that allows tissue-specific transcription factor target identification in intact *C. elegans* animals without cell isolation. We employ this method to recover putative targets in the epidermis for two transcription factors, the HES1 homologue LIN-22 and the NR5A1/2 nuclear hormone receptor NHR-25. Using single-molecule FISH (smFISH), we validate TaDa predictions and reveal a role for these transcription factors in promoting cell differentiation, as well as an unusual link between a HES factor and the Wnt signalling pathway.

Our results expand our understanding of the epidermal gene network and highlight the power of TaDa to dissect the architecture of tissue-specific gene regulatory networks.

Keywords: DamID, transcription factors, LIN-22, HES1, NHR-25, NR5A1, NR5A2, seam cells, epidermis, stem cells

Introduction

Development of animals and plants involves the reproducible formation of complex 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 (TFs) are prominent players as they participate in gene regulatory networks that govern robust cell fate specification, often acting in a combinatorial way (Reilly et al., 2020). Therefore, elucidating the architecture of developmental gene networks by identifying the participating TFs, as well as their biologically relevant targets, is fundamental for increasing our mechanistic understanding of how specialised cells and tissues are formed (Davidson, 2010; Oliveri and Davidson, 2007).

An important type of cells in the *C. elegans* epidermis are 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 at each larval stage generate differentiated hypodermal or neuronal cells, while maintaining the total seam cell number through self-renewal (Sulston and Horvitz, 1977). The highly reproducible nature of *C. elegans* development allows the use of this model to reveal the mechanisms underlying the balance between cell proliferation and differentiation (Joshi et al., 2010). 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 the nucleus and, along with the TF POP-1/T-cell factor/lymphoid enhancer factor (TCF/LEF), activates Wnt target genes (Sawa and

77 Korswagen, 2013). The Wnt/β-catenin asymmetry (WβA) is a modified version of the
78 Wnt pathway adapted for the purposes of asymmetric cell division, where asymmetric
79 distribution of Wnt pathway components polarises the mother cell and leads to
80 asymmetric inheritance of the potential for Wnt pathway activation in the two daughter
81 cells following division (Lam and Phillips, 2017).

82 Besides Wnt signalling, a number of conserved TFs have been identified to
83 influence cell fate decisions in the epidermis. For example, RNT-1 is the *C. elegans*
84 Runx homolog and has been shown to act together with its binding partner BRO-1/
85 CBFβ to promote symmetric seam cell divisions (Kagoshima et al., 2007; van der Horst
86 et al., 2019). Mutations in Runx genes and CBFβ are known to cause various
87 leukaemias in humans (Cameron and Neil, 2004), and mutations or overexpression in
88 *C. elegans* lead to seam cell loss or tissue hyperplasia respectively (Nimmo et al.,
89 2005). Another example is the HES1 homologue and basic helix-loop-helix (bHLH) TF
90 LIN-22, which acts to suppress neurogenesis in seam cells (Wrischnik and Kenyon,
91 1997) and maintain correct patterning, possibly by antagonising the Wnt signalling
92 pathway (Katsanos et al., 2017). The GATA family of TFs is linked to various types of
93 cancer in humans (Zheng and Blobel, 2010), and related factors in *C. elegans*, such
94 as ELT-1 and EGL-18, are thought to specify seam cell fate (Gorrepati et al., 2013;
95 Koh and Rothman, 2001; Smith et al., 2005). Finally, TFs of the nuclear hormone
96 receptor (NHR) family have been reported to be play a role in epidermal development
97 (Antebi, 2006; Miyabayashi et al., 1999). A key example is the NR5A1/2 homologue
98 NHR-25, which regulates the establishment of cell-to-cell contacts in the seam, while
99 it acts more broadly in vulval development, molting and neurogenesis in the T lineage
100 (Chen et al., 2004; Gissendanner and Sluder, 2000; Hajduskova et al., 2009; Hayes et
101 al., 2006; Shao et al., 2013; Šilhánková et al., 2005).

102 Despite the gain in knowledge on individual TFs playing a role in epidermal
103 development, we still have a limited understanding of the underlying gene network and
104 how this drives the decision between cell proliferation and differentiation. This is
105 largely due to the fact that most identified components are still disconnected to each
106 other, or when interactions are known, these remain at the genetic level. Identification
107 of direct TF targets has been predominantly pursued by ChIP experiments in *C.*
108 *elegans* (Araya et al., 2014; Celniker et al., 2009; Kudron et al., 2018). DNA adenine
109 methyltransferase identification (DamID) (van Steensel et al., 2001; van Steensel and
110 Henikoff, 2000), offers an alternative way to reveal targets by fusing a TF of interest
111 to the Dam methyltranferase from *E. coli*. This enzyme mediates site-specific
112 methylation in the adenine of GATC sequences therefore it can leave methylation
113 marks to the DNA wherever a TF binds (Schuster et al., 2010; van de Walle et al.,
114 2020). A key requirement in DamID methods is keeping the levels of Dam expression
115 very low to avoid toxicity and saturated methylation of DNA (Pindyurin et al., 2016; van
116 Steensel and Henikoff, 2000). An additional challenge in multicellular systems like *C.*
117 *elegans* is to characterise DNA-protein interactions with tissue-specific resolution. This
118 has been previously achieved using recombinase-based systems to allow expression
119 of Dam fusions at low levels defined by the basal transcription from non-induced heat-
120 shock promoters (Gómez-Saldivar et al., 2020; Harr et al., 2020; Pindyurin et al.,
121 2016). Alternatively, Targeted DamID (TaDa) allows tissue-specific target
122 identification (Southall et al., 2013) and has been successfully employed in *Drosophila*
123 to dissect mechanisms of neuronal fate determination (Sen et al., 2019; Vissers et al.,
124 2018), as well as in mammalian stem cell lines to characterise the binding profile of
125 pluripotency factors (Cheetham et al., 2018). Low levels of Dam-fusion expression are
126 achieved in TaDa with a specific transgene configuration where an unrelated primary

127 ORF (usually *mCherry)* is introduced followed by 2 STOP codons and a frameshift
128 preceding the Dam-fusion as a second ORF (Figure 1A). Due to the universal property
129 of eukaryotic ribosomes to reinitiate translation after a STOP codon at a reduced
130 frequency (Kozak, 2001, 1987), this transgene design results in low levels of
131 expression of the Dam-TF fusion in the tissue of interest as defined by the specific
132 promoter used (Figure 1A).

133 In this study, we demonstrate that TaDa is a powerful method to identify TF
134 targets in *C. elegans*. We employ TaDa to find targets of LIN-22 and NHR-25 in the
135 epidermis. Using single molecule FISH (smFISH), we validate changes in target
136 expression upon perturbation of LIN-22 or NHR-25 activity. Our results suggest a role
137 for these two TFs in promoting cell differentiation by repressing stem cell promoting
138 factors. We also reveal a direct link between LIN-22 and the Wnt signalling pathway
139 through the Frizzled receptor *lin-17*. Our findings expand our knowledge of the gene
140 network underlying epidermal cell fate decisions and provide a methodological
141 framework for resolving regulatory networks in specific nematode tissues.

142

143 **Results**

144 **TaDa circumvents Dam-associated toxicity and allows the recovery of TF-**
145 **specific methylation profiles.**

146 To enrich our understanding of the gene network underlying epidermal cell fate
147 patterning, we chose to study two TFs, the bHLH factor LIN-22 and the nuclear
148 hormone receptor NHR-25, which are both key players in epidermal development
149 (Chen et al., 2004; Gissendanner and Sluder, 2000; Wrischnik and Kenyon, 1997). We
150 decided to focus on LIN-22 because previous work implicated this factor in Wnt-
151 dependent seam cell patterning, although its direct targets remained unknown

152 (Katsanos et al., 2017; Wrischnik and Kenyon, 1997). On the other hand, NHR-25
153 targets had been previously studied using ChIP-seq (Araya et al., 2014; Shao et al.,
154 2013), although direct targets in the epidermis were not known. To profile TF targets
155 by TaDa, we constructed transgenes consisting of TF-Dam fusions under the *wrt-2*
156 promoter, which drives expression primarily in the seam cells. We included a *C.*
157 *elegans* optimised *mCherry* as the primary ORF, followed by 2 STOP codons and an
158 extra nucleotide for frame-shift before the TF-Dam (*lin-22:dam* and *nhr-25:dam)* or
159 control (*NLS-GFP:dam)* fusions (Figure 1B). We also designed a versatile backbone
160 to allow the assembly of N- or C- terminal fusions of any other transcription factor with
161 Dam and using any promoter of interest (Figure S1A).

162 Induced ubiquitous Dam expression by a heat-shock promoter has been shown to
163 produce saturated methylation and toxicity (Schuster et al., 2010; van Steensel and
164 Henikoff, 2000), but the effect of tissue-specific expression of Dam-fusions had not
165 been previously investigated in *C. elegans*. To test this, we constructed *lin-22:dam*
166 fusions with and without *mCherry* as the primary ORF (Figure 1B), and inserted these
167 into the genome as single-copy transgenes. We found that transgenic lines lacking
168 *mCherry* showed developmental defects, as well as significantly reduced brood size
169 compared to animals containing the *mCherry* primary ORF, which appeared
170 comparable to wild-type controls (Figure 1C, D). Aberrant phenotypes in the absence
171 of the primary ORF were likely due to high levels of Dam expression because they
172 were observed in strains carrying both *lin-22:dam* and *NLS-GFP:dam* fusions, and
173 correlated with increased amount of DNA methylation (Figure S1B). We therefore
174 conclude that the TaDa transgene configuration reduces toxicity in *C. elegans*, which
175 is consistent with what has been reported in other systems (Southall et al., 2013; van
176 Steensel and Henikoff, 2000).

177 We then tested the single-copy transgenics harbouring TaDa transgenes for
178 expression and functionality of each fusion. The expression of *mCherry* was used as
179 proxy to confirm the spatial localisation of the TF-Dam fusions. Microscopy at the L4
180 stage revealed low *mCherry* expression in the seam cells, as expected for single copy
181 transgenes (Figure S2A). Despite the TaDa design that results in low level of
182 expression of the TF-Dam fusions, their functionality can still be assessed using a
183 multi-copy transgenesis assay, where an increase in copy number may allow to rescue
184 loss-of-function mutations or induce ectopic phenotypes. For example, *lin-22(icb38)*
185 mutants show an increase in the number of postdeirid (PDE) neurons, and this
186 phenotype was suppressed in transgenics carrying the *lin-22:dam* construct as an
187 extrachromosomal array (Figure S2B). Furthermore, the *nhr-25:dam* fusion was
188 sufficient to disrupt seam cell patterning as a multi-copy array (Figure S2C).
189 We sequenced amplicons derived from methylated genomic sequences for both
190 TFs at two developmental stages (L2 and L4) and processed the data using the
191 damidseq pipeline (Marshall and Brand, 2015). We generated normalised aligned read
192 count maps to evaluate replicate reproducibility and assess the level of correlation
193 between samples. High level of reproducibility was observed for all fusions between
194 biological replicates (Figure S3A, B) with higher Pearson correlation coefficients for
195 comparisons within a given *TF:dam* fusion (Figure S3C). Importantly, control *NLS-*
196 *GFP:dam* samples were found to cluster separately from TF-Dam samples using
197 principal component analysis (Figure S3D), and separate clustering was observed for
198 the *lin-22:dam* and *nhr-25:dam* samples (Figure S3E). These results highlight that TF-
199 Dam fusions produce specific and distinct methylation patterns compared to control
200 fusions.
201

202 **NHR-25 and LIN-22 binding profiles overlap with regulatory regions of the**
203 **genome.**

204 To investigate the profile of signal enrichment across the genome, we compared
205 *TF:dam* and *NLS-GFP:dam* sample pairs and calculated the normalised
206 $\log_2$(*TF:dam/NLS-GFP:dam*) ratio scores per GATC fragment based on normalised
207 aligned read count maps (an example covering 1Mb of sequence is shown in Figure
208 S3F). To address whether the observed signal enrichment is likely to reflect TF binding,
209 we focused on putative targets of LIN-22 and NHR-25 proposed by genetic analysis or
210 ChIP-seq experiments respectively. For example, we observed significant enrichment
211 in the *lin-22:dam* profiles near the Frizzled receptor *lin-17*, the Hox gene *mab-5* and
212 the *C. elegans achaete-scute* homologue *lin-32*, which are reported to be genetic
213 interactors of *lin-22* (Katsanos et al., 2017; Wrischnik and Kenyon, 1997) (Figure 2A).
214 With regard to the *nhr-25:dam* profiles, significant enrichment was found in regions
215 near the putative target genes *idh-1* and *rpl-3,* as well as the *nhr-25* locus itself,
216 supporting previous evidence for self-regulation (Shao et al., 2013) (Figure 2B).
217 Qualitative inspection of the enrichment profiles suggested a preference for enrichment
218 in intergenic regions (Figure 2A, B). Indeed, aggregate genome-wide signal profiles
219 showed increased average enrichment scores in upstream sequences directly
220 proximal to the transcriptional start site (TSS) of protein coding genes (Figure 2C).

221 Statistical processing of the signal profiles was used to identify significant peaks.
222 We found 1965 and 1972 peaks for *lin-22:dam* at the L2 and L4 stage, and 2044 and
223 2169 peaks for *nhr-25:dam,* a complete list of which is shown in Table S1. Hierarchical
224 clustering of the localisation and score of those peaks that lie between 5 kb upstream
225 and 1kb downstream of genes, was consistent with a preference for signal enrichment
226 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 3A). 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 3A). In stark contrast, a very small percentage of peaks, (4.2% - 9%) were found exclusively within exons, despite the fact that exons have a double median size compared to introns (Spieth et al., 2014). A complete list of assigned peaks can be found in Table S2.

To consolidate the link between the identified peaks and putative 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 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). We finally investigated the overlap between our data and tissue-specific open chromatin elements at the L4 stage described in a recent study (Serizay et al., 2020). We found significant overlap with hypodermally-enriched accessible chromatin (10.8% - 41.5%, $p < 0.0001$) in

252 comparison to other tissues like neurons (3.4% - 2.4%) or the germline (0.8% - 1.7%),
253 where the overlap was not significant. Taken together, the overall localisation profiles
254 of TaDa peaks strongly support that TaDa signal is likely to reflect genuine LIN-
255 22/NHR-25 binding sites.
256
257 **Comparisons of peak localisation profiles across methods and transcription**
258 **factors.**
259 To compare our TaDa results against other methods, we used two published ChIP-
260 seq datasets for NHR-25 binding at L1 and L2 stage from Shao *et al*., 2013 (L1) and
261 Araya *et al*., 2014 (L2). Initial qualitative assessment of the signal tracks showed some
262 promising agreement (Figure 4A). At a genome-wide level, aggregate signal of the *nhr-*
263 *25:dam* L2 and L4 over all ChIP-seq L1 and L2 peaks, or vice verca, exhibited strong
264 overlapping localisation (Figure 4B), supporting the similarity of peak profiles across
265 methods. Both the L2 and L4 *nhr-25:dam* peaks showed significant localisation
266 overlaps by Monte Carlo simulations across the genome with the 683 peaks of the
267 NHR-25 ChIP-seq L1 and the 5980 peaks of the ChIP-seq L2 datasets (Figure 4C).
268 For example, we found a substantial overlap with the L2 ChIP-seq dataset, with
269 approximately 37% (726 peaks) and 38.5% (835 peaks) of the total L2 and L4 *nhr-*
270 *25:dam* peaks overlapping respectively. It is of note that in 75% of the 726 *nhr-25:dam*
271 L2 peaks overlapping the ChIP-seq L2 peaks, the overlap included the highest
272 enriched GATC fragment of the TaDa peak.
273 We then examined the overlap between the TaDa peak profiles identified for the
274 two TFs. Interestingly, *lin-22:dam* signal showed an increased average preference to
275 map onto NHR-25 TaDa peaks in comparison to adjacent regions and the same was
276 true for *nhr-25:dam* signal on LIN-22 peaks (Figure 4D). Since LIN-22 and NHR-25

277 both participate in epidermal development, it is conceivable that this similarity may
278 reflect genuine proximity or overlap of TF binding sites. Alternatively, the shared peaks
279 could also reflect genomic sites where promiscuous binding of multiple TFs may occur.
280 Such regions, usually referred to as High-Occupancy Target (HOT) regions have been
281 previously determined in ChIP-seq experiments (Araya et al., 2014). TaDa peaks
282 showed significant (p< 0.0001 by Monte Carlo simulations) yet smaller overlap with
283 HOT regions compared to ChIP-seq (8% for *nhr-25:dam* as opposed to 34% forn NHR-
284 25 ChIP-seq L2 peaks). Furthermore, out of the 2167 HOT regions found in L2 (Araya
285 et al., 2014), 13% were overlapping with TaDa NHR-25 L2 peaks, as opposed to 83%
286 with NHR-25 ChIP-seq L2 peaks.

287 To further investigate the overlap between the TaDa profiles for the two TFs, all
288 the pairwise comparisons were made and found to be non-random by Monte Carlo
289 simulations. Around 60% of peaks of each profile overlapped across the two stages for
290 the same TF, whereas <33% overlapped across TFs (Figure 4E left). To avoid p-value
291 inflation when interrogating the entire genome while TF binding sites are expected to
292 localise on promoters, overlaps and their statistical significance were re-calculated for
293 promoter regions only (defined here as 5 kb upstream to 500 bp downstream of TSS).
294 Again, we found promoter overlaps to be highly significant (Figure 4E right) and
295 represented the majority of the genome-wide overlaps (67% - 72%). In addition, less
296 than a third (24.3% -32.5%) of the overlaps in promoters across TF peaks occurred in
297 HOT regions, suggesting that LIN-22 and NHR-25 may indeed share some targets in
298 epidermal development.

299 To dissect the tissue-specificity of the binding profile of LIN-22 and NHR-25, we
300 compared it to peak profiles for various TFs from ChIP-seq experiments conducted at
301 L2 from the modERN database (Kudron et al., 2018). We included key TFs known to

302 act specifically in the epidermis, such as ELT-1 and ELT-3, along with others that play
303 broader roles in development including neurogenesis to investigate the global peak
304 localisation pattern these TFs exhibit. Profile-wide comparisons of peak localisation
305 between factors based on overlap and proximity statistics can provide a measure of
306 co-association and statistically rank factors to highlight the broader similarity of their
307 binding (Chikina and Troyanskaya, 2012). Previous calculations of co-association
308 matrices have shown that TFs clustering together often regulate the same targets or
309 act in the same tissue (Araya et al., 2014; Kudron et al., 2018). We found that our
310 NHR-25 and LIN-22 TaDa datasets cluster together with the epidermal ELT-1 and ELT-
311 3 TFs, while ChIP-seq and TaDa profiles for NHR-25 did not cluster together (Figure
312 4F). Since NHR-25 also acts in non-epidermal cells, which is likely to be captured in
313 ChIP-seq but not in the TaDa profiles, this may explain the distinct positioning of the
314 NHR-25 binding profile depending on the method via which this profile was acquired
315 (Figure 4F). These comparisons highlight the value of TaDa in revealing TF binding
316 within a tissue of interest.
317
318 **LIN-22 and NHR-25 binding motif identification based on TaDa peaks.**
319 Next, we used the identified TaDa peaks to discover putative motifs that are
320 associated to TF binding. Sequences restricted to the overlap between L2 and L4
321 peaks for each factor were used to increase the probability that these contain binding
322 sites. The *de novo* identified motif for NHR-25 was consistent with the one previously
323 reported by ChIP-seq and confirmed by functional studies (Araya et al., 2014;
324 Barkoulas et al., 2016) (Figure S4A). When this motif was run against a database of
325 known motifs (Fornes et al., 2020), it showed significant similarity to those of the *nhr-*
326 *25* human orthologue NR5A1 (p=3.04×10^{-6}) and mouse orthologue *Nr5a2* (p=2.35×10^{-}

327 ^{5}) (Figure S4B). The same analysis was carried out for LIN-22 and identified a motif
328 that matches an E-box sequence (5'-CANNTG-3') (Figure S4A), which matches
329 reported motifs for the human orthologue HES1 (Lichtenberg et al., 2018). Comparison
330 to known motifs showed significant similarity to that of the human HEY1 (Hes-related
331 Family bHLH with YRPW Motif 1) factor (p=3.29×10^{-3}), and HLH-1/ *MyoD* (p=2.45×10^{-}
332 ^{4}), which *Hes1* is known to antagonise for sites and binding partners (Sasai et al., 1992)
333 (Figure S4B).

334 To assess if the identified motifs broadly represent preference for TF binding as
335 determined in TaDa signal profiles, aggregate signal was mapped onto motif sites. For
336 the NHR-25 motif, the *nhr-25:dam* L2 and L4 signal showed increased average
337 preference for regions that include the motif as opposed to neighbouring sequences,
338 while the *lin-22:dam* signal did not show any enrichment (Figure S4C). The reverse
339 relationship was observed when the signal profiles were mapped onto the LIN-22 motif
340 (Figure S4D). Some limited enrichment was observed for *nhr-25:dam* L2 data over the
341 LIN-22 motif, which may be due to the fact that this motif is more noisy (Figure S4A),
342 thus more frequent in the genome (1.5 times more frequent than the NHR-25 motif).
343

344 **Characterisation of LIN-22 and NHR-25 targets.**

345 To identify LIN-22 and NHR-25 putative targets, TaDa peaks were assigned to
346 neighbouring genes. This assignment resulted in a set of 2809 genes for LIN-22 at L2
347 and 2833 genes at L4, and 3552 genes for NHR-25 at L2 and 3724 genes at L4 (Figure
348 5A, B and Table S2). The majority of the identified genes were shared between the
349 two stages for each factor, with >63% of genes in any dataset being shared between
350 L2 and L4 (Figure 5A, B). LIN-22 controls aspects of epidermal development by
351 repressing neurogenesis and influencing ray formation in males (Wrischnik and

352 Kenyon, 1997). These functions were reflected in related GO terms for the identified
353 target gene sets at L2 and L4 (Figure 5C-E). In addition, the L2 dataset was found to
354 be enriched for genes related to the Wnt signalling pathway (KEGG adjusted *p*-values
355 of 3.28×10^{-5}). The L2 and L4 datasets of putative NHR-25 targets were also enriched
356 for multiple GO terms related to known NHR-25 biological functions (Chen et al., 2004;
357 Chisholm and Hsiao, 2012; Gissendanner and Sluder, 2000; Hayes et al., 2006). For
358 example, terms for structural constituents of the cuticle and molting cycle were found
359 to be amongst the most significantly enriched (Figure 5F-H).

360 We then sought to investigate how the genes identified via TaDa compare to those
361 identified with other methods. To this end, we first compared the NHR-25 TaDa-
362 identified targets with NHR-25 ChIP-seq datasets and found very significant overlap
363 for all pairwise or higher order dataset intersections (Figure S5A). For example, the
364 overlap with the ChIP-seq L2 dataset containing 7438 genes contained 62% of L2 and
365 61.6% of the L4 TaDa genes. We reasoned that some of the non-overlapping genes
366 from ChIP-seq may reflect targets of NHR-25 outside the epidermis that are not
367 captured by TaDa. Consistent with this hypothesis, tissue enrichment analysis for
368 genes exclusively found by ChIP-seq showed higher enrichment for reproductive
369 tissues compared to the epidermis (Figure S5B, C).

370 As there are currently no ChIP-seq datasets available for LIN-22, we intersected
371 the TaDa-identified genes with *C. elegans* orthologs from ChIP-seq dataset for HES1,
372 the human homologue of *lin-22*. We found statistically significant overlap containing
373 genes that showed enrichment for the same GO terms previously identified in the TaDa
374 gene sets alone, indicating conservation in regulatory interactions of LIN-22/HES1
375 across species (Figure S5D). A significant overlap was also found when the LIN-22
376 TaDa target genes were intersected with a list of 52 *in silico* predicted downstream

377 genetic interactors of *lin-22* (Zhong and Sternberg, 2006). These included *lin-32* and
378 *mab-5,* but also other genes known to participate in seam cell development like *unc-*
379 *62* (Hughes et al., 2013), *nmy-2* (Ding and Woollard, 2017) or *rnt-1*, for which there
380 was no prior evidence for regulation by LIN-22 (Figure S5D and Table S2).

381 We finally assessed the overlap between the identified targets for the two TFs.
382 Interestingly, significant overlap was found in all pairwise comparisons across factors
383 with 37%- 46.4% of putative LIN-22 target genes overlapping NHR-25 datasets and
384 29.3%-35.3% of putative NHR-25 target genes overlapping LIN-22 datasets (Figure
385 S5E). The intersection between all TaDa datasets contained 663 genes, which showed
386 enrichment for GO terms related to neurogenesis and development (Figure S5E).
387 These results indicate that the functions of the two TFs during epidermal development
388 are likely to be executed by a combination of distinct and shared target genes.

389

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

391 To validate putative interactions predicted by TaDa, we focused on selected
392 candidates that are known to participate in the epidermal gene network. With regard to
393 LIN-22, we focused on *rnt-1*, *cki-1* and *lin-17,* which showed significant signal
394 enrichment in their promoter regions (Figure 2A and 6A). First, *rnt-1* is known to
395 promote symmetric proliferative seam cell divisions (Nimmo et al., 2005). We found
396 that *rnt-1* expression by smFISH is significantly increased in V1-V4 lineages in *lin-*
397 *22(icb49)* null mutants, indicating that LIN-22 is likely to act as a repressor of *rnt-1*
398 (Figure 6B, C). Second, *cki-1* encodes a cyclin-dependent kinase inhibitor protein
399 (Buck et al., 2009; Hong et al., 1998). Comparison of *cki-1* expression between wild-
400 type and *lin-22(icb49)* mutants revealed a significant reduction in transcript levels in

401 V1-V4 lineages (Figure 6D, E). This finding suggests that LIN-22 may activate *cki-1*
402 expression to modulate the differentiation programme.
403 Third, TaDa revealed two major sites of LIN-22 signal enrichment on the *lin-17*
404 promoter (Figure 2A), which largely overlapped with two conserved regions (termed
405 CRE1 and CRE2) between *C. elegans* and related *Caenorhabditis* species as revealed
406 by Vista analysis (Frazer et al., 2004) (Figure 6F). To test the significance of these *cis*-
407 regulatory elements, we constructed transcriptional reporters containing the CRE1 or
408 CRE2 sequence fused to a minimal core promoter driving the expression of histone
409 bound GFP. Multi-copy arrays were created for each element and introduced into the
410 *lin-22(icb49)* mutant background. Both reporters were sufficient to drive expression in
411 the posterior epidermis in wild-type animals, while expression expanded to more
412 anterior epidermis in the *lin-22(icb49)* mutant background (Figure 6G, H, I). Taken
413 together, these data indicate that LIN-22 is likely to bind to these regulatory elements
414 on the *lin-17* promoter to repress *lin-17* expression. It is of note that LIN-22 was found
415 to possibly target other Wnt receptors (*mom-5),* as well as Wnt ligands (*cwn-2*), Wnt
416 secretion factors *(mom-1, mig-14*) and components of the signal transduction
417 machinery (*lit-1*, *bar-1*) (Figure S6), indicating that LIN-22 may regulate the Wnt
418 pathway at various levels of the signalling cascade.
419 With regard to putative targets of NHR-25, we focused on *egl-18* and *elt-1,* which
420 showed signal enrichment in their promoter region (Figure 7A). These genes encode
421 GATA TFs that promote seam cell fate and had not been previously linked to regulation
422 by NHR-25. In particular, *egl-18* is a target of the Wnt signalling pathway that is known
423 to maintain seam cell fate (Gorrepati et al., 2013; Koh and Rothman, 2001). Expansion
424 of the *egl-18* expression domain to the anterior seam cell daughters that normally adopt
425 the hypodermal differentiation programme has been shown to correlate with ectopic

426 maintenance of the seam cell fate (Gorrepati et al., 2013; Katsanos et al., 2017).
427 Furthermore, *elt-1* is thought to be the master epidermal fate regulator in *C. elegans*
428 and is known to regulate *nhr-25* in the embryo (Chisholm and Hsiao, 2012; Gilleard
429 and Mcghee, 2001). To test whether *elt-1* and *egl-18* are likely to be targets of NHR-
430 25, we carried out smFISH quantifications in control and *nhr-25* RNAi treated animals
431 at the L3 division stage. The efficacy of the RNAi treatment was confirmed by scoring
432 for an increase in terminal seam cell number (Figure 7B). In *nhr-25* RNAi treated
433 animals, an overall increase in *egl-18* expression in V1-V4 daughters was observed
434 (Figure 7C, D), indicating that NHR-25 is likely to be a repressor of *egl-18.* Similarly,
435 *elt-1* expression was found to be increased in anterior daughters of the V1-V4 lineages
436 in *nhr-25* RNAi treated animals (Figure 7E, F). These findings suggest that NHR-25
437 promotes the differentiation programme by directly repressing key seam cell fate-
438 promoting factors.
439

440 **Discussion**

441 **LIN-22 and NHR-25 targets provide insights into epidermal cell fate specification.**

442 Our study refines the position of *lin-22* and *nhr-25* in the epidermal gene
443 network. Based on our findings and previous literature (Brabin et al., 2011; Cassata et
444 al., 2005; Chisholm and Hsiao, 2012; Hughes et al., 2013; Katsanos et al., 2017; Koh
445 and Rothman, 2001; Thompson et al., 2016; van der Horst et al., 2019), we present an
446 updated gene network underlying cell fate decisions in the epidermis (Figure 8). In light
447 of our TaDa findings, we propose that LIN-22 and NHR-25 play a prominent role in
448 mediating correct hypodermal differentiation in the epidermis. NHR-25 was known to
449 influence seam cell patterning by regulating seam cell shape and establishment of cell-
450 to-cell contacts post cell division (Chen et al., 2004; Gissendanner and Sluder, 2000;

451 Šilhánková et al., 2005), although its link to cell fate specification was not well
452 understood. Our findings suggest a role in the specification of the hypodermal cell fate
453 through direct suppression of core seam cell specifying factors, such as *egl-18* and *elt-*
454 *1.*

455 We previously proposed that LIN-22 influences seam cell divisions by
456 antagonising Wnt signalling through regulation of the expression of the Frizzled
457 receptor *lin-17* (Katsanos et al., 2017). LIN-22 is also known to supress neurogenesis
458 in V1-V4 seam cells through suppression of the Hox gene *mab-5,* as well as the pro-
459 neuronal factor *lin-32* (Wrischnik and Kenyon, 1997). However, in all these cases, there
460 was no prior evidence for a direct interaction. Our TaDa signal profiles reveal that LIN-
461 22 is likely to directly bind to *mab-5* and *lin-32* upstream sequences in the seam cells.
462 Although *Hes*-related factors are commonly associated with Notch signalling in other
463 systems (Kageyama et al., 2007), we found an unusual link between LIN-22 and Wnt
464 signalling in *C. elegans*, potentially through direct regulation of multiple components
465 within this signalling pathway. LIN-22 was also found to repress *rnt-1,* the Runx
466 homologue of *C. elegans,* which in complex with BRO-1 promote seam cell fate and
467 symmetric divisions by repressing *pop-1* (van der Horst et al., 2019). This interaction
468 is consistent with the observation of ectopic symmetric divisions in *lin-22* mutants
469 (Katsanos et al., 2017). LIN-22 may activate the cell cycle inhibitor *cki-1*. This gene is
470 expressed in the seam cells and *cki-1* RNAi increases seam cell number (Buck et al.,
471 2009; Hong et al., 1998). Since *cki-1* drives G1 arrest allowing differentiation to
472 progress (Hong et al., 1998; Matus et al., 2015), reduction of *cki-1* expression may
473 contribute to the supernumerary seam cells observed in *lin-22* mutants. Canonical *Hes*
474 factors are thought to generally act as repressors (Kageyama et al., 2007), for example
475 the mammalian homologue of *cki-1*, $p27^{Kip1}$ has been shown to be directly repressed

476 by Hes1 in mice (Murata et al., 2005). In contrast to canonical HES factors, LIN-22
477 lacks a Groucho interacting domain (Schlager et al., 2006), so it may act as a repressor
478 or activator depending on available co-factors or competition with other TFs for binding.
479

480 **Expanding the *C. elegans* toolkit for transcription factor target identification**

481 We introduce in this study TaDa as a method to identify TF targets in a tissue-
482 specific manner in *C. elegans*. Tissue-specific or constitutive levels of expression of
483 Dam-fusions have been known to lead to toxicity, which we also confirmed here in the
484 case of the *C. elegans* epidermis (Schuster et al., 2010; Southall et al., 2013; van
485 Steensel and Henikoff, 2000). The TaDa transgene configuration overcomes this
486 obstacle by minimising the levels of Dam expression in the tissue of interest. We show
487 that TaDa can be used in early larval stages and dividing cell populations, such as the
488 seam cells at the L2 stage, while still leading to robust identification of putative targets.

489 Currently, ChIP-seq is the most commonly used method to identify TF targets
490 in *C. elegans*, with many datasets available based on large-scale projects (Celniker et
491 al., 2009; Kudron et al., 2018). We present here evidence that TaDa is comparable to
492 ChIP-seq in identifying target genes, while it offers two key advantages. First, TaDa
493 allows the identification of TF targets in a specific tissue of interest without cell isolation,
494 as opposed to all tissues where the TF is expressed (Aughey et al., 2019; Aughey and
495 Southall, 2016). Second, TaDa requires substantially less material than ChIP-seq. We
496 found approximately 2000 animals to be sufficient to allow the recovery of enough
497 methylated DNA from the seam cells, which represent 3% of the total cells of *C.*
498 *elegans*. This is also consistent with observations in mammalian cell lines, where a lot
499 more cells were used to identify targets for the same TF by ChIP-seq than TaDa
500 (Cheetham et al., 2018).

501 When direct comparison across methods could be made, the majority of TaDa
502 putative target genes (>61%) were also identified by stage-matched ChIP-seq. The
503 NHR-25 ChIP-seq L2 dataset was approximately three times larger in terms of peaks
504 compared to TaDa L2 (2044 peaks in TaDa compared to 5980 in ChIP-seq). Although
505 tissue specificity is likely to contribute to this difference, there are also alternative
506 explanations. Both ChIP-seq and DamID methods have inherent biases due to the
507 protocol and reagents used including the quality of the Dam fusions (Ramialison et al.,
508 2017), issues with antibody affinity or PCR biases (Meyer and Liu, 2014). A key
509 limitation in all DamID-based methods is the dependence on availability of GATC sites,
510 which could hinder detection of some targets. However, the average length of GATC
511 fragments in *C. elegans* is 367 bp and the median 204 bp, so depletion of GATC sites
512 is unlikely to be pervasive enough to significantly undermine detection of targets. TaDa
513 peak profiles may be distinct to ChIP-seq in other ways too, for example they were
514 found to be less inclusive of HOT regions (Teytelman et al., 2013), with only 13% of
515 L2 HOT regions being represented in NHR-25 TaDa in contrast to 83% in ChIP-seq.
516 While TaDa peaks tend to be broader than those obtained from ChIP-seq experiments,
517 we demonstrate that they can be used to identify TF binding motifs, for example based
518 on GATC fragments that show the highest local enrichment.
519 In summary, TaDa can be a powerful approach to dissect complex TF
520 behaviours related to tissue-specific target binding. This includes for example the HLH
521 family of transcription factors in *C. elegans* that are thought to bind different targets in
522 different tissues depending on their dimerising partners (Grove et al., 2009). TaDa
523 predictions can be followed up with single cell resolution using smFISH in appropriate
524 mutant or silenced backgrounds. This experimental combination can be useful to add
525 quantitative strengths to newly described interactions, thereby facilitating mathematical

526 modelling of developmental gene networks (Piano et al., 2006). The revised epidermal
527 network represents a framework for future experiments to build upon. Given the
528 conserved nature of some of the participating factors, this network can inform about
529 interactions that underlie more broadly stem cell behaviour to dissect the
530 developmental logic of robust stem cell fate patterning.
531

532 **Acknowledgments**

533 We thank Tony Southall, Colin McClure and Gabriel Aughey for discussions on setting
534 up the method. We thank members of the Barkoulas lab and David Matus for
535 comments on the manuscript. Some *C. elegans* strains were provided by the CGC,
536 which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).
537 We acknowledge the support from the European Commission [ROBUSTNET-
538 639485]. DK was the recipient of an Imperial College President's PhD scholarship.
539

540 **Author contributions**

541 DK carried out all experiments and data analysis. MB supervised the work. DK and
542 MB wrote the manuscript.
543

544 **Declaration of Interests**

545 The authors declare no competing interests.
546

547 **Figure Legends:**

548 **Figure 1: TaDa transgene design prevents animal toxicity.**

549 **(A)** Schematic showing the bicistronic design of TaDa with a primary ORF of *mCherry*
550 followed by two STOP codons and a frameshift, which permits low levels of tissue-
551 specific expression of a fusion between the protein of interest (POI) and Dam. **(B)**

552 Illustration of the key features of single-copy transgenes used in this study for LIN-22
553 and NHR-25 target identification by TaDa (left). Transgenes used to assess the
554 requirement of an *mCherry* primary ORF to prevent toxicity and high methylation levels
555 are shown on the right. **(C)** Representative images of adult wild type (WT) and
556 transgenic animals carrying TaDa fusions. Note that animals carrying *lin-22:dam* or
557 *NLS-GFP:dam* fusions, that is in the absence of the primary *mCherry* ORF, show
558 aberrant phenotypes. White arrows point to internal tissue outside the animal body.
559 **(D)** Quantification of brood size in the above strains. Note defect in brood size in
560 animals carrying *lin-22:dam* and *NLS-GFP:dam* fusions (n=15). Scale bars in C are
561 100 µm. In D error bars indicate standard error of the mean and black stars indicate
562 statistically significant differences in the mean with a one-way ANOVA followed by a
563 Dunnett's test (* $p<0.05$, **** $p<0.0001$).
564
565 **Figure 2: LIN-22 and NHR-25 binding signal is enriched in upstream gene**
566 **regulatory regions**. **(A-B)** Examples of signal profiles in selected regions associated
567 with putative target genes. Shaded regions indicate statistically significant peaks
568 (FDR<0.05). The Y-axes represent $\log_2$*(TF:dam*/*NLS-GFP:dam)* scores (Data range
569 for *lin-22:dam*: -1 – 3.5, and for *nhr-25:dam*: -3 – 8). Scale bar length is 5 kb as
570 indicated. **(C)** Aggregation plots depicting average enrichment scores in 10 bp bins for
571 regions of equal length across all of the specified genomic features indicated on the
572 X-axis. Strong enrichment preference is seen upstream to gene regions. Plots show 5
573 kb upstream of the TSS of genes to 1 kb downstream of the TES, with gene bodies
574 pushed into a 2 kb pseudo-length. Y-axes are z-scores for the plotted sequence length
575 and shaded areas represent 95% confidence intervals. **(D)** Heatmaps representing the

576 hierarchically clustered localisation and enrichment score of all statistically significant
577 peaks (FDR<0.05) within 5 kb upstream and 1 kb downstream of a gene.
578
579 **Figure 3: LIN-22 and NHR-25 peaks largely overlap genomic regions with**
580 **regulatory potential.**
581 **(A)** Pie charts indicate the proportions of peaks residing in various genomic locations.
582 Peaks were assigned to the single closest gene when their centre coordinate was
583 found within 6 kb upstream and 1 kb downstream of the TSS and TES respectively of
584 a gene. Ratios above pie charts show the number of assigned peaks to the total
585 number of significant peaks found. The proportions of peaks within genes with
586 exclusive intron or exon localisation are indicated under the pie charts. **(B)**
587 Aggregation plots of open chromatin signal from ATAC-seq L2 and L4 data (Jänes et
588 al., 2018) over LIN-22 and NHR-25 TaDa L2 and L4 peaks respectively. Note
589 increased chromatin openness at the sites of peaks. Y-axes are z-scores for the
590 plotted sequence length and shaded areas represent 95% confidence intervals. ±5 kb
591 around the peak centres is shown. **(C)** Proportions of total peaks for each indicated
592 TF at each stage overlapping with different categories of regulatory annotated open
593 chromatin elements. A single TF peak may overlap with more than one open chromatin
594 element.
595
596 **Figure 4: Peak localisation profiles align with ChIP-seq data and reveal overlap**
597 **between the two transcription factors.**
598 **(A)** Representative snapshots showing agreement between *nhr-25:dam* TaDa data
599 and two ChIP-seq signal profiles from L1 (Shao et al., 2013) and L2 (Araya et al.,
600 2014) staged animals over *nhr-25*, *idh-1* and *egl-18*. Shaded regions indicate

601 statistically significant peaks (FDR<0.05). The Y-axes for TaDa samples represent
602 $\log_2$*(nhr-25:dam/NLS-GFP:dam)* scores with data range: -3 – 8 while for ChIP-seq the
603 fragment pile-up per million reads score with data range: 0-10 for L1 and 0-1000 for
604 L2. Scale bar is 5 kb as indicated. **(B)** Aggregation plots showing enrichment of *nhr-*
605 *25:dam* L2 and L4 signal over centre of NHR-25 ChIP-seq L1 (left) and L2 (right) peaks
606 and of NHR-25 ChIP-seq L1 and L2 signal over centre of NHR-25 TaDa L2 (left) and
607 L4 (right) peaks. **(C)** Graph indicating the number of peaks from *nhr-25:dam* L2 and
608 L4 samples that overlap with the L1 or with the L2 NHR-25 ChIP-seq peaks datasets.
609 **(D)** Aggregation plots of *lin-22:dam* L2 and L4 signal over NHR-25 TaDa L2 and L4
610 peaks (top) and *nhr-25:dam* L2 and L4 signal over LIN-22 TaDa L2 and L4 peaks. All
611 show increased signal for the peak location. **(E)** Graph showing the numbers of peaks
612 for each colour-coded sample that overlap with the peaks of the sample indicated on
613 the X-axis either across the whole genome (left) or restricted to promoter regions 5 kb
614 upstream to 500 bp downstream of the TSS (right). **(F)** Hierarchical clustering of
615 symmetrized co-association values for all pairwise comparisons between peak
616 localisation patterns of the indicated TFs at the L2 stage. Peak profiles are taken from
617 previous ChIP-seq experiments (Kudron et al., 2018) or our TaDa results. Note the
618 separate clustering of TaDa peaks together with epidermal factors. In (B) and (D), ±5
619 kb around the peak centres have been plotted and the Y-axes represent z-scores for
620 the plotted sequence length and shaded areas represent 95% confidence intervals. In
621 (C) and (E), the exact number of peaks with overlaps is written above the bars and *p*-
622 values from Monte Carlo simulations for the statistical significance of overlaps is
623 shown inside the bar.
624

625 **Figure 5: Enriched GO terms for LIN-22 and NHR-25 putative targets relate to**
626 **their biological functions**.
627 **(A-B)** Venn diagrams for putative target intersections between genes identified at L2
628 and L4 for the *lin-22:dam* (A) and *nhr-25:dam* fusions (B). The intersection in both
629 cases is significant with a hypergeometric distribution test $p < 0.0001$). **(C-H)** Plots of
630 selected significantly enriched terms from GO term analysis for the genes in the L2
631 (C,F), L4 (E,H) or intersection datasets (D,G). For both TFs at all stages, relevant
632 terms are recovered related to neurogenesis, development or molting.
633
634 **Figure 6: LIN-22 activates *cki-1* and represses *rnt-1 and lin-17.***
635 **(A)** *lin-22:dam* signal forming significant peaks (shaded regions) around genes
636 identified as putative LIN-22 targets. Y-axes are $\log_2$*(lin-22:dam/NLS-GFP:dam)*
637 scores with data range: -1 – 3.5. **(B)** Representative *cki-1* smFISH images from WT
638 and *lin-22(icb49)* animals at the late L1 stage. **(C)** Quantification of *cki-1* mRNA spots
639 in H0-V4 seam cells, showing a significant reduction of expression in V1-V4 cells
640 ($50 \leq n \leq 126$). **(D)** Representative *rnt-1* smFISH images from WT and *lin-22(icb49)*
641 animals at the late L1 stage. **(E)** Quantification of *rnt-1* mRNA spots in H0-V4 seam
642 cells, showing a significant increase in expression in V1-V4 cells ($65 \leq n \leq 167$). **(F)** Vista
643 analysis of the *lin-17* promoter identified two conserved elements (*CRE1* and *CRE2*)
644 between *C. elegans, C. briggsae* and *C. brenneri*. The position of the elements is
645 shown on the *C. elegans* promoter sequence as a reference. Note that these elements
646 are both overlapping GATC fragments with high TaDa enrichment scores (shown as
647 red in the heatmap). **(G-H)** Representative brightfield and fluorescence images of L4
648 transgenic animals carrying transcriptional reporters for the *lin-17 CRE1* (*icbEx177*
649 transgene) (G) and *CRE2* sequences (*icbEx180* transgene) (H) fused to *GFP:H2B* in

650 WT and the *lin-22(icb49)* mutant background. Note that expression is restricted to few
651 posterior cells in WT and expands to the anterior epidermis in the mutants. White
652 arrowheads indicate epidermal GFP expression. *lin-22(icb49)* mutants also carry a
653 *dat-1p::GFP* reporter labelling dopaminergic neurons, which is distinguishable by the
654 visible axons. **(I)** Quantification of the proportion of transgenic animals with epidermal
655 expression anterior to the vulva in WT (n=35 for *CRE1* and n=24 for *CRE2*) and *lin-*
656 *22(icb49)* mutants (n=47 for *CRE1* and n=20 for *CRE2*). Error bars indicate the
657 standard error of the proportion. Black stars show statistically significant differences
658 with a Fisher's exact test, *** $p<0.001$. In B and D, seam cells are labelled with
659 *SCMp:GFP* and black spots correspond to investigated mRNAs. Scale bars are 5 kb
660 in A, 10 µm in B and D, 100 µm in G and H.
661
662 **Figure 7: NHR-25 represses *egl-18* and *elt-1.***
663 **(A)** *nhr-25:dam* signal forming significant peaks (shaded regions) around genes
664 identified as putative NHR-25 targets. **(B)** Seam cell number comparison between
665 control and *nhr-25* RNAi treated animals used for smFISH (n≥36). **(C)** Representative
666 *egl-18* smFISH images of control and *nhr-25* RNAi treated animals during the L3
667 division. **(D)** Quantification of *egl-18* mRNA spots in the V1-V4 daughter cells following
668 the L3 division (60≤n≤88) showing an overall increase in expression upon *nhr-25*
669 RNAi. **(E)** Representative *elt-1* smFISH images of control and *nhr-25* RNAi treated
670 animals during the L3 division. **(F)** Quantification of *elt-1* mRNA spots in the V1-V4
671 daughter cells following the L3 division showing a significant increase in expression in
672 the anterior daughters (Vn.papa and Vn.pppa, 82≤n≤116). In C and E, seam cells are
673 labelled with *SCMp:GFP* and black spots correspond to investigated mRNAs. Scale

674 bars are 5 kb in A and 10 μm in C and E. Arrowheads in C and E indicate instances of
675 strong expression in anterior daughter cells.
676
677 **Figure 8: A combined gene network of epidermal cell fate interactions.**
678 Blue and magenta dashed lines indicate new interactions predicted by TaDa for LIN-
679 22 and NHR-25 respectively. Activation or repression of targets is shown based on
680 smFISH validation when known. Black dashed lines indicate published genetic
681 interactions that is yet unknown whether they are direct or not (except for the link
682 between *elt-1* and *bro-1*, which is likely to be direct). Hypodermal fate is shown in grey
683 and seam cell fate in green.
684
685 **Materials and methods**
686 ***C. elegans* maintenance**
687 The *C. elegans* strains used in this study were maintained on standard Nematode
688 Growth Medium (NGM) plates and were grown monoxenically on a lawn of *E. coli*
689 OP50 at 20 °C. For TaDa experiments, strains were grown for at least two generations
690 on a lawn of a *dam⁻/dcm⁻ E. coli* mutant of the K12 strain (New England biolabs,
691 C2925). The laboratory reference N2 strain was used as the reference strain in this
692 study. A complete list of strains used in this study is available in Table S3.
693 **Molecular cloning**
694 To construct a TaDa backbone plasmid for epidermal expression of TFs fused
695 upstream of Dam, the pCFJ151 universal MosSCI vector (Frokjaer-Jensen et al., 2014)
696 was digested with BcuI/BspTI enzymes. The promoter of *wrt-2* was amplified from N2
697 lysate with oligos PB16 and PB7, the *C. elegans* optimised *mCherry* was amplified
698 from the pAA64 (Barkoulas et al., 2016) plasmid using oligos PB8 and PB17, the *Dam*

699 sequence was amplified from the pUAST attB LT3 Dam plasmid (gift by Tony Southall)
700 using oligos PB18 and PB13 and the *unc-54 3' UTR* was amplified from N2 lysate
701 using oligos PB14 and PB15. All 4 fragments and the digested backbone were inserted
702 in a multi-fragment Gibson assembly reaction to produce the pPB7(*wrt-*
703 *2p::mCherry::Dam::unc-54 3'UTR* + *cb-unc-119*) plasmid. The XmaJI site between
704 *mCherry* and *Dam* was digested to linearise the vector and allow the in-frame to *Dam*
705 insertion of the *nhr-25* coding sequence amplified using oligos PB19 and PB20 from
706 N2 cDNA to produce the pPB10(*wrt-2p::mCherry::nhr-25:Dam::unc-54 3'UTR* + *cb-*
707 *unc-119*) plasmid. To construct the seam cell driven *lin-22:Dam* fusion and the *NLS-*
708 *GFP:Dam* control for the TF TaDa experiments, the *lin-22* gene was amplified with
709 oligos DK11 and DK12 from fosmid WRM0627dG07 while *NLS-GFP* was amplified
710 from plasmid pPD93_65 (Fire Lab vector Kit) using oligos DK15 and DK16. Both
711 amplicons were inserted upstream and in-frame with *Dam* by Gibson assembly in an
712 XmaJI linearised pPB7 vector like above. The resulting plasmids produced were
713 pDK4(*wrt-2p::mCherry::lin-22:Dam::unc-54 3'UTR* + *cb-unc-119*) and pDK8(*wrt-*
714 *2p::mCherry::nhr-25:Dam::unc-54 3'UTR* + *cb-unc-119*). To test the importance of the
715 *mCherry* primary ORF to the viability of animals and methylation levels, versions of the
716 *lin-22:Dam* and *NLS-GFP:Dam* TaDa constructs without *mCherry* were produced. In
717 more detail, the pPB7 plasmid was digested with BcuI/MunI, the 4085 bp and 6141 bp
718 fragments were excised, extracted and kept. *lin-22* was amplified from pDK4 using
719 DK102 and DK11. The 2 digestion fragments, *the lin-22* amplicon and the repair oligo
720 DK103 were all inserted into a Gibson reaction to produce pDK49(*wrt-2p::lin-*
721 *22:Dam::unc-54 3'UTR* + *cb-unc-119*) which was then digested with BcuI/XmaJI to
722 remove *lin-22* and insert via Gibson assembly a DK108 and DK15 amplified fragment

723 of NLS-GFP from pDK8 to generate pDK50(*wrt-2p::NLS-GFP:dam::unc-54 3'UTR* +
724 *cb-unc-119*).
725 For ease of future applications, a versatile TaDa vector called pDK7 was constructed.
726 Briefly, the *att* recombination cassette of pDest R4-R3 (Invitrogen) was amplified
727 including the *attR4* site, *ccdb* and *CamR* genes but excluding the *attR3* site using the
728 oligos DK17 and DK18 including half of the *attL1* site sequence on the 3' DK18 primer.
729 *mCherry* was amplified from pPB7 using oligos DK19 and DK20 carrying the other half
730 of the *attL1* site on the 5' of DK19. *dam* was amplified from pPB7 with oligos DK21 and
731 DK22. The *unc-54 3'UTR* was amplified from pPB7 with oligos DK23 and DK24. All 4
732 fragments were inserted in a Gibson assembly reaction along with BcuI/BspTI doubly
733 digested pCFJ151 vector to generate pDK7(*attR4-L1::mCherry::Dam-myc::unc-54*
734 *3'UTR* + *cb-unc-119*).
735 For the *lin-17 CRE1* and *CRE2* transcriptional reporters the oligos DK115 and DK116
736 along with DK118 and DK119 were used to amplify each of the respective regions from
737 N2 lysate. The *Δpes-10* core promoter was amplified from L3135 using either the *CRE1*
738 or the *CRE2* compatible forward primers DK117 and DK120 along with the DK107
739 reverse and was cloned along with the respective CRE amplicon in a NheI/XmaJI
740 digested pDK16 to create pDK59(*lin-17CRE1:: Δpes-10::GFPo-H2B::unc-54 3'UTR* +
741 *cb-unc-119*) and pDK60(*lin-17CRE2:: Δpes-10::GFPo-H2B::unc-54 3'UTR* + *cb-unc-*
742 *119*). A complete list of the oligos used in this study is presented in Table S4.
743 **Single copy transgenesis**
744 Single-copy transgenic lines were produced using the *Mos1*-mediated single-copy
745 insertion (MosSCI) method (Frokjaer-Jensen et al., 2014). Briefly, ~30 day-one adult
746 animals of the EG6699 strain with a *Mos1* transposon insertion on chromosome II
747 (*ttTi5605* locus) showing the uncoordinated (Unc) phenotype were injected for each

748 transgene insertion. All the MosSCI injection mixes comprised of 50 ng/µl of a universal
749 MosSCI vector carrying the transgene of interest flanked by the *ttTi5605* left and right
750 recombination arms along with plasmids harbouring the *Mos1* transposase (pCFJ601
751 at 50 ng/µl), a heat-shock inducible *peel-1* toxin (pMA122 at 10 ng/µl) and co-injection
752 markers (pGH8 at 5 ng/µl, *myo-2::dsRed* at 2.5 ng/µl, *myo-3::mCherry* at 5 ng/µl). Post-
753 injection, animals were kept at 25 °C until plates were completely starved. The heat-
754 shock treatment that follows was performed at 34 °C for 3.5 hours, after which the
755 plates were allowed to recover for 3 hours at room temperature before "reverse
756 chunking" was performed, where NGM chunks from the lawn of a new plate were
757 placed on top of the starved, treated plate with the OP50 lawn facing upwards
758 (O'Connell, 2010). The next day the top of lawns were screened for *unc-119(-)* rescued
759 animals with absence of co-injection markers, which were transferred on different NGM
760 plates per injected P0. Homozygous lines were confirmed molecularly for the presence
761 of single-copy transgene insertions using oligos NM3880 and NM3884. A complete list
762 of the transgenes produced for this study along with injection mix make-up information
763 is available in Table S5.

764 **Microscopy**

765 For seam cell imaging, live animals were mounted on fresh 2% agarose pads
766 containing 100 µM NaN_3 for immobilisation on glass slides. The slides were then
767 imaged using either an AxioScope A1 (Zeiss) upright epifluorescence microscope with
768 a Light Emitting Diode (LED) light source fitted with a RETIGA R6™ camera (Q
769 IMAGING) controlled via the Ocular software (Q IMAGING) or on an inverted Ti-eclipse
770 fully motorised epifluorescence microscope (Nikon) with a metal halide light source
771 fitted with an iKon M DU-934, 1024 x 1024 CCD-17291 camera (Andor) controlled via
772 the NIS-Elements software (Nikon). Scoring of the terminal seam cell or postdeirid

773 neuron number phenotype was performed in late-L4 or early adult animals, carrying
774 the *SCMp::GFP* or *dat-1p::GFP* markers. The lateral side most proximal to the
775 objective was counted for every animal.
776 To perform smFISH, large populations of animals were synchronised by bleaching and
777 were subsequently grown at 20 °C for 18 hours to reach the late L1 and 35 hours for
778 the L3 asymmetric division stage. Animals were fixed in 4% formaldehyde (Sigma-
779 Aldrich) in 1x PBS (Ambion) for 45 min on a vertical Stuart™ Rotating disk (Cole-
780 Palmer). They were washed with 1.5 ml 1x PBS twice before being stored at 4 °C in
781 70% Ethanol for at least 24 hours for permeabilisation. Hybridisation was performed in
782 100 µl of buffer (100 mg/ml dextran sulphate (Sigma-Aldrich), 10% formamide in 2x
783 SSC) at 30 °C for 16 hours with1 µl of a diluted in water probe added. The probe
784 dilution was between 1:5 and 1:50 of a custom-made mixture of 21-48 Cy5-labelled
785 oligonucleotides (Biomers) targeting the gene of interest. A complete list of probes and
786 their dilution used in this study along with their sequences is available in Table S4.
787 Animals were washed in a solution of 10% formamide, 2x SSC, stained with DAPI and
788 resuspended for imaging in 100 µl GLOX buffer (0.4% glucose, 10 mMTris-HCl in 2x
789 SSC) supplemented with 1 µl of 3.7 mg/ml glucose oxidase (Sigma-Aldrich) and 1 µl
790 of 5 mg/ml Catalase (Sigma-Aldrich). Imaging was performed using the Nikon set-up
791 described above using the seam cells closest to the objective lens as homing
792 coordinates to acquire 17 Z-stack slices with a step of 0.8 µm for each of the DAPI,
793 Cy5 and GFP channels (Semrock). Acquisition was performed using a 100x oil
794 immersion objective with exposure set at 100 ms for DAPI at 1/32 reduced light
795 intensity, 3 s for Cy5 and 300 ms for GFP. Analysis was performed using a custom
796 MATLAB (MathWorks) pipeline (Barkoulas et al., 2013). In brief, selected animal DAPI
797 and GFP images were used to annotate seam cells and draw regions of interest (ROIs)

798 around the nuclei for at least 5 slices within which smFISH spots would be counted.
799 An animal specific threshold for spot detection was set by manually sampling spots.

800 **TaDa wet lab protocol**

801 Strains for TaDa experiments were transferred onto *dam⁻/dcm⁻* plates by spot
802 bleaching and two biological replicates in separate plates were processed
803 simultaneously. For each strain and replicate, nine 55 mm plates fully populated by
804 gravid adults were used for large-scale egg preparation with isolated embryos seeded
805 in new *dam⁻/dcm⁻* plates that were incubated at 20 °C. Half of the resulting
806 synchronised populations were collected in a 15 ml centrifuge tube after 24 hours at
807 the L2 stage and the other half after 48 hours at the L4 stage using 2 ml M9 buffer per
808 plate. All collected populations underwent extensive washing by centrifuging animals
809 at 1200 g for 3 min, removing the supernatant and washing the pellet (~100 µl) with 10
810 ml M9 at least 5 times. Animal pellets were frozen at -20 °C before gDNA extraction.
811 For gDNA extraction pellets were lysed using 750 µl of Cell Lysis Solution (QIAGEN)
812 containing 100 µg/ml of Proteinase K on a heat-block at 55 °C shaking at 500 RPM for
813 16 hours overnight. Lysates were treated with 4 µl of 100 mg/ml RNase A (QIAGEN)
814 at 37 °C shaking at 500 RPM for 3 hours. In turn, 250 µl of Protein Precipitation Solution
815 (QIAGEN) was added to each sample and were incubated on ice for 5 min followed by
816 vigorous vortexing for at least 30 sec, another 5 min incubation on ice and a
817 centrifugation at 6000 g for 10 min at 4 °C. The supernatant for each sample was
818 treated with isopropanol, washed with 70% ethanol and the pellet was air-dried for 1
819 hour. DNA pellets were hydrated with 55 µl of UltraPure™ distilled water and were left
820 for 48 hours at 4 °C for the DNA pellet to dissolve.
821 For isolation and amplification of the GATC methylated gDNA fragments the protocol
822 followed here is an adapted version of the one presented in (Marshall et al., 2016) for

823 TaDa in *Drosophila* with minor alterations. Of the extracted gDNA samples above, a
824 total amount of 5 μg was transferred into a new 1.5 ml tube and was brought to 43 μl
825 with the addition of UltraPure™ Distilled Water. For those samples where 5 μg were
826 not available, 43 μl of the original sample were transferred. To each of those tubes 5
827 μl of 10x CutSmart Buffer (New England Biolabs) and 2 μl of DpnI restriction enzyme
828 (New England Biolabs) were added and mixed by gentle flicking to prevent shearing of
829 gDNA. The samples were incubated at 37 °C for 16 hours overnight and were in turn
830 cleaned-up using the QIAquick PCR Purification kit (QIAGEN) and eluted with 40 μl of
831 50 °C water. 20 μl of each clean digestion product were split equally in two 0.2 ml PCR
832 tubes (15 μl in each) and 4 μl of *Adaptor ligation buffer* (5x T4 DNA Ligase Buffer (New
833 England Biolabs) and 10 μM of the dsAdR adaptor along with 1 μl (400 U) T4 DNA
834 Ligase (New England Biolabs) were added in each. The samples were then incubated
835 at 16 °C for 2 hours, followed by 10 min at 65 °C in a thermocycler. The double-
836 stranded adaptor dsAdR was initially prepared by mixing equal volumes of 100 μM of
837 the single stranded oligos AdRT and AdRb in a 1.5ml tube, immersing in a boiling-hot
838 water-bath and letting it to cool down to room temperature to allow for gradual
839 annealing. Following the adaptor ligation each sample was mixed with 20 μl of a 2x
840 DpnII Digestion Buffer and 10 units (1 μl) of DpnII restriction enzyme (New England
841 Biolabs) mastermix and were incubated for 3 hours at 37 °C. At this stage for each
842 original sample two 40 μl digestion reaction products were available. For methylated
843 DNA amplification by PCR each one of these products was mixed with 118 μl of *DamID*
844 *PCR buffer* and 2 μl (10 units) of MyTaq™ DNA polymerase (Bioline) and were
845 aliquoted at 40 μl in 4 different 0.2 ml PCR tubes. The *DamID PCR buffer* consisted of
846 1.36x MyTaq™ Buffer (Bioline) and 1.06 μM of the DamID PCR primer (Adr_PCR) that

847 anneals on the adaptor sequence. In total for each original gDNA sample 8 PCR
848 reactions were performed using the following cycling program:
849 Single cycle of steps 1-4: 72 °C for 10 min, 94 °C for 30 sec, °C for 5 min, 72 °C for 15
850 min
851 3 Cycles of steps 5-7: 94 °C for 30 sec, 65 °C for 1 min, 72 °C for 10 min
852 21 Cycles of steps 8-10: 94 °C for 30 sec, 65 °C for 1 min, 72 °C for 2 min
853 Final extension step: 72°C 5 min, Slow cool down to room temperature and storing at
854 10 °C.
855 The number of cycles for steps 8-10 was increased from 17 described in (Marshall et
856 al., 2016) to 21, which is more commonly used in previous DamID experiments in *C.*
857 *elegans* (Askjaer et al., 2014). The 8 reactions per sample were pooled and cleaned-
858 up using QIAquick PCR Purification kit (QIAGEN). To remove the adaptor sequences
859 from the resulting PCR products up to 2.5 µg of product was transferred to a new 1.5
860 ml tube and was digested with AlwI restriction enzyme (New England Biolabs) at 37
861 °C for 16 hours. The products were cleaned-up again using the QIAquick PCR
862 Purification kit. These final purified amplicons were sent to GENEWIZ for library
863 preparation and Next Generation Sequencing (NGS) using the Illumina HiSeq platform.
864 **Calculation of TaDa signal profiles for LIN-22 and NHR-25 binding**
865 FASTQ files representing single-end reads for each sample and replicate, were initially
866 assessed using the fastq-stats perl script (available at
867 https://github.com/owenjm/damid_misc/blob/master/fastq-stats) for uncut adaptors,
868 primer dimer and internal GATC content as a post-sequencing quality control step for
869 the wet lab executed protocol. The sequencing reads were mapped on the *C. elegans*
870 genome, sequence alignment read-count maps were generated and normalised
871 $\log_2$(*TF:dam/NLS-GFP:dam*) ratio scores were calculated per GATC fragment of the

872 genome using the perl script damidseq_pipeline v1.4.5 (Marshall and Brand, 2015)
873 (available at https://github.com/owenjm/damidseq_pipeline). The pipeline was used
874 calling Bowtie 2 v2.3.4 (Langmead and Salzberg, 2012) for alignment to *C. elegans*
875 bowtie indices from genome assembly WBcel235 (available from illumina iGenomes
876 page), Samtools v1.9 (Li et al., 2009) for alignment manipulations and a GATC
877 fragment file with the coordinates of all GATC fragments across the *C. elegans* genome
878 in gff format, built from a WBcel235 FASTA file (available at
879 https://www.ensembl.org/Caenorhabditis_elegans/Info/Index) using the
880 gatc.track.marker.pl script (available at https://github.com/owenjm
881 /damidseq_pipeline). The number of usable reads that have been acquired by the
882 experiments presented here and map only once to the genome varied from over 6
883 million up to ~30 million reads per sample with the genome coverage being between
884 ~9 and 45 times. From the 2 replicates per TF-Dam fusion and control-Dam fusion all
885 pairwise genome-wide $\log_2$(*TF:dam/NLS-GFP:dam*) calculations were performed and
886 averaged into a single signal profile of $\log_2$(*TF:dam/NLS-GFP:dam*) enrichment scores
887 per GATC of the genome for each TF at each developmental stage. Genomic
888 coordinate files produced and used throughout this study were converted between
889 formats (BED, GFF, Bedgraph, BigWIg, Wig, GTF) using Excel, the Convert between
890 GTrack/BED/WIG/bedGraph/GFF/FASTA files tool of the Galaxy powered GSuite
891 Hyperbrowser (elixir) (at https://hyperbrowser.uio.no/hb/!mode=advanced) and the
892 UCSC browser binaries bedGraphToBigWig, BigWigToBedGraph, bigWigToWig. BED
893 and GFF signal and feature track files were visualised using the SignalMap NimbleGen
894 software (Roche).

895 **Peak-calling and gene-assignment**

896 Identification of statistically significant enriched peaks across the genome for each TF
897 and developmental stage was performed using the perl script find_peaks (available at
898 https://github.com/owenjm/find_peaks) with an FDR<0.05 and default settings with the
899 averaged $\log_2$(*TF:dam/NLS-GFP:dam*) signal profiles as input. The output is a list of
900 genomic interval coordinates for statistically significant peaks with a peak enrichment
901 score and an FDR value. Significant peaks were assigned to genes using UROPA
902 (Kondili et al., 2017) as a web tool (available at http://loosolab.mpi-
903 bn.mpg.de/UROPA_GUI/) with Caenorhabditis_elegans.WBcel235.99.gtf (from
904 http://www.ensembl.org/Caenorhabditis_elegans/Info/Index) as the genome
905 annotation file. Peaks were assigned to genes on any strand when their centre
906 coordinate was positioned up to 6 kb upstream of a gene start site or 1 kb downstream
907 of the end site. The location of the peak relative to the gene was assigned based on
908 the full length of the peak and the strand of the gene. To avoid discarding valid
909 regulatory relationships and due to the compactness of the *C. elegans* genome no
910 prioritisation was set for peaks when multiple assignments were available and all are
911 reported.
912 For the NHR-25 L1 ChIP-seq dataset (Shao et al., 2013) used in this study for peak
913 localisation comparisons, raw sequencing data (GEO number: GSE44710) were
914 processed into significant peak profiles using MACS2 on Galaxy
915 (https://usegalaxy.eu/). NHR-25 and input bedgraph signal tracks were inserted into
916 the MACS2 bdgcmp tool (default: Poisson *p*value algorithm) to deduct noise and
917 identify NHR-25 specific signal. The output was then used with the MACS2
918 bdgpeakcall tool (cutoff 1.0, min-length 200 and max-gap 30) to generate genome-
919 wide profiles of significant peaks. Profiles representing different replicates were
920 merged using bedtools merge with averaged heights for peaks that overlap.

921 **Pearson's correlation and principal component analysis**

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

928 **Aggregation plots and heatmaps of signal localisation**

929 Signal localisation preference around given genomic features presented as
930 aggregation plots or heatmaps were generated using the SeqPlots GUI application
931 (Stempor and Ahringer, 2016) with specific settings mentioned individually for each
932 presented result. Aggregation plots represent signal averages for 10 bp bins in regions
933 of varying but specified length around positional features of the genome. For genes all
934 of their start and end coordinates, based on the largest transcript and used here as the
935 transcriptional start site (TSS) and transcriptional end site (TES) of genes, are
936 anchored to two positions of the X-axis and their genic sequence is pushed or stretched
937 to a pseudo-length of usually 2 kb. For other features the midpoint coordinate is used
938 to align all to the same position on the X-axis which then extents upstream and/or
939 downstream of that region. For each position around the feature an average is
940 calculated across all the features to generate the aggregation plot line with a shaded
941 area representing the 95% confidence interval. When z-scores are presented on the
942 Y-axes, those have been calculated as deviations from the mean signal seen across
943 the plotted region. In heatmaps each line represents each occurrence of a genomic
944 feature indicated on the X-axis along with a surrounding region of a given length also

945 indicated on the X-axis. Colourscales indicate the positional enrichment score
946 calculated as averages per 10 bp bins.

947 **Assessment of overlaps between sets of genomic intervals or gene sets**

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

970 were performed either on http://nemates.org/MA/progs/overlap_stats.html or using the
971 R software package *SuperExactTest* (Wang et al., 2015) respectively. For both tests
972 when sets of coding genes are compared the size of the sampling pool was set to
973 20191, the number of annotated coding genes in the WBcel235.99 release.
974 Representation of overlaps was either in the form of Venn diagrams generated using
975 http://bioinformatics.psb.ugent.be/webtools/Venn/ or in the form of the output of the
976 *SuperExactTest* package.

977 **TF motif identification by TaDa**

978 Identification of motifs from transcription factor TaDa peaks was done here using
979 HOMER (Heinz et al., 2010). The top 200 peaks with the highest averaged enrichment
980 score were used for the analysis. Peak interval files were used as input for the
981 findMotifsGenome.pl script using the ce11 genome assembly, masking of the
982 sequences and the option to analyse the size of sequences provided by the interval
983 file (options: ce11 -size given -mask). The logos presented here for motifs identified
984 using homer were generated after converting the homer positional weight matrix into a
985 transfac matrix using the RSAT (Nguyen et al., 2018) Metazoa convert matrix tool
986 (http://rsat.sb-roscoff.fr/convert-matrix_form.cgi) and importing to Weblogo3 (Crooks
987 et al., 2004) (http://weblogo.threeplusone.com/) for logo drawing. Identification of
988 similar known motifs to the *de novo* identified motifs was performed using the Tomtom
989 tool of MEME-suite (Gupta et al., 2007). The default parameters were used and the
990 interrogated motif matrices were compared against the JASPAR core 2018 non-
991 redundant database.

992 **Gene set enrichment analysis**

993 Gene sets identified in this study were assessed for enriched gene ontology (GO)
994 terms or association with tissue specific expression using the worbase.org Enrichment

995 Analysis tool (Angeles-Albores et al., 2016; David Angeles-Albores, Raymond Y. N.
996 Lee, n.d.) (https://wormbase.org/tools/enrichment/tea/tea.cgi), using a *q*-value
997 threshold of 0.1. For significant GO terms presented here the $-\log_{10}q$ value is plotted.
998 Association of gene sets with biological pathways was evaluated using the gProfiler
999 gOst tool (https://biit.cs.ut.ee/gprofiler/gost) and a g:SCS calculated significance
1000 threshold of 0.05.

1001 **RNAi**

1002 Knockdown of *nhr-25* was performed by feeding animals on lawns of bacteria
1003 expressing dsRNA targeting *nhr-25* (Source Bioscience). Bacteria were grown
1004 overnight and then seeded directly onto NGM plates containing 1 μM IPTG, 25 μg/ml
1005 ampicillin and 6.25 μg/ml tetracycline. 5 L4 animals of the strain of interest were
1006 transferred on RNAi plates and were allowed to lay progeny that were phenotyped at
1007 the stage of interest. Control treatments were performed in parallel, with precisely the
1008 same experimental conditions, by feeding animals on lawns of the same strain of
1009 HT115 bacteria containing an empty vector.

1010 **Statistical analysis**

1011 Statistical analysis for comparisons between datasets that is not covered in the above
1012 paragraphs was performed using GraphPad prism 7 (www.graphpad.com). To test
1013 differences in the mean between seam cell scoring or smFISH counting datasets an
1014 unpaired two-tailed t-test was performed when the comparison was between two
1015 datasets and a one-way Analysis of Variance (ANOVA) was performed when multiple
1016 datasets were compared. One-way ANOVA was followed by a Dunnet's post hoc test
1017 when the mean of multiple datasets was compared to that of a control. The significance
1018 threshold used throughout is $p<0.05$.

1019 **Data availability**

1020 All raw sequence files and processed signal files have been deposited in the National
1021 Center for Biotechnology Information Gene Expression Omnibus (accession number:
1022 GSE163243).
1023
1024 **Supplementary Tables**
1025 **Table S1.** Genome coordinates of significant TaDa peaks for NHR-25 and LIN-22 at
1026 the L2 and L4 stage.
1027 **Table S2.** Assigned genes per TaDa peak for NHR-25 and LIN-22 at the L2 and L4
1028 stage.
1029 **Table S3.** Strains used in this study.
1030 **Table S4.** Primers used in this study (general oligos and smFISH probes).
1031 **Table S5.** List of transgenes and injection mixes used in this study.
1032
1033 **Supplemental Figure Legends**
1034 **Figure S1: Assembly and assessment of *in vivo* methylation by TaDa fusions.**
1035 **(A)** Graphic illustration of the design of a versatile *C. elegans* universal TaDa cloning
1036 platform with its key features. From left to right the plasmid contains: universal MosSCI
1037 recombination sites (ttTi5605_R and L), an LR attR4-attL1 Gateway cloning site for
1038 promoter insertion, an *mCherry* primary ORF followed by 2 x STOP codons, indicated
1039 in red, a frameshift in yellow and a unique XmaJI restriction site followed by *dam*,
1040 unique restriction cites for PaeI, PacI, NheI and an in-frame STOP codon prior to an
1041 *unc-54 3'UTR*. **(B)** Example of amplification products from methylated gDNA extracted
1042 from strains carrying the *dam* fusions described in this study and a WT dam^{-} strain
1043 fed on dam^{-} *E. coli.* Note that all samples except for the *dam*$^{-}$ strain show a
1044 pronounced 2 kb to 200 bp smear. Wells for all samples were loaded with the same

1045 volume of reaction product. Also note the marked increase in methylation observed in
1046 the absence of *mCherry* as a primary ORF. Strains grown on *dam*$^{+}$ OP50 also show
1047 extensive amplification due to methylated bacterial DNA.
1048
1049 **Figure S2: Validation of expression and functionality of TaDa transgenes.**
1050 **(A)** Confirmation of single-copy transgene expression in the *wrt-2* domain (seam cells)
1051 using *mCherry* expression (primary ORF). Animals were imaged at the L4 stage and
1052 white arrows mark weak expression in the seam, as expected for single copy
1053 transgenes. **(B)** Overexpression of the *lin-22:dam* construct as a multi-copy transgene
1054 (*icbEx54*) in the *lin-22(icb38)* mutant partially rescues the supernumerary PDE
1055 neurons phenotype labelled by *dat-1p::GFP* (n$\geq$31). **(C)** Overexpression of the *nhr-*
1056 *25:dam* fusion as a multi-copy transgene (*icbEx184*) in the *nhr-25(ku217)* mutant
1057 significantly increased the mean seam cell number (n$\geq$31). Scale bars in A are 20 µm.
1058 In B and C error bars indicate standard error of the mean and black stars indicate
1059 statistically significant differences in the mean with a t-test (* p<0.05, **** p<0.0001).
1060
1061 **Figure S3: TaDa replicate reproducibility and fusion-dependent methylation**.
1062 **(A-B)** Pearson correlation heatmaps based on normalised aligned read count maps
1063 for the LIN-22 (A) and NHR-25 (B) TaDa experiments. The correlation coefficient for
1064 each pairwise comparison is printed in each cell of the heatmaps. All Dam-fusions
1065 show strong replicate reproducibility and high within-fusion correlation, indicated by
1066 high correlation coefficients. Note that TF and control samples show low correlation
1067 between them and cluster separately. **(C)** Summary heatmap of Pearson correlations
1068 between all samples for the TF TaDa performed in this study. Note the high correlation
1069 between control samples from different experiments and low correlation with distinct

1070 separate clustering between the *lin-22:dam* and *nhr-25:dam* samples. Samples
1071 belonging to the LIN-22 or NHR-25 experiment are indicated with purple or blue font
1072 respectively. **(D)** Principal component analysis on normalised aligned read count maps
1073 for all samples shows distinct grouping between TF and control fusions. **(E)** Principal
1074 component analysis only on *lin-22:dam* and *nhr-25:dam* replicates shows distinct
1075 grouping between the two TF samples. Samples in D and E are colour-coded as per
1076 the shared key in the middle. **(F)** Example of averaged signal enrichment profiles for
1077 *lin-22:dam* and *nhr-25:dam* fusions across a 1 Mb region of chromosome I (locations
1078 of protein coding genes are indicated as black lines / boxes. Signal enrichment forms
1079 distinct peaks for each TF that show some similarity between stages. The Y-axes
1080 represent $\log_2$*(TF:dam/NLS-GFP:dam)* scores. Scale bar length is 2 Mb.
1081
1082 **Figure S4: NHR-25 and LIN-22 binding motifs identified using TaDa peaks.**
1083 **(A)** *De novo* identified DNA motifs for NHR-25 (left) and LIN-22 (right) binding from
1084 TaDa peaks. These motifs were found in 41% of the LIN-22 L2, 33% of the LIN-22 L4,
1085 47% of the NHR-25 L2 and 51% of the NHR-25 L4 total TaDa peaks, with the co-
1086 occurrence being statistically significant using Monte Carlo simulation (*p*-values <
1087 0.0001 in all cases)**. (B)** Similar known motifs to the TaDa-identified NHR-25 (left) and
1088 LIN-22 (right) motifs are shown from available databases. **(C-D)** Aggregation plots of
1089 TaDa signal for all TFs and available stages over instances of the NHR-25 TaDa motif
1090 (C) and the LIN-22 TaDa motif (D). Both TFs show stronger preference for their
1091 respective motif compared to the alternative motif. Regions of ±5 kb around the peak
1092 centres have been plotted, the y-axes represent mean enrichment scores for the
1093 plotted sequences and shaded areas represent 95% confidence intervals.
1094

1095 **Figure S5: TaDa identified targets for NHR-25 and LIN-22 show significant**
1096 **overlaps between them and with other published datasets.**
1097 **(A)** Barplot showing the size of all possible intersections between: *nhr-25:dam*
1098 identified target genes at L2/L4, NHR-25 ChIP-seq-identified target genes at L1 (Shao
1099 et al., 2013) and NHR-25 ChIP-seq peaks at L2 (Araya et al., 2014) assigned to genes
1100 using the same method used in this study. Selected enriched GO terms for gene sets
1101 from the large intersections between TaDa and ChIP-seq L2 are shown. **(B-C)** Tissue-
1102 enrichment analysis in the gene set of putative targets that are common between
1103 NHR-25 L2 ChIP-seq and TaDa (B) or exclusively identified by ChIP-seq (C). In both
1104 cases, the top 15 most overrepresented tissues are shown. **(D)** Barplot showing the
1105 size of all possible intersections between: the identified target genes for *lin-22:dam* L2
1106 and L4, genes predicted to be downstream genetic interactors of *lin-22* (Zhong and
1107 Sternberg, 2006) and the *C. elegans* orthologues of HES1 targets based on ChIP-seq
1108 (Encode project accession: ENCSR109ODF). Selected enriched GO terms for gene
1109 sets from the intersections between orthologues of HES1 ChIP-seq targets and TaDa
1110 targets are shown. Genes in the pairwise intersection between TaDa and predicted
1111 interactors are listed in full above the bars. **(E)** Plot showing all possible intersections
1112 between TaDa-identified target genes for LIN-22 and NHR-25 at L2 and L4. Selected
1113 enriched GO terms are shown for the genes common in all datasets. In A, D the size
1114 of each individual gene set is printed at the bottom right of each graph. In A, D and E
1115 the statistical significance of each intersection assessed by a Fisher's exact test is
1116 colour-coded as shown in the key (yellow, orange, red marks significant overlaps).
1117
1118 **Figure S6: TaDa reveals a link between LIN-22 and multiple components of the**
1119 **Wnt signaling pathway.**

1120 Signal profiles from *lin-22:dam* showing enrichment forming statistically significant
1121 peaks (shaded regions) on sequences associating with the Wnt component-encoding
1122 genes *mom-1, mig-14, cwn-2, lit-1, bar-1* and *mom-5.* Y-axes are $\log_2$*(lin-22:dam*/*NLS-*
1123 *GFP:dam)* scores with data range: -1 – 3.5.
1124
1125 **References:**

1126 Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW. 2016. Tissue enrichment
1127 analysis for *C. elegans* genomics. *BMC Bioinformatics* **17**:366.
1128 doi:10.1186/s12859-016-1229-9
1129 Antebi A. 2006. Nuclear hormone receptors in *C. elegans*. *WormBook* 1–13.
1130 doi:10.1895/wormbook.1.64.1
1131 Araya CL, Kawli T, Kundaje A, Jiang L, Wu B, Vafeados D, Terrell R, Weissdepp P,
1132 Gevirtzman L, MacE D, Niu W, Boyle AP, Xie D, Ma L, Murray JI, Reinke V,
1133 Waterston RH, Snyder M. 2014. Regulatory analysis of the *C. elegans* genome
1134 with spatiotemporal resolution. *Nature* **512**:400–405. doi:10.1038/nature13497
1135 Askjaer P, Ercan S, Meister P. 2014. Modern techniques for the analysis of
1136 chromatin and nuclear organization in *C. elegans*. *WormBook* 1–35.
1137 doi:10.1895/wormbook.1.169.1
1138 Barkoulas M, van Zon JS, Milloz J, van Oudenaarden A, Félix MA. 2013.
1139 Robustness and Epistasis in the *C. elegans* Vulval Signaling Network Revealed
1140 by Pathway Dosage Modulation. *Dev Cell* **24**:64–75.
1141 doi:10.1016/j.devcel.2012.12.001
1142 Barkoulas M, Vargas Velazquez AM, Peluffo AE, Félix MA. 2016. Evolution of New
1143 cis-Regulatory Motifs Required for Cell-Specific Gene Expression in
1144 *Caenorhabditis*. *PLoS Genet* **12**:1–23. doi:10.1371/journal.pgen.1006278

1145 Brabin C, Appleford PJ, Woollard A. 2011. The *Caenorhabditis elegans* gata factor
1146 elt-1 works through the cell proliferation regulator *bro-1* and the fusogen *eff-1* to
1147 maintain the seam stem-like fate. *PLoS Genet*.
1148 doi:10.1371/journal.pgen.1002200
1149 Buck SH, Chiu D, Saito RM. 2009. The cyclin-dependent kinase inhibitors, *cki-1* and
1150 *cki-2*, act in overlapping but distinct pathways to control cell cycle quiescence
1151 during *C. elegans* development. *Cell Cycle* **8**:2613–2620.
1152 doi:10.4161/cc.8.16.9354
1153 Cameron ER, Neil JC. 2004. The Runx genes: Lineage-specific oncogenes and
1154 tumor suppressors. *Oncogene* **23**:4308–4314. doi:10.1038/sj.onc.1207130
1155 Cassata G, Shemer G, Morandi P, Donhauser R, Podbilewicz B, Baumeister R.
1156 2005. ceh-16/engrailed patterns the embryonic epidermis of *Caenorhabditis*
1157 *elegans*. *Development* **132**:739–749. doi:10.1242/dev.01638
1158 Celniker SE, Dillon LAL, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis
1159 M, Lai EC, Lieb JD, MacAlpine DM, Micklem G, Piano F, Snyder M, Stein L,
1160 White KP, Waterston RH. 2009. Unlocking the secrets of the genome. *Nature*
1161 **459**:927–930. doi:10.1038/459927a
1162 Cheetham SW, Gruhn WH, Ameele J Van Den, Krautz R, Southall TD. 2018.
1163 Targeted DamID reveals differential binding of mammalian pluripotency factors.
1164 *Development*. doi:10.1242/dev.170209
1165 Chen Z, Eastburn DJ, Han M. 2004. The *Caenorhabditis elegans* Nuclear Receptor
1166 Gene *nhr-25* Regulates Epidermal Cell Development. *Mol Cell Biol* **24**:7345–
1167 7358. doi:10.1128/mcb.24.17.7345-7358.2004
1168 Chikina MD, Troyanskaya OG. 2012. An effective statistical evaluation of chipseq
1169 dataset similarity. *Bioinformatics* **28**:607–613. doi:10.1093/bioinformatics/bts009

1170 Chisholm AD, Hsiao TI. 2012. The *Caenorhabditis elegans* epidermis as a model
1171 skin. I: Development, patterning, and growth. *Wiley Interdiscip Rev Dev Biol*.
1172 doi:10.1002/wdev.79
1173 Crooks G, Hon G, Chandonia J, Brenner S. 2004. WebLogo: a sequence logo
1174 generator. *Genome Res* **14**:1188–1190. doi:10.1101/gr.849004.1
1175 David Angeles-Albores, Raymond Y. N. Lee JC and PWS (2017). n.d. Phenotype
1176 and gene ontology enrichment as guides for disease modeling in C. elegans.
1177 *Manuscr Prep*.
1178 Davidson EH. 2010. Emerging properties of animal gene regulatory networks. *Nature*
1179 **468**:911–920. doi:10.1038/nature09645
1180 Ding SS, Woollard A. 2017. Non-muscle myosin II is required for correct fate
1181 specification in the *Caenorhabditis elegans* seam cell divisions. *Sci Rep* **7**:3524.
1182 doi:10.1038/s41598-017-01675-7
1183 Ferré Q, Charbonnier G, Sadouni N, Lopez F, Kermezli Y, Spicuglia S, Capponi C,
1184 Ghattas B, Puthier D. 2019. OLOGRAM: determining significance of total
1185 overlap length between genomic regions sets. *Bioinformatics* **36**:1920–1922.
1186 doi:10.1093/bioinformatics/btz810
1187 Fornes O, Castro-Mondragon JA, Khan A, Van Der Lee R, Zhang X, Richmond PA,
1188 Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G,
1189 Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW,
1190 Mathelier A. 2020. JASPAR 2020: Update of the open-Access database of
1191 transcription factor binding profiles. *Nucleic Acids Res* **48**:D87–D92.
1192 doi:10.1093/nar/gkz1001
1193 Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA:
1194 Computational tools for comparative genomics. *Nucleic Acids Res* **32**:273–279.

1195 doi:10.1093/nar/gkh458
1196 Frokjaer-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella M,
1197 Pozniakovski A, Moerman DG, Jorgensen EM. 2014. Random and targeted
1198 transgene insertion in C. elegans using a modified Mosl transposon. *Nat*
1199 *Methods* **11**:529–534. doi:10.1038/nmeth.2889.Random
1200 Fuxman Bass JI, Tamburino AM, Mori A, Beittel N, Weirauch MT, Reece-Hoyes JS,
1201 Walhout AJM. 2014. Transcription factor binding to *Caenorhabditis elegans* first
1202 introns reveals lack of redundancy with gene promoters. *Nucleic Acids Res*
1203 **42**:153–162. doi:10.1093/nar/gkt858
1204 Gilleard JS, Mcghee JD. 2001. Activation of Hypodermal Differentiation in the
1205 *Caenorhabditis elegans* Embryo by GATA Transcription Factors ELT-1 and
1206 ELT-3. *Mol Cell Biol* **Vol. 21**:2533–2544. doi:10.1128/MCB.21.7.2533
1207 Gissendanner CR, Sluder AE. 2000. *nhr-25*, the *Caenorhabditis elegans* ortholog of
1208 ftz-f1, is required for epidermal and somatic gonad development. *Dev Biol*
1209 **221**:259–272. doi:10.1006/dbio.2000.9679
1210 Gleason JE, Eisenmann DM. 2010. Wnt signaling controls the stem cell-like
1211 asymmetric division of the epithelial seam cells during *C. elegans* larval
1212 development. *Dev Biol*. doi:10.1016/j.ydbio.2010.09.005
1213 Gómez-Saldivar G, Osuna-Luque J, Semple JI, Glauser DA, Jarriault S, Meister P.
1214 2020. Tissue-Specific Transcription Footprinting Using RNA Pol DamID (RAPID)
1215 in *Caenorhabditis elegans*. *Genetics* genetics.303774.2020.
1216 doi:10.1534/genetics.120.303774
1217 Gorrepati L, Thompson KW, Eisenmann DM. 2013. *C. elegans* GATA factors EGL-
1218 18 and ELT-6 function downstream of Wnt signaling to maintain the progenitor
1219 fate during larval asymmetric divisions of the seam cells. *Development*

1220 **140**:2093–102. doi:10.1242/dev.091124
1221 Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity
1222 between motifs. *Genome Biol* **8**. doi:10.1186/gb-2007-8-2-r24
1223 Hajduskova M, Jindra M, Herman MA, Asahina M. 2009. The nuclear receptor NHR-
1224 25 cooperates with the Wnt/β-catenin asymmetry pathway to control
1225 differentiation of the T seam cell in *C. elegans*. *J Cell Sci* **122**:3051–3060.
1226 doi:10.1242/jcs.052373
1227 Harr JC, Schmid CD, Muñoz-Jiménez C, Romero-Bueno R, Kalck V, Gonzalez-
1228 Sandoval A, Hauer MH, Padeken J, Askjaer P, Mattout A, Gasser SM. 2020.
1229 Loss of an H3K9me anchor rescues laminopathy-linked changes in nuclear
1230 organization and muscle function in an Emery-Dreifuss muscular dystrophy
1231 model. *Genes Dev* **34**:560–579. doi:10.1101/gad.332213.119
1232 Hayes GD, Frand AR, Ruvkun G. 2006. The *mir-84* and *let-7* paralogous microRNA
1233 genes of *Caenorhabditis elegans* direct the cessation of molting via the
1234 conserved nuclear hormone receptors NHR-23 and NHR-25. *Development*
1235 **133**:4631–4641. doi:10.1242/dev.02655
1236 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C,
1237 Singh H, Glass CK. 2010. Simple combinations of lineage-determining
1238 transcription factors prime cis-regulatory elements required for macrophage and
1239 B cell identities. *Mol Cell* **38**:576–589. doi:10.1016/j.molcel.2010.05.004
1240 Hong Y, Roy R, Ambros V. 1998. Developmental regulation of a cyclin-dependent
1241 kinase inhibitor controls postembryonic cell cycle progression in *Caenorhabditis*
1242 *elegans*. *Development* **125**:3585–3597.
1243 Hughes S, Brabin C, Appleford PJ, Woollard A. 2013. CEH-20/Pbx and UNC-
1244 62/Meis function upstream of *rnt-1*/Runx to regulate asymmetric divisions of the

1245 *C. elegans* stem-like seam cells. *Biol Open* **2**:718–27.
1246 Jänes J, Dong Y, Schoof M, Serizay J, Appert A, Cerrato C, Woodbury C, Chen R,
1247 Gemma C, Huang N, Kissiov D, Stempor P, Steward A, Zeiser E, Sauer S,
1248 Ahringer J. 2018. Chromatin accessibility dynamics across *C. elegans*
1249 development and ageing. *Elife* **7**:267–272. doi:10.7554/eLife.37344
1250 Joshi PM, Riddle MR, Djabrayan NJ V, Rothman JH. 2010. *Caenorhabditis elegans*
1251 as a model for stem cell biology. *Dev Dyn* **239**:1539–1554.
1252 doi:10.1002/dvdy.22296
1253 Kageyama R, Ohtsuka T, Kobayashi T. 2007. The Hes gene family: Repressors and
1254 oscillators that orchestrate embryogenesis. *Development* **134**:1243–1251.
1255 doi:10.1242/dev.000786
1256 Kagoshima H, Nimmo R, Saad N, Tanaka J, Miwa Y, Mitani S, Kohara Y, Woollard
1257 A. 2007. The *C. elegans* CBF homologue BRO-1 interacts with the Runx factor,
1258 RNT-1, to promote stem cell proliferation and self-renewal. *Development*
1259 **134**:3905–3915. doi:10.1242/dev.008276
1260 Katsanos D, Koneru SL, Mestek Boukhibar L, Gritti N, Ghose R, Appleford PJ,
1261 Doitsidou M, Woollard A, van Zon JS, Poole RJ, Barkoulas M. 2017. Stochastic
1262 loss and gain of symmetric divisions in the *C. elegans* epidermis perturbs
1263 robustness of stem cell number. *PLoS Biol* **15**:1–31.
1264 doi:10.1371/journal.pgen.1006278
1265 Koh K, Rothman JH. 2001. ELT-5 and ELT-6 are required continuously to regulate
1266 epidermal seam cell differentiation and cell fusion in *C. elegans*. *Development*
1267 **128**:2867–2880.
1268 Kondili M, Fust A, Preussner J, Kuenne C, Braun T, Looso M. 2017. UROPA: A tool
1269 for Universal RObust Peak Annotation. *Sci Rep* **7**:1–12. doi:10.1038/s41598-

1270 017-02464-y
1271 Kozak M. 2001. Constraints on reinitiation of translation in mammals. *Nucleic Acids*
1272 *Res* **29**:5226–5232. doi:10.1093/nar/29.24.5226
1273 Kozak M. 1987. Effects of intercistronic length on the efficiency of reinitiation by
1274 eucaryotic ribosomes. *Mol Cell Biol* **7**:3438–45.
1275 doi:10.1128/MCB.7.10.3438.Updated
1276 Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, Kirkey
1277 M, Hammonds AS, Gersch J, Ammouri H, Wall ML, Moran J, Steffen D,
1278 Szynkarek M, Seabrook-Sturgis S, Jameel N, Kadaba M, Patton J, Terrell R,
1279 Corson M, Durham TJ, Park S, Samanta S, Han M, Xu J, Yan KK, Celniker SE,
1280 White KP, Ma L, Gerstein M, Reinke V, Waterston RH. 2018. The modern
1281 resource: genome-wide binding profiles for hundreds of *Drosophila* and
1282 *Caenorhabditis elegans* transcription factors. *Genetics* **208**:937–949.
1283 doi:10.1534/genetics.117.300657
1284 Lam AK, Phillips BT. 2017. Wnt signaling polarizes *C. elegans* asymmetric cell
1285 divisions during developmentResults and Problems in Cell Differentiation. pp.
1286 83–114. doi:10.1007/978-3-319-53150-2_4
1287 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*
1288 *Methods* **9**:357–359. doi:10.1038/nmeth.1923
1289 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G,
1290 Durbin R. 2009. The Sequence Alignment/Map format and SAMtools.
1291 *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352
1292 Lichtenberg KH de, Funa N, Nakic N, Ferrer J, Zhu Z, Huangfu D, Serup P. 2018.
1293 Genome-Wide Identification of HES1 Target Genes Uncover Novel Roles for
1294 HES1 in Pancreatic Development. *bioRxiv* 335869. doi:10.1101/335869

1295 Marshall OJ, Brand AH. 2015. Damidseq-pipeline: An automated pipeline for
1296 processing DamID sequencing datasets. *Bioinformatics* **31**:3371–3373.
1297 doi:10.1093/bioinformatics/btv386
1298 Marshall OJ, Southall TD, Cheetham SW, Brand AH. 2016. Cell-type-specific
1299 profiling of protein-DNA interactions without cell isolation using targeted DamID
1300 with next-generation sequencing. *Nat Protoc* **11**:1586–98.
1301 doi:10.1038/nprot.2016.084
1302 Matus DQ, Lohmer LL, Kelley LC, Schindler AJ, Kohrman AQ, Barkoulas M, Zhang
1303 W, Chi Q, Sherwood DR. 2015. Invasive Cell Fate Requires G1 Cell-Cycle
1304 Arrest and Histone Deacetylase-Mediated Changes in Gene Expression. *Dev*
1305 *Cell* **35**:162–174. doi:10.1016/j.devcel.2015.10.002
1306 Miyabayashi T, Palfreyman MT, Sluder a E, Slack F, Sengupta P. 1999. Expression
1307 and function of members of a divergent nuclear receptor family in
1308 *Caenorhabditis elegans*. *Dev Biol* **215**:314–331. doi:10.1006/dbio.1999.9470
1309 Murata K, Hattori M, Hirai N, Shinozuka Y, Hirata H, Kageyama R, Sakai T, Minato
1310 N. 2005. Hes1 directly controls cell proliferation through the transcriptional
1311 repression of p27Kip1. *Mol Cell Biol* **25**:4262–4271.
1312 doi:10.1128/mcb.25.10.4262-4271.2005
1313 Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio
1314 R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, Van
1315 Helden J, Medina-Rivera A, Thomas-Chollier M. 2018. RSAT 2018: Regulatory
1316 sequence analysis tools 20th anniversary. *Nucleic Acids Res* **46**:W209–W214.
1317 doi:10.1093/nar/gky317
1318 Nimmo R, Antebi A, Woollard A. 2005. *mab-2* encodes RNT-1, a *C. elegans* Runx
1319 homologue essential for controlling cell proliferation in a stem cell-like

1320 developmental lineage. *Development* **132**:5043–5054. doi:10.1242/dev.02102
1321 O'Connell KF. 2010. Reverse chunking, a simple and effective method for identifying
1322 *unc-119(+)* transgenics. *Worm Breeder's Gaz* **Volume 18**.
1323 Oliveri P, Davidson EH. 2007. Built to run, not fail. *Science (80-)* **315**:1510–1511.
1324 doi:10.1126/science.1140979
1325 Pindyurin A V., Pagie L, Kozhevnikova EN, Van Arensbergen J, Van Steensel B.
1326 2016. Inducible DamID systems for genomic mapping of chromatin proteins in
1327 *Drosophila*. *Nucleic Acids Res* **44**:5646–5657. doi:10.1093/nar/gkw176
1328 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S,
1329 Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-
1330 sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165.
1331 doi:10.1093/nar/gkw257
1332 Reilly MB, Cros C, Varol E, Yemini E, Hobert O. 2020. Unique homeobox codes
1333 delineate all the neuron classes of C. elegans. *Nature* **584**:595–601.
1334 doi:10.1038/s41586-020-2618-9
1335 Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. 1992. Two
1336 mammalian helix-loop-helix factors structurally related to *Drosophila hairy* and
1337 *Enhancer of split*. *Genes Dev* **6**:2620–2634. doi:10.1101/gad.6.12b.2620
1338 Sawa H, Korswagen HC. 2013. Wnt signaling in *C. elegans*. *WormBook* 1–18.
1339 doi:10.1895/wormbook.1.
1340 Schlager B, Röseler W, Zheng M, Gutierrez A, Sommer RJ. 2006. HAIRY-like
1341 transcription factors and the evolution of the nematode vulva equivalence group.
1342 *Curr Biol* **16**:1386–1394. doi:10.1016/j.cub.2006.06.058
1343 Schuster E, McElwee JJ, Tullet JMA, Doonan R, Matthijssens F, Reece-Hoyes JS,
1344 Hope IA, Vanfleteren JR, Thornton JM, Gems D. 2010. DamID in *C. elegans*

1345 reveals longevity-associated targets of DAF-16/FoxO. *Mol Syst Biol* **6**:1–6.
1346 doi:10.1038/msb.2010.54
1347 Sen SQ, Chanchani S, Southall TD, Doe CQ. 2019. Neuroblast-specific chromatin
1348 landscapes allows the integration of spatial and temporal cues during
1349 *Drosophila* neurogenesis. *Elife* **8**:399–404. doi:10.7554/eLife.44036
1350 Serizay J, Dong Y, Jänes J, Chesney M, Cerrato C, Ahringer J. 2020. Distinctive
1351 regulatory architectures of germline-active and somatic genes in C. elegans.
1352 *Genome Res* **31**:1752–1765. doi:10.1101/gr.265934.120
1353 Shao J, He K, Wang H, Ho WS, Ren X, An X, Wong MK, Yan B, Xie D,
1354 Stamatoyannopoulos J, Zhao Z. 2013. Collaborative regulation of development
1355 but independent control of metabolism by two epidermis-specific transcription
1356 factors in *Caenorhabditis elegans*. *J Biol Chem* **288**:33411–33426.
1357 doi:10.1074/jbc.M113.487975
1358 Šilhánková M, Jindra M, Asahina M. 2005. Nuclear receptor NHR-25 is required for
1359 cell-shape dynamics during epidermal differentiation in *Caenorhabditis elegans*.
1360 *J Cell Sci* **118**:223–232. doi:10.1242/jcs.01609
1361 Smith JA, McGarr P, Gilleard JS. 2005. The *Caenorhabditis elegans* GATA factor *elt-*
1362 *1* is essential for differentiation and maintenance of hypodermal seam cells and
1363 for normal locomotion. *J Cell Sci* **118**:5709–5719. doi:10.1242/jcs.02678
1364 Southall T, Gold K, Egger B, Davidson C, Caygill E, Marshall O, Brand A. 2013. Cell-
1365 type-specific profiling of gene expression and chromatin binding without cell
1366 isolation: Assaying RNA pol II occupancy in neural stem cells. *Dev Cell*.
1367 doi:10.1016/j.devcel.2013.05.020
1368 Spieth J, Hillier LW, Wilson RK. 2005. Evolutionarily conserved elements in
1369 vertebrate , insect , worm , and yeast genomes. *Genome Res* 1034–1050.

1370 doi:10.1101/gr.3715005
1371 Spieth J, Lawson D, Davis P, Williams G, Howe K. 2014. Overview of gene structure
1372 in *C. elegans*. *WormBook* 1–18. doi:10.1895/wormbook.1.65.2
1373 Stempor P, Ahringer J. 2016. SeqPlots - Interactive software for exploratory data
1374 analyses, pattern discovery and visualization in genomics. *Wellcome Open Res*
1375 **1**:14. doi:10.12688/wellcomeopenres.10004.1
1376 Sulston JE, Horvitz HR. 1977. Post-embryonic cell lineages of the nematode,
1377 *Caenorhabditis elegans*. *Dev Biol* **56**:110–156. doi:10.1016/0012-
1378 1606(77)90158-0
1379 Thompson KW, Joshi P, Dymond JS, Gorrepati L, Smith HE, Krause MW,
1380 Eisenmann DM. 2016. The Paired-box protein PAX-3 regulates the choice
1381 between lateral and ventral epidermal cell fates in *C. elegans*. *Dev Biol*
1382 **412**:191–207. doi:10.1016/j.ydbio.2016.03.002
1383 van de Walle P, Muñoz-Jiménez C, Askjaer P, Schoofs L, Temmerman L. 2020.
1384 DamID identifies targets of CEH-60/PBX that are associated with neuron
1385 development and muscle structure in *Caenorhabditis elegans*. *PLoS One* **15**:1–
1386 19. doi:10.1371/journal.pone.0242939
1387 van der Horst SEM, Cravo J, Woollard A, Teapal J, van den Heuvel S. 2019. *C.*
1388 *elegans* Runx/CBFβ suppresses POP-1 TCF to convert asymmetric to
1389 proliferative division of stem cell-like seam cells. *Development* **146**:1–14.
1390 doi:10.1242/dev.180034
1391 van Steensel B, Delrow J, Henikoff S. 2001. Chromatin profiling using targeted DNA
1392 adenine methyltransferase. *Nat Genet* **27**:304–8. doi:10.1038/85871
1393 van Steensel B, Henikoff S. 2000. Identification of in vivo DNA targets of chromatin
1394 proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**:424–428.

1395 doi:10.1038/74487

1396 Vissers JHA, Froldi F, Schröder J, Papenfuss AT, Cheng LY, Harvey KF. 2018. The
1397 Scalloped and Nerfin-1 transcription factors cooperate to maintain neuronal cell
1398 fate. *Cell Rep* **25**:1561-1576.e7. doi:10.1016/j.celrep.2018.10.038

1399 Wang M, Zhao Y, Zhang B. 2015. Efficient Test and Visualization of Multi-Set
1400 Intersections. *Sci Rep* **5**:1–12. doi:10.1038/srep16923

1401 Wrischnik L a, Kenyon CJ. 1997. The role of *lin-22, a hairy/enhancer of split*
1402 homolog, in patterning the peripheral nervous system of *C. elegans*.
1403 *Development* **124**:2875–2888.

1404 Zheng R, Blobel GA. 2010. GATA transcription factors and cancer. *Genes and*
1405 *Cancer* **1**:1178–1188. doi:10.1177/1947601911404223

1406 Zhong W, Sternberg PW. 2006. Genome-wide prediction of *C. elegans* genetic
1407 interactions. *Science (80-)* **311**:1481–1484. doi:10.1126/science.1123287

1408

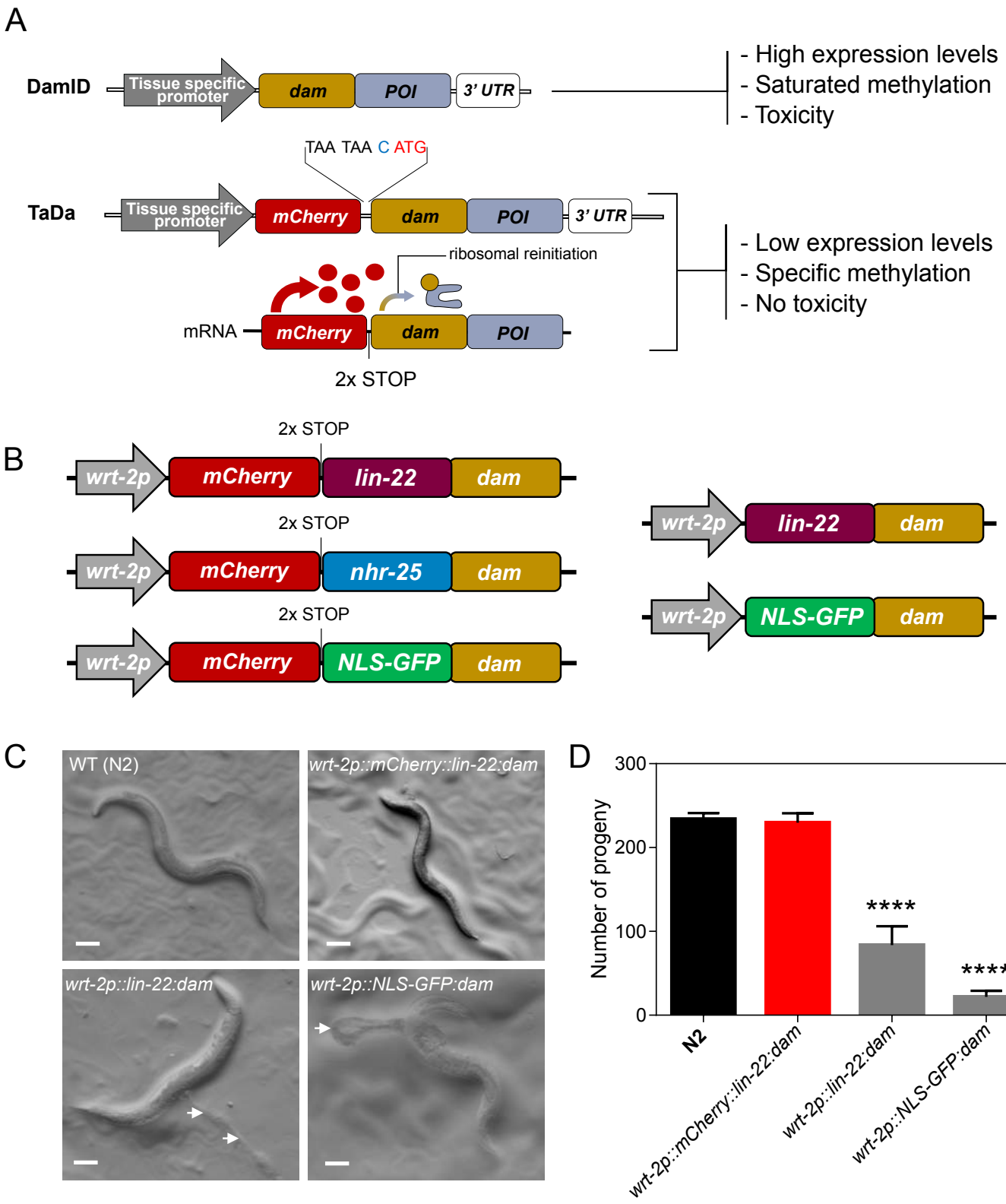
A
DamID
Tissue specific promoter
dam
POI
3' UTR
- High expression levels
- Saturated methylation
- Toxicity
TAA TAA C ATG
TaDa
Tissue specific promoter
mCherry
dam
POI
3' UTR
ribosomal reinitiation
mRNA
mCherry
dam
POI
2x STOP
- Low expression levels
- Specific methylation
- No toxicity
B
2x STOP
wrt-2p
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 1

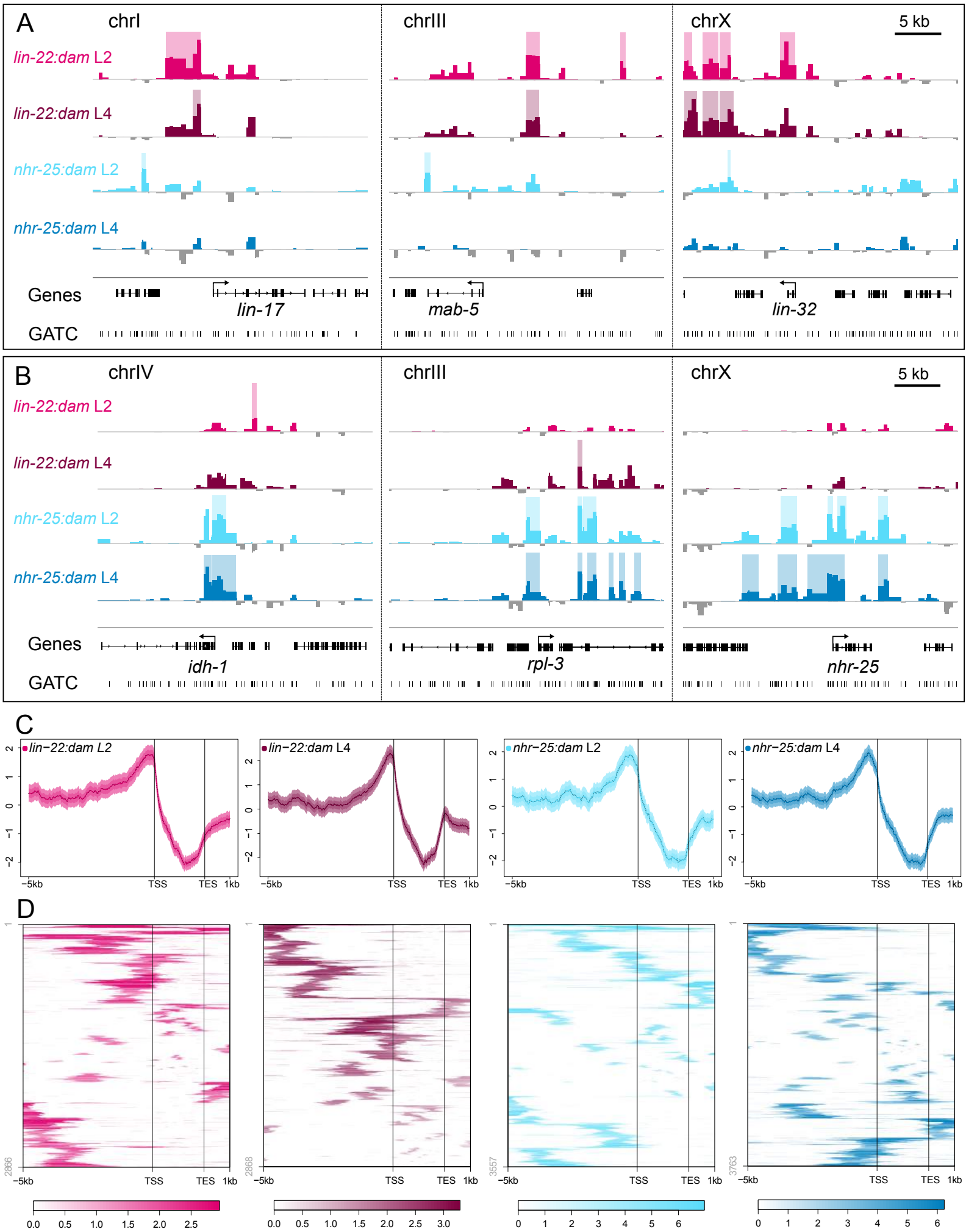
A
chrI
chrIII
chrX
5 kb
lin-22:dam L2
lin-22:dam L4
nhr-25:dam L2
nhr-25:dam L4
Genes
lin-17
mab-5
lin-32
GATC
B
chrIV
chrIII
chrX
5 kb
lin-22:dam L2
lin-22:dam L4
nhr-25:dam L2
nhr-25:dam L4
Genes
idh-1
rpl-3
nhr-25
GATC
C
lin−22:dam L2
lin−22:dam L4
nhr−25:dam L2
nhr−25:dam L4
−5kb
TSS
TES
1kb
D
2866
2868
3557
3763
0.0 0.5 1.0 1.5 2.0 2.5
0.0 0.5 1.0 1.5 2.0 2.5 3.0
0 1 2 3 4 5 6
0 1 2 3 4 5 6

Figure 2

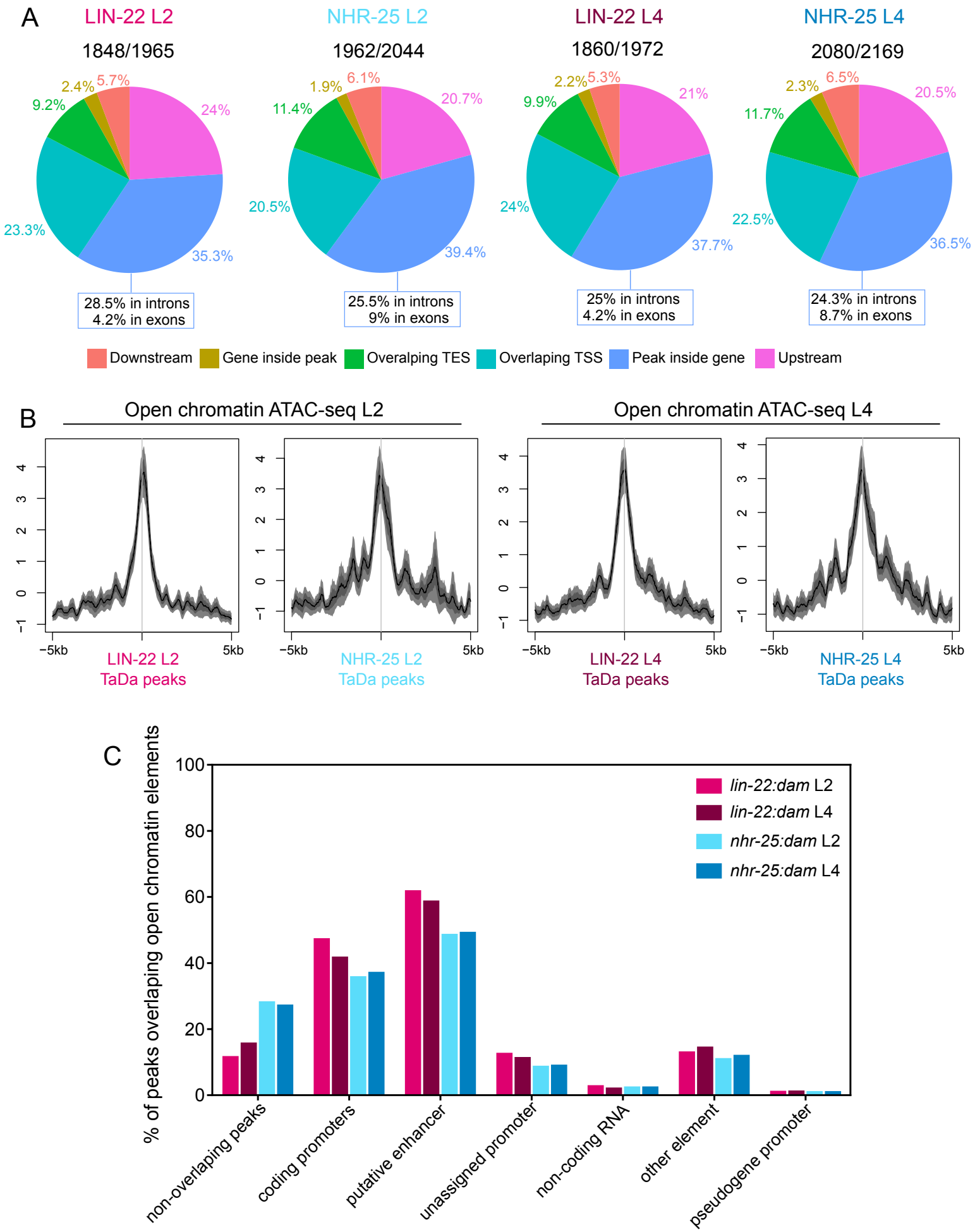
A
LIN-22 L2
1848/1965
2.4%
5.7%
9.2%
24%
23.3%
35.3%
28.5% in introns
4.2% in exons
NHR-25 L2
1962/2044
1.9%
6.1%
11.4%
20.7%
20.5%
39.4%
25.5% in introns
9% in exons
LIN-22 L4
1860/1972
2.2%
5.3%
9.9%
21%
24%
37.7%
25% in introns
4.2% in exons
NHR-25 L4
2080/2169
2.3%
6.5%
11.7%
20.5%
22.5%
36.5%
24.3% in introns
8.7% in exons
Downstream
Gene inside peak
Overalping TES
Overlaping TSS
Peak inside gene
Upstream
B
Open chromatin ATAC-seq L2
Open chromatin ATAC-seq L4
−5kb
5kb
LIN-22 L2
TaDa peaks
NHR-25 L2
TaDa peaks
LIN-22 L4
TaDa peaks
NHR-25 L4
TaDa peaks
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 3

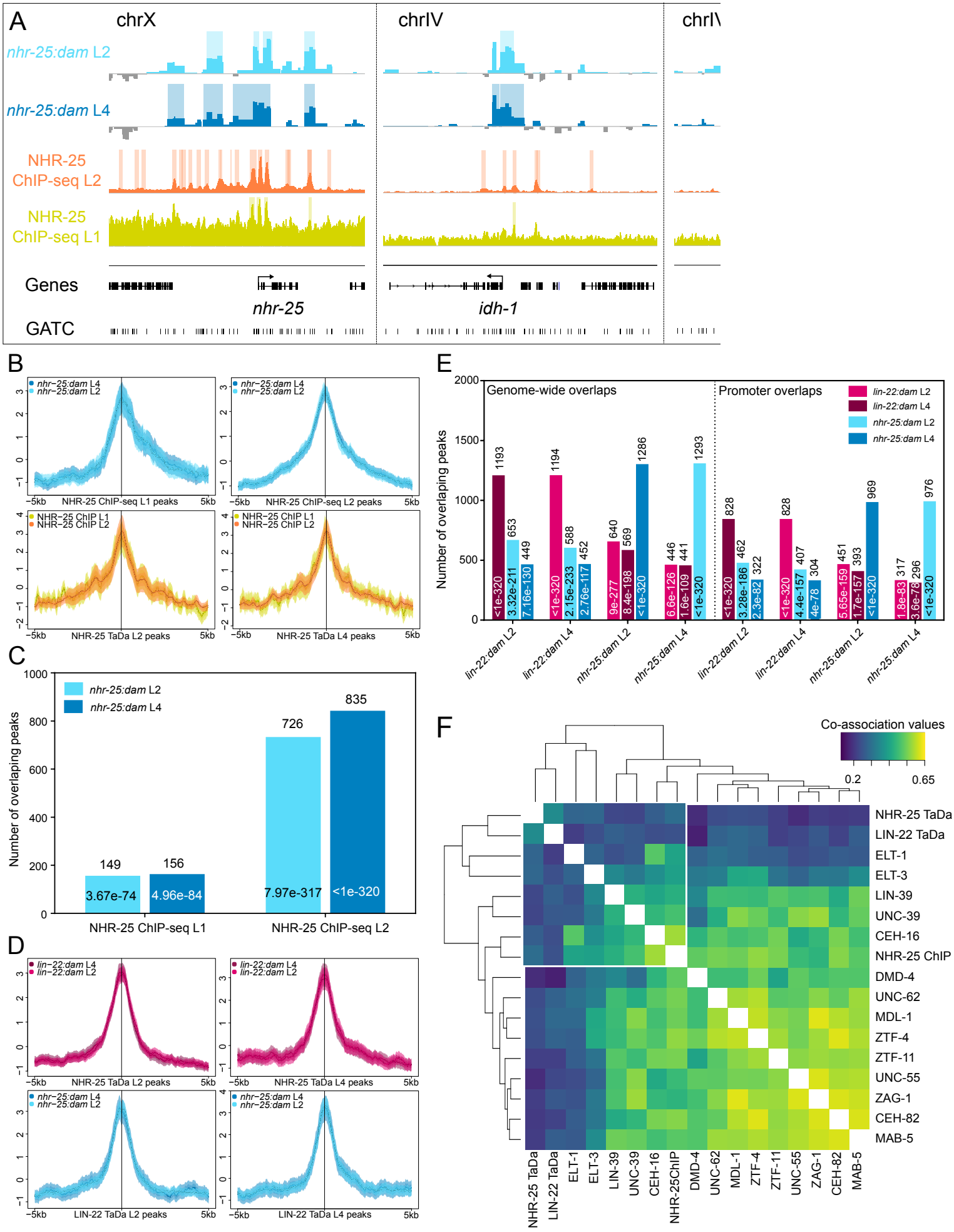
A
chrX
chrIV
nhr-25:dam L2
nhr-25:dam L4
NHR-25 ChIP-seq L2
NHR-25 ChIP-seq L1
Genes
nhr-25
idh-1
GATC
B
nhr−25:dam L4
nhr−25:dam L2
NHR−25 ChIP L1
NHR−25 ChIP L2
−5kb
5kb
NHR-25 ChIP-seq L1 peaks
NHR-25 ChIP-seq L2 peaks
NHR-25 TaDa L2 peaks
NHR-25 TaDa L4 peaks
C
nhr-25:dam L2
nhr-25:dam L4
Number 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
LIN-22 TaDa L2 peaks
LIN-22 TaDa L4 peaks
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
<1e-320
3.32e-211
7.16e-130
1194
588
452
<1e-320
2.15e-233
2.76e-117
640
569
1286
9e-277
8.4e-198
<1e-320
446
441
1293
6.6e-126
1.6e-109
<1e-320
828
462
322
<1e-320
3.28e-186
2.3e-82
828
407
304
<1e-320
4.4e-157
4e-78
451
393
969
5.65e-159
1.7e-157
<1e-320
317
296
976
1.8e-83
3.6e-76
<1e-320
F
Co-association values
0.2
0.65
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

Figure 4

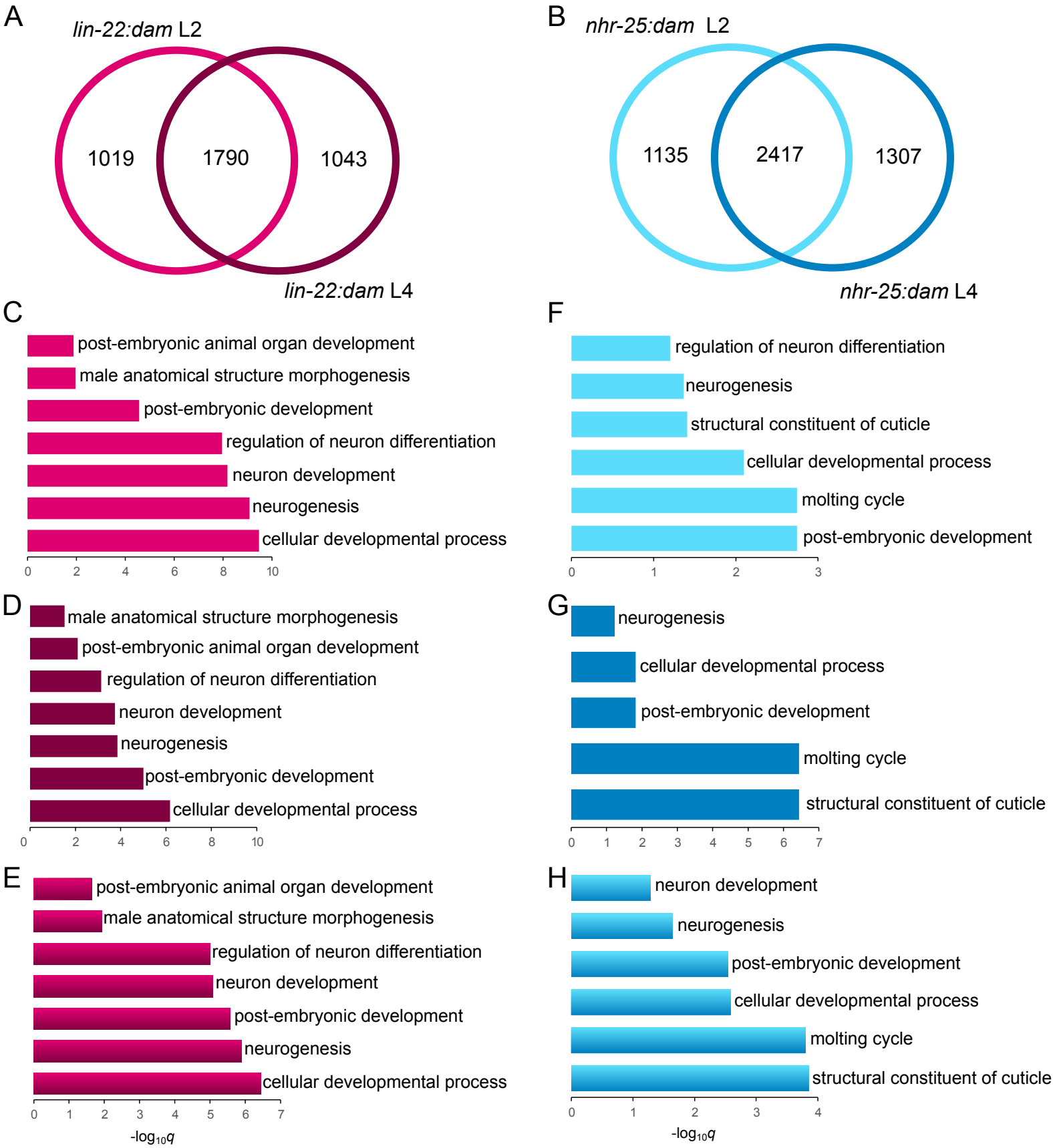
A
lin-22:dam L2
1019
1790
1043
lin-22:dam L4
B
nhr-25:dam L2
1135
2417
1307
nhr-25:dam L4
C
post-embryonic animal organ development
male anatomical structure morphogenesis
post-embryonic development
regulation of neuron differentiation
neuron development
neurogenesis
cellular developmental process
0 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 5

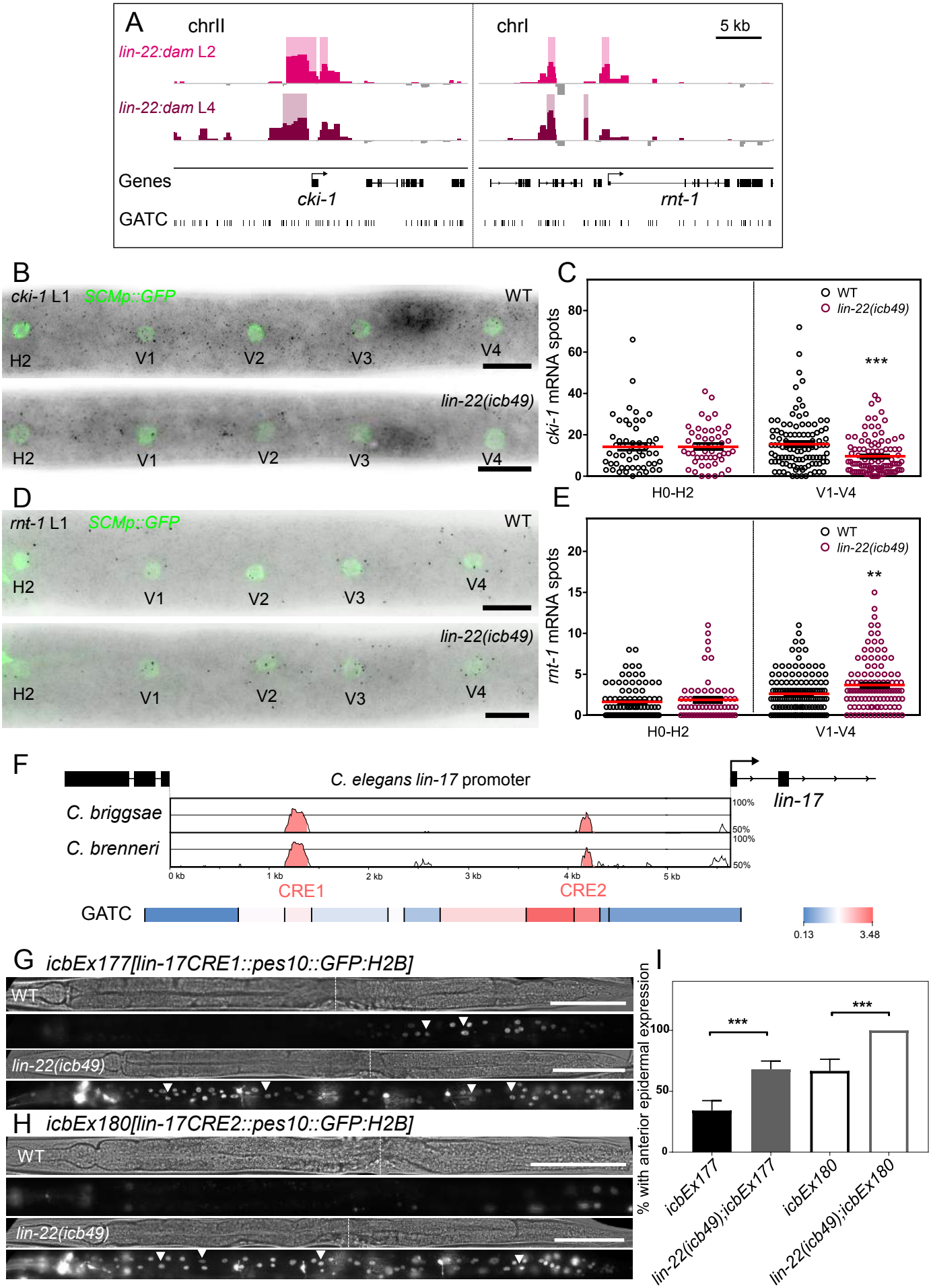
A
chrII
chrI
5 kb
lin-22:dam L2
lin-22:dam L4
Genes
cki-1
rnt-1
GATC
B
cki-1 L1
SCMp::GFP
WT
H2
V1
V2
V3
V4
lin-22(icb49)
C
cki-1 mRNA spots
WT
lin-22(icb49)
***
H0-H2
V1-V4
D
rnt-1 L1
SCMp::GFP
E
rnt-1 mRNA spots
**
F
C. elegans lin-17 promoter
lin-17
C. briggsae
C. brenneri
100%
50%
0 kb
1 kb
2 kb
3 kb
4 kb
5 kb
CRE1
CRE2
GATC
0.13
3.48
G icbEx177[lin-17CRE1::pes10::GFP:H2B]
H icbEx180[lin-17CRE2::pes10::GFP:H2B]
I
% with anterior epidermal expression
icbEx177
lin-22(icb49);icbEx177
icbEx180
lin-22(icb49);icbEx180

Figure 6

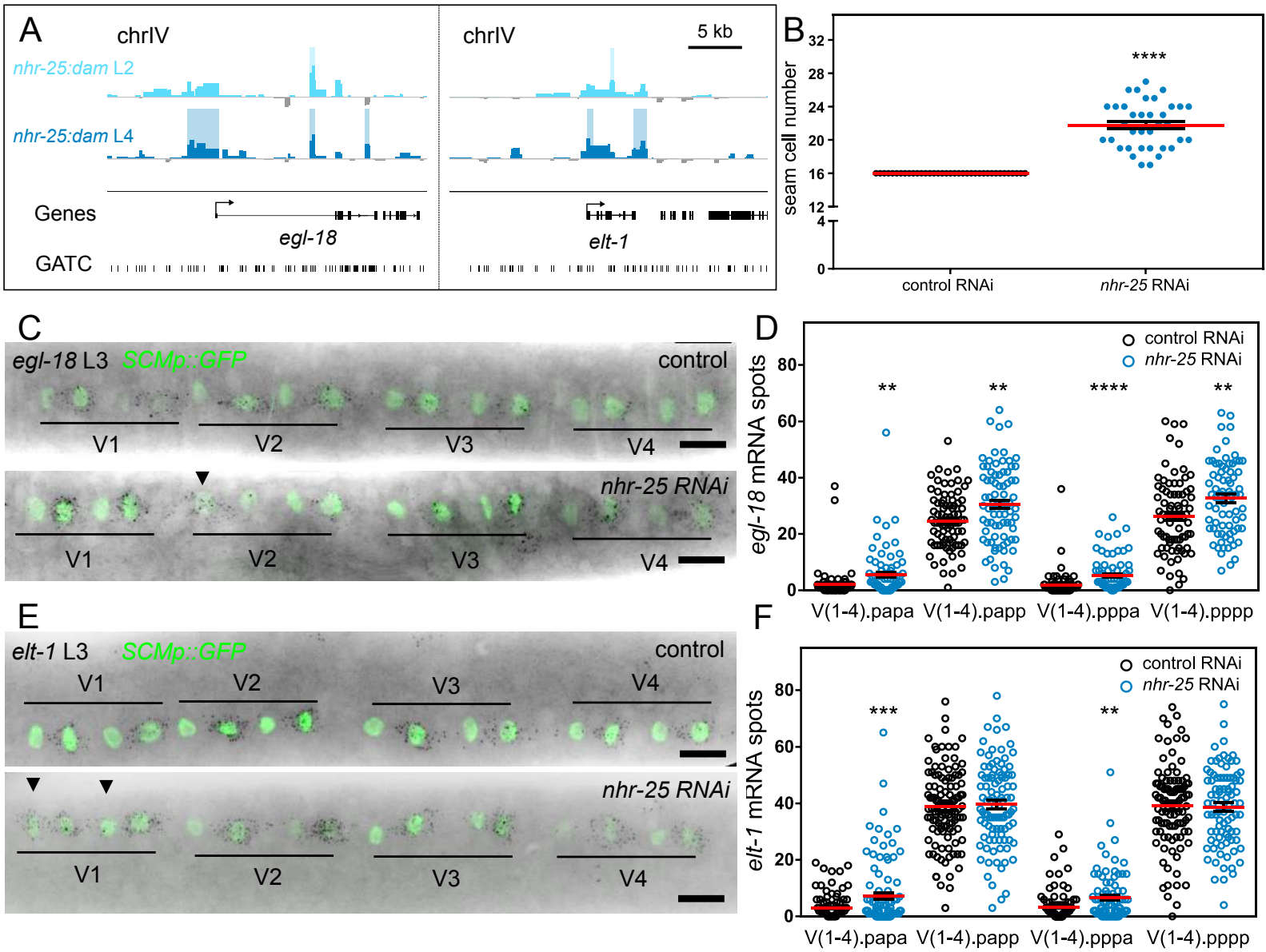
A
chrIV
nhr-25:dam L2
nhr-25:dam L4
Genes
egl-18
GATC
chrIV
5 kb
elt-1
B
seam cell number
****
control RNAi
nhr-25 RNAi
C
egl-18 L3 SCMp::GFP
control
V1
V2
V3
V4
nhr-25 RNAi
D
egl-18 mRNA spots
control RNAi
nhr-25 RNAi
**
**
****
**
V(1-4).papa
V(1-4).papp
V(1-4).pppa
V(1-4).pppp
E
elt-1 L3 SCMp::GFP
control
nhr-25 RNAi
F
elt-1 mRNA spots
control RNAi
nhr-25 RNAi
***
**

Figure 7

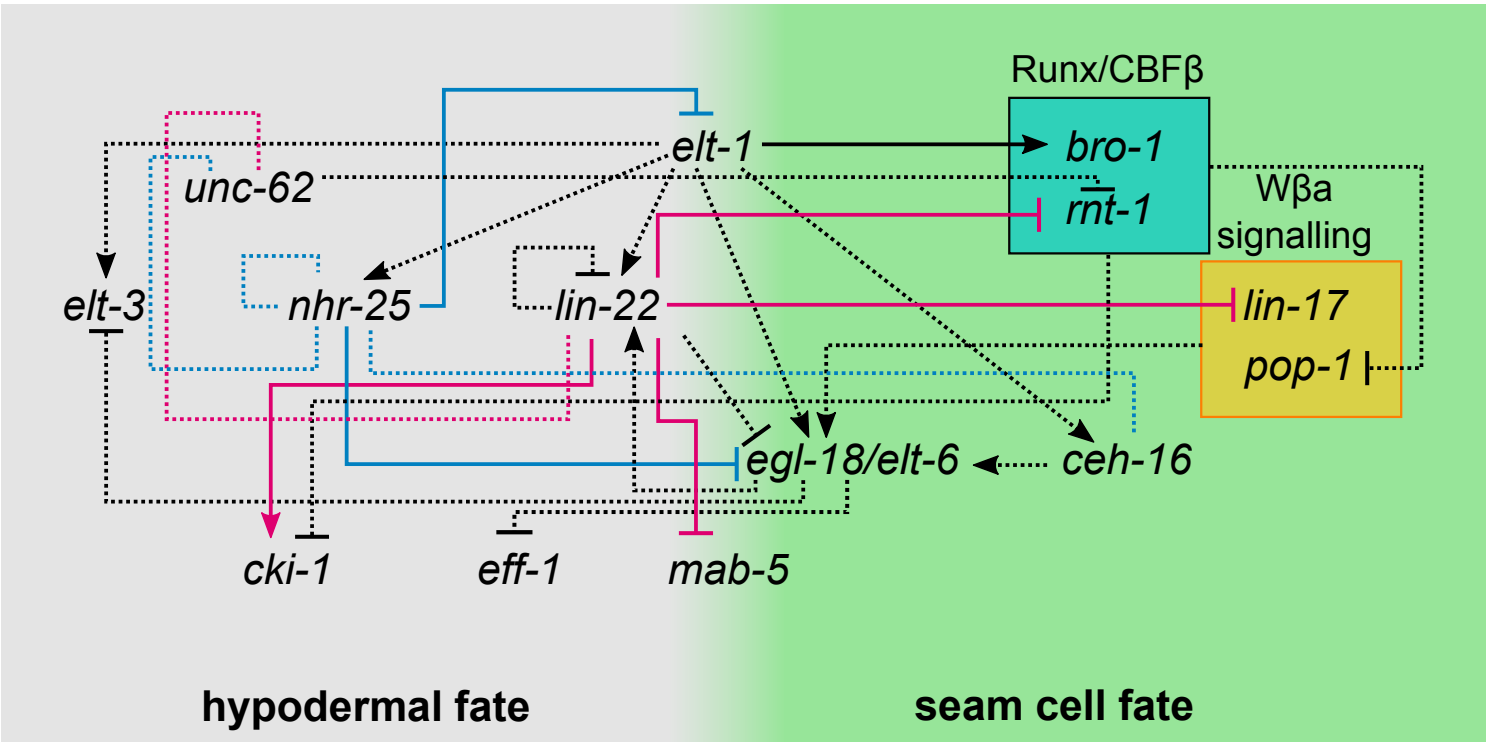
Runx/CBFβ
bro-1
rnt-1
elt-1
unc-62
Wβa
signalling
elt-3
nhr-25
lin-22
lin-17
pop-1
egl-18/elt-6
ceh-16
cki-1
eff-1
mab-5
hypodermal fate
seam cell fate

Figure 8