G3, 2022, 12(6), jkac091
https://doi.org/10.1093/g3journal/jkac091
Advance Access Publication Date: 12 May 2022
Investigation

Identifying the *Caenorhabditis elegans* vulval transcriptome

Qi Zhang [1,4], Heather Hrach [2,3], Marco Mangone [2,3] and David J. Reiner [1,*]

[1]Department of Translational Medical Science, Institute of Biosciences and Technology, Texas A&M Health Science Center, Texas A&M University, Houston, TX 77030, USA,
[2]Molecular and Cellular Biology Graduate Program, Arizona State University, Tempe, AZ 85281, USA,
[3]Virginia G. Piper Center for Personalized Diagnostics, The Biodesign Institute at Arizona State University, Tempe, AZ 85281, USA
[4]Present address: Department of Integrative Biology and Pharmacology, McGovern Medical School, The University of Texas Health Science Center at Houston, 6431 Fannin St, Houston, TX 77030, USA.

*Corresponding author: Department of Translational Medical Science, Institute of Biosciences and Technology, Texas A&M Health Science Center, Texas A&M University, Houston, TX 77030, USA. Email: dreiner@tamu.edu

Abstract

Development of the *Caenorhabditis elegans* vulva is a classic model of organogenesis. This system, which starts with 6 equipotent cells, encompasses diverse types of developmental event, including developmental competence, multiple signaling events to control precise and faithful patterning of three cell fates, execution and proliferation of specific cell lineages, and a series of sophisticated morphogenetic events. Early events have been subjected to extensive mutational and genetic investigations and later events to cell biological analyses. We infer the existence of dramatically changing profiles of gene expression that accompanies the observed changes in development. Yet, except from serendipitous discovery of several transcription factors expressed in dynamic patterns in vulval lineages, our knowledge of the transcriptomic landscape during vulval development is minimal. This study describes the composition of a vulva-specific transcriptome. We used tissue-specific harvesting of mRNAs via immunoprecipitation of epitope-tagged poly(A) binding protein, PAB-1, heterologously expressed by a promoter known to express GFP in vulval cells throughout their development. The identified transcriptome was small but tightly interconnected. From this data set, we identified several genes with identified functions in development of the vulva and validated more with promoter-GFP reporters of expression. For one target, *lag-1*, promoter-GFP expression was limited but a fluorescent tag of the endogenous protein revealed extensive expression. Thus, we have identified a transcriptome of *C. elegans* vulval lineages as a launching pad for exploration of functions of these genes in organogenesis.

Keywords: vulval transcriptome; *Caenorhabditis elegans*; *lin-31*; *lag-1*; *mbl-1*; *toe-1*; *shc-1*; *F27A3.4*

Introduction

Organogenesis involves an elaborate series of developmental events that encompass much of the spectrum of developmental biology. This process is presumed to be accompanied by multiple incidences of dynamic spatiotemporal changes in gene expression. Because of sequential developmental decisions, the mechanisms of many later steps can be masked by experimental perturbation of earlier steps, making developmental genetic analysis of the entirety challenging to analyze.

The *Caenorhabditis elegans* vulva is a classic system for the genetic investigation of organogenesis (Sternberg 2005). Thus far, most analysis has focused on the initial patterning of the 6 vulval precursor cells (VPCs). These roughly equipotent cells are located in an anterior-to-posterior line along the ventral midline of the animal (Fig. 1a). Vulval fates are induced in VPCs by an EGF-like signal emanating from the anchor cell (AC), in the nearby ventral gonad, to form a pattern of 3˚-3˚-2˚-1˚-2˚-3˚ cell fates, with EGFR and Notch receptors functioning centrally in the developmental process (Shin and Reiner 2018). Each cell type undergoes a stereotyped series of cell divisions and subsequent lineal cell behaviors. Understanding of mechanisms of VPC fate specification mostly ends at the level of transcription factors downstream of these EGFR and Notch signals. An exception to this trend is the chance identification of genes whose expression occurs in various sublineages of the vulva and defined a gene regulatory network of transcription factors (Inoue *et al.* 2002; Kirouac and Sternberg 2003; Inoue *et al.* 2005; Ririe *et al.* 2008). Despite the extensive study of initial cell fate patterning in the vulva, a mechanistic understanding of development after this point, including proliferation and morphogenesis, remains to be characterized (Sternberg 2005).

Our group has refined a method to isolate and sequence tissue-specific transcriptomes at high-resolution (Blazie *et al.* 2015, 2017). This method, which we named PAT-Seq, takes advantage of the ability of the *C. elegans* ortholog of poly(A)-binding protein, PAB-1, to bind poly(A) tails of mature mRNAs. In this method, heterologously expressed PAB-1 protein is tagged with a 3×FLAG epitope and fused to the green fluorescence protein, GFP, for visualization purposes. The construct is then expressed in

Received: March 07, 2022. **Accepted:** April 11, 2022

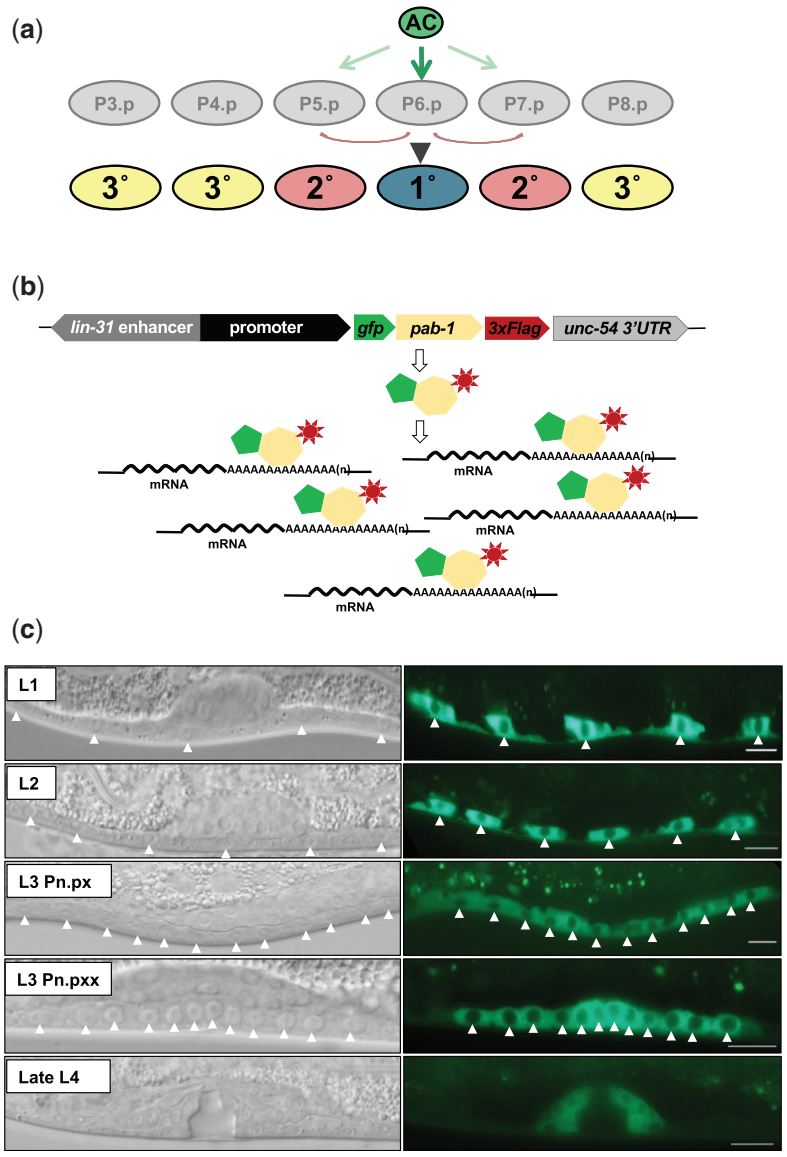

Fig. 1. Transgenic scheme to identify the vulval transcriptome. a) A schematic of VPC fate development. Signal from the anchor cell induces six competent VPCs (P3.p—P8.p) to assume one of three cell fates: 1°, 2°, and 3°, with 1° closest to the AC and 2°s flanking the 1°. 1° and 2° cells undergo three rounds of division to generate 22 vulval cells—eight for the 1° lineage and seven for each of two 2° lineages—which undergo morphogenesis to form the mature vulva in the adult. b) We generated transgenes, integrated extrachromosomal arrays *reIs27* and *reIs28*[$P_{lin\text{-}31}$:: *gfp*::*pab*-1::3×*FLAG*::*unc*-54 *3′UTR*], to serve as "+PAB-1" bait to identify vulval-specific transcripts. c) Tracking DIC (left) and GFP expression (right) from the *reIs28* transgene through stages of larval development. (We also observed expression in a handful of unidentified small cells, possibly neurons, in the head and tail; Fig. S1). The stage of animals is shown in insets. Arrowheads indicate nuclei of cells in vulval lineages. Arrows are not used at the late L4 stage because there are 22 cells in close proximity and some cells are outside the plane of focus. The control transgene, *reIs30*, lacked sequences encoding *pab-1* but still expressed GFP::3×FLAG in vulval lineages (not shown). Scale bars = 10 μm.

selected *C. elegans* tissues using defined tissue-specific promoters. Transgenic animals are then recovered, their lysate is subjected to immunoprecipitation with an anti-FLAG antibody, and the attached mRNAs are sequenced. This method is reliable and produces consistent results with minimal background noise, even using small tissues (Blazie *et al.* 2017).

Here, we applied the PAT-Seq method to isolate, sequence, and define the transcriptome of the *C. elegans* vulva throughout development. Overall, the vulva transcriptome is smaller than other tissues. Selected identified genes were further analyzed by generating extrachromosomal transgenes containing transcriptional GFP fusions as reporters for promoter expression; some of these reporters were expressed in VPCs and all in other tissues. A GFP reporter for one gene, *lag-1*, revealed expression in VPCs as expected from prior mechanistic studies. Still, the expression in the animal was far more limited than expected based on mutant phenotypes. We used CRISPR to tag the endogenous *lag-1* gene. We observed widespread and dynamic expression of LAG-1 protein, including in VPCs. Our analysis uncovers a large set of genes expressed in VPCs and provides a stepping-off point for further exploration of the genetic basis of organogenesis.

Materials and methods

Caenorhabditis elegans culturing and handling

Strains were derived from the wild-type Bristol N2 parent strain and grown at 20°C (Brenner 1974). Nomenclature conforms to that of the field (Horvitz *et al.* 1979). Crosses were performed

using standard methods and details are available upon request. Names and genotypes of strains used in this study are listed in Supplementary Table 3.

Generation of plasmids and CRISPR strains

Plasmids containing sequences encoding GFP::PAB-1::3×FLAG "+PAB-1" bait and GFP::3×FLAG "−PAB-1" control were generated through the following steps. First, we inverse PCR linearized vector pB255 vector (10,873 bp; Tan *et al.* 1998) with *lin31* promoter and enhancer using the primer pair QZ35f and QZ36r (see Supplementary Table 3; the requirement of inverted *lin-31* coding sequences to function as an enhancer were included in the original promoter::GFP reporter for *lin-31*; Tan *et al.* 1998). Second, the GFP::PAB-1::3xFLAG fragment (3,130 bp) was PCR amplified from plasmid p221 using primer pair QZ17f and QZ23r. Third, with overlapping homology arms included in both of these primers, Gibson assembly cloning kit (NEB) was applied to ligate vector and GFP::PAB-1::3×FLAG fragment to generate plasmid pQZ2. Fourth, pQZ2 was amplified by inverse PCR using primers QZ37r and QZ38f to delete the *pab-1* gene sequences to generate plasmid pQZ3. Downstream of all inserts is the *unc-54* 3′UTR contained in many *C. elegans* vectors.

Plasmids pQZ2 (GFP::PAB-1::3×FLAG) and pQZ3 (GFP::3×FLAG) were injected with the pPD118.33 ($P_{myo\text{-}2}$::*GFP*) coinjection marker in a mix of linearized plasmid and digested genomic DNA designed to mitigate silencing observed with heterologous *lin-31* promoter (A. Fire, personal communication; R.E.W. Kaplan and D. Reiner, unpublished): 1 ng/µl SacII-linearized $P_{lin\text{-}31}$-harboring plasmid, 0.25 ng/µl pPD118.33 ($P_{myo\text{-}2}$::*gfp*), 50 ng/µl EcoRV-digested and column purified *C. elegans* wild-type genomic DNA. Transgenes were tracked through crosses using a Nikon stereofluorescence microscope, based on their green pharynges. Unstable extrachromosomal transgenic lines with moderate levels of pharyngeal GFP were integrated with UV irradiation at the L4 stage using a 2400 UV crosslinker (Stratagene). UV dosage was calibrated by a dosage curve and selecting a dosage just below that which confers sterility. F2 progeny of irradiated animals were screened for 100% stable inheritance of green pharynges, resulting in integrated transgenes *reIs27*($P_{lin\text{-}31}$::*gfp*::*pab-1*::3×*FLAG*), *reIs28* ($P_{lin\text{-}31}$::*gfp*::*pab-1*::3×*FLAG*). These integrated arrays were outcrossed to the N2 wild type 4x, resulting in strains DV3507 and DV3509, respectively, and *reIs30* ($P_{lin\text{-}31}$::*gfp*::3×*FLAG*), resulting in strain DV3520. Expression was confirmed by epifluorescence and no silencing of vulval signal was ever observed.

Transcriptional promoter::GFP fusion plasmids were generated by amplifying regulatory sequences upstream of the initiating ATG codon and cloning into plasmid pPD95.67 digested with restriction enzymes HindIII and XmaI. Extrachromosomal arrays harboring GFP reporters were generated by microinjecting N2 wild-type animals with reporter plasmids at 40 ng/µl and coinjection marker pCFJ90 ($P_{myo\text{-}2}$::*mCherry*) at 1 ng/µl.

Fluorescence microscopy

Some animal handling was performed using a Nikon SMZ18 stereofluorescence microscope with 1.0x and 1.6x objectives, hybrid light transmitting base, GFP filter cube and a Xylis LED lamp. For slide-based imaging live animals were mounted in M9 buffer containing 2% tetramisole on slides with a 3% NG agar pad. DIC and epifluorescence images were acquired using a Nikon Eclipse Ni microscope and captured using NIS-Elements AR 4.20.00 software. Confocal images were acquired using DIC optics and fluorescence microscopy using a Nikon A1si confocal microscope with a 488 nm laser. Captured images were processed using NIS Elements Advanced research, version 4.40 (Nikon).

Immunoblotting

Animals were washed from plates and boiled in 4% SDS loading buffer at 95°C for 2 min to prepare lysates. Lysates were separated on 4–15% SDS gels (Bio-Rad), transferred to PVDF membrane (EMD Millipore Immobilon), and probed with mouse anti-FLAG antibody (Sigma-Aldrich #F1804) or monoclonal mouse anti-α-tubulin antibody (Sigma-Aldrich #T6199) diluted 1:2,000 in blocking solution overnight. Following primary incubation, blots were incubated with goat anti-mouse HRP-conjugated secondary antibody (MilliporeSigma 12-349) diluted 1:5,000 in blocking solution for 1 h. Immunoblots were then developed using ECL kit (Thermo Fisher Scientific) and X-ray film (Phenix).

RNA immunoprecipitation

The transgenic *C. elegans* strains used for RNA immunoprecipitations were maintained at 20°C on nematode growth media (NGM) agar plates seeded with OP50-1. Animals were then harvested, suspended, and crosslinked in 0.5% paraformaldehyde solution for 1 h at 4°C as previously described (Blazie *et al.* 2015, 2017; Hrach *et al.* 2020). We used an animal pellet of approximately 1 ml for each immunoprecipitation. Following centrifugation, animals were pelleted at 1,500 rpm, washed twice with M9 buffer, and flash-frozen in an ethanol-dry ice bath. The recovered pellets were thawed on ice and suspended in 2 ml of lysis buffer (150 mM NaCl, 25 mM HEPES, pH 7.5, 0.2 mM dithiothreitol (DTT), 30% glycerol, 0.0625% RNAsin, 1% Triton X-100; (Blazie *et al.* 2015). The lysate was subjected to sonication (Fisher Scientific) for 5 min at 4°C (amplitude 20%, 10 s pulses, 50 s rest between pulses) and centrifuged at 21,000 × g for 15 min at 4°C. One milliliter of supernatant was added per 100 µl of Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) and incubated overnight at 4°C in a tube rotisserie rotator (Barnstead international). The mRNA immunoprecipitation step was carried out as previously described (Blazie *et al.* 2015, 2017; Hrach *et al.* 2020). At the completion of the immunoprecipitation step, the precipitated RNA was extracted using Direct-zol RNA Miniprep Plus kit (R2070, Zymo Research), suspended in nuclease-free water, and quantified. Each RNA IP was performed in duplicate to produce 2 technical replicates for each of the following samples: DV3507, DV3509, and DV3520 (total 6 immunoprecipitations).

cDNA library preparation and sequencing

We prepared 6 mixed-stages cDNA libraries from the following worm strains: DV3507, DV3509, and DV3520. Each cDNA library was prepared using 100 ng of precipitated mRNAs. We used the SPIA (Single Primer Isothermal Amplification) technology to prepare each cDNA library (IntegenX and NuGEN, San Carlos, CA, USA) as previously described (Blazie *et al.* 2015, 2017). Briefly, the cDNA was sheared using a Covaris S220 system (Covaris, Woburn, MA, USA), and the resultant fragments were sequenced using the HiSeq platform (Illumina, San Diego, CA, USA). We obtained between 70M (+PAB-1) to 21M (−PAB-1) mappable reads across all 6 datasets. Because the DV3520 (−PAB-1) negative control strain does not express the *pab-1*::3×FLAG transgene, most of the immunoprecipitated mRNAs are likely non-specific, either binding to GFP, 3×FLAG, or beads used in immunoprecipitation. Consequently, we expected lower numbers of sequencing reads to be produced using the −PAB-1 strain.

Bioinformatics analysis

Raw reads mapping

The FASTQ files corresponding to the 2 datasets and the control with each corresponding replicate (total 6 datasets), were mapped to the *C. elegans* gene model WS250 using the Bowtie 2 algorithm (Langmead and Salzberg 2012) with the following parameters: –local -D 20 -R 3 -L 11 -N 1 -p 40 –gbar 1 -mp 3. The mapped reads were then converted into a bam format and sorted using SAMtools software using standard parameters (Li *et al.* 2009). We processed ∼500 M reads obtained from all our datasets combined and obtained a median mapping success of ∼90%.

Cufflinks/Cuffdiff analysis

Expression levels of genes obtained in each dataset were estimated from the bam files using the Cufflinks software (Trapnell *et al.* 2010). We calculated the fragment per kilobase per million bases (FPKM) number in each experiment and performed each comparison using the Cuffdiff algorithm (Trapnell *et al.* 2010). We obtained the expression levels by dividing the median *fpkm* value of our 2 technical replicates minus the median *fpkm* value of our 2 negative controls. We then use the ≥0.5 cutoffs to assign a given gene in the vulval data set. The results are shown in Supplementary Fig. 2 and Supplementary Table 2 using scores obtained by the Cuffdiff algorithm (Trapnell *et al.* 2010) and plotted using the CummeRbund package.

Network analysis

The network shown in Fig. 2c was constructed parsing the 1,671 hits identified in this study using the STRING algorithm (v. 11.5) (Szklarczyk *et al.* 2021), run with standard parameters. We have used only protein-protein interactions. The produced network possesses 1,666 nodes and 10,989 edges, with an average node degree of 13.2 and an average local clustering coefficient of 0.321.

The predicted miRNA targeting network was constructed extracting the longest 3′UTR sequence of the top 1,000 genes identified in our study, converting the sequences in FASTA format, and parsing the file using the MIRANDA algorithm (Enright *et al.* 2003) and the mature *C. elegans* miRNA list from miRBase release v22.1 (Griffiths-Jones *et al.* 2006) using stringent parameters (-strict -sc -1.2). MIRANDA produced 1,128 predicted targets for 114 mature *C. elegans* miRNAs. Both networks were uploaded to the Network Analyst online software (Xia *et al.* 2015) to produce the network images shown in Fig. 2c (Griffiths-Jones *et al.* 2006).

Promoter analysis

We used custom Perl scripts to extract 2,000 nt from the transcription start site for the top 90 genes identified in this study. We then used different custom Perl scripts to calculate the nucleotide distribution. The transcription factor predictions were produced parsing these promoters to the Simple Enrichment Analysis script from the MEME suite software (Bailey *et al.* 2015).

Results

Engineering the vulval cell lineage for transcriptomic sampling

Identifying a tissue-specific transcriptome requires expressing a bait protein specifically in the tissue of interest. The *lin-31* gene encodes the *C. elegans* ortholog of FoxB transcription factors. LIN-31 is a terminal selector protein that functions in collaboration with LIN-1/Ets and the Mediator Complex to mediate induction of vulval fates (Miller *et al.* 1993, 1996; Beitel *et al.* 1995; Tan *et al.* 1998; Hart *et al.* 2000; Grants *et al.* 2016; Underwood *et al.* 2017). Yet, unlike LIN-1, genetic perturbation of LIN-31 is described as impacting only the VPCs (Miller *et al.* 1993, 2000; Tan *et al.* 1998). Furthermore, the promoter of *lin-31* drives GFP expression chiefly in the VPCs (Tan *et al.* 1998). Consequently, we used the *lin-31* promoter to transgenically express bait protein in the VPCs throughout development.

The *C. elegans* ortholog of polyadenylation binding protein 1, PAB-1, specifically binds poly(A) tails of mature mRNAs and can be used to immunoprecipitate mRNAs from whole-RNA preparations (Blazie *et al.* 2015, 2017). Tissue-specific expression of PAB-1 with GFP and an epitope tag has been used to identify tissue-specific transcriptomes from large tissues like neurons, intestine, hypodermis, and muscle, as well as smaller subsets of specific neuron types (Blazie *et al.* 2017).

We cloned a sequence encoding a fusion of GFP, PAB-1, and 3×FLAG epitope into a vector containing the *lin-31* promoter (Fig. 1b). We generated transgenic extrachromosomal arrays, randomly integrated arrays into the genome to generate *reIs27* and *reIs28*, both consisting of $P_{lin\text{-}31}$::*GFP*::*PAB-1*::3×*FLAG* + $P_{myo\text{-}2}$::*GFP* (i.e. "+PAB-1"), and outcrossed to the wild type N2 strain to generate DV3507 and DV3509, respectively. We similarly generated *reIs30*, expressing $P_{lin\text{-}31}$::*GFP*::3×*FLAG* + $P_{myo\text{-}2}$::*GFP* (i.e. "−PAB-1") and outcrossed to the wild type to generate DV3520. Critically, *reIs30* expresses control protein lacking the PAB-1 sequences.

Published anti-LIN-31 immunohistochemistry described expression of endogenous LIN-31 in the broader Pn.p group, including nonVPCs P1,2.p and P9,10,11.p, during the L1 stage before expression is limited to VPCs P3.p-P8.p (Tan *et al.* 1998). In contrast, our construct only expressed in VPCs and vulval lineages: analysis of *reIs28*($P_{lin\text{-}31}$::*GFP*::*PAB-1*::3×*FLAG* + $P_{myo\text{-}2}$::*GFP*)-bearing animals using DIC and epifluorescence microscopy revealed expression of GFP in vulval lineages throughout larval development, from the first (L1) to the fourth (L4) larval stages and young adult (YA) (Fig. 1c). *reIs27* and *reIs30* similarly expressed GFP in vulval lineages post-embryonically. We also observed additional expression in a small number of unidentified cells in the head and tail (Supplementary Fig. 1). The resulting "+PAB-1" bait- and "−PAB-1" control-expressing transgenes were subsequently used for identification of the transcriptome of the vulval lineages. However, we note that expression in other cells would identify a transcriptome of mixed lineages (see Discussion). We anticipate that subsequent validation of putative VPC-specific genes via promoter::gfp fusion analysis should determine which are expressed in the vulval lineages.

Identifying the transcriptome of the vulval lineage

We have prepared 2 independent transgenic strains expressing our vulva-specific pulldown construct ("+PAB-1" biological replicates; DV3507 and DV3509) and one control strain in which we deleted sequences encoding PAB-1 (DV3520; "−PAB-1" negative control). We performed each immunoprecipitation in duplicate (technical replicates), processing a total of 6 samples. We obtained approximately ∼90M mappable reads for each biological and technical replicate and ∼30M mappable reads for our negative control strain DV3520. We could map more than 90% of the total reads across all samples (Supplementary Fig. 2a). The results obtained with our biological replicates correlate well (Supplementary Fig. 2, b and c).

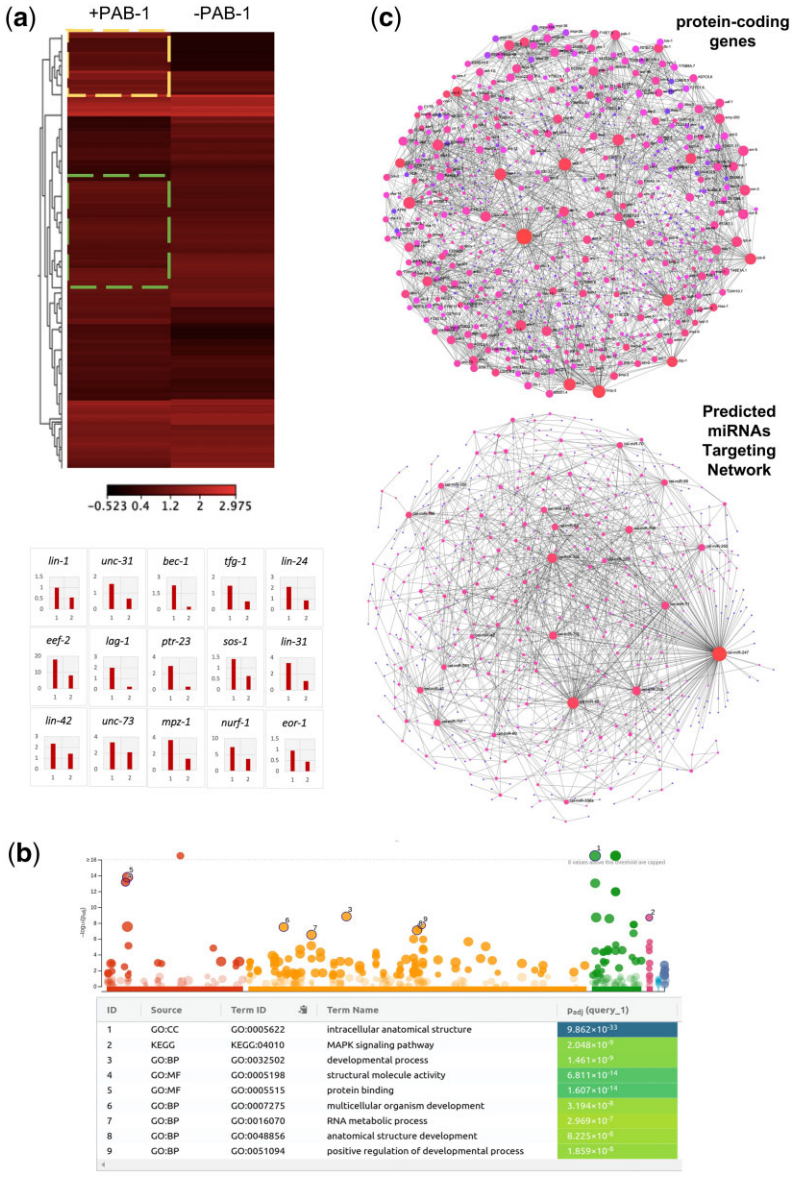

Fig. 2. The vulval transcriptome. a) A Heatmap comparison of the vulva transcriptome (+PAB-1; median FPKM value of DV3507 and DV3509), as compared to its negative control (−PAB-1; DV3520 negative control). Upper or lower boxes with dotted lines in the heat map indicate genes that were either up- or downregulated in the vulva dataset, respectively. Several upregulated genes are shown in the bar chart below the heatmap. b) Summary of the GO term analysis produced by the vulva transcriptome dataset. c) The vulva protein-coding gene interactome (top) and the predicted miRNA targeting network (bottom).

The *C. elegans* vulva transcriptome

Using our PAT-Seq approach, we were able to map 1,671 protein-coding genes in the *C. elegans* vulva, which corresponds to 8.2% of all *C. elegans* protein-coding genes (20,362 protein-coding genes; WS250; Fig. 2a and Supplementary Tables 1 and 2). As expected, a GO term analysis highlights "intracellular anatomical structure," "MAP kinase pathway," "developmental processes," which are all entries consistent with the tissue of origin of our immunoprecipitated RNAs (Fig. 2b). While several of our top hits are genes with unknown functions, (e.g. *F27D4.4*, *Y65A5A*, *fipr-1*, and *F49B2.3*), many others have been previously linked to vulval development or morphogenesis. For example, the *C. elegans* ortholog of translation elongation factor 2 EEF-2 (Fraser *et al.* 2000), a GTP-binding protein required for embryogenesis and vulval morphogenesis, expressed during all stages of development; the coiled-coil domain protein GEI-4, is required for embryonic viability, fertility, and vulval morphogenesis (Poulin *et al.* 2005); and the *C. elegans* ortholog of Drosophila NURF301, NURF-1, a member of the NURF chromatin remodeling complex, which is also known to regulate vulval development (Andersen *et al.* 2006) (Fig. 2 and Supplementary Table 1). We also detected several known transcription factors, including *lin-22*, *lin-31*, *eor-1*, *lag-1*, and *pop-1*, all known to be expressed and functioning in vulval lineages (Fig. 2 and Supplementary Table 1).

In addition to the well-known vulval marker *lin-31*, other genes mutated to a lineage-defective phenotypes during development of the vulva were identified by our sequencing effort, including *lin-22*, *lin-24*, *lin-41*, and *lin-42*. *lin-22* encodes an ortholog of human HES1 and HES6 [hes family bHLH transcription factors 1 and 6; (Schlager *et al.* 2006). *lin-24* was originally identified in a

screen for mutations that result in altered vulval cell lineages and is expressed in vulval lineages (Ferguson *et al.* 1987; Galvin *et al.* 2008). *lin-41* and *lin-42* are genes involved in the heterochronic pathway to regulate developmental switches that occur in multiple tissues, including the vulva (Tennessen *et al.* 2006; Parry *et al.* 2007; Fig. 2 and Supplementary Table 1).

We also identified 23 genes that, when mutated, confer defective locomotion (Uncoordinated; "Unc"). Some of these, like UNC-31, an ortholog of the human CADPS (calcium-dependent secretion activator) required for the Ca^{2+}-regulated exocytosis of secretory vesicles, would be presumed to be abundant contaminating transcripts associated with the neurons in which our "+PAB-1" bait protein is also expressed. And yet UNC-31 is expressed in certain vulval sublineages (Speese *et al.* 2007). UNC-32 is a vacuolar proton-transporting ATPase (V-ATPase), expressed in the vulva in adults, that contributes to a protruding vulva phenotype when depleted (Oka *et al.* 2001; Pujol *et al.* 2001; Shephard *et al.* 2011).

Other "Unc" genes are more conventionally associated with the development of the vulva. UNC-73 is an ortholog of mammalian TRIO (Steven *et al.* 1998), a guanyl-nucleotide exchange factor (GEF) that stimulates GTP-loading on Rho family small GTPases like CED-10/Rac and MIG-2/RhoG, which control cytoskeletal dynamics. UNC-73, CED-10, and MIG-2 regulate vulval morphogenesis and extensive axonal growth cone and cell migration events (Steven *et al.* 1998; Kishore and Sundaram 2002). UNC-62 is the ortholog of *Drosophila homothorax* (mammalian Meis/Prep), a transcription factor that regulates the development of many tissues, including vulval lineages (Yang *et al.* 2005).

Other previously identified genes with functions in the vulva are SQV-6, which is similar to the human protein xylosyltransferase involved in modification of proteoglycan cores, localizes in the Golgi and endoplasmic reticulum membranes, and is required for vulval morphogenesis (Hwang *et al.* 2003), and HMP-2, a β-Catenin required for epithelial cell migration and elongation during embryo morphogenesis and necessary for vulva morphogenesis (Costa *et al.* 1998; Hoier *et al.* 2000, Fig. 2 and Supplementary Table 1). A novel feature of *C. elegans* is that transcriptional and cytoskeletal functions of β-catenin are performed by distinct paralogs, BAR-1 and HMP-2, respectively, rather than the same protein as in other systems (Eisenmann 2005).

The network shaped by our identified genes, although small, results in an interconnected protein interactome (Fig. 2c).

miRNA targets

MicroRNAs have been found to be involved in the morphogenesis of the vulva. *let-7* is expressed in the vulval tissue and found to target the 3′UTRs of several genes, including the 3′UTR of the *lin-41* heterochronic gene, to prevent vulval rupturing (Ecsedi *et al.* 2015), and in the 3′UTR of *let-60*, which encodes the Ras ortholog, and the genes for mammalian Ras orthologs (Grosshans *et al.* 2005). We sought to identify potential miRNAs and their targets expressed in the vulva. We parsed our dataset using the MIRANDA software (Enright *et al.* 2003) and identified 1,128 predicted targets for 114 mature *C. elegans* miRNAs (Fig. 2c and Supplementary Table 2). The most abundant miRNA targets in our study are those of miR-247, potentially targeting 157 vulval-expressed genes, followed by those of miR-85 (85 genes) and miR-255 (56 genes). Sequences upstream of the initiating ATG for *mir-241* drive a GFP reporter in the vulva (Martinez *et al.* 2008). Unfortunately, our PAT-Seq method was not designed to identify miRNAs, and more experiments need to be performed to validate the presence of these miRNAs in this tissue.

Promoter analysis

Next, we aimed to study the promoter composition of the genes detected in our study with the goal of identifying novel *cis*-acting elements potentially used by vulva-specific transcription factors. We have extracted DNA regions 500 bp upstream and 100 bp downstream from the transcription start of our top 100 genes detected in our study (Supplementary Fig. 2). We identified 3 motifs. The first motif, CAACCTGC, is recognized by the human transcription factor TCF12, a basic helix-loop-helix (bHLH) factor that in humans regulates lineage-specific gene expression through the formation of heterodimers with other bHLH E-proteins using mainly the ERK and WNT signaling pathway (Belle and Zhuang 2014; Supplementary Fig. 4b, top panel). Use of Ortholist 2 (Kim *et al.* 2018) suggests HLH-2 is similar to TCF12. HLH-2 was not detected in our vulval transcriptome but RNAi depletion causes a protruding vulva phenotype (Ceron *et al.* 2007). The second motif, CAATTAA, in humans is targeted by Hmx2, a Homeodomain transcription factor and plays an important role in organ morphogenesis and development during embryogenesis; Supplementary Fig. 4b, middle panel). Hmx2 is similar to MLS-2 in *C. elegans*, though it was not detected in our vulval transcriptome and MSL-2 is not known to function in vulval development. The third motif, CCACGCCCAC, in humans is targeted by SP3, a 3-zinc finger Kruppel-related transcription factor that stimulates or represses the transcription of numerous genes (Supplementary Fig. 4b, middle panel). SP3 is similar to SPTF-2 and SPTF-3 in *C. elegans*. These were not found in our vulval transcriptome, but mutant SPTF-3 confers a bivulva phenotype that reflects reversed polarity of the P7.p lineage (Ulm *et al.* 2011).

Validating selected candidate genes hypothesized to be expressed in VPC lineages

Our PAT-Seq analysis, based on a comparison of data sets generated with $P_{lin\text{-}31}$::GFP::PAB-1::3×FLAG "+PAB-1" vs $P_{lin\text{-}31}$::GFP::3×FLAG "−PAB-1" control, generated a set of genes potentially expressed in VPCs. Notably, the expression of this set of genes is not expected to be exclusive to VPCs and may also be expressed in other tissues.

To validate our approach, we selected candidate genes identified in this study for analysis with promoter::GFP transgenes to ascertain whether they are expressed in VPCs. We cloned sequences upstream of the ATG initiator methionine codon for several genes into vector pPD95.67 with 2xNLS::GFP (nuclear localization signal) and generated extrachromosomal arrays harboring these clones (Fig. 3a). Given the interests of our research program, we focused on genes potentially regulating signaling and/or developmental biology, with some randomly selected genes included.

The *lag-1* gene encodes the *C. elegans* ortholog of *Drosophila* Suppressor of Hairless/Su(H), the central nuclear transcriptional regulator downstream of the Notch receptor. Because of the role of LIN-12/Notch in the induction of 2° VPC fate, expression of LAG-1 is expected in VPCs. Indeed, we observed GFP expression driven by *lag-1a* promoter sequences in VPCs (Fig. 3b, panel i). However, despite the broad use of LIN-12/Notch and GLP-1/Notch in development (reviewed in Priess 2005; Greenwald and Kovall 2013), we did not observe expression from the *lag-1a* promoter in other tissues, including the germline and embryo. This observation reinforces the notion that sequences defined as "promoters"

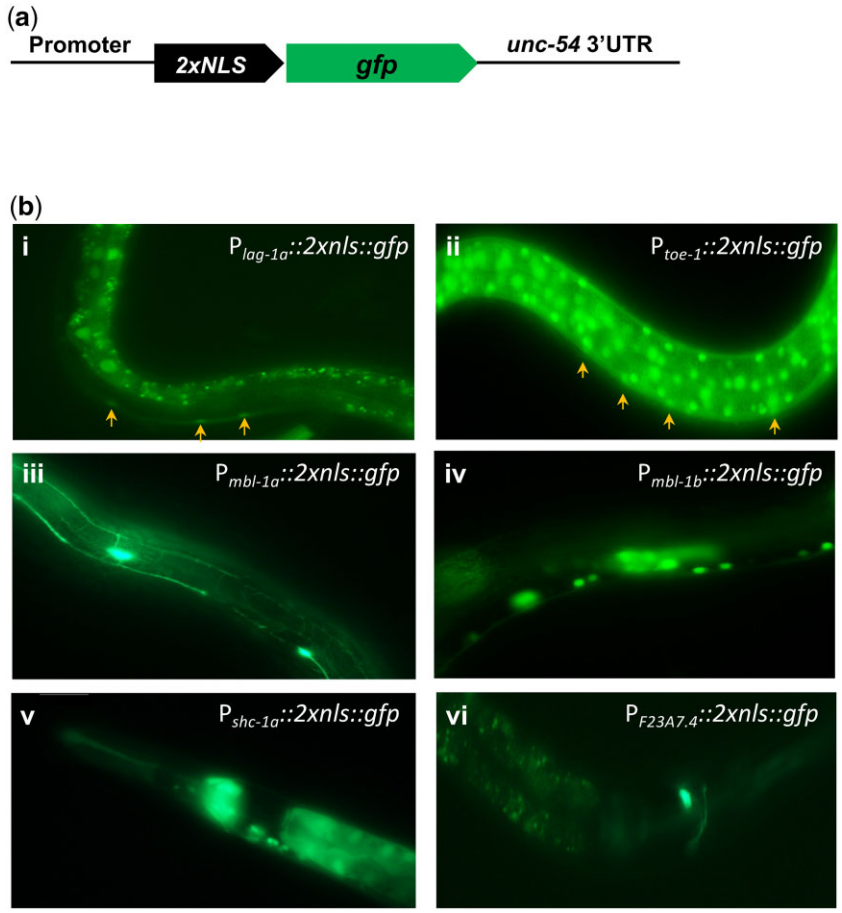

Fig. 3. Promoter::GFP expression patterns of selected genes in the VPC-expressed data set. A) A general schematic of promoter sequences cloned in front of *2xNLS::gfp* and the *unc-54* 3′UTR. The solid line represents upstream sequences and dotted line represents inverted *lin-31* coding sequences that contain an enhancer in the pB255 plasmid (Tan *et al.* 1998). B) Epifluorescence images of extrachromosomal transgenic lines generated for this study. Arrows indicate GFP expression in VPCs. i) The *lag-1a* promoter drives expression in VPCs but not in other tissues expected to express LAG-1. ii) The *toe-1* promoter drives expression at high levels in hypodermis and other cell types, including VPCs but not neighboring ventral cord neurons. iii) The *mbl-1a* promoter is expressed in the touch response neurons (TRNs). iv) The *mbl-1b* promoter drives expression in ventral neurons and putative neurons in the head but not VPCs. v) The *shc-1a* drives expression in head neurons and intestinal cells but not VPCs. vi) The promoter of F23A4.7 drives expression in a head neuron but not VPCs.

are limited by an arbitrary distance upstream of initiator methionine codons. Key regulatory sequences are likely to be present further upstream, in coding sequences, or downstream of codon sequences (e.g. a repressive element for the *egl-1* gene located 5.6 kb downstream of the termination codon of *egl-1*; Conradt and Horvitz 1999).

Sequences upstream of *toe-1* (Target Of ERK kinase MPK-1) (Arur *et al.* 2009), a putative nucleolar protein also identified in this study drove expression of GFP strongly in nuclei of many cells, including VPCs (Fig. 3b, panel ii).

mbl-1, which encodes an RNA-binding protein from 2 promoters, *a* and *b*, showed diverse expression from the 2 promoters. Sequences upstream of *mbl-1a* drove expression in a subset of neurons while those upstream of *mbl-1b* drove expression in the vicinity of the ventral nerve cord but not in VPCs (Fig. 3b, panels iii and iv). Sequences upstream of the *a* isoform of *shc-1*, a signaling adaptor protein [SHC (Src Homology domain C-terminal) adaptor ortholog], drove expression in the pharynx, intestine, and neurons but not VPCs (Fig. 3b, panel v). Finally, we tested the promoter of a not yet characterized gene, *F23A7.4*, which encodes a *C. elegans*-specific protein. Although in our dataset, we were unable to detect its expression in VPCs but detected its expression in unidentified neurons (Fig. 3b, panel vi).

CRISPR tagged *lag-1* gene reveals expression broadly throughout development

We speculated that the reason why we were unable to detect *lag-1* expression in the germline and embryos—but still detect it in the vulva—was because this gene possesses four splice variants that differ at their 5′ end but share the same 3′ end (Fig. 4a). We hypothesized that perhaps each of these isoforms possess differential tissue localization, and our cloned promoter region, which was specific to the *a* isoform, while driving strong vulva expression, was not enough to drive the expression of other *lag-1* isoforms, perhaps expressed in germline and embryos. To further explore the "missing" expression from our cloned *lag-1a* putative promoter sequence in the germline and embryos, we used CRISPR technology to tag the endogenous *lag-1* gene at the 3′ end with sequences encoding mNeonGreen (mNG) fluorescent protein and a 3×FLAG epitope. We expected to detect full-length protein fusions regardless of the use of different promoters at the 5′ end. Specifically, we used the "self-excising cassette" (SEC) method for two-step positive-negative selection from a single plasmid and injection (Dickinson *et al.* 2015, Fig. 4b). Proper insertion was validated by PCR and western blotting of worm lysates probed with anti-FLAG (for FLAG-tagged LAG-1) and anti-α-tubulin (control). Anti-FLAG antibodies recognized 2 major tagged

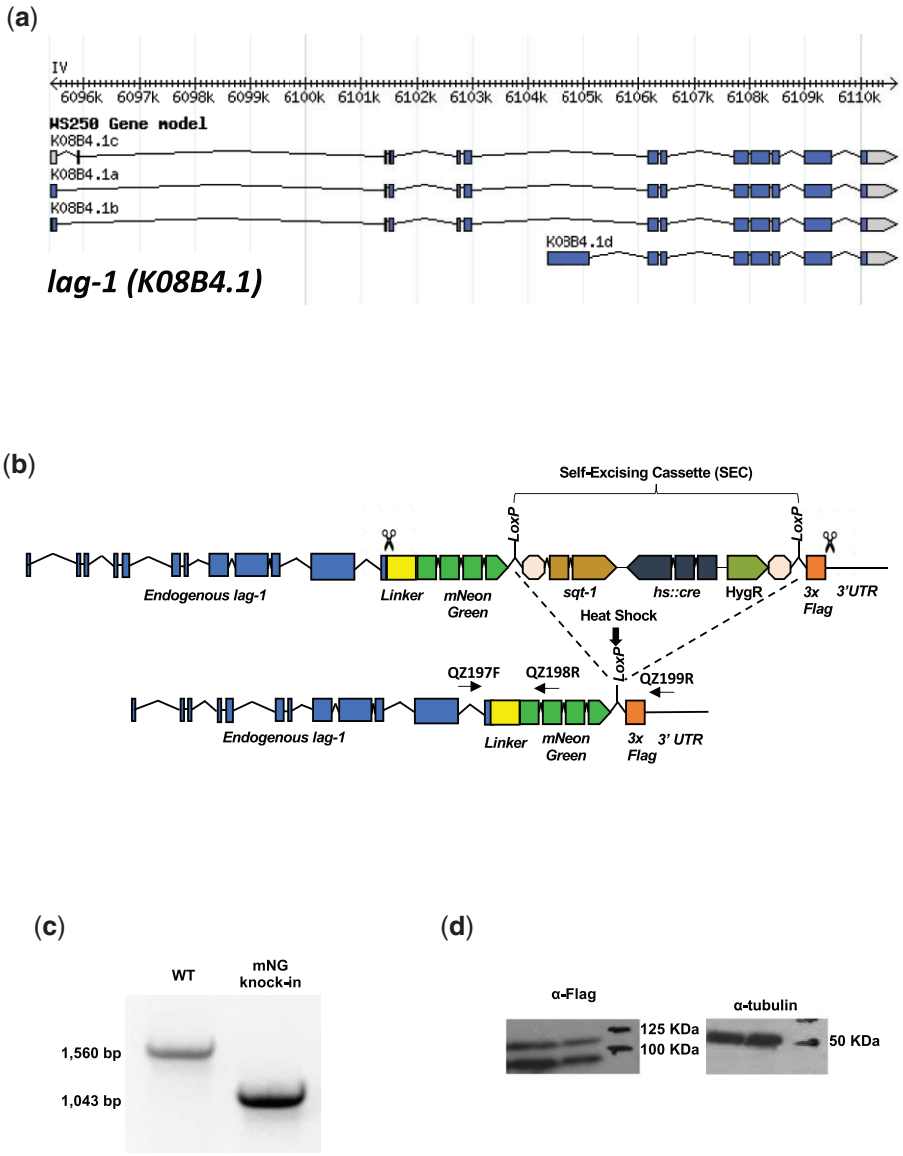

Fig. 4. CRISPR tagging strategy for the *lag-1* C-terminus. a) WormBase gene model for *lag-1* (K08B4.1) on LGI. By RNAseq analysis in WormBase, the *d* isoform is a small minority. b) CRISPR tagging strategy using the SEC approach (Dickinson, *et al.* 2015). Detection primers are denoted by "QZ" (see Supplementary Table 3) PCR detection of wild-type and homozygous insertion bands. d) Western blot detection using anti-FLAG antibody of endogenous LAG-1::mNeonGreen::3×FLAG protein; the tag portion of the protein is predicted to be 28.8 kD. Isoform D with tag is predicted to be 116.5 kD, while isoforms A, B, and C are predicted to be 103.7, 103.5, and 98.5 kD, respectively. We detected 2 general band species but due to gel smiling it was unclear how well they matched predicted sizes. Control anti-tubulin antibody detected the expected 50 kD bands (left 2 lanes) and the 50 kD marker (right).

protein products (Fig. 4c). While this study was in progress, another group described the expression of tagged LAG-1, with similar results (Luo *et al.* 2020).

As expected based on functional studies (Christensen *et al.* 1996; Yoo *et al.* 2004; Greenwald 2005; Kimble and Crittenden 2005; Priess 2005; Yoo and Greenwald 2005), we observed LAG-1::mNeonGreen expression broadly and localized to nuclei throughout vulval development, from L1 through L4 stages and in neighboring uterine lineages (Fig. 5). We also observed dynamic LAG-1 expression in various embryonic cells (Fig. 6). LAG-1::mNeonGreen expression was also observed broadly throughout the animal at various stages (Fig. 7, a and b). In conjunction with the description of other expression patterns derived from endogenous genes tagged by CRISPR from our lab (Rasmussen *et al.* 2018; Shin *et al.* 2018, 2019; Duong *et al.* 2020; Rasmussen and Reiner 2021), we concluded that promoter::GFP fusions are limiting. Positive results are likely informative, but the absence of expression often can be a result of regulatory sequences missing from transcriptional expression reporters.

Discussion

The *C. elegans* vulva is an effective model for the study of signaling cues and morphogenic processes required in development to produce an organ. Although numerous studies thus far have highlighted fundamental processes and key genes involved in its morphogenesis, the composition of its transcriptome and its interactome are still not known. Here, we have used PAT-Seq, a

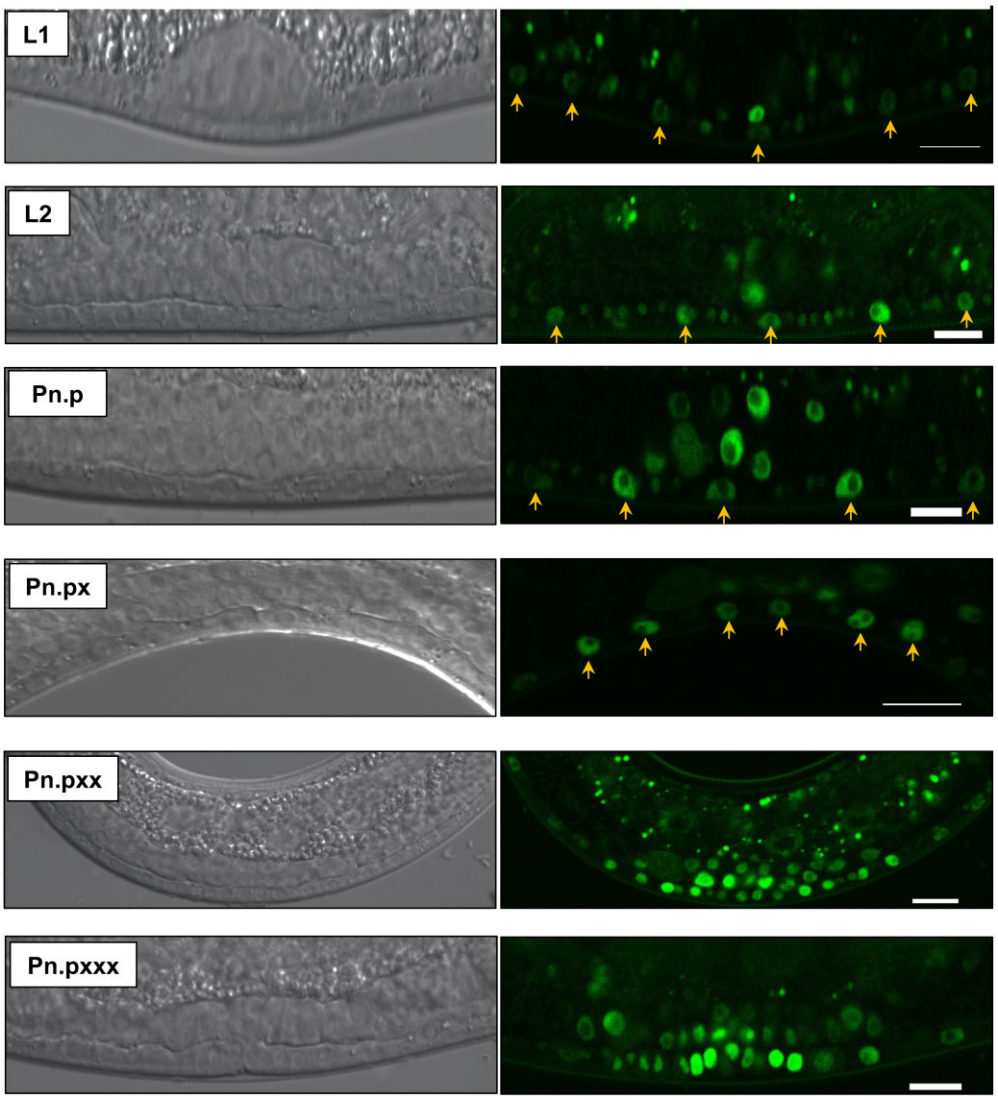

Fig. 5. Expression of tagged endogenous LAG-1::mNeonGreen in VPCs and the ventral gonad. We observed expression and nuclear localization of green fluorescence in the VPCs, vulval lineages and ventral gonad throughout larval development. Notably, we consistently observed decreased but not eliminated nuclear green fluorescence in 1° lineage descendants (P6.px and P6.pxx). Each "x" in lineage notation indicates daughters of an original Pn.p cell. Stage of development is noted on the left. Arrows indicate Pn.p or Pn.px cells at the ventral midline of the animal, later vulval lineages are self-evident. The ventral gonad is not indicated but is directly above the vulval lineages. Scale bars = 10 μm.

method that allows the isolation of high-quality tissue-specific transcriptomes, to sequence and study the *C. elegans* vulval transcriptome. We have identified 1,671 high-quality *bona fide* genes expressed in this tissue and developed accurate miRNA targeting predictions in this dataset. Overall, the vulva dataset is smaller than other tissues we have profiled in the past (Supplementary Fig. 3). Within its transcriptome, we defined 39 transcription factors, 49 kinases, 50 membrane-associated proteins and 118 genes containing an RNA binding domain (Supplementary Table 1). Our promoter analysis in Supplementary Fig. 4 identified 3 specific DNA elements enriched in promoters of vulva-transcribed genes, which are targeted by transcription factors previously not associated with vulva development in worms (TCF12, Hmx2, and Sp3).

Unfortunately, there are no available *C. elegans* vulva datasets we could use to compare our results, and we cannot conclusively pinpoint all the genes expressed in this organ. Importantly, we have sequenced two independently generated transgenic animal lines (biological replicates DV3507 and DV3509), with a technical replicate each, and subtracted the genes identified in the sequencing results of our negative control (DV3520), which is unable to bind poly(A) tails, to thus isolate transcripts specific to the vulva. Our PCA analysis (Supplementary Fig. 2c) shows that our two biological replicates correlate well with each other, suggesting little contamination.

Using transgenes harboring promoter::GFP transcriptional fusion reporters, we also were able to validate putative targets identified in our study (e.g. *lag-1* and *toe-1*), as being expressed in the VPCs, while for others (e.g. *shc-1*, *mbl-1*, and *F23A7.4*) we were unable to detect expression in VPCs, perhaps because of false-positive candidates or the insufficiency of the promoter::GFP

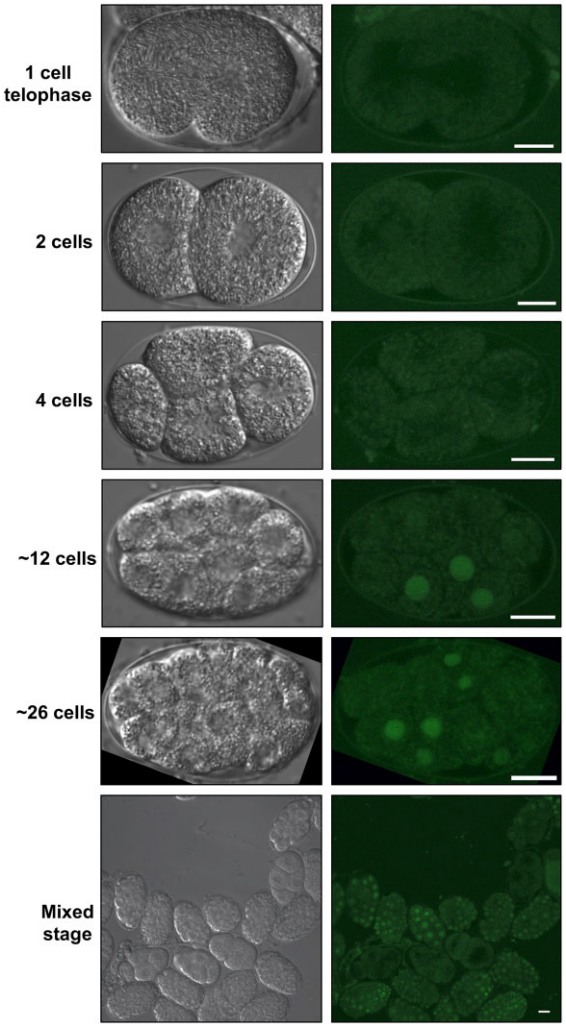

Fig. 6. Dynamic regulation of LAG-1::mNeonGreen expression in embryos. Expression of LAG-1::mNeonGreen may be absent until the 12-cell stage, and then is expressed at various levels. After, the expression was observed in nuclei of subsets of cells. The stage of development is noted on the left. Scale bars = 10 μm.

transgenes in reflecting the full expression patterns of genes. One validated target, *lag-1*, exhibited limited expression via promoter::GFP fusion analysis (Fig. 3), but our CRISPR tagging of the endogenous protein revealed spatiotemporally broad and dynamic expression (Figs. 5–7).

A *caveat* to our analysis is that the *lin-31* promoter sequences derived from the plasmid pB255 (Tan *et al.* 1998) also drive expression of GFP in 2 to 3 small cells, perhaps neurons, each in the head and tail. We have been unable to identify these cells, though could likely do so using a comprehensive label for neurons in the worm (Yemini *et al.* 2021). But the presence of additional nonvulval cells, albeit of smaller collective volume than the vulval cells, indicates that a subset of our identified genes is likely specific to nonvulval lineages. This set cannot be discriminated at present but likely represents a significant source of false-positive candidates in our transcriptome set. A solution to this conundrum would be to perform deletion analysis on the *lin-31* promoter to identify regulatory DNA sequences specific to these ancillary cells but not vulval lineages. Such analysis would position us to perform PAT-SEQ that is more specific to vulval lineages.

Another *caveat* is the fusion of 3° VPC cells to the hyp7 syncytium after initial patterning of VPC cell fates. VPCs are specialized hypodermal cells surrounded by nonspecialized hypodermis, called the hyp7, a syncytium comprised of many fused hypodermal cells. 1° and 2° cells (Fig. 1a) go through stereotyped series of cell divisions, but nonvulval 3° cells divide once and fuse to the surrounding syncytium. The release of "+PAB-1" protein into the general hyp7 syncytium may result in identification of transcripts specific to the hyp7. But we expect the concentration of "+PAB-1" protein in the hyp7 after fusion of the 3° daughters to be relatively low, and GFP in the hyp7 was not observed after the 3° fusion at the L3 stage. Unlike improvement of the *lin-31* promoter used to express "+PAB-1" protein, we foresee no plan for working around this limitation to our approach.

A final limitation to our approach is the complexity of the vulval system over time. Our bait "+PAB-1" protein and control "−PAB-1" protein proteins were expressed from the L1 to young adult stages (Fig. 1c), yet we obtained a sample of mixed stage populations. During this time, developmental competence of P3.p-P8.p is established through the actions of multiple transcription factors including Homeobox proteins (Clandinin *et al.* 1997; Wagmaister *et al.* 2006a, 2006b; Myers and Greenwald 2007; Takacs-Vellai *et al.* 2007), unexpectedly early MPK-1/ERK signaling observed in the L2 VPCs prior to conventional induction in the L3 (de la Cova *et al.* 2017), the VPCs have at least 5 signals activated [3 major and 2 modulatory (Gleason *et al.* 2006; Nakdimon *et al.* 2012; Shin *et al.* 2019) reviewed in Shin and Reiner (2018), fusion of 3°s (Shemer and Podbilewicz 2002)], 3 rounds of distinctive and highly reproducible cell divisions specific to 1° and 2° lineages specifically (Sulston and Horvitz 1977; Braendle and Felix 2008), polarity of 2° lineages (Inoue *et al.* 2004; Gleason *et al.* 2006; Green *et al.* 2007, 2008; Kidd *et al.* 2015), and a sophisticated series of morphogenetic events ending with joining of the vulva and uterus (Hagedorn and Sherwood 2011; Cohen *et al.* 2020; Spiri *et al.* 2022) to form a tube through which eggs are laid and males deposit sperm in the spermatheca. The developmental complexity of this system, probably reflected by the complexity of transcriptional changes, surely must exceed the resolution of our PAT-SEQ analysis, probably by a great margin.

Yet it is important that we pilot this technology in the vulval system to be able to refine our analysis in the future. A more specific vulva promoter driving bait "+PAB-1" protein and control "−PAB-1" protein proteins coupled with tight synchrony of animal populations may yield coherent temporal "slices" of gene expression in the vulval system over time. Even though such an approach would not be able to distinguish between different vulval lineages, deconvoluting gene expression in this manner may shed important light on the process of organogenesis and identify specific candidates for further analysis later in development. Such PAT-SEQ analysis with more refined temporal lysates could also be performed with "+PAB-1" bait protein expressed in specific lineages or sublineages after initially patterning, or in backgrounds where patterning signals

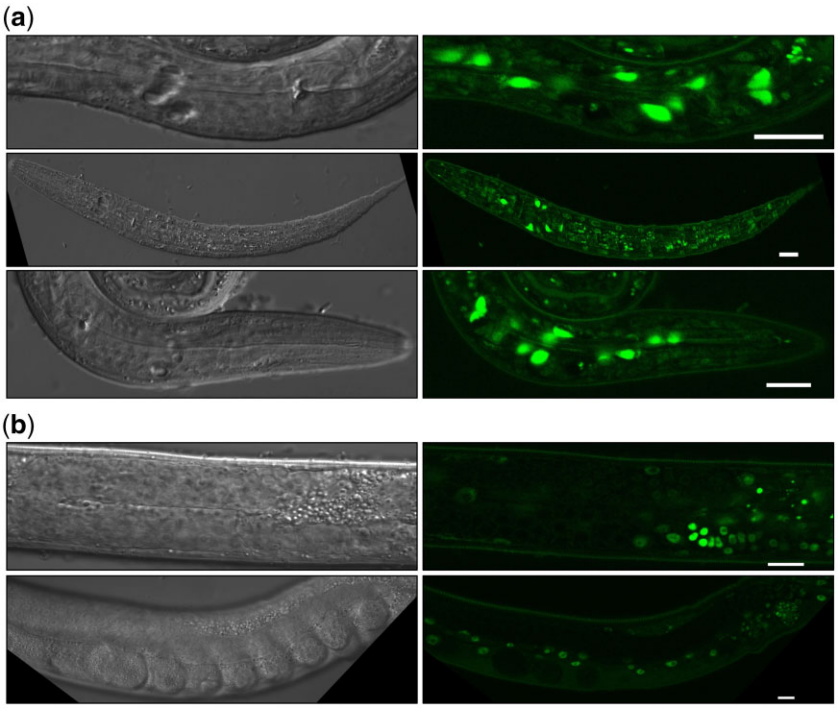

Fig. 7. Other expression of endogenous LAG-1::mNeonGreen. a) Strong expression was observed in a variety of cells in the head. b) Expression was observed in the somatic germline but not in proximal germ cells. We never observed expression in the distal tip cell, but we did observe faint expression in expression in distal-most germ nuclei in the mitotic and transition zones, perhaps reflecting GLP-1/Notch signaling to these nuclei among the germline syncytium. Expression was also observed in sperm. Scale bars = 10 μm.

are altered mutationally, to identify sets of transcriptional client genes that respond to those signals.

Data availability

Raw reads were submitted to the NCBI Sequence Read Archive (http://trace.ncbi.nlm.nih.gov/Traces/sra/) with BioProject ID: PRJNA811605 and Submission ID: SUB11135968. The results of our analyses are available in Excel format as Supplementary Table 1.

Supplemental material is available at G3 online.

Acknowledgments

WormBase was used routinely. We thank members of the Reiner lab for discussion and helpful comments on the manuscript.

DJR and MM designed the experiments. QZ performed the experiments described in Figs. 1 and 3–7. HH performed the lin-31 PAT-Seq immunoprecipitations and prepared the sequencing reactions. MM performed the bioinformatic analysis. MM and DR analyzed the data. MM and DJR led the analysis and interpretation of the data, assembled the figures, and wrote the manuscript in collaboration with HH. All authors read and approved the final manuscript.

Funding

This work was supported by NIH grant 1R21HD090707 to DJR and 1R01GM118796 to MM. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

Conflicts of interest

None declared.

Literature cited

Andersen EC, Lu X, Horvitz HR. *C. elegans* ISWI and NURF301 antagonize an Rb-like pathway in the determination of multiple cell fates. Development. 2006;133(14):2695–2704.

Arur S, Ohmachi M, Nayak S, Hayes M, Miranda A, Hay A, Golden A, Schedl T. Multiple ERK substrates execute single biological processes in *Caenorhabditis elegans* germ-line development. Proc Natl Acad Sci USA. 2009;106(12):4776–4781.

Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43(W1):W39–W49.

Beitel GJ, Tuck S, Greenwald I, Horvitz HR. The *Caenorhabditis elegans* gene lin-1 encodes an ETS-domain protein and defines a branch of the vulval induction pathway. Genes Dev. 1995;9(24):3149–3162.

Belle I, Zhuang Y. E proteins in lymphocyte development and lymphoid diseases. Curr Top Dev Biol. 2014;110:153–187.

Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in *Caenorhabditis elegans* intestine and muscles. BMC Biol. 2015;13:4.

Blazie SM, Geissel HC, Wilky H, Joshi R, Newbern J, Mangone M. Alternative polyadenylation directs tissue-specific miRNA targeting in *Caenorhabditis elegans* somatic tissues. Genetics. 2017;206(2):757–774.

Braendle C, Felix MA. Plasticity and errors of a robust developmental system in different environments. Dev Cell. 2008;15(5):714–724.

Brenner S. The genetics of *Caenorhabditis elegans*. Genetics. 1974;77(1):71–94.

Ceron J, Rual JF, Chandra A, Dupuy D, Vidal M, van den Heuvel S. Large-scale RNAi screens identify novel genes that interact with the *C. elegans* retinoblastoma pathway as well as splicing-related components with synMuv B activity. BMC Dev Biol. 2007;7:30.

Christensen S, Kodoyianni V, Bosenberg M, Friedman L, Kimble J. lag-1, a gene required for lin-12 and glp-1 signaling in *Caenorhabditis elegans*, is homologous to human CBF1 and Drosophila Su(H). Development. 1996;122(5):1373–1383.

Clandinin TR, Katz WS, Sternberg PW. *Caenorhabditis elegans* HOM-C genes regulate the response of vulval precursor cells to inductive signal. Dev Biol. 1997;182(1):150–161.

Cohen JD, Sparacio AP, Belfi AC, Forman-Rubinsky R, Hall DH, Maul-Newby H, Frand AR, Sundaram MV. A multi-layered and dynamic apical extracellular matrix shapes the vulva lumen in *Caenorhabditis elegans*. eLife. 2020;9:e57874.

Conradt B, Horvitz HR. The TRA-1A sex determination protein of *C. elegans* regulates sexually dimorphic cell deaths by repressing the egl-1 cell death activator gene. Cell. 1999;98(3):317–327.

Costa M, Raich W, Agbunag C, Leung B, Hardin J, Priess JR. A putative catenin-cadherin system mediates morphogenesis of the *Caenorhabditis elegans* embryo. J Cell Biol. 1998;141(1):297–308.

de la Cova C, Townley R, Regot S, Greenwald I. A real-time biosensor for ERK activity reveals signaling dynamics during *C. elegans* cell fate specification. Dev Cell. 2017;42(5):542–553.e4.

Dickinson DJ, Pani AM, Heppert JK, Higgins CD, Goldstein B. Streamlined genome engineering with a self-excising drug selection cassette. Genetics. 2015;200(4):1035–1049.

Duong T, Rasmussen NR, Ballato E, Mote FS, Reiner DJ. The Rheb-TORC1 signaling axis functions as a developmental checkpoint. Development. 2020;147:dev181727.

Ecsedi M, Rausch M, Großhans H. The let-7 microRNA directs vulval development through a single target. Dev Cell. 2015;32(3):335–344.

Eisenmann DM. Wnt signaling. WormBook. 2005;1.7.1:1–17. doi:10.1895/wormbook.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003;5(1):R1.

Ferguson EL, Sternberg PW, Horvitz HR. A genetic pathway for the specification of the vulval cell lineages of *Caenorhabditis elegans*. Nature. 1987;326(6110):259–267.

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature. 2000;408(6810):325–330.

Galvin BD, Kim S, Horvitz HR. *Caenorhabditis elegans* genes required for the engulfment of apoptotic corpses function in the cytotoxic cell deaths induced by mutations in lin-24 and lin-33. Genetics. 2008;179(1):403–417.

Gleason JE, Szyleyko EA, Eisenmann DM. Multiple redundant Wnt signaling components function in two processes during *C. elegans* vulval development. Dev Biol. 2006;298(2):442–457.

Grants JM, Ying LT, Yoda A, You CC, Okano H, Sawa H, Taubert S. The mediator kinase module restrains epidermal growth factor receptor signaling and represses vulval cell fate specification in *Caenorhabditis elegans*. Genetics. 2016;202(2):583–599.

Green JL, Inoue T, Sternberg PW. The *C. elegans* ROR receptor tyrosine kinase, CAM-1, non-autonomously inhibits the Wnt pathway. Development. 2007;134(22):4053–4062.

Green JL, Inoue T, Sternberg PW. Opposing Wnt pathways orient cell polarity during organogenesis. Cell. 2008;134(4):646–656.

Greenwald I. LIN-12/Notch signaling in *C. elegans*. WormBook. 2005;1.10.1:1–16. doi:10.1895/wormbook.

Greenwald I, Kovall R. Notch signaling: genetics and structure. WormBook. 2013;1.10.2:1–28. doi:10.1895/wormbook.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:D140–D144.

Grosshans H, Johnson T, Reinert KL, Gerstein M, Slack FJ. The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in *C. elegans*. Dev Cell. 2005;8(3):321–330.

Hagedorn EJ, Sherwood DR. Cell invasion through basement membrane: the anchor cell breaches the barrier. Curr Opin Cell Biol. 2011;23(5):589–596.

Hart AH, Reventar R, Bernstein A. Genetic analysis of ETS genes in *C. elegans*. Oncogene. 2000;19(55):6400–6408.

Hoier EF, Mohler WA, Kim SK, Hajnal A. The *Caenorhabditis elegans* APC-related gene apr-1 is required for epithelial cell migration and Hox gene expression. Genes Dev. 2000;14(7):874–886.

Horvitz HR, Brenner S, Hodgkin J, Herman RK. A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol Gen Genet. 1979;175(2):129–133.

Hrach HC, O'Brien S, Steber HS, Newbern J, Rawls A, Mangone M. Transcriptome changes during the initiation and progression of Duchenne muscular dystrophy in *Caenorhabditis elegans*. Hum Mol Genet. 2020;29(10):1607–1623.

Hwang HY, Olson SK, Brown JR, Esko JD, Horvitz HR. The *Caenorhabditis elegans* genes sqv-2 and sqv-6, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. J Biol Chem. 2003;278(14):11735–11738.

Inoue T, Oz HS, Wiland D, Gharib S, Deshpande R, Hill RJ, Katz WS, Sternberg PW. *C. elegans* LIN-18 is a Ryk ortholog and functions in parallel to LIN-17/Frizzled in Wnt signaling. Cell. 2004;118(6):795–806.

Inoue T, Sherwood DR, Aspöck G, Butler JA, Gupta BP, Kirouac M, Wang M, Lee P-Y, Kramer JM, Hope I, *et al.* Gene expression markers for *Caenorhabditis elegans* vulval cells. Gene Expr Patterns. 2002;2(3–4):235–241.

Inoue T, Wang M, Ririe TO, Fernandes JS, Sternberg PW. Transcriptional network underlying *Caenorhabditis elegans* vulval development. Proc Natl Acad Sci USA. 2005;102(14):4972–4977.

Kidd AR, III, Muniz-Medina V, Der CJ, Cox AD, Reiner DJ. The *C. elegans* Chp/Wrch ortholog CHW-1 contributes to LIN-18/Ryk and LIN-17/Frizzled signaling in cell polarity. PLoS One. 2015;10(7):e0133226.

Kim W, Underwood RS, Greenwald I, Shaye DD. OrthoList 2: a new comparative genomic analysis of human and *Caenorhabditis elegans* genes. Genetics. 2018;210(2):445–461.

Kimble J, Crittenden SL. Germline proliferation and its control. WormBook. 2005;1.13.1:1–14. doi:10.1895/wormbook

Kirouac M, Sternberg PW. cis-Regulatory control of three cell fate-specific genes in vulval organogenesis of *Caenorhabditis elegans* and *C. briggsae*. Dev Biol. 2003;257(1):85–103.

Kishore RS, Sundaram MV. ced-10 Rac and mig-2 function redundantly and act with unc-73 trio to control the orientation of vulval cell divisions and migrations in *Caenorhabditis elegans*. Dev Biol. 2002;241(2):339–348.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079.

Luo KL, Underwood RS, Greenwald I. Positive autoregulation of *lag-1* in response to LIN-12 activation in cell fate decisions during *C. elegans* reproductive system development. Development. 2020;147: dev193482.

Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJ. Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. Genome Res. 2008;18(12):2005–2015.

Miller LM, Gallegos ME, Morisseau BA, Kim SK. lin-31, a *Caenorhabditis elegans* HNF-3/fork head transcription factor homolog, specifies three alternative cell fates in vulval development. Genes Dev. 1993;7(6):933–947.

Miller LM, Hess HA, Doroquez DB, Andrews NM. Null mutations in the lin-31 gene indicate two functions during *Caenorhabditis elegans* vulval development. Genetics. 2000;156(4):1595–1602.

Miller LM, Waring DA, Kim SK. Mosaic analysis using a ncl-1 (+) extrachromosomal array reveals that lin-31 acts in the Pn.p cells during *Caenorhabditis elegans* vulval development. Genetics. 1996; 143(3):1181–1191.

Myers TR, Greenwald I. Wnt signal from multiple tissues and lin-3/EGF signal from the gonad maintain vulval precursor cell competence in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2007; 104(51):20368–20373.

Nakdimon I, Walser M, Frohli E, Hajnal A. PTEN negatively regulates MAPK signaling during *Caenorhabditis elegans* vulval development. PLoS Genet. 2012;8(8):e1002881.

Oka T, Toyomura T, Honjo K, Wada Y, Futai M. Four subunit a isoforms of *Caenorhabditis elegans* vacuolar H+-ATPase. Cell-specific expression during development. J Biol Chem. 2001;276(35): 33079–33085.

Parry DH, Xu J, Ruvkun G. A whole-genome RNAi Screen for *C. elegans* miRNA pathway genes. Curr Biol. 2007;17(23):2013–2022.

Poulin G, Dong Y, Fraser AG, Hopper NA, Ahringer J. Chromatin regulation and sumoylation in the inhibition of Ras-induced vulval development in *Caenorhabditis elegans*. EMBO J. 2005;24(14): 2613–2623.

Priess JR. Notch signaling in the *C. elegans* embryo. WormBook. 2005; 1.4.1:1–16. doi:10.1895/wormbook

Pujol N, Bonnerot C, Ewbank JJ, Kohara Y, Thierry-Mieg D. The *Caenorhabditis elegans* unc-32 gene encodes alternative forms of a vacuolar ATPase a subunit. J Biol Chem. 2001;276(15): 11913–11921.

Rasmussen NR, Dickinson DJ, Reiner DJ. Ras-dependent cell fate decisions are reinforced by the RAP-1 small GTPase in *Caenorhabditis elegans*. Genetics. 2018;210(4):1339–1354.

Rasmussen NR, Reiner DJ. Nuclear translocation of the tagged endogenous MAPK MPK-1 denotes a subset of activation events in *C. elegans* development. J Cell Sci. 2021;134:jcs258456.

Ririe TO, Fernandes JS, Sternberg PW. The *Caenorhabditis elegans* vulva: a post-embryonic gene regulatory network controlling organogenesis. Proc Natl Acad Sci USA. 2008;105(51):20095–20099.

Schlager B, Roseler W, Zheng M, Gutierrez A, Sommer RJ. HAIRY-like transcription factors and the evolution of the nematode vulva equivalence group. Curr Biol. 2006;16(14):1386–1394.

Shemer G, Podbilewicz B. LIN-39/Hox triggers cell division and represses EFF-1/fusogen-dependent vulval cell fusion. Genes Dev. 2002;16(24):3136–3141.

Shephard F, Adenle AA, Jacobson LA, Szewczyk NJ. Identification and functional clustering of genes regulating muscle protein degradation from amongst the known *C. elegans* muscle mutants. PLoS One. 2011;6(9):e24686.

Shin H, Braendle C, Monahan KB, Kaplan REW, Zand TP, Mote FS, Peters EC, Reiner DJ. Developmental fidelity is imposed by genetically separable RalGEF activities that mediate opposing signals. PLoS Genet. 2019;15(5):e1008056.

Shin H, Kaplan REW, Duong T, Fakieh R, Reiner DJ. Ral signals through a MAP4 Kinase-p38 MAP kinase cascade in *C. elegans* cell fate patterning. Cell Rep. 2018;24(10):2669–2681.e5.

Shin H, Reiner DJ. The signaling network controlling *C. elegans* vulval cell fate patterning. J Dev Biol. 2018;6:30.

Speese S, Petrie M, Schuske K, Ailion M, Ann K, Iwasaki K, Jorgensen EM, Martin TF. UNC-31 (CAPS) is required for dense-core vesicle but not synaptic vesicle exocytosis in *Caenorhabditis elegans*. J Neurosci. 2007;27(23):6150–6162.

Spiri S, Berger S, Mereu L, DeMello A, Hajnal A. Reciprocal EGFR signaling in the anchor cell ensures precise inter-organ connection during *Caenorhabditis elegans* vulval morphogenesis. Development. 2022;149(1):dev199900.

Sternberg PW. Vulval development. In: The C. elegans Research Community WormBook, editor. WormBook. 2005.

Steven R, Kubiseski TJ, Zheng H, Kulkarni S, Mancillas J, Ruiz Morales A, Hogue CW, Pawson T, Culotti J. UNC-73 activates the Rac GTPase and is required for cell and growth cone migrations in *C. elegans*. Cell. 1998;92(6):785–795.

Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Dev Biol. 1977;56(1):110–156.

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605–D612.

Takacs-Vellai K, Vellai T, Chen EB, Zhang Y, Guerry F, Stern MJ, Muller F. Transcriptional control of Notch signaling by a HOX and a PBX/EXD protein during vulval development in *C. elegans*. Dev Biol. 2007;302(2):661–669.

Tan PB, Lackner MR, Kim SK. MAP kinase signaling specificity mediated by the LIN-1 Ets/LIN-31 WH transcription factor complex during *C. elegans* vulval induction. Cell. 1998;93(4):569–580.

Tennessen JM, Gardner HF, Volk ML, Rougvie AE. Novel heterochronic functions of the *Caenorhabditis elegans* period-related protein LIN-42. Dev Biol. 2006;289(1):30–43.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–515.

Ulm EA, Sleiman SF, Chamberlin HM. Developmental functions for the *Caenorhabditis elegans* Sp protein SPTF-3. Mech Dev. 2011; 128(7–10):428–441.

Underwood RS, Deng Y, Greenwald I. Integration of EGFR and LIN-12/Notch signaling by LIN-1/Elk1, the Cdk8 kinase module, and SUR-2/Med23 in vulval precursor cell fate patterning in *Caenorhabditis elegans*. Genetics. 2017;207(4):1473–1488.

Wagmaister JA, Gleason JE, Eisenmann DM. Transcriptional upregulation of the *C. elegans* Hox gene lin-39 during vulval cell fate specification. Mech Dev. 2006a;123(2):135–150.

Wagmaister JA, Miley GR, Morris CA, Gleason JE, Miller LM, Kornfeld K, Eisenmann DM. Identification of cis-regulatory elements from the *C. elegans* Hox gene lin-39 required for embryonic expression

and for regulation by the transcription factors LIN-1, LIN-31 and LIN-39. Dev Biol. 2006b;297(2):550–565.

Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc. 2015;10(6):823–844.

Yang L, Sym M, Kenyon C. The roles of two *C. elegans* HOX co-factor orthologs in cell migration and vulva development. Development. 2005;132(6):1413–1428.

Yemini E, Lin A, Nejatbakhsh A, Varol E, Sun R, Mena GE, Samuel ADT, Paninski L, Venkatachalam V, Hobert O. NeuroPAL: a multicolor atlas for whole-brain neuronal identification in *C. elegans*. Cell. 2021;184(1):272–288.e11.

Yoo AS, Bais C, Greenwald I. Crosstalk between the EGFR and LIN-12/Notch pathways in *C. elegans* vulval development. Science. 2004; 303(5658):663–666.

Yoo AS, Greenwald I. LIN-12/Notch activation leads to microRNA-mediated down-regulation of Vav in *C. elegans*. Science. 2005; 310(5752):1330–1333.

Communicating editor: M. Zetka