Tissue-Specific Transcription Footprinting Using RNA Pol DamID (RAPID) in Caenorhabditis elegans

Georgina Gómez-Saldivar,*1 Jaime Osuna-Luque,*1,‡ J-M, Jennifer I. Semple,* Dominique A. Glauser,*2 Sophie Jarriault,*1 and Peter Meister*1

*Department of Biology, University of Fribourg, Fribourg 1700, Switzerland, †Cell Fate and Nuclear Organization, Institute of Cell Biology, University of Bern, 3012, Switzerland, ‡IGBMC, DBSC Dept/INSERM U1258/ CNRS UMR7104/Université de Strasbourg, 67400 France, and §Graduate School for Cellular and Biomedical Sciences, University of Bern, 3012, Switzerland

ORCID IDs: 0000-0003-3264-3222 (G.G.-S.); 0000-0001-5701-4129 (J.O.-L.); 0000-0001-7221-9560 (J.I.S.); 0000-0001-7221-9560 (D.A.G.); 0000-0003-2847-1675 (S.J.); 0000-0002-6230-4216 (P.M.)

ABSTRACT Differential gene expression across cell types underlies development and cell physiology in multicellular organisms. Caenorhabditis elegans is a powerful, extensively used model to address these biological questions. A remaining bottleneck relates to the difficulty to obtain comprehensive tissue-specific gene transcription data, since available methods are still challenging to execute and/or require large worm populations. Here, we introduce the RNA Polymerase DamID (RAPID) approach, in which the Dam methyltransferase is fused to a ubiquitous RNA polymerase subunit to create transcriptional footprints via methyl marks on the DNA of transcribed genes. To validate the method, we determined the polymerase footprints in whole animals, in sorted embryonic blastomeres and in different tissues from intact young adults by driving tissue-specific Dam fusion expression. We obtained meaningful transcriptional footprints in line with RNA-sequencing (RNA-seq) studies in whole animals or specific tissues. To challenge the sensitivity of RAPID and demonstrate its utility to determine novel tissue-specific transcriptional profiles, we determined the transcriptional footprints of the pair of XXX neuroendocrine cells, representing 0.2% of the somatic cell content of the animals. We identified 3901 candidate genes with putatively active transcription in XXX cells, including the few previously known markers for these cells. Using transcriptional reporters for a subset of new hits, we confirmed that the majority of them were expressed in XXX cells and identified novel XXX-specific markers. Taken together, our work establishes RAPID as a valid method for the determination of RNA polymerase footprints in specific tissues of C. elegans without the need for cell sorting or RNA tagging.

KEYWORDS single cell type gene expression analysis; C. elegans targeted DamID; RNA polymerase footprinting; ONT long read sequencing

DIFFERENTIAL gene expression across cell types encompasses both key determinants and markers of cell identity. Cataloging these differences can provide critical insights and entry points in research aiming at elucidating the mechanisms controlling fundamental biological processes, such as organismal development and cell/tissue physiology. Caenorhabditis elegans is a widely used model animal, particularly well-suited for integrative studies bridging our understanding across the molecular, cellular, and organismal levels. With its transparent body, C. elegans was the first animal used to analyze tissue-specific transcription in vivo with GFP reporters (Chalfie et al. 1994), an approach still extensively used. In contrast to the versatility of the model for individual gene expression analysis, more holistic approaches such as tissue-specific transcriptomics still remain relatively challenging, in particular in postembryonic animals due to the tough cuticle and the difficulty to isolate intact tissue or cell types.

Two main general strategies have been developed to analyze specific tissues/cell types in C. elegans. A first general strategy involves the purification of tissue-specific messenger...
RNAs (mRNAs) from whole animals, which we will call “RNA tagging/pulling” hereafter. The most widely used RNA tagging/pulling method rely on the immunoprecipitation of a tagged poly-A binding protein-1 (FLAG:PA-B-1) expressed in a specific cell type and cross-linked to RNA (Roy et al. 2002; Kunitomo et al. 2005; Pauli et al. 2006; Hrach et al. 2020). One of the latest versions of this method is referred to as polyA tagging and sequencing (PAT-seq; Blazie et al. 2015, 2017). The PAB-based method is technically demanding, in particular for the analysis of a small number of cells (Takayama et al. 2010) and has been shown to be associated with significant background noise (Von Stettina et al. 2007; Ma et al. 2016). Two more recent alternatives to PAB-based mRNA tagging are tissue-specific translating ribosome affinity purification (TRAP; Gracia and Calarco 2017; Rhoades et al. 2019) and trans-splicing-based RNA tagging (SRT; Ma et al. 2016). TRAP analysis focuses on ribosome-engaged mRNAs recovered after cross-linking and immunoprecipitation of a tagged ribosome subunit. TRAP allowed the identification of genes expressed in a specific neuron type representing only two cells per animal (Rhoades et al. 2019). SRT uses a modified SL1 splice leader expressed in target tissues, which is trans-spliced to the transcripts by the cellular machinery (Ma et al. 2016). While SRT bypasses the noise inherent in immunoprecipitation procedures, it has so far only been applied to large tissues and the approach is limited to SL1-associated transcripts (62% of the C. elegans genes; Yang et al. 2017).

The second general strategy relies on animal disruption and cell isolation, or dissociation, followed by the in vitro culturing or sorting of labeled cells (Von Stettina et al. 2007; Spencer et al. 2011), or nuclei (Haenni et al. 2012; Steiner et al. 2012) before transcriptomic and analysis, and which will call “dissociation-based” methods. The tough cuticle in larval stages and a fortiori in adults constitutes a significant obstacle, and initial studies focused on embryonic cells which were more easily dissociated. More recent protocols combining FACS and RNA-seq were successfully used to analyze major tissues (Kaletsky et al. 2018) as well as neuronal subsets in adults (Wang et al. 2015; down to six neurons per animal; Kaletsky et al. 2016). Combined with single-cell RNA-sequencing (scRNA-seq), large-scale dissociation-based studies can address transcript profiles in multiple cell types at the same time (Cao et al. 2017; Packer et al. 2019), including the analysis of individual neuron types (Hammarlund et al. 2018; Lorenzo et al. 2020). However, how efficiently specific cell types, especially rare cells, can be purified varies depending on their morphologies, how fragile they are, and the developmental stage considered, suggesting that dissociation-independent methods could be useful complementary approaches.

We investigated whether we could use DNA adenine methylation identification (DamID) to footprint actively transcribed genes in specific cell types in whole worms. This approach has been successfully used in Drosophila to determine transcribed genes in rare brain cell types (Southall et al. 2013). DamID relies on the low-level expression of a fusion between the Escherichia coli Dam methyltransferase and a protein of interest, here a subunit of the RNA polymerase (Pol). Binding of the latter to DNA leads to the methylation of GATC sites in the vicinity of the binding site (Figure 1A). After DNA extraction, methylated GATCs are specifically cleaved by the restriction enzyme DpnI and amplified using adapter-mediated PCR before sequencing (Supplemental Material, Figure S1A). The adaptation of the method to C. elegans seemed promising since endogenous adenine methylation/demethylation is very rare in worms and not targeted to GATC motifs (Greer et al. 2015). Moreover, the GATC site frequency is expected to provide good spatial resolution. Indeed, C. elegans has 269,049 GATC sequences per haploid genome, corresponding to an average of one site for every 374 bp and a median of 210 bp (Gómez-Saldívar et al. 2016a). In C. elegans, DamID has been used to study the genomic footprint of the DAF-16 transcription factor (Schuster et al. 2010), and large-scale interactions between the genome and the nuclear periphery (Towbin et al. 2012; Sharma et al. 2014; Cabianca et al. 2019; Harr et al. 2020).

Here, we describe the RNA Polymerase DamID (RAPID) approach for transcriptional footprinting in specific tissues in C. elegans in both embryonic blastomeres and in young adults. Using a fusion between a small subunit present in all three RNA Pols, we show that the technique can be used on both fluorescantly sorted blastomere cells and DNA isolated from entire young adults using cell-type-specific expression generated by Cre/lox. To test the versatility of the method, we determined the Pol footprints in three different tissues at each stage. In young adults, these tissues represent between 10 and 0.2% of the somatic cells of the animal. We show that meaningful transcriptional patterns can be recovered using this technique, in line with previously used RNA tagging/pulling and dissociation-based methods. We further explore how this technique can be used to discover tissue-specific markers and report the identification of eight new reporters expressed in adult XXX cells. Additionally, as a result of the phasing-out of older sequencers able to sequence amplicons of different sizes as produced by DamID-sequencing, we show that new long-read sequencing methods can be used to sequence DamID libraries, which makes it possible to carry out the experiments, from DNA extraction to amplicon sequencing, in less than a week, without any external sequencing facility.

Materials and Methods

Plasmids and transgenic strains

All plasmids and worm strains are listed in Tables S1 and S2. Tissue-specific DamID plasmids were generated using Gibson assembly. All DamID plasmids were integrated as single copies using MosSCI either on chromosome II or IV (see details in Table S1; Frøkjaer-Jensen et al. 2008). Strains and plasmids are available upon request.

Worm growth

Worms were grown on solid NGM, seeded with OP50 bacteria for maintenance culture and genetic crosses. All worm
cultures were grown at 20°C. For cell sorting, animals were grown on peptone plates seeded with HB101, synchronized twice by hypochlorite treatment. Young adults containing up to 10 embryos were recovered 66 hr after plating synchronized L1 larvae. For DamID experiments in young adults, worms were grown on NGM seeded with Dam-negative E. coli GM48 for at least two generations. Around 4000 synchronized L1s were seeded onto 100-mm plates (1000–1200 per plate) and collected 53 hr later. Worms were washed extensively with M9 (at least 10 times) and distributed in aliquots of 30 μl, removing the excess liquid. Samples were snap-frozen and stored at −80°C.

**Sorting of embryonic blastomeres**

Synchronized gravid adult hermaphrodites containing 8–10 eggs were treated with hypochlorite. Eggs were then incubated for 3 hr in M9 at 25°C until they reached the 1.5-fold stage. Eggs were transferred to 500 μl egg buffer (25 mM HEPES pH 7.3, 118 mM NaCl, 48 mM KCl, 2 mM CaCl2, 2 mM MgCl2; Sigma-Aldrich, St. Louis, MO) and pelleted 1 min at 2000 rpm. The supernatant was then aspirated, leaving 100 μl of buffer with the pellet. Then, 500 U of chitinase (C8241-25U; Sigma-Aldrich, St. Louis, MO) was added and the mixture was resuspended and further incubated for 1 hr at room temperature. Chitinase was neutralized with 800 μl Leibovitz medium. Disaggregated embryos were recovered by centrifugation at 3000 rpm for 5 min at 4°C. Embryos were then dissociated into isolated blastomeres by pipetting up and down with a P1000 pipette, up to 150 times until dissociation was complete. The cell population was filtered with a Millex-SV syringe 5-μm filter (SLSV025LS; Millipore, Bedford, MA). A total of 500 fluorescent cells (one technical replicate) were sorted using a BD FACSaria Fusion (8000 events/second; 85-μm nozzle) in sterile PCR tubes containing 1 μl of pick buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl2, 137 mM NaCl). Collected samples were then frozen in liquid nitrogen times before further processing. After DamID processing and PCR (see below), two technical replicates were pooled for sequencing library preparation.

**DamID amplification, library preparation, and sequencing**

For sorted cells, frozen samples were lysed by addition of 2 μl of lysis buffer (10 mM TrisAc, 10 mM MgAc, 50 mM KAc, 0.67% Tween 20, 0.67% Igepal + 1 mg/ml Proteinase K) and incubated for 2 hr at 60°C before Proteinase K inactivation at 95°C for 15 min. DamID amplicons were obtained as previously described (Gomez-Saldivar et al. 2016b), using 30 PCR cycles.

For young adults, DamID was performed on 500 ng gDNA extracted from the animals using DNAeasy Blood and Tissue Kit (#69504; QIAGEN, Valencia, CA). Two replicates for each stage and cell type were processed. DamID amplicons were obtained as previously described (Gomez-Saldivar et al. 2016a), with 20 PCR cycles for worm-wide DamID, 22 PCR cycles for muscle DamID, 22–24 PCR cycles for intestine DamID, and 24–26 PCR cycles for DamID of XXX cells. New Illumina patterned flow cell technology does not allow sequencing of amplicons larger than 600 bp. Both types of DamID PCR amplicons were therefore sequenced using nanopore sequencing. After AMPure XP (Beckman Coulter) purification with 1.8 bead volume, DamID PCR amplicons were directly used for nanopore library barcoding and preparation using NBD-104 and LSK-109 kits (Oxford Nanopore Technologies). Libraries were sequenced to obtain at least 1 million reads per library on MinION, using R9.4.1 flow cells. We compared the performance of both sequencing approaches by sequencing the same libraries with both old paired-end HiSeq2500 Illumina flow cells and nanopore sequencers. Coverage obtained by both approaches were similar (see snapshot of mapped reads in Figure S1E). However, nanopore sequencing allowed the direct sequencing of longer amplicons (Figure S1F), most likely because Illumina cluster amplification does not work well on molecules longer than 800 bp, even on old-generation (HiSeq2500) flow cells. We compared whole animal DamID sample libraries generated using previously described strains expressing either a free GFP-Dam fusion or a perinuclear lamin fusion (Sharma et al. 2014). Statistical comparison of both techniques at the single restriction fragment level gave a Pearson correlation coefficient between short paired-end (PE) and long reads between 0.84 and 0.86 (Figure S1G).

**Bioinformatic analysis**

Nanopore sequences were base-called and demultiplexed using guppy 3.6 with the high-accuracy model, before mapping to the ce11 genome using minimap2 (Li 2018). Reads were considered as DamID amplicons when both ends mapped ±8 bp from a genomic GATC motif. Filtered libraries were then used to call Pol footprinting values using the damidseq pipeline (Marshall and Brand 2015), using bam files (parameters: --bamfiles--extend_reads = 0). The damidseq pipeline normalizes for GATC accessibility using a free Dam (in our case Dam::GFP) promoter. DamID amplicons were therefore sequenced using nanopore sequencing (ChIP-seq) tracks, fastq files were downloaded from the Gene Expression Omnibus (GEO), mapped to ce11, normalized for sequencing depth with a pseudocount of 8 and log2-normalized to the input. Final figure construction was made using IGV and Adobe Illustrator.

**Array construction and microscopy**

Reporter plasmids were assembled using the three-fragment MultiSite Gateway system (Thermo Fisher Scientific). Promoters of candidate genes were PCR amplified from N2 genomic DNA. Primer sequences and promoter lengths are described in Table S3. Candidate promoters were cloned into pDONR-P4-P1R vector (Invitrogen, Carlsbad, CA) by
Gateway cloning, generating slot 1 entry vectors (Table S2). Final expression plasmids were generated by Gateway cloning between promoters (slot 1), mNeonGreen fluorescent ORF (slot 2; dg353; Hostettler et al. 2017), and pDEST-R4-R3. The control localization/expression plasmid [sdf-9p::NLS::wrmScarlet] was generated through a Gateway combination reaction between dg801 (slot 1), dg651 (slot 2; Marques et al. 2019), pMH473 (slot 3), and pDEST-R4-R3. Finally, promoter::mNeonGreen and sdf-9p::NLS::wrmScarlet plasmids were co-injected into N2 young adult hermaphrodites at 20 ng/μl each, with 20 ng/μl of dg9 (unc-122p::RFP; red coelomocyte) as a co-injection marker (#8938; Addgene; Miyabayashi et al. 1999). Two stable lines of each candidate promoter were selected (Table S1). For imaging, young adult animals identified based on the red coelomocyte reporter expression were transferred to a 2% agarose pad containing 0.01% sodium azide. Worms were then imaged on a Zeiss Axiovert microscope with a ×20 objective driven by Visiview, using a Photometrics Coolsnap Myo CCD camera with GFP and RFP settings (60 planes spaced by 2 μm), plus a middle plane section in DIC settings. Using Fiji, the acquired optical stacks were then partially z-projected to capture the worm section containing the XXX cells, identified using the sdf-9p::NLS::wrmScarlet signal. Final figure construction was made using Adobe Illustrator. For each promoter except asp-9, two independent extrachromosomal arrays were scored in at least five animals.

Identification of detected and unique tissue-specific genes and validation

Gene coverage of RAPID profiles are listed in Table S4 (blastosomeres) and Table S5 (young adults) (Sheet 1–2). Detected genes in a tissue were defined as genes that are significantly expressed [false discovery rate (FDR) < 0.05] in one replicate of one tissue (Table S4, sheet 3 and Table S5, sheet 3). Genes consistently detected in a tissue were defined as genes that are significantly expressed (FDR <0.05) in both replicates of one tissue (Table S4, sheet 4 and Table S5, sheet 4). Representative examples of consistently detected genes used in Figure 2C were selected from the CeNGEN project (C. elegans Neuronal Gene Expression Map & Network; https://cengen.shinyapps.io/SCeNGEA/; Hammarlund et al. 2018) and visualized using IGV (https://igv.org/).

Unique tissue-specific genes were defined as genes significantly and differentially expressed relative to the other two tissues (Table S4, sheet 5 and Table S5, sheet 5). Analysis of the uniqueness of the genes from each tissue was performed using the JavaScript library jvenn (http://bioinfo.genotoul.fr/jvenn).

To ensure an unbiased comparison between genes identified as expressed in muscle with RAPID and the set of genes identified with SRT (Ma et al. 2016), only SL1 trans-spliced genes were selected, based on the intersection with the 10,589 mRNAs annotated as SL1-trans-spliced by modENCODE (Allen et al. 2011), generating a subset of 3477 genes. For the comparison between the genes identified in the intestine with RAPID and the set of genes identified with fluorescence-activated nuclear sorting (FANS; Haenni et al. 2012), downregulated genes were eliminated from the list and only genes found as expressed and upregulated (total 9169 genes) were analyzed. For the comparison between the genes identified with RAPID and the set of genes identified with FACS (Kaletsky et al. 2018) in muscle and intestine, we directly used the expressed genes list from Table S1, which contained 7690 and 9603 genes, respectively. Chi-squared statistical tests were performed using the GraphPad QuickCalc (https://www.graphpad.com/quickcalc/; accessed August 2020).

Tissue-specific prediction analysis

Tissue expression prediction analyses were performed using the top 500 statistically significant genes that were uniquely enriched in muscle and intestine (Table S5, sheet 6), selecting a multigene search within the Tissue-specific Expression Predictions for C. elegans program, version 1.0 (http://worm-tissue.princeton.edu/search/multi). For the tissue expression prediction test using the RAPID muscle-SL1 group, the parameters published by Ma et al. (2016) were used.

Gene ontology analysis

Gene ontology (GO) analyses were performed on unique tissue-specific gene lists. GO terms and q-values were obtained using gProfiler (https://biit.cs.ut.ee/gprofiler/gost; version e100_eg47_p14_7733820) with the g:SCS multiple testing correction method, applying a significance threshold of 0.05 (Table S5, sheet 7; Raudvere et al. 2019).

Functional enrichment analysis for genes detected with different methods was performed using WormCat (http://www.wormcat.com/index). Annotations classified into category 2 were visualized with heatmap diagrams.

Candidate selection for reporter analysis

Candidate genes were selected for follow-up promoter analyses with two goals in mind: demonstrating that they are indeed transcribed in XXX cells, and identifying markers with expression restricted to XXX cells. Since XXX cells represent only 0.2% of the C. elegans cellular content, we reasoned that XXX-specific markers should produce very little or no signal in the whole animal samples. Therefore, out of the 862 XXX-enriched genes, with a significant RAPID signal in XXX, but not in muscle or in intestine samples, we further removed genes detected in at least one worm-wide sample. We obtained a refined list of 275 XXX-marker gene candidates, which still included the known XXX cell markers daf-9 and eak-4 (Table S5, sheet 8). From this candidate list, we selected dhs-17 as an uncharacterized gene with a plausible link to XXX cell function, plus 11 random candidates. To limit caveats related to operons, we excluded a candidate if another transcript was located <200 bp upstream. The XXX-specific RAPID signals (RNA Pol occupancy values) in the
12-gene subset (average = 1.35, SD = 0.59) were similar to those in the starting 275-gene set (average = 1.47, SD = 0.64; \( P = 0.16 \) by Student’s t-test).

**Data availability**

Embryonic and adult DamID-sequencing data are available under the GEO accession number GSE157418. Supplemental material available at figshare: https://doi.org/10.25386/genetics.13072355.

**Results**

**RNA Pol footprinting using RAPID**

The RAPID approach relies on the expression of trace levels of the *E. coli* Dam fused to a subunit of the RNA Pols, and the analysis of its DNA occupancy to evaluate the transcriptional state of genes. We first tested whether we could use AMA-1, the largest catalytic subunit of RNA Pol II, as reported in targeted DamID, done in neuronal lineages in *Drosophila* (Southall et al. 2013). However, the Dam signal was too weak, possibly due to the fact that the expression level of this very large fusion protein under transcriptional control of the uninduced *hsp-16.2* promoter was too low (not shown), or that the localization of the Dam domain within the *C. elegans* Pol complex was not favorable to access its DNA substrate. We therefore replaced AMA-1 with RPB-6, the *C. elegans* homolog of the RNA Pol subunit F present in all three RNA Pols (D. Katsanos, M. Ferrando-Marco, I. Razzaq, T. Southall, and M. Barkoulas; unpublished data). RPB-6 together with AMA-1 and RPB-2, forms a “clamp” that retains DNA near the active center of Pol II (Cramer et al. 2000), stabilizing the transcription on the DNA template. In contrast to AMA-1 attempts, we obtained a strong DamID amplicon PCR signal in animals carrying a single-copy transgene of the *rpb-6::Dam* fusion. As these animals had not been subjected to heat shock, they therefore expressed only low levels of the Dam fusion, yet the methylation levels were sufficient to perform DamID. Amplicons were sequenced using an ONT MinION nanopore device (see Materials and Methods for details and experimental validation of the technique). After signal normalization with GFP::Dam data to control for overall chromatin accessibility, we could detect general patterns of methylation that were consistent with Pol-dependent methylation (Figure 1, D–F) and a good correlation at the gene level between replicates (Figure 1, B, D, and F).

We compared the RAPID profiles with published recent RNA Pol II ChIP-seq data sets in young adult animals (Garrido-Lecca et al. 2016; Kalinava et al. 2017; Miki et al. 2017). Visually, the patterns appeared similar (see examples of profiles in Figure 1, D and F). Even if the nature of the data are not identical, we performed a comparative analysis between the data sets for all genes, down-scaling the resolution of the ChIP-seq data to GATC fragment (Figure 1B). RAPID and RNA Pol ChIP-seq show a significant correlation (\( R = 0.38–0.46 \)). Overall, inter-replicate correlations were higher between the RAPID replicates (\( R = 0.86 \)) and between the RNA Pol II ChIP-seq experiments (\( R = 0.63–0.97 \)). The difference between the RAPID and RNA Pol II ChIP-seq profiles most likely results from the different approach to recover DNA and from the fact that RAPID in adults represent a picture of cumulative transcriptional activity, while the RNA Pol II ChIP-seq represents a Pol occupancy snapshot at the young adult stage.

To get a more quantitative understanding of the relation between RAPID and transcript abundance, we compared RAPID with RNA-seq data, using one of the published young adult RNA-seq data sets (Miki et al. 2017). We ranked all genes based on their RAPID signal and calculated the average RNA-seq signal across 30 equally sized bins. We observe a clear correlation between RAPID and RNA-seq: genes with high RAPID signal are highly expressed as determined by RNA-seq, while lowly expressed genes harbor low RAPID levels (Figure 1C).

To understand how the methylation signal spreads along the body of the genes, we constructed metagene plots over all 20,000 *C. elegans* genes for RAPID and RNA Pol II ChIP-seq experiments. ChIP-seq signals show a characteristic profile, as the Pol accumulates at the transcription start site (TSS) and the transcription end site (TES) (Figure 1E). RAPID signal increases steadily from the TSS to the TES, where it peaks before decreasing from the TES onward (Figure 1E). This 3’ end accumulation is similar to the one observed in ChIP-seq. In contrast, the difference between ChIP-seq and RAPID profiles (absence of a 5’ peak in RAPID) is very likely due to the localization of the RPB-6::Dam fusion inside the RNA Pol complex. Located on the opposite side of the AMA-1/RPB-2 complex relative to the DNA strands, the presence of the preinitiation complex at the promoter greatly restricts access to DNA [for review, see Cramer (2004); Schier and Taatjes 2020]. Once the Pol switches to elongation, leaving the preinitiation complex on the promoter, Dam is likely to gain access to the DNA and efficiently methylates the transcribed region, as observed in the RAPID profiles.

We also examined the profile in genes transcribed by Pol I and Pol III. As expected from the differential inclusion of AMA-1 and RPB-6 subunits in the different Pols, RAPID with RPB-6 also labeled genes transcribed by RNA Pol I and III. RAPID showed high enrichments on the ribosomal RNA genes at the end of chromosome I (Figure S2A), as well as a majority of previously characterized small nucleolar RNAs (snoRNAs) transcribed by RNA Pol III (46 out of 57; Ikegami and Lieb 2013). snoRNA are sometimes located in introns of Pol II-transcribed genes, making it difficult to differentiate whether the RAPID signal originates from the snoRNA transcription or from the overlapping gene. Nevertheless, 24 out of 44 RAPID-positive snoRNA genes could be unambiguously assigned to Pol III transcription as there was no overlap (Figure S2B). In contrast, the corresponding signal was markedly weaker in the young adult RNA Pol II ChIP-seq data sets and very high for the RNA Pol III subunit RPC-1 (performed in embryos, Figure S2B; Ikegami and Lieb 2013).

RNA Polymerase Footprinting in Nematodes 935
Taken together, these data show that the RAPID method is suitable to reveal RNA Pol footprints, serving as indirect indications of transcriptional activity by the three different RNA Pols.

**RAPID in FACS-sorted embryonic cells**

Since RAPID could label transcribed genes in whole animals, our next goal was to adapt the method for tissue-specific analysis. We first explored whether RAPID could be used to footprint transcription in specific cell types of the embryo purified using fluorescence sorting. Using strains constitutively expressing Dam fusions, we sorted 1000 cells for 3 different tissues of various abundance: body wall muscle, intestine, and the rectal Y cell (80, 20, and 1 cell per embryo, respectively; Figure S3A and Table S4). RAPID could produce

---

**Figure 1** RAPID: RNA Pol II DamID scheme. (A) Dam::rpb-6 bound to all three RNA polymerase types modifies proximal GATC motifs. DamID analysis provides RNA polymerase footprints in vivo. (B) Gene-level correlations in young adult whole animals of RAPID and RNA polymerase ChIP-seq using different antibodies [a: Garrido-Lecca et al. (2016); b: Kalinava et al. (2017); c: Miki et al. (2017)]. (C) Comparison of RAPID signal with mRNA expression profiling. All genes are represented on the x-axis, ranked from left to right, based on RAPID signal in the entire animal (shown on left y-axis). Averages transcripts per million (tpm; calculated with salmon) using data set from study c above (ribosomal RNA depleted) were calculated, using the genes falling into each bin of 690 genes on the x-axis (values on right y-axis). (D and F) Profiles at large scale (D) and gene scale (housekeeping genes, F) of RAPID and the different RNA polymerase II ChIP-seq studies cited in B. (E) Metagene plot of the RNA polymerase II ChIP-seq and RAPID signals on WS270 genes. TES, transcription end site; TSS, transcription start site.
footprints for all samples (Figure S3C), and the reproducibility of the RAPID signal at the gene level between replicas was good for the intestinal sample \((r = 0.78)\) but lower for muscle and Y cell samples \((r = 0.46\) and 0.24, respectively; Figure S4A). When using an FDR < 0.05, we detected a total of 4986, 3165, and 4819 genes with a high RAPID signal in

Figure 2  Tissue-specific expression profiles using RAPID by Cre/lox recombination in young adult animals. (A) Experimental system for tissue-specific expression of Dam fusions and RAPID analysis. (B) Venn diagram of overlap between expressed genes identified using RAPID in three different tissues. (C) RAPID profiles for previously characterized genes expressed in a tissue-specific manner in intestine, muscle, and the XXX neuroendocrine pair.
the muscle, intestinal, and Y data sets, respectively; and 1566, 1570, and 840 consistently detected genes (FDR < 0.05 in both replicates) in these tissues (Table S4). In these experimental conditions, RAPID appears suitable to identify transcription footprints of cell-type-specific genes, such as the intestinal genes cpr-6 (cathepsin B; Pauli et al. 2006) and elo-2 (fatty acid elongase ELOVL; Han et al. 2017), or the Glutathione S-Transferase gene gst-4 expressed in muscle, as well as the muscle-specific WDR1 homolog unc-78 (Figure S3C; Mohri and Ono 2003; Hasegawa et al. 2007). Comparison of the footprints obtained for all three tissues allowed us to define 395 Y-specific genes (Figure S3B), and among them, shallow Y-specific RAPID signal was visible for ceh-6 and sox-2, two transcription factors previously shown to be involved in Y-to-PDA transdifferentiation initiation (Figure S3C; Kagias et al. 2012). In contrast, we observed some unexpected RAPID footprints for genes that we expected to be expressed in a tissue-specific manner. For example, the myosin myo-3 was observed in all three tissues (Figure S3D). Conversely, some genes which are known to be expressed in tissue-specific manner did not show high RAPID values, e.g., the intestine-specific asp-1 and spp-5 genes or the muscle-specific pat-10 gene (Figure S3D). These differences might have a biological significance: most transcription patterns have been defined in older animals, yet those genes might be expressed more broadly, or not at all, in embryonic blastomeres. Alternatively, the lack of a robustly established transcription pattern might hinder faithful footprinting. Indeed, RAPID requires the enzymatic action of the methyltransferase to build up the signal, while footprints are removed at each replication round in rapidly dividing blastomeres. Together with the stochasticity of single methylation events, this might decrease the precision of the probe. In agreement with this, intestinal cells, most of which are born beyond selected tissue markers, we next engaged three more comprehensive approaches. These approaches focused on muscle and intestine, since no large-scale gene expression data are available in XXX cells.

**Comparison of tissue-specific RAPID with available data sets in muscle and intestine**

To expand the validation of our RAPID footprint gene lists beyond selected tissue markers, we next engaged three more comprehensive approaches. These approaches focused on muscle and intestine, since no large-scale gene expression data are available in XXX cells.

First, we focused on a narrower set of genes with high RAPID footprints present exclusively in one of the tissues. These gene lists with the top 500 unique genes of each tissue, were analyzed with a computational method that predicts tissue-specific enrichment, using a multigene query (http://worm-tissue.princeton.edu/; Chikina et al. 2009). The tissue prediction tool showed that muscle-only and intestine-only RAPID footprinted genes gave a high prediction score for muscle and intestine, respectively (Figure 3, A and B). In
contrast, their scores were much lower for nontargeted tissues (including neurons, hypodermis, and germline).

Second, we compared the genome-wide RAPID footprint signal levels with expression levels of genes determined by RNA-seq in FACS-sorted body wall muscle and intestinal cells (Kaletsky et al. 2018). As for whole animal samples (Figure 1C), we observed a correlation between RNA-seq levels and RAPID signals for both tissues (Figure 3, C and D). Similar correlations were observed when comparing RAPID signal with scRNA-seq data (Figure S8), with the caveat that the scRNA-seq experiment was performed in the second larval stage, not in young adults (Cao et al. 2017).

Third, we compared the RAPID-detected gene sets with those reported for adult muscle and intestine in RNA-seq-based studies [RNA-seq of FACS-sorted cells by Kaletsky et al. (2018), SRT (Ma et al. 2016), FANS (Haenni et al. 2012), or PAT-seq (Blazie et al. 2017)]. A majority of genes with detectable RAPID footprints were also identified with these tissue-specific transcriptomic methods. We found that 90% of the muscle RAPID hits (Figure 3E and Figure S5) and 87% of the intestinal RAPID hits were detected by at least one of the other methods (Figure 3F and Figure S5). Remarkably, the overlap between the genes detected by RAPID and other methods is in the same range as that found between other methods [our own data in Table S5, and similar analyses in Haenni et al. (2012); Ma et al. 2016]. Although these results are encouraging and suggest that RAPID hits are genuinely expressed in these tissues, we note that the number of genes detected with RAPID is overall lower than that detected with other methods, especially in the intestine, where FACSsequencing and PAT-seq detect twice as many genes as RAPID (Figure SSC). To evaluate what kind of genes were not
identified by RAPID under our experimental conditions, we performed GO analyses on the common gene sets (identified by two methods) and method-specific gene sets (identified by one but not the other method) (Figures S6 and S7 and Table S6). For this purpose, we used the WormCat analysis tool (Holdorf et al. 2020), specifically developed to categorize and visualize C. elegans GO data. For the RAPID/FACS and RAPID/PAT-seq common gene sets, we mostly found GO categories related to widely expressed genes (Development, Transmembrane proteins, Signaling, Metabolism…) and categories that one would expect in these tissues (Figures S6 and S7). For example, significantly enriched categories in muscle included Muscle function and Cytoskeleton; and in the intestine, Metabolism of lipids, Trafficking of vesicles, Endocytosis, and Lysosomes (Figure S7). GO categories associated with RAPID-missed genes (only-FACS, only-PAT-seq in Figure S7) and FACS/PAT-seq common genes largely overlapped with those from the RAPID/FACS and RAPID/PAT-seq common genes, and contained additional very general categories (such as Transcription, DNA, Ribosome, Development, or mRNA processing). They also included less expected categories (such as Extracellular material: collagen, Neuronal function: neuropeptide and Major sperm protein in muscle; and Muscle function, Cytoskeleton actin function, and Major sperm protein in intestine). These latter categories may point to possible contaminations by other tissues, a known phenomenon in cell-sorting and cross-linking-based methods (Von Stetina et al. 2007; Haenni et al. 2012; Spencer et al. 2014).

Taken together, the results of our comparative analyses with previous gene expression studies indicate that, even if the number of RAPID-detected genes tends to be lower than for RNA-seq-based methods, RAPID identifies tissue-specific RNA Pol transcriptional activity that largely matches the specific transcriptional output in these tissues.

**RAPID identifies genes expressed in XXX cells**

XXX cells are neuroendocrine cells derived from hypodermal embryonic progenitors. They express enzymes required for steroid hormone synthesis, such as DAF-9, and are implicated in the control of dauer formation during development (Gerisch et al. 2001; Ohkura et al. 2003). As we still have limited information on the set of genes expressed in XXX cells, as well as on the role of those cells in adults, our XXX cell-transcribed gene data set could represent a valuable resource for future studies. We thus conducted additional experiments to confirm the validity of RAPID profiles to infer gene transcription in those cells and identify new markers.

A GO term analysis highlighted many terms related to secretory and neural functions in the gene set specific for XXX cells, while they were absent from the muscle- and intestine-enriched gene sets (Table S5, sheet 7). These results are expected for neuroendocrine cells, with functional properties very similar to those of neurons.

To further confirm the validity of RAPID profiles to infer gene transcription, we selected 12 genes with high RAPID signal in XXX cells compared to intestine, muscle, and whole animals, by comparing their profiles in the four tissues (Figure S6). We cloned their putative promoters in front of a mNeon-Green fluorescent reporter (Hostettler et al. 2017) and created transgenic lines bearing extrachromosomal arrays by gonad microinjection, together with an XXX-specific red reporter driven by the sdf-9 promoter (the promoter that we previously used for XXX-specific RAPID). Out of the 12 promoters, 8 drove expression in the XXX cells at various levels (asic-1, dhs-17, F14H8.2, C06A1.3, asp-9, B0034.5, F52E1.5, and nlp-15; Figure 4), one was impossible to score (T22E5.6) because it also drove high expression in the pharynx located very close to the XXX cells, and three did not show detectable expression in XXX cells (in adults; mab-9, twk-39, and lgc-36). The absence of expression in the latter group could indicate that these genes represent false positives in the RAPID analysis, that the transcription was active only at earlier developmental time points, or that the arbitrarily defined promoters failed to reflect the endogenous transcriptional activity. Even if the subset of genes tested with the reporter follow-up analysis is small, we could calculate a gross estimate of the true positive rate of RAPID in XXX cells. Based on the 8:11 ratio (8 positive genes/promoters out of 11 interpretable readouts), a resampling bootstrap simulation indicated that the true positive rate should range between 45 and 88% (95% confidence interval, Figure S10). By extrapolation, this would correspond to a range of ~1100–2100 genes expressed in XXX and identified via RAPID.

Interestingly, in half of the reporters expressed in XXX cells (four out of eight), we observed only limited or no signal in other cell types. In particular, reporters for dhs-17 and F14H8.2 were predominantly expressed in XXX cells and sequence homology suggests plausible XXX-specific roles for these two genes. Indeed, dhs-17 encodes an uncharacterized member of the short-chain dehydrogenase/reductase family, which is potentially involved in steroid metabolism (Zhang et al. 2013), and may therefore contribute to this specific neuroendocrine function of XXX cells. As for F14H8.2, it is a paralog of eak-4 selectively expressed in XXX cells and regulating dauer formation (Hu et al. 2006).

Taken together, the results of our reporter gene analysis indicate that the RAPID approach applied to only two cells per animal successfully identifies actively transcribed genes in these cells, including genes with relatively low expression and novel cell-specific markers.

**Discussion**

Determining the transcriptional profile of specific cell types has been a major hurdle for C. elegans researchers, as most research is carried out on intact, living animals. Recent technological advances now allow the acquisition of a comprehensive view of the genes expressed in individual cell types. Two main RNA-based approaches have been used in the community to this aim: “RNA tagging/pulling” approaches that allow to sample gene expression directly from whole, live animals, either through affinity purification or pull-down
Figure 4 Promoters of genes identified as expressed in XXX cells using RAPID drive fluorescent reporter expression in XXX cells. Left panel: partial z-projections of adult heads showing expression of sdf-9p::mScarlet in XXX cells. Middle panel: partial z-projections of adult heads showing the expression pattern of mNeonGreen reporter under transcriptional control of the promoters of indicated genes. Right panel: merge of fluorescent channels within DIC images validating the expression of mentioned genes in XXX cells. Bars, 10 μm.
enrichment; and “dissociation-based” methods, requiring tissue dissociation followed by either purification of cells/nuclei by FACS or individual cell isolation.

Here, we adapt a third type of approach originally developed in *Drosophila*, based on RNA Pol footprinting on DNA (Southall et al. 2013). RAPID is based on DNA modification by DNA adenine methyltransferase fused with a RNA Pol subunit, detecting active transcribed genes footprinting by DamID (van Steensel et al. 2001). RAPID is a simple, fast, and cost-effective technique to identify transcribed genes in individual tissues or cells, requiring only the knowledge of one tissue- or cell-specific promoter. We discuss below the specificities and limits of RAPID for transcriptional profiling, in comparison with previously published methods (summarized in Table S7).

First, as RAPID is based on the extraction of DNA from entire animals, it avoids in vitro cell manipulations and cell isolation-induced transcriptome modifications, similarly to RNA-tagging/pulling techniques. Second, since methylation occurs only in cells in which the Dam fusion is expressed after Cre/lox recombination, RAPID yields signals with a high specificity toward targeted tissues. Our comparative analyses with other methods for muscle and intestine, as well as the validation of our results with XXX cells, indeed indicate that a majority of RAPID hits are genuinely expressed in these tissues. Third, the very low level of signal background noise in RAPID makes it sensitive enough to function with a low number of worms (4000, at least a 10-fold reduction compared to other methods), even for a tissue representing two cells per animal. This reduces the time necessary for worm population growth. Additionally, as each examined tissue only requires a tissue-specific Cre driver (many of which already exist; Kage-Nakadai et al. 2014; Ruijtenberg and van den Heuvel 2015), and as the library preparation is relatively simple to execute, RAPID makes it possible to process several conditions/tissues in parallel and is accessible to a neophyte researcher in the field. Fourth, RAPID is versatile, as minimal modifications will be required to target other DNA-interacting proteins, such as transcription factors or chromatin modifiers, to analyze tissue- or cell-specific genome-wide binding (Marshall and Brand 2017; Aughey et al. 2018). With the lower cost of nanopore sequencing systems, it is even conceivable that individual laboratories purchase their own equipment, further cutting down on waiting time and indirect costs.

The advantages described above come at a cost, in particular on how quantitative, dynamic, and comprehensive RAPID data are. Our comparison of the overlap between transcriptomic data sets from different methods suggests that a certain portion of genes is uniquely detected by each method (Figure S5), an indication that no single approach truly captures the complete transcriptome or that variable experimental conditions affect the genes identified as expressed. However, when compared to all other RNA-seq based methods (Figure S5), RAPID identified overall less genes as expressed. On the one hand, it is possible that affinity purification or dissociation-based methods could detect a number of genes due to contaminants and experiment-induced gene expression, or stored messengers inherited from mother cells. On the other hand, RAPID has a number of inherent technical limitations, which may hinder a comprehensive and unbiased genome interrogation.

First, RAPID relies on the GATC density per gene, and a lower density will yield a lower signal, as longer DpnI restriction fragments will be less efficiently amplified by the DamID PCR. In agreement, genes detected using FACS/RNA-seq, but not by RAPID, have on average a lower GATC density (data not shown; Kaletsky et al. 2018). Increasing the material amount and/or sequencing coverage is expected to dampen this type of bias. Second, another consequence of the methylation is that the dynamic range of RAPID compared to RNA-seq is expected to be lower: once a gene sequence is fully methylated, further transcription will not lead to increased methylation levels of the DNA, leading to a plateau effect. Conversely, as RAPID uses minute traces of the Dam fusions, lowly expressed genes will rarely be bound by a Pol containing the Dam fusion, and their sequence will be rarely methylated. As our quantitative comparison reveals (Figures 1C and 3B), these issues are partially relieved by the stochasticity of the methylation between different cells in different animals, leading to a correlation between RAPID levels and expression levels determined by RNA-seq. Increasing sequencing coverage should allow both the retrieval of more lowly expressed genes and an improvement in the signal dynamic range. For certain applications, such as identifying tissue-specific genes to serve as molecular markers, a lower dynamic range can actually be an advantage as compared to RNA-seq methods, as the large range of mRNA molecule number in a cell requires very deep sequencing to detect more lowly expressed genes. Third, RAPID signal on DNA will not discriminate between gene isoforms created by alternative splicing, since the RNA Pol will progress over the whole intronic and exonic regions regardless of whether they are retained in the final mRNA product. Fourth, RAPID signal is a stochastic average of the methylation by Pols since the last DNA replication (which erases the methylation signal). The comparison between embryonic blastomeres and young adult tissues (with older postmitotic cells) highlights the increase in reproducibility of RAPID as animals age (Figure S4). Genes detected by RAPID are likely to represent cell-specific genes that are expressed over longer time scales in the cells under study, rather than genes transiently expressed in response to the environment. While this phenomenon enables RAPID to efficiently identify cell-type-specific genes (Tables S4 and S5), it will also limit its ability to report quantitative variations over shorter time scales. Finally, in its present form, RAPID is not well suited to analyze transcriptome dynamics, as the methylation signal can only be erased by DNA replication. RAPID could be further improved by timing the expression of the Dam fusions. We tested auxin-mediated Dam degradation. However, degradation was not complete enough and the remaining low levels of Dam fusions were sufficient to create a RAPID signal (data not shown). Alternatively, degron-tagged Cre recombinases might provide a more reliable way to time the expression of the
multiple genotypes have to be assessed. Synchronization of large populations is tedious, notably when the existing methods to analyze comprehensive gene expression are limited by their dependence on cell type-specific markers and further study their biology. RAPID is unlikely to capture this type of phenomenon, which could turn into a desirable feature when one wants to mitigate this variability, e.g., in studies comparing hard-to-synchronize genetic backgrounds.

In conclusion, we believe that RAPID is a useful addition to the existing methods to analyze comprehensive gene expression at the cell-type level, allowing one to identify new and specific markers and further study their biology. RAPID is an easy, scalable, entry-level method to target rare or less known cell types, cell types difficult to purify because of their morphology or the developmental stage targeted, and when synchronization of large populations is tedious, notably when multiple genotypes have to be assessed.

Acknowledgments

We are grateful to Laurence Bulliard, Cihan Enci, and Lisa Schild for expert technical support; Marc Hammarlund, Piali Sengupta, and Sanders van der Heuvel for the gift of material; and Michalis Barkoulas for sharing preliminary data. Some strains were provided by the CGC (Caenorhabditis Genetics Center), which is funded by the National Institutes of Health Office of Research Infrastructure Programs (P40-OD-010440). The study was supported by the Swiss National Science Foundation (grants IZCNZ0-174703/SBF1_C16.0013, BSSG10-155764, and PP0OP3_150681 to D.A.G.; ZCSZ0-174641/SBF1_C15.0076, PP0P3_159320, and 31003A_176226 to P.M.), the ERC CoG (European Research Council Consolidator Grant) PlastiCell grant #648960 to S.J., and a COST Action (European Cooperation in Science and Technology; BM1408).

Literature Cited

Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of C. elegans trans-splicing. Genome Res. 21: 255–264. https://doi.org/10.1101/gr.113811.110
Aughey, G. N., A. Estacio Gomez, J. Thomson, H. Yin, and T. D. Southall, 2018 CATaDa reveals global remodelling of chromatins accessibility during stem cell differentiation in vivo. eLife 7: e32341. https://doi.org/10.7554/eLife.32341
Blazie, S. M., C. Babb, H. Wilky, A. Rawls, J. G. Park et al., 2015 Comparative RNA-seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. BMC Biol. 13: 4. https://doi.org/10.1186/s12951-015-0116-6
Blazie, S. M., H. C. Geissel, H. Wilky, R. Joshi, J. Newbern et al., 2017 Alternative polyadenylation directs tissue-specific miRNA targeting in Caenorhabditis elegans somatic tissues. Genetics 206: 757–774. https://doi.org/10.1534 genetics.116.196774

Cabinca, D. S., C. Muñoz-Jiménez, V. Kalck, D. Gaidatzis, J. Padeken et al., 2019 Active chromatin marks drive spatial sequestration of heterochromatin in C. elegans nuclei. Nature 569: 734–739. https://doi.org/10.1038/s41586-019-1243-y
Cao, J., J. S. Packer, V. Ramani, D. A. Cusanovich, C. Huynh et al., 2017 Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357: 661–667. https://doi.org/10.1126/science.aam8940
Challife, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, 1994 Green fluorescent protein as a marker for gene expression. Science 263: 802–805. https://doi.org/10.1126/science.8303295
Chikina, M. D., C. Huttonhower, C. T. Murphy, and O. G. Troyanskaya, 2009 Global prediction of tissue-specific gene expression and context-dependent gene networks in Caenorhabditis elegans. PLOS Comput. Biol. 5: e1000417. https://doi.org/10.1371/journal.pcbi.1000417
Cramer, P., 2004 RNA polymerase II structure: from core to functional complexes. Curr. Opin. Genet. Dev. 14: 218–226. https://doi.org/10.1016/j.gde.2004.01.003
Cramer, P., D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis et al., 2000 Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288: 640–649. https://doi.org/10.1126/science.288.5466.640
Frøkjær-Jensen, C., M. W. Davis, C. E. Hopkins, B. J. Newman, J. M. Thummel et al., 2008 Single-copy insertion of transgenes in Caenorhabditis elegans. Nat. Genet. 40: 1375–1383. https://doi.org/10.1038/ng.248
Garrido-Lecca, A., T. Saldi, and T. Blumenthal, 2016 Localization of RNAPII and 3′ end formation factor CstF subunits on C. elegans genes and operons. Transcription 7: 96–110. https://doi.org/10.21542/2016.1168509
Gerisch, B., C. Weitzel, C. Kobert-Eisermann, V. Rottiers, and A. Antebi, 2001 A hormonal signaling pathway influencing C. elegans metabolism, reproductive development, and life span. Dev. Cell 1: 841–851. https://doi.org/10.1016/S1534-5807(01)00085-5
Gómez-Saldívar, G., P. Meister, and P. Askjaer, 2016a DamID analysis of nuclear organization in Caenorhabditis elegans. Methods Mol. Biol. 1411: 341–358. https://doi.org/10.1007/978-1-4939-3530-7_22
Gómez-Saldívar, G., P. Meister, and P. Askjaer, 2016b DamID of nuclear envelope proteins in C. elegans, The Nuclear Envelope, Methods and Protocols, edited by S. Shackleton, P. Collas, and E. C. Schirmer. Springer, Berlin, Germany. pp 341–358. https://doi.org/10.1007/978-1-4939-3530-7_22
Gracia, X., and J. A. Calarco, 2017 Cell type-specific transcriptome profiling in C. elegans using the translating ribosome affinity purification technique. Methods 126: 130–137. https://doi.org/10.1016/j.ymeth.2017.06.023
Greer, E. L., M. A. Blanco, L. Gu, E. Sendinc, J. Liu et al., 2015 DNA methylation on N6-adenine in C. elegans. Cell 161: 868–878. https://doi.org/10.1016/j.cell.2015.04.005
Haenni, S., Z. Ji, M. Hoque, N. Rust, H. Sharpe et al., 2012 Analysis of C. elegans intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3′-end-seq. Nucleic Acids Res. 40: 6304–6318. https://doi.org/10.1093/nar/gks282
Hammarlund, M., O. Hober, D. M. Miller, and N. Sestan, 2018 The CeNGEN project: the complete gene expression map of an entire nervous system. Neuron 99: 430–433. https://doi.org/10.1016/j.neuron.2018.07.042
Han, S., E. A. Schroeder, C. G. Silva-García, K. Hebestreit, W. B. Mair et al., 2017 Mono-unsaturated fatty acids link H3K4me3 modifiers to C. elegans lifespan. Nature 544: 185–190. https://doi.org/10.1038/nature21686
Hart, J. C., C. D. Schmid, C. Muñoz-Jiménez, R. Romero-Bueno, V. Kalck et al., 2020 Loss of an H3K9me anchor rescues laminopathy-linked changes in nuclear organization and muscle function in Drosophila.
Pauli, F., Y. Liu, Y. A. Kim, P. J. Chen, and S. K. Kim, 2006 Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in C. elegans. Development 133: 287–295. https://doi.org/10.1242/dev.02185

Pindyurin, A. V., L. Pagie, E. N. Kozhevnikova, J. van Arensbergen, and B. van Steensel, 2016 Inducible DamID systems for genomic mapping of chromatin proteins in Drosophila. Nucleic Acids Res. 44: 5646–5657. https://doi.org/10.1093/nar/gkw176

Raudvere, U., L. Kolberg, I. Kuzmin, T. Arak, P. Adler et al., 2019 Profi lé: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47: W191–W198. https://doi.org/10.1093/nar/gkz269

Redolfi, J., Y. Zhan, C. Valdes-Quezada, M. Kryzhanovska, I. Guerreiro et al., 2019 DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. Nat. Struct. Mol. Biol. 26: 471–480. https://doi.org/10.1038/s41594-019-0231-0

Rhoades, J. L., J. C. Nelson, I. Nwabudike, S. K. Yu, I. G. McLachlan et al., 2019 ASICs mediate food responses in an enteric serotonergic neuron that controls foraging behaviors. Cell 176: 85–97.e14. https://doi.org/10.1016/j.cell.2018.11.023

Roeder, T., M. Stanišak, C. Gelhaus, I. Bruchhaus, J. Grötzing et al., 2010 Caenopores are antimicrobial peptides in the nematode Caenorhabditis elegans instrumental in nutrition and immunity. Dev. Comp. Immunol. 34: 203–209. https://doi.org/10.1016/j.dci.2009.09.010

Roy, P. J., J. M. Stuart, J. Lund, and S. K. Kim, 2002 Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature 418: 975–979 [corrigenda: Nature 450: 128 (2007)]. https://doi.org/10.1038/nature01012

Ruijtenberg, S., and S. van den Heuvel, 2015 G1/S inhibitors and the SWI/SNF complex control cell-cycle exit during muscle differentiation. Cell 162: 300. https://doi.org/10.1016/j.cell.2015.06.013

Schier, A. C., and D. J. Taatjes, 2020 Structure and mechanism of the RNA polymerase II transcription machinery. Genes Dev. 34: 465–488. https://doi.org/10.1101/gad.335679.119

Schuster, E., J. J. McElwee, J. M. Tullet, R. Doonan, F. Matthijssens et al., 2010 DamID in C. elegans reveals longevity-associated targets of DAF-16/FoxO. Mol. Syst. Biol. 6: 399. https://doi.org/10.1038/msb.2010.54

Sharma, R., D. Jost, J. Kind, G. Gómez-Saldívar, B. van Steensel et al., 2014 Differential spatial and structural organization of the X chromosome underlies dosage compensation in C. elegans. Genes Dev. 28: 2591–2596. https://doi.org/10.1101/gad.248864.114

Southall, T. D., K. S. Gold, B. Egger, C. M. Davidson, E. E. Cagigill et al., 2013 Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. Dev. Cell 26: 101–112. https://doi.org/10.1016/j.devcel.2013.05.020

Spencer, W. C., G. Zelles, J. D. Watson, S. R. Henz, K. L. Watkins et al., 2011 A spatial and temporal map of C. elegans gene expression. Genome Res. 21: 325–341. https://doi.org/10.1101/gr.114595.110

Spencer, W. C., R. McWhirter, T. Miller, P. Strasbourger, O. Thompson et al., 2014 Isolation of specific neurons from C. elegans larvae for gene expression profiling. PLoS One 9: e112102. https://doi.org/10.1371/journal.pone.0112102

Steiner, F. A., P. B. Talbert, S. Kasinathan, R. B. Deal, and S. Henikoff, 2012 Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Res. 22: 766–777. https://doi.org/10.1101/gr.131748.111

Takayama, J., S. Faumont, H. Kunitomo, S. R. Lockery, and Y. Iino, 2010 Single-cell transcriptional analysis of taste sensory neuron pair in Caenorhabditis elegans. Nucleic Acids Res. 38: 131–142. https://doi.org/10.1093/nar/gkp868

Therepenova, I., L. Bhattacharyya, C. S. Rubin, and J. H. Freedman, 2000 Aspartic proteases from the nematode Caenorhabditis elegans. Structural organization and developmental and cell-specific expression of asp-1. J. Biol. Chem. 275: 26359–26369. https://doi.org/10.1074/jbc.M000956200

Towbin, B. D., C. González-Aguilera, R. Sack, D. Gaidatzis, V. Kalck et al., 2012 Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell 150: 934–947. https://doi.org/10.1016/j.cell.2012.06.051

van Steensel, B., J. Delrow, and S. Henikoff, 2001 Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. 27: 304–308. https://doi.org/10.1038/85871

Von Stetina, S. E., J. D. Watson, R. M. Fox, K. L. Olszewski, W. C. Spencer et al., 2007 Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C. elegans nervous system. Genome Biol. 8: R135. https://doi.org/10.1186/gb-2007-8-7-r135

Wang, J., R. Kaletsky, M. Silva, A. Williams, L. A. Haas et al., 2015 Cell-specific transcriptional profiling of ciliated sensory neurons reveals regulators of behavior and extracellular vesicle biogenesis. Curr. Biol. 25: 3232–3238. https://doi.org/10.1016/j.cub.2015.10.057

Widmer, Y. F., A. Bicolor, R. Bruggmann, and S. G. Sprecher, 2018 Regulators of long-term memory revealed by mushroom body-specific gene expression profiling in Drosophila melanogaster. Genetics 209: 1167–1181. https://doi.org/10.1534/genetics.118.301106

Yang, Y.-F., X. Zhang, X. Ma, T. Zhao, Q. Sun et al., 2017 Transsplicing enhances translational efficiency in C. elegans. Genome Res. 27: 1525–1535. https://doi.org/10.1101/gr.20150.115

Zhang, Y., X. Zou, Y. Ding, H. Wang, X. Wu et al., 2013 Comparative genomics and functional study of lipid metabolic genes in Caenorhabditis elegans. BMC Genomics 14: 164. https://doi.org/10.1186/1471-2164-14-164

Communicating editor: O. Hobert

RNA Polymerase Footprinting in Nematodes 945