Supplemental Methods

Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation

Michael S. Werner[1], Bogdan Sieriebriennikov[1], Neel Prabh[1], Tobias Loschko[1], Christa Lanz[1] and Ralf J. Sommer[1]

Nematode synchronization and growth for ChIP-, ATAC-, and RNA-seq

P. pacificus were grown for three healthy generations on OP50 bacteria. Then, to increase homozygosity, a single worm was passaged to 10 x 10 cm plates. After 5 days, 15-20 worms were split to 100 x 10 cm plates spotted with 700 µl OP50 and grown 5-6 days (at 7 days the presence of too many late stage J1s that are still resistant to bleach prevents consistent synchronization). Worms were washed with water into 50 ml conical tubes, and bleach-treated (10 minutes, 30% NaOH/bleach)(Stiernagle 2006) to isolate eggs and J1 larva. Eggs-J1 were then filtered through a 140 µM nylon net (Millipore) to remove carcasses, then centrifuged 500 x g, 1 minute to pellet. The pellet was washed by removing supernatant with a pipette and re-suspending in 3 ml M9, then re-centrifuged 500 x g, 30-60 seconds. Supernatant was removed, and the egg-J1 pellet was re-suspended in 100 µl M9 buffer x the number of plates grown (i.e. 20 ml for 200 plates). Eggs-J1 were added drop-wise to new 10 cm agar plates (100 µl each) spotted with 700 µl OP50 and grown to young adults (approximately 64 hours).

ChIP-seq

Most antibodies used conformed to the ChIP-Seq guidelines outlined by ENCODE/modENCODE (Landt et al. 2012). Specifically, we applied three levels of quality control: (1) immunoblot specificity (primary evidence) from previous work in the lab (Serobyan et al. 2016) or the manufacturer (Diagenode or abcam), except for H3K9ac and H3K36me3, (2) expected annotation enrichment (i.e. H3K4me3/H3K27ac overlap with each other and at promoters), and (3) independent biological replication of each antibody (see below). Note, ChIP with H4K20me1 and H2AZ.2 antibodies did not successfully co-precipitate sufficient DNA for library preparation (Supplemental Fig. 2F). A list of antibodies used is provided in Supplemental Table 3. Two to three independent biological replicates of native ChIP-seq of histone modifications in *P. pacificus* were performed by a combination of native ChIP (Brand et al. 2008) with nematode nuclear isolation (Steiner et al. 2012), with the exception that we did not observe

an increase in specificity with hydroxyapatite purification, and therefore this step was not used. We first disrupted the worm cuticle using liquid nitrogen and mortar and pestle adapted from Steiner et al., 2012, followed by resuspension in 10 ml nuclei purification buffer (NPB: 10 mM Tris at pH 7.5, 40 ml NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.2 mM DTT, 0.5 mM PMSF, 0.5 mM spermidine, 0.25 mM spermine, 0.1% Triton X-100). Cells were lysed by applying 30-50 strokes of dounce-homoginization, then cell and cuticle debris were removed by slow centrifugation (100 x g, 2 minutes, room temperature). Nuclei in the supernatant were purified over a 7.5 ml sucrose cushion (10 mM Hepes, pH 7.5, 30 % w/v sucrose, 1.5 mM $MgCl_2$) for 12 minutes, 4° C, 1,300 x g. Then the sucrose cushion was removed by pipette, and nuclei were washed by resuspension in 1 ml NPB and centrifuged 10 minutes, 4° C, 1,300 x g. Nuclei concentration was assessed by A260 nanodrop absorption of a 1:10 dilution in NBP, then 1M $CaCl_2$ was added to 5 mM final concentration and incubated 37° C for 5 minutes to 'warm up' the sample, and then 0.015 µl MNase (NEB)/µg chromatin was added and incubated 37° C for six minutes to digest chromatin to predominately mono-nucleosomes (Supplemental Fig. S2B). Then 100 mM EGTA was added to 20 mM final concentration to stop the reaction, and 5 M NaCl was added drop-wise to 400 mM final concentration while vortexing on the lowest setting. Then nuclei were incubated 30 minutes 4° C with rotation to extract nucleosomes from nuclei, and centrifuged max speed, 4° C, 5 minutes to pellet nuclear debris. Resulting mono and di-nucleosomes in the supernatant were re-quantified by nanodrop, and diluted to 10-20 ng/µl in ChIP Buffer 1 (25 mM Tris pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 10% glycerol, 0.1% NP-40, 100 µM PMSF, 50 µg/ml BSA), and immunoprecipitated with histone modification-specific antibodies conjugated to 20 µl para-magnetic Dynabeads (Invitrogen) for 10 minutes. 2.5 µg of chromatin (~250-500 µl) was used as input per IP which is a typical amount from 20 x 10 cm plates, and 5 µg Ab were conjugated to 20 µl beads before hand per IP. Washing of IPs was performed for 1 minute with rotation at 4° C, followed by 1 minute magnetic separation, including three tube changes with the following buffers: 2x with 600 µl ChIP Buffer 2 (25 mM Tris pH 7.5, 5 mM

$MgCl_2$, 300 mM KCl, 10% glycerol, 0.1% NP-40, 100 µM PMSF, 50 µg/ml BSA), 1x with ChIP Buffer 3 (10 mM Tris pH 7.5, 250 mM LiCl, 1 mM EDTA, 0.5% Na-Deoxycholate, 0.5% NP-40, 100 µM PMSF, 50 µg/ml BSA), 1x with ChIP Buffer 1, and 1x with TE buffer. Immunoprecipated nucleosomes were eluted in 50 µl elution buffer (50 mM Tris pH 7.5, 1 mM EDTA, 1% SDS) 55° C for 5 minutes. Supernatant was transferred to a new tube and purified by adding 2 µl of 5 M NaCl (200 mM final), 1 µl of 500 mM EDTA (10 mM final), and 1 µl of 20 mg/ml Proteinase K (0.4 mg/ml final) and incubating 55° C for 2 hours, followed by 3 volumes of Ampure XP bead addition and purification, and elution in 50 µl TE buffer. Recovered DNA was verified for specificity by qPCR (Supplemental Fig. S2D), and then converted to Illumina sequencing libraries using a TruSeq Nano kit, and sequenced on an Illumina HiSeq 3000.

Iso-Seq

In order to increase representation of stage- and condition-specific transcripts, RNA used for the Iso-Seq experiment was collected from worms of different ages and grown in two different culture conditions, which are known to influence multiple aspects of nematode physiology, such as body morphology and development of feeding structures(Werner et al. 2017). First, eggs were synchronized by bleaching (see above, Stiernagle 2016). After bleach-synchronization, eggs were hatched in one of the two following conditions. Half of the eggs were added to 10 cm agar plates with Nematode Growth Medium (NGM) spotted with 1 mL of overnight culture of Escherichia coli OP50. Another half were transferred to 25 mL conical flasks with 10 mL of S-medium that contained re-suspended bacteria in the amount corresponding to 100 mL of an overnight culture with OD600 of 0.5. Flasks with liquid cultures were incubated on a shaking platform at 180 rpm. Worms were collected from both solid and liquid cultures at three time points - 24 h, 55-60 h and 75 h. Samples collected at 24 h contained a mixture of J2 and J3 (early juvenile stages) and samples collected at 55-60 h contained a mixture of J4 (late juvenile stage) and young adults. At 75 h, the majority of animals were gravid adults. They were

bleached as described above to extract embryos and J1 (egg-bound juvenile stage). RNA was extracted from all samples separately using TRIzol Reagent (Invitrogen) including 3x freeze-thaw cycles, and chloroform, and RNA was purified from the aqueous phase using RNA Clean and Concentrator kit (Zymo Research). Extracted RNA was quantified using Nanodrop One C (Thermo Scientific) and quality was verified using capillary electrophoresis performed on Bioanalyzer 2100 in combination with RNA 6000 Nano Kit (Agilent Technologies). After the quality control step, equal amounts of RNA from different time points were pooled. Two separate pools were made for worms grown on agar plates and in the liquid medium. 1 ug of RNA was taken from each pool and first-strand cDNA synthesis was performed using SMARTer PCR cDNA Synthesis Kit (Clontech Laboratories). These cDNA samples were labelled as "direct" and presumably only contained cDNA reverse-transcribed from polyadenylated transcripts. Another 4 ug of RNA was taken from each pool and split into "rRNA-depleted" and "control" samples. The "rRNA-depleted" samples were purified from ribosomal RNA (rRNA) using Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat) (Illumina) and both samples were *in vitro* polyadenylated using E. coli Poly(A) Polymerase (New England Biolabs) to enrich the polyA RNA profile for transcripts that are not, or under-polyadenylated *in vivo*. cDNA synthesis was performed as described above, and the efficiency of rRNA depletion was verified by comparing relative abundances of rRNA in the "rRNA-depleted" and the "control" samples using real-time quantitative PCR (qPCR)(Supplemental Fig. S1A). Primers targeted 18S rRNA, 28S rRNA and a reference gene *Ppa-cdc-42*(Schuster and Sommer 2012), and qPCR was performed on a LightCycler 480 Instrument using LightCycler 480 SYBR Green I Master (Roche). "Direct" and "rRNA-depleted" cDNA was further converted into SMRTbell libraries following the guidelines provided by Pacific Biosciences. In short, cDNA was amplified by PCR using PrimeSTAR GXL Polymerase (Clontech Laboratories). Optimal number of cycles (12 cycles) was identified based on band intensity and size distribution of amplification products run on a 1% agarose gel (Supplemental Fig. S1B). Amplified cDNA was repaired from damage

using SMRTbell Damage Repair Kit-SPv3 (Pacific Biosciences) and ligated to sequencing adapters, followed by exonuclease treatment using SMRTbell Template Prep Kit 1.0-SPv3. Optional size selection step was omitted. Libraries were sequenced on the Sequel System using Sequel Binding Kit 2.0, Sequel Sequencing Kit 2.0, Sequencing Primer v3 and Sequel SMRT Cell 1M v2 Trays (Pacific Biosciences). Each sample ("direct" and "rRNA-depleted" for liquid culture, "direct" and "rRNA-depleted" for agar plates) was run on two SMRT Cells in separate runs, totaling to 2 trays of 4 SMRT Cells. The following parameters were used for library loading and sequencing: MagBead loading at 20 pM (first run) or 30 pM (second run), 120 min immobilization, 120 min pre-extension, 600 min movie. Sequencing of one SMRT Cell containing "rRNA-depleted" sample for agar plates failed as a result of a manufacturing defect. The rest of SMRT Cells produced the sequencing output of 2.1-4.3 Gb each with mean polymerase read length of 17-25 kb. SMRT Link software version 4.0.0 (Pacific Biosciences) was used to convert subreads to circular consensus sequences and identify full-length non-chimeric reads.

ATAC-seq

To obtain regions of open chromatin we followed the omni-ATAC protocol (Corces et al. 2017) with purified worm nuclei from ≥15 µl of worm pellet. After the Iodixanol centrifugation, a 200 µl nuclear band was mixed with 1 ml ATAC-RSB/0.1% Tween-20 solution, and centrifuged 10 minutes, 500 x g, 4°C. Typically a pellet was not visible at this stage, however 12.5 µl was pipetted from the bottom of the tube and used for the transposon reaction. The duration and amount of transposase enzyme were the same as described in the Corces et al. protocol. After purification, ATAC libraries were size-selected using a BluePippin (Sage Science), and sequenced on site on an Illumina HiSeq3000. Technical replicates of each biological replicate were prepared with different illumina adapters, and after mapping the resulting .bam files were merged.

153

154 **Bioinformatic data analysis**

155 Iso-Seq reads were mapped to the El Paco assembly of the *P. pacificus* genome (Rödelsperger
156 et al. 2017) using GMAP (Wu and Watanabe 2005). Reads of the same samples from different
157 sequencing runs were merged, sorted and indexed using SAMtools (Li et al. 2009)(2009). Gene
158 annotations were derived from Stringtie (Pertea et al. 2016) using guided assembly ('-G' option)
159 with the El Paco reference for both 'direct' and 'rRNA-depleted total RNA'. Gene annotations
160 were then compared to El Paco for gene length and exon number using the density function in
161 R. Isoforms were assessed by converting .gtf annotations to .bed with a constant numerical 5th
162 column (1000), then parsing overlapping transcripts using BEDTools merge and performing
163 'sum' on the 5th column, and selecting genes that had greater than 1000 in this column.
164 Isoforms with the same start and stop coordinates were selected (again using BEDTools merge
165 and selecting genes with 'distinct' coordinates = 1), and a histogram of distinct coordinates was
166 plotted in R. Iso-Seq coverage of evolutionary gene classes was assessed by BEDTools
167 coverage, and density of exons and gene lengths was determined as before with the entire
168 annotation, but broken down by class.

169 For ChIP- and ATAC-seq, Illumina Fastq reads were aligned to the El Paco reference
170 genome (Rödelsperger et al. 2017) using Bowtie 2. The output .bam files were converted to
171 .sam using samtools and peaks were called independently for each biological replicate with
172 MACS2 (macs2 callpeak --broad -t Ab.bam -c Input.bam -f BAMPE -g 15e7 -n Ab_BP -B -m 2
173 50) (Supplementary Fig. S3). Input samples were processed and sequenced identically for each
174 biological replicate, and used as background control in MACS2 analysis. For duplicate data,
175 overlapping BroadPeaks were merged to create a genome-wide histone peak data set, and for
176 triplicate data (H3K4me1 and H3K4me3) we performed a multiIntersect (BEDTools), which
177 identified 6,081 H3K4me3 and 6,720 H3K4me1 locations that exhibited a peak in each of the
178 three replicates (Supplemental Table S1, ex. Supplemental Fig. S4A). To assess replicability

between duplicate data, we determined the fraction of overlapping peaks compared to the total amount of peaks (displayed as a weighted Venn Diagram in Supplemental Fig. S4B). For triplicate data we determined overlapping peaks for each sample with either of the other two samples, and for simplicity displayed the percent unique per replicate as an un-weighted Venn Diagram (weighted Venn Diagrams for three samples are mathematically impossible)(Supplemental Fig. S4B). Most of the Ab ChIP-seq and ATAC-seq replicates exhibited between 70-90% replicability of the smaller (fewer peaks) sample for duplicate data, or the average for triplicate data (Supplemental Fig. S4C). We set a minimum threshold of ≥50% overlap, which led to the removal of H2bub (27.5% overlap), H3K9ac (40.9% overlap), and H3K79me2 (49.8% overlap) data from further analysis. Although H3K9me3 exhibits relatively poor reproducibility to the other remaining samples (54% compared to >70%), we chose to keep it as 1) its broad distribution can make peak calling challenging, and 2) although H3K9me3 antibodies are generally less specific (Nishikori et al. 2012; Hattori et al. 2013), ChIP-seq data sets from multiple samples and organisms suggest it can still provide relevant information to distinguish facultative vs. constitutive heterochromatin (Trojer and Reinberg 2007).

Mapped replicate reads were normalized by coverage (depth x alignment rate) using samtools sub-sampling (samtools view –h –b –s), then combined using samtools 'merge' for chromatin state annotation with ChromHMM (Ernst and Kellis 2012) with a strict Poisson threshold (-strictthresh) and bin center (-center) options in 'BinarizeBam', and eight chromatin states (LearnModel binarized_files std_out 8 gene__assembly). Candidate state annotations were derived from previous observations in other organisms (Ernst et al. 2011; ENCODE Project Consortium 2012; Zentner et al. 2011; Consortium et al. 2015). Heatmaps of promoters and enhancers for different histone marks were made in R with combined replicate .bams using 'Pheatmap' from a matrix of per locus densities created in HOMER using 'annotatePeaks' -ghist. Rows were clustered from high to low densities, and colored from high to low log2(coverage)+1, the same scale was used for per mark for both promoters and enhancers. Sequence motifs

found in promoters and enhancers were obtained from HOMER using 'findMotifs' with standard options, and data presented are from a *de novo* motif search. Genomic loci presented are derived from Integrated Genome Viewer (IGV) images of combined replicate .bam files. Density of histone and ATAC-seq reads across the TSS of each evolutionary gene class were calculated in HOMER using combined .bams of replicates, and normalized by the high and low values from each data set. Meta-gene profiles of promoter and active enhancer locations relative to gene-bodies (+/-20 percent from 'slopBed') were obtained using the HOMER tool 'makeMetaGeneProfile.pl'.

Stranded RNA-seq data from two biological adult replicates were prepared by NEBNext Ultra Directional RNA-seq for Illumina kits, and mapped to El Paco using HISAT2 with standard parameters. Average expression (FPKM) from the two biological replicates of the reference gene annotation or evolutionary gene classes was determined by Stringtie and Ballgown(Pertea et al. 2016), and plotted in R by gene count. Expression groups were binned according to approximate inflection points. Distances in base pairs between evolutionary gene class TSSs to the nearest promoter (chromatin state 2) or enhancer (chromatin states 1 and 8) were obtained with the 'closestBed' function from BEDTools, and ordered and plotted in R in kilobase (kb).

Supplemental References

Brand M, Rampalli S, Chaturvedi C-P, Dilworth FJ. 2008. Analysis of epigenetic modifications of chromatin at specific gene loci by native chromatin immunoprecipitation of nucleosomes isolated using hydroxyapatite chromatography. *Nat Protoc* **3**: 398–409.

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

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**: 959–962.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

234 Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and
235 characterization. *Nat Methods* **9**: 215–216.

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

239 Hattori T, Taft JM, Swist KM, Luo H, Witt H, Slattery M, Koide A, Ruthenburg AJ, Krajewski K,
240 Strahl BD, et al. 2013. Recombinant antibodies to histone post-translational modifications.
241 *Nat Methods* **10**: 992–995.

242 Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P,
243 Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and
244 modENCODE consortia. *Genome Res* **22**: 1813–1831.

245 Nishikori S, Hattori T, Fuchs SM, Yasui N, Wojcik J, Koide A, Strahl BD, Koide S. 2012. Broad
246 Ranges of Affinity and Specificity of Anti-Histone Antibodies Revealed by a Quantitative
247 Peptide Immunoprecipitation Assay. *Journal of Molecular Biology* **424**: 391–399.

248 Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis
249 of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**: 1650–1667.

250 Rödelsperger C, Meyer JM, Prabh N, Lanz C, Bemm F, Sommer RJ. 2017. Single-Molecule
251 Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model
252 Organism Pristionchus pacificus. *Cell Rep* **21**: 834–844.

253 Schuster LN, Sommer RJ. 2012. Expressional and functional variation of horizontally acquired
254 cellulases in the nematode Pristionchus pacificus. *Gene* **506**: 274–282.

255 Serobyan V, Xiao H, Namdeo S, Rödelsperger C, Sieriebriennikov B, Witte H, Röseler W,
256 Sommer RJ. 2016. Chromatin remodelling and antisense-mediated up-regulation of the
257 developmental switch gene eud-1 control predatory feeding plasticity. *Nat Commun* **7**:
258 12337.

259 Steiner FA, Talbert PB, Kasinathan S, Deal RB, Henikoff S. 2012. Cell-type-specific nuclei
260 purification from whole animals for genome-wide expression and chromatin profiling.
261 *Genome Res* **22**: 766–777.

262 Stiernagle T. 2006. *Maintenance of C. elegans, WormBook, ed. The C. elegans Research*
263 *Community*. WormBook.

264 Trojer P, Reinberg D. 2007. Facultative Heterochromatin: Is There a Distinctive Molecular
265 Signature? *Mol Cell* **28**: 1–13.

266 Werner MS, Sieriebriennikov B, Loschko T, Namdeo S, Lenuzzi M, Dardiry M, Renahan T,
267 Sharma DR, Sommer RJ. 2017. Environmental influence on Pristionchus pacificus mouth
268 form through different culture methods. *Sci Rep* **7**: 7207.

269 Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA
270 and EST sequences. *Bioinformatics* **21**: 1859–1875.

271 Zentner GE, Tesar PJ, Scacheri PC. 2011. Epigenetic signatures distinguish multiple classes of
272 enhancers with distinct cellular functions. *Genome Res* **21**: 1273–1283.

273 2009. The Sequence Alignment/Map format and SAMtools. **25**: 2078–2079.
274 http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19505943&retmod
275 e=ref&cmd=prlinks.

276