Supplemental Experimental Procedures

Strain culture

We maintained *C. elegans* N2 animals on nematode growth media (NGM) plates seeded with either HB101 or OP50 bacterial cultures as described (Brenner, 1974).

Recombinant Mos1 insertions

All injections were carried out at room temperature on a custom built injection microscope with a gliding stage mounted on an inverted Zeiss microscope. We generated recombinant Mos1 insertions as described (Frøkjær-Jensen et al., 2014). Genomic insertion sites were identified by inverse PCR, Sanger sequencing, and mapping to the *C. elegans* genome. Strain names, exact insertion sites, and the flanking genomic DNA sequence are listed in **Table S1**.

MosSCI insertions

We generated MosSCI insertions as previously described (Frøkjær-Jensen et al., 2008, 2012) into an expanded set of universal MosSCI landing sites as previously described (Frøkjær-Jensen et al., 2014). We further validated the position of each landing site with oligos specifically designed to each insertion (oligo sequences are in **Table S1**). We generated a total of 15 new universal MosSCI insertions sites to complement the previously published universal insertion sites (6 sites) and MosSCI sites (6 sites). Strain names and insertion sites of all insertion sites are listed in **Table S1**.

Imaging

Germline fluorescence

We scored germline expression as all-or-none and cannot exclude that some strains categorized as partially or fully silenced had low levels of GFP expression that was undetectable by eye or that some strains with full germline expression were brighter than others. Quantification and strain information is included in **Table S2**.

Somatic fluorescence

Strains with somatic transgene variegation was based on injection of approximately ~200 animals and screening an estimated 800 independent strains for full or partial transgene silencing at the first larval stage (L1). Qualitatively, most strains were reproducibly bright and only a subset of strains displayed a "mottled" appearance with expression in a limited and variable number of somatic cell. Picking a single animal from each injection, we isolated 40 strains in which most cells exhibited decreased fluorescence (no strains were fully silenced) and 160 control strains with bright expression; of these we could unambiguously determine the transgene insertion sites in 32 partially silenced and 138 control strains. Plates with synchronized L1 populations were scored (blinded) for expression of the P*eft-3*:tdTomato:H2B transgene on a fluorescence dissection microscope. The strains were grouped into five arbitrary groups starting from the brightest to dimmest strains: Class 1 (4), Class 2 (3), Class 3 (2), Class 4 (1) and Class 5 (0) and given a numerical value corresponding to each category (in parenthesis). Quantification and strain information is included in **Table S2**.

Automated quantification of Somatic Gene Expression

L1 imaging, cell annotation and fluorescence quantification were performed as described previously (Liu et al., 2009; Long et al., 2009). Mixed-stage populations of several thousand transgenic nematodes where measured by using the COPAS-Profiler2 (Union Biometrica) as described previously (Dupuy et al., 2007). The larval stage of animals was estimated based on the time of flight through the flow cell.
Quantification and strain information is included in **Table S2**.

Single molecule fluorescence in situ hybridization (smFISH)

Hybridization: we ordered 37 custom Quasar 670 probes against the coding region of GFP (**Table S1**) and ready-made Stellaris RNA FISH hybridization and wash buffers (Biosearch Technologies, Petaluma, CA). We used the manufacturer's protocol for *C. elegans* larvae on dissected germlines from adult animals raised at 25°C. Imaging: prior to imaging stained animals were mounted overnight in ProLong Gold Antifade Mountant® (Thermo Fisher Scientific, Waltham, MA). The samples were imaged on an Eclipse Ni Microscope (Nikon Instruments), at 100x magnification with an oil immersion lens (Plan Apo delta, NA = 1.45) and a Zyla sCMOS camera (Andor Technology Ltd, Belfast, UK). Image stacks of dissected germlines were acquired sequentially with GFP, DAPI, and Cy5 filter sets and with LED excitation (Lumencor, Beaverton, OR). Quantification: The image stacks were deconvolved with NIS Elements Advanced Research software (Nikon Instruments). Max intensity projections of 5 slices were blinded and scored visually for the presence or absence of smFISH spots and GFP fluorescence; subsequently the images were quantified with the software StarSearch (Arjun Raj laboratory, U. Penn, http://rajlab.seas.upenn.edu/StarSearch/launch.html) with analysis settings fixed at the standard setting.

Molecular Biology

We generated all plasmids by standard molecular biology techniques, including Gateway Cloning (Invitrogen, CA), Gibson Assembly (Gibson et al., 2009), and Golden Gate cloning (Engler et al., 2009) which allowed us to simultaneously inserted three or four introns into a GFP backbone construct. The standard gfp was described in (Fire et al., 1998). The germline optimized *gfp* was optimized for high expression (Redemann et al., 2011) and we removed piRNA homology stretches with less than 4 mismatches based on complementarity to published piRNAs (Batista et al., 2008; Das et al., 2008; Gu et al., 2012). Several constructs used the two common germline promoters P*mex-5* (Zeiser et al., 2011) and P*pie-1* (Reese et al., 2000). Annotated Genbank sequences are included in **Data S1**.

Total RNA and small RNA sequencing

Total RNAs: We grew synchronized populations of animals at 25°C from bleached embryos and picked 20 young adult animals for RNA isolation. RNA isolation was performed by vortexing animals in Trizol for 20 minutes following guidelines from Johnstone et al. (ed. Hope, 1999). Ribosomal RNAs were depleted by incubating the total RNA sample with a mix of 94 DNA oligos (AF-NJ-16 through AF-NJ-107, AF-NJ-150 and AF-NJ-151) complementary to rRNA and samples were treated with Hybridase Thermostable RNase H (Epicentre, Illumina, San Diego, CA) followed by TURBO DNase treatment (Ambion, ThermoFisher Scientific, Waltham, MA) to remove oligos. We generated sequencing libraries with a SMARTer Stranded kit (Clontech Laboratories, Mountain View, CA) and sequenced the libraries on a miSeq instrument (Illumina, San Diego, CA). The reads were processed to remove bases corresponding to template switching oligos in the R1 read and bases from random hexamer priming in the R2 read. We aligned reads that mapped uniquely with no mismatches to the *C. elegans* protein coding transcriptome (WS245) with the program STAR (Dobin et al., 2013) and used custom Python scripts to count reads to each gene.

Small RNAs: We grew synchronized populations of animals at 25°C from bleached embryos to the young adult stage and washed off one large plate of animals for small RNA isolation. Small RNAs were isolated by freeze grinding worms in liquid nitrogen and small RNAs were captured with a mirVana kit (Ambion, ThermoFisher Scientific, Waltham, MA). We generated sequencing libraries with a TruSeq small RNA kit and sequenced the libraries on a miSeq instrument (Illumina, San Diego, CA). The reads were trimmed for adapter sequence at the 3' end and reads that mapped uniquely with no mismatches to the *C. elegans* protein coding transcriptome (WS245) or the transgene insertion in each strain, were aligned with the program STAR (Dobin et al., 2013). We used custom Python scripts to count reads to each gene for all samples.

Data analysis

We performed data analysis in R v.3.1.2 (R Core Team, 2014) with the IDE R Studio (RStudio Team, 2015) and the packages RColorBrewer (Neuwirth, 2014) and hexbin (Carr et al., 2015). Germline expression was based on RNA sequencing of isolated sperm (Ma et al., 2014), RNA sequencing of dissected germlines from *fem-3* and *fog-2* mutant animals (Ortiz et al., 2014), RNA sequencing of isolated single cell oocytes (Stoeckius et al., 2014), and SAGE analysis of dissected germlines from wildtype animals (Wang et al., 2009).

Statistical analysis

We performed statistical analyses with GraphPad Prism 6 (La Jolla, CA) and specific statistical test are described in the legends of each figure. In general, germline expression data did not follow a normal distribution with stochastic silencing often being an "all-or-none" phenomenon for each strain; we therefore preferentially used rank tests. For all experiments with more than two samples we first analyzed the data set for overall statistically significant differences (ANOVA) and subsequently compared the indicated datasets with corrections for multiple comparisons applied to the stated P values.

PATC analysis

We used the PATC algorithm, described in Fire et al., (2006), to quantify PATCs in individual genes and introns (**Table S3**). In brief, the PATC algorithm analyzes DNA sequences for two characteristics: (1) the presence of A_n/T_n clusters within 5-basepair segments, and (2) the relative spacing of any such clusters. The algorithm assigns high scores to DNA sequences that contain "perfect" A_n/T_n clusters (AAAAA or TTTTT) spaced by exactly one helical DNA turn (10 basepairs). Lower scores are assigned to DNA sequences with less than perfect A_n/T_n clusters (for example, GAAAA and TTTAT) and with clusters spaced shorter or further apart than one turn of the DNA helix (9, 11, or 12 basepairs). The algorithm starts at a single basepair in the analyzed DNA sequences and continues to extend along the sequence as long as additional A_n/T_n clusters are found approximately one helical DNA turn ahead. In practice, this is done by assigning higher positive scores for "good' A_n/T_n clusters and penalties for G/C-rich clusters and non-canonical DNA spacing of clusters. The algorithm terminates the extension when the total score starts decreasing. Any given sequence of DNA is therefore associated with a continuous PATC score, where large positive values indicate the presence of highly phased A_n/T_n clusters along one face of the DNA helix.

By setting a minimal threshold score, the algorithm can be used to define discrete PATC-rich regions that contain regularly spaced A_n/T_n clusters. Less than 0.1% of random DNA sequences generate a PATC score above 60 (Fire et al., 2006) and we have therefore used that as a cut-off to define endpoints for PATC-rich regions. We note that this threshold was chosen to minimize false positive PATC signals and was not justified by any evidence of a meaningful discrete biological threshold for PATCs. We have indicated PATC-rich regions with PATC scores above 60 below transgenes in figures with black boxes. The discrete regions can be visualized by loading the continuous PATC values contained in **Data S2** in a genome browser (for example, Integrative Genomics Viewer (https://www.broadinstitute.org/igv/)) and setting the min/max display values to 59 and 60, respectively.

Because the PATC algorithm is "greedy" and keeps extending along DNA sequences, there is no meaningful upper limit on absolute PATC scores (other than the length of chromosomes). It is therefore useful to have a measure of PATCs that does not scale with length, for example to compare PATCs in introns of different lengths. To normalize, the PATC density is defined as the total sum of PATC values for the sequence divided by the length of the sequence (Fire et al., 2006); the upper limit for the maximum PATC-density approaches 2190 for perfect A_n/T_n clusters spaced 10 basepairs apart for extended stretches. As an example, one of the endogenous introns (intron 1 from wbgene00021132) inserted into *gfp* had a total PATC score of 136,732 and a length of 281 basepairs for a PATC density of 487 (**Table S3**). For the same intron, there was a discrete stretch of 270 basepairs with a PATC score above 60 ("length of PATC-rich region") and the overall PATC frequency of the intron was 96% (270bp/280bp). By definition, the PATC density contains information about the spacing and composition of A_n/T_n clusters whereas the PATC frequency only depends on an arbitrary threshold level (here chosen = 60). We primarily used the PATC density to compare different sequence elements (genes and introns) within the same species, and the PATC frequency to compare between species (for example, between *C. elegans* and *C. briggsae*).

We also developed a slightly modified PATC algorithm and we refer to the updated algorithm as $PATC_{balanced}$. The modified algorithm allows subtraction of "off-helical" A_n/T_n signals in an effort to reduce false-positive PATC signals in repeat regions and A/T rich genomes. The PATC values of whole genome sequences were calculated with the balanced algorithm. We have generated continuous traces of balanced PATC scores with a 25 base pair resolution of the *C. elegans* (WS245) and *C. briggsae* (WS245) genomes in bigwig format and included these in **Data S2**.

The PATC algorithm was written in Pascal (source code in Fire et al. (2006)) and compiled on an Apple computer with the program fpc (http://www.freepascal.org/). The algorithm is quick and can analyze the full genome sequence of *C. elegans* (~100 MB) in approximately 15 minutes on a standard personal computer.

Analysis of *C. elegans* and *C. briggsae* orthologs

The analysis was based on WS245 builds of *C. elegans* and *C. briggsae* genomes. Ortholog pairs, the calculated PATC values, expression level in female oocytes from (Stoeckius et al., 2014), and the classification of genes into intra-chromosomal and intra-chromosomal are included in **Table S4**. Because the transition between high and low recombination frequency (on chromosome arms and centers, respectively) approximates the transition between high and low density of repressive histone marks, we used recombination frequency in *C. elegans* and *C. briggsae* (Ross et al., 2011) to classify genes as "central" versus "arm". Although most gene transpositions in nematodes are intra-domain and occur within the same chromosome (Ross et al., 2011), we could also detect inter-domain gene transfer events. We limited our analysis to unique ortholog pairs that: (1) had moved within the same autosome, (2) were not within 1MB of a center to arm transition, and (3) the length of the coding regions were comparable (within 34 amino acids). We furthermore divided the orthologs into germline expressed or non-germline expressed based on expression in female *C. elegans* oocytes (Stoeckius et al., 2014). With these criteria we identified 2255 intra-domain ortholog retentions (1865 genes in center, 390 genes on arm) and 455 inter-domain ortholog shifts (334 genes that are central in *C. elegans* and peripheral in *C. briggsae*, and 121 genes that are central in *C. briggsae* and peripheral in *C. elegans*).

Nematode genomes were obtained from Wormbase May 2015. Nematode phylogeny is based on NCBI taxonomy and generated with PhyloT and Interactive Tree of Life (Letunic and Bork, 2011). Unique orthologs were determined based on the EnsemblCompara Gene Tree algorithm (Vilella et al., 2009).

Supplemental References

Carr, D., Lewin-Koh, ported by N., Maechler, M., and Sarkar, contains copies of lattice functions written by D. (2015). hexbin: Hexagonal Binning Routines.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. 29, 15–21.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS ONE 4, e5553.

Fire, A., Ahnn, Kelly, W., Harfe, B., Kostas, S., Hsieh, J., Hsu, M., and Xu, S. (1998). GFP applications in C. elegans. In the book Green Fluorescent Protein: Properties, Applications, and Protocols (eds. Chalfie and Kain) (NY: John Wiley and Sons).

Frøkjær-Jensen, C., Davis, M.W., Ailion, M., and Jorgensen, E.M. (2012). Improved Mos1-mediated transgenesis in C. elegans. Nat. Methods 9, 117–118.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345.

Gu, W., Lee, H.-C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and Mello, C.C. (2012). CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors. Cell 151, 1488–1500.

ed. Hope, I.A. (1999). C. elegans: A Practical Approach: A Practical Approach. (Oxford University Press).

Letunic, I., and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39, W475–W478.

Long, F., Peng, H., Liu, X., Kim, S.K., and Myers, E. (2009). A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat. Methods 6, 667–672.

Ma, X., Zhu, Y., Li, C., Xue, P., Zhao, Y., Chen, S., Yang, F., and Miao, L. (2014). Characterisation of Caenorhabditis elegans sperm transcriptome and proteome. BMC Genomics 15, 168.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes.

R Core Team (2014). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Redemann, S., Schloissnig, S., Ernst, S., Pozniakowsky, A., Ayloo, S., Hyman, A.A., and Bringmann, H. (2011). Codon adaptation-based control of protein expression in C. elegans. Nat. Methods 8, 250–252.

RStudio Team (2015). RStudio: Integrated Development Environment for R (Boston, MA: RStudio, Inc.).

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19, 327–335.

Wang, X., Zhao, Y., Wong, K., Ehlers, P., Kohara, Y., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hansen, D. (2009). Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. BMC Genomics 10, 213.

Zeiser, E., Frøkjær-Jensen, C., Jorgensen, E., and Ahringer, J. (2011). MosSCI and gateway compatible plasmid toolkit for constitutive and inducible expression of transgenes in the C. elegans germline. PloS One 6, e20082.