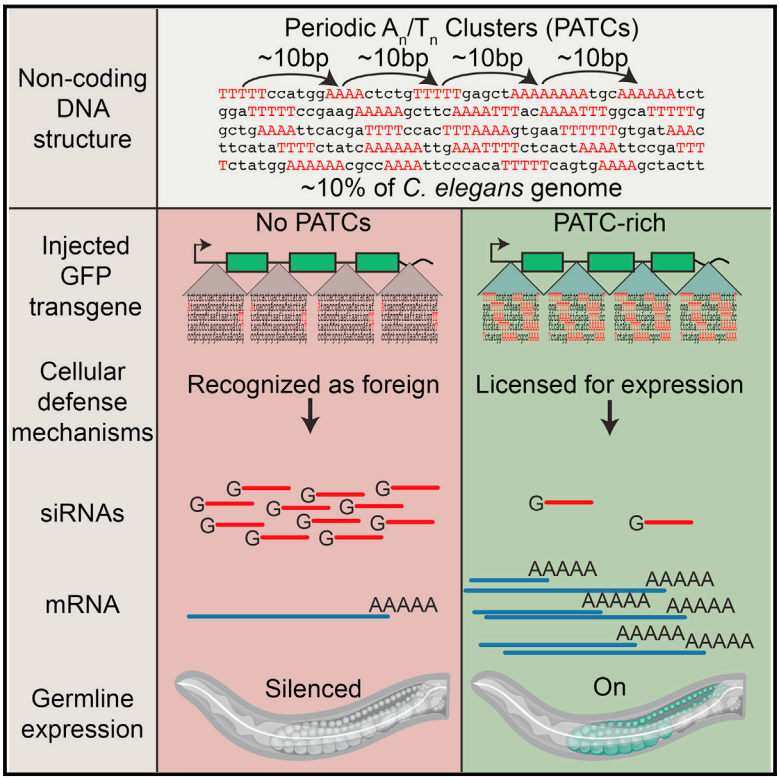
Article

Cell

An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the *C. elegans* Germline

Graphical Abstract



Highlights

- Periodic, non-coding DNA can prevent transgenes from stochastic silencing in germline
- Non-coding content of genes is shaped by genomic context and heterochromatin domains
- Conditioning of active DNA may allow cells to distinguish foreign from host genes

Authors

Christian Frøkjær-Jensen, Nimit Jain, Loren Hansen, ..., Denis Dupuy, Erik M. Jorgensen, Andrew Z. Fire

Correspondence

jorgensen@biology.utah.edu (E.M.J.), afire@stanford.edu (A.Z.F.)

In Brief

A non-coding DNA sequence pattern can prevent epigenetic silencing of transgenes in germ cells, revealing a functional role for an abundant class of non-coding DNA and a possible mechanism by which cells may recognize and silence foreign DNA.

Frøkjær-Jensen et al., 2016, Cell *166*, 343–357
July 14, 2016
http://dx.doi.org/10.1016/j.cell.2016.05.072

CellPress

Article

Cell

An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the *C. elegans* Germline

Christian Frøkjær-Jensen,[1,2,3] Nimit Jain,[4] Loren Hansen,[2] M. Wayne Davis,[1] Yongbin Li,[8] Di Zhao,[8] Karine Rebora,[6] Jonathan R.M. Millet,[6] Xiao Liu,[5,8] Stuart K. Kim,[5,7] Denis Dupuy,[6] Erik M. Jorgensen,[1,*] and Andrew Z. Fire[2,7,*]

[1]Department of Biology, Howard Hughes Medical Institute, University of Utah, Salt Lake City, UT 84112, USA
[2]Department of Pathology, Stanford University, Stanford, CA 94305, USA
[3]Department of Biomedical Sciences and Danish National Research Foundation Centre for Cardiac Arrhythmia, University of Copenhagen, 2200 Copenhagen N, Denmark
[4]Department of Bioengineering, Stanford University School of Medicine, Stanford, CA 94305, USA
[5]Department of Developmental Biology, Stanford University Medical Center, Stanford, CA 94305, USA
[6]IECB, University of Bordeaux, Laboratoire ARNA-INSERM, U869, 33600 Pessac, France
[7]Department of Genetics, Stanford University, Stanford, CA 94305, USA
[8]School of Life Sciences, Tsinghua University, Beijing 100084, China
*Correspondence: jorgensen@biology.utah.edu (E.M.J.), afire@stanford.edu (A.Z.F.)
http://dx.doi.org/10.1016/j.cell.2016.05.072

SUMMARY

Cells benefit from silencing foreign genetic elements but must simultaneously avoid inactivating endogenous genes. Although chromatin modifications and RNAs contribute to maintenance of silenced states, the establishment of silenced regions will inevitably reflect underlying DNA sequence and/or structure. Here, we demonstrate that a pervasive non-coding DNA feature in *Caenorhabditis elegans*, characterized by 10-base pair periodic A_n/T_n-clusters (PATCs), can license transgenes for germline expression within repressive chromatin domains. Transgenes containing natural or synthetic PATCs are resistant to position effect variegation and stochastic silencing in the germline. Among endogenous genes, intron length and PATC-character undergo dramatic changes as orthologs move from active to repressive chromatin over evolutionary time, indicating a dynamic character to the A_n/T_n periodicity. We propose that PATCs form the basis of a cellular immune system, identifying certain endogenous genes in heterochromatic contexts as privileged while foreign DNA can be suppressed with no requirement for a cellular memory of prior exposure.

INTRODUCTION

Invasive DNA derived from viruses, retrotransposons, and DNA transposons constitute a substantial challenge to organisms. Uncontrolled replication of transposable elements will compromise the host's genome (Malone and Hannon, 2009) and consequently, cellular defense mechanisms have evolved to detect and silence foreign DNA. These mechanisms are particularly well developed in germ cells, where deleterious changes will impact the fitness of subsequent generations. In eukaryotes, several classes of small RNAs (~20–30 nucleotides long) form complexes with Argonaute proteins to silence foreign nucleic acids by degrading target mRNAs (Zamore et al., 2000) and by transcriptional silencing via RNA-directed heterochromatin formation (Volpe et al., 2002). In germ cells, genome surveillance is mediated in part by a large class of small RNAs (Piwi-interacting RNAs [piRNAs]) that interact with Argonautes from the Piwi clade (e.g., Aravin et al., 2007; Batista et al., 2008; Das et al., 2008).

Small RNAs act as a recognition system for transcription from invasive DNA, but introduce the danger of silencing endogenous genes. In *Caenorhabditis elegans*, several potentially-related mechanisms have been proposed to protect endogenous genes against silencing. At a chromatin level, activating H3K36 histone marks deposited on germline-expressed genes provide a positive feed-forward mechanism promoting expression in subsequent generations (Rechtsteiner et al., 2010). Operating alongside and/or in parallel, RNA-based protection systems have been shown to use maternal RNAs to license forthcoming expression of genes (Johnson and Spence, 2011). A licensing mechanism based on small RNAs associated with the CSR-1 Argonaute protein has been proposed to protect endogenous genes from piRNA-mediated silencing (e.g., Shirayama et al., 2012). Notably, these mechanisms would only propagate germline expression decisions made in prior generations. What other features, such as genome position or DNA sequences, determine initial licensing for germline expression has remained an open question. Specifically, how can a gene be selected (or rejected) for expression if there is no chromatin mark or small RNA population to indicate whether the gene was expressed in prior generations? To probe this question, we focused on de novo expression and silencing of transgenes in the germline of *C. elegans*.

Transgenes have been a useful tool to determine the effects of large-scale genome organization in several organisms, most notably *Drosophila* (Elgin and Reuter, 2013). In *C. elegans*, transgenes are notoriously difficult to express in the germline, with

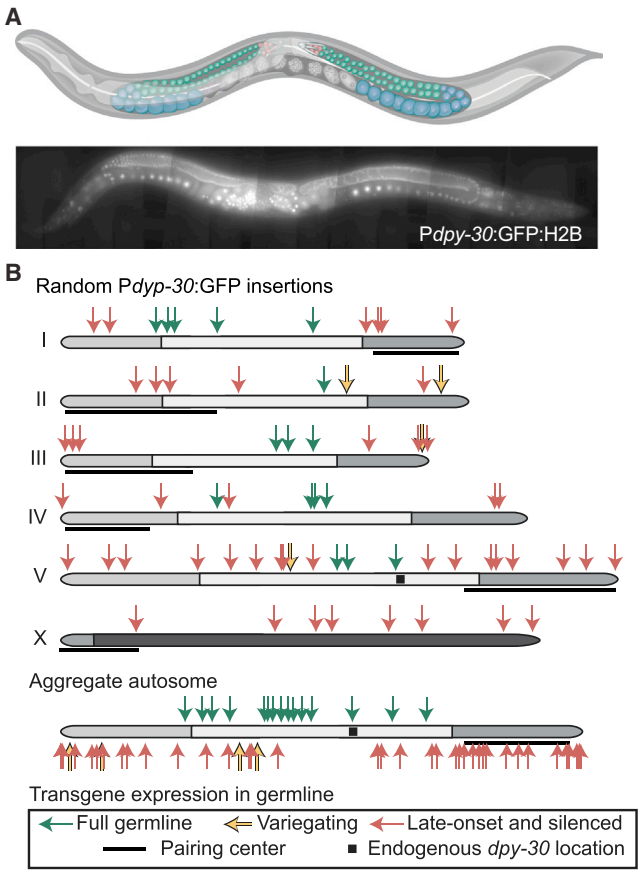


Figure 1. Many P*dpy-30*:GFP Transgenes Are Specifically Silenced in the Germline

(A) Top: schematic of *C. elegans* hermaphrodite with the female germline highlighted. Red, mitotic cells; green, early meiotic cells; blue, late meiotic cells. Bottom: composite fluorescence image of animal expressing a P*dpy-30*:GFP transgene (42× magnification, scale bar, 50 μm). Graphic of *C. elegans* modified from "*Caenorhabditis elegans* hermaphrodite adult-en.sv" by K.D. Schroeder from Wikimedia Commons under a CC-BY-SA 3.0 license.

(B) Top: germline expression at 25°C of P*dpy-30*:GFP transgenes inserted randomly by Mos1 transposition. Genomic insertion sites, transgene copy number, and somatic expression of all insertions were previously verified (Frøkjær-Jensen et al., 2014). Germline fluorescence is indicated by colored arrows (key at bottom). Black squares indicate the endogenous location of *dpy-30*, gray shading indicates enrichment of repressive histone modifications (Liu et al., 2011), and pairing centers are indicated with black lines (MacQueen et al., 2005). Bottom: aggregated normalized autosomes aligned with pairing centers to the right and P*dpy-30*:GFP insertions.

See also Figure S1 and Table S2 for details of classification.

rapid silencing of episomal DNA (Kelly et al., 1997) and progressive silencing of many single-copy genomic transgenes (Shirayama et al., 2012). Here, we study a stochastic process in which some single-copy insertions variegate in somatic cells and are frequently silenced in germ cells. Variegation and silencing mirror a chromosomal pattern that corresponds to the organization of the genome into broad domains (e.g., Liu et al., 2011). We find that a non-coding DNA structure, called periodic A_n/T_n clusters (PATCs) (Fire et al., 2006) can license transgenes for expression in the germline. Thus, PATCs constitute an abundant class (comprising ~10% of the *C. elegans* genome) of functionally important non-coding DNA in nematodes that may safeguard endogenous genes from silencing in repressive chromatin environments. We propose that lack of PATCs in foreign DNA may be one characteristic used by nematodes to silence invasive genetic material.

RESULTS

Germline Expression Is Sensitive to Large-Scale Genome Organization

To explore transcriptionally permissive and repressive genomic regions for germline expression, we inserted a ubiquitously expressed transgene (P*dpy-30*:GFP:H2B) into random locations by transposition (Frøkjær-Jensen et al., 2014) (Figure 1A; Table S1). Of 67 insertion strains analyzed, all had visible somatic expression but most (51 strains) had limited germline expression. We categorized germline expression from animals maintained at 25°C into three classes: complete germline expression (mitosis, early, and late meiosis), early germline silencing (fluorescence only visible in late meiosis), and full germline silencing (no visible expression in any germ cells) (Figures 1A and S1). Additionally, some strains showed transgene variegation and were categorized as "variable."

Strains with early or full germline silencing appeared to cluster in non-random chromosomal patterns (Figure 1B). First, all insertions into the X chromosome showed early germline silencing (Figure S1). This pattern is consistent with broad inactivation of the X chromosome in the early meiotic germline by homologs of the Polycomb repressive complex 2 (PRC2) (Fong et al., 2002) and de-repression at late meiotic stages (Kelly et al., 2002). Second, 15 of 25 strains with transgenes inserted within the central 50% of autosomes were expressed, whereas only 1 of 30 transgenes inserted into the distal 25% of autosome arms were expressed in the germline ($p < 0.001$, Fisher's exact test) (Figure 1B). This pattern of expression is in agreement with the observed central and distal autosomal domains based on recombination frequency (Brenner, 1974; Rockman and Kruglyak, 2009), histone modifications (Gu and Fire, 2010; Liu et al., 2011), and heterochromatin protein 1 (HP1) distribution (Garrigues et al., 2015). Targeted insertion of a P*dpy-30*:GFP transgene (Frøkjær-Jensen et al., 2008, 2014) showed a similar pattern of germline expression and frequent position-dependent stochastic silencing on autosome arms (Figure S1).

Transgene Expression in the Soma Is Position-Dependent and Variegates

Although we were primarily interested in germline-specific silencing, we also tested whether somatic transgene expression was position-dependent. We generated random insertions with a strong ubiquitous promoter *eft-3* (*eef-1A.1*) that expressed a bright red fluorophore (*tdTomato*) and screened transgenic animals for low expression at the L1 stage. Qualitatively, most strains showed reproducibly bright expression, and we were unable to isolate any animals with fully silenced transgenes. However, a subset of strains (40 of ~800 insertions) showed "mottled" expression in a limited number of somatic cells

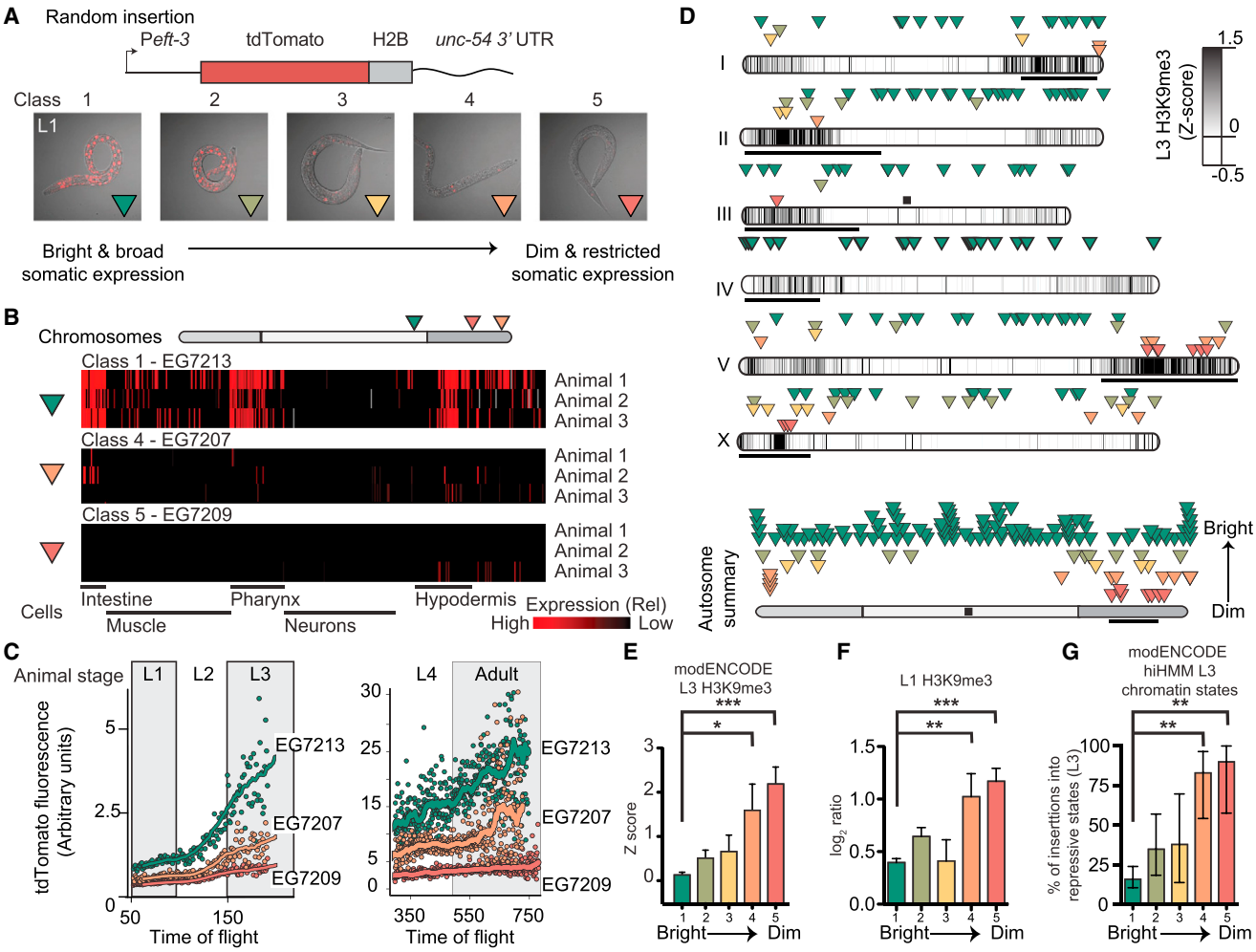


Figure 2. Somatic Transgene Silencing Is Frequent on Chromosome Arms

(A) Top: schematic of the P*eft-3*:*tdTomato* transgene. Bottom: overlay of tdTomato fluorescence and transmitted light images (identical exposure) of representative L1 animals from each expression class.

(B) Automated analysis of single cell fluorescence in L1 animals (Liu et al., 2009) of one bright and two dimmer P*eft-3*:*tdTomato*:*h2b* insertions. Each horizontal line represents replicate imaging of different animals from the same strain. The brightness of each red vertical line indicates the level of tdTomato fluorescence in individual, identified cells. Gray indicates failed cell identification. The cells are clustered based on general classes of tissue and indicated below.

(C) tdTomato fluorescence intensity at different larval stages (L1 to adult) based on flow cytometry (Dupuy et al., 2007). Average peak tdTomato fluorescence is indicated for each animal with a dot and the line indicates a smoothed average of peak fluorescence.

(D) Top: genomic location and average brightness of each strain (green, bright and broad expression; red, dim and restricted expression). Darker chromosome shading indicates higher H3K9me3 density at the L3 stage (Liu et al., 2011), black bars indicate the approximate location of pairing centers (MacQueen et al., 2005), and the black box indicates the endogenous location of *eft-3*. Bottom: aggregate normalized autosome with all insertions.

(E) Average H3K9me3 *Z* score in L3 animals (Liu et al., 2011) in a 2-kb interval centered on insertions from each class. Average ± SEM. Statistics: Mann-Whitney test (**p < 0.01, ***p < 0.001).

(F) Average H3K9me3 level (this work) in starved L1 animals in a 2-kb interval centered on insertions from each class. Log2 ratio between sample and input. Average ± SEM. Statistics: Mann-Whitney test (**p < 0.01, ***p < 0.001).

(G) Percentage of inserts from each class (class 1–class 5) that was inserted into repressive chromatin states (Polycomb or heterochromatin, states 10–13) identified at the L3 stage based on hierarchical non-parametric machine-learning (hiHMM) (Ho et al., 2014). Error bars indicate 95% confidence interval. Statistics: Fischer's exact test of class 1 (brightest) compared to classes 2–5 (dimmer) (**p < 0.01).

See also Figure S2 and Table S2 for higher resolution images and expression quantification.

(Figure 2A). We scored the fluorescence qualitatively at the L1 stage by assigning animals to five expression classes (class 1 [brightest] to class 5 [most silenced]) (Figure 2A; Table S2).

Insertions into chromosome centers generally belonged to the two brightest classes whereas most dim transgenes were inserted into distal regions (Figure 2D). Genes silenced in the L1 stage corresponded to local genomic environments (±1 kb) previously found to be significantly enriched in repressive H3K9me3 histone marks at a later larval (L3) stage (Liu et al., 2011) (Figure 2E). We generated an H3K9 methylation dataset from L1

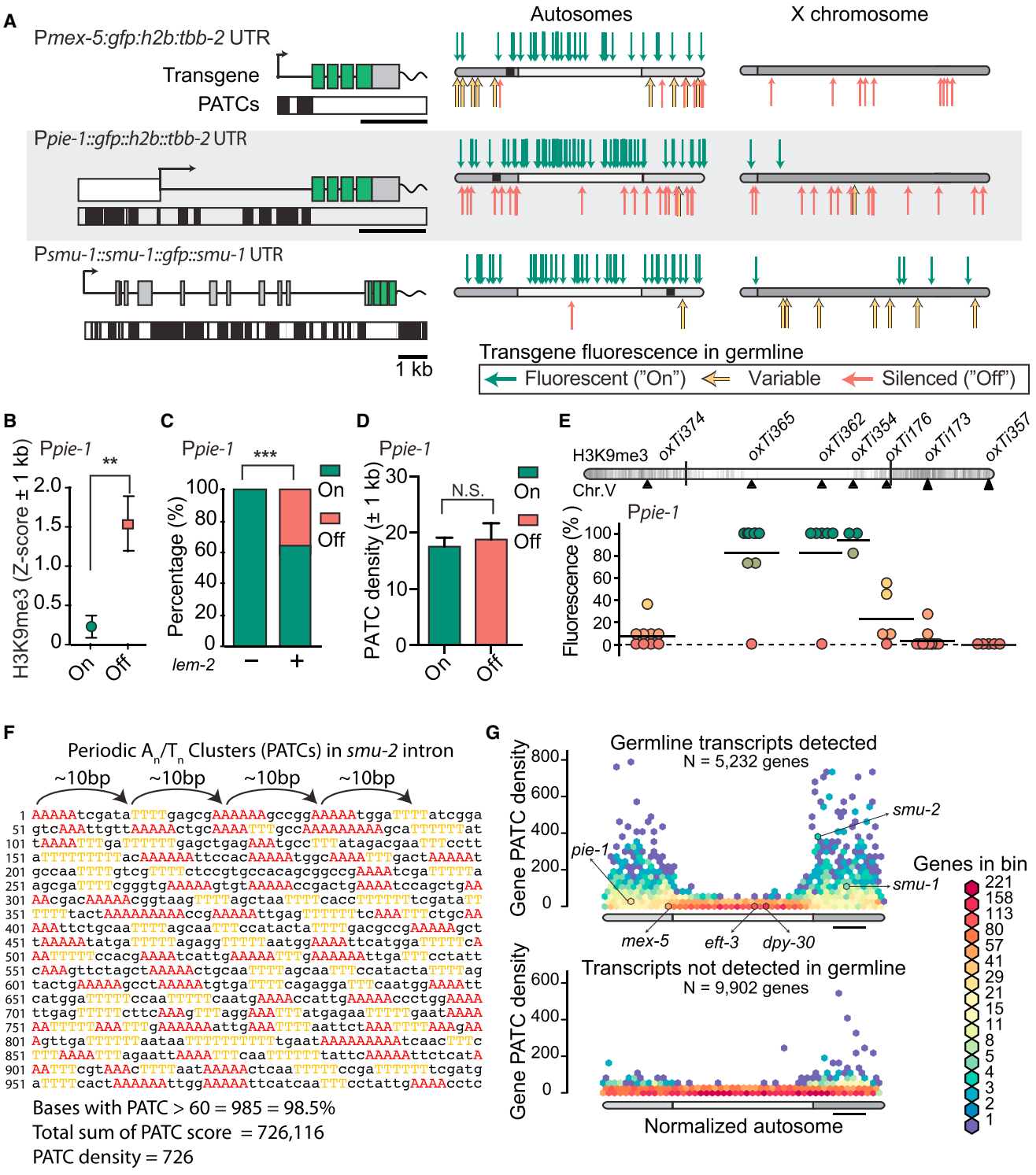


Figure 3. Transgenes Containing PATCs Are Less Frequently Silenced

(A) Expression of P*pie-1:gfp*, P*mex-5:gfp* and *smu-1:gfp* transgenes. Left: transgene schematics with PATCs >60 indicated below as black boxes. Right: location and germline expression of insertions on aggregated autosomes and the X chromosome. Endogenous locations of *pie-1*, *mex-5*, and *smu-1* are indicated with black squares.

(B) Local chromatin environment (2-kb interval centered on insertion sites) near P*pie-1* insertions. H3K9me3 signal from early embryos (Liu et al., 2011), which have been used as a proxy for germline tissue (Rechtsteiner et al., 2010). Mean ± SEM. Statistics: Mann-Whitney test (**p < 0.01).

(legend continued on next page)

animals and verified this association between transgene expression and H3K9me3 histone marks (Figure 2F). Finally, silenced insertions were frequently inserted into chromatin identified as repressive based on global chromatin marks at the L3 stage (Polycomb repressed and heterochromatin states, hiHMM 10–13) (Ho et al., 2014) (Figures 2G and S2C).

We also measured transgene variegation with quantitative imaging techniques on strains from class 1, class 4, and class 5. One method automatically identified and assigned fluorescence levels to 363 of the 558 cells from fixed L1 animals (Liu et al., 2009). L1 imaging detected expression in significantly more cells and higher levels of expression in the class 1 strain compared to class 4 and class 5 strains (Figures 2B and S2A) and a higher degree of variation in dim strains compared to bright strains was evident (e.g., E lineage, Figure S2B). A second technique relied on imaging live worms across different developmental stages by flow cytometry (Dupuy et al., 2007), which confirmed higher transgene expression in class 1 compared to class 4 and class 5 at the L1 stage (EG7213 = 19.2 ± 0.4, EG7207 = 11.3 ± 0.2, EG7209 = 7.9 ± 0.1, mean ± SEM) (Figure 2C). At later larval stages, the two class 1 and class 4 strains increased in absolute fluorescence, whereas the one class 5 strain remained mostly dim throughout development (Figure 2C). Thus, transgene silencing in early larval stages is not necessarily a permanent state, possibly tracking developmental changes in local chromatin environments (Meister et al., 2010).

In sum, we observed evidence for transgene variegation and an influence of large-scale genome domains on somatic expression. At the same time, only a small subset of insertions variegated. These data extend the long-standing observation that it is easier to express transgenes in the soma compared to the germline of *C. elegans* (e.g., Kelly et al., 1997).

Transgene Silencing in the Germline Depends on Promoter Elements

The silencing character of any given chromosomal region might conceivably be universal (so that any insertion would be shut down) or insertion-dependent (capable of silencing some transgenes but not others). We examined whether position-dependent germline silencing was transgene-dependent by using two germline-specific promoters: P*mex-5* and P*pie-1*. Similar to P*dpy-30*:GFP (Figure 1B), P*mex-5*:GFP and P*pie-1*:GFP transgenes were mostly silenced on the X chromosome and frequently silenced on autosome arms compared to centers (GFP-positive: P*mex-5*, 11 of 17 arm, 25 of 25 center, $p < 0.01$; P*pie-1*: 28 of 48 arm, 34 of 37 center, $p < 0.01$; Fischer's test) (Figure 3A). P*pie-1*:GFP insertions were more frequently silenced in genomic regions with high levels of H3K9me3 (early embryo [EE]) (Liu et al., 2011) (Figure 3B) and in regions that immunoprecipitate with a nuclear lamina protein (LEM-2) (Ikegami et al., 2010) (Figure 3C). However, the *mex-5* and *pie-1* promoters were significantly more active compared to P*dpy-30* from autosome arms (GFP-positive insertions: P*mex-5*, 11 of 17 arm; P*pie-1*: 28 of 48 arm; P*dpy-30*: 2 of 31 arm, $p < 0.01$, Fischer's, Figure 3C) and from repressive hiHMM chromatin states (Ho et al., 2014) (Figure S3F).

Targeted insertions of P*pie-1*:GFP showed a similar pattern of germline expression and stochastic silencing (GFP-positive insertions: 13 of 18 center, 0 of 24 arm, $p < 0.01$, Fischer's) (Figure 3E). One insertion site (*oxTi176*) near the transition between domains of low and high H3K9me3 showed highly variable P*pie-1*:GFP (Figure 3E) and P*mex-5*:GFP (Figures S3I) expression; it is possible that insertions into chromosomal locations bordering heterochromatin are particularly prone to stochastic silencing and position effect variegation, similar to what has been observed in *Drosophila* (Elgin and Reuter, 2013).

These data provide support for genome position as a strong determinant of germline expression and suggest that some germline promoters are more resistant to silencing imposed by heterochromatic domains.

A Transgene Rich in Periodic A/T Clusters Is Expressed from Repressive Chromatin Domains

To resolve how endogenous genes are protected from stochastic silencing in the germline, we looked for a common DNA character that might safeguard genes from the surrounding repressive heterochromatin environment. Periodic A_n/T_n clusters (PATCs) are an abundant class of non-coding DNA that are enriched in germline expressed genes on autosome arms (Fire et al., 2006) and are anti-correlated with H3K9 methylation (Gu

(C) Local chromatin interactions with nuclear lamin (2-kb interval) near P*pie-1* insertions. Nuclear lamin interactions based on chromatin immunoprecipitation sequencing (ChIP-seq) on mixed stage embryos with an antibody against the transmembrane nuclear protein *lem-2* (Ikegami et al., 2010). Mean ± SEM. Fischer's exact test (***$p < 0.001$).

(D) Local PATC density (2-kb interval) near P*pie-1* insertions. Mean ± SEM. Statistics: Mann-Whitney test (N.S., not significant).

(E) Germline expression of targeted, single-copy P*pie-1*:GFP:H2B:*tbb-2* UTR insertions into universal MosSCI sites on Chr. V (see also Figure S1). Circles indicate independent transgene insertions that are color coded for each insertion's expression in the germline (11 animals scored, horizontal bar = mean of independent insertions). Darker chromosome shades correspond to higher H3K9me3 density in early embryos (Liu et al., 2011).

(F) Example of a periodic A_n/T_n cluster (PATC) from intron 3 of *smu-2*. Clusters of three, four, or five adjacent As and Ts are colored. Clusters of As and Ts are separated by ~10 bps (approximately one helical DNA turn) and short A_n and T_n clusters therefore align along one face of the DNA helix over an extended region (here, 1 kb). The PATC algorithm (Fire et al., 2006) assigns a PATC value to every nucleotide of a DNA sequence; higher values indicate that the nucleotide is part of an extended DNA stretch ("PATC-rich region") with many clusters of A_n/T_n clusters in perfect 10-bp register. Less than 0.1% of nucleotides in a random DNA sequence reach a PATC value of 60. The PATC density is defined as the average PATC value of nucleotides in a sequence. See Fire et al. (2006) and the Supplemental Information for details.

(G) Comparison of the PATC density of autosomal, protein-coding genes as a function of chromosome position. Top: germline-expressed genes (based on >2 fragments per kilobase of exon per million fragments mapped [FPKM] expression in 1-cell oocytes) (Stoeckius et al., 2014). Bottom: genes with no detectable germline expression. Gene distributions were hexagonally binned on a logarithmic frequency scale. A subset of genes used in this study are indicated with arrows. See also Figure S3 and Table S3.

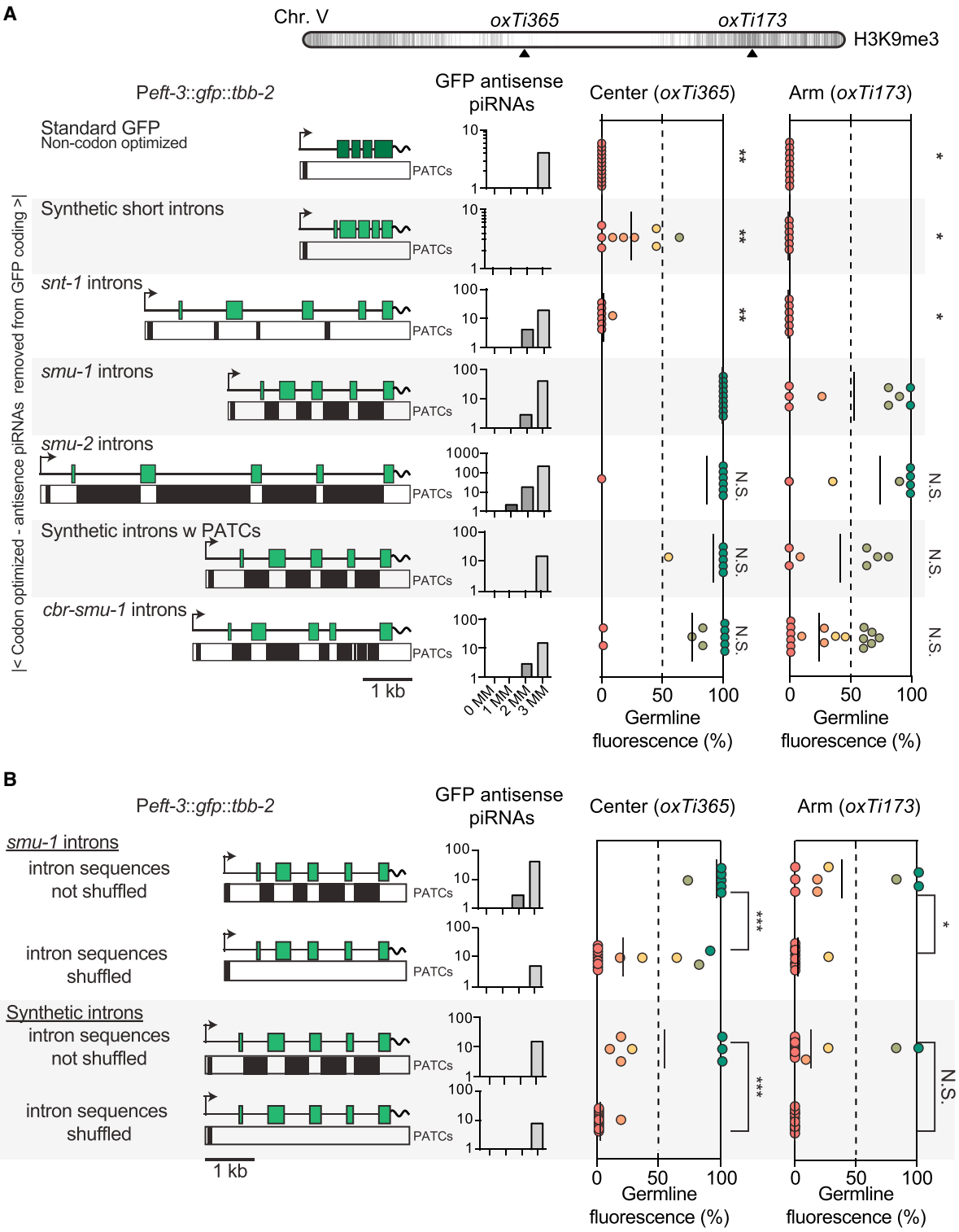


Figure 4. PATCs in GFP Introns Reduce Germline Silencing

All transgenes were expressed in somatic cells. P*eft-3*:*gfp* transgenes with *smu-1* introns and synthetic PATCs in (A) and (B) are the same transgene constructs; however, independent insertions were generated in parallel to the shuffled transgene insertions for comparison under identical conditions.

(A) Germline expression of transgenes inserted into chromosome V at 25°C. Top: location of MosSCI insertion sites on Chr. V. Left: P*eft-3*:*gfp* transgenes with PATCs >55 indicated as black boxes underneath. Center: piRNA homology of transgenes allowing 0, 1, 2, or 3 mismatches (MM). Right: germline expression from

(legend continued on next page)

and Fire, 2010). PATCs are composed of short clusters of adenines and thymines spaced ~10 base pairs (bps) apart, thereby "coating" one face of the DNA helix with A_n/T_n clusters over extended runs (Figure 3F). It is notable that the relatively silencing-resistant *pie-1* and *mex-5* promoters contain PATCs (Figure 3A) whereas P*dpy-30* does not (not shown). We confirmed the previously established positive association between germline expression and PATCs (Fire et al., 2006) with more recent gene models (WS245) and gene expression profiles from isolated germlines and single-cell oocytes (Ortiz et al., 2014; Stoeckius et al., 2014; Wang et al., 2009) (Figures 3G and S3).

As a starting point for investigating the role of PATCs, we generated random insertions of a *gfp*-tagged *smu-1* gene (*smu-1*:*gfp*). *smu-1* has moderately strong overall PATC content (Figure 3G) with PATCs distributed across the endogenous promoter, gene body, and 3′ UTR (Figure 3A). *smu-1*, and the related gene *smu-2*, have the highly unusual property that simple, extrachromosomal arrays (hereditary, highly repetitive episomal DNA structures) with these genes are readily expressed in both soma and germ cells (Spartz et al., 2004; Spike et al., 2001). When inserted randomly, all *smu-1*:*gfp* insertions were expressed in somatic cells and in the germline when cultured at 25°C (with one exception, which is likely a damaged transgene insertion). Notably, despite some initial silencing, X-linked *smu-1*:GFP transgene inserts were also expressed in the full germline after propagation for three to four generations (Figure 3A). Germline de-silencing over time is not a general feature of transgenes; X-linked P*mex-5* insertions remained silenced over the same number of generations (five of five P*mex-5* strains, Table S2). These observations are unlikely to be explained by relative promoter strength: *smu-1* expression is generally low compared to *mex-5*, *pie-1*, and *dpy-30* expression, as measured by endogenous gene expression (Stoeckius et al., 2014) or visual inspection of transgene fluorescence.

If PATCs contribute to prevent silencing, then insertion into a PATC-rich chromatin environment might also promote germline expression? We analyzed the local PATC content near P*pie-1*:*gfp* insertions and found no positive association between high PATC-content and germline expression (Figure 3D). We also observed no association between somatic transgene expression (P*eft-3*:*tdTomato*) and local PATC environment (Figure S3J), suggesting that insertion into A_n/T_n clusters does not in itself protect from silencing.

These data show that at least one PATC-rich transgene (*smu-1*) is remarkably resistant to germline silencing and suggest that, if PATCs permit germline expression, only do so when they are part of the transgene itself.

PATCs in Introns of *gfp* Reduces Stochastic Germline Silencing

Are PATCs sufficient to safeguard transgenes from gene silencing? To test the effect of PATCs in a consistent context we used a ubiquitous promoter (P*eft-3*) and 3′ UTR (*tbb-2*) with few PATCs. We expressed a *gfp* with minimal piRNA homology that had been optimized for high expression (Figure 4A). Only the PATC content within introns was varied; in particular, the initial 68 base pairs of *gfp* were kept invariant to minimize possible effects on translation efficiency and all intronic splice junctions were identical between transgenes to minimize possible differences in silencing caused by spliceosome stalling (Dumesic et al., 2013). A standard *gfp* and optimized *gfp*s with short synthetic introns or introns from a neuronal gene *snt-1* were frequently silenced in the germline from the center of Chr. V (*oxTi365*, insertion at 25°C) (Figure 4A). In contrast, PATC-rich introns from *smu-1*, *smu-2*, or the *Caenorhabditis briggsae* ortholog of *smu-1* (*cbr-smu-1*) significantly reduced stochastic transgene silencing. In the repressive chromatin environment on the arm of Chr. V (*oxTi173*), we observed a similar pattern of stochastic silencing, except that all transgenes were expressed at lower frequency (Figure 4A). Similarly, silencing of a *gfp:cdk-1* transgene was reduced when fused to an optimized GFP containing PATCs (Figure S4C).

Are the intronic PATCs responsible for reduced germline silencing, or do the introns harbor other signals that increase expression, for example, germline-specific enhancers? First, no single *smu-2* intron increased the frequency of germline expression (Figure S4A). Second, a *gfp* with *smu-1* introns inserted with a minimal promoter (*pes-10*) did not result in visible germline or somatic expression (data not shown). These data argue against the presence of strong enhancers in the introns. Third, we generated synthetic introns with PATCs by gene synthesis. P*eft-3*:*gfp* transgenes with synthetic PATCs showed partial but significant resistance to germline silencing in both permissive and repressive chromatin domains (Figure 4A). Fourth, increasing the number of synthetic PATC introns reduced stochastic gene silencing in repressive environments, suggesting an additive effect of PATCs (Figure S4B), although one of our synthetic introns consistently decreased expression. Fifth, the ability of PATC-rich *smu-1* introns or synthetic introns to prevent germline silencing was lost when we shuffled the intron sequences to eliminate the A-T clusters but maintained overall nucleotide composition (Figure 4B). This indicates that the base pair composition or specific length of these introns does not in itself improve germline expression.

By contrast, we were unable to confidently demonstrate that PATCs reduce somatic transgene variegation. We inserted P*eft-3*:tdTomato transgenes with no introns (cDNA),

insertions into *oxTi365* and *oxTi173*, respectively. The standard *gfp* (dark green) contains three synthetic introns and was not codon-optimized for *C. elegans* (Fire et al., 1998). All other *gfp*s (light green) were *C. elegans* codon-optimized and homologies with less than four mismatches to piRNAs were removed. All codon-optimized *gfp* transgenes were identical except for the four introns. "Synthetic introns" and "synthetic introns w PATCs" were generated by gene synthesis. Statistics: Kruskal-Wallis test ANOVA comparing all insertions at a given genomic location. Dunn's multiple comparison t test against *gfp* with *smu-1* introns (*$p < 0.05$, **$p < 0.01$).

(B) Germline expression of transgenes with semi-randomly shuffled intron sequences (shuffling done with rules to prevent novel consensus splice site motifs while maintaining base pair composition of introns). Statistics: Mann-Whitney rank test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$; N.S., not significant).

See also Figure S4.

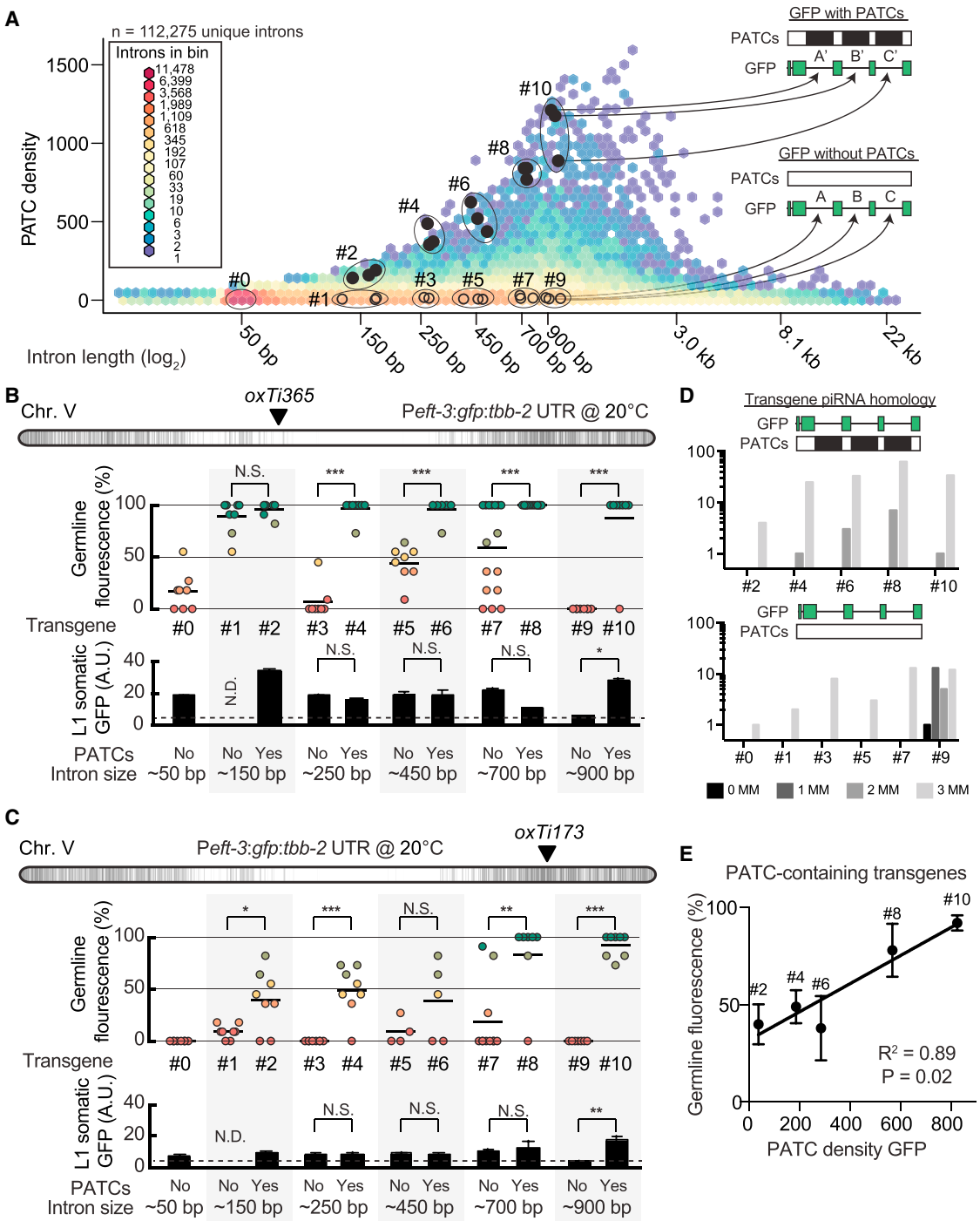


Figure 5. PATC Algorithm Can Identify Introns that Reduce Germline Silencing

(A) Hexagon plot showing the PATC density of all predicted, unique introns extracted from protein-coding genes (WS245). Introns tested in (B) are indicated by closed black circles (high PATC density) and open circles (low PATC density). Each individual intron with high PATC content (top) was matched to an intron with low PATC content (bottom) based on two parameters: (1) intron length, and (2) germline expression of the parent gene (both within 10%).

(B) Germline (top) and somatic (bottom) expression at 20°C from single-copy P*eft-3:gfp* transgene insertions at a central, permissive chromosome location (*oxTi365*). Statistics: Mann-Whitney test (*p < 0.05, **p < 0.01, ***p < 0.005; N.S., not significant; N.D., no data).

(C) Germline (top) and somatic (bottom) expression at 20°C from single-copy P*eft-3:gfp* transgene insertions at a distal, repressive chromosome location (*oxTi173*). Statistics: Mann-Whitney test (*p < 0.05, **p < 0.01, ***p < 0.005; N.S., not significant; N.D., no data).

(legend continued on next page)

a codon-optimized tdTomato with short synthetic introns or introns from *smu-1* with PATCs into central (*oxTi365*) and distal (*oxTi173*) locations. Somatic expression quantified by visual classification, with flow cytometry, and with automated identification in L1 animals showed at most a very modest increase in expression from transgenes with PATCs and only at later larval stages (Figures S4D and S4E).

In sum, native and synthetic PATC-rich introns placed in a foreign coding region can reduce position-dependent silencing in the germline.

PATC-Rich Introns across a Range of Lengths and from Many Genes Reduce Germline Silencing

Introns from *smu-1* and *smu-2* could efficiently reduce transgene silencing but were derived from genes that are unusually resistant to germline silencing (Spartz et al., 2004; Spike et al., 2001). Furthermore, we maintained animals at 25°C, a temperature that empirically promotes germline expression of transgenes (Strome et al., 2001) but also reduces fecundity and is above the thermal tolerance of some *C. elegans* isolates (e.g., the Bergerac isolate) (Hirsh et al., 1976). To more fully characterize PATC introns, we investigated the role of intron length, intron diversity, and temperature on germline expression with five pairs of GFP with or without PATCs in introns. To select introns, we analyzed the PATC density of all protein-coding introns individually (112,275 introns, WS245): introns have a median length of 69 bps and a prominent peak of PATC-rich introns near ~900 bp (Figure 5A). We selected 30 introns spanning lengths from ~150 bp to ~900 bp that were derived from 29 different endogenous genes and inserted three introns into each P*eft-3*:gfp transgene. Each individual PATC-containing intron (solid black circles) was matched to an intron with no PATCs (open circles); the introns were matched to within 10% based on intron length and germline expression of the endogenous genes containing the two introns (Table S3).

At the central insertion site (*oxTi365*), we observed variable germline expression at 20°C from transgenes with no PATCs (Figure 5B). In contrast, all five transgenes with PATCs were expressed at high frequencies (Figure 5B). The effects of PATCs on somatic transgene expression were mixed; transgenes with PATCs were generally well expressed but only one PATC-rich transgene had significantly higher expression than a poorly expressed matched control (Figure 5B). In repressive chromatin (*oxTi173*), we observed a strong and consistent effect of PATCs: transgenes containing PATCs were less frequently silenced in the germline (four of five matched transgenes), and there was no pervasive enhancement of somatic expression (Figure 5C). These differences did not generally appear to be caused by differences in piRNA homology, with one possible exception (900 bp, non-PATC) (Figure 5D). We note that only transgenes with the longest introns (700–900 bps) were expressed at high frequency in repressive chromatin and we observed a good association (R^2 = 0.89) between PATC content and resistance to germline silencing (Figure 5E). It is possible that longer introns or higher PATC densities are required to efficiently prevent germline silencing within highly repressive chromatin.

In sum, these data demonstrate a consistent ability for PATC-rich introns to reduce germline silencing and small effects, if any, on somatic transgene expression.

mRNA Is Depleted and Small Antisense RNAs Are Enriched in Strains with Silenced Transgenes

To determine at what stage transgenes were silenced we performed single molecule RNA fluorescence in situ hybridization (smFISH) (Raj et al., 2008) and RNA sequencing experiments on active and silenced P*eft-3*:*gfp* insertions. We observed many individual *gfp* transcripts (diffraction limited cytoplasmic spots; Figure 6A) and frequent transcriptional foci (brighter nuclear spots; Figure S5) by smFISH against *gfp* in germlines with GFP expression (PD1538). In contrast, we detected few transcripts and no transcriptional foci in GFP-negative germlines from wild-type animals (negative control) or from fully silenced strains (e.g., PD1540). The same was true for a transgene with PATCs that was fully silenced (PD1539) or in GFP-negative animals from a transgene that was infrequently silenced (PD1537) (Figures 6A and S5).

To examine silencing on a bulk level, we sequenced total RNA isolated from synchronized young adult hermaphrodites and observed a depletion of *gfp* mRNA in animals with silenced P*eft-3*:*gfp* transgenes (Figure 6B). A strain with frequent GFP expression in the germline (PD1537) had ~10-fold more transcripts than animals with a fully silenced PATC-rich *gfp* (PD1539) or a fully silenced *gfp* with no PATCs (PD1540). We did not capture unspliced *gfp* pre-mRNAs sequences in RNA samples from strains with active or silenced transgenes. In combination with the lack of detectable transcripts in the nucleus by smFISH, we found no evidence for accumulation of unspliced transcripts in the germ cells of animals with silenced GFPs.

To further understand the mechanisms involved in transgene silencing, we isolated and sequenced populations of small RNAs from synchronized young adult hermaphrodites (Figure 6C). Animals carrying silenced GFPs with PATCs (PD1539) or lacking PATCs (PD1540) were 10- to 20-fold-enriched for detectable small antisense RNAs against GFP (primarily 21G–23G RNAs) compared to a strain with frequent GFP expression (PD1537). Fully silenced *gfp* transgenes (i.e., full stochastic silencing from the time of insertion) with or without PATCs were indistinguishable based on the level of *gfp* mRNA and small anti-sense RNAs (PD1539 versus PD1540) (Figures 6B and 6C).

Thus, PATCs within introns of a foreign gene confer significant but incomplete protection from stochastic silencing, in agreement with visible germline fluorescence (Figures 4, 5, and 6). It is possible these observations reflect a lack of PATCs in *Peft-3* and the *tbb-2* 3′ UTR; perhaps, full protection from stochastic silencing also requires PATCs in regions flanking the coding sequence, as observed for the *smu-1* transgene (Figure 3A).

(D) piRNA homology with zero (0MM), one (1MM), two (2MM), or three (3MM) mismatches to the sequence of each numbered *gfp*.

(E) Linear correlation between germline expression and PATC content for PATC-rich transgenes inserted at the repressive *oxTi173* location. The p value indicates the statistical significance for a positive slope of the linear fit.

See also Table S3.

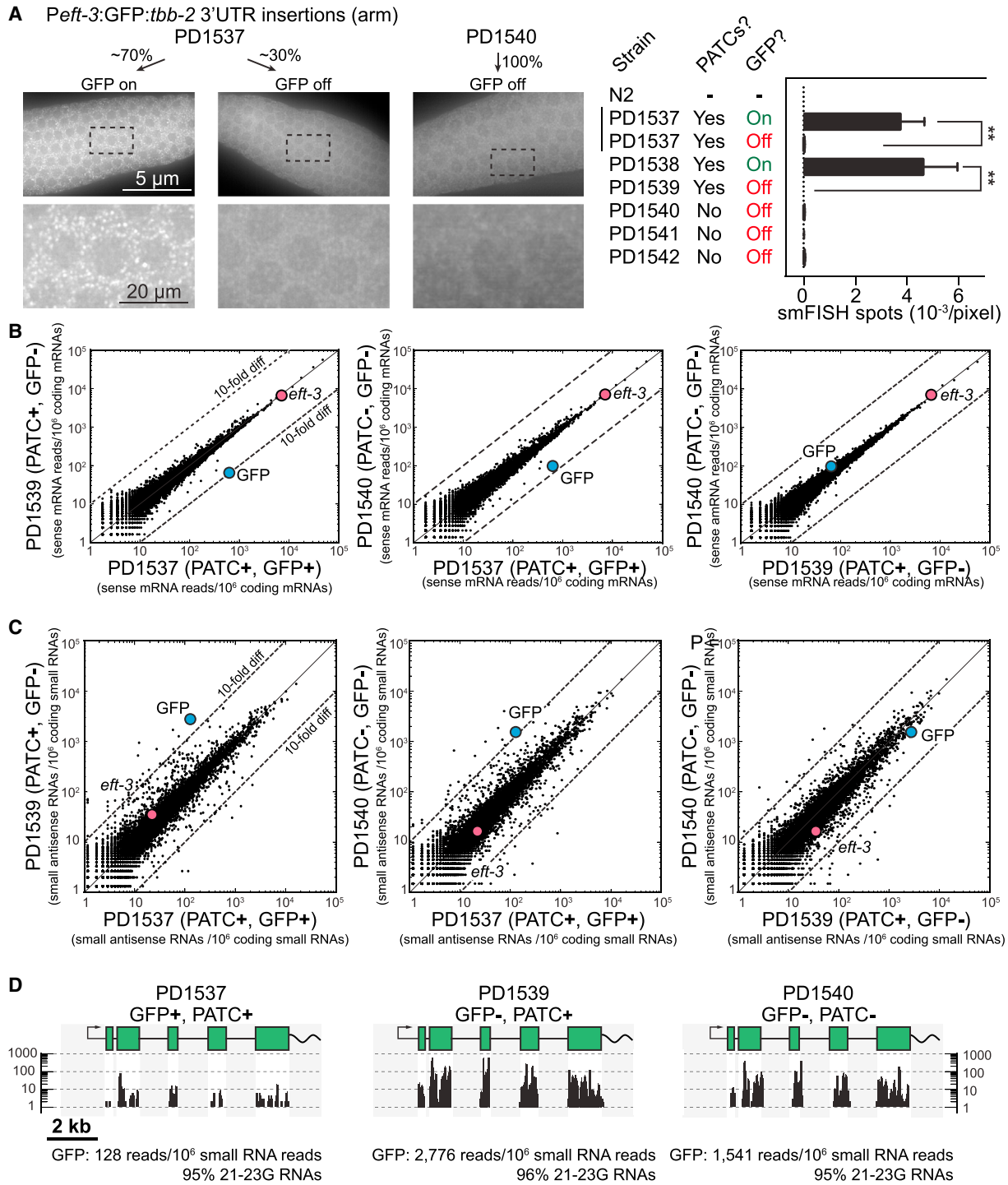


Figure 6. Transgene mRNA Expression and Small RNA Populations

(A) Left: representative images from single molecule fluorescence in situ hybridization (smFISH) on the germline from animals with frequent GFP expression (~70% fluorescent germlines, PD1537) and a fully silenced strain (PD1540) from P*eft-3*:*gfp* transgenes. Individual white spots indicate diffraction-limited single mRNA transcripts (Raj et al., 2008) and germline nuclei are visible as darker circles. Right: quantification (blinded) of the number of smFISH spots (mean ± SEM) from transgenic animals and N2 animals. In many cases, germlines with silenced *gfp* showed easily distinguishable GFP expression in somatic cells and smFISH spots in those tissues. Statistics: ANOVA, post-test Sidak's multiple comparison test (**p < 0.01).

(B) Sense mRNA expression of young adult animals with frequent germline expression (PD1537) or complete germline silencing (PD1539, PD1540) of P*eft-3*:*gfp* transgenes. Statistical tests: 2-proportion *Z* test (two-tailed), number of *gfp* transcripts in PD1537 versus PD1539, p ≅ 0; PD1537 versus PD1540, p ≅ 0. RNAs

(legend continued on next page)

In sum, these data suggest that transgenes are most likely transcriptionally silenced with the abundance of small RNAs potentially indicative of an RNAi-like mechanism maintaining, or potentially initiating, the silenced state via secondary 22G small interfering RNAs (siRNAs).

Evolutionary Adjustment of PATC Content for Different Genomic Environments

Genomes are under selective pressure and we expect functionally important sequence characteristics to be evolutionarily conserved in closely related species. PATCs are well conserved in the *Caenorhabditis* genus (5%–11% of the total genome sequences are PATC-rich), whereas conservation is mixed in more distantly related nematodes (Figure 7A; Table S4). Outside of nematodes, most genomes do not contain comparable frequencies of PATCs, although some distantly related organisms have PATC-like structures in their genomes (e.g., the centipede *Strigamia maritima*) (Figure 7B; Table S4).

If PATCs are functionally important, we might also predict the PATC content of genes to change in response to changes in chromatin environment, for example those caused by large-scale genome rearrangements. To test this prediction, we compared *C. elegans* and *C. briggsae*, whose most recent common ancestor existed ~20 million years ago (Ross et al., 2011). The overall genomic PATC-content is similar in the two species (Figure 7A) and PATCs from *cbr-smu-1* were able to safeguard transgenes in *C. elegans* (Figure 4A) suggesting functional conservation. Based on recombination frequencies (Ross et al., 2011) a chromosome structure with distinct center and arm domains is conserved in *C. briggsae,* and PATCs are similarly enriched on autosomal arms (Figure S6A). We analyzed unique *C. elegans* and *C. briggsae* ortholog pairs and determined their PATC content as a function of genomic location and expression in the germline (Figure 7C). Orthologs pairs that remain on arms have longer introns (Figure 7D) and higher PATC frequency (Figure 7E) compared to ortholog pairs that remained at a central location over evolution. Ortholog pairs that change chromatin domain show reciprocal changes: the ortholog residing in repressive chromatin has ~3-fold longer introns (Figure 7D) and ~4-fold higher PATC content (Figure 7E) than the ortholog residing at a central domain. Examples of these large changes in intron size and PATCs as a function of genome position are illustrated for three ortholog pairs in Figures 7F, 7G, S6B, and S6C.

To investigate how PATCs may be generated and/or maintained, we analyzed a large set of single nucleotide variants (SNVs) identified in chemically mutagenized and natural variant strains of *C. elegans* (Thompson et al., 2013) (Figure S7). For PATC regions, the underlying periodicity can be used to associate each SNV with a phase (0–10 using A_n/T_n cluster starts as defining phase 0) (Fire et al., 2006). Normalized mutation frequencies that result in a net loss of G/C content (G → A and C → T transitions and G → T and C → A transversions) within PATCs are most frequent where the G/C content is lowest. In contrast, mutations that increase G/C content are enriched in areas with already high G/C content. These two mutational profiles are offset by ~5 bps, yielding a tendency to strengthen the density of PATCs within regions already contained in A_n/T_n-rich clusters (Figure S7). These results are consistent with models (e.g., Holmquist, 1994) in which large-scale genomic domains in nematodes are relatively fixed and genes adapt under evolutionary pressure or due to net mutagenic bias to the chromatin domains they are embedded in.

In sum, changes in non-coding DNA reveal a surprisingly dynamic character (on an evolutionary time-scale) with PATCs generated or eliminated and concomitant intron expansion or contraction for genes expressed in the germline.

DISCUSSION

In many cases, epigenetic control of gene expression is initiated by small RNAs and maintained by proteins bound to DNA. Here, we show that distributed sequences of non-coding DNA can safeguard genes from epigenetic gene silencing. Specifically, these DNA structures are comprised of clusters of A/T sequences arranged so that they are on a single face of the DNA molecule. These periodic A_n/T_n clusters (PATCs) are found within introns or in intergenic regions and can promote germline expression of transgenes in repressive environments. We propose that PATCs similarly promote expression of endogenous genes from repressive chromatin environments.

Genome Domains and Position Effect Variegation in *C. elegans*

Using a synthetic transposon, we probed the *C. elegans* genome for chromatin environments affecting transgene expression. The chromatin domains for transgene expression described here are consistent with the large-scale structural stratification observed for other genome features:

(1) *C. elegans* autosomes are partitioned into broad central and distal regions. Here, we present functional evidence that these broad genomic domains influence gene expression. Transgenes inserted into distal autosomal regions are frequently silenced in the germline and subject to position effect variegation in somatic cells. Consistent with alternating regions of repressive and permissive chromatin marks (e.g., Gu and Fire, 2010; Liu et al., 2011) transgene silencing on arms is not uniform but rather a heterogeneous mix of active and silenced insertions.

were aligned against all protein coding genes (WS245) and normalized to uniquely aligned RNAs. Colored dots indicate *gfp* and *eft-3* expression. All strains had GFP expression in somatic cells.

(C) Small antisense RNA expression in young adult animals with germline expressed (PD1537) or silenced (PD1539, PD1540) P*eft-3*:*gfp* transgenes. Statistics: 2-proportion *Z* test (two-tailed), number of *gfp* transcripts in PD1537 versus PD1539, $p \cong 0$; PD1537 versus PD1540, $p \cong 0$. RNAs were aligned against all protein coding genes (WS245) and normalized to uniquely aligned RNAs. Colored dots indicate *gfp* and *eft-3* expression. All strains had GFP expression in somatic cells.

(D) Unique alignments of small RNAs (>95% 21G–23G RNAs) detected against *gfp*. The native introns in *gfp* (250 bp), the *eft-3* promoter, and *tbb-2* 3′UTR are also present in the *C. elegans* genome and therefore reads could not be aligned uniquely to these regions (shaded gray).

See also Figure S5.

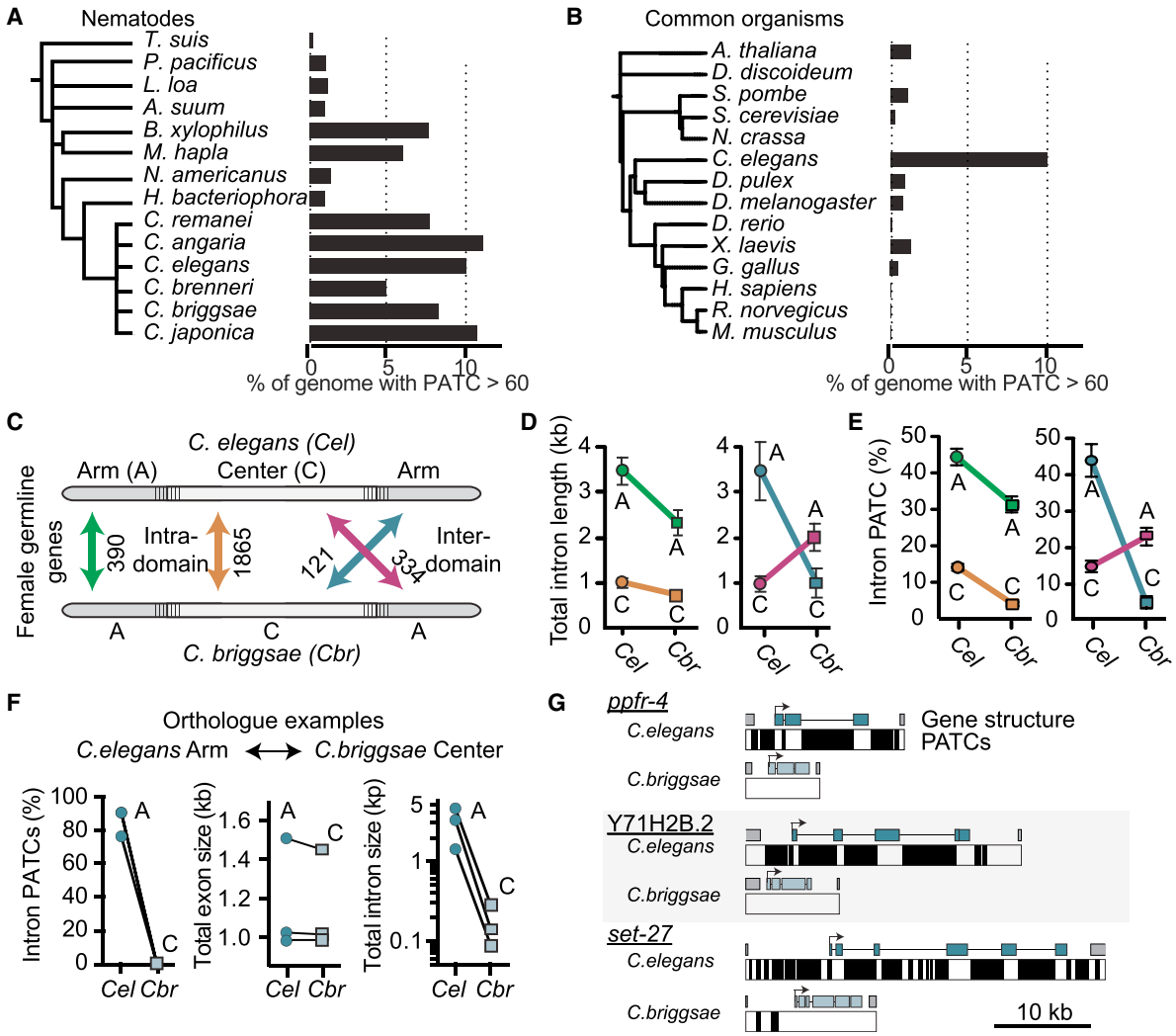


Figure 7. *C. elegans* and *C. briggsae* Intron Size and PATC Content Contract and Expand with Chromosomal Location

(A) PATC content of select nematode genomes. Frequency of DNA with a PATC value >60.

(B) PATC signal in commonly studied genetic model organisms and the human genome sequence. Frequency of DNA with a PATC value >60.

(C) Schematic of the unique orthologs analyzed. One class ("intra-domain") contains orthologs that reside on the arm (green) or the center (brown) of both *C. elegans* and *C. briggsae*. A second class contains "inter-domain" orthologs (blue, red) that have moved between chromatin domains in the two species.

(D) Comparison of the total intron length of genes for each ortholog class.

(E) Comparison of the percentage of intronic bases with an average PATC >60 for each ortholog class.

(F) An example of the PATC content, exon length, and intron size for three genes residing in a distal repressive chromatin domain in *C. elegans* and a central permissive domain in *C. briggsae* on Chr. III.

(G) Gene structure and PATCs >60 of the three genes from (F). The DNA sequence from the last exon of the upstream gene ("5′ gene") to the first intron of the downstream gene ("3′ gene") are included.

See also Figures S6, S7, Table S4, and Data S2.

(2) The *C. elegans* X chromosome is largely inactivated in early meiotic stages of the gonad (Fong et al., 2002; Kelly et al., 2002). We observed expression that closely mimicked this pattern: most X chromosome insertions were silenced in the early germline. There were two interesting exceptions. First, two P*pie-1* insertions near the left tip of the chromosome were expressed. The left tip of the X chromosome exhibits several features characteristic of distal regions of autosomes and is distinct from the remainder of the X chromosome (Fire et al., 2006; Fong et al., 2002; Kelly et al., 2002). Second, P*smu-1* insertions anywhere on the X chromosome were initially silenced but de-silenced over a few generations. Genes on the X chromosome are largely devoid of PATCs, which may reflect the unique constraints on X-linked genes caused by X chromosome inactivation in the germline and somatic sex dosage compensation (Meyer, 2010). These data suggest that the *smu-1*:GFP transgene, possibly due to

PATCs, can overcome chromosome-scale Polycomb-mediated epigenetic silencing.

A Role for an Abundant Class of Non-coding DNA in Nematodes

An analysis of the large RNA polymerase II gene (*ama-1*) identified periodic sequence features (A-tracts), which were predicted to result in significant DNA bending and explain the unusual migration of *C. elegans* genomic DNA on electrophoretic gels (VanWye et al., 1991). Subsequent analysis on the completed *C. elegans* genome sequence identified a pervasive periodic 10 bp motif of A_n/T_n clusters that were associated with germline expressed genes (Fire et al., 2006). Here, we present experimental support for how genes can acquire PATCs in repressive chromatin environments through biased mutations and evidence supporting a causal role for PATCs in permitting germline expression. Transgenes containing promoters with PATCs (P*pie-1*:GFP and P*mex-5*:GFP) were less prone to silencing than a transgene with few PATCs in the promoter (P*dpy-30*:GFP). PATC distributed throughout the entire construct (*smu-1*:GFP) or only in the coding region (various GFPs) significantly reduced germline silencing compared to transgenes lacking PATCs. These effects were consistent across different PATCs: *C. elegans* and *C. briggsae* introns, synthetic introns, short and long introns, and from a multitude of different genes. Our data support a generally permissive role for PATCs in allowing germline expression in contrast to an instructive signal that directly drives germline expression; we imagine that distributed PATCs within a gene allow "proper" DNA access for transcriptional regulation (enhancers, promoters, transcriptional elongation, splicing, etc.). Importantly, transgenes with many internal PATCs are not immune from stochastic silencing; once silenced, we could not distinguish a PATC-rich transgene from a PATC-poor transgene based on fluorescence imaging, smFISH and RNA sequencing.

PATCs May Protect Endogenous Genes from Silencing

Why are PATCs necessary? We have studied the role of PATCs in the context of transgenes but we propose that their natural role is to protect endogenous genes from otherwise non-discriminating silencing mechanisms. There are several described mechanisms by which transgenes are silenced in *C. elegans*: histone methylation, RNAi, RNA epigenetic (RNAe), and unpaired chromosomes in meiosis (Kelly et al., 2002; Ketting et al., 1999; Leopold et al., 2015; Shirayama et al., 2012; Tabara et al., 1999). PATCs may function in parallel to protective pathways that are proposed to depend on a balance between silencing piRNAs bound to the PRG-1 Argonaute and activating small RNAs bound to the Argonaute protein CSR-1 (Shirayama et al., 2012). A balance between permissive and repressive pathways does not imply full activation or complete silencing; for example, expression of a P*mex-5*:GFP transgene was increased 5-fold in *prg-1* mutants (Leopold et al., 2015). More generally, a class of endogenous genes with complementarity to 22G small RNAs were derepressed in *prg-1* mutants, suggesting that piRNAs do not exclusively repress foreign DNA (Lee et al., 2012). Unless mechanisms to "reboot" expression exist, antagonistic repressive and permissive pathways that monitor expression in prior generations would be prone to enter a negative feedback loop, which would ultimately lead to a silenced state. Endogenous structures that counteract silencing, such as PATCs, may constitute a fail-safe mechanism to prevent negative feedback loops from forming. PATCs may be particularly important for endogenous genes located in repressive genomic environments, such as the autosomal arms, where un-impeded spreading of repressive histone marks from adjacent regions may bias the balance toward the repressive pathway. Because no homologs of canonical boundary elements or insulators have been identified in *C. elegans* (Heger et al., 2009), one possible role for distributed PATCs is to prevent heterochromatin spreading.

PATCs May Defend the Genome against Viral or Transposon DNA

What are the possible benefits to nematodes for evolving PATC structures? We propose that this unusual DNA structure may be an important component of a genomic immune system that protects the nematode from viral or transposon invasion. A cellular immune system must recognize self versus non-self. We suggest a model in which PATCs can protect endogenous genes from silencing by counteracting silencing pathways. For endogenous genes, the PATC signature can be incorporated into non-coding DNA with no effect on protein sequence. For such a defense system to be efficient, insertion into a PATC-rich chromatin domain should not confer anti-silencing properties; this requirement is consistent with the observed broad distribution of PATCs in endogenous genes and silencing of transgenes inserted into local PATC-rich domains. In the constant arms race between hosts and parasitic DNA elements, it would be costly for invasive DNA to evolve structures that mimic PATCs to avoid silencing, especially if PATCs are specific to nematodes. This protection of self may have allowed nematodes to evolve an aggressive defense system that includes large, repressive genome domains and a piRNA system that tolerates target mismatches. In this model, invasive foreign DNA lacking an elaborate structural feature of endogenous genes are silenced by default (even in central, less repressive regions of the genome) and uncontrolled propagation of invasive DNA is limited. Notably, such a genomic defense system would not require a cellular memory of prior exposure to any particular foreign DNA sequence or structure.

EXPERIMENTAL PROCEDURES

Transgene insertions

Random miniMos insertions were generated in *unc-119*(*ed3*) animals and targeted mosSCI insertions in animals with Mos1 elements at defined locations (Frøkjær-Jensen et al., 2008, 2014). See Table S1 for strains and insertion sites.

Imaging

Germline fluorescence was classified by visual inspection on a fluorescence dissection microscope blind to strain identity (Table S2). Automated imaging of somatic fluorescence was performed on confocal microscopy images of fixed L1 animals (Liu et al., 2009) or mixed stage profiling of live animals on a COPAS-profiler2 (Dupuy et al., 2007) (Table S2). smFISH was performed according to the manufacturer's protocol (Biosearch Technologies) on

dissected germlines, imaged at 100× magnification and deconvolved. mRNA and transcriptional foci were quantified blind to strain identity.

Molecular Biology

All plasmids were generated by standard techniques and annotated GenBank sequences are included in Data S1.

Total mRNA and Small RNA Sequencing

We isolated mRNA and small RNAs from animals grown at 25°C and sequenced libraries on a miSeq instrument (Illumina). We aligned reads to *C. elegans* protein coding sequences (WS245): NCBI Short Read Archive: SRP072711.

PATC Analysis

Individual genes and introns (Table S3) were analyzed with the original PATC algorithm (Fire et al., 2006) and whole genomes were analyzed with a modified algorithm that minimizes off-helical A_n/T_n signals (Table S4). See Data S2 for 25 bp resolution PATC signals for *C. elegans* (WS245) and *C. briggsae* (WS245) genomes and Data S3 for intron sequences incorporated in *gfp*.

ACCESSION NUMBERS

The accession number for the small RNA and total mRNA sequences reported in this paper is NCBI Short Read Archive: SRP072711.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, four tables, and three data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.05.072.

AUTHOR CONTRIBUTIONS

Conceptualization, C.F.J. and A.Z.F.; Software, L.H., M.W.D., and A.Z.F.; Formal Analysis, C.F.J. and A.Z.F.; Investigation, C.F.J., N.J., L.H., Y.L., D.Z., K.R., J.R.M.M., X.L., and D.D.; Writing – Original Draft, C.F.J.; Writing – Review & Editing, C.F.J., E.M.J., and A.Z.F.; Supervision, C.F.J., X.L., S.K.K., D.D., E.M.J., and A.Z.F.; Funding Acquisition, C.F.J., E.M.J., and A.Z.F.

ACKNOWLEDGMENTS

We thank J. Feldman, A. Villeneuve, S. Strome, A. Rechtsteiner, J. Lieb, K. Ikegami, C. Engert, S. Klemm, T. Machacek, V. Jantsch, C. Girard, G. Maro, D. Nix, the A.Z.F. and E.M.J. lab members, and K. Hoe for expert technical assistance. This work was supported by the Carlsberg Foundation (C.F.J.), Direktør Ib Henriksens Foundation (C.F.J.), a Stanford Graduate Fellowship (N.J.), NIH grants R01GM095817 (E.M.J) and R01GM37706 (A.Z.F.), and the Howard Hughes Medical Institute (E.M.J.).

Received: July 16, 2015
Revised: March 31, 2016
Accepted: May 18, 2016
Published: June 30, 2016

REFERENCES

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol. Cell *31*, 67–78.

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

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol. Cell *31*, 79–90.

Dumesic, P.A., Natarajan, P., Chen, C., Drinnenberg, I.A., Schiller, B.J., Thompson, J., Moresco, J.J., Yates, J.R., 3rd, Bartel, D.P., and Madhani, H.D. (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. Cell *152*, 957–968.

Dupuy, D., Bertin, N., Hidalgo, C.A., Venkatesan, K., Tu, D., Lee, D., Rosenberg, J., Svrzikapa, N., Blanc, A., Carnec, A., et al. (2007). Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. Nat. Biotechnol. *25*, 663–668.

Elgin, S.C.R., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb. Perspect. Biol. *5*, a017780.

Fire, A., Kelly, W.G., Hsu, M., Xu, S., Ahnn, J., Harfe, B.D., Kostas, S.A., and Hsieh, J. (1998). The uses of green fluorescent protein in *Caenorhabditis elegans*. In Green Fluorescent Protein: Properties, Applications, and Protocols, M. Chalfie and S. Kain, eds. (John Wiley and Sons), pp. 153–168.

Fire, A., Alcazar, R., and Tan, F. (2006). Unusual DNA structures associated with germline genetic activity in Caenorhabditis elegans. Genetics *173*, 1259–1273.

Fong, Y., Bender, L., Wang, W., and Strome, S. (2002). Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans. Science *296*, 2235–2238.

Frøkjær-Jensen, C., Davis, M.W., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.-P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat. Genet. *40*, 1375–1383.

Frøkjær-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovsky, A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat. Methods *11*, 529–534.

Garrigues, J.M., Sidoli, S., Garcia, B.A., and Strome, S. (2015). Defining heterochromatin in C. elegans through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. Genome Res. *25*, 76–88.

Gu, S.G., and Fire, A. (2010). Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma *119*, 73–87.

Heger, P., Marin, B., and Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Mol. Biol. *10*, 84.

Hirsh, D., Oppenheim, D., and Klass, M. (1976). Development of the reproductive system of Caenorhabditis elegans. Dev. Biol. *49*, 200–219.

Ho, J.W.K., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. Nature *512*, 449–452.

Holmquist, G.P. (1994). Chromatin self-organization by mutation bias. J. Mol. Evol. *39*, 436–438.

Ikegami, K., Egelhofer, T.A., Strome, S., and Lieb, J.D. (2010). Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome Biol. *11*, R120.

Johnson, C.L., and Spence, A.M. (2011). Epigenetic licensing of germline gene expression by maternal RNA in C. elegans. Science *333*, 1311–1314.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227–238.

Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.-H., Kim, S.K., Villeneuve, A.M., and Reinke, V. (2002). X-chromosome silencing in the germline of C. elegans. Development *129*, 479–492.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

Lee, H.-C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell *150*, 78–87.

Leopold, L.E., Heestand, B.N., Seong, S., Shtessel, L., and Ahmed, S. (2015). Lack of pairing during meiosis triggers multigenerational transgene silencing in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA *112*, E2667–E2676.

Liu, X., Long, F., Peng, H., Aerni, S.J., Jiang, M., Sánchez-Blanco, A., Murray, J.I., Preston, E., Mericle, B., Batzoglou, S., et al. (2009). Analysis of cell fate from single-cell gene expression profiles in C. elegans. Cell *139*, 623–633.

Liu, T., Rechtsteiner, A., Egelhofer, T.A., Vielle, A., Latorre, I., Cheung, M.-S., Ercan, S., Ikegami, K., Jensen, M., Kolasinska-Zwierz, P., et al. (2011). Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res. *21*, 227–236.

MacQueen, A.J., Phillips, C.M., Bhalla, N., Weiser, P., Villeneuve, A.M., and Dernburg, A.F. (2005). Chromosome sites play dual roles to establish homologous synapsis during meiosis in C. elegans. Cell *123*, 1037–1050.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656–668.

Meister, P., Towbin, B.D., Pike, B.L., Ponti, A., and Gasser, S.M. (2010). The spatial dynamics of tissue-specific promoters during C. elegans development. Genes Dev. *24*, 766–782.

Meyer, B.J. (2010). Targeting X chromosomes for repression. Curr. Opin. Genet. Dev. *20*, 179–189.

Ortiz, M.A., Noble, D., Sorokin, E.P., and Kimble, J. (2014). A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans. G3 (Bethesda) *4*, 1765–1772.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura, H., Lieb, J.D., and Strome, S. (2010). The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. PLoS Genet. *6*, e1001091.

Rockman, M.V., and Kruglyak, L. (2009). Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. *5*, e1000419.

Ross, J.A., Koboldt, D.C., Staisch, J.E., Chamberlin, H.M., Gupta, B.P., Miller, R.D., Baird, S.E., and Haag, E.S. (2011). Caenorhabditis briggsae recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. *7*, e1002174.

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65–77.

Spartz, A.K., Herman, R.K., and Shaw, J.E. (2004). SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. Mol. Cell. Biol. *24*, 6811–6823.

Spike, C.A., Shaw, J.E., and Herman, R.K. (2001). Analysis of smu-1, a gene that regulates the alternative splicing of unc-52 pre-mRNA in Caenorhabditis elegans. Mol. Cell. Biol. *21*, 4985–4995.

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., Herzog, M., Selbach, M., and Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. EMBO J. *33*, 1751–1766.

Strome, S., Powers, J., Dunn, M., Reese, K., Malone, C.J., White, J., Seydoux, G., and Saxton, W. (2001). Spindle dynamics and the role of γ-tubulin in early Caenorhabditis elegans embryos. Mol. Biol. Cell *12*, 1751–1764.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123–132.

Thompson, O., Edgley, M., Strasbourger, P., Flibotte, S., Ewing, B., Adair, R., Au, V., Chaudhry, I., Fernando, L., Hutter, H., et al. (2013). The million mutation project: a new approach to genetics in Caenorhabditis elegans. Genome Res. *23*, 1749–1762.

VanWye, J.D., Bronson, E.C., and Anderson, J.N. (1991). Species-specific patterns of DNA bending and sequence. Nucleic Acids Res. *19*, 5253–5261.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

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.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25–33.

Cell

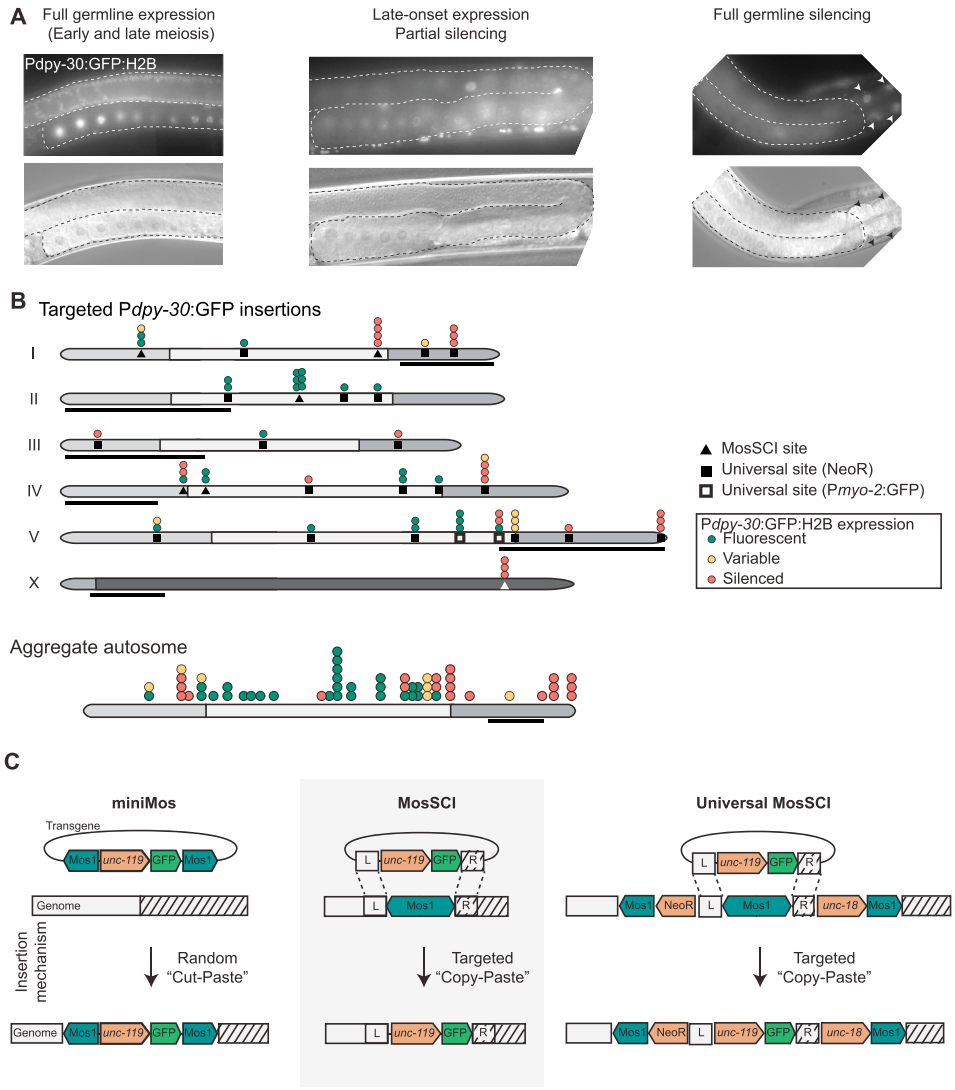


Figure S1. Classes of Germline Expression and Overview of Transgenic Methods, Related to Figure 1

(A) Examples of different fluorescence patterns observed from P*dpy-30*:GFP:H2B:*tbb-2* UTR transgene insertions. GFP fluorescence (top) and Differential Interference Contrast (DIC) transmitted light images (bottom) imaged with a 42x oil immersion objective are shown. The outline of the germline is indicated with dotted lines. Left. Full germline expression in early and late germline. Middle. Partial silencing: Visible germline expression is observed starting at the pachytene to diplotene transition (just before the "gonad bend"), which coincides with a described transition from the most intense H3K9me3 antibody staining (pachytene) to a rapid remodeling and disappearance of H3K9me3 antibody binding (Kelly et al., 2002). Right. No visible germline expression. Somatic GFP expression in four intestinal cells is indicated with arrowheads (and broad GFP expression is visible in whole animals).

(B) Overview of P*dpy-30*:GFP:H2B insertions into 27 MosSCI sites or Universal MosSCI sites used for targeted insertion. Black lines indicate the chromosome pairing site (MacQueen et al., 2005). Multiple independent insertions were obtained at 16 sites; 7 transgene insertions showed "variable" germline expression, where some animals derived from a unique insertion had "full" germline expression and other siblings had "late-onset" or "no germline" expression. Expression from five MosSCI sites was previously characterized and included for completeness (Frøkjær-Jensen et al., 2014).

(C) Schematics of the three different methods used to insert transgenes into the genome. MosSCI and Universal MosSCI rely on a "copy-paste" mechanism, which is inherently error-prone (Frøkjær-Jensen et al., 2008). Mos1 transposition relies on a "cut-paste" mechanism that is less error-prone (Frøkjær-Jensen et al., 2014). miniMos. Mos1 transposition relies on transposition of cargo carried inside a transposon ("miniMos"). Mos1 transposition inserts randomly into TA dinucleotides in the genome. MosSCI. MosSCI insertions were generated by excision of a Mos1 transposon and repair from an injected transgene. The repair process inserts a single copy of the transgene (here "GFP") and selection marker (here "*cbr-unc-119*(+)") to a unique genomic location (Frøkjær-Jensen et al., 2008). Universal MosSCI. Universal MosSCI insertions were generated by the same principle but relied on a standard ("universal") set of landing sites. The landing sites contain an internal Mos1 element that can be remobilized and each can be targeted individually by a single transgene (Frøkjær-Jensen et al., 2014).

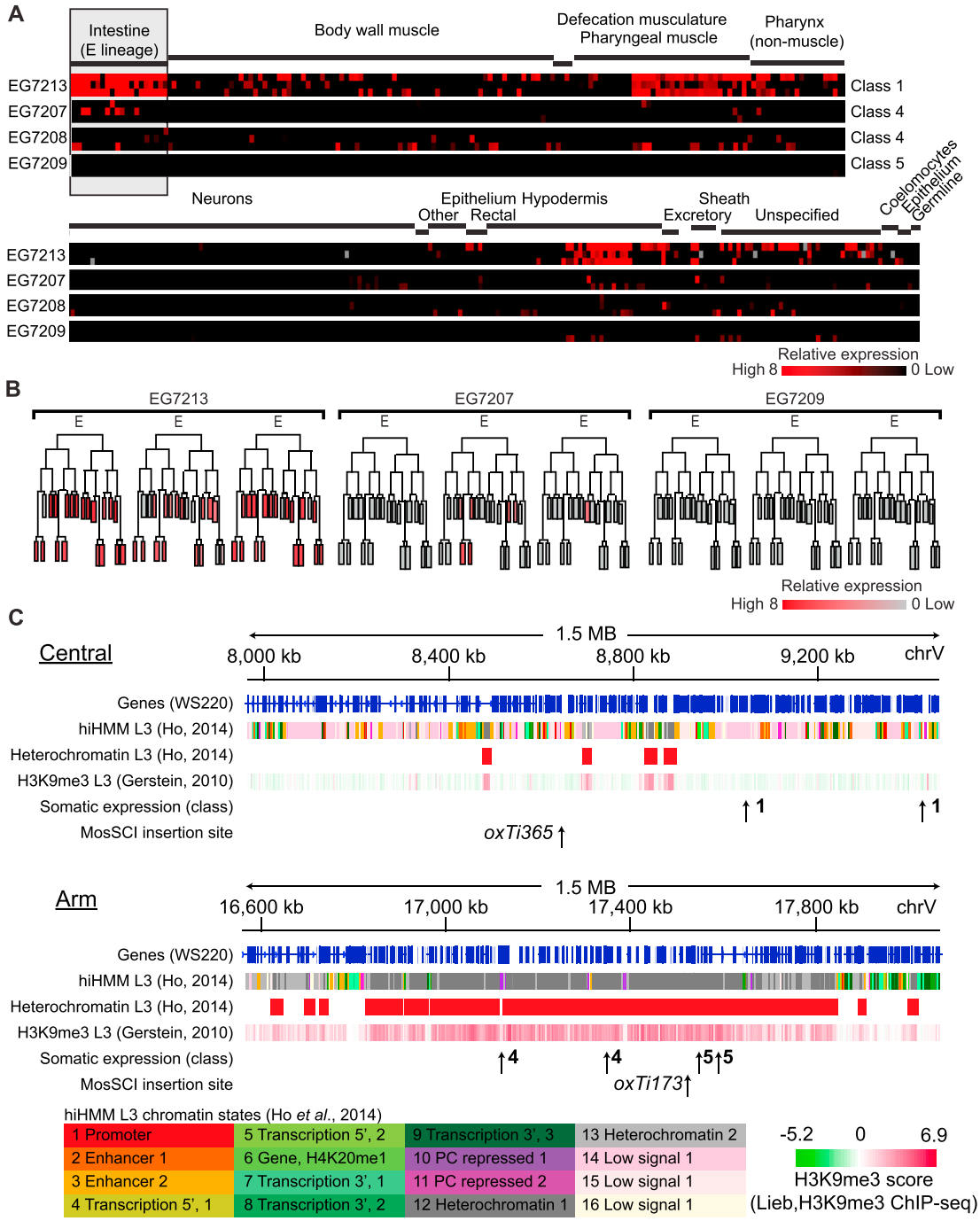


Figure S2. Automated Imaging of Somatic Fluorescence from Broadly Expressed tdTomato Transgenes, Related to Figure 2

(A) Automated quantification of P*eft-3*:*tdTomato*:*h2b*:*unc-54* 3′ UTR expression from confocal images of fixed larval stage 1 (L1) animals (Liu et al., 2009). The data from Figure 2 was repeated at higher magnification and one additional Class 4 strain (EG7208) was included. Single cell expression data is included in Table S2.

(B) Example of expression variegation in the E cell lineage of the intestine from the P*eft-3*:*tdTomato* transgene. The developmental time-course of the E cell lineage (Sulston et al., 1983) is included for illustration purposes; no time-course of tdTomato expression was performed.

(C) Example of the local hiHMM (Ho et al., 2014), heterochromatin domains (Ho et al., 2014), and H3K9me3 modifications (Gerstein et al., 2010) near a central (*oxTi365*) and distal (*oxTi173*) mosSCI insertion site. The somatic expression class of random P*eft-3*:tdTomato insertions is indicated by arrows and a number. Key indicates the hiHMM classes defined by Ho et al. (2014).

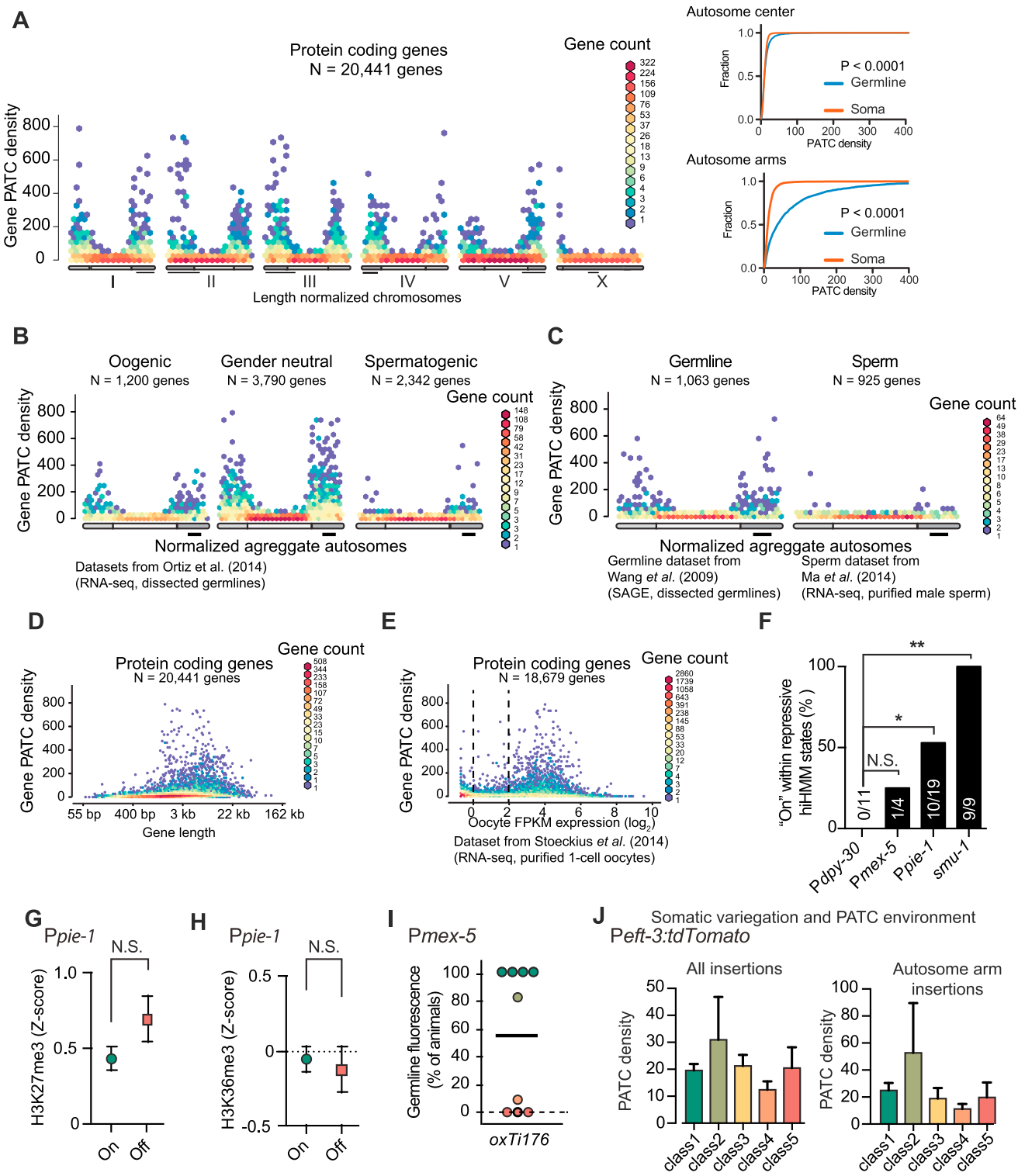


Figure S3. PATCs Are Primarily Enriched in Genes that Reside on Autosome Arms and Are Expressed in the Hermaphrodite Germline, Related to Figure 3

PATCs are composed of short clusters of adenines and thymines spaced approximately 10 basepairs apart and coat one face of the DNA helix with A_n/T_n clusters over extended runs. Previously, we had shown that PATCs are positively associated with germline expression (Fire et al., 2006) and here we confirm these observations with recent gene models (WS245) and new gene expression profiles primarily from RNA-seq datasets (Ma et al., 2014; Ortiz et al., 2014; Stoeckius

(legend continued on next page)

et al., 2014; Wang et al., 2009). We also note that that PATC-rich regions tend toward decreased interaction with autosome centers when assayed by Hi-C chromosome capture protocols (Gabdank et al., 2016).

(A) Left. Genic PATC density is plotted as a function of genome position. The PATC density was calculated over the full coding region of all predicted protein-coding genes (WS245); for genes with multiple isoforms the highest PATC density was included in the plot. Right. Cumulative distribution of PATC density for genes residing at autosome centers (central half) or autosome arms (outer 25% of each autosome) for genes not detected in the germline ("soma," RPKM < 1) or expressed in the germline ("germline," RPKM > 5). Statistical comparison with Kolmogorov-Smirnov test. PATC values for all genes (protein-coding and non-coding) are included in Table S3.

(B) Genic PATC density and autosome location for genes with observed reads showing statistically significant enrichment in oogenic (left), gender neutral (middle), or spermatogenic (right) tissue based on transcriptomes (RNA-seq) of germlines isolated from mutant animals with oocytes (*fog-2*) or sperm (*fem-3*) (Ortiz et al., 2014). For aggregating autosomes, lengths were normalized and each chromosome oriented with pairing center facing right.

(C) Genic PATC density of transcripts detected with 9 or more tags in dissected hermaphrodite germlines (left) by Serial Analysis of Gene Expression (SAGE) (Wang et al., 2009) and in purified male sperm for transcripts with RPKM > 0.159 by RNA sequencing (Ma et al., 2014).

(D) Genic PATC density of protein-coding genes as a function of transcript length.

(E) Genic PATC density of protein-coding genes as a function of transcript abundance detected in purified 1-cell oocytes (Stoeckius et al., 2014).

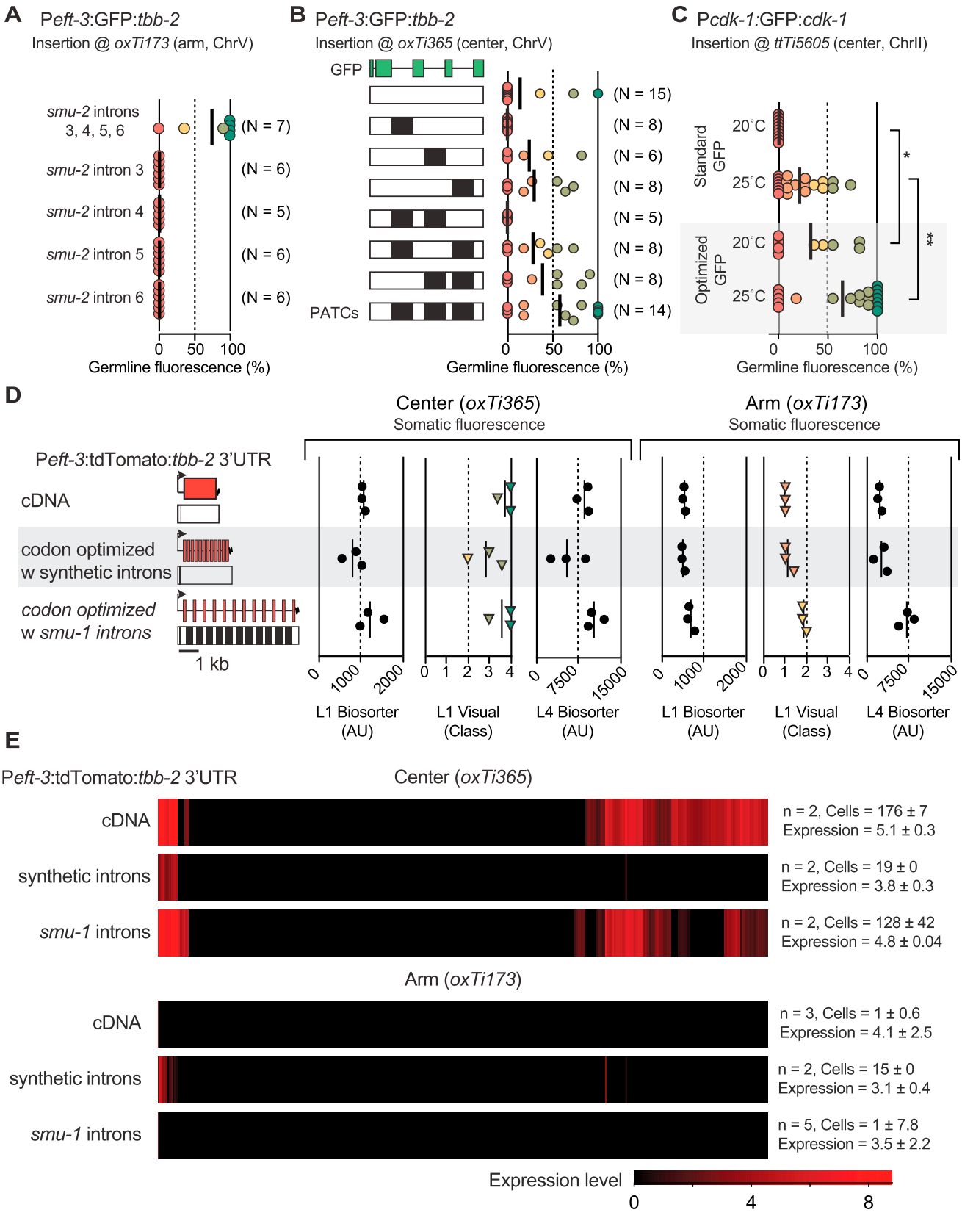
(F) Proportion of germline expressed inserts (indicated with green arrows in Figure 1B and Figure 3A) embedded in repressive chromatin states (Polycomb or heterochromatin, states 10-13) classified in early embryos by the modENCODE consortium (Ho et al., 2014). See Table S2. Statistics: Fischer's exact test (N.S. = no significant, *$p < 0.05$, **$p < 0.01$).

(G) Local (±1kb from insertion site) H3K27me3 environment near P*pie-1* insertions. Statistical test: Mann Whitney. N.S. = not significant.

(H) Local (±1kb from insertion site) H3K36me3 environment near P*pie-1* insertions. Statistical test: Mann Whitney. N.S. = not significant.

(I) Germline fluorescence of P*mex-5:gfp:h2b* transgenes into the mosSCI site *oxTi176* at the border between low and high levels of the repressive H3K9me3 mark on Chr. V.

(J) Local (±1 kb from insertion site) PATC density near P*eft-3*:TdTomato insertions.



(legend on next page)

Figure S4. Effect of PATCs on Germline and Somatic Transgene Expression, Related to Figure 4

(A) Single copies of P*eft-3*:*gfp*:*tbb-2* UTR transgenes with four or single *smu-2* introns (all single introns inserted at the same location in *gfp*) were inserted into the *oxTi173* landing site on the arm of Chr. V. Germline fluorescence was scored after propagation for two generations at 25°C. The *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).

(B) Single copies of P*eft-3*:*gfp*:*tbb-2* UTR transgenes with three intron positions containing combinations of introns with synthetic PATCs or the scrambled versions of the same introns (see PATC schematic under transgene). Transgenes were inserted into the *oxTi365* landing site at the center of Chr. V. Germline fluorescence was scored after propagation for two generations at 20°C.

(C) Quantification of germline expression of MosSCI insertions of *cdk-1* transgenes at *ttTi5605* on chromosome II. The *cdk-1* transgene with a standard *gfp* contains the same promoter, coding region, and 3′ UTR as a transgene that was particularly prone to silencing (Shirayama et al., 2012). The optimized *gfp* is codon optimized and contains introns from *smu-1*. The transgenes were inserted at 20°C or 25°C. All identified strains with insertions were propagated at 25°C for at least two generations prior to imaging. Statistics: Pairwise Mann-Whitney t test, *$p < 0.05$, **$p < 0.01$.

(D) Quantification of somatic expression from P*eft-3*:*tdTomato* transgenes with no introns (top), with short synthetic introns lacking PATCs (middle), or with longer introns containing *smu-1* introns with PATCs (bottom). The cDNA transgene is identical to the transgene randomly inserted by transposition in Figure 2. Independent single-copy transgene integrations were scored for somatic fluorescence across life stages with a COPAS worm sorter ("Biosorter," L1 & L4 on graph) and classified by eye on a dissection microscope ("visual score"). The tdTomato coding region contains an exact duplication ("tandem dimer") and both synthetic and *smu-1* introns were repeated within transgenes; thus all three transgenes are relatively repetitive.

(E) Automated cellular expression expression pattern and expression level in L1 animals using the CellExplorer method. The figures show average intensities across several animals; the number of independent animals, the number of cells with non-zero expression, and the average level of expression in cells with expression is indicated to the right. Primary data in Table S2.

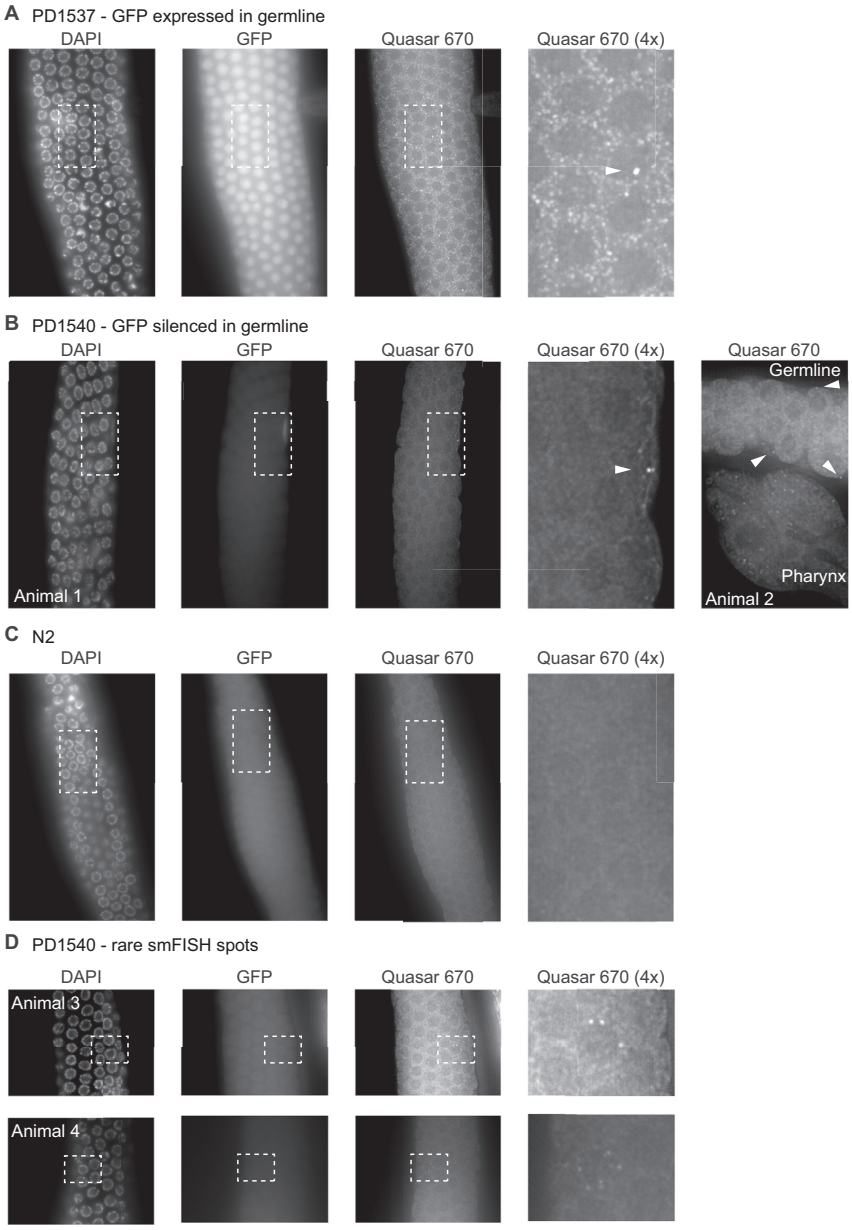


Figure S5. Supporting Figure for Single Molecule FISH and Small RNAs, Related to Figure 6

(A) smFISH image of a fixed germline from the strain PD1537 with GFP expressed in the germline. The DAPI stain shows all the nuclei in the germline, the GFP signal shows nuclear expression throughout the germline, and the Quasar 670 signal identifies widespread spots corresponding to individual *gfp* mRNAs. An example of a brighter smFISH spot localized to the nucleus is indicated with a white arrowhead; these spots are interpreted as transcriptional foci and evidence for active transcription (Raj et al., 2008).

(B) smFISH images of a fixed germlines from the strain PD1540 with a GFP that is silenced in the germline. smFISH spots are absent in the germline and only visible near the GFP positive cell of the somatic gonad (arrowheads, animal 1 and 2) or in muscle cells of the pharynx (animal 2).

(C) Fixation and images from N2 wild-type animals lacking GFP and smFISH signal.

(D) Examples of rare smFISH spots in PD1540 animals with silenced GFPs. In animal 3 and animal 4, a few cells have *gfp* mRNAs near individual germline nuclei; because GFP is also expressed in the somatic gonad, we cannot exclude that these spots identify expression in somatic sheath cells, which partially envelop germline nuclei in the syncytial gonad (Hall et al., 1999).

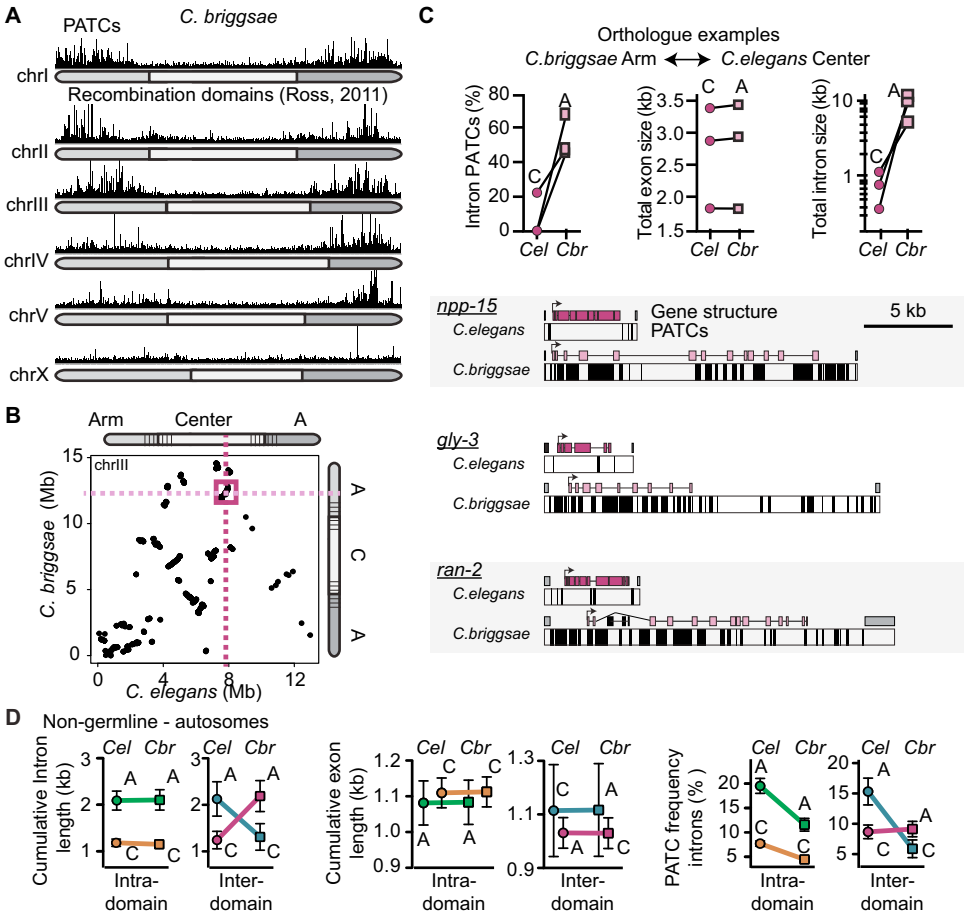


Figure S6. Phylogenetic PATC Conservation, Related to Figure 7

(A) Schematic overview of the spatial distribution of PATCs in the *C. briggsae* genome (CB4). The chromosomes are color coded corresponding to chromosome domains with low (light gray) and high (darker gray) recombination frequency (Ross et al., 2011).

(B) *C. elegans* and *C. briggsae* synteny plot of orthologs on chromosome III. The chromosome schematics show the genomic limits used to classify genes and boundary regions (hatched) that were omitted from the analysis. Box shows location of genes for (C).

(C) Examples of genes moving between the center of a *C. elegans* autosome and the arm of a *C. briggsae* autosome. Top. Intronic PATC content, cDNA size ("total exon size") and the accumulated size of introns (log scale). Bottom. Gene structure and PATC content of the three example.

(D) Comparison of orthologs for which gene expression is not detected in the female *C. elegans* germline (Stoeckius et al., 2014). *C. elegans* genes are indicated by a circle and *C. briggsae* genes with a square. Values represent the mean value and error bars indicate the 95% confidence interval.

The source data are included in Table S4.

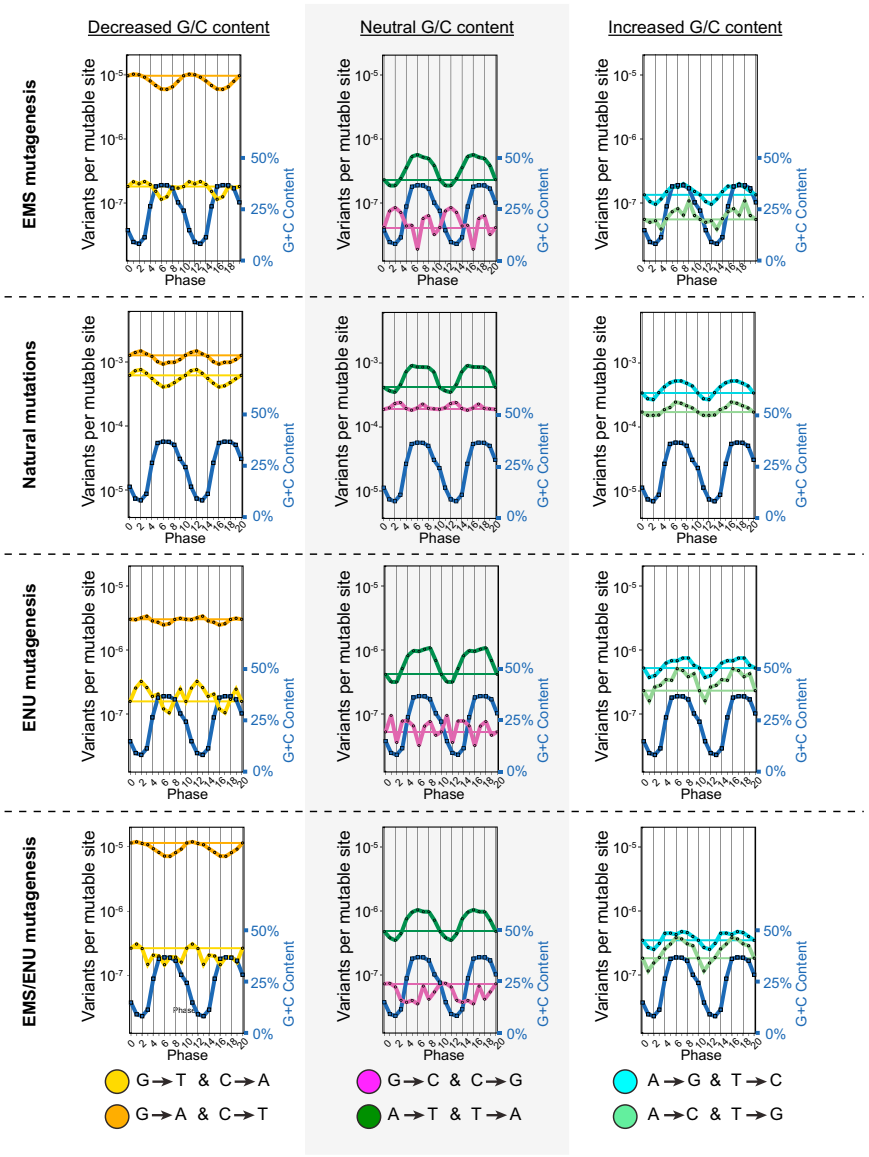


Figure S7. Chemical and Random Mutagenic Bias Strengthen PATCs, Related to Figure 7

To examine mutational spectra as a function of PATC phasing in *C. elegans*, we took advantage of an extensive project by several groups in which mutagenized and wild *C. elegans* isolates with natural variants were sequenced (Thompson et al., 2013). In particular, that work provides a curated list of observed mutations and their genomic positions for 2,007 mutagenized strains and 40 wild isolates. We identified mutations that fell within PATCs and associated each mutation with a phase (0-10). The phase is defined by the PATC algorithm (Fire et al., 2006) and corresponds to radial positions on the DNA helix, with maximal ~10bp periodic alignment of A_n and T_n starting points defined as phase 0. Mutagenized strains were classified into four groups based on the cause of mutations (chemical mutagens: EMS, ENU, or an EMS/ENU mixture, and natural sequence changes in wild *C. elegans* isolates). Mutations were classified based on changes relative to the base identity in the non-mutagenized canonical strain VC2010 (a defined derivative of the original Brenner N2 stock (Brenner, 1974) that was the starting point for the mutagenesis efforts of Thompson et al., 2013) and assigned to different classes based on the mutant base transition (e.g., C- > T or A- > T). For chemically mutagenized strains, a directionality of the transition could be unambiguously assigned because all strains were derived from VC2010. For wild strains, no such definitive assignment of parental and mutant base was possible (i.e., the direction of the transition could not be assigned). However, Thompson et al. (2013) assembled the list of mutations by removing commonly observed mutants within a category, which presumably removed many ancestral mutations, while identifying the majority of single nucleotide variants arising more recently in natural variant strains. For each position in the phase, we calculated the mutational frequency for each base individually (for example, how often does the base G at phase 0 make each of the transitions: G - > A, G - > T, or G - > C) for each mutagen. We normalized the observed transition frequency at each position in the phase to the total number of bases eligible for the transition at that position (in the previous example, we divided the three transition frequencies for G at phase 0 by the total number of Gs at phase 0 in the parent VC2010 strain). This allowed a precise comparison of transition frequencies as a function of phase, independent of the relative basepair composition at each position in the phase. For reference, the G+C composition in the wild-type reference strain VC2010 is shown for each position in the phase (identical in all panels), allowing inference of whether a class of mutation either supports or antagonizes the existing sequence composition preference at each position. Plots are shown here traversing two full (duplicate)

(legend continued on next page)

phases roughly corresponding to two full turns of the DNA helix (21 base pairs total). A thin horizontal line for each mutational class indicates frequencies of recovered mutations of that class among SNVs not in PATC regions. A tendency for (G/C) rich areas to get richer while poor areas get poorer would certainly contribute to the maintenance (and potentially establishment) of PATC character (Holmquist, 1994). It is possible that this bias is caused by repair that derives from distinct local environments caused by structural alignment of germline DNA on a surface (Gu et al., 2013; Thoma, 1999). The inability to observe the precise mutagenic conditions encountered during early evolution of the species preclude any simple determination of whether such mutagenic biases alone are responsible for the PATC character.

Cell, Volume *166*

Supplemental Information

An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the *C. elegans* Germline

Christian Frøkjær-Jensen, Nimit Jain, Loren Hansen, M. Wayne Davis, Yongbin Li, Di Zhao, Karine Rebora, Jonathan R.M. Millet, Xiao Liu, Stuart K. Kim, Denis Dupuy, Erik M. Jorgensen, and Andrew Z. Fire

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.

Table S1

Figure 1	Plasmid
Panel A	
miniMos	
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ641
Panel C	
mosSCI plasmids	
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ632
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ634
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ636
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ638
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ635
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ633
Figure 2	
Panel A	
miniMos	
Peft-3:tdTomato:H2B:unc-54 UTR	pCFJ453
Figure 3	
Panel A	
Minimos	
Pmex-5:GFP:H2B:tbb-2UTR	pCFJ1024
Ppie-1:GFP:H2B:tbb-2UTR (unc-119)	pCFJ747
Psmu-1:smu-1:gfp:smu-1UTR	pCFJ889
Panel E	
MosSCI	
Ppie-1:GFP:H2B:tbb-2UTR	pCFJ164
Figure 4	
Panel A	
MosSCI	
Peft-3:GFP:tbb-2UTR (standard)	pCFJ1504
Peft-3:GFP:tbb-2UTR (codon optimized, synthetic introns)	pCFJ1137
Peft-3:GFP:tbb-2UTR (codon optimized, snt-1 introns)	pCFJ1505
Peft-3:GFP:tbb-2UTR (codon optimized, smu-1 introns)	pCFJ1415
Peft-3:GFP:tbb-2UTR (codon optimized, smu-2 introns)	pCFJ1320
Peft-3:GFP:tbb-2UTR (codon optimized, synthetic introns w PATCs)	pCFJ1495
Peft-3:GFP:tbb-2UTR (codon optimized, cbr-smu-1 introns)	pCFJ1645
Panel B	
Peft-3:GFP:tbb-2UTR (codon optimized, smu-1 introns)	pCFJ1415
Peft-3:GFP:tbb-2UTR (codon optimized, shuffled smu-1 introns,)	pCFJ2414
Peft-3:GFP:tbb-2UTR (codon optimized, synthetic introns w PATCs)	pCFJ1495
Peft-3:GFP:tbb-2UTR (codon optimized, shuffled synthetic introns w PATCs)	pCFJ2415
Figure 5	
Panel A	
MosSCI	
#0 Peft-3 | ceGFP(1 intron, 3 syntrons, NLS(2)) | tbb-2 3UTR	pCFJ2406
#1 Peft-3 | ceGFP(150 bp, non-PATC, NLS) | tbb-2 3UTR	pCFJ2407
#2 Peft-3 | ceGFP(150 bp, wPATC, NLS) | tbb-2 3UTR	pCFJ2412

#3 Peft-3 | ceGFP(250 bp, non PATC, NLS) | tbb-2 3UTR	pCFJ2408
#4 Peft-3 | ceGFP(250 bp, wPATC, NLS) | tbb-2 3UTR	pCFJ2413
#5 Peft-3 | ceGFP(450 bp, non PATC, NLS) | tbb-2 3UTR	pCFJ2409
#6 Peft-3 | ceGFP(450bp introns PATCs, NLS) | tbb-2 3UTR	pCFJ2371
#7 Peft-3 | ceGFP(700 bp, non PATC, NLS) | tbb-2 3UTR	pCFJ2410
#8 Peft-3 | ceGFP(700bp introns PATCs, NLS) | tbb-2 3UTR	pCFJ2373
#9 Peft-3 | ceGFP(900 bp, non PATC, NLS) | tbb-2 3UTR	pCFJ2411
#10 Peft-3 | ceGFP(900bp introns PATCs, NLS) | tbb-2 3UTR	pCFJ2375

Figure S1

Panel A

MosSCI

Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ632
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ634
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ636
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ638
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ635
Pdpy-30:GFP:H2B:tbb-2UTR	pCFJ633

Figure S3

Panel I

MosSCI

Pmex-5:GFP:H2B:tbb-2UTR	pCFJ196

Figure S4

Panel A

MosSCI

Peft-3:GFP:tbb-2UTR (smu-2 introns, 4x)	pCFJ1320
Peft-3:GFP:tbb-2UTR (smu-2 intron 3)	pCFJ1597
Peft-3:GFP:tbb-2UTR (smu-2 intron 4)	pCFJ1602
Peft-3:GFP:tbb-2UTR (smu-2 intron 5)	pCFJ1599
Peft-3:GFP:tbb-2UTR (smu-2 intron 6)	pCFJ1601

Panel B

MosSCI

Peft-3:GFP:tbb-2UTR (sh - sh - sh)	pCFJ2427
Peft-3:GFP:tbb-2UTR (syn - sh - sh)	pCFJ2429
Peft-3:GFP:tbb-2UTR (sh - syn - sh)	pCFJ2430
Peft-3:GFP:tbb-2UTR (sh - sh - syn)	pCFJ2431
Peft-3:GFP:tbb-2UTR (syn - syn - sh)	pCFJ2432
Peft-3:GFP:tbb-2UTR (syn - sh - syn)	pCFJ2434
Peft-3:GFP:tbb-2UTR (sh - syn - syn)	pCFJ2433
Peft-3:GFP:tbb-2UTR (syn - syn - syn)	pCFJ2428

Panel C

MosSCI

Pcdk-1:GFP:cdk-1:cdk-1UTR (standard)	pCFJ1349
Pcdk-1:GFP:cdk-1:cdk-1UTR (codon optimized, five smu-1 introns)	pCFJ1350

Panel D

MosSCI

Peft-3:tdTomato(cDNA):tbb-2 3'UTR	pCFJ1102
Peft-3:tdTomato(synthetic introns):tbb-2 3'UTR	pCFJ1955
Peft-3:tdTomato(smu-1 introns):tbb-2 3'UTR	pCFJ1612

Table S1- universal MosSCI sites										
Transposon insertion	**Co-insertion marker**	**Chromosome**	**Genomic position (WS220)**	**Type of chromosome domain**	**Germline expression from Pdpy-30:GFP:H2B**	**Germline expression "class"**	**Insertion site and homozygosity validation post-inverse PCR**	**Injection strain**	**Genotype**	
Previously published universal mosSCI sites (Frøkjær-Jensen et al., 2014)										
oxTi185	unc-18(+) + NeoR	I	6503678	Central autosome	1 strain: Early germline all animals	On	Yes	EG7483	oxTi185[pCFJ687 unc-18(+)] I ; unc-119(ed3) III ; oxEx1850[cb-unc-119 (KAN) 1kb ladder]	
oxTi179	unc-18(+) + NeoR	II	9833484	Central autosome	1 strain: Early germline all animals	On	Yes	EG7791	oxTi179[pCFJ687 unc-18(+)] II ; unc-119(ed3) III ; oxEx1797[cb-unc-119(+) Kan]	
oxTi444	unc-18(+) + NeoR	III	7014337	Central autosome	1 strain: Early germline all animals	On	Yes	EG7794	oxTi444uni(unc-18) III unc-119(ed3) III ; oxEx1808[cb-unc-119(+)]	
oxTi177	unc-18(+) + NeoR	IV	13048923	Autosome are with no pairing center	1 strain: Early germline all animals	On	Yes	EG7797	unc-119(ed3) III ; oxTi177[pCFJ687 unc-18(+)] IV ; oxEx1781[unc-119(+)]	
oxTi365	unc-18(+) + NeoR	V	8643273	Central autosome	1 strain: Early germline all animals	On	Yes	EG7800	unc-119(ed3) III ; oxTi365 V ; oxEx1777[unc-119(+)]	
oxTi354	unc-18(+) + Pmyo-2:GFP	V	13783532	Central autosome	3 strains: 2 Strains early germline all animals, 1 strain 10/11 early, 1/11 non-fluorescent	On	Position not verified. Homozygosity validated by homozygous Pmyo-2:GFP transmission.	EG7801	oxTi354[Pmyo-2::GFP::H2B ttTi5605 unc-18(+)] ; oxEx1852[cb-unc-119 (KAN) 1kb ladder]	
New universal mosSCI sites										
oxTi369	unc-18(+) + NeoR	I	12924953	Autosome arm w pairing center	1 strain: 2/11 early, 9/11 non-fluorescent	Variable	Yes	EG7788	oxTi369 I ; unc-119(ed3) III ; oxEx1878[cb-unc-119(+)]	Non-unique insertion site (ChrI: 12924953 or ChrI: 12926593
oxTi363	unc-18(+) + NeoR	I	13941402	Autosome arm w pairing center	3 strains: All animals non-flourescent	Off	Yes	EG7789	oxTi363 I ; unc-119(ed3) III ; oxEx1872[cb-unc-119(+)]	
oxTi367	unc-18(+) + NeoR	II	5800506	Central autosome	2 strains: Early germline all animals	On	Yes	EG7790	oxTi367 II ; unc-119(ed3) III ; oxEx1780[unc-119(+)]	
oxTi174	unc-18(+) + NeoR	II	10980908	Central autosome	1 strain: Early germline all animals	On	Yes	EG7792	oxTi174[pCFJ687 unc-18(+)] II ; unc-119(ed3) III ; oxEx1796[cb-unc-119(+) Kan]	
oxTi172	unc-18(+) + NeoR	III	1276196	Autosome arm w pairing center	1 strain: Late germline all animals (X-like)	Off (X-like)	Yes	EG7793	xTi172[pCFJ687 unc-18(+)] III unc-119(ed3) III ; oxEx1794[cb-unc-119(+) Kan]	
oxTi180	unc-18(+) + NeoR	III	11670865	Autosome are with no pairing center	1 strain: Late germline all animals (X-like)	Off (X-like)	Yes	EG7482	oxTi180[pCFJ687 unc-18(+)] III unc-119(ed3) III ; oxEx1849[cb-unc-119 (KAN) 1kb ladder]	
oxTi188	unc-18(+) + NeoR	IV	8545894	Central autosome	1 strain: All animals non-flourescent.	Off	Yes	EG7795	unc-119(ed3) III ; oxTi188[pCFJ687 unc-18(+)] IV ; oxEx1802[cb-unc-119(+) Kan]	
oxTi376	unc-18(+) + NeoR	IV	11796364	Autosome arm w pairing center	2 strains: Early germline all animals	On	Yes	EG7796	unc-119(ed3) III ; oxTi376 IV ; oxEx1874[cb-unc-119(+)]	
oxTi381	unc-18(+) + NeoR	IV	14643762	Autosome are with no pairing center	4 strains. Strain 1: 1/11 early, 10/11 late germline. Strain 2: 2/11 early, 9/11 late germline. Strain 3+4: 11/11 late germline.	Variable/Off (X-like)	Yes	EG7798	unc-119(ed3) III ; oxTi381 IV ; oxEx1875[cb-unc-119(+)]	
oxTi374	unc-18(+) + NeoR	V	3339184	Autosome are with no pairing center	2 strains. Strain 1: early germline all animals. Strain 2: 3/11 early, 4/11 late, 4/11 non-fluorescent.	Variable	Yes	EG7799	unc-119(ed3) III ; oxTi374 V ; oxEx1873[cb-unc-119(+)]	
oxTi362	unc-18(+) + NeoR	V	12238606	Central autosome	2 strains: Early germline all animals	On	Yes		unc-119(ed3) III ; oxTi362 V ; oxEx1782[unc-119(+)]	
oxTi351	unc-18(+) + Pmyo-2:GFP	V	15240028	Central autosome (near transition)	3 strains. Strain 1: 11/11 early germline. Strain 2: 11/11 late germline. Strain 3: 11/11 non-fluorescent.	Variable	Position not verified. Homozygosity validated by homozygous Pmyo-2:GFP transmission.	EG7802	unc-119(ed3) III ; oxTi351[Pmyo-2::GFP::H2B ttTi5605 unc-18(+)] ; oxEx1828[cb-unc-119(+)]	
oxTi176	unc-18(+) + NeoR	V	15383969	Central autosome	3 strains: All strains variable with approx 50% early germline and 50% non-fluorescent.	Variable	Yes	EG7803	unc-119(ed3) III ; oxTi176[pCFJ687 unc-18(+)] V ; oxEx1807[cb-unc-119(+)]	
oxTi173	unc-18(+) + NeoR	V	17523246	Autosome arm w pairing center	1 strain: late germline all animals.	Off (X-like)	Yes	EG7804	unc-119(ed3) III ; oxTi173[pCFJ687 unc-18(+)] V ; oxEx1795[cb-unc-119(+) Kan]	
oxTi357	unc-18(+) + NeoR	V	20921413	Autosome arm w pairing center	3 strains. 2 strains late germline all animals. 1 strain non-fluorescent.	Off (X-like)	Yes	EG7805	unc-119(ed3) III ; oxTi357 V ; oxEx1876[cb-unc-119(+)]	
Previously published MosSCI sites (Frøkjær-Jensen et al., 2008)										
ttTi5605	None	II	8420090	Central autosome	6 strains: Early germline all strains	On	N.A.	EG6699	ttTi5605 II; unc-119(ed3) III	
cxTi10882	None	IV	4237760	Autosome arm w pairing center	3 strains. Strain 1: Early germline all animals. Strain 2: Late germline all animals. Strain 3: Non-fluorescent germline.	Variable	N.A.	EG6700	unc-119(ed3) III; cxTi10882 IV	

Previously published MosSCI sites (Frøkjær-Jensen et al., 2012)										
ttTi4348	None	I	2851043	Autosome are with no pairing center	3 strains. Strain 1+2: Early germline all animals. Strain 3: Variable germline expression (early, late, non-fluorescent)	Variable	N.A.	EG6701	ttTi4348 I; unc-119(ed3) III	
ttTi4391	None	I	11269568	Autosome arm w pairing center	4 strains. All animals non-fluorescent germline.	Off	N.A.	EG6702	ttTi4391 I; unc-119(ed3) III	
cxTi10816	None	IV	5014687	Central autosome	2 strains. Early germline all animals.	On	N.A.	EG6703	unc-119(ed3) III; cxTi10816 IV	
ttTi14024	None	X	15574632	X chromosome	3 strains. Strain 1+2: Late germline. Strain 3: non-fluorescent germline.	Off (X-like)	N.A.	EG6705	unc-119(ed3) III; ttTi14024 X	