Long-term experimental evolution reveals purifying selection on piRNA-mediated control of transposable element expression

Ulfar Bergthorsson[*1], Caroline J. Sheeba[*2,3], Anke Konrad[1], Tony Belicard[2,3], Toni Beltran[2,3], Vaishali Katju[^1], Peter Sarkies[^2,3]

^Joint corresponding
*co-first authors
[1] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77845, USA
[2] MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN UK
[3] Institute of Clinical Sciences, Imperial College London, Du Cane Road, London W12 0NN UK
^Vaishali Katju vkatju@cvm.tamu.edu
^ Peter Sarkies psarkies@imperial.ac.uk

Abstract

Transposable elements (TEs) are an almost universal constituent of eukaryotic genomes. In animals, Piwi-interacting small RNAs (piRNAs) and repressive chromatin often play crucial roles in preventing TE transcription and thus restricting TE activity. Nevertheless, TE content varies widely across eukaryotes and the dynamics of TE activity and TE silencing across evolutionary time is poorly understood. Here we used experimentally evolved populations of *C. elegans* to study the dynamics of TE expression over 400 generations. The experimental populations were evolved at three different population sizes to manipulate the efficiency of natural selection versus genetic drift. We demonstrate increased TE expression relative to the ancestral population, with the largest increases occurring in the smallest populations. We show that the transcriptional activation of TEs within active regions of the genome is associated with failure of piRNA-mediated silencing, whilst desilenced TEs in repressed chromatin domains retain small RNAs. Additionally, we find that the sequence context of the surrounding region influences the propensity of TEs to lose silencing through failure of small RNA-mediated silencing. Together, our results show that natural selection in *C. elegans* is responsible for maintaining low levels of TE expression, and provide new insights into the epigenomic features responsible.

Introduction

Transposable elements (TEs) are almost ubiquitous across eukaryotic genomes (Chuong et al., 2017). Their ability to replicate independently of the host genome, coupled with the existence of multiple copies liable to ectopic recombination means they present a potential threat to genome stability. Moreover, TEs pose a threat to genome function as new integrations can disrupt genes or gene regulatory elements.

48 As a result, organisms have evolved sophisticated control strategies, which protect the
49 genome from TE proliferation. Across eukaryotes, short (20-33 nucleotides) small
50 RNAs play an important role in the suppression of TE activity. Within animals, Piwi-
51 interacting small RNAs (piRNAs) are paramount in the TE defence armoury (Siomi et
52 al., 2011). piRNAs are produced from defined genomic loci named piRNA clusters
53 and after processing, associate with the Piwi subfamily of argonaute proteins
54 (Brennecke et al., 2007). They recognise TEs through sense-antisense base pairing
55 and target TEs for transcriptional and post-transcriptional silencing(Siomi et al.,
56 2011). In many model organisms, piRNAs are essential for fertility through their role
57 in controlling TE proliferation in the germline (Weick and Miska, 2014)
58
59 The nematode *Caenorhabditis elegans* is a well-established model for small-
60 RNA mediated silencing. piRNAs in *C. elegans* are unusual in that the two piRNA
61 clusters on Chromosome IV are composed of individual RNA polymerase II (RNA
62 pol II) transcription loci where each piRNA has its own upstream motif (Batista et al.,
63 2008; Das et al., 2008; Ruby et al., 2006; Wang and Reinke, 2008). piRNA clusters
64 are located within H3K27me3-rich chromatin, which, together with cis-acting RNA
65 pol II pausing sequences downstream of the piRNA, enforce production of ~28
66 nucleotide piRNA precursors(Beltran et al., 2019). piRNA precursors are further
67 trimmed to result in mature 21 nucleotide piRNAs with a Uracil as the first nucleotide
68 (21U-RNAs), which associate with the *C. elegans* Piwi protein PRG-1 (Batista et al.,
69 2008; Das et al., 2008; Wang and Reinke, 2008). Downstream of PRG-1, piRNA
70 silencing relies on a nematode-specific class of secondary small RNAs known as
71 22G-RNAs(Das et al., 2008). 22G-RNA synthesis is carried out by RNA-dependent
72 RNA polymerases using the target RNA as a template, following initiation by piRNA
73 target recognition(Pak and Fire, 2007). 22G-RNAs bind to Argonaute proteins and
74 lead to transcriptional and post-transcriptional silencing of target RNAs (Yigit et al.,
75 2006). Additionally 22G-RNAs can be transmitted transgenerationally (Buckley et
76 al., 2012) and as a result piRNA-initiated silencing can persist for many generations
77 even after piRNAs themselves are removed by mutating PRG-1 (Ashe et al., 2012;
78 Luteijn et al., 2012; Shirayama et al., 2012). Consequently whilst removal of piRNAs
79 alone has mild effects on TE expression, combining mutations of PRG-1 with
80 mutations disrupting the 22G-RNA biogenesis machinery leads to reactivation of
81 several TEs (de Albuquerque et al., 2015; Phillips et al., 2015)

82

83 Despite the universality of TEs across eukaryotes, there is striking variability
84 both in TE content and TE expression across species. Several interacting factors
85 have been proposed to account for this. First, silencing mechanisms differ between
86 organisms. For example, the entire piRNA pathway has been lost independently
87 multiple times in nematodes (Sarkies et al., 2015), and was lost in parasitic flatworms
88 (Fontenla et al., 2017; Skinner et al., 2014) and in dust mites (Mondal et al., 2018).
89 Second, it is possible that some TEs may have beneficial consequences through their
90 ability to act as reservoirs for evolutionary novelty. For example up to 60% of
91 human-specific enhancers may be TE-derived (Rebollo et al., 2011) and TE insertions
92 have been proposed to substantially rewire the human immune cell transcriptome
93 (Imbeault et al., 2017). TEs themselves may also be co-opted into developmental
94 programs. For example, transcription of L1 RNA is observed at the 2-cell stage in
95 mouse embryogenesis where it may have a direct role in coordinating gene expression
96 programs (Percharde et al., 2018). Furthermore, TEs may serve as a genome-wide
97 source of regulatory elements (Chuong et al., 2017). Despite these examples, TEs are
98 overall considered to be detrimental to fitness and beneficial TE insertions appear
99 overrepresented due to the effects of natural selection in weeding out deleterious
100 insertions (Simonti et al., 2017). The diversity of TEs across evolution may thus
101 reflect population genetics factors such as population structure and effective
102 population size. For example, even moderately deleterious TE insertions might
103 become fixed in very small populations as the intensity of selection decreases with
104 reduced population size. In agreement with this model, a recent large-scale study
105 across nematodes concluded that genetic drift was likely responsible for differences in
106 TE content across nematodes (Szitenberg et al., 2016).
107

108 In the context of these potential models to explain the diversity in TE content,
109 it is important to understand the extent to which the balance between TE expression
110 and TE regulation is under selection. One way to study this is to use a mutation
111 accumulation (MA) framework in which replicate lines descended from a single
112 common ancestor are propagated under a regime of drastic population bottlenecks for
113 several hundred generations (Halligan and Keightley, 2009; Katju and Bergthorsson,
114 2019). The maintenance of these lines at a minimal population size attenuates the
115 efficacy of selection, thereby enabling the accumulation of a large, unbiased sample

116 of spontaneous mutations under conditions of genetic drift which can subsequently be
117 identified and their fitness effects investigated. Previously, MA lines have been used
118 to investigate the rate and spectrum of TE copy-number changes in *Saccharomyces*
119 *cerevisiae*, *C. elegans* and *Drosophila melanogaster*, thus providing estimates of the
120 rate of TE transposition (Adrion et al., 2017; Bast et al., 2019; Bégin and Schoen,
121 2006, 2007). However, TE expression is not necessarily directly linked to TE copy-
122 number and may have independent fitness consequences. The opportunity to
123 interrogate genome-wide RNA expression has been exploited to investigate the effect
124 of mutations on protein-coding gene expression divergence (Denver et al., 2005;
125 Landry et al., 2007; Rifkin et al., 2005). Here, we extend this approach to investigate
126 the effect of spontaneous mutations on TE expression divergence.
127
128 We created spontaneous MA lines of *C. elegans* that were descended from a
129 single worm ancestor and propagated for ~400 generations under three population
130 size treatments of $N = 1$, 10 and 100 individuals per generation (Katju et al., 2015).
131 The varying population size treatment in the experiment permitted a manipulation of
132 the strength of selection, with the $N = 1$ lines evolving under close to neutral
133 conditions (minimal selection) and an incremental increase in the strength of selection
134 with increasing population size. We employed this framework to investigate how TE
135 expression evolves under conditions of near neutrality and under the influence of
136 increasing selection intensity. We show that overall TE expression increases in MA
137 lines with the smallest population size. We further show that expression increase
138 results in part from failure of piRNA-mediated silencing. Intriguingly, differences in
139 the responses of different TEs to reduced piRNA-mediated silencing depend on the
140 chromatin environment of the TE loci, such that TEs in repressed chromatin domains
141 largely remain silent due to epigenetic memory imparted by 22G-RNAs, whilst in
142 active chromatin domains, increased TE expression is much more likely to occur.
143 Together our results demonstrate for the first time that robust control of TE
144 expression is under selection in animals. Importantly further, our results provide new
145 insight into how the chromatin environment interacts with piRNA-mediated silencing
146 to control TE expression.
147
148
149

150 **Results**

151

152 *Relaxed selection leads to increased TE expression*

153

154 In order to assess the effect of selection on TE expression, we generated
155 mutation accumulation (MA) lines and propagated them by randomly selecting *N*
156 individuals at each generation, where *N* was either 1, 10 or 100 (Katju et al., 2015).
157 Henceforth we refer to the three conditions as *N*.1, *N*.10 or *N*.100 (Figure 1A). After
158 409 generations, we isolated RNA and performed RNA sequencing to investigate TE
159 expression. The fundamental difference between these conditions is the size of the
160 bottleneck that the population is subjected to in each generation. However, it is
161 inevitable that the population density will also vary between the three population size
162 treatments, resulting in environmental differences that could introduce variation in
163 gene expression among the experimental lines maintained at these three population
164 sizes. To counteract this, all lines were maintained in similar conditions including no
165 population bottleneck differences for three generations prior to RNA extraction.
166 Therefore, the most likely explanation for any differences in TE expression that we
167 observe is the different strength of purifying selection between the three population
168 size treatments during MA.

169

170 MA lines from all three population size treatments showed an increase in total
171 TE expression relative to the pre-MA ancestral control; moreover, *N*.1 had higher
172 total TE expression than *N*.10 or *N*.100. Across increasing population sizes, we
173 observed a monotonic decrease in total TE expression (Jonckheere test for ordered
174 medians; henceforth Jh, $p = 0.02$; Figure 1B). Similarly, linear regression analysis
175 showed a significant negative correlation between increasing population size and TE
176 expression (linear regression; henceforth LR, $p = 0.02$ Figure 1B). The mean
177 expression across all changes in TEs normalized to the pre-MA ancestral control
178 showed a significant tendency to decrease as the population size increased (Jh, $p =$
179 0.003; LR, $p = 0.02$; Supplemental Figure 1A).

180

181 Protein-coding gene expression diverges during MA in a variety of model
182 organisms including *C. elegans* (Denver et al., 2005; Rifkin et al., 2005; Landry et al.,
183 2007; Hodgins-Davis et al., 2015). In addition to an overall increase, TE expression

184 also appears to show broader distributions in the *N*.1 lines compared to the *N*.10 and
185 *N*.100 lines (Supplemental Figure 1B). To test this directly, we estimated the
186 variation in the expression of each individual TE and each individual gene. To
187 control for potential changes in the mean expression, which can affect noise, we
188 calculated the Fano factor [var(x)/mean(x)] within *N*.1, *N*.10 and *N*.100 lines
189 separately. Fano factors for TEs and genes were higher in *N*.1 lines than in *N*.10 or
190 *N*.100 lines (TEs: Wilcoxon paired test $p = 0.015$ and 0.004 for *N*.10 and *N*.100;
191 genes: Wilcoxon paired test $p < 1\times10^{-16}$ for both *N*.10 and *N*.100; Supplemental
192 Figures 1C and D). To control for the possibility that the larger number of *N*.1 lines
193 might lead to higher variance, we calculated TE Fano factors from 1,000 subsets of
194 five *N*.1 lines and all 252 subsets of five *N*.10 lines and compared these to the five
195 *N*.100 lines. This showed the same trend as the full dataset (Supplemental Figure 1E).
196 To further investigate the variation in expression, we calculated the total variance in
197 the change in expression of all TEs or all genes between each line and the pre-MA
198 ancestral control. Variance in the differences in both TE and gene expression
199 increased with smaller population sizes (Jh, $p = 0.008$ and 0.009, respectively)
200 (Supplemental Figure 1 F,G). However, importantly, there was no correlation in
201 overall variance between TEs and genes in the same line (Supplemental Figure 1H),
202 showing that TE expression and gene expression diverge independently.
203
204 In order to understand loss of repression of TE expression in more detail, we
205 classified TEs into different families using RepeatMasker annotations and compared
206 the median TE expression for each TE family across different lines. TE expression
207 was compared between *N*.1 and *N*>1 lines and tested for monotonic median increase
208 with decreasing population size. As expected, the majority of TEs that showed a
209 significant difference between *N*.1 and combined *N*.10 and *N*.100 lines had increased
210 expression in *N*.1 compared to the other lines (Figure 1C and Supplemental Figure 2).
211 However, individual TEs displayed different patterns of expression change. Some
212 TEs, notably the DNA transposon Turmoil2, showed more consistent increases across
213 the *N*.1 lines compared to the *N*.10 and *N*.100 lines (Figure 1D). Indeed, the majority
214 of the total effect on TE expression seen in Figure 1B could be attributed to one TE
215 family, the Turmoil2 TEs, which showed a large expression increase across the
216 majority of the *N*.1 lines (Figure 1C, D). Contrastingly some TEs, notably the
217 Mariner family DNA transposon Tc1, showed a burst-like pattern of expression where

218 a large increase in expression was observed in a few *N*.1 lines while retaining low
219 expression in the remainder *N*.1 lines as well as the *N*>1 lines (Figure 1E).
220
221 We next investigated the timecourse of TE desilencing during propagation of
222 the MA lines. We performed gene expression analysis by RNA-Seq on 11 *N*.1 lines
223 at 25 and 100 MA generations. Median total TE expression showed a highly
224 significant increase with increasing numbers of generations (Jh, $p = 2.6\times10^{-5}$; Figure
225 1F). Linear regression analysis confirmed a positive relationship between increased
226 numbers of generations and increased TE expression (LR, $p = 4.4\times10^{-3}$; Figure 1F).
227 Different TEs showed different kinetics of desilencing. Turmoil2 showed a positive
228 relationship between the number of MA generations and expression (LR $p = 3.9\times10^{-3}$)
229 and a monotonic increase in median expression (Jh, $p = 2.6\times10^{-3}$; Figure 1G).
230 Contrastingly, Tc1 desilencing did not show a positive linear relationship between the
231 number of MA generations and expression (LR, $p = 0.14$; Figure 1H) though there
232 was a significant increase in median expression (Jh, $p = 4.2\times10^{-6}$).
233
234 We further investigated whether the expression of TEs in individual MA lines
235 correlated with the expression of other TEs. The majority of TEs showed little
236 correlation with the expression of other TEs (Figure 2A, B). The TEs with the most
237 statistically significant increases in expression, Turmoil2 and non-LTR
238 retrotransposons of the LINE2 family (Figure 1C and Supplemental Figure 2)
239 clustered together (Figure 2B) suggesting coregulation of these TE families despite
240 their different mechanisms of replication.
241
242 *Expression of TEs is weakly associated with increased copy-number*
243
244 TEs are capable of replicating independently of the host genome and thus their
245 copy-number might change across MA lines. We sequenced the genomes of the MA
246 lines after 400 generations and mapped the reads to consensus TE sequences thereby
247 obtaining estimates of copy-number variation (CNV) for each TE family. Median TE
248 copy-number increased with decreasing population size (Jh, $p = 8.0\times10^{-4}$; LR, $p =$
249 4.2×10^{-3} Figure 3A). Across all lines, there was a significant positive correlation (LR,
250 $p = 0.01$) between increased copy-number and increased expression, although this was

251 driven by *N.*1 lines and not apparent in the *N.*10 and *N.*100 lines (Figure 3B).
252 Furthermore, Turmoil2 elements, which exhbitied the largest changes in expression,
253 displayed no correlation between expression changes and copy-number increases
254 (Figure 3C). We conclude that increased TE copy-number was not the primary cause
255 of increased TE expression. Moreover, increased expression of specific TEs did not
256 always lead to increased copy-number.
257
258 *Alterations in small RNA levels are associated with TE expression changes*
259
260 We investigated whether changes in regulation of TE expression could explain
261 the loss of silencing observed during mutation accumulation. In *C. elegans*, piRNAs
262 and 22G-RNAs are important small RNA classes involved in TE silencing (de
263 Albuquerque et al., 2015; Bagijn et al., 2012; Das et al., 2008; Phillips et al., 2015).
264 To test whether piRNAs are important in the loss of silencing of TEs in the *N.*1 lines,
265 we remapped recently published cross-linking immunoprecipitation (CLIP) data
266 (Shen et al., 2018) to identify TE transcripts that are bound by piRNA-Piwi
267 complexes *in vivo*. Approximately 25% of TE transcripts with RNA-Seq reads were
268 targeted by piRNAs. TEs targeted by piRNAs showed a statistically significant
269 increase in total expression in the *N.*1 lines (Jh, $p = 0.02$; LR, $p = 0.03$; Figure 4A)
270 whilst TEs that were not targeted by piRNAs were not significantly altered (Jh, $p =$
271 0.47; LR, $p = 0.27$; Figure 4B).
272
273 We tested whether defective piRNA-mediated silencing might account for
274 increased expression of TEs targeted by piRNAs. piRNAs in *C. elegans* are expressed
275 from individual promoters as a result of RNA polymerase II transcription (Billi et al.,
276 2013; Cecere et al., 2012; Gu et al., 2012). We considered two potential mechanisms
277 that might give rise to altered piRNAs. One possibility is that mutations in the
278 piRNA sequences might occur in individual lines, which might affect their ability to
279 recognise transposable elements and thus interfere with TE silencing. There was a
280 trend for *N.*1 lines to have more mutations in piRNAs than *N.*10 or *N.*100 lines
281 (Supplemental Figure 3). However, the trend was not significant (Jh, $p = 0.12$).
282 Moreover, on average, we identified only 1.1 piRNA sequences with mutations across
283 the *N.*1 lines. Thus, we conclude that the increase in TE expression is unlikely to be
284 related to mutations in specific piRNAs.

285

286 We next considered the expression of individual piRNA loci. We identified
287 piRNA loci with significantly altered median expression between *N*.1 lines and *N*.10
288 and *N*.100 lines combined using the Wilcoxon unpaired test. A small percentage of
289 piRNA loci showed significant differences but overall there was no trend for these
290 loci to show reduced expression in the *N*.1 lines. Indeed, these loci were more likely
291 to have higher expression in the *N*.1 lines (Figure 4C). Thus, changes in piRNA
292 expression are unlikely to explain the changes we observed in TE expression.
293

294 22G-RNAs act downstream of piRNAs to bring about target silencing(Das et
295 al., 2008). Surprisingly, although 22G-RNAs silence TEs, we found that the total
296 levels of 22G-RNAs mapping to TEs were increased in lines with smaller population
297 sizes, although this increase was on the border of significance (Jh, $p = 0.05$; LR, $p =$
298 0.07; Figure 4D). To examine this in more detail, we analysed 22G-RNAs at
299 individual TEs. The majority of TEs showing significantly increased 22G-RNAs in
300 *N*.1 lines were non-LTR transposons (Figure 4E). Notably, these TEs were not
301 significantly elevated in transcript levels in the *N*.1 lines (Figure 1C; Supplemental
302 Figure 2). Thus, the global increase in 22G-RNA levels does not relate directly to the
303 increased transcript levels of TEs in *N*.1 lines.
304

305 To compare directly how increased expression of TEs relates to 22G-RNA
306 mediated silencing, for each TE we compared 22G-RNA levels in lines with increased
307 expression to 22G-RNA levels in lines without increased expression. The only TE
308 showing alterations in both expression and 22G-RNA levels was Turmoil2. The
309 expression changes of Turmoil2 relative to the ancestral control across all lines
310 correlated inversely with changes in 22G-RNA levels (LR, $p = 2.0\times10^{-5}$, Figure 4F).
311 Thus, the increased transcript levels of Turmoil2 elements is associated with reduced
312 22G-RNAs. We conclude that the increased transcript levels of some TEs may be
313 caused by reduced small RNA-mediated silencing, but that some increases in TE
314 transcript levels occur independently of 22G-RNA changes.
315

316 *Chromatin environment is associated with the relationship between 22G-RNA levels*
317 *and TE expression*

318

319 22G-RNAs interact with chromatin modifying factors to control expression of
320 TEs (McMurchy et al., 2017). piRNA-mediated silencing has been directly linked to
321 the generation of H3K9me2/3 marked nucleosomes ("classical heterochromatin") and
322 this has been proposed to be important for transcriptional silencing induced by
323 piRNAs (Ashe et al., 2012; McMurchy et al., 2017; Shirayama et al., 2012;
324 Woodhouse et al., 2018). Additionally, it is becoming clear that a large proportion of
325 the autosomal DNA in *C. elegans* can be divided into active domains, containing
326 H3K36me3 and germline expressed genes and regulated domains, containing
327 H3K27me3 marked nucleosomes and silent genes (Evans et al., 2016; Liu et al.,
328 2011). These domains are largely stable through development, including in the adult
329 germline (Evans et al., 2016; Rechtsteiner et al., 2010). We examined the influence
330 of these three types of chromatin on the response of TEs to mutation accumulation.
331 TEs in regulated domains and in classical heterochromatin showed no significant
332 overall increase in expression (Figure 5A, B). The rare examples of *N*.1 lines
333 showing significantly higher expression of TEs in classical heterochromatin
334 corresponded to lines in which Tc1 reactivation occurred, consistent with enrichment
335 of Tc1 elements within these regions (McMurchy et al., 2017). Contrastingly, TEs in
336 active domains had significantly higher expression in *N*.1 than *N*>1 lines (Wilcoxon
337 unpaired test, $p = 0.02$). There was also a trend towards decreased median expression
338 with increasing population size although this was on the border of significance (Jh, p
339 $= 0.07$; LR, $p = 0.08$; Figure 5C).
340

341 We next examined 22G-RNAs mapping to TEs across different chromatin
342 domains. TEs in repressed domains showed significantly increased levels of 22G-
343 RNAs in *N*.1 lines relative to *N*.10 and *N*.100. However, in active domains there was
344 no significant change in 22G-RNA levels (Figure 5D-F). Thus, increased 22G-RNAs
345 occur predominantly in TEs within repressed chromatin.
346

347 *AT-rich sequences in TEs reduce the generation of 22G-RNAs*
348

349 A recent study has demonstrated that silencing of both transgenes and
350 endogenous genes by 22G-RNAs is inhibited by a high content of periodic repeats of
351 AT-rich sequences, known as PATCs (Frøkjær-Jensen et al., 2016). We tested how

352 PATC density within TEs was associated with changes in their expression under
353 reduced selection. High PATC density corresponded to reactivation of TEs (Jh, p =
354 0.02; LR, p = 0.02) whereas TEs with low PATC density did not show an increase in
355 expression (Figure 6A). Contrastingly, only TEs with low PATC density showed
356 significantly increased 22G-RNAs in *N*.1 lines relative to *N*.10 and *N*.100 (Jh, p =
357 0.02; LR, p = 0.01; Figure 6B). Importantly this effect was specific to PATC
358 sequences as GC-content alone had no significant effect on either TE expression or
359 small RNA generation (Supplemental Figure 4A, B). We conclude that low PATC
360 density is required for 22G-RNA generation, which may be required to restrain TE
361 activation. We tested whether the chromatin environment modulated the effect of
362 PATC sequences on TE reactivation in MA lines. Importantly, PATC content was
363 similar in TEs across active, classic heterochromatin and regulated domains
364 (Supplemental Figure 4C). 22G-RNAs were significantly increased in low PATC
365 regions within regulated domains (Jh, p = 0.02; LR, p = 0.11) and classical
366 heterochromatin domains (Jh, p = 0.06; LR, p = 0.04) but not in active domains
367 (Figure 6C, D). In contrast, there were no significant changes in 22G-RNA levels in
368 high PATC regions within these domains (Figure 6C, D). Thus, the generation of
369 increased 22G-RNAs against TEs in MA lines is associated with both low PATC
370 content and a repressive chromatin environment.
371
372 **Discussion**
373
374 Our analysis of how the interplay between TE expression and TE silencing
375 factors changes over 400 generations at small population sizes provides the first clear
376 view of how TE expression diverges under reduced selection in animals.
377 Additionally, closer analysis of how TE control mechanisms are affected in the MA
378 lines offers new insight into the fundamental mechanisms of TE silencing in *C.*
379 *elegans*, underlining the ability of experimental evolutionary studies to derive
380 fundamental molecular insights. Here we discuss each of these aspects of our work in
381 turn.
382
383 *The effect of selection on TE expression*
384

385 Here, we demonstrate that in *C. elegans*, TEs drift to higher expression under
386 conditions of strong genetic drift and reduced efficiency of selection. This is
387 consistent with the action of purifying selection on TE expression. Only a single TE,
388 Vingi-1, showed the opposite trend of lower expression in *N*.1 lines relative to the
389 *N*.10 or *N*.100 lines (Figure 1B). However, the expression of this element was actually
390 lower in the ancestral control than in the *N*.10 or *N*.100 lines. Thus, the significance of
391 this observation is unclear. Our observations suggest that the expression of most TEs
392 is largely detrimental and TE expression is under purifying selection. It remains a
393 formal possibility that TE activation is beneficial under fluctuating environmental
394 conditions or low fitness as opposed to the stable environment of the laboratory.
395
396 An important point raised by our results is that that not all expression
397 increases of TEs are linked to increasing copy-number; indeed many lines with very
398 high expression of specific TEs display no evidence of increased copy-number at all.
399 This suggests that many TEs replicate inefficiently in *C. elegans* such that even very
400 large increases in expression levels do not automatically result in increased copy-
401 numbers. This also implies that TE expression may be detrimental without directly
402 posing a threat to genome integrity, potentially through effects on endogenous gene
403 expression networks or through toxicity of repetitive RNA within the cell (Simon et
404 al., 2014).
405
406 Phenotypic analyses of previous *C. elegans* MA experiments suggest that the
407 decline in fitness in $N = 1$ lines results primarily from a few mutations with large
408 effects (Estes et al., 2004; Halligan et al., 2003; Katju et al., 2015, 2018; Keightley
409 and Cabellero, 1997). Similar results have been obtained in experimental evolution
410 studies in *D. melanogaster* (Ávila and García-Dorado, 2002) and bacteria (Dillon and
411 Cooper, 2016; Heilbron et al., 2014). In our study, purifying selection at larger
412 population sizes (*N*.10 and *N*.100) would eliminate such large-effect mutations and
413 indeed, our *N*.10 and N. 100 lines exhibited no evidence of fitness reduction over the
414 course of successive bottlenecking for 409 generations (Katju et al., 2015, 2018).
415 Analysis of gene expression data from MA experiments from *C. elegans*, *D.*
416 *melanogaster* and *S. cerevisiae* concluded that large effect mutations are also
417 responsible for changes in protein-coding gene expression (Hodgins-Davis et al.,
418 2015). However, our results for TE expression do not seem to fit with this model

419 because overall TE expression, which is largely driven by Turmoil2 and non-LTR
420 elements (see Figure 1), increases gradually with time across the *N*.1 lines and is also
421 increased, although less so, in *N*.10 and *N*.100 lines. The expression of Tc1 is an
422 exception to the overall trend as it is not affected in *N*.10 or *N*.100 lines but a small
423 number of *N*.1 lines show markedly increased Tc1 expression. Thus, Tc1 reactivation
424 may be dominated by a few mutations with large effect. This difference might be
425 related to the different mechanism of silencing of Tc1 compared to Turmoil2
426 elements as discussed further below.
427
428
429 *Weakened piRNA silencing is responsible for increased expression of TEs under*
430 *relaxed selection.*
431
432 Our investigations of the molecular mechanisms behind reduced silencing of
433 TEs in MA lines strongly suggest that defective piRNA silencing is a major culprit.
434 Only piRNA-targeted TEs show increased expression in MA lines, and, whilst
435 piRNAs themselves do not seem to change significantly in MA lines, the levels of
436 22G-RNAs that act as effectors of piRNA silencing are perturbed at the Turmoil2 TEs
437 that show increased expression. Why is piRNA silencing so vulnerable to mutation
438 accumulation? TE silencing and TE activation in organisms are likely in a precarious
439 equilibrium due to a constant evolutionary arms race between TEs and their host
440 genome. As a result, many mutations could converge on the piRNA pathway to
441 throw TE silencing out of balance.
442
443 *New insights into the role of chromatin in the piRNA pathway*
444
445 Our analysis of how mutation accumulation affects TE silencing provides
446 novel insights into how the chromatin environment of TEs might affect piRNA-
447 mediated silencing in *C. elegans* (Figure 7). Importantly, TEs in repressive chromatin
448 are much less prone to reactivation than those in active chromatin regions.
449 Mechanistically this is because 22G-RNAs in repressive chromatin regions are stable
450 or even increased, whilst 22G-RNAs mapping to TEs in active regions are reduced in
451 MA lines with increased expression. This result may also explain why reactivation of
452 Tc1 elements occurs less frequently than Turmoil2 elements, because Tc1 elements

453 are predominantly located in repressed domains and are therefore silenced more
454 robustly.
455
456 What is the mechanism whereby silencing memory is supported in repressed
457 chromatin regions? The simplest possibility is that silencing and generation of 22G-
458 RNAs are directly promoted by repressive chromatin modifications. In line with this
459 possibility, a mutually reinforcing loop between H3K9 methylation and small RNAs
460 is well documented in fission yeast (Bühler, 2009), and H3K9 methylation factors
461 contribute to silencing of transgenes in *C. elegans* (Ashe et al., 2012; McMurchy et
462 al., 2017; Shirayama et al., 2012; Woodhouse et al., 2018) although the situation is
463 more complicated for endogenous genes (Ni et al., 2014). However, our observations
464 hold equally well for H3K27me3-repressed chromatin, which has not been directly
465 linked to 22G-RNA silencing. We propose therefore that the nuclear small RNA
466 pathway responds differently depending on whether surrounding genes are active or
467 repressed to detect and quell aberrant gene activation. This model will be of interest
468 for further mechanistic investigation of small-RNA mediated silencing in *C. elegans*.
469
470
471
472

473 *Figure Legends*

474

475 **Figure 1** Increased expression of transposable elements in mutation accumulation
476 lines. **A** Diagram of the mutation accumulation (MA) experimental design with
477 different population sizes. **B** Overall transposable element expression in MA lines as
478 a function of population size. **C** Volcano plot for individual TE genes coloured
479 according to family. The x axis shows the median difference between *N*.1 and
480 combined *N*.10 and *N*.100 lines such that higher values correspond to higher
481 expression in smaller population sizes. The p-value is a Wilcoxon unpaired test
482 comparing the median in *N*.1 lines to *N*.10 and *N*.100. **D** Expression changes in
483 Turmoil2 elements. The box shows interquartile range with a line at median and the
484 whiskers extend to the furthest point ≤ 1.5 times the interquartile range from the
485 median. **E** Expression changes in Tc1 elements. Boxplot as in D. **F** Total TE
486 expression after different numbers of generations of mutation accumulation in lines
487 with a population size of 1. **G** Turmoil2 expression after different numbers of
488 generations of mutation accumulation with a population size of 1. **H** Tc1 expression
489 after different numbers of generations of mutation accumulation with a population
490 size of 1.

491

492 **Figure 2** A heatmap of expression of individual TE elements illustrating significant
493 differences in expression between different population size treatments. **B** A heatmap
494 illustrating the significance in correlation in expression between different TEs across
495 all the MA lines. The colour intensity shows the significance of the Spearman's rank
496 correlation coefficient with blue showing a negative correlation and orange a positive
497 correlation.

498

499 **Figure 3** Relationship between TE copy-number and expression in MA lines . **A**
500 Increase in TE copy-number is associated with stronger genetic drift (weaker
501 selection). Data shows mean of TE copy-number across all TEs normalized to the
502 mean across all lines. **B** Correlation between total change in TE copy-number across
503 all TEs and total TE expression. **C** Lack of correlation between Turmoil2 RNA levels
504 and copy-number changes in MA lines.

505

506 **Figure 4** Perturbed 22G-RNAs are associated with changes in TE expression in MA
507 lines. **A and B** Comparison of differences in TE expression in MA lines of different
508 population sizes for piRNA targeted and non-piRNA targeted TEs, respectively. **C**
509 Total piRNA levels in MA lines across different population sizes. **D** Levels of 22G-
510 RNAs mapping to all TEs in MA lines at different population sizes. **E** Volcano plot
511 showing the gradient of change in 22G levels for different TEs relative to the
512 population size against their *p*-values on the y-axis. The gradient of change was
513 estimated by linear regression. Negative values on the x-axis correspond to higher
514 22G-RNA levels in lines with smaller population sizes.
515 **F** 22G-RNAs mapping to Turmoil2 plotted against Turmoil2 expression across all
516 MA lines.
517
518 **Figure 5** Chromatin environment controls alterations in piRNA mediated silencing in
519 MA lines. **A, B, and C** Expression changes in TEs in Regulated (H3K27me3),
520 Classic Heterochromatin (H3K9me2) and Active (H3K36me3) chromatin domains.
521 *P*-values are Wilcoxon unpaired test. **D, E, and F** Changes in TE-mapping 22Gs in
522 Regulated, Classic Heterochromatin and Active domains.
523
524 **Figure 6 PATC content influences 22G-RNA generation in MA lines**
525 **A** Expression differences in TEs in four equal-size bins of 21 TEs with decreasing
526 PATC content. **B** 22G-RNAs across the bins used in A. **C** and **D** Stratification of
527 bins from A and B into regulated (H3K27me3) chromatin, active (H3K36me3)
528 chromatin, and heterochromatin (H3K9me2) domains. High PATC is the top bin and
529 low PATC is the lowest bin.
530
531 **Figure 7 Chromatin and PATC together influence TE control in MA lines**
532 Model for how chromatin environment contributes to TE desilencing in MA lines.
533 Left- repressed chromatin domains, taking H3K27me3 enriched domains as an
534 example. Right, Active chromatin domains.
535
536
537 *Supplemental Figure Legends*
538
539 **Supplemental Figure 1**

540 **A** Mean change in expression relative to the ancestral control across all TEs for each
541 line. **B** Violin plots showing distribution of changes in TE expression across all TEs
542 for each line. Lines are ordered by increasing mean TE expression change (L to R).
543 **C** Fano factor of individual TE transcript levels across all lines of the indicated
544 population size. **D** Fano factor in individual protein-coding gene transcript levels
545 across all lines of the indicated population size. **E** Fano factor of TE transcript levels
546 in 1,000 samples of five *N*.1 lines, 253 samples of five *N*.10 lines [the maximum] and
547 all five *N*.100 lines. **F** Total variance in transcript level differences between each TE
548 and its corresponding value in the starting population across lines of different
549 population size. **G** Total variance of transcript level differences between each
550 protein-coding gene and the starting population in lines of the indicated population
551 size. Boxplots for A-D are as in Figure 1D. **H** Variance in TE transcript changes
552 relative to starting population compared to the variance in protein-coding gene
553 transcript changes in the same line.
554
555 **Supplemental Figure 2**
556 Volcano plot showing the logarithm (base 10) *p*-value of a linear model relating
557 expression of TEs to population size on the y-axis to the gradient of the linear model
558 on the x-axis.
559
560 **Supplemental Figure 3**
561 Total piRNA reads showing one base pair mismatch to the reference sequence divided
562 by the total number of mismatched loci, in MA lines across different population sizes.
563
564 **Supplemental Figure 4**
565 **A** GC content does not affect reactivation of TEs in the *N*.1 lines. Bins with high to
566 low GC content, left to right. **B** GC content does not affect the change in 22G-RNA
567 levels. Bins with high to low GC content, left to right. **C** No clear difference in the
568 proportion of TEs from different chromatin domains within bins of different PATC
569 content. **D** Expression of TEs from regulated, heterochromatic and active chromatin
570 regions in the top PATC bin. **E** Expression of TEs from regulated, heterochromatic
571 and active chromatin regions in the lowest PATC bin.
572
573 **Additional Files**

R code and input files required to generate the plots in the figures will be released along with final publication and are available upon reasonable request to psarkies@imperial.ac.uk

578 **Methods**

579

580 *Spontaneous Mutation Accumulation (MA) Experiment at Varying Population Sizes*
581 *and Its Theoretical Underpinnings*

582

583 The descendants of a single wild-type Bristol (N2) hermaphrodite were used
584 to establish 35 MA lines. Twenty MA lines were propagated through single individual
585 descent, while ten lines were bottlenecked to ten individuals, and five lines were
586 bottlenecked to 100 individuals every generation (Katju et al., 2015, 2018; Konrad et
587 al., 2018)(Katju et al., 2015, 2018; Konrad et al. 2018). The size of these bottlenecks
588 was ensured through careful counting of randomly chosen L4 larvae selected to
589 establish a new generation every four days. All populations were grown on NGM
590 plates (Nematode Growth Medium) at 20˚C (Figure 1A). The lines were propagated
591 through 409 generations or until extinction under standard laboratory conditions.
592 Populations of size $N = 1$ (N.1 lines) and 10 (N.10 lines) were maintained on 60×15
593 mm petri dishes seeded with 250μl of *Escherichia coli* (OP50) in YT medium. $N =$
594 100 (or N.100) populations were housed on 90×15 mm petri dishes seeded with
595 750μl OP50 in YT medium.

596

597 The fitness effects of mutations range continuously from lethal to deleterious
598 to neutral to beneficial. In small populations, beneficial mutations can be lost and
599 detrimental mutations fixed by random chance events (genetic drift). The loss or
600 fixation of mutations depend upon both their selection coefficients (s) and the
601 effective population size, N_e. It has been shown that for sexually reproducing diploids,
602 the dynamics of mutations with $|s| << 1/2N_e$ are dominated by random genetic
603 drift(Kimura, 1962). Therefore, small populations subjected to attenuated selection
604 and an increased magnitude of genetic drift can potentially accumulate mutations with
605 extremely large effects in addition to ones with moderate to very slight effects. With
606 increasing population size, the efficiency of natural selection is increased. The
607 differences in populations size in this MA experiment alters the relative importance of
608 genetic drift versus natural selection in the fixation or loss of mutations, with genetic
609 drift having the greatest influence in $N = 1$ lines and natural selection having greater

610 influence in populations that were bottlenecked at 10 and 100 individuals each
611 generation.
612
613 *RNA Library Preparation, Sequencing, and Analysis of Transcript Abundance*
614
615 The library preparation and RNA-sequencing procedures have previously been
616 described in detail (Konrad et al., 2018). Briefly, we isolated one, two, and three
617 individuals each of the *N*.1, *N*.10, and *N*.100 lines, respectively. These 55 worms, as
618 well as one individual from the ancestral population, were each sequestered to NGM
619 plates seeded with OP50, where they were allowed to self-fertilize and reproduce at
620 20ºC. Three offspring worms at the L4 larval stage were isolated from each of the F_1
621 populations to serve as biological replicates. These 168 individual worm samples
622 were allowed to reproduce for three generations to yield enough tissue for RNA
623 extraction. A standard bleaching protocol was used to collect gravid eggs from adults
624 in order to generate synchronized populations of L1 larvae. Total RNA was isolated
625 from L1 larvae via the Qiagen RNeasy Mini Kit. The Nanodrop 2000, Qubit 3.0
626 Fluorometer, and an Agilent RNA Analyzer were used to evaluate the quality of the
627 RNA samples, and an Illumina TruSeq RNA library Prep Kit v2 was used with
628 standard procedures to prepare the RNA sequencing libraries for each sample at the
629 Texas A&M University Genomics and Bioinformatics Services Center. The RNA was
630 fragmented and Illumina adapters were annealed for amplification. Size selected
631 cDNA fragments were isolated via a Qiagen Gel Extraction Kit. Finally, sequencing
632 was performed on the Illumina HiSeq 4000 platform with default quality filters.
633
634 Demultiplexing and prefiltering of the sequencing reads was performed based
635 on default Illumina QC protocols. Reads containing abnormally short insert lengths
636 were removed, and adapters were discarded from the reads. The raw RNA-sequencing
637 reads in fastq format were aligned to the protein-coding transcriptome of *C. elegans*
638 (Wormbase reference N2 genome version WS247) using TopHat(Trapnell et al.,
639 2009) via the "very sensitive" bowtie2 algorithm with a maximum of one mismatch in
640 the anchor region for each spliced alignment and a minimum and maximum intron
641 length of 20 and 3,000 bp, respectively. Cufflinks(Trapnell et al., 2010) with default
642 settings and gene annotations from the N2 genome version WS247 was used to
643 estimate the relative transcript abundance for each protein-coding gene. All following

644 analyses were focused on FPKM values calculated on the per gene level. The relative
645 transcript abundances (FPKM) from the three biological replicates for each original
646 sample were averaged to get mean relative transcript abundance for each gene in that
647 sample.
648
649 *Small RNA sequencing*
650
651 MA lines were synchronised using hypochlorite treatment and embryos were
652 isolated after 12 hours. RNA was extracted using trizol and small RNA libraries were
653 prepared using the Illumina Small RNA sequencing kit as described
654 previously(Sarkies et al., 2015). Small RNAs were aligned to a genome built using
655 bowtie from a fasta file containing all piRNAs, miRNAs, ncRNAs and genes
656 including TEs, extracted from Wormbase (WS264; ce11), requiring perfect mapping.
657 Reads mapping to the sense strand of ncRNAs and miRNAs were extracted using
658 bedtools intersect –c –S. We used DEseq using ncRNAs and miRNAs to extract size
659 factors. 22G-RNAs mapping to TEs and genes were extracted using a custom Perl
660 script and the number of 22G-RNAs mapping antisense to each gene and TE was then
661 counted using bedtools intersect –c -s. The 22G-RNAs were then normalized to the
662 size factors from ncRNAs and miRNAs combined.
663
664 *TE copy-number analysis*
665
666 DNA sequence reads were aligned to a genome built using bowtie2-build from
667 TE consensus sequences extracted from repbase combined with all coding sequences.
668 Bowtie2 was used to map PE reads to this genome and the read count mapping to
669 each CDS or TE was obtained using bedtools intersect –c.
670
671
672 *Computational analysis of TE expression and small RNA analysis*
673
674 TEs from WS264 were annotated using Repeatmasker(Smit et al., 2017). All
675 data analysis was conducted using the R environment for statistical analysis
676 (www.Rproject.com). Details of the individual analyses are documented in the R
677 markdown file accompanying this manuscript (additional file 1), with the raw data

678 tables required to run these programs in as a zipped file (additional file 2). Previously
679 published datasets containing Chromatin domain annotations from Early Embryo
680 ChiP-Seq were taken from Evans et al. (2016) updated to WS264 using liftover
681 (https://genome.ucsc.edu/). Small RNA sequencing data from reactivation of small
682 RNA pathways in the presence or absence of piRNAs was taken from Phillips et al.
683 (2015) and aligned to the *C. elegans* genome as described above. The average PATC
684 score for each TE was calculated by taking the average PATC score across the
685 element from per-base sliding window genome-wide PATC scores from (Frøkjær-
686 Jensen et al., 2016).

687 **REFERENCES**
688

689 Adrion, J.R., Song, M.J., Schrider, D.R., Hahn, M.W., and Schaack, S. (2017).
690 Genome-wide estimates of transposable element insertion and deletion rates in
691 *Drosophila melanogaster*. Genome Biol. Evol. *9*, 1329–1340.

692 de Albuquerque, B.F.M., Placentino, M., and Ketting, R.F. (2015). Maternal piRNAs
693 are essential for germline development following de novo establishment of endo-
694 siRNAs in *Caenorhabditis elegans*. Dev. Cell *34*, 448–456.

695 Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C.,
696 Doebley, A.L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J., et al. (2012). PiRNAs can
697 trigger a multigenerational epigenetic memory in the germline of *C. elegans*. Cell
698 *150*, 88–99.

699 Ávila, V., and García-Dorado, A. (2002). The effects of spontaneous mutation on
700 competitive fitness in *Drosophila melanogaster*. J. Evol. Biol. *15*, 561–566.

701 Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach,
702 N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of
703 *Caenorhabditis elegans* piRNAs. Science (80-.). *337*, 574–578.

704 Bast, J., Jaron, K.S., Schuseil, D., Roze, D., and Schwander, T. (2019). Asexual
705 reproduction reduces transposable element load in experimental yeast
706 populations. Elife *8*.

707 Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D.,
708 Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs
709 interact to form the piRNA complex required for fertility in *C. elegans*. Mol. Cell
710 *31*, 67–78.

711 Bégin, M., and Schoen, D.J. (2006). Low impact of germline transposition on the
712 rate of mildly deleterious mutation in *Caenorhabditis elegans*. Genetics *174*,
713 2129–2136.

714 Bégin, M., and Schoen, D.J. (2007). Transposable elements, mutational
715 correlations, and population divergence in *Caenorhabditis elegans*. Evolution (N.
716 Y). *61*, 1062–1070.

717 Beltran, T., Barroso, C., Birkle, T.Y., Stevens, L., Schwartz, H.T., Sternberg, P.W.,
718 Fradin, H., Gunsalus, K., Piano, F., Sharma, G., et al. (2019). Comparative
719 epigenomics reveals that rna polymerase ii pausing and chromatin domain
720 organization control nematode piRNA biogenesis. Dev. Cell *48*, 793-810.e6.

721 Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K.
722 (2013). a conserved upstream motif orchestrates autonomous, germline-
723 enriched expression of *Caenorhabditis elegans* piRNAs. PLoS Genet. *9*.

724 Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and
725 Hannon, G.J. (2007). Discrete small rna-generating loci as master regulators of
726 transposon activity in *Drosophila*. Cell *128*, 1089–1103.

727 Buckley, B.A., Burkhart, K.B., Gu, S.G., Spracklin, G., Kershner, A., Fritz, H., Kimble,
728 J., Fire, A., and Kennedy, S. (2012). A nuclear Argonaute promotes
729 multigenerational epigenetic inheritance and germline immortality. Nature *489*,
730 447–451.

731 Bühler, M. (2009). RNA turnover and chromatin-dependent gene silencing.
732 Chromosoma *118*, 141–151.

733 Cecere, G., Zheng, G.X.Y., Mansisidor, A.R., Klymko, K.E., and Grishok, A. (2012).
734 Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Mol.
735 Cell *47*, 734–745.

736 Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of
737 transposable elements: From conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

738 Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig,
739 A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and
740 piRNAs act upstream of an endogenous siRNA pathway to suppress tc3
741 transposon mobility in the *Caenorhabditis elegans* germline. Mol. Cell *31*, 79–90.

742 Denver, D.R., Morris, K., Streelman, J.T., Kim, S.K., Lynch, M., and Thomas, W.K.
743 (2005). The transcriptional consequences of mutation and natural selection in
744 Caenorhabditis elegans. Nat. Genet. *37*, 544–548.

745 Dillon, M.M., and Cooper, V.S. (2016). The fitness effects of spontaneous
746 mutations nearly unseen by selection in a bacterium with multiple
747 chromosomes. Genetics *204*, 1225–1238.

748 Estes, S., Phillips, P.C., Denver, D.R., Thomas, W.K., and Lynch, M. (2004).
749 Mutation accumulation in populations of varying size: the distribution of
750 mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics *166*,
751 1269–1279.

752 Evans, K.J., Huang, N., Stempor, P., Chesney, M.A., Down, T.A., and Ahringer, J.
753 (2016). Stable *Caenorhabditis elegans* chromatin domains separate broadly
754 expressed and developmentally regulated genes. Proc. Natl. Acad. Sci. *113*,
755 E7020--E7029.

756 Fontenla, S., Rinaldi, G., Smircich, P., and Tort, J.F. (2017). Conservation and
757 diversification of small RNA pathways within flatworms. BMC Evol. Biol.

758 Frøkjær-Jensen, C., Jain, N., Hansen, L., Davis, M.W., Li, Y., Zhao, D., Rebora, K.,
759 Millet, J.R.M., Liu, X., Kim, S.K., et al. (2016). An abundant class of non-coding DNA
760 can prevent stochastic gene silencing in the *C. elegans* germline. Cell *166*, 343–
761 357.

762 Gu, W., Lee, H.-C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and Mello,
763 C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped
764 small RNAs as C. elegans piRNA precursors. Cell *151*, 1488–1500.

765 Halligan, D.L., and Keightley, P.D. (2009). Spontaneous mutation accumulation

766 studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. *40*, 151–172.

767 Halligan, D., Peters, A., and Keightley, P. (2003). Estimating numbers of EMS-
768 induced mutations affecting life history traits in *Caenorhabditis elegans* in
769 crosses between inbred sublines. Genet. Res. *82*, 191–205.

770 Heilbron, K., Toll-Riera, M., Kojadinovic, M., and MacLean, R.C. (2014). Fitness is
771 strongly influenced by rare mutations of large effect in a microbial mutation
772 accumulation experiment. Genetics *197*, 981–990.

773 Hodgins-Davis, A., Rice, D.P., and Townsend, J.P. (2015). Gene expression evolves
774 under a house-of-cards model of stabilizing selection. Mol. Biol. Evol. *32*, 2130–
775 2140.

776 Imbeault, M., Helleboid, P.-Y., and Trono, D. (2017). KRAB zinc-finger proteins
777 contribute to the evolution of gene regulatory networks. Nature *543*, 550–554.

778 Katju, V., and Bergthorsson, U. (2019). Old trade, new tricks: insights into the
779 spontaneous mutation process from the partnering of classical mutation
780 accumulation experiments with high-throughput genomic approaches. Genome
781 Biol. Evol. *11*, 136–165.

782 Katju, V., Packard, L.B., Bu, L., Keightley, P.D., and Bergthorsson, U. (2015).
783 Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis*
784 *elegans* with varying effective population sizes. Evolution (N. Y). *69*, 104–116.

785 Katju, V., Packard, L.B., and Keightley, P.D. (2018). Fitness decline under osmotic
786 stress in *Caenorhabditis elegans* populations subjected to spontaneous mutation
787 accumulation at varying population sizes. Evolution (N. Y). *72*, 1000–1008.

788 Keightley, P.D., and Cabellero, A. (1997). Genomic mutation rates for lifetime
789 reproductive output and lifespan in Caenorhabditis elegans. Proc. Natl. Acad. Sci.
790 USA *94*, 3823–3827.

791 Kimura, M. (1962). On the probability of fixation of mutant genes in a population.
792 Genetics *47*, 713-719

793 Konrad, A., Flibotte, S., Taylor, J., Waterston, R.H., Moerman, D.G., Bergthorsson,
794 U., and Katju, V. (2018). Mutational and transcriptional landscape of spontaneous
795 gene duplications and deletions in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci.
796 *115*, 7386–7391.

797 Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J., and Hartl, D.L. (2007). Genetic
798 properties influencing the evolvability of gene expression. Science (80-.). *317*,
799 118–121.

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

804 Luteijn, M.J., van Bergeijk, P., Kaaij, L.J.T., Almeida, M.V., Roovers, E.F., Berezikov,
805 E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in
806 *Caenorhabditis elegans*. EMBO J. *31*, 3422–3430.

807 McMurchy, A.N., Stempor, P., Gaarenstroom, T., Wysolmerski, B., Dong, Y.,
808 Aussianikava, D., Appert, A., Huang, N., Kolasinska-Zwierz, P., Sapetschnig, A., et
809 al. (2017). A team of heterochromatin factors collaborates with small RNA
810 pathways to combat repetitive elements and germline stress. Elife *6*.

811 Mondal, M., Klimov, P., and Flynt, A.S. (2018). Rewired RNAi-mediated genome
812 surveillance in house dust mites. PLoS Genet. *14*.

813 Ni, J.Z., Chen, E., and Gu, S.G. (2014). Complex coding of endogenous siRNA,
814 transcriptional silencing and H3K9 methylation on native targets of germline
815 nuclear RNAi in *C. elegans*. BMC Genomics *15*, 1157.

816 Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary
817 effectors during RNAi in C. *elegans*. Science (80-.). *315*, 241–244.

818 Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A.,
819 Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-
820 nucleolin partnership regulates early development and ESC Identity. Cell *174*,
821 391-405.e19.

822 Phillips, C.M., Brown, K.C., Montgomery, B.E., Ruvkun, G., and Montgomery, T.A.
823 (2015). PiRNAs and piRNA-dependent siRNAs protect conserved and essential *C.*
824 *elegans* genes from misrouting into the RNAi Pathway. Dev. Cell *34*, 457–465.

825 Rebollo, R., Romanish, M.T., and Mager, D.L. (2011). Transposable Elements: an
826 abundant and natural source of regulatory sequences for host genes. Annu. Rev.
827 Genet. *46*, 120913153128008.

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

832 Rifkin, S.A., Houle, D., Kim, J., and White, K.P. (2005). A mutation accumulation
833 assay reveals a broad capacity for rapid evolution of gene expression. Nature
834 *438*, 220–223.

835 Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel,
836 D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional
837 microRNAs and endogenous siRNAs in *C. elegans*. Cell *127*, 1193–1207.

838 Sarkies, P., Selkirk, M.E., Jones, J.T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B.,
839 Ardila-Garcia, A., Fast, N.M., Schiffer, P.M., et al. (2015). Ancient and novel small
840 RNA pathways compensate for the loss of piRNAs in Multiple Independent
841 Nematode Lineages. PLoS Biol. *13*.

842 Shen, E.-Z., Chen, H., Ozturk, A.R., Tu, S., Shirayama, M., Tang, W., Ding, Y.-H., Dai,

843 S.-Y., Weng, Z., and Mello, C.C. (2018). Identification of piRNA binding sites
844 reveals the argonaute regulatory landscape of the *C. elegans* germline. Cell *172*,
845 937-951.e18.

846 Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte, D., and Mello, C.C.
847 (2012). piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans*
848 germline. Cell *150*, 65–77.

849 Simon, M., Sarkies, P., Ikegami, K., Doebley, A.L., Goldstein, L.D., Mitchell, J.,
850 Sakaguchi, A., Miska, E.A., and Ahmed, S. (2014). Reduced Insulin/IGF-1 signaling
851 restores germ cell immortality to *Caenorhabditis elegans* Piwi mutants. Cell Rep.
852 *7*, 762–773.

853 Simonti, C.N., Pavličev, M., and Capra, J.A. (2017). Transposable Element
854 exaptation into regulatory regions is rare, influenced by evolutionary age, and
855 subject to pleiotropic constraints. Mol. Biol. Evol. *34*, 2856–2869.

856 Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small
857 RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246.

858 Skinner, D.E., Rinaldi, G., Koziol, U., Brehm, K., and Brindley, P.J. (2014). How
859 might flukes and tapeworms maintain genome integrity without a canonical
860 piRNA pathway? Trends Parasitol.

861 Smit, A., Hubley, R., and Green, P. (2017). RepeatMasker Open-4.0.6 2013-2015 .
862 Http://Www.Repeatmasker.Org.

863 Szitenberg, A., Cha, S., Opperman, C.H., Bird, D.M., Blaxter, M.L., and Lunt, D.H.
864 (2016). Genetic drift, not life history or RNAi, determine long-term evolution of
865 Transposable Elements. Genome Biol. Evol. *8*, 2964–2978.

866 Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice
867 junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

868 Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J.,
869 Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). How Cufflinks works. Nat.
870 Biotechnol. *28*, 511–515.

871 Wang, G., and Reinke, V. (2008). A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs
872 during spermatogenesis. Curr. Biol. *18*, 861–867.

873 Weick, E., and Miska, E. a (2014). piRNAs : from biogenesis to function. Genes
874 Dev. 1–41.

875 Woodhouse, R.M., Buchmann, G., Hoe, M., Harney, D.J., Low, J.K.K., Larance, M.,
876 Boag, P.R., and Ashe, A. (2018). Chromatin modifiers SET-25 and SET-32 Are
877 required for establishment but not long-term maintenance of transgenerational
878 epigenetic inheritance. Cell Rep. *25*, 2259-2272.e5.

879 Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.-C.G., Tolia, N.H., Joshua-Tor, L.,
880 Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the *C. elegans*

881 Argonaute family reveals that distinct Argonautes act sequentially during RNAi.
882 Cell *127*, 747–757.

883

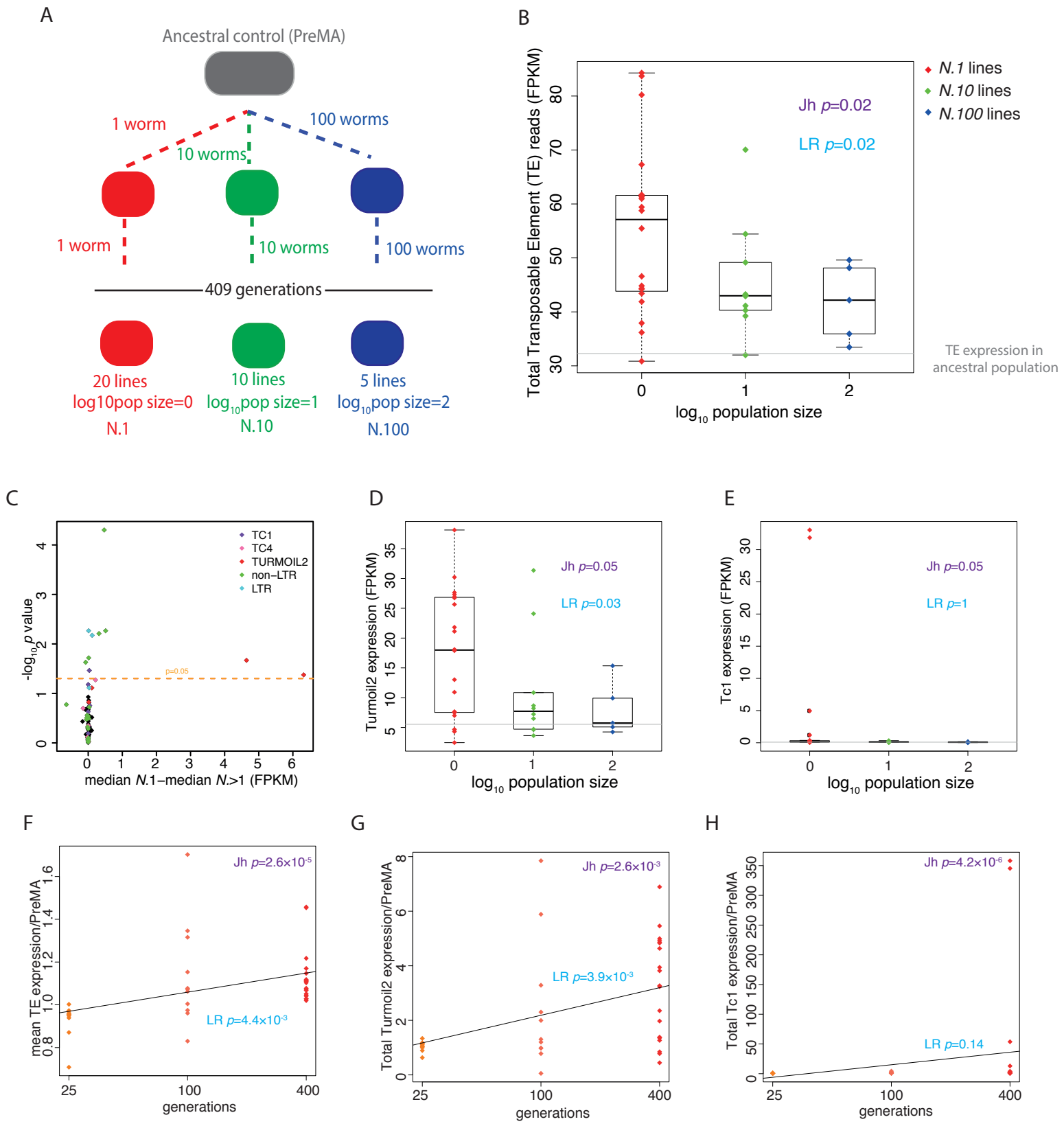

A
Ancestral control (PreMA)
1 worm
10 worms
100 worms
1 worm
10 worms
100 worms
409 generations
20 lines
log10pop size=0
N.1
10 lines
log10pop size=1
N.10
5 lines
log10pop size=2
N.100
B
Total Transposable Element (TE) reads (FPKM)
log10 population size
N.1 lines
N.10 lines
N.100 lines
Jh p=0.02
LR p=0.02
TE expression in ancestral population
C
-log10 p value
median N.1−median N.>1 (FPKM)
TC1
TC4
TURMOIL2
non−LTR
LTR
p=0.05
D
Turmoil2 expression (FPKM)
log10 population size
Jh p=0.05
LR p=0.03
E
Tc1 expression (FPKM)
log10 population size
Jh p=0.05
LR p=1
F
mean TE expression/PreMA
generations
Jh p=2.6×10-5
LR p=4.4×10-3
G
Total Turmoil2 expression/PreMA
generations
Jh p=2.6×10-3
LR p=3.9×10-3
H
Total Tc1 expression/PreMA
generations
Jh p=4.2×10-6
LR p=0.14

Figure 1

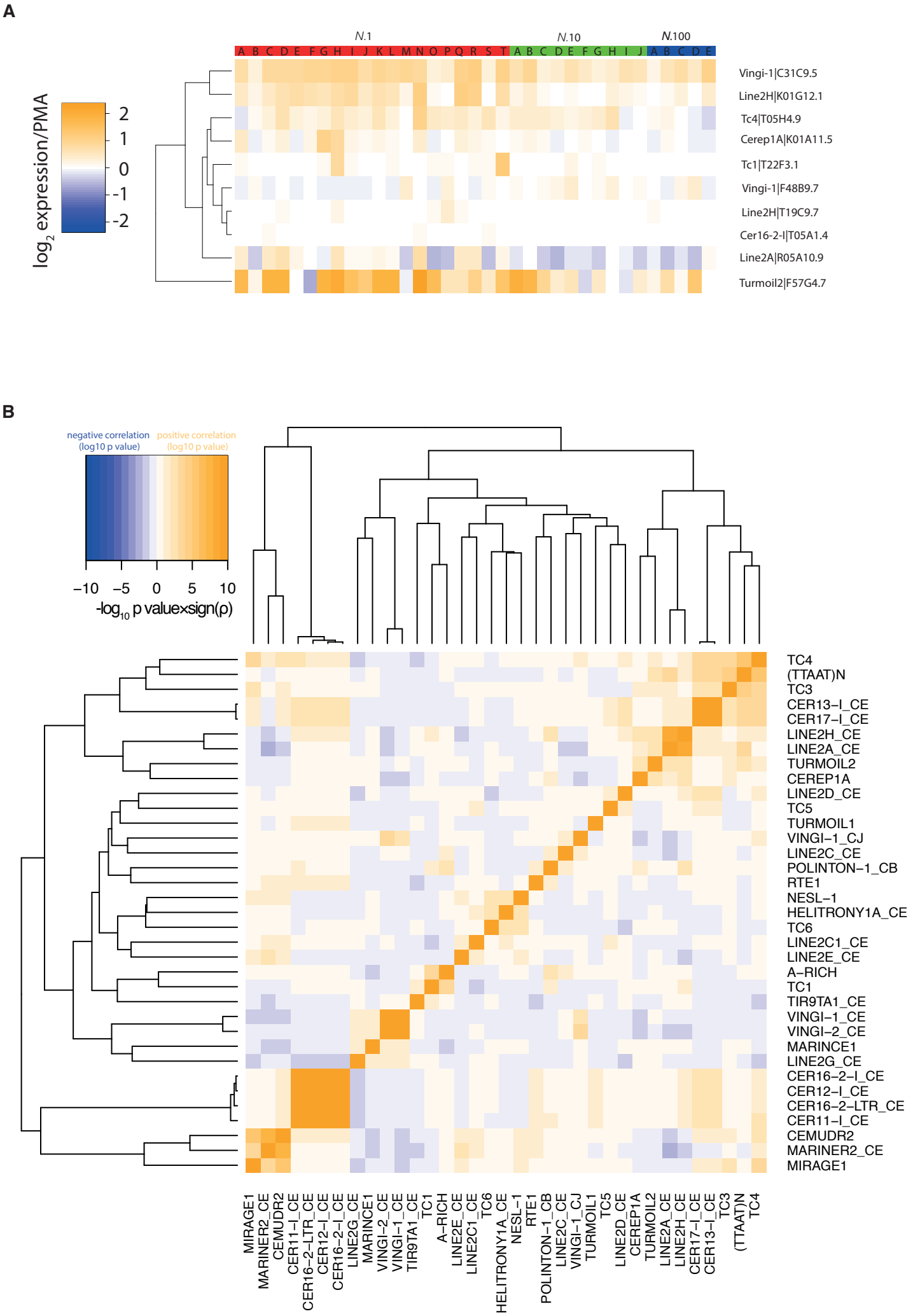

A
N.1
N.10
N.100
A B C D E F G H I J K L M N O P Q R S T A B C D E F G H I J A B C D E
log2 expression/PMA
2
1
0
-1
-2
Vingi-1|C31C9.5
Line2H|K01G12.1
Tc4|T05H4.9
Cerep1A|K01A11.5
Tc1|T22F3.1
Vingi-1|F48B9.7
Line2H|T19C9.7
Cer16-2-I|T05A1.4
Line2A|R05A10.9
Turmoil2|F57G4.7
B
negative correlation
(log10 p value)
positive correlation
(log10 p value)
−10 −5 0 5 10
-log10 p value×sign(ρ)
TC4
(TTAAT)N
TC3
CER13−I_CE
CER17−I_CE
LINE2H_CE
LINE2A_CE
TURMOIL2
CEREP1A
LINE2D_CE
TC5
TURMOIL1
VINGI−1_CJ
LINE2C_CE
POLINTON−1_CB
RTE1
NESL−1
HELITRONY1A_CE
TC6
LINE2C1_CE
LINE2E_CE
A−RICH
TC1
TIR9TA1_CE
VINGI−1_CE
VINGI−2_CE
MARINCE1
LINE2G_CE
CER16−2−I_CE
CER12−I_CE
CER16−2−LTR_CE
CER11−I_CE
CEMUDR2
MARINER2_CE
MIRAGE1
MIRAGE1
MARINER2_CE
CEMUDR2
CER11−I_CE
CER16−2−LTR_CE
CER12−I_CE
CER16−2−I_CE
LINE2G_CE
MARINCE1
VINGI−2_CE
VINGI−1_CE
TIR9TA1_CE
TC1
A−RICH
LINE2E_CE
LINE2C1_CE
TC6
HELITRONY1A_CE
NESL−1
RTE1
POLINTON−1_CB
LINE2C_CE
VINGI−1_CJ
TURMOIL1
TC5
LINE2D_CE
CEREP1A
TURMOIL2
LINE2A_CE
LINE2H_CE
CER17−I_CE
CER13−I_CE
TC3
(TTAAT)N
TC4

Figure 2

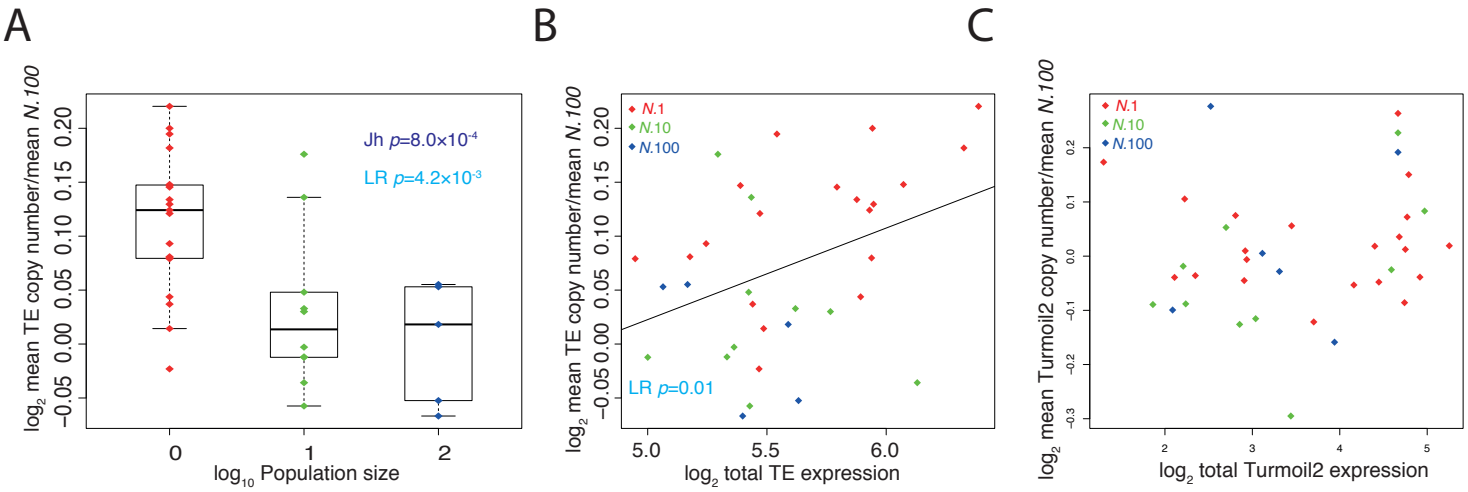

A
B
C
log2 mean TE copy number/mean N.100
Jh p=8.0×10-4
LR p=4.2×10-3
0
1
2
log10 Population size
N.1
N.10
N.100
LR p=0.01
5.0
5.5
6.0
log2 total TE expression
log2 mean Turmoil2 copy number/mean N.100
log2 total Turmoil2 expression

Figure 3

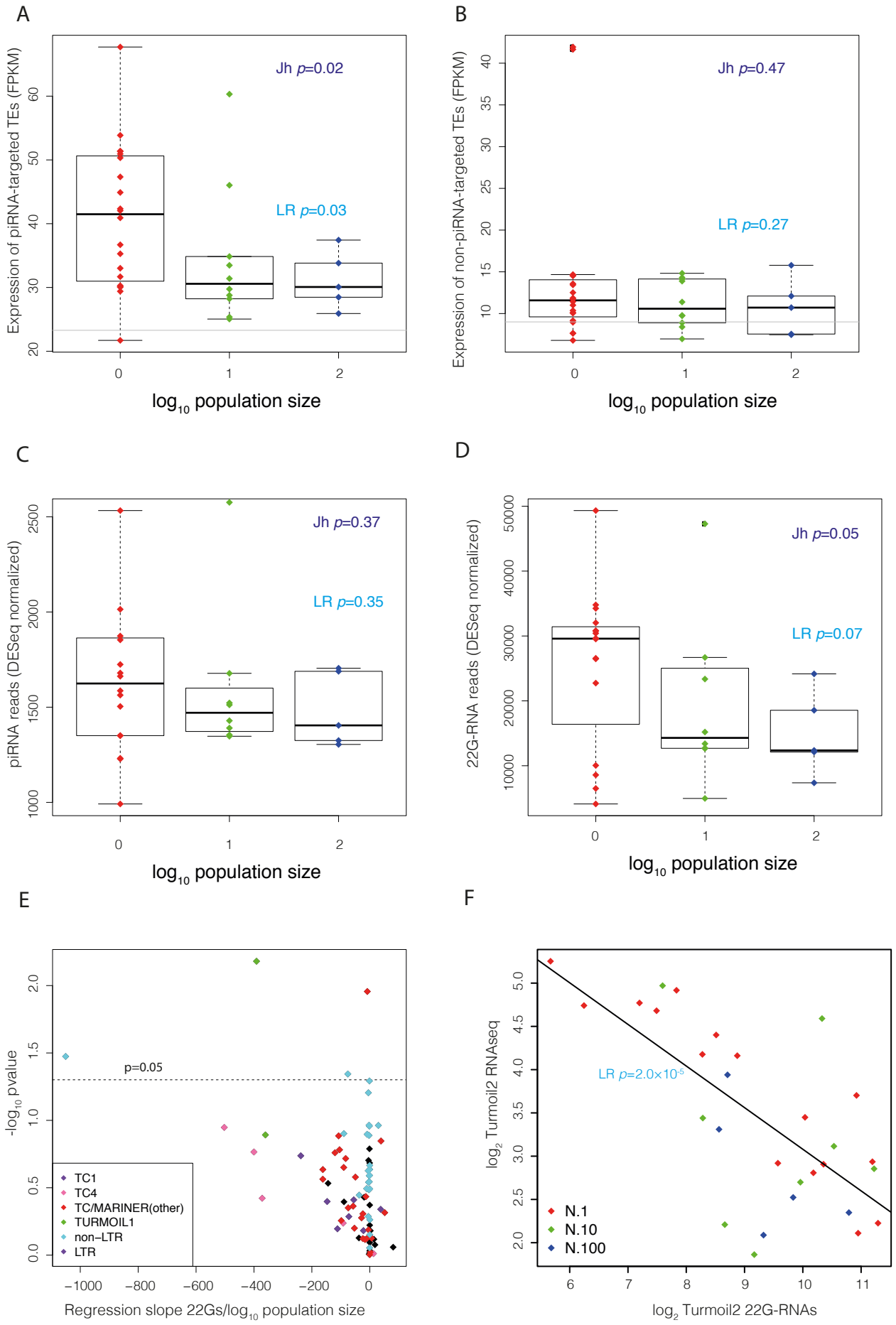

A
Expression of piRNA-targeted TEs (FPKM)
Jh p=0.02
LR p=0.03
log10 population size
B
Expression of non-piRNA-targeted TEs (FPKM)
Jh p=0.47
LR p=0.27
log10 population size
C
piRNA reads (DESeq normalized)
Jh p=0.37
LR p=0.35
log10 population size
D
22G-RNA reads (DESeq normalized)
Jh p=0.05
LR p=0.07
log10 population size
E
-log10 pvalue
p=0.05
TC1
TC4
TC/MARINER(other)
TURMOIL1
non−LTR
LTR
Regression slope 22Gs/log10 population size
F
log2 Turmoil2 RNAseq
LR p=2.0×10-5
N.1
N.10
N.100
log2 Turmoil2 22G-RNAs

Figure 4

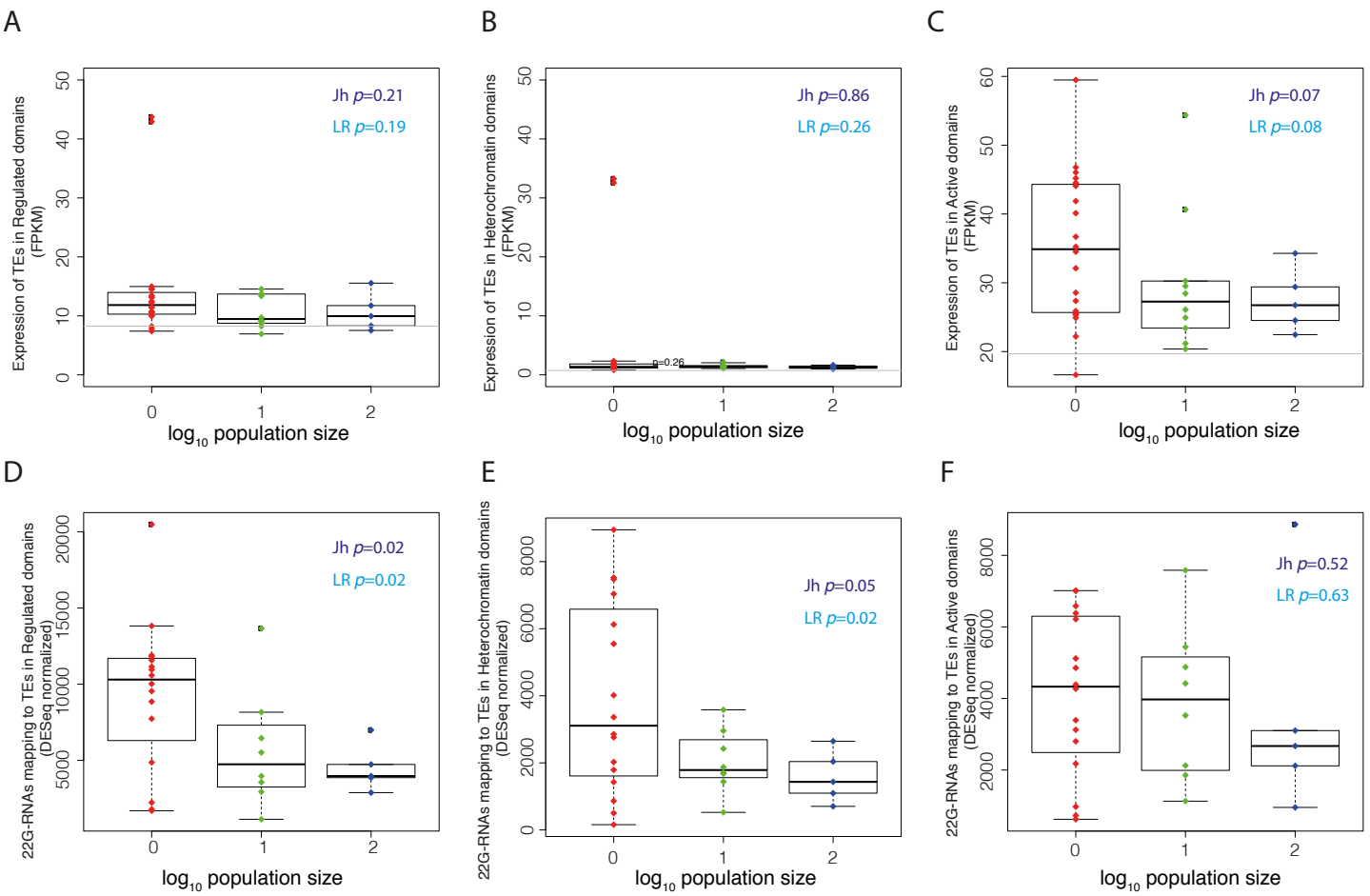

A
Expression of TEs in Regulated domains (FPKM)
Jh p=0.21
LR p=0.19
log10 population size
B
Expression of TEs in Heterochromatin domains (FPKM)
Jh p=0.86
LR p=0.26
p=0.26
log10 population size
C
Expression of TEs in Active domains (FPKM)
Jh p=0.07
LR p=0.08
log10 population size
D
22G-RNAs mapping to TEs in Regulated domains (DESeq normalized)
Jh p=0.02
LR p=0.02
log10 population size
E
22G-RNAs mapping to TEs in Heterochromatin domains (DESeq normalized)
Jh p=0.05
LR p=0.02
log10 population size
F
22G-RNAs mapping to TEs in Active domains (DESeq normalized)
Jh p=0.52
LR p=0.63
log10 population size

Figure 5

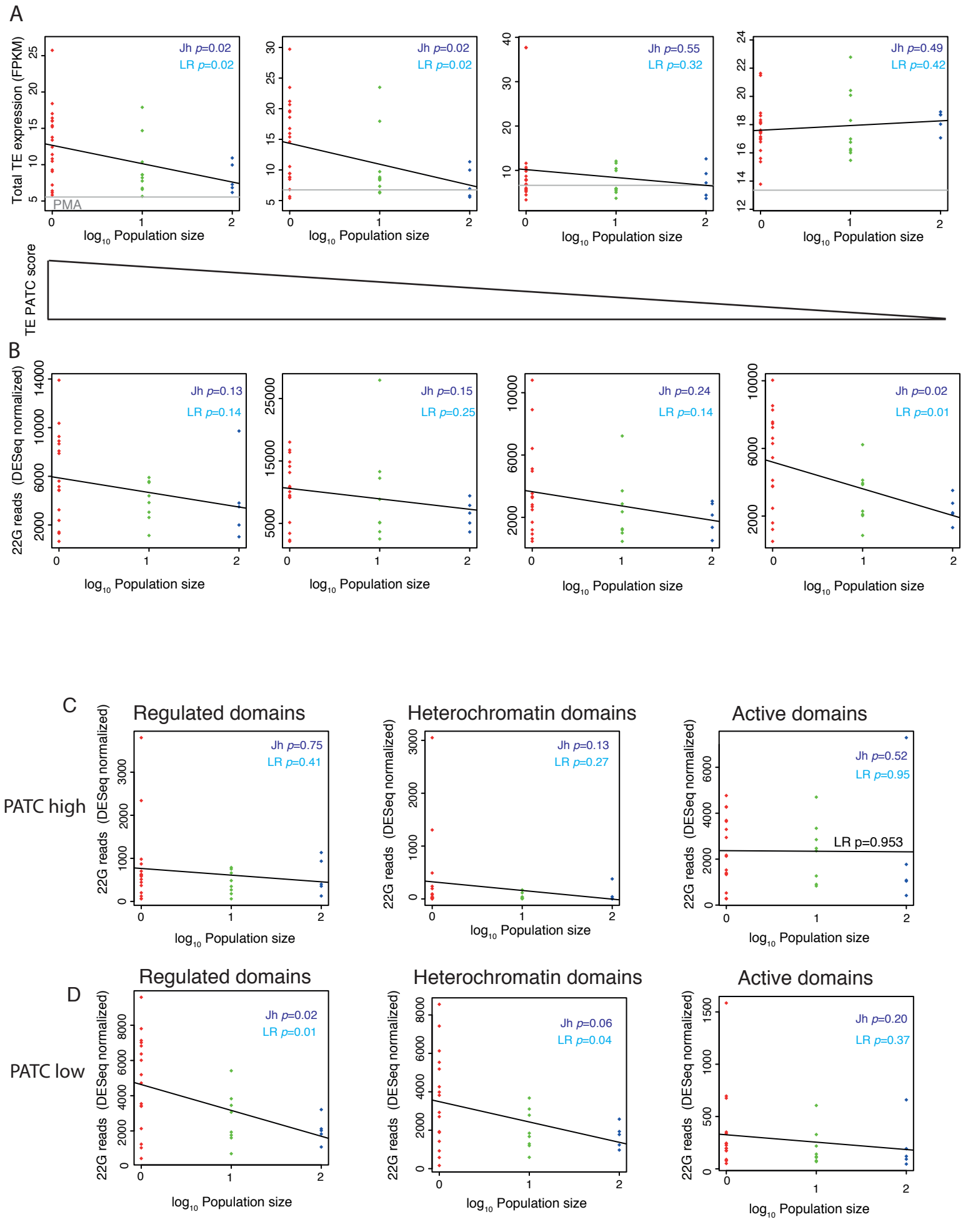

A
Total TE expression (FPKM)
Jh p=0.02
LR p=0.02
PMA
Jh p=0.02
LR p=0.02
Jh p=0.55
LR p=0.32
Jh p=0.49
LR p=0.42
log10 Population size
TE PATC score
B
22G reads (DESeq normalized)
Jh p=0.13
LR p=0.14
Jh p=0.15
LR p=0.25
Jh p=0.24
LR p=0.14
Jh p=0.02
LR p=0.01
C
PATC high
Regulated domains
Jh p=0.75
LR p=0.41
Heterochromatin domains
Jh p=0.13
LR p=0.27
Active domains
Jh p=0.52
LR p=0.95
LR p=0.953
D
PATC low
Regulated domains
Jh p=0.02
LR p=0.01
Heterochromatin domains
Jh p=0.06
LR p=0.04
Active domains
Jh p=0.20
LR p=0.37

Figure 6

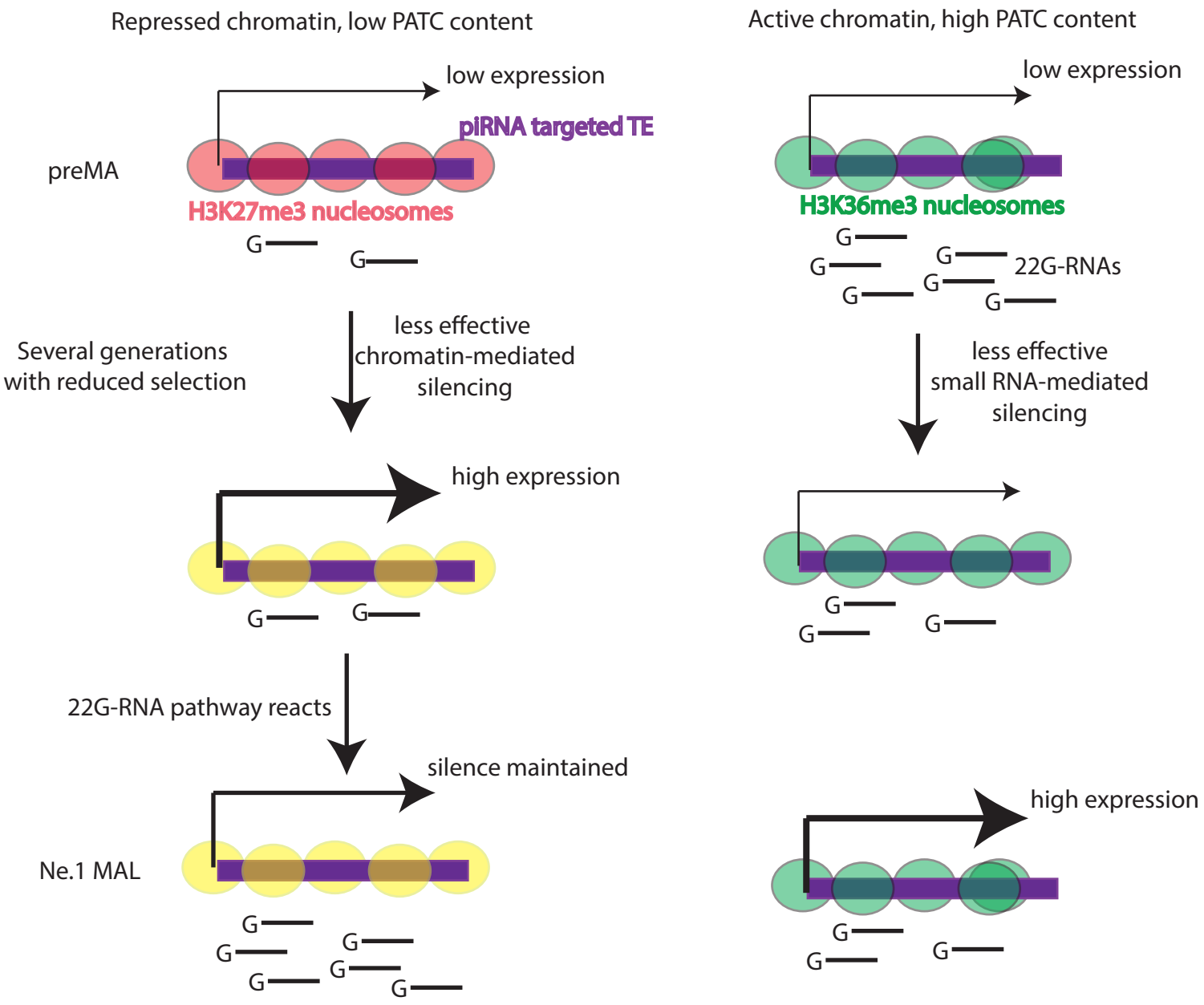
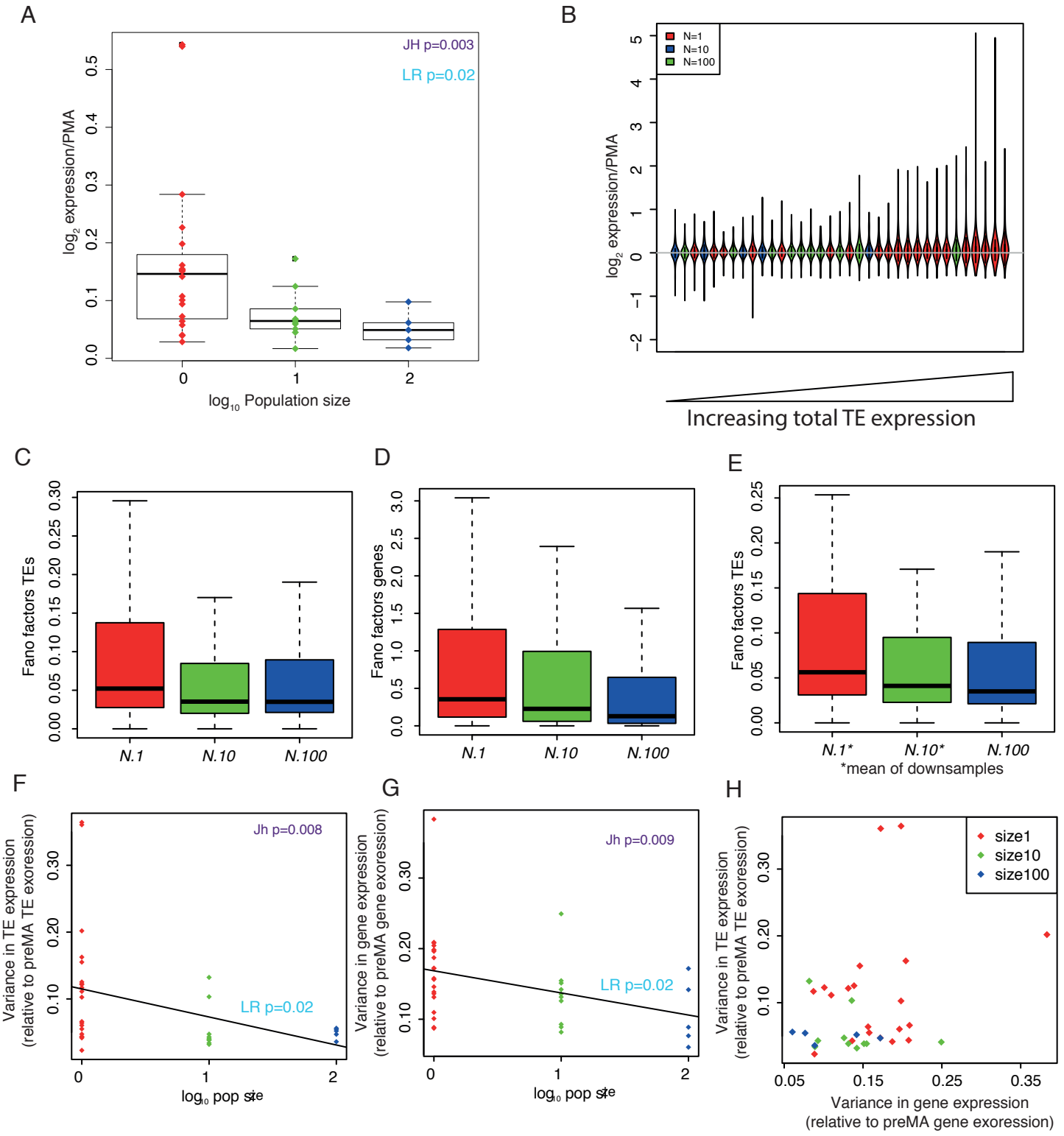
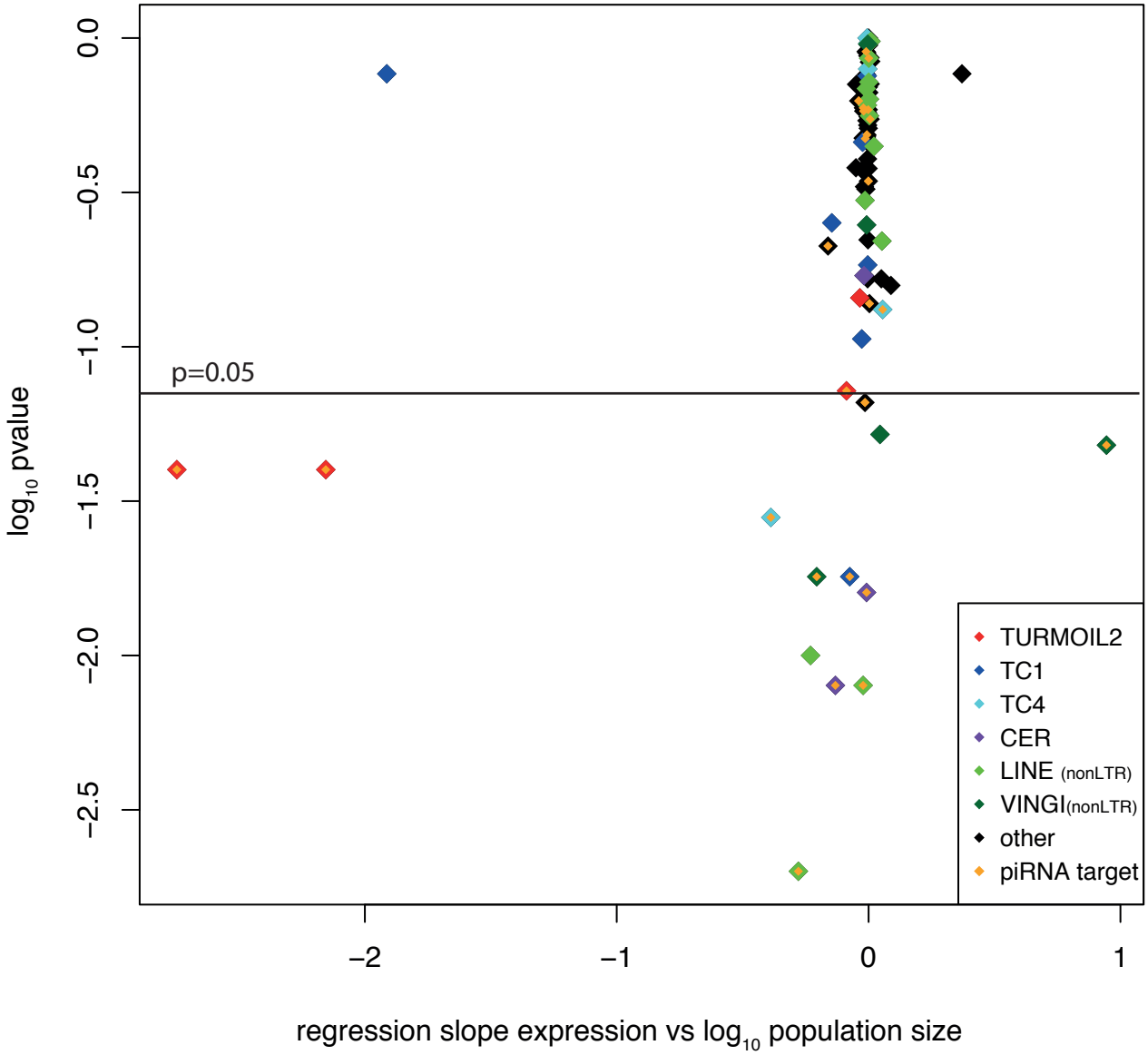

Repressed chromatin, low PATC content
Active chromatin, high PATC content
low expression
piRNA targeted TE
preMA
H3K27me3 nucleosomes
G
low expression
H3K36me3 nucleosomes
22G-RNAs
Several generations
with reduced selection
less effective
chromatin-mediated
silencing
less effective
small RNA-mediated
silencing
high expression
22G-RNA pathway reacts
silence maintained
Ne.1 MAL
high expression

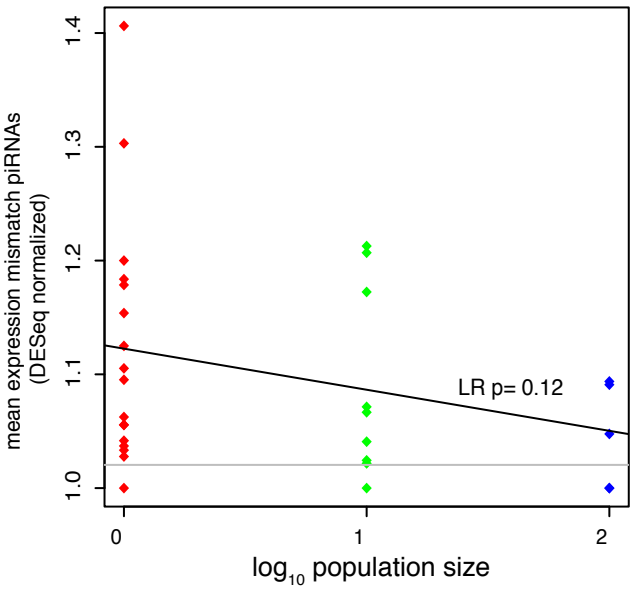
Figure 7



Supplemental Figure 1



Supplemental Figure 2



Supplemental Figure 3

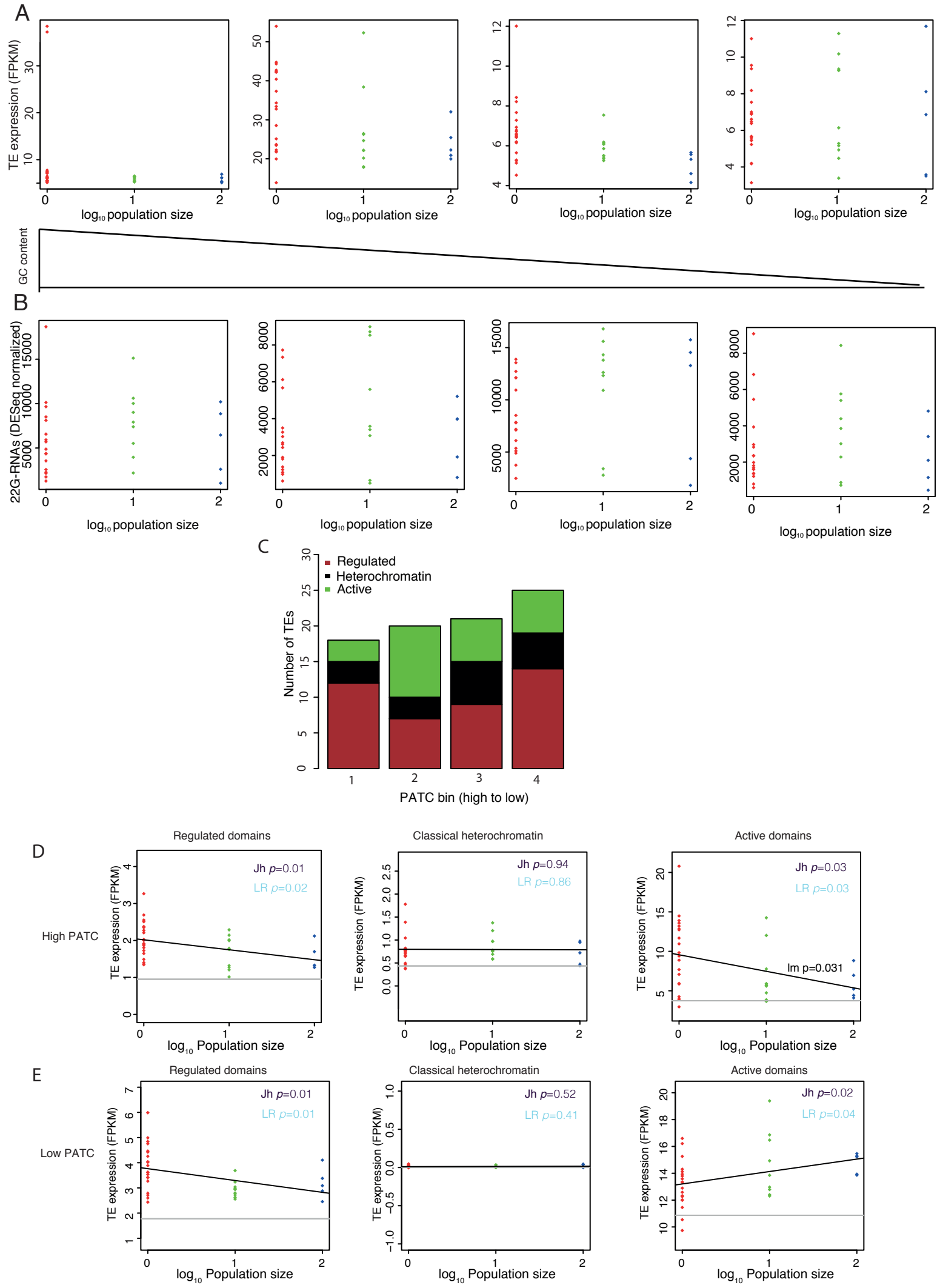

A
TE expression (FPKM)
log10 population size
GC content
B
22G-RNAs (DESeq normalized)
log10 population size
C
Regulated
Heterochromatin
Active
Number of TEs
PATC bin (high to low)
D
High PATC
Regulated domains
Classical heterochromatin
Active domains
Jh p=0.01
LR p=0.02
Jh p=0.94
LR p=0.86
Jh p=0.03
LR p=0.03
lm p=0.031
TE expression (FPKM)
log10 Population size
E
Low PATC
Regulated domains
Classical heterochromatin
Active domains
Jh p=0.01
LR p=0.01
Jh p=0.52
LR p=0.41
Jh p=0.02
LR p=0.04
TE expression (FPKM)
log10 Population size

Figure S4