Identification of the *Telomere elongation* mutation in *Drosophila*

Hemakumar M. Reddy [1, 2] , **Thomas A. Randall** [3] , **Radmila Capkova Frydrychova** [Corresp., 4, 5, 6] , **James M. Mason** [6, 7]

[1] Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, RTP, USA

[2] Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of Medicine, Gainesville, Florida, USA

[3] Integrative Bioinformatics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

[4] Institute of Entomology, Biology Centre AS CR, v.v.i., Ceske Budejovice, Czech Republic

[5] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic

[6] National Institute of Environmental Health Sciences, Laboratory of Molecular Genetics, Research Triangle Park, North Carolina, USA

[7] National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Corresponding Author: Radmila Capkova Frydrychova
Email address: radmila.frydrychova@hotmail.com

Background. Telomeres in *Drosophila melanogaster* are similar to those of other eukaryotes in terms of their function, although they are formed by non-LTR retrotransposons instead of telomerase-based short repeats. The length of the telomeres in Drosophila depends on the number of copies of these transposable elements. A dominant mutation, Tel1, causes a several-fold elongation of telomeres.

Methods. In this study we identified the Tel1 mutation by a combination of transposon-induced, site-specific recombination and next generation sequencing.

Results. Recombination located Tel1 to a 15 kb region in 92A. Comparison of the DNA sequence in this region with the Drosophila Genetic Reference Panel of wild type genomic sequences delimited Tel1 to a 3 bp deletion inside intron 8 of Ino80.

Discussion. The mapped Tel1 mutation (3-bp deletion found in Ino80) did not appear to affect the quantity or length of the Ino80 transcript. Tel1 causes a significant reduction in transcripts of CG18493, a gene nested in an intron 8 of Ino80, which is expressed in ovaries and expected to encode a serine-type peptidase.

1 **Identification of the *Telomere elongation* mutation in *Drosophila***

2 Hemakumar M. Reddy [1, 3], Thomas A. Randall [2], Radmila Capkova Frydrychova [1, 4,5] * and
3 James M. Mason [1,6]

4 [1] Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences,
5   Research Triangle Park, North Carolina, 27713;
6 [2] Integrative Bioinformatics, National Institute of Environmental Health Sciences, Research
7   Triangle Park, North Carolina, 27713; thomas.randall@nih.gov
8 [3] Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of
9   Medicine, Gainesville, Florida 32608; hemakumar.m@gmail.com
10 [4] Institute of Entomology, Biology Centre AS CR, v.v.i., Ceske Budejovice, Czech Republic;
11   Radmila.Frydrychova@hotmail.com
12 [5] Faculty of Science, University of South Bohemia, Ceske Budejovice, Czech Republic
13 [6] National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina,
14   27713; flymason312@gmail.com
15
16 ***** Correspondence: Radmila.Frydrychova@hotmail.com; Tel.: +420-387 775 322
17

18   **ABSTRACT**

19   **Background.** Telomeres in *Drosophila melanogaster* are similar to those of other eukaryotes in
20   terms of their function, although they are formed by non-LTR retrotransposons instead of
21   telomerase-based short repeats. The length of the telomeres in Drosophila depends on the
22   number of copies of these transposable elements. A dominant mutation, Tel1, causes a several-
23   fold elongation of telomeres.

24   **Methods**. In this study we identified the Tel1 mutation by a combination of transposon-induced,
25   site-specific recombination and next generation sequencing.

26   **Results.** Recombination located Tel1 to a 15 kb region in 92A. Comparison of the DNA
27   sequence in this region with the Drosophila Genetic Reference Panel of wild type genomic
28   sequences delimited Tel1 to a 3 bp deletion inside intron 8 of Ino80.

29   **Discussion.** The mapped Tel1 mutation (3-bp deletion found in Ino80) did not appear to affect
30   the quantity or length of the Ino80 transcript. Tel1 causes a significant reduction in transcripts of
31   CG18493, a gene nested in an intron 8 of Ino80, which is expressed in ovaries and expected to
32   encode a serine-type peptidase.

33

34   **Keywords:** Drosophila melanogaster; telomere; next generation sequencing; transposon
35   induced recombination
36

37

38

39

40

41

42

INTRODUCTION

Telomeres in all eukaryotes are functionally similar, although structural differences between some species exist (Frydrychova and Mason 2013). Linear chromosome ends are not replicated completely, and telomeres must compensate this end replication problem by adding new sequences at the chromosome end. The majority of eukaryotes use a specialized reverse transcriptase, telomerase, which adds a short, tandemly repeated DNA sequence to chromosome ends for telomere elongation (Sfeir and de Lange 2012). Insects in the order Diptera lack both telomerase and the short terminal repeats found in other organisms. The telomeres of Drosophila instead contain three families of non-long terminal repeat (LTR) retrotransposons, *HeT-A, TART* and *TAHRE* (jointly termed HTT), which transpose specifically to chromosome ends and attach using their 3' oligo(A) tails. Among these three families of elements *HeT-A* is most abundant, comprising as much as 80-90% of the total number of elements (Capkova Frydrychova et al. 2009). Telomeric chromatin consists of the HTT elements and different proteins that are bound to them (Andreyeva et al. 2005). The rate of transposition of these HTT elements may depend on an equilibrium between the level of their expression and the chromatin bound proteins (Takács et al. 2012; Silva-Sousa et al. 2013).

Telomere elongation may also be accomplished by terminal gene conversion using a neighboring telomere as a template (Roth et al. 1997; Kahn et al. 2000; Mason et al. 2008). The mechanism of telomere elongation by gene conversion has also been observed in yeast and in certain human cancers and immortalized mammalian cells, in which overall telomere length increases in the absence of telomerase activity. This mechanism is called alternative lengthening of telomeres (ALT) (Henson et al. 2002; Neumann and Reddel 2002). A recent study showed that this mechanism may also be found in normal mammalian somatic cells (Neumann et al. 2013).

The genes involved in telomere elongation and the mechanisms of elongation are not well studied in Drosophila. Two independent studies in Drosophila identified dominant factors *Tel* (Siriaco et al. 2002) and *E(tc)* (Melnikova and Georgiev 2002), which developed telomeres several fold longer than controls, to the extent that these differences can be observed microscopically in polytene chromosomes. Mutations in *Su(var)205*, which encodes the HP1a protein, and deficiencies for components of the Ku70-Ku80 complex are also dominant telomere elongation mutations. While *Su(var)205* mutants seem to increase both HTT transposition frequencies and terminal gene conversion, and Ku deficiencies increase gene conversion, the specific mechanisms by which telomere elongation occurs in these mutants are not understood (Capkova Frydrychova et al. 2008).

Early efforts to map *Tel* (Siriaco et al. 2002) allowed meiotic recombination between the *Tel¹* bearing chromosome and a multiply marked chromosome; the results located *Tel¹* at 69 on the genetic map, which translates roughly to 92 on the cytogenetic map. Meiotic mapping of *E(tc)* (Melnikova and Georgiev 2002) indicated that this gene is in the same vicinity. In the present study we took advantage of the observation that there is no meiotic recombination in Drosophila males. Thus, site specific genetic recombination induced by double strand breaks that result from the excision of DNA transposons can be identified (Chen et al. 1998). We used both *P* elements and *Minos* transposons to induce recombination, which allowed the localization of *Tel¹* to a 15 kb region in the middle of the right arm of chromosome 3 (3R) at 92A. Whole genome sequencing resulted in the identification of many single nucleotide polymorphisms (SNPs) and small insertion/deletion polymorphisms (indels) in the *Tel¹* bearing genome relative to the reference sequence. Comparison of the *Tel¹* genomic sequence with a collection of inbred

89   lines of the Drosophila Genetic Reference Panel (DGRP) (Mackay et al. 2012) eliminated all of
90   these SNPs and most of the indels, and mapped *Tel¹* to a 3 bp deletion (TGT) at 3R:15,221,791
91   in the middle of intron 8 of *Ino80*. Examination of the expression of neighboring genes in
92   oocytes shows that only *CG18943* activity is significantly altered in the *Tel¹* mutant.

93

MATERIAL & METHODS

Mapping by site-specific recombination

96   Transposon-induced male recombination was performed as per the mating scheme reported
97   earlier (Chen et al. 1998; Zhai et al. 2003). The chromosome carrying the *Tel¹* mutation was
98   marked with two mutations with eye color phenotypes, *st* and *ca*. A *st Tel ca* chromosome was
99   made heterozygous in males with a *P* element and the Δ2-3 transposase. Recombinant
100   chromosomes bearing either *st* or *ca* were collected and put into stocks. Stock generation zero
101   occurs at the time the stock was established, two generations after the homozygous recombinants
102   were obtained. These stocks were maintained for a further 12 generations, flies from
103   homozygous recombinants at generation zero, six, nine and twelve were collected and frozen for
104   DNA isolation in order to assay *HeT-A* copy number over time.
105   *Minos* element-induced male recombination mapping is the same as the *P* element-induced
106   male recombination procedure, except that the P{hsILMiT}2.4 transposase was used in place of
107   Δ2-3. The P{hsILMiT}2.4 transposase is under the control of a heat shock promoter, therefore
108   the larvae generated by the cross of heterozygous *st Tel ca*/*Minos* males, were exposed to heat
109   shock at 37⁰C in a water bath for 1 hr daily from day two to day six post egg laying (Metaxakis
110   et al. 2005).
111

Genome sequencing, mapping to reference and de novo assembly

113   DNA was isolated from approximately 30-40 adult flies by standard procedures of lysis,
114   phenol:chloroform extraction and ethanol precipitation. The DNA pellet was resuspended in TE
115   buffer (10 mM Tris, 0.1 mM EDTA, pH 7.8). DNA quality and concentration were estimated
116   using a Qubit dsDNA BR Assay Kit and measured by Qubit 2.0 Fluorometer (Life
117   Technologies), as per the manufacturer's protocol. Five micrograms of DNA was taken for
118   library preparation.
119   All Illumina genome sequence generated for this project can be found in BioProject
120   Accession PRJNA255315 at NCBI. Genome sequencing was done on an Illumina GA IIx
121   sequencer following standard protocols by the NIH Intramural Sequencing Center. For a detailed
122   step-wise protocol for library preparation and genome sequencing, see Supplementary methods.
123   For *Tel* 66,654,840 paired reads of 101 bp length were obtained, and for *y w* 76,757,736 reads
124   were obtained, representing a genome coverage of 48X and 55X, respectively. Genomic reads
125   for each strain were mapped to the *D. melanogaster* reference by two methods. First, all the
126   reads were imported into CLC Genomics Workbench 4.8 and mapped using the parameters Min
127   distance = 150, Max distance = 10,000 to the Genbank annotated chr3R of dm3 (Release 5 from
128   ftp.ncbi.nih.gov). Second, the same raw reads were mapped to the same reference sequence with
129   bwa 0.6.0 (Li and Durbin 2010) using default settings.

130   The genomic reads were also assembled *de novo* by two methods. First, in CLC Genomics
131   Workbench 4.8 using parameters of Min distance = 150, Max distance = 2000 which generated
132   the highest N50 for *Tel* (28.8 kb). Then in ABySS 1.2.3 (Simpson et al. 2009) a range of kmers
133   from 25-65 was tested using the *Tel* sequence, with a kmer setting of 45 generating the highest
134   N50 (45.3 kb). This setting was subsequently used for both genomic assemblies; all other ABySS
135   settings were default.

Variant detection

137   SNPs and indels (referred to as deletion insertion polymorphisms, DIPs, by CLC Genomics
138   Workbench) were identified from mapped reads in comparison to the reference genome by two
139   methods. First, using CLC Genomics Workbench 4.8 with the SNP and DIP Detection tools at
140   default settings. Separately, the bwa assemblies were imported into CLC as bam files, and both
141   SNP and indel detection were performed on these assemblies as above. As this software is not
142   trained for detecting large indels (>5 bp), we scanned a large mapped region of 79 kb (chr3R:
143   15,151,000 -15,230,000) manually and identified additional indels, which had not been detected
144   by the CLC software. To complement this SNP/indel analysis by CLC, the same assemblies
145   (CLC and bwa) were also analyzed for SNPs and indels using the pileup program of SAMtools
146   1.6 (Li et al. 2009).
147   Separately, the contigs spanning the 79 kb region of interest were extracted from each of the
148   *de novo* assemblies and were aligned to the corresponding reference region with MAFFT 6.849
149   (Neron et al. 2009) and manually inspected for indels.

Comparison to DGRP data

151   Files containing the SNPs identified in 162 DGRP (Mackay e2012) lines on chr3R were
152   downloaded (Freeze 1, August 2010 release; http://www.hgsc.bcm.tmc.edu/content/drosophila-
153   genetic-reference-panel) and the chromosomal coordinates of those 159 strains of normal
154   telomere length were compared with the SNPs identified in the *Tel* genome assembly (SNPs
155   identified by either CLC or pileup). Any SNP identified in the *Tel* genome that was also
156   identified in the SNP collection from DGRP was ruled out as possibly causing the Tel
157   phenotype.
158   As there was no indel data for DGRP lines in Freeze 1, each indel found in the *Tel* genome
159   assembly, as described above, was compared to the DGRP data. For a subset of eight DGRP
160   lines the Illumina fastq sequence was downloaded from SRA (SRP000694, Lines 40, 85, 177,
161   321, 352, 405, 426, 802), imported into CLC Genomics Workbench and assembled to the chr3R
162   reference as above, and indel detection was performed. Any indel identified in the *Tel* genome
163   and also found in one or more of these DGRP lines was ruled out as potentially causative. For
164   the indels discovered by manual inspection of both the *Tel* assembly to reference and the *Tel de*
165   *novo* assemblies a separate local *de novo* assembly strategy was used for comparison to a subset
166   of the DGRP population. Using 200-300 bp of reference sequence around a candidate indel as
167   bait, BLAT (Kent 2002) was used to identify individual reads covering this region from the fasta
168   sequence of a given DGRP line. These reads were then extracted, assembled and compared to
169   both the *Tel* and reference sequences. Any manually identified indel also found in one or more of
170   these DGRP lines was ruled out as a candidate mutation.

Real-time PCR

172　DNA was isolated from 20-30 flies by using DNeasy Blood & Tissue Kit (Qiagen) columns as
173　per the manufacturer's protocol. For large numbers of samples, DNA was isolated from 10 flies
174　of each line using Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) as per
175　the manufacturer's protocol. DNA isolation steps were handled by Biomek 4000 Liquid
176　Handling System (Beckman Coulter robotic system). DNA was eluted in 50 ul water. The DNA
177　concentration was estimated by using NanoDrop 2000 (Thermo Scientific) and diluted to a
178　concentration of 10 ng/µl, using sterile water.
179　Primers used for real-time PCR are:
180　　　RpS17-F: 5′AAGCGCATCTGCGAGGAG3′,
181　　　RpS17-R: 5′CCTCCTCCTGCAACTTGATG3′,
182　　　HeT-9D4GAG-ORF-F: 5′TTGTCTTCTCCTCCGTCCACC3′,
183　　　HeT-9D4GAG-ORF-R: 5′GAGCTGAGATTTTTCTCTATGCTACTG3′,

184　Predicted sizes of amplicons are 195 bp for RpS17, 152 bp for the HeT-9 D4 GAG-ORF.
185　GenBank accession number for *HeT-A* element 9D4 is X68130 and for RpS17 is M22142 (Török
186　et al. 2007; Wei et al. 2017). An aliquot of 20 ng of each DNA sample was taken for quantitative
187　PCR using 50 nM of each primer and 5 µl of 2X Power SYBR green PCR Master Mix (Applied
188　Biosystems) in a 10 µl reaction volume. These samples were amplified under the following
189　conditions: 95^0 C for 10 min (polymerase activation), followed by 40 cycles containing
190　denaturation at 95^0 C for 15 sec, and annealing/extension at 60^0 C for 1 min. Real-time PCR was
191　run using ABI Prism 7900 HT Sequence detection system (Applied Biosystems)
192　　　Competitive threshold (Ct) values for each sample were collected for *HeT-A* primers (9D4
193　element GAG ORF) and for control Rps17 (ribosomal protein17) primers. Delta Ct values for
194　each sample were calculated by normalizing *HeT-A* Ct values to control Ct values and graphed
195　using Microsoft Excel. Each DNA sample was run in triplicates to estimate average Ct values.

196　**RNA isolation and cDNA synthesis**

197　RNA was isolated by using RNeasy plus Mini kit (Qiagen) as per the manufacturer's protocol.
198　Ovaries from 2-3 day old adult females were removed and placed in lysis buffer (RLT buffer)
199　containing 1% β-mercaptoethanol. While dissecting the ovaries the tubes were kept in dry ice.
200　Immediately after dissection ovaries in RLT buffer were stored at -80^0 C until RNA was isolated.
201　Samples were passed through gDNA eliminator mini spin columns (Qiagen) after lysis to
202　eliminate DNA contamination. RNA samples were eluted in sterile DEPC-treated RNase free
203　water. RNA concentrations were estimated using a NanoDrop 2000 (Thermo Scientific). cDNA
204　was synthesized from 1 µg of RNA using Superscript III First-strand synthesis system for RT-
205　PCR (Life Technologies) as per the manufacturer's protocol. Twenty µl of cDNA was diluted
206　five-fold with water, and 2 µl of each diluted cDNA was taken for quantitative real time PCR to
207　measure the level of transcript expression. Each Q-RT-PCR was repeated three times, with
208　triplicate reactions at each time, with each set of gene primers. Mean and standard deviation
209　values of the three replicates were used to estimate expression levels. Significance of differential
210　expression was determined by an unpaired two-tailed t test.
211　Primers used for transcript expression analysis are:
212　　　Ino80 (CG31212): 169 bp
213　　　For: 5'- TGCCGAAGATGAGGACGAAGTAG -3'
214　　　Rev: 5'- AAAGAAGGATGTGGAGAACGAGC -3'
215　　　CG3581: 107 bp

216   For: 5'- TGAAAAAGGCACAGTGGAGGG -3'
217   Rev: 5'- GAACTTGTCAACACAGGGTATGGG -3'
218   CG31404: 112 bp
219   For: 5'- GCTTTTGGAACACACTTGGGC -3'
220   Rev: 5'- GCACCTTGGGTTTTACGAACAATG -3'
221   CG31245: 126 bp
222   For: 5'- AAAAGCCAGGACCGATGAGC -3'
223   Rev: 5'- CAACAATGGGTTTGCTATCTCGC -3'
224   CR34285: 171 bp
225   For: 5'- ACAGTTCCTCAAGCAATGGCAG -3'
226   Rev: 5'- GGAACCCCAGCCCAAATCAATC -3'
227   CG3734: 106 bp
228   For: 5'- AGGCATTTCTTTGGCTCTGTTG -3'
229   Rev: 5'- ATTCTGAGTGGTTGGCAGTGGTGG -3'
230   CG18493: 230 bp
231   For: 5'- GTGAAAATGGCTGCTCTGCG -3'
232   Rev: 5'- TGATAGGTTTGCGTGTTGCTCG -3'
233   CG3739: 131 bp
234   For: 5'- GAATGCGTTCGTCAAATCTCTGAG -3'
235   Rev: 5'- TGGCGTTGTTGCTGTCATCG -3'
236   CG31244: 166 bp
237   For: 5'- TGGACAGGACCAAGACTACGAAAC -3'
238   Rev: 5'- CGCTTTACGCTCGGTATCTCTG -3'
239   Ino80_Exon8-9: 148 bp
240   For: 5'-TGGGAATTAATTTAACAGCCGCC-3'
241   Rev: 5'-ATGGTTCCTTTGCAAATCAGTCG-3'

242

RESULTS

Transposase induced male recombination mapping

245 As there is no meiotic recombination in Drosophila males, it is possible to identify site-specific
246 recombination events generated by transposable elements that transpose by a cut-and-paste
247 mechanism and leave a double strand DNA break in their wake (Chen et al. 1998; Zhai et al.
248 2003). In general, our procedure was similar to previous work (Chen et al. 1998; Zhai et al.
249 2003); in particular it involved generating heterozygous *st Tel ca*/transposon males, inducing
250 transposition with an exogenous transposase and recovering recombinant *st* or *ca* chromosomes
251 to be tested for the Tel phenotype. Initial efforts to map the *Tel[1]* mutation by male recombination
252 were limited by the paucity of useful transposon insertions in the surrounding chromosomal
253 region and continued as new insertion chromosomes became available. The assay used to
254 identify *Tel[1]* on the recombinant chromosomes changed over time. The assay used in early
255 recombination experiments (Table 1, Round 1) used a cytogenetic analysis of heterozygous
256 chromosomes exposed to *Tel* for two years (Siriaco et al. 2002). Later, relative telomere length
257 was estimated by measuring relative copy number of the open reading frame (ORF) of *HeT-A* at

258 zero, six, nine, and twelve generations after a recombinant stock was established (Török et al.
259 2007; Wei et al. 2017).
260   The initial round of mapping, using seven *P* element insertions lying in the 91-93
261 cytogenetic region (Table 1), showed that *Tel¹* mapped between the *P{PZ}Dl⁰⁵¹⁵¹* and *P{SUPor-*
262 *P}CG16718^{KG06218}* transposon insertion sites (Hereafter we refer to the transposon insertions
263 simply with their allele designation; full names are listed in Table 1). The physical location is
264 3R:15,151,940 to 15,467,496, a region of 316 kb. This region showed a surprising paucity of *P*
265 element insertions.
266   As new *P* element insertions became available, we used three transposons lying within this
267 316 kb region (Table 1, Round 2; Figure 1A). All three *st*-bearing recombinant chromosomes
268 generated using *P* element insertion *d10097* showed telomere elongation from generation 0 to
269 12, whereas the three *ca*-bearing recombinant chromosomes from the same *P* element did not
270 show significant telomere elongation (Figure 1B). Thus, *Tel¹* lies to the left of this *P* element
271 insertion (3R:15,229,135). Similar results were obtained for recombinants from *P* element
272 insertions *EY10678* and *d03320* (see Figure S1), both of which are to the right of *d10097* (Figure
273 1A). These results mapped *Tel* to 3R: 15,151,940 to 15,229,135, a region of ~77 kb.
274 *Minos* elements were also used to induce recombination (Table 1), although *Minos* elements had
275 not previously been shown to induce recombination in males. Two *Minos* insertions, *MB02141*
276 and *MB0163*, lying to the right of *d10097* in the 316 kb region showed similar results (see Figure
277 S1), indicating that they are situated to the right of *Tel¹* as expected. Two *Minos* insertions in the
278 77 kb region were selected for further mapping studies (Table 1, Round 3; Figure 1). Even after
279 heat shock, these *Minos* elements generated only a few recombinant males. We obtained only
280 one *st*-bearing recombinant chromosome from each of these *Minos* transposons (Table 1). The *st*
281 recombinants for *MI03112* and *MI02316* showed no evidence of telomere elongation from
282 generation zero through twelve (Figure 1B). This result eliminated the region to the left of
283 *MI02316* as containing *Tel¹* and mapped *Tel¹* to a 15 kb region (3R:15,214,101 to 15,229,135).

Telomere length in transposon insertion lines

285 We measured relative *HeT-A* copy number in the transposon lines used to induce site-specific
286 recombination. Q-PCR analysis showed that all of the lines, except one bearing *MB09416*, had
287 telomeres comparable in length to the Oregon R control (Figure 2). The relative *HeT-A* copy
288 number in the *MB09416* insertion stock was highly elevated and similar to that of *Tel¹*. The
289 *MB09416 Minos* element is inserted at 3R:15,224,448, which is in intron 8 of *Ino80*, and within
290 the 15 kb *Tel¹* region identified above. As it is likely that the high *HeT-A* copy number in this
291 line might interfere with the ability to observe an increase in telomere length, this transposon was
292 not used in the mapping of *Tel¹*. It is also possible that genome of the *MB09416* line carries a
293 genetic factor, either at the insertion site or elsewhere, that has a phenotype similar to that of *Tel¹*
294 and might therefore confound the analysis.

Effect of Tel copy number on telomere length

296 Different deficiencies and duplications spanning region 92A3 (Figure 3A), where *Tel* was
297 mapped, were analyzed for an effect on telomere length by measuring relative *HeT-A* copy
298 number by Q-PCR in the respective stocks. None of the deficiency and duplication stocks
299 showed longer telomeres (Figure 3B). Thus, it appears that neither a 50% increase nor a 50%
300 decrease in copy number of the region around *Tel* had an effect on telomere length.

SNP and indel identification

The genomes of the three strains *Tel, y¹ w¹* and *E(tc)* were sequenced using the Illumina GAIIx platform. As *Tel¹* appeared in a wild-caught strain that had not been outbred to any laboratory flies (Siriaco et al. 2002), there was no useful wild type control. *E(tc)*, however, appeared in a *y¹ w¹* laboratory stock (Melnikova and Georgiev 2002); we therefore used a *y¹ w¹* stock as a wild type control. Upon sequencing, the *E(tc)* genome appeared to be highly heterozygous. We therefore assumed the stock was contaminated and did not pursue it further. The other two stocks were sequenced to 48X and 55X coverage, respectively. Both genomes were assembled to reference using CLC Genomics Workbench and bwa. Also, two *de novo* assemblies, using CLC Genomics Workbench and ABySS, were generated. Concurrently with the *Minos* transposon recombination mapping, a 79 kb genomic region of 3R: 15,151,000 to 15,230,000, roughly the region between inserts *05151* and *d10097,* extended slightly on either end, was analyzed for SNP and indel variations with the above assemblies (Figure S2). Within this region, there were 626 SNPs and 88 indels identified on the *Tel¹* chromosome compared to the reference using the CLC Genomics SNP and DIP Detection analysis (Table 2). A similar number of variations (586 SNPs and 80 indels) were also found in the *y¹ w¹* control strain. After eliminating common variations between *Tel* and *y¹ w¹* in this region, we are left with 332 SNPs and 53 indels that appear to be unique to the *Tel¹* genome in this 79 kb region.

Current variant calling tools are only proficient at defining small indels (1-5 bp) (Krawitz et al. 2010; O'Rawe et al. 2013). To detect larger indels, we scanned *Tel¹* genomic assemblies manually and detected 13 polymorphisms of 5 bp or larger (Table 2). These large indels, present in the *Tel¹* genome but not in *y¹ w¹*, were analyzed by PCR with primers flanking these indels and by Sanger sequencing of PCR products (Supplementary results, see Figure S4). A limitation of the assembly to reference strategy for variant identification is that potential novel insertions not present in the reference sequence will not be detected by this approach. To search for such variations, we aligned the *de novo* assemblies of the *Tel¹* genome to the reference. Manually scanning this alignment identified an additional 14 insertions not found by the above methods (Table 2).

Comparison of variations to DGRP data

To differentiate natural polymorphisms among these remaining SNP and indel variations found in the *Tel¹* genome, we compared them with the genomes available from the DGRP (Mackay et al. 2012), a collection of wild-caught, inbred Drosophila strains whose genomes have been sequenced. Our hypothesis is that if any variant found in the *Tel¹* genome is also found in a DGRP line having normal length telomeres that variant can be ruled out as causing the Tel phenotype. As a first step, all DGRP lines were tested for relative *HeT-A* copy number as a proxy for telomere length. The *HeT-A* copy number data for these strains fit a log-normal distribution, with three outliers that had copy numbers higher than the *Tel¹* strain (Figure 4). These three lines, RAL-161, -703 and -882, have been excluded from the following discussion and described elsewhere (Wei et al. 2017).

As the remaining DGRP lines have what we consider to be normal telomere length, close to that found in the Oregon R control, the SNPs identified from Freeze1 (August 2010) of the DGRP lines (Mackay et al. 2012) were compared to the SNPs in the *Tel¹* genome identified by the CLC SNP detection software. All the SNPs found in the 79 kb region around *Tel¹* were also found in the DGRP lines (Table 2). Thus, all the SNPs found in the *Tel¹* genome are natural polymorphisms with little expected effect on telomere elongation. No indel data were available

346   for DGRP lines in Freeze1. We therefore identified indels in a selection of eight DGRP lines for
347   the 79 kb region of interest. All except two indels found on the *Tel^1* chromosome, a deletion of C
348   at 3R:15,162,997 and a deletion of TGT at 3R:15,221,789-91, were also found in one or more
349   lines from the DGRP collection (Table 2). The deletion of C is located in a large intergenic
350   region, 11 kb to the right of *Dl* and more than 16 kb to the left of *CG43203*, while the deletion of
351   TGT is located in intron 8 of *Ino80*. The latter is the only indel specific to the *Tel^1* genome in the
352   15 kb region of interest and therefore was identified as the *Tel^1* mutation.

353   **Comparison to modENCODE data**

354   RNA sequence coverage for the 15 kb *Tel* region was analyzed by comparison with the
355   modENCODE database, including stage and tissue specific transcript expression levels. This
356   analysis shows that the candidate *Tel^1* mutation was not included in a transcript at any stage in
357   any tissue (Figure 5) and suggests that *Tel^1* could be acting to alter the expression of other
358   transcripts near or within the *Ino80* locus. The UCSC Genome browser map for this 15 kb region
359   was examined for sequence conservation among Drosophila species and other insect species.
360   This analysis showed that, even though the candidate *Tel^1* mutation is noncoding, it is in a well-
361   conserved region, similar in the level of conservation to neighboring coding regions (see Figure
362   S3). The transposon insertion site of *MB09416*, which is carried in a strain with elongated
363   telomeres, however is not conserved (see Figure S3).

364   **Transcript analysis**

365   The transcript levels from the ovaries of *Tel^1* and Oregon R were analyzed for 9 genes in the
366   vicinity of the *Tel^1* mutation (TGT deletion), which include *Ino80* and those found within its
367   introns. Quantitative PCR with cDNA from these two lines showed that there is no significant
368   difference (p > 0.05) in the expression levels between the two strains for most of these genes
369   (Figure 6A). The transcript level of *Ino80* around the exon 8-9 junction that spans the intron 8,
370   where *Tel^1* mutation is located, also showed an intact transcript with normal expression similar to
371   other parts of the transcript, indicating that the *Tel^1* mutation does not interfere with local
372   splicing (Figure 6A). *CG18493*, however, showed a 15 fold lower expression in *Tel^1* compared
373   to the control (p = 0.0006) and *CG3734* showed a slight reduction of expression in *Tel^1* ovaries
374   (p = 0.0103). Given that a 50% reduction in *Tel^+* copy number appears to have no effect on
375   telomere length (Figure 3), and that after a Bonferroni correction the small decrease seen for
376   *CG3734* expression (Figure 6A) may not be considered significant, it seems likely that
377   expression of this gene is not relevant for the Tel phenotype. To analyze further the significance
378   of the reduction in the *CG18493* transcript, we repeated this comparison with more control lines
379   [Oregon R, Canton S and DGRP line RAL-513, which has a *HeT-A* copy number near the mean
380   of the DGRP distribution (Figure 4)]. The long-telomere line *MB09416* was also examined. This
381   comparison confirmed the differences seen in the expression levels of *CG18493* between *Tel^1*
382   and Oregon R and extended it to the two other genetically unrelated controls (Figure 6B). Thus,
383   the *Tel^1* mutation may affect expression of *CG18493* but not any of the other neighboring genes.
384   Further, the *MB09416* insertion strain exhibits significantly reduced *CG18493* expression (p <
385   0.05) compared with all three controls. Thus, both *Tel^1* and the *MB09416* insertion are associated
386   with a significant reduction in expression of *CG18493*, and based on their proximity to each
387   other may cause this reduction by similar mechanisms.

388   **DISCUSSION**

Mechanism of telomere elongation

The stability of telomeres in Drosophila depends on terminin and non-terminin telomeric proteins. The terminin proteins Moi, Ver, HipHop, and HOAP are found only at telomeres, whereas non-terminin proteins HP1, the ATM and ATR kinases, and the proteins of MRN complex also have biological roles apart from their involvement in telomere maintenance and structure (Mason et al. 2008; Raffa et al. 2013). Mutations in any of the genes encoding these proteins cause telomere fusions and abnormal cell divisions. However, only mutations in HP1 (Savitsky et al. 2002) and the Ku70/Ku80 complex (Metaxakis et al. 2005) are associated with telomere elongation. The exact mechanism and the genes involved in telomere length homeostasis in Drosophila are largely unknown. There are reports of RNAi control over *HeT-A*, *TART* and *TAHRE* transcript levels (Savitsky et al. 2006), but the exact mechanism of its involvement in telomere length homeostasis is not well characterized. Two dominant mutations *Tel* and *E(tc)* showed telomere elongation (Melnikova and Georgiev 2002; Siriaco et al. 2002). One study reported that the *E(tc)* mutation is associated with elevated rates of gene conversion in telomeric regions; whereas the *Tel* mutation was associated with both transposition of the telomeric retrotransposons and gene conversion (Proskuryakov and Melnikova 2008), but not with the transcription of telomeric retrotransposons (Pineyro et al. 2011). The data thus suggest the involvement of *Tel* in a recombination pathway. The *Tel* mutation has no known phenotype other than telomere elongation (Siriaco et al. 2002; Walter et al. 2007). The genetic mapping of this mutation and identification of a candidate *Tel* gene will give us insight into telomere length homeostasis in Drosophila and possibly the recombination-mediated telomere maintenance mechanisms, such as the ALT pathway found in some human cancer cells. Although the ALT pathway is suggested for some human cancers, the molecular details remain still unknown (Nabetani and Ishikawa 2011).

Mapping Tel using transposon-induced male recombination

The *Tel[1]* mutation in Drosophila was previously localized by meiotic recombination to 69 on the genetic map (Siriaco et al. 2002). Meiotic recombination in Drosophila occurs only in females, but it is possible to induce site-specific recombination in males using transposons and use this for mapping (Chen et al. 1998; Ryder and Russell 2003). A collection of more than 15,000 publicly available *P* element insertions means that in many regions a resolution of 5-10 kb is possible for *P* element-induced recombination (Spradling et al. 1995; Bellen et al. 2004).

One major drawback of the *P* element, however, is its strong bias for insertion into some genes (hot spots) and against insertion into others (cold spots). The region around *Tel* is a cold spot for *P* element insertions. *Minos*, a member of the *Tc1/Mariner* family of transposable elements, is active in diverse organisms and cultured cells; it produces stable integrants in the germ line of several insect species, in the mouse, and in human cells (Metaxakis et al. 2005). To expand the usefulness of transposon mapping in Drosophila, collections of other transposable elements with different insertional specificities, such as *Minos* (Franz and Savakis 1991; Loukeris et al. 1995), have been introduced (Bellen et al. 2004). A recent analysis of *Minos* elements found a generally uniform distribution in the genome (Venken et al. 2011). We used available *Minos* elements to refine our mapping of the *Tel[1]* mutation and show for the first time that these transposons can induce recombination events useful for this purpose. This approach localized *Tel[1]* to a region of 15 kb.

Genome sequencing and DGRP resources for Tel[1] mapping

433 The molecular lesion associated with *Tel[1]* was identified by deep sequencing of the *Tel[1]* genome
434 and analyzing this sequence for novel SNP and indel variants not found in the DGRP lines
435 (Mackay et al. 2012). After comparing the variants in the genome bearing *Tel[1]* with DGRP
436 polymorphisms we ruled out all SNPs and all but one indel in intron 8 of *Ino80* as candidates for
437 *Tel[1]*. Thus, the combination of formal genetics and next generation sequencing resulted in the
438 identification of the molecular defect in *Tel* as a 3 bp deletion (TGT) at 3R:15,221,789-91.

439 To our knowledge, this is the first study using the DGRP collection to map a Mendelian trait
440 in *D. melanogaster*. The *Tel[1]* mutant used in this study was caught near Gaiano, Bergamo in
441 northern Italy, likely prior to 1946 (Siriaco et al. 2002). It is of interest that the natural genetic
442 diversity captured by DGRP in a Raleigh, North Carolina population was of sufficient diversity
443 to identify all of the SNPs and all but two indels within our 79 kb region defined by the
444 transposon mapping. This suggests that the DGRP will be an important general resource for
445 genetic mapping of genes in *Drosophila melanogaster*, even from strains not closely related to
446 the standard reference isolates.

Transposon insertion MI{ET1}Ino80[MB09416] shows elongated telomeres

448 Flies carrying the *Minos* transposon *MB09416* show long telomeres similar to those of the *Tel[1]*
449 mutant strain. This insertion site is 2660 bp to the right of the TGT deletion in the same intron of
450 *Ino80*. However, flies bearing nearby *Minos* insertions have normal length telomeres, suggesting
451 that neither the *Minos* elements themselves nor the genomes in which they reside contribute to
452 the formation of long *HeT-A* arrays. Rather, it appears the specific *MB09416* insertion may be
453 causative. Expression of *CG18493* in this transposon insertion stock is reduced significantly
454 relative to the control, although not to the same extent as *Tel[1]*. This supports the idea that the
455 telomeres seen in *Tel[1]* and *MB09416* may result from reduced *CG18493* expression.

Effect of Tel[1] on transcript expression in ovaries

457 This study identified a 3 bp deletion (TGT) in intron 8 of *Ino80* in the middle of chromosome
458 arm *3*R as the most likely causative factor for the telomere elongation phenotype found in the
459 *Tel[1]* strain. The Drosophila *Ino80* gene has 14 exons that spans 34,159 bp and produces a 5,243
460 nt long transcript, which encodes a single 1,638 aa protein. Intron 8 of *Ino80* has four other genes
461 nested inside it, while intron 12 includes another three genes. The Drosophila *Ino80* gene has a
462 regulatory role in DNA binding, and mutants in this gene showed its role in regulation of gene
463 expression (Bhatia et al. 2010). The protein is part of the Ino80 complex and present in nuclear
464 chromatin (Klymenko et al. 2006). Based on its sequence structure it is predicted to have a role
465 in ATP binding and helicase activity. Homozygous deletion mutations *Ino80[Δ3]* and *Ino80[Δ4]*, with
466 3 kb and 4 kb deletions respectively, spanning intron 11 and exon 12 of *Ino80*, are late
467 embryonic lethals (Bhatia et al. 2010). This suggests that *Ino80* has an essential, non-redundant
468 function during Drosophila development. The *Tel[1]* mutation has no known phenotype other than
469 telomere elongation, suggesting the 3 bp deletion in intron 8 does not inactivate the normal
470 function of *Ino80*. The results in this study also show that the integrity of the exon 8-9 splice
471 junction of the *Ino80* transcript is not affected, as its expression is at a normal level in
472 comparison to other exons in the ovary. The *Tel[1]* mutation might have a role at a specific stage in
473 the germ line that allows the telomeres to elongate. The possibility of a truncated/altered *Ino80*
474 transcript in the germ line or a novel transcript (non-coding RNA or small RNA) found in intron
475 8 of *Ino80* cannot be ruled out.

476   Of all the genes within the *Ino80* locus, only *CG18493* showed a significant repeatable
477 reduction in expression in *Tel[1]*. A slight reduction in the expression of *CG18493* observed in the
478 *MB09416* stock also supports the hypothesis that the long telomere phenotype observed in both
479 the lines could be caused by a reduction in expression of *CG18493*. This gene has no known
480 function, but based on the sequence information is predicted to have a serine-type peptidase
481 activity involved in proteolysis. It is expected to produce a transcript of 1,566 nt with a protein of
482 480 aa at embryonic stages 13-16 in embryonic/larval midgut. The TGT deletion is located 4 kb
483 from the 3' end of CG18493 with one gene between them, and *MB09416* is located more than 5
484 kb downstream of CG18493 with two intervening genes, suggesting *Tel[1]* and *MB09416* might
485 alter regulation of this locus by disrupting enhancers or other regulators. Further studies are
486 needed to elucidate the mechanism of this regulation.

487
488
489 **CONCLUSIONS**

490 This study mapped the Tel1 mutation in *Drosophila melanogaster* by *P* elements and *Minos*
491 transposons induced male recombination and subsequent analysis on telomere elongation
492 measurements over generations to a 15 kb region in the middle of the right arm of chromosome 3
493 (3R) at 92A. Further analysis by whole genome sequencing of Tel1 and comparison of variants
494 in the mapped region to Drosophila Genetic Reference Panel (DGRP) genomes eliminated all the
495 SNPs and most of the indels, and mapped Tel1 to a 3 bp deletion (TGT) at 3R:15,221,791 in the
496 middle of intron 8 of Ino80. This approach can be used to identify any mutation to a nucleotide
497 level from the entire genome of Drosophila, even if that mutation is not part of a coding region.

498

504 **REFERENCES**

505 Andreyeva EN, Belyaeva ES, Semeshin VF, et al. (2005) Three distinct chromatin domains in
506   telomere ends of polytene chromosomes in Drosophila melanogaster Tel mutants. J Cell Sci
507   118:5465–77. doi: 10.1242/jcs.02654

508 Bellen HJ, Levis RW, Liao G, et al. (2004) The BDGP gene disruption project: single transposon
509   insertions associated with 40% of Drosophila genes. Genetics 167:761–81. doi:
510   10.1534/genetics.104.026427

511 Bhatia S, Pawar H, Dasari V, et al. (2010) Chromatin remodeling protein ino80 has a role in
512   regulation of homeotic gene expression in Drosophila. Genes Cells 15:725–735.

513 Capkova Frydrychova R, Mason JM, Archer TK (2008) HP1 is distributed within distinct
514   chromatin domains at Drosophila telomeres. Genetics 180:121–31. doi:
515   10.1534/genetics.108.090647

516   Capkova Frydrychova R, Biessmann H, Mason JM (2009) Regulation of telomere length in
517   Drosophila. Cytogenet Genome Res 122:356–364. doi: 10.1159/000167823

518   Franz G, Savakis C (1991) Minos, a new transposable element from drosophila hydei, is a
519   member of the tc1-like family of transposons. Nucleic Acids Res 19:6646.

520   Capkova Frydrychova R., Mason J. (2013) Telomeres: Their structure and maintenance. in:
521   David Stuart (ed.) The Mechanisms of DNA Replication. InTech, Open Access Publisher,
522   Rijeka, Croatia, pp. 423-443. DOI: 10.5772/51356

523   Henson JD, Neumann AA, Yeager TR, Reddel RR (2002) Alternative lengthening of telomeres
524   in mammalian cells. Oncogene 21:598–610. doi: 10.1038/sj/onc/1205058

525   Chen B, Chu T, Harms E, et al. (1998) Mapping of Drosophila Mutations Using Site-Specific
526   Male Recombination. Genetics 149:157–163.

527   Kahn T, Savitsky M, Georgiev P (2000) Attachment of HeT-A sequences to chromosomal
528   termini in Drosophila melanogaster may occur by different mechanisms. Mol Cell Biol
529   20:7634–42.

530   Kent W (2002) Blat--the blast-like alignment tool. Genome Res. 12: 656-664.

531   Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M, Muller J
532   (2006) A polycomb group protein complex with sequence-specific DNA-binding and
533   selective methyl-lysine-binding activities. Genes & development 20, 1110-1122.

534   Krawitz P, Rodelsperger C, Jager M, Jostins L, Bauer S, Robinson PN (2010) Microindel
535   detection in short-read sequence data. Bioinformatics 26, 722-729.

536   Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform.
537   Bioinformatics 26:589–95.

538   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
539   Genome Project Data Processing, S. (2009) The sequence alignment/map format and
540   samtools. Bioinformatics 25, 2078-2079.

541   Loukeris TG, Arca B, Livadaras I, Dialektaki G, Savakis C. (1995) Introduction of the
542   transposable element minos into the germ line of drosophila melanogaster. Proceedings of
543   the National Academy of Sciences of the United States of America 92, 9485-9489.

544   Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y,
545   Magwire MM, Cridland JM et al. (2012) The Drosophila melanogaster genetic reference
546   panel. Nature 482, 173-178.

547   Mason JM, Frydrychova RC, Biessmann H (2008) Drosophila telomeres: an exception providing
548   new insights. Bioessays 30:25–37.

549   Melnikova L, Georgiev P (2002) Enhancer of terminal gene conversion, a new mutation in
550   drosophila melanogaster that induces telomere elongation by gene conversion. Genetics
551   162:1301–1312.

552   Metaxakis A, Oehler S, Klinakis A, Savakis C (2005) Minos as a Genetic and Genomic Tool in
553   Drosophila melanogaster. Genetics 171:571–581. doi: 10.1534/genetics.105.041848

554   Nabetani A, Ishikawa F (2011) Alternative lengthening of telomeres pathway: Recombination-
555   mediated telomere maintenance mechanism in human cells. J Biochem 149:5–14.

556   Neron B, Menager H, Maufrais C, et al. (2009) Mobyle: A new full web bioinformatics
557   framework. Bioinformatics 24:3005–3011.

558   Neumann AA, Reddel RR (2002) Telomere maintenance and cancer ? look, no telomerase. Nat.
559   Rev. Cancer 2:

560   Neumann AA, Watson CM, Noble JR, et al. (2013) Alternative lengthening of telomeres in
561   normal mammalian somatic cells. Genes Dev 18:18–23. doi:
562   10.1101/gad.205062.112.Freely

563   O'Rawe J, Jiang T, Sun G, et al. (2013) Low concordance of multiple variant-calling pipelines:
564   Practical implications for exome and genome sequencing. Genome Med. 5:

565   Pineyro D, Lopez-Panades E, Lucena-Perez, M. Casacuberta E (2011) Transcriptional analysis
566   of the het-a retrotransposon in mutant and wild type stocks reveals high sequence variability
567   at drosophila telomeres and other unusual features. BMC Genomics 12:573.

568   Proskuryakov KA, Melnikova LS (2008) Functional separation of genetic factors telomere
569   elongation (tel) and enhancer of terminal gene conversion (e(tc)) involved in telomere
570   elongation in Drosophila melanogaster. Dokl Biochem Biophys 421:199–203.

571   Raffa GD, Ciapponi L, Cenci G, et al. (2013) Organization and evolution of Drosophila terminin:
572   Similarities and differences between Drosophila and human telomeres. Front Oncol *3*, 112.

573   Roth CW, Kobeski F, Walter MF, Biessmann H (1997) Chromosome end elongation by
574   recombination in the mosquito Anopheles gambiae. Mol Cell Biol 17:5176–83.

575   Ryder E, Russell S (2003) Transposable elements as tools for genomics and genetics in
576   Drosophila. Brief Funct Genomic Proteomic 2:57–71.

577   Savitsky M, Kravchuk O, Melnikova L, Georgiev P (2002) Heterochromatin Protein 1 Is
578   Involved in Control of Telomere Elongation in Drosophila melanogaster. 22:3204–3218.
579   doi: 10.1128/MCB.22.9.3204

580   Savitsky M, Kwon D, Georgiev P, Kalmykova A, Gvozdev V (2006) Telomere elongation is
581   under the control of the RNAi-based mechanism in the Drosophila germline. 345–354. doi:
582   10.1101/gad.370206.

583   Sfeir A, de Lange T (2012) Removal of shelterin reveals the telomere end-protection problem.
584   Science 336:593–7.

585   Silva-Sousa R, Varela MD, Casacuberta E (2013) The Putzig partners DREF, TRF2 and KEN
586   are involved in the regulation of the Drosophila telomere retrotransposons, HeT-A and
587   TART. Mob DNA 4:18. doi: 10.1186/1759-8753-4-18

588   Simpson J, Wong K, Jackman S, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel
589   assembler for short read sequence data. Genome Res 19:1117–23.

590   Siriaco GM, Cenci G, Haoudi A, Champion LE, Zhou C, Gatti M, Mason JM (2002) Telomere
591   elongation (tel), a new mutation in Drosophila melanogaster that produces long telomeres.
592   Genetics 160, 235-245

593   Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM (1995) Gene disruptions using p
594   transposable elements: An integral component of the Drosophila genome project. PNAS
595   92:10824–10830.

596   Takács S, Biessmann H, Reddy HM, Mason JM, Torok T (2012) Protein Interactions on
597   Telomeric Retrotransposons in Drosophila. IntJ Biol Sci 10:1055 –1061. doi:
598   10.7150/ijbs.4460

599   Török T, Benitez C, Takács S, Biessmann H (2007) The protein encoded by the gene
600   proliferation disrupter (prod) is associated with the telomeric retrotransposon array in
601   Drosophila melanogaster. Chromosoma 116:185–95. doi: 10.1007/s00412-006-0090-4

602   Venken KJ, Schulze KL, Haelterman NA, et al. (2011) Mimic: A highly versatile transposon
603   insertion resource for engineering Drosophila melanogaster genes. Nat Methods 8:737–743.

604   Walter MF, Biessmann MR, Benitez C, Torok T, Mason JM, Biessmann H. (2007) Effects of
605   telomere length in drosophila melanogaster on life span, fecundity, and fertility.
606   Chromosoma 116:41–51.

607   Wei KH, Reddy HM, Rathnam C, Lee J, Lin D, Ji S, Mason JM, Clark AG, Barbash DA (2017)
608   A pooled sequencing approach identifies a candidate meiotic driver in Drosophila. Genetics
609   206:451–465.

610   Zhai RG, Hiesinger PR, Koh T, et al. (2003) Mapping Drosophila mutations with molecularly
611   defined P element insertions. PNAS 100:10860–10865.

612

613   **FIGURE LEGENDS**

614   **Figure 1. Localization of *Tel¹* by site-specific recombination. A.** The upper chromosome
615   map shows the candidate genes between two P element insertion sites, *05151* and *05113*
616   (vertical green lines). This region was identified as containing *Tel¹* based on Table 1, round
617   1. Positions of transposons used for further mapping are indicated by green arrows. The *Tel¹*
618   mutation is boxed in red. The lower chromosome map shows expansion of the 92A3 region.
619   **B**. Graphs showing the change in relative *HeT-A* copy number (telomere length) in
620   recombinants of *Tel¹/MI02316*, *Tel¹/MI03112* and *Tel¹/d10097* over 12 generations. The *st*
621   recombinants are shown as red lines; *ca* recombinants as purple lines. These data delimit
622   *Tel¹* to a 15 kb between inserts *MI02316* and *d10097* (shown as red rectangles in Figure 1A).

623   **Figure 2. Telomere length in transposon insertion stocks.** Q-PCR analysis of *HeT-A* copy
624   number in different transposon insertion stocks used for mapping *Tel¹* mutation. Error bars

625   represent standard deviation measured from the triplicate Q-PCR results. *MB09416* was not
626   used for subsequent site-specific recombination mapping.

627   **Figure 3. Dosage effects of Tel on *HeT-A* copy number. A.** The cytogenetic and physical
628   map of genomic region 91C-91D is displayed, highlighting the 15 kb *Tel[1]* region as a
629   vertical box. The lower part of the panel shows the extent of chromosome deficiencies. The
630   bottom rectangle, *Dp(3:3)cam30T*, is a duplication for this region. Dotted lines beyond this
631   rectangle show that the duplication extends beyond the represented region. **B.** Shows the
632   relative *HeT-A* copy number in stocks of the aberrations shown in A. The highlighted box
633   represents a duplication that includes mapped region (*Dp(3;3)cam30T* covers 90C-93C) and,
634   another duplication of a neighboring region of the genome *(Dp(3;3)cam35* covers 67C5-
635   69A5) as a control. The mean from three replicates is represented here and error bars
636   represent standard deviation.

637   **Figure 4. Telomere lengths in DGRP lines.** A bar graph shows the log normal distribution
638   of telomere lengths among the 162 DGRP lines measured. The blue arrow shows the
639   position of the Oregon-R control, and the red arrow shows the position of *Tel[1]*. Three lines
640   have *Het-A* copy numbers that exceed three standard deviations from the mean. These are
641   RAL-161, -703 and -882. The red curve indicates the expected distribution.

642   **Figure 5. Genetic activity and conservation of the 15 kb Tel1 region. A**. The genes found
643   in this region are aligned with molecular coordinates. Minos insertions used for mapping are
644   shown in cyan triangles. Minos insertion MB09416, which showed telomere elongation, is
645   highlighted in red square. **B.** The University of California Santa Cruz genome browser map
646   highlights sequence conservation in this region among different insect species. **C.** A
647   developmental transcriptome analysis for the same region as determined by the
648   modENCODE project is also shown. The red vertical line spanning all three panels indicates
649   the position of the TGT deletion.

650   **Figure 6. Transcript analysis in the genes near *Tel[1]*. A.** The histogram represents the
651   relative transcript levels from Tel1 mutant (red) and Oregon-R (blue) adult ovaries. The
652   genes analyzed are found near *Tel[1]* (TGT mutation), which include *Ino80* and the genes
653   within its introns. The last bar represents the expression of Ino80 gene spanning exons 8-9,
654   around the position of the TGT deletion in intron 8. The expression levels of *CG31244* and
655   CG31245 are very low, similar to the modENCODE data. **B.** The relative transcript level of
656   CG18493 was measured among different control lines. Standard laboratory strains Oregon R
657   and Canton S were tested as well as DGRP line RAL-513, which has a chromosomal *HeT-A*
658   copy number near the DGRP mean. As the *MB09416* insertion is very close to the *Tel* TGT
659   deletion, it was tested at the same time and shows a reduction in expression of *CG18493*
660   compared to all of the controls.

661

Figure 1(on next page)

Localization of *Tel¹* by site-specific recombination.

A. The upper chromosome map shows the candidate genes between two P element insertion sites, *05151* and *05113* (vertical green lines). This region was identified as containing *Tel¹* based on Table 1, round 1. Positions of transposons used for further mapping are indicated by green arrows. The *Tel¹* mutation is boxed in red. The lower chromosome map shows expansion of the 92A3 region. **B.** Graphs showing the change in relative *HeT-A* copy number (telomere length) in recombinants of *Tel¹/MI02316*, *Tel¹/MI03112* and *Tel¹/d10097* over 12 generations. The *st* recombinants are shown as red lines; *ca* recombinants as purple lines. These data delimit *Tel¹* to a 15 kb between inserts *MI02316* and *d10097* (shown as red rectangles in Figure 1A).
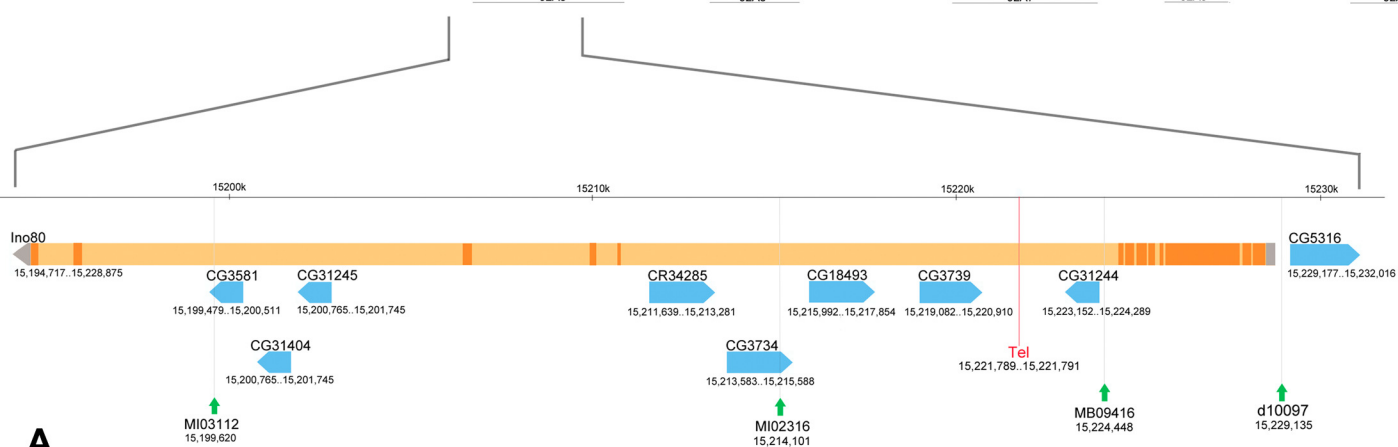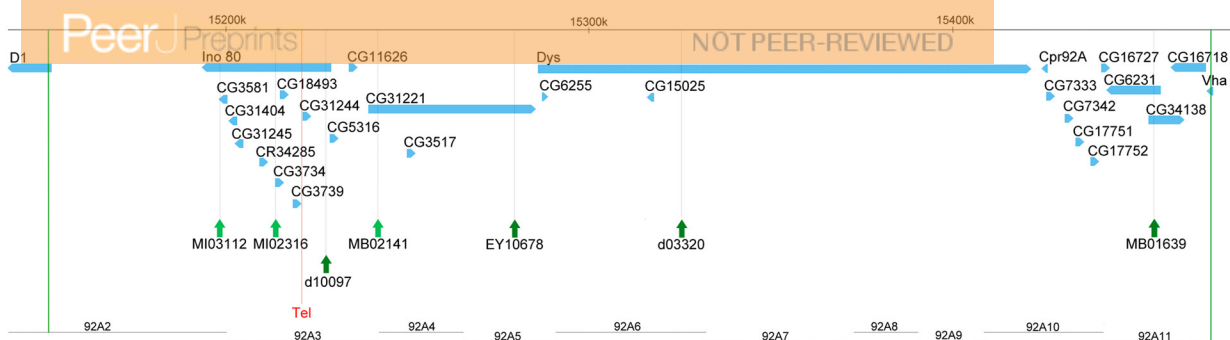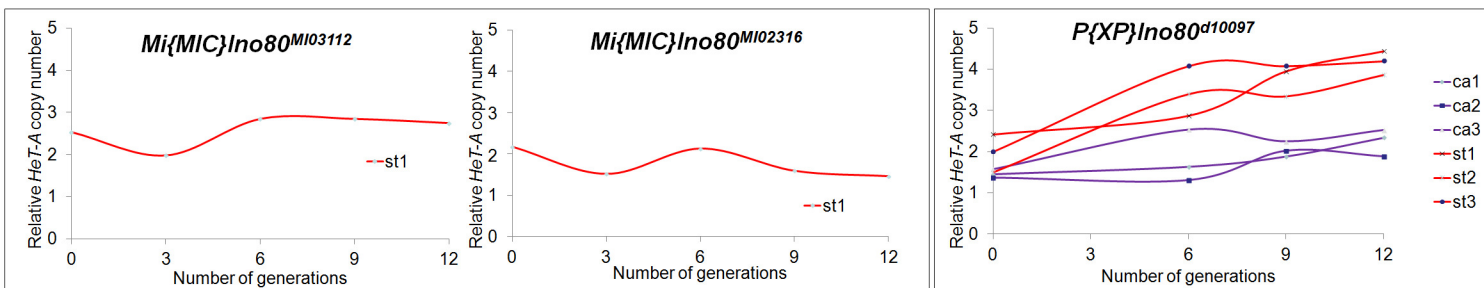
Figure 2(on next page)

Telomere length in transposon insertion stocks.

Q-PCR analysis of *HeT-A* copy number in different transposon insertion stocks used for mapping *Tel*[1] mutation. Error bars represent standard deviation measured from the triplicate Q-PCR results. *MB09416* was not used for subsequent site-specific recombination mapping.
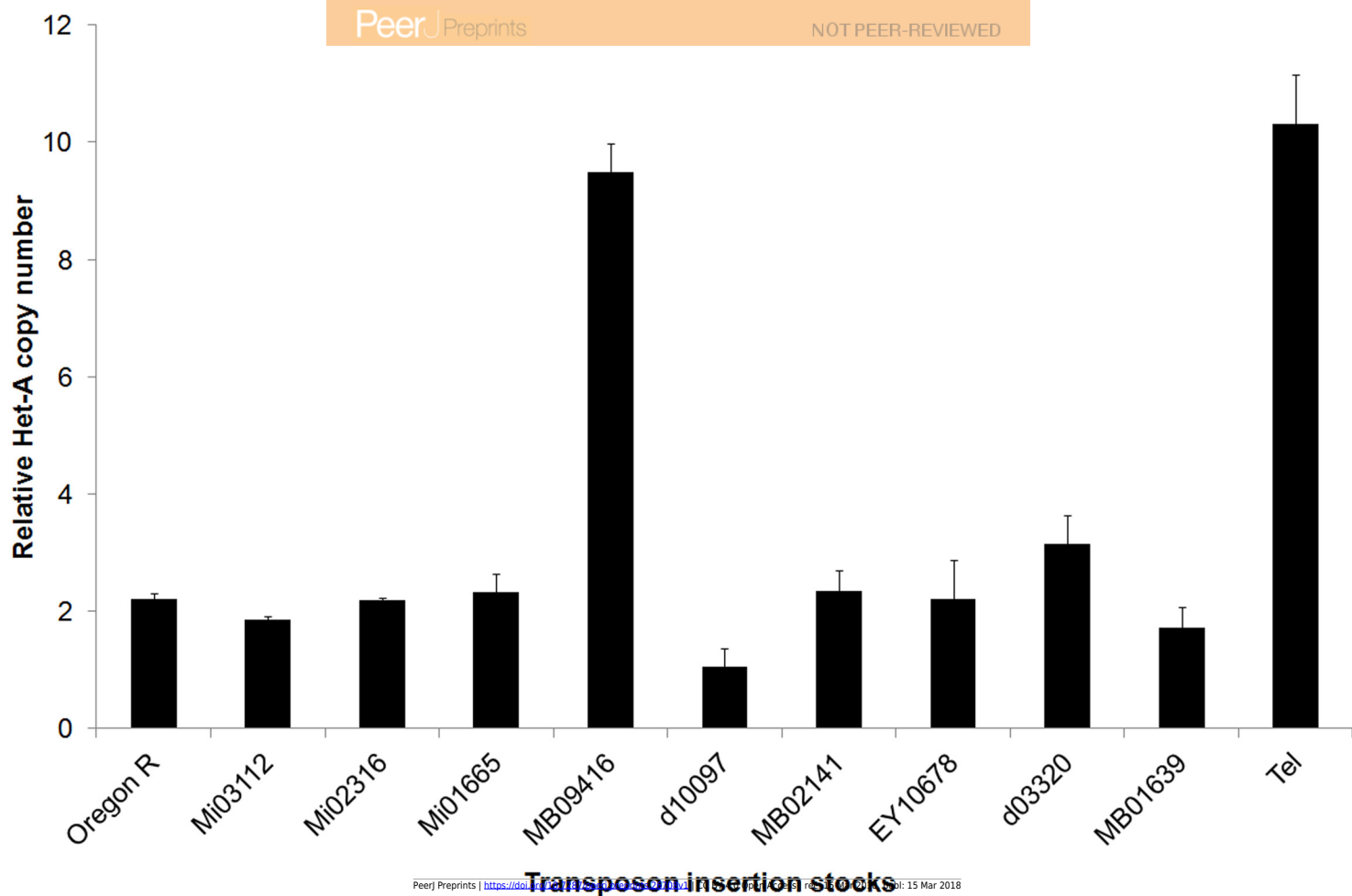
Figure 3(on next page)

Dosage effects of Tel on *HeT-A* copy number.

A. The cytogenetic and physical map of genomic region 91C-91D is displayed, highlighting the 15 kb *Tel[1]* region as a vertical box. The lower part of the panel shows the extent of chromosome deficiencies. The bottom rectangle, *Dp(3:3)cam30T*, is a duplication for this region. Dotted lines beyond this rectangle show that the duplication extends beyond the represented region. B. Shows the relative *HeT-A* copy number in stocks of the aberrations shown in A. The highlighted box represents a duplication that includes mapped region (*Dp(3;3)cam30T* covers 90C-93C) and, another duplication of a neighboring region of the genome *(Dp(3;3)cam35* covers 67C5-69A5) as a control. The mean from three replicates is represented here and error bars represent standard deviation.
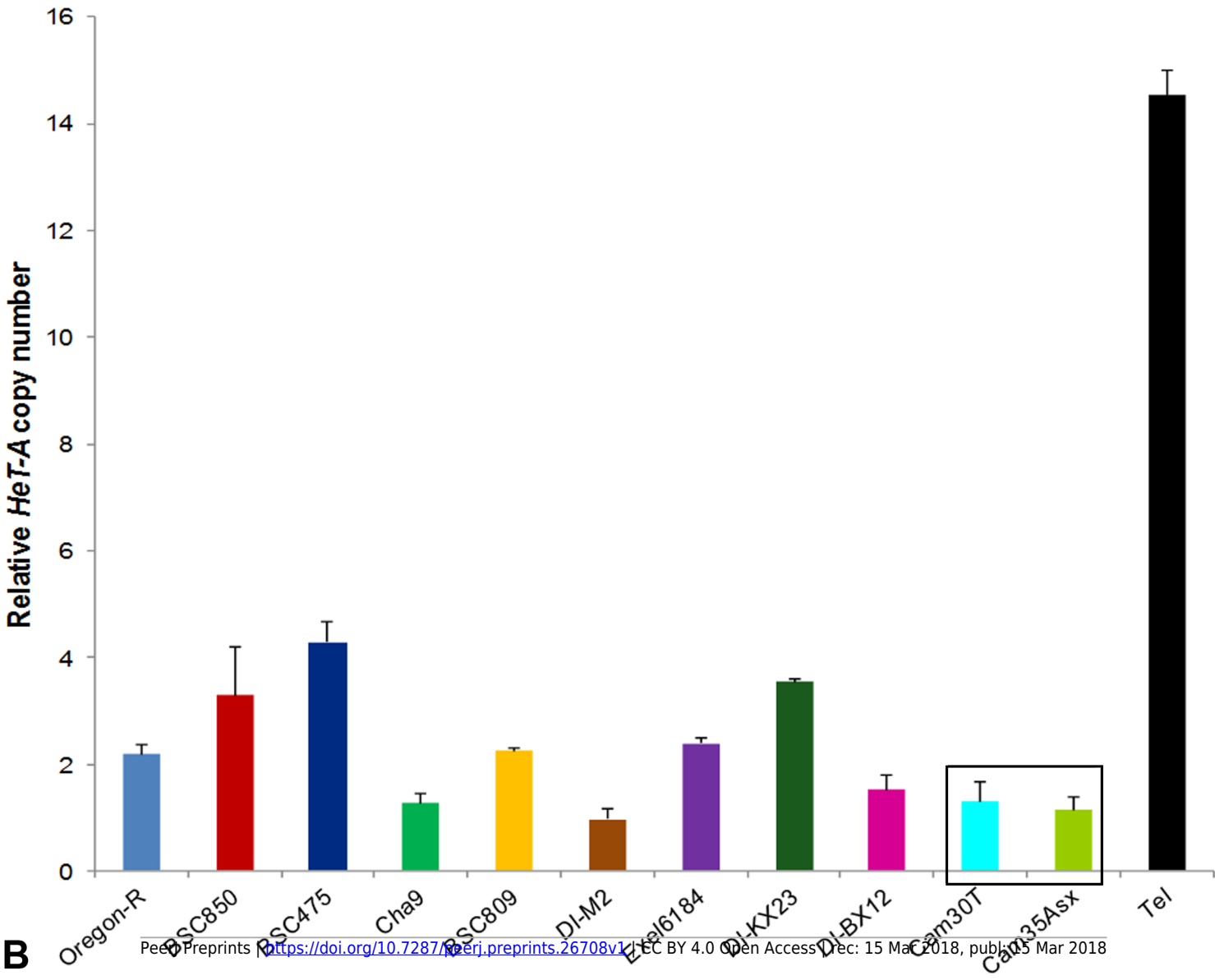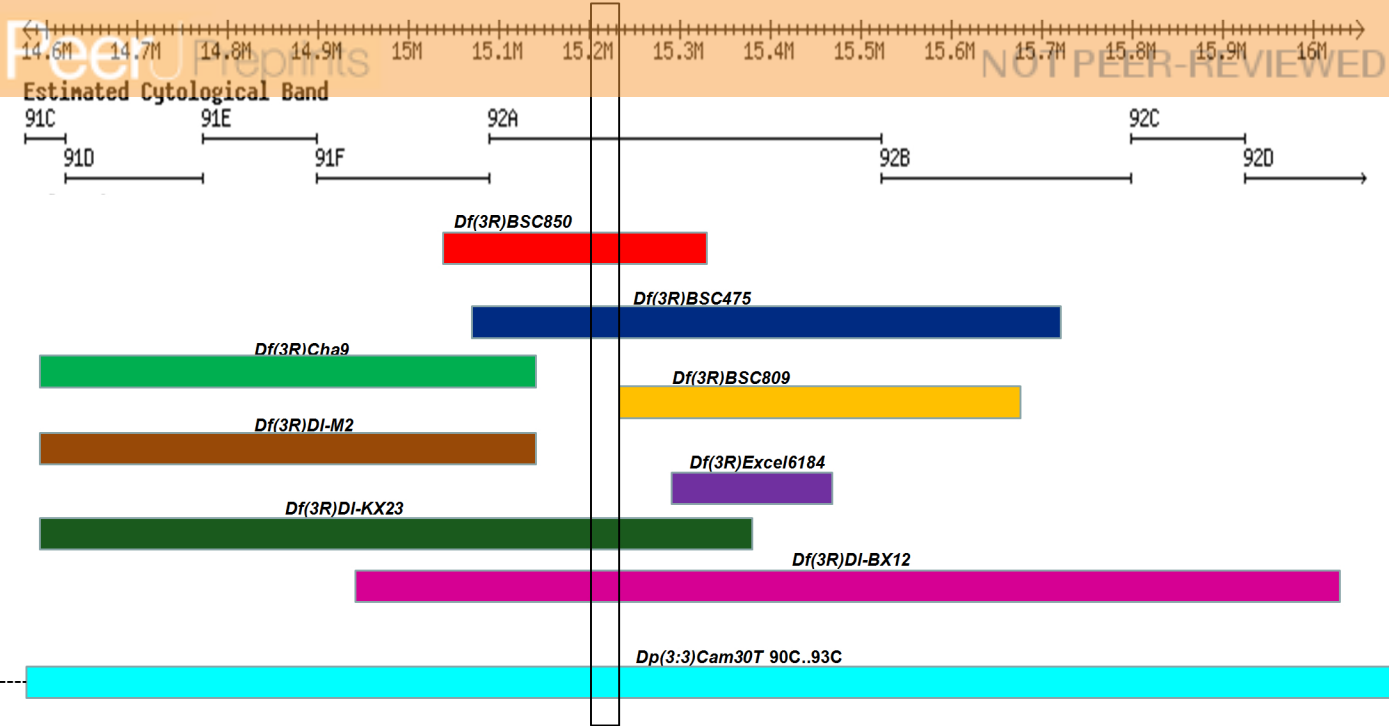
B

Figure 4(on next page)

Telomere lengths in DGRP lines.

A bar graph shows the log normal distribution of telomere lengths among the 162 DGRP lines measured. The blue arrow shows the position of the Oregon-R control, and the red arrow shows the position of Tel^1. Three lines have Het-A copy numbers that exceed three standard deviations from the mean. These are RAL-161, -703 and -882. The red curve indicates the expected distribution.

|

Oregon-R

Tel

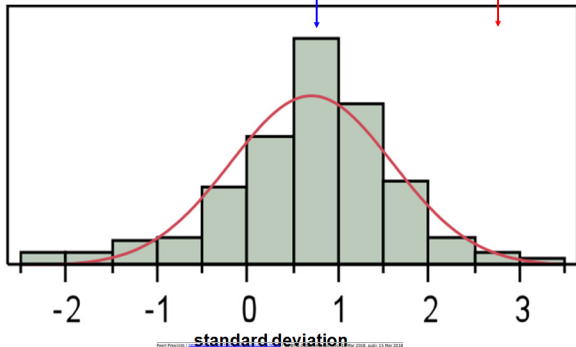number of strains

standard deviation

Figure 5(on next page)

Genetic activity and conservation of the 15 kb Tel1 region.

A. The genes found in this region are aligned with molecular coordinates. Minos insertions used for mapping are shown in cyan triangles. Minos insertion MB09416, which showed telomere elongation, is highlighted in red square. B. The University of California Santa Cruz genome browser map highlights sequence conservation in this region among different insect species. **C.** A developmental transcriptome analysis for the same region as determined by the modENCODE project is also shown. The red vertical line spanning all three panels indicates the position of the TGT deletion.

A
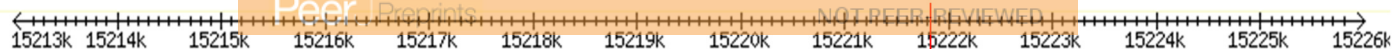
pseudogene:CR34285-RB
Ino80-RA

CG3734-RA   CG18493-RA   CG3739-RB   CG31244-RA
CG3734-RB

Transgenic insertion site
Mi{MIC}Ino80[MI02316]   Mi{MIC}Ino80[MI01665]
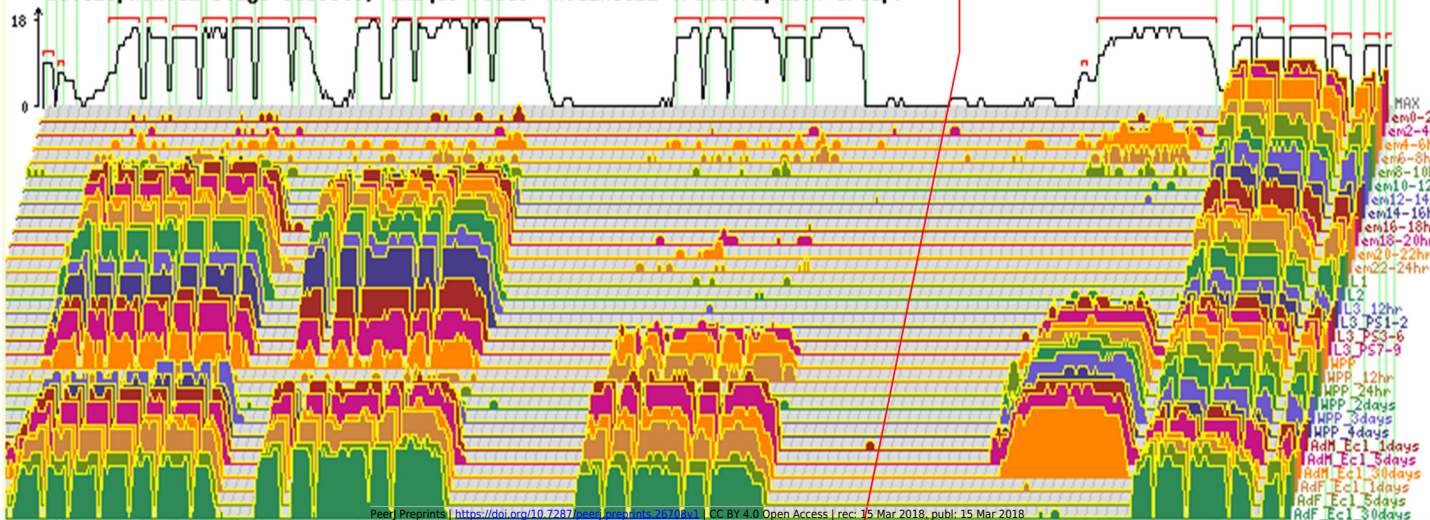   Mi{ET1}Ino80[MB0941

B

12 Flies, Mosquito, Honeybee, Beetle Multiz Alignments & phastCons Scores

Conservation

C

Developmental stage subsets, unique reads (modENCODE Transcription Group)

18

0

MAX
em0-2
em2-4h
em4-6h
em6-8h
em8-10h
em10-12
em12-14h
em14-16h
em16-18h
em18-20hr
em20-22hr
em22-24hr
L2
L3_12hr
L3_PS1-2
L3_PS3-6
L3_PS7-9
WPP
WPP_12hr
WPP_24hr
WPP_2days
WPP_3days
WPP_4days
AdM_Ecl_1days
AdM_Ecl_5days
AdM_Ecl_30days
AdF_Ecl_1days
AdF_Ecl_5days
AdF_Ecl_30days

Figure 6(on next page)

Transcript analysis in the genes near *Tel*[1].

A. The histogram represents the relative transcript levels from Tel1 mutant (red) and Oregon-R (blue) adult ovaries. The genes analyzed are found near *Tel*[1] (TGT mutation), which include *Ino80* and the genes within its introns. The last bar represents the expression of Ino80 gene spanning exons 8-9, around the position of the TGT deletion in intron 8. The expression levels of *CG31244* and CG31245 are very low, similar to the modENCODE data. **B.** The relative transcript level of CG18493 was measured among different control lines. Standard laboratory strains Oregon R and Canton S were tested as well as DGRP line RAL-513, which has a chromosomal *HeT-A* copy number near the DGRP mean. As the *MB09416* insertion is very close to the *Tel* TGT deletion, it was tested at the same time and shows a reduction in expression of *CG18493* compared to all of the controls.
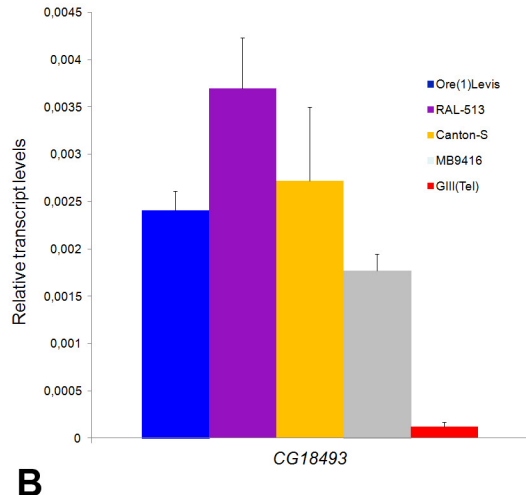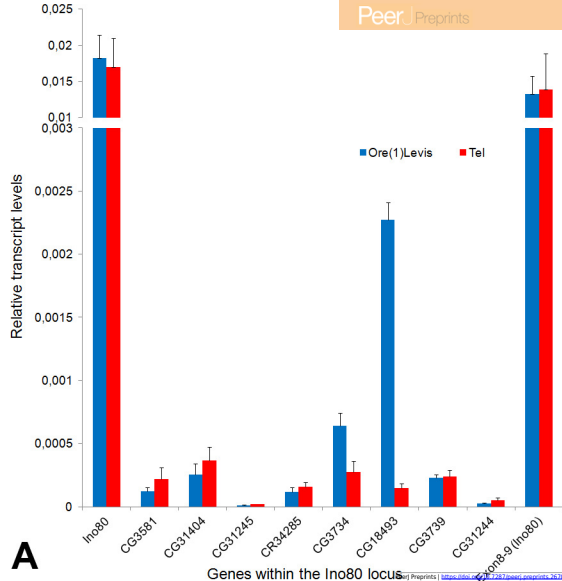
A

B

Table 1(on next page)

Recombinants obtained from each transposon used for site specific recombination

[a] Transposons were used to induce recombination as they became available. Succeeding rounds used slightly different procedures as described in the text. [b] Estimated cytological band as reported in FlyBase. [c] Nucleotide position in the genomic sequence of chromosome arm 3R. [d] The position of Tel^1 relative to the insertion site.

1   **Table 1.** Recombinants obtained from each transposon used for site specific recombination

Round[a]	Transposon insertion	Cytology[b]	Coordinate[c]	No. Recombinants	Tel position[d]
1	$P\{PZ\}sqz^{02102}$	91F4	14991598	3	Right
1	$P\{lacW\}vib^{j5A6}$	91F12	15052087	6	Right
1	$P\{PZ\}l(3)05820^{05820}$	91F12	15055218	3	Right
1	$P\{PZ\}Dl^{05151}$	92A2	15151940	2	Right
3	$Mi[4]Ino80^{MI03112}$	92A3	15199620	1	Right
3	$Mi[4]Ino80^{MI02316}$	92A3	15214101	1	Right
2	$P[26]Ino80^{d10097}$	92A3	15229135	6	Left
2	$Mi\{ET1\}CG31221^{MB02141}$	92A3	15239869	5	Left
2	$P\{EPgy2\}CG31221^{EY10678}$	92A5	15279376	4	Left
2	$P[26]Dys^{d03320}$	92A6	15324651	3	Left
2	$MI\{ET1\}CG6231^{MB01639}$	92A11	15455374	6	Left
1	$P\{SUPor-P\}CG16718^{KG06218}$	92A11	15467496	3	Left
1	$P\{PZ\}Vha13^{05113}$	92A11	15469740-15470048	3	Left
1	$P\{hsneo\}l(3)neo50^{1}$	92B3	15662593	3	Left

2   [a] Transposons were used to induce recombination as they became available. Succeeding rounds used slightly
3   different procedures as described in the text.

4   [b] Estimated cytological band as reported in FlyBase.

5   [c] Nucleotide position in the genomic sequence of chromosome arm *3*R.

6   [d] The position of Tel^1 relative to the insertion site.

7

Table 2 (on next page)

Table 2. SNPs and indels found in the *Tel1* genome relative to the standard reference sequence of chromosome arm 3R between coordinates 15,151,000 and 15,230,000

[1] There were 159 DGRP lines used in the SNP comparison (Freeze1 data) and eight used in the indel comparison.

1   **Table 2.** SNPs and indels found in the *Tel1* genome relative to the standard reference sequence of chromosome
2   arm 3R between coordinates 15,151,000 and 15,230,000

Polymorphisms	SNPs	Indels
Identified by CLC-Genomics	626	88
Identified by manual comparison	-	13
Identified by *de novo* assembly	-	14
Total	**626**	**115**
Not in DGRP[a]	0	2

3   [1] There were 159 DGRP lines used in the SNP comparison (Freeze1 data) and eight used in the indel comparison.

4