Targeting of \( P \) element reporters to heterochromatic domains by transposable element 1360 in \( Drosophila melanogaster \)

Authors:
Kathryn L Huisinga\(^*\), Nicole C Riddle\(^\dagger\), Wilson Leung\(^\ddagger\), Shachar Shimonovich\(^\dagger \dagger\), Stephen McDaniel\(^\ddagger \dagger\), Alejandra Figueroa-Clarevega\(^\ddagger \dagger\), and Sarah CR Elgin\(^\ddagger\)

Affiliations:
\(^*\) Department of Science and Mathematics, Malone University, Canton OH 44709

\(^\dagger\) Department of Biology, The University of Alabama at Birmingham, Birmingham AL 35294

\(^\ddagger\) Biology Department, Washington University in St. Louis, St. Louis MO 63130

\(^\dagger \dagger\) Present address: Department of Surgery, SUNY Downstate Health Sciences Center, Brooklyn, NY 11203

\(^\ddagger \dagger\) Present address: Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

\(^\ddagger\) Present address: Department of Molecular and Cell Biology, University of California – Berkeley, Berkeley, CA 94720
Running title: 1360 can target heterochromatin

Key words: heterochromatin; transposable elements; position effect variegation; chromatin; Drosophila

Corresponding authors:
Kathryn L. Huisinga
Assistant Professor of Biochemistry
Malone University
2600 Cleveland Ave NW
Canton OH 44709
khusinga@malone.edu
(330) 471-8356

Nicole C. Riddle
Assistant Professor of Biology
The University of Alabama at Birmingham
Birmingham, AL 35294
riddlenc@uab.edu
(205) 975-4049

Sarah C. R. Elgin
Viktor Hamburger Professor of Arts & Sciences
Professor of Biology
Washington University in St. Louis
St. Louis, MO 63130
selgin@wustl.edu
(314) 935-5348
ABSTRACT

Heterochromatin is a common DNA packaging form employed by eukaryotes to constitutively silence transposable elements. Determining which sequences to package as heterochromatin is vital for an organism. Here, we use Drosophila melanogaster to study heterochromatin formation, exploiting position effect variegation, a process whereby a transgene is silenced stochastically if inserted in proximity to heterochromatin, leading to a variegating phenotype. Previous studies identified the transposable element 1360 as a target for heterochromatin formation. We use transgene reporters with either one or four copies of 1360 to determine if increasing local repeat density can alter the fraction of the genome supporting heterochromatin formation. We find that including 1360 in the reporter increases the frequency with which variegating phenotypes are observed. This increase is due to a greater recovery of insertions at the telomere-associated sequences (~50% of variegating inserts). In contrast to variegating insertions elsewhere, the phenotype of telomere-associated sequence insertions is largely independent of the presence of 1360 in the reporter. We find that variegating and fully expressed transgenes are located in different types of chromatin, and variegating reporters in the telomere-associated sequences differ from those in pericentric heterochromatin. Indeed, chromatin marks at the transgene insertion site can be used to predict the eye phenotype. Our analysis reveals that increasing the local repeat density (via the transgene reporter) does not enlarge the fraction of the genome supporting heterochromatin formation. Rather, additional copies of 1360 appear to target the reporter to the telomere-associated sequences with greater efficiency, thus leading to an increased recovery of variegating insertions.
INTRODUCTION

_In vivo_ regulation of gene expression occurs in the context of chromatin, the complex structure formed by DNA, histones, and a variety of associated proteins. Two basic types of chromatin were originally distinguished, euchromatin and heterochromatin, based on cytological staining behavior during the cell cycle (Heitz 1928). Euchromatin and heterochromatin differ in a number of characteristics, including biochemical makeup and their effect on the expression of transgene reporters. Euchromatin is generally accessible for transcription, contains the majority of the genes, and is characterized by biochemical marks associated with transcription, such as high levels of histone acetylation, transcriptional activators, and RNA polymerase II. In contrast, heterochromatin contains few genes and has a high repeat density. Heterochromatin is characterized by low levels of histone acetylation, high levels of histone 3 lysine 9 (H3K9) methylation, and the presence of “silencing” proteins such as heterochromatin protein 1 (HP1) (Beisel and Paro 2011). While this model of two basic chromatin types has been very successful in explaining many experimental observations, recent genome-wide profiling of chromatin structure has demonstrated that there are many chromatin sub-types, some of which combine aspects of euchromatin and heterochromatin [work in Drosophila: (Filion et al. 2010; Kharchenko et al. 2010; Riddle et al. 2011)].

Reporter transgenes can be used to probe the chromatin context of the genome and to examine the impact of chromatin structure on gene expression. Expression levels differ based on the transgene insertion site, reflecting both local regulatory elements and
larger packaging domains. When a normally euchromatic reporter transgene is inserted into heterochromatic regions of the genome, a variegating expression pattern is observed, due to the transgene being silenced in some of the individual cells in which it would normally be active. This phenomenon, termed position effect variegation (PEV), appears to be a general characteristic, determined by the chromatin domain, not the reporter [for a review focused on PEV in Drosophila see (ELGIN AND REUTER 2013)]. In *Drosophila melanogaster*, studies using a transgene construct encoding the white gene (essential for red eye color) driven by the hsp70 heat-shock promoter (*hsp70-white*) have shown that a variegating (PEV) phenotype is observed when the transgene is inserted into any of the major heterochromatic domains: the pericentric regions, the telomeres, the Y chromosome, and the fourth chromosome (WALLRATH AND ELGIN 1995). In the small fourth chromosome (~1% of the genome), transgene reporters with variegating expression and with full expression are interspersed (SUN *et al.* 2000). A careful examination of this domain revealed a correlation between the presence of the repetitive element 1360 and the variegating phenotype (SUN *et al.* 2004). Follow-up studies found that incorporating 1360 sequences into the *hsp70-white* transgene construct could increase the gene silencing observed at a subset of insertion sites (HAYNES *et al.* 2006; SENTMANAT AND ELGIN 2012). However, these studies also showed that the presence of 1360 was not sufficient to induce PEV – and heterochromatin formation – in general, as full expression, rather than variegation, is observed for insertions of the 1360-*hsp70-white* transgene at most sites in the euchromatic chromosome arms (HAYNES *et al.* 2006; SENTMANAT AND ELGIN 2012).
The available data suggest that PEV is caused by spreading of adjacent heterochromatin to encompass the reporter gene (Weiler and Wakimoto 1995; Vogel et al. 2009). The effectiveness of this spreading is thought to be stochastic, thus leading to the variegating phenotype. As noted above, proximity to high repeat density regions is key, even if a repeated element such as 1360 is included in the transgene (Wallrath and Elgin 1995; Sun et al. 2000; Haynes et al. 2006; Sentmanat and Elgin 2012). While the importance of a high density of repetitive elements is well supported, genetic evidence indicates that there are several different mechanisms operating in the various heterochromatic domains of the Drosophila genome (Phalke et al. 2009). Studies of PEV have identified approximately 150 genes that either enhance or suppress the variegating phenotypes (Schotta et al. 2003), and these genes differ in their impact on the same reporter in different domains (Phalke et al. 2009).

The mechanisms that target heterochromatin formation are poorly understood. In addition to the H3K9 histone methyltransferases and HP1a, some of the various RNAi pathway components appear to be involved in heterochromatin formation (Pal-Bhadra et al. 2004; Girton and Johansen 2008). Specifically the piRNA pathway is important for the silencing of some, but not all transposable elements (TEs) with associated heterochromatin formation, as mutations in various components in the piRNA pathway can alter such TE silencing (Brower-Toland et al. 2007; Yin and Lin 2007; Kavi et al. 2008; Fagegaltier et al. 2009; Moshkovich and Lei 2010; Wang and Elgin 2011; Gu and Elgin 2013). In particular, the deletion of piRNA homology sites within 1360 in a 1360-hsp70-white transgene can alter its PEV phenotype, and similar results have been
obtained with Invader4 (Sentmanat and Elgin 2012). These findings raise the question of whether repeat content and local sequence context influence PEV independent of the piRNA pathway or whether these processes are linked, potentially providing a targeting mechanism.

To further investigate the mechanism(s) leading to heterochromatin formation and to evaluate the influence of the local repeat content versus sequence context, we carried out an extensive screen using two P-based constructs with the hsp70-white reporter: P{T1} has one copy of 1360, while P{T4} has four copies of 1360, altering the local TE content. 110 insertion lines with variegating eye color were recovered, with a significantly higher frequency of variegating lines observed when additional copies of 1360 are present. We find that the genomic distributions of insertion sites of transgenes with one or with four copies of 1360 are similar, indicating that providing additional repetitive sequences locally does not expand the regions of the genome supporting heterochromatin formation. Rather, the transgene with additional 1360 repeats appears to be targeted to heterochromatic regions of the genome more efficiently, in particular to the telomere-associated sequences (TAS) adjacent to the telomeres of chromosome arms 2R and 3R. In contrast to many 1360 transgenes in pericentric-proximal domains, most transgenes inserted in the TAS do not exhibit an altered phenotype in response to the removal of 1360 sequences, indicating that while the 1360 sequences increase the targeting of the transgene reporter to the TAS, they are not required for maintaining the variegating phenotype. While chromatin marks do not appear to predict a transgene’s response to the removal of 1360 sequences, we identify H3K9me3, HP1a, and
H3K23ac as chromatin components that can be used to predict whether a genomic domain will yield a variegating phenotype or full expression on insertion of this reporter. These and prior results suggest that not only is 1360 recognized as a target for silencing in pericentric and fourth chromosome domains, it is also being recognized and targeted for insertion into specific TAS heterochromatic domains.
RESULTS

Additional copies of 1360 in an hsp70-white reporter lead to a higher frequency of lines exhibiting PEV.

Previous mobilizations of the \( P\{T1\} \) reporter yielded one insertion line (a line carrying a single copy of our reporter transposable element; Figure 1A) with a variegating eye phenotype and 17 insertions with a solid eye color phenotype (HAYNES et al. 2006). To address the impact of incorporating additional 1360 sequences into the hsp70-white reporter, we constructed a second \( P \) element containing four copies of 1360 upstream of the hsp70-white reporter (\( P\{T4\} \); Figure 1A). Large-scale mobilizations of \( P\{T1\} \) and \( P\{T4\} \) led to recovery of 110 insertion lines with a variegating eye phenotype, 38 for \( P\{T1\} \) and 72 for \( P\{T4\} \). (See Supplemental Figure S1 for the crossing scheme and Table 1 for summary statistics; insertion lines carrying \( P\{T1\} \) are designated “9M1-x”, while insertion lines carrying \( P\{T4\} \) are designated “9M4-x”.) In addition, many solid red-eyed insertion lines, as well as several fly lines with a solid dark-orange to pale-yellow eye color (non-red, solid; NRS), were recovered. The fraction of lines exhibiting a variegating eye was increased approximately three-fold for \( P\{T4\} \) relative to \( P\{T1\} \) (12% vs. 4%; \( p <0.0001 \), Chi-square test; Figure 1B). The percentage of variegating \( P\{T4\} \) transgene insertions on the 2\(^{nd} \) and 3\(^{rd} \) chromosomes was increased relative to the \( P\{T1\} \) transgene, but the percentage on the heterochromatic 4\(^{th} \) chromosome was very similar (Table 2). Thus, additional copies of 1360 led to the recovery of a larger fraction of lines exhibiting PEV, with insertion sites specifically on chromosomes 2 and 3. Here we present an analysis that includes 211 new insertion lines derived from mobilization (54 \( P\{T1\} \) and 157 \( P\{T4\} \)), plus all of our \( P \) element transformation lines (33
lines), as well as 19 \( P\{T1\} \) insertion lines published previously (Haynes et al. 2006), a total of 263 lines (Supplemental Tables S1 and S2).

**1360-containing reporters exhibiting PEV map to established sites of heterochromatin.**

To examine the genome-wide distribution of the \( P\{T1\} \) and \( P\{T4\} \) reporters, we mapped the transgene insertion sites in the \( D. \) melanogaster genome. From the 211 stocks generated in the mobilization screen and 33 additional \( P\{T1\} \) insertion lines from a round of \( P \) element transformation, 60 new \( P\{T1\} \) lines and 111 new \( P\{T4\} \) lines were confirmed to contain undamaged, single copy inserts which were mapped with high confidence to a single genomic location (Supplemental Table S1; see Materials and Methods for details). A map of all new insertion sites is shown in Figure 2. Consistent with previous results, the insertion sites for most red-eyed lines map to the euchromatic chromosome arms. In addition, three red-eyed lines are located on chromosome 4, a genomic domain where red and variegating inserts can be found in close proximity (Sun et al. 2004). Most variegating insertions are found in known heterochromatic domains of the Drosophila genome. PEV-inducing insertion sites are within two cytological regions of the centromeres of chromosomes 2 and 4 (i.e. cytological regions 38-42 & 101-102) or in subtelomeres of chromosome arms 2R and 3R (60F and 100E). Four variegating \( P\{T1\} \) lines have insertion sites located in the histone gene cluster (HIS-C) at cytological region 39D3-39D5, a known heterochromatic locus. Three additional PEV lines (two \( P\{T1\} \) and one \( P\{T4\} \)) have reporter insertion sites located outside these regions on the euchromatic chromosome arms. Thus, the overall distribution of
variegating $P\{T1\}$ and $P\{T4\}$ reporters is very similar to the distribution observed with the original $hsp26$-$pt$-$hsp70$-$white$ transgene reporter (Wallrath and Elgin 1995).

**Insertion into the TAS on chromosomes 2R and 3R occurs at a higher frequency if additional 1360 elements are included in the reporter construct.**

One aspect of the insertion site distribution is different from that observed previously with the original $hsp26$-$pt$-$hsp70$-$white$ transgene reporter: enrichment at subtelomeric regions [also seen by (Sentmanat and Elgin 2012)]. Specifically, the mapping analysis reveals that 40% (12/30) of $P\{T1\}$ and 59% (31/53) of $P\{T4\}$ transgenes exhibiting variegating eye color map to the subtelomeric regions of chromosomes 2R and 3R (Figure 2). These insertions map to the TAS, subtelomeric arrays of Invader4 derived sequences (Figure 3; Supplemental Figure S2). This recovery of additional variegating inserts in the TAS for $P\{T4\}$ compared to $P\{T1\}$ indicates that the $P\{T4\}$ transgene shows enhanced targeting to TAS ($p=0.0060$, Pearson’s Chi-square test). To explore this, we mapped the relative locations of the $P\{T1\}$ and $P\{T4\}$ reporters onto an alignment of the chromosome 2R and 3R TAS, done with the published sequence for the X chromosome TAS (Karpen and Spradling 1992) and the consensus Invader4 LTR (Figure 3). All of our reporters are inserted in the internal tandem repeats, which have been identified previously as a hotspot for $P$ element insertion (Karpen and Spradling 1992). These data suggest that including additional copies of 1360 in the $P$ element reporter increases its preference for this insertion hotspot. Note that the 1360 element does not share sequence similarity with the Invader4 repeats, or with the $P$ element target site previously identified (Linheiro and Bergman 2008) (see Discussion).
Given that \( P\{T4\} \) does not contain any sequence elements that are not also present in \( P\{T1\} \), the enhanced targeting of \( P\{T4\} \) to the TAS is not DNA sequence dependent per se, but reflects the TE copy number, suggesting a dosage-dependent recognition process.

**The eye phenotype of most transgenes inserted into TAS is independent of 1360.**

Next, we examined the dependency of the eye color phenotype on the presence of 1360 sequences in the reporter \( P \) element construct. \( P\{T1\} \) and \( P\{T4\} \) were designed to allow for the removal of the 1360 sequences by the FLP recombinase. We crossed 81 variegating and 20 solid eye color insertion lines (41 \( P\{T1\} \) and 60 \( P\{T4\} \); Table 3) to flies containing FLP recombinase (for details of the cross see Supplemental Figure S3). Based on visual inspection, none of the red-eyed lines exhibited changes in eye pigment levels, and fly lines with a NRS eye phenotype similarly were unaffected (Table 3 and Supplemental Table S3). Of the variegating inserts tested, 27% of the \( P\{T1\} \) and 33% of the \( P\{T4\} \) lines exhibited suppression of the PEV phenotype (shown by an increase in eye pigment levels) when the 1360 element was excised, the expected response (Figure 4A). Most of these lines did not switch to a full red eye phenotype, but continued to exhibit some level of variegation (Figure 4B). However, one of the 30 \( P\{T1\} \) PEV lines tested (3%) and six of the 51 \( P\{T4\} \) PEV lines (12%) switched to an apparently uniform red eye. All reporters that respond to 1360 excision with increased eye pigment were located either in the pericentromeric regions or on the 4th chromosome (Supplemental Table S3).
In contrast to the effect observed in the pericentric region and on the 4th chromosome, none of the variegating insertions in the TAS regions were suppressed by excision of the 1360 element. In fact, six reporter lines inserted into TAS (plus one additional reporter outside of the TAS) exhibited a slight enhancement of PEV upon 1360 excision (Figure 4C), an effect that is generally stronger in males than in females. This finding suggests a potential role for 1360 in impeding heterochromatin formation in the TAS.

Insertion sites for enhanced and unaffected reporters are closely interspersed in the TAS and can even be in the same location relative to the consensus TAS repeat, as is the case for 9M1-827 and 9M1-912 (chromosome 3R: 27,899,586; Figure 3). However, due to the repetitive nature of the TAS, we cannot confirm whether the different lines actually have reporter insertions in the same repeat. Thus, the effects of removal of 1360 on the reporter phenotype are varied and context dependent, with reporters in the pericentric heterochromatin and fourth chromosome showing the anticipated loss of silencing, while reporters in the TAS showed no change or an enhancement of silencing. Clearly the mechanism of silencing in these domains must differ.

**TAS insertions have a different chromatin context than other variegating insertions.**

To investigate factors impacting variegation of a given reporter, we analyzed the surrounding environment, examining the enrichment of chromatin marks [histone modifications and chromosomal proteins; data from the modENCODE project (KHARCHENKO et al. 2010)], the repeat characteristics (class and family), the transcription start site (TSS) density, and the density of sequenced piRNAs and
endosiRNAs (BRENNECKE et al. 2007; YIN AND LIN 2007; CZECH et al. 2008; KAWAMURA et al. 2008; SAI TO et al. 2009; AMERES et al. 2010). For the analysis of chromatin marks we used available data from cell culture lines, adult heads, embryos, and 3rd instar larvae. The ideal situation would be to compare chromatin marks in the eye imaginal discs (which give rise to the eyes), but as these data are very difficult to obtain, 3rd instar larvae are often used as a proxy. A comparison of the pattern of key active and repressive marks seen in 3rd instar larvae with that in the BG3 cell line shows a strong correspondence between the two for the regions surrounding our reporters (Supplemental Figure 5, A-C; correlations range from 0.6 to 0.86). Examination of the relationships among all 113 data sets (chromatin/histone marks in the various tissues assayed by modENCODE) shows a strong correlation across the different tissue types for the mark recognized by a given antibody at the sites of our reporters: the average correlation for datasets for the same mark from different tissues is 0.77, while the average correlation across all datasets is 0.12 (Supplemental Figure 5D). Because of this strong correlation for a given mark among the datasets, we proceeded to examine the chromatin marks using all available genome-wide data; most of these data are from cell lines (i.e. BG3 and S2 cells).

Hierarchical clustering was used to group the variegating lines tested for the effect of 1360 excision based on the surrounding chromatin marks, and the additional data types were then mapped onto the resulting clusters (Figure 5). The TAS sites show a different pattern of histone modifications and chromatin marks compared to the sites in other repeat-rich regions of the genome that support variegation. TAS insertion sites
are enriched for marks typically associated with Polycomb regulation, including PC, SCE, PCL, E(Z), and H3K27me3, as well as for HP1a. This enrichment for Polycomb marks appears to be a general characteristic of the TAS regions (Andreyeva et al. 2005). It is much more pronounced at the 3rd chromosome TAS than the 2nd chromosome TAS, allowing for the separation of the 2R and 3R TAS insertion sites into different clusters. Additionally, there is some enrichment of H3K4me2 and HP1B at the chromosome 3R TAS.

Among the non-TAS insertion sites, the analysis identified several subclusters based on the chromatin marks. Most variegating reporters inserted into regions other than the TAS were found in chromatin characterized by marks typical of heterochromatin, including HP1a, H3K9me2, H3K9me3, and sometimes SU(VAR)3-9 (Figure 5). Examining the chromatin state of these locations [as defined by Kharchenko and colleagues (Kharchenko et al. 2010)], we find that these reporters fall into “state 7”, typical of pericentric heterochromatin. One exception to this pattern is a small subset of variegating reporters that cluster together on chromosome 2L (Supplemental Figure S4C). These reporters are located closer to TSSs, in particular Lamp1, and are enriched for marks of transcriptional activation such as RNA pol II, CHRO, HP1B, and H3K4me2, which puts these reporters into chromatin state 1, the typical state for promoters and TSS regions (Kharchenko et al. 2010). The chromatin context of these reporters is similar to that of some 4th chromosome variegating reporters, observed in BG3 cells to be in states 1 or 2 (associated with the TSS and transcription elongation, respectively) while embedded in regions resembling the pericentric heterochromatin.
However, the actual state of these sites in the critical cells for the eye phenotype (developing eye disc) is not known. Most of the variegating reporter locations are enriched for both piRNA sequences (piRNA defined by association with the PAZ/PIWI family of proteins PIWI, AUB, or AGO3) and endosiRNA sequences (siRNA defined by association with the Argonaute family member AGO2). piRNA and endosiRNA matching sequences are greatly enriched around the TAS insertion sites; the insertion sites in the pericentric regions and the 4th chromosome are also enriched for small RNA sequences, but to a lesser extent (Figure 5, panel 3). Thus, similar to the chromatin marks examined, the distribution of small RNAs indicates that the TAS insertion sites represent a heterochromatic domain distinct from pericentric heterochromatin. Many of the variegating lines are inserted into piRNA loci (Supplemental Table S3). It appears that the inclusion of 1360 in the construct biases toward insertion into clusters of piRNA sequences.

**Chromatin marks are a strong predictor of the variegating and red eye phenotypes.**

Given the different genomic distribution of red and variegating insertion sites, we extended the chromatin analysis to examine whether the chromatin marks could be used to predict eye phenotypes. The screen shots from the genome browser in Figure 6A and 6B illustrate the different chromatin environments in which the transgene insertions can be found. K-means clustering was used to assign each reporter (identified by the insertion site) into one of five clusters based on the 113 chromatin datasets (from tissue culture cells and different life stages). The percentage of
reporters with each eye phenotype (i.e. red, variegating, and NRS) within each of these clusters was then tabulated (Figure 6). The 20kb window size – compared to the 1kb window size used for Figure 5 – was chosen to obtain the best possible separation of red and variegating lines into distinct clusters. While variegating and red insertion lines show significantly different distributions among the clusters (Chi-square test p<0.0001), the differences in the distribution of the NRS and red insertion lines are not statistically significant (Pearson’s Chi-square test n.s.). The red-eyed lines comprise two main clusters: Cluster 1 insertions are in the typical “euchromatic chromatin,” enriched for marks associated with transcription, including H3K4me2, RNA pol II, and a variety of histone acetylation marks; Cluster 2 sites lack most marks examined and are reminiscent of state 9 identified by Kharchenko and colleagues (KHARCHENKO et al. 2010). The variegating inserts fall into three clusters: Cluster 4 sites represent typical heterochromatin, enriched for H3K9me2/3, HP1a, SU(VAR)3-9, and SU(VAR)3-7 (state 7). The large majority of these inserts are in heterochromatic “islands,” small regions highly enriched in repetitious sequences, near the pericentric heterochromatin at the base of chromosome arms 2L and 2R, but not contiguous with the pericentric heterochromatin, and thus not delineated as such in the modENCODE analysis (RIDDLE et al. 2011). This cluster also includes the lines with insertion sites in chromosome 4. The insertion sites in the TAS domains split into two clusters: Cluster 3 (chromosome 3R TAS) and cluster 5 (chromosome 2R TAS). Cluster 3 is strongly enriched for Polycomb family marks such as PC, H3K27me3, E(Z), and SCE, while in this larger window size (20kb compared to 1kb for Figure 5) the chromosome 2R TAS sites in Cluster 5 lack enrichment of these marks. A few variegating insertions do not follow this
general pattern, as they group with the red-eyed lines in clusters 1 or 2. This group includes three inserts that map to the euchromatic chromosome arms, and four inserts that are in repeat-rich domains, but whose chromatin marks are more “euchromatin-like” in the cell and tissue samples analyzed.

A Random Forest analysis of the chromatin marks surrounding each insertion site produces a classifier that can correctly predict the eye phenotype of 89% of the insertion sites. Similar to the observations from the K-means analysis, the Multi-Dimensional Scaling (MDS) plot of the proximity matrix shows that the red and variegating insertion sites form distinct clusters, while the NRS reporters are interspersed within both clusters (Figure 7A). Consistent with the MDS plot, the confusion matrix shows a much higher rate of misclassification for the NRS lines (0.71) compared to the red and variegating lines (0.04 and 0.06, respectively). The variable importance plot shows that the most important marks for the classification are H3K9me3, HP1a, and H3K23ac (Figure 7B). H3K9me3 and HP1a are well known marks enriched in heterochromatic domains, and here are associated with the variegating reporters. The importance of H3K23ac is not as obvious from the K-means analysis; however, it is depleted in the three clusters that contain the variegating insertion lines (Figure 6A-C). Thus, both the Random Forest analysis and the K-means clustering analysis indicate that in addition to the enrichment of H3K9me3 and HP1a, regions supporting PEV are also depleted for H3K23ac. Most importantly, using this Random Forest classifier, we can predict with high fidelity the likely eye phenotype of
reporter insertions at each position of the genome where the modENCODE data are available.

The increase in variegating reporter lines recovered using $P\{T4\}$ compared to $P\{T1\}$ is due to targeting of the reporter to genome regions supporting heterochromatin formation.

Mobilization of the $P\{T4\}$ construct resulted in a significant increase in the rate of recovery of variegating reporter lines relative to the $P\{T1\}$ construct (Pearson’s Chi-square test, $p<0.0001$). Two possible mechanisms could account for such a result: an enhanced ability to induce or promote heterochromatin formation in regions of the genome that normally do not form heterochromatin (i.e. heterochromatin “expansion”), or a bias for insertion into heterochromatic regions of the genome. In the first case (heterochromatin expansion), one would expect the genomic locations of variegating $P\{T4\}$ reporters to be more dispersed than the genomic locations of variegating $P\{T1\}$ reporters; $P\{T4\}$ reporters would be located further from known heterochromatic regions (pericentromeres, subtelomeres and 4th chromosome) and/or in regions of the genome with a lower overall repeat density. However, a comparison of the genomic distribution of insertion sites for variegating $P\{T1\}$ and $P\{T4\}$ reporters shows substantial overlap (Figure 2, Supplemental Table S3). In both cases more than 90% of the insertion sites resulting in variegation (93% for $P\{T1\}$ and 98% for $P\{T4\}$) are located in regions previously identified as supporting a variegating phenotype, including the subtelomeric (TAS) regions of chromosomes 2R and 3R, the 4th chromosome, and pericentromeric regions on the 2nd chromosome (CRYDERMAN et al. 1998).
Next, we compared the repeat density of the genomic regions surrounding the insertion sites, using multiple window sizes (Figure 8A and Supplemental Figure S5). The analysis shows that there is no significant difference between the \(P\{T1\}\) and \(P\{T4\}\) insertions in total repeat density (left panel, Figure 8A, Kruskal-Wallis chi-squared = 0.37473, df = 1, p-value = 0.5404). Using a 20kb window, we find that for \(P\{T1\}\), the insertion sites of 34/38 red-eyed lines map to regions with less than 15% total repeat density as do 45/51 red-eyed \(P\{T4\}\) lines. In contrast, 30/32 \(P\{T1\}\) lines and 50/51 \(P\{T4\}\) lines with a variegating eye phenotype have insertion sites that map to regions with greater than 15% total repeat density. Examining the density of individual repeat classes (Figure 8A), we find that the density of “simple repeats” in the sequences surrounding variegating reporters is lower than in the regions surrounding solid-eye reporters (Kruskal-Wallis chi-squared = 52.973, df =1, p-value = 3.382e-13). Comparing the \(P\{T1\}\) and \(P\{T4\}\) insertion sites, the density of “DNA transposon” and “retrotransposon” appears to be similar (Kruskal-Wallis chi-squared = 0.11198, df = 1, p-value = 0.7379 for “DNA transposon”, and Kruskal-Wallis chi-squared = 0.24701, df = 1, p-value = 0.6192 for “retrotransposon”). The density of “satellite repeats” (which is the classification of the TAS) is significantly greater surrounding the variegating inserts than surrounding the red or NRS inserts (Kruskal-Wallis chi-squared = 72.085, df = 1, p-value < 2.2e-16). However, the slightly higher density of satellite repeats associated with \(P\{T4\}\) variegating lines is not significantly different from the \(P\{T1\}\) variegating lines (Kruskal-Wallis test not significant p-value=0.1019; Figure 8A). Excluding insertions
mapping to the TAS, we find that the densities of all repetitive element classes are similar between \( P\{T1\} \) and \( P\{T4\} \) variegating lines (Figure 8B).

Taken together, these data indicate that while the variegating reporters are preferentially found in domains with a high repeat density, the regions surrounding the \( P\{T1\} \) and \( P\{T4\} \) reporters have similar repeat density. Thus, the increased recovery of variegating insertions obtained using the \( P\{T4\} \) reporter is not due to the establishment of heterochromatic domains at new sites in the genome. [Formation of “new heterochromatin” initially appeared to be a reasonable hypothesis, given that the total repeat density of the \( P\{T4\} \) reporter is >30% - within the range of the total repeat density in the regions flanking all variegating transgenes and the average for the fourth chromosome, which appears heterochromatic (Figure 8A).] Thus the alternative hypothesis becomes attractive: that the significantly higher frequency in the recovery of variegating lines observed for \( P\{T4\} \) (12% as compared to 4%, Figure 1B) can be attributed to an increased rate of insertion into high repeat density regions, specifically the 2R and 3R TAS regions.
DISCUSSION

To expand our knowledge of the factors controlling heterochromatin formation, we carried out an extensive genetic screen using an hsp70-white reporter construct containing 1360 sequences. 1360 is a remnant of the DNA transposon ProtopP, and has previously been shown to influence reporter PEV (Haynes et al. 2006; Sentmanat and Elgin 2012). PEV generally occurs if reporter transgenes are inserted into regions of high repeat density, leading us to explore the importance of repeat density for 1360-associated PEV and reporter silencing. Mobilization of transgenes with one copy or four copies of 1360 both resulted in a higher recovery rate of variegating inserts than that observed previously with a similar transgene that did not contain any 1360 sequence (~1%) (Wallrath and Elgin 1995). Including TE copies within the construct resulted in recovery of 3.7% variegating inserts with one copy [similar to what was seen by Sentmanat and colleagues (Sentmanat and Elgin 2012)] and 11.7% variegating inserts with four copies. However, with either transgene, we still recovered many red-eyed reporter lines with insertion sites distributed along the euchromatic arms, indicating that even the presence of four copies of 1360 is insufficient to induce de novo heterochromatin formation and induce PEV at all genomic locations, despite the fact that the local repeat density due to the 1360 sequences exceeds 30% (i.e. the average repeat density of the heterochromatic fourth chromosome). Analysis of the flanking genomic sequence context confirms that variegating transgenes are typically in regions of high repeat density (~20-40% in a 20kb window), but exceptions occur, where a red eye phenotype is observed despite high repeat density in the vicinity of the insertion site. We also identified three lines with a variegating eye phenotype from transgene
insertion into regions with very low repeat density (less than 5% in a 20kb window). Therefore, while repeat density is strongly correlated with the variegating phenotype, proximity to a repetitive sequence (e.g. a TE) is insufficient by itself to induce stable silencing of the reporter sufficient to give a variegating eye phenotype. Hence, other as yet unidentified features must play a role as well.

However, there is at least one region of the genome where the addition of an extra TE can tip the balance between heterochromatin and euchromatin. Sentmanat and Elgin have shown that for a modified version of \( P\{T1\} \) the silencing effect of 1360 at such locations seems to depend on the presence of two short sequences with similarity to piRNAs (Sentmanat and Elgin 2012). They favor a model where the piRNA pathway, relying on local low levels of transcription, silences the reporter transgene (Sentmanat and Elgin 2012). The suggested requirement for an embedded piRNA sequence is consistent with our conclusion that repeat density \textit{per se} is an insufficient predictor of transgene silencing. However, as with repeat density, piRNA sequence density by itself is also an insufficient predictor of transgene phenotypes (Figure 6, Supplemental Table S4), indicating that multiple conditions have to be met for stable heterochromatin to form, possibly including low levels of transcription coming from the transgene locus at the correct developmental time, when heterochromatin is being established. A model that depends on some transcription to achieve silencing, apparently by generating a transcript, which can hybridize with the available piRNAs to initiate silencing events, is well established in the yeast \textit{S. pombe} (Martienssen and Moazed 2015).
Mapping the insertion sites of variegating reporters recovered in our screen revealed that for both the \( P\{T1\} \) and the \( P\{T4\} \) transgenes the majority of the insertion sites are in regions of the genome previously described as supporting heterochromatin formation and PEV. These include chromosome 4 and subtelomeric (TAS) sequences, as well as the now more clearly defined pericentric heterochromatic “islands.” However, we identified very few variegating lines inserted into the centric heterochromatic regions recently defined by epigenomic analysis (RIDDLE \textit{et al.} 2011), suggesting that these domains are either inaccessible to the transposable element, or cause complete silencing, and so were not recovered in the screen. We did identify the histone gene cluster HIS-C on chromosome arm 2R as a region supporting PEV, providing further evidence that at least part of this domain is packaged as heterochromatin [for additional evidence, see (VAN STEENSEL \textit{et al.} 2001; NER \textit{et al.} 2002)]. Histone genes are very unusual as they are maintained in a silent state throughout most of the cell cycle, with the exception of S phase, when they are highly expressed (MARZLUFF AND DURONIO 2002). This expression pattern is reminiscent of what has been described for the centromeric repeats in both yeast and plants, where expression occurs from these repeats during S phase, a step that is necessary for heterochromatin formation at these sites (MAY \textit{et al.} 2005; CHEN \textit{et al.} 2008).

Notably absent from our reporter collection are variegating lines with insertion sites at the base of chromosome 3, despite the fact that variegating reporter lines with insertion sites in this region have been previously identified (WALLRATH AND ELGIN 1995; YAN \textit{et al.} 2002; MOSHKOVICH AND LEI 2010). Most pericentric variegating reporters we
identified on chromosome 2 fall into heterochromatic “islands” – repeat-rich regions close to, but not continuous with the pericentric heterochromatin. These “islands” of H3K9me- and HP1a-enriched chromatin have not been observed on chromosome 3; if they are specific targets for our reporter, this might explain our failure to recover variegating lines with insertion sites in the pericentric domain of chromosome 3. We also failed to recover lines with insertion sites mapping to the subtelomeres (TAS sequences) of chromosome arms 2L or 3L, despite previous reports of PEV in these regions (WALLRATH AND ELGIN 1995). In contrast, we observe a clear bias towards insertion into the subtelomeres of chromosomes 2R and 3R, where 40% and 59% of our insertion sites are located for \( P\{T1\} \) and \( P\{T4\} \) respectively. While the sequences at the subtelomeres of chromosomes 2L and 3L share some similarity to the 1.8kb X chromosome TAS, alignment of the published chromosome 2L TAS to the 1.8kb X chromosome TAS indicates that the overlap is in a region distinct from the alignment of the 2R/3R/Inv4LTR to the 1.8kb X TAS (Supplemental Figure S2A) (WALTER et al. 1995), and does not include the Invader4 LTR regions where most \( P\{T1\} \) and \( P\{T4\} \) insertions are found. One contributing factor may be the \( P \) element target sequence identified in previous studies (LINHEIRO AND BERGMAN 2008). However, the different TAS repeats showed no clear enrichment or depletion for this specific sequence motif, making it unlikely that a simple sequence-based targeting mechanism is responsible for the targeting bias we observe for the 1360-containing transgenes.

Our study also confirms that the TAS regions are distinct genomic domains. Multiple previous studies have shown that the PEV exhibited by reporters in subtelomeric
regions is distinct from the PEV exhibited by reporters in pericentric and/or 4th chromosome heterochromatin. Many of the known suppressors of PEV, such as mutations in HP1a [i.e. Su(var)205], do not have a dominant effect on variegation at the subtelomeres, leading to a separate name, TPE (Telomere Position Effect) being assigned. Instead, Phalke and colleagues find that variegation of reporters inserted into Invader4 LTR/TAS repeats depends on MT2, JIL-1 [encoded by Su(var)3-1], and SU(VAR)4-20, the histone methyltransferase responsible for generating H4K20 methylation (Phalke et al. 2009). Earlier reports examining variegating hsp70-white transgenes in the telomeric regions of 2L and 3R showed sensitivity to mutations in Su(z)2 and Psc, pointing to a role for the Polycomb system (Cryderman et al. 1999), as later confirmed by the modENCODE data (Kharchenko et al. 2010). In this study, we find that none of the subtelomeric reporters shows increases in eye pigment level upon excision of 1360, in contrast to the mixed responses observed for insertions in other domains, again demonstrating that the subtelomeric domains are distinct. This finding suggests that, while the transgene targets the 1360-tagged reporter element to the TAS for insertion, the repetitious 1360 sequences in the transgene do not add to heterochromatin formation in these regions. The observation that some TAS insertions show decreased eye pigment levels upon excision of the 1360 sequences implies that the presence of these sequences might actually impede silencing of the reporter in the TAS.

The chromatin analysis also supports the concept of the TAS as distinct domains. Hierarchical clustering reveals that the TAS-inserted reporters reside in a distinct
chromatin environment. They tend to show less enrichment for classical heterochromatin marks such as H3K9me2, H3K9me3, and HP1a, and instead tend to display marks generally associated with the Polycomb system of regulation. Interestingly, some classical heterochromatic marks are also present; H3K9me3 and HP1a are enriched at the TAS in 3\textsuperscript{rd} instar larvae. These findings are supported by a recent Proteomic analysis of isolated chromatin segments (PICH), which found SU(VAR)3-9 (one of the three H3K9 methyltransferases) along with TRL and PC (Polycomb system components) associated with the TAS sequences (Antao et al. 2012). As noted above, PEV studies have found that inserts into the TAS are generally unaffected by mutations in HP1a [i.e. Su(var)205 (Cryderman et al. 1999)], despite its presence at the TAS. A subset of Su(var)3-9 mutations, on the other hand, are known suppressors of TPE (Donaldson et al. 2002; Doheny et al. 2008), as are some of the components of the Polycomb system (Cryderman et al. 1999; Doheny et al. 2008). Together, these data suggest that the chromatin composition at the TAS is distinct from other heterochromatic regions in the Drosophila genome, and suggest a dual silencing mechanism, involving both HP1a/H3K9me3 and Pc/H3K27me3.

While the chromatin analysis successfully distinguished TAS insertion sites from insertion sites in other regions of the genome, and allowed successful prediction of eye phenotype based on the chromatin landscape of the insertion site, none of the marks examined allowed us to predict a reporter’s response to the removal of the 1360 sequences. Other than the uniform lack of response on 1360 excision from variegating reporters in TAS regions, there are no clear DNA sequence or chromatin characteristics
that demarcate the PEV inserts that are affected by 1360 excision from those that are unaffected. Our analysis included over 100 chromatin datasets as well as various small RNA datasets, none of which allowed us to demarcate affected and unaffected non-TAS variegating reporters. Possible explanations for this failure could be that the appropriate mark has not been included in the data sets available for analysis, or that the mark is subtly changing during fly development, and we have not examined the correct stage or tissue to detect a pattern. Alternatively, the difference in the effect of 1360 excision might be highly sensitive to subtle genetic background differences between reporter lines. While all crosses were preformed in parallel, it is impossible to monitor all of the chromosomes in every cross, and thus subtle differences likely exist between the reporter lines. Because the genomic locations of some variegating reporters that differ in 1360 impact are very close together (within 1bp), we favor the later explanation for the inability to identify a factor or set of factors to explain the differences observed between insertion lines. Variegation is a balance between two states and can be affected by over 150 suppressors and enhancers, any of which could affect the impact of 1360 excision, but could not be accounted for in this analysis.

Our analysis also shows that the increased frequency of PEV observed with \( P\{T1\} \) and \( P\{T4\} \) insertions can be attributed to targeting to repetitious regions, rather than a physical “expansion” of regions supporting heterochromatin formation. \( P\{T1\} \) and \( P\{T4\} \) insertion sites resulting in a variegating phenotype occupy the same regions of the genome; the additional copies of 1360 present in \( P\{T4\} \) do not alter the genomic distribution nor is the repeat density in the surrounding genomic sequences lower than
for \( P\{T1\} \) insertion sites. Instead, the increase in frequency of variegating inserts (11% in \( P\{T4\} \) versus 4% in \( P\{T1\} \)) is likely due to “targeting” the \( P \) element to heterochromatic domains, especially to the TAS regions, as there is a much larger fraction of all variegating \( P\{T4\} \) insertion sites in TAS repeats than is the case for the \( P\{T1\} \) variegating lines. Targeting of transgene constructs derived from mobile elements is common, although the targeting observed here is unusual as it is dependent on the “inner part” of the transgene, rather than the \( P \) element ends necessary for transposition (CRAIG 1997). Mobile elements usually have small target sites, specific sequence motifs that they recognize. \( P \) elements have a weak consensus motif, but insertions occur frequently at other sites as well. In addition, the \( P \) element has a well-known bias towards the 5’-ends of genes (SPRADLING et al. 1995), presumably because chromatin packaging at these sites allows greater accessibility. However, previous studies with a PRE containing \( P \) element construct have demonstrated that the sequences contained within the “inner part” of the transgene can impose a bias on the insertion profile of the transgene (KASSIS et al. 1992; KASSIS 2002), which appears to be the case with 1360 as well. In Arabidopsis thaliana, an endogenous TE has been reported to be specifically targeted to heterochromatic insertion sites (TSUKAHARA et al. 2012), but the mechanism(s) is unclear. In addition, recent work on hybrid dysgenesis in D. melanogaster has shown that new TE insertions in dysgenic flies with a compromised TE silencing system are biased towards localization in heterochromatin and/or piRNA clusters (KHURANA et al. 2011). Such a mechanism may be important for mounting a defense against an invading TE, incorporating it into piRNA generating sites so that it will be silenced by that system, potentially by post-transcriptional mechanisms.
as well as by heterochromatin packaging. The mechanism(s) leading to the observed bias is unknown, nor is it known if this bias operates under normal conditions or if it is restricted to the special circumstances present in dysgenic flies.

Thus, while there are precedents for TE targeting to specific genomic regions, how the presence of 1360 sequences in the P element transgene alters its targeting properties remains to be resolved. Several different models can be envisioned. A logical possibility is a protein-protein interaction model, whereby proteins bound to the 1360 sequence could interact with proteins at the TAS domains, thereby biasing the insertion into this region. This model implies that heterochromatin- or heterochromatin-interacting proteins bind the P element during transposition, presumably through recognition of the 1360 elements. Additional alternatives are a DNA-DNA interaction, DNA-RNA interaction, or protein-DNA interaction between these two genomic regions. A mechanism utilizing nucleic acid recognition (DNA-DNA or DNA-RNA base pairing) appears unattractive, given the lack of DNA sequence similarity in this case. Further studies of this interesting potential defense mechanism will be needed to distinguish among these possibilities.
MATERIALS AND METHODS

Transgene constructs

1360(Sentmanat and Elgin 2012) was selected for use in these constructs due to its high representation in siRNA libraries and due to the frequent recovery of a variegating transgene inserted into its proximity on the fourth chromosome [for details see (Haynes et al. 2006)]. The P element constructs P[FRT-1X 1360-FRT hsp70-white], referred to as P{T1} (Figure 1A), and P[FRT-4X 1360-FRT hsp70-white]), referred to as P{T4} (Figure 1A), are derived from transformation vector pA412 (also known as P[W]) from V. Pirrotta. 1360 (FlyBase, Release 5.32) was amplified from genomic DNA by PCR using primers recognizing unique flanking sequences (5’-CCC ACT GAT GAT ACA GCA AT-3’, 5’-CCG TGG TTT GAC TGT AGT TA-3’) cloned into the TOPO® TA cloning vector pCR® 2.1-TOPO (Invitrogen), and subcloned using EcoRI sites into pLITMUS28 (New England Biolabs, Ipswich, MA). 1360 fragments were isolated from the pLITMUS28 clone and allowed to self-ligate creating a mixture of monomers, multimers and circularized fragments. The reaction products were subsequently incubated with EcoRI-digested pFRT (a gift from E. Gracheva). The recovered clones carried a single copy or four tandem copies of 1360 between the FRT sites. The FRT-flanked 1360 sequence was inserted into pA412 (Wallrath and Elgin 1995) upstream of the hsp70-white reporter gene.

Fly stocks

All stocks were raised at 25°C with 70% humidity on cornmeal-sucrose media
supplemented with yeast (SHAFFER et al. 1994) unless otherwise stated. Stock w1118; snaSC0/SM6b, P[70FLP]7 (BL#6876) containing a copy of the FLP recombinase under the control of a heat-shock promoter, and stock w*; ry^{506} Sb^T P(Δ2-3)99B/TM6B, TbT containing the Δ2-3 P element transposase, were both obtained from the Bloomington Drosophila Stock Center (BL#1798).

**P element transgenesis and mobilization**

The generation of the T190 series of fly lines has been described previously (HAYNES et al. 2006). Drosophila embryo germline P-mediated transformation was carried out using standard methods (RUBIN AND SPRADLING 1982). In-house injections yielded four fly lines carrying P{T1} (T1-36C, -32E, -90E, -60F) and one fly line carrying P{T4} (T4-12E). Additional injections of P{T1} performed by Genetic Services (Cambridge, MA) generated all other T1-series fly stocks. This transformation yielded 32 additional insertion lines, 22 of which mapped to a single location in the fly genome (Supplemental Table S2 and S5).

The 9M series of Drosophila lines was generated by mobilizing P{T1} insertions T1-37M and T1-21M1 and P{T4} insertion T4-12E, all located on the X chromosome (Supplemental Figure S1). Females homozygous for the P{T1} or P{T4} transgene were crossed to w*; ry^{506}, Sb^T, P(Δ2-3)99B/TM6B, TbT males carrying the P element transposase gene (F0). Male progeny from this cross carrying the P element and the Δ2-3 transposase chromosome were crossed to y, w^{67c2}; net; Sb{sbd}; sv{apa-pol} females, a stock marked with recessive mutations on each chromosome to allow for mapping of
new $P$ element insertions recovered (F1). Male progeny which carried the $P(T1)$ or $P(T4)$ construct (determined by the presence of $w^+$), but lacked the $Sb^I$, $\Delta2-3$ chromosome were backcrossed to females of the $yw^{67c23}; net; Sb^{abd}, sv^{spa-pol}$ stock (F2), and the $P$ element was genetically mapped by scoring $hsp70$-white expression relative to the recessive markers in the offspring. Initially, all non-red eyed lines were retained in addition to the first 50 red-eyed lines recovered from $P(T1)$ and $P(T4)$. To establish stable stocks, insertion lines retained from the screen were crossed to appropriate marked chromosomes ($CyO$ for the 2\textsuperscript{nd}, $TM3$, $Sb^I$ for the 3\textsuperscript{rd}, $ci^D$ for the 4\textsuperscript{th}). Sibling progeny carrying the $P$ element over the marked chromosome were mated to generate homozygous stocks. If this cross produced offspring lacking the marked chromosome, the line was considered homozygous viable (Supplemental Table S3).

**Eye pigment analysis**

All eye phenotypes were assessed by crossing the $P$ element reporter stock males to $yw^{67c23}$ virgin females and examining the eye phenotype in 3-5 day old male and female offspring separately. Photos were taken of individual flies, and a quantitative measure of pigment level was obtained using acid-extracted eye pigment measured at OD\textsubscript{480} [adapted from (Khesin and Leibovitch 1978)]. For eye pigment assays, five flies taken at random were used per assay, with four biological replicates per genotype.

**Removal of 1360 from the transgene construct using FLP**

The main focus of this work was on insertion sites that lead to the variegating eye phenotype; a random subset of fly strains exhibiting red and non-red solid eye color was
analyzed as a control. The crossing scheme used to generate derivative lines lacking the 1360 portion of the transgene construct (-1360) is shown in Supplemental Figure S3. Males carrying a P{T1} or P{T4} insert were crossed to w\textsuperscript{1118}, sna\textsuperscript{Scy}/SM6b, P\{70FLP\}7 virgin females carrying a FLP transgene under the control of the heat-shock promoter hsp70 (Cross 1). The flies were mated for five days, transferring the parents to a new vial after each 24h period. To excise the 1360 sequences, FLP recombinase expression was induced by heat shock (37°C for 1h per day from day three until pupation, day seven); one 24h collection of progeny was not heat treated to generate sibling lines retaining the 1360 portion of the transgene (+1360) in the same genetic background. To generate stable lines and eliminate the FLP chromosome, adult males carrying the P{T1} or P{T4} reporter and the SM6b, P\{70FLP\}7 chromosome were crossed with virgin females from one of the following stocks depending upon the reporter location: y, w\textsuperscript{67c23}; T(2;3)ap\textsuperscript{xa}, ap\textsuperscript{xa}/CyO; + (2\textsuperscript{nd} chromosome insertions), y, w\textsuperscript{67c23}; T(2;3)ap\textsuperscript{xa}, ap\textsuperscript{xa}/+; TM3, Sb\textsuperscript{+} (3\textsuperscript{rd} chromosome insertions), or y, w\textsuperscript{67c23}, sv\textsuperscript{spa-pol}/In(4)ci\textsuperscript{D}, ci\textsuperscript{D} pan\textsuperscript{ciD} (4\textsuperscript{th} chromosome insertions) (Cross 2). Fifteen individual male progeny (ten from the heat shocked vials and five from the no heat shock vial) were selected to generate derivative lines and mated to virgin females from the same stock used in the previous cross (Cross 3). After mating, the adult male was removed from the vial and analyzed for the excision of 1360 by PCR (forward and reverse primers [F – CGT ATT GAG TCT GAG TGA GAC AGC GA; R – GTT TAG CTT GTT CAG CTG CGC TTG] as shown in Supplemental Figure S7A). For each individual transgene insertion, males from three to four -1360 lines and one to three +1360 lines were mated to virgin female siblings heterozygous for the P element/Balancer chromosome to
establish a stable stock (Cross 4a). Eye pigment analysis for +1360 and -1360 lines was carried out as described above.

**Data analysis: comparing +/- 1360 lines**

Data from males and females were analyzed separately. Mean and standard deviation of eye pigment level (OD$_{480}$) was calculated for all $P\{T1\}$ and $P\{T4\}$ insertions with +1360 and -1360 derivative lines. In addition, for each insertion site, a fold change (mean pigment level -1360 / mean pigment level +1360) was calculated. A line was considered affected by the loss of 1360 if the mean +/- the standard deviation of the -1360 lines was greater than (or less than) the mean +/- the standard deviation of the +1360 lines. For some insertions, the effect observed in the males and females differed (see Supplemental Table S3). In these cases, if the effect of removal of 1360 was greater than 2-fold in either sex, the line was considered affected, whereas if it was less than 2-fold it was considered unaffected. If the fold change was only greater than 2-fold in one sex, this was noted in Supplemental Table S3.

**Verification of 1360 integrity**

Genomic DNA was prepared from fly stocks containing a transgene insertion using standard methods (SAMBROOK AND RUSSELL 2001). For $P\{T1\}$ reporter lines, PCR was performed with the following primers (forward primer: 5’-CGTATTGAGTCTGAGTGAGACAGCGA-3’; reverse primer: 5’-GGTTAGCTTTCTGAGTTCAGACGCTTTG-3’; see Supplemental Figure S7). Products were analyzed by agarose gel electrophoresis to confirm the fragment size. For $P\{T4\}$
reporter lines, genomic DNA was digested with Clal, subjected to Southern blotting (SAMBROOK AND RUSSELL 2001), and probed with a $^{32}\text{P}$-labeled fragment of the reporter gene (for probe location see Supplemental Figure S7A). The blots were visualized by autoradiography, and the size of the fragment of the transgene containing the 1360 elements was estimated. Note that the white probe also hybridized to the endogenous $w^{67c23}$ locus. For results from $P\{T4\}$ reporter lines see Table S6.

Fine-mapping of new $P\{T1\}$ and $P\{T4\}$ insertions

Mapping of $P$ element inserts was carried out by using inverse PCR amplification of the genomic sequence flanking the 5' $P$ end of each insert as described in (SUN et al. 2004) with the following primers: 5'-AGA CGA AAT GAA CCA CTC GGA ACC-3' and 5'-CTT CGG CTA TCG ACG GGA CCA CCT TA-3'). The amplified DNA was sequenced and mapped by BLAST (ALTSCHUL et al. 1990) to the *D. melanogaster* genome (Release 5). If a single map location could be determined, site specific primers located in the genomic flanking sequences were designed and used in conjunction with primers from within the $P$ element to confirm the location of the insertion site.

Verification of single $P$ element insertion

Genomic DNA was prepared from fly stocks containing the transgene construct using standard methods, digested with Clal, subjected to Southern blotting (SAMBROOK AND RUSSELL 2001), and probed with a $^{32}\text{P}$-labeled fragment from the 3' end of the *white* gene (Supplemental Figure 3A). The blots were visualized by autoradiography and the
number of bands determined. The white probe also hybridizes to the endogenous \( w_{67c23} \) locus, providing an internal standard.

**Analysis of TAS sequences**

A multiple sequence alignment of the TAS sequences was generated using ClustalW2 (available online at [http://www.ebi.ac.uk/Tools/msa/clustalw2/] (CHENNA et al. 2003). The following parameters were used: DNA weight matrix: ClustalW; Gap Open: 25; Gap Extension: 0.5; Gap Distances: 5; No End Gaps: False; Iteration: tree; Numiter: 1; Clustering: NJ. The sequences used for the alignment were 2R TAS_1.001kb - 2R: 21,145,570-21,146,570; 3R TAS_984bp - 3R: 27,898,888-27,899,871; INVADER4 LTR_346bp [from Repbase (JURKA 2000)], and the HETRP_DM_1.872kb satellite sequence (from Repbase). The HETRP_DM sequence is also known as the X-TAS 1.8kb repeat; it is identical to the 1.8B repeat described by Karpen and Spradling (KARPEN AND SPRADLING 1992). Note that in the Drosophila genome reference sequence there is one additional partial repeat at the 2R TAS - 2R: 21,146,571-21,146,708 and seven additional repeats (two partial and five full-length) at the 3R TAS - 3R: 27,898,670-27,898,887; 27,899,872-27,900,855; 27,900,856-27,901,839; 27,901,840-27,902,823; 27,902,824-27,903,807; 27,903,808-27,904,791; 27,904,792-27,905,053. In addition, there are several regions within the unmapped scaffolds (U and UExtra) in the *D. melanogaster* release 5 assembly that have sequences very similar to the 3R TAS_984bp repeat.

**Genomics analysis**
Identifying Drosophila transposon fragments with RepeatMasker

RepeatMasker [version open-3.2.8; (SMIT et al. 1996-2010)] was run at the most sensitive settings (-s) using the cross match search engine (version 1.090518; http://www.repeatmasker.org) and the Drosophila repeat library from Repbase [release 14.09; (JURKA 2000)] as the library to identify repeats present in the flanking genomic sequences adjacent to the \( P\{T1\} \) and \( P\{T4\} \) insertions of interest (window size: 1kb, 5kb, 10kb, or 20kb). As there were no large differences observed in total repeat density between the different window sizes (see Supplemental Figure S5 for 1kb, 5kb, and 10kb), a 20kb window was used in the detailed analysis.

Chromatin analysis

Data sets produced by the modENCODE project “Genome-Wide Mapping of Chromosomal Proteins in Drosophila” (G.Karpen, PI) were utilized in the chromatin analysis (KHARCHENKO et al. 2010). Data used initially were from S2 cells, 2-4h and 14-16h embryos, 3rd instar larvae, and fly heads. As datasets from different cell types showed a strong positive correlation for a given mark in the regions analyzed (Supplemental Figure S6), we included all available genome-wide data. The supplemental figures show the data from all the tissue types available, while the figures in the main paper only include a subset of data for each antibody. All data sets are listed in Supplemental Table S7.

TSS
Gene annotations from FlyBase 5.29 were extracted from the pre-computed files at the FlyBase FTP site (ftp://ftp.flybase.net/releases/FB2010_06/dmel_r5.29/). The distance and count calculations for each 9M insert were generated against three different types of features: the subset of genes with annotated untranslated regions (UTRs), all mRNAs, and all gene spans. Because not all of the genes in *D. melanogaster* have annotated UTRs, the start site of each feature may not correspond to the actual transcription start site of the gene. Consequently, the subset of genes with annotated UTRs is the most specific among the feature types analyzed. Additionally, since some genes have more than one isoform, and therefore more than one TSS, we calculated the distance relative to each isoform and to only the first isoform of each gene (i.e. the isoform that had the shortest distance from the insertion site). The results from this analysis are summarized in Supplemental Table S4. In addition to calculating the distance to the closest TSS, we also calculated the number of TSS instances (referred to as TSS density) within a 1kb, 5kb, 10kb, and 20kb window surrounding each insert using the three feature types described above (also in Supplemental Table S4).

**Small RNAs**

Analyses of the distribution of small RNA sites surrounding each insert were limited to the small RNA libraries that have been recovered by immunoprecipitation with specific argonaute family proteins. The ten small RNA libraries used in this analysis are available through the NCBI Gene Expression Omnibus (GEO) database under the following accession numbers: GSM154620, GSM154621, GSM154622, GSM231091, GSM266765, GSM280086, GSM280087, GSM280088, GSM378200, GSM466489.
The genomic region surrounding each insert (in 1kb, 5kb, 10kb, or 20kb windows) was extracted from the Release 5 *D. melanogaster* genome assembly. Custom Perl scripts were used to identify exact matches (in both strands) between the sequences in each small RNA library and the regions surrounding the inserts. For each region, we tabulated the total number of bases covered by small RNAs from each library and used this information to derive the "small RNA library coverage" for each region (i.e. number of bases covered by reads in the small RNA library / total size of region). The results from this analysis are summarized in Supplemental Table S4.

**Random Forest analysis**

Random Forest analysis (R package randomForest, available at http://cran.r-project.org/web/packages/randomForest/index.html; A. Liaw and M. Wiener 2002) was used with the genome-wide chromatin mapping data sets (mean M-values from all the replicates) for each 9M line to predict eye phenotypes. We performed the classification with the following parameters: formula=eye_phenotype ~ .; ntree=5000; mtry=10; proximity=true. A large number of trees were used (ntree = 5000) to ensure that the variable importance and proximity values were stable. We decided to use the default value (i.e. square root of the number of variables) for the number of variables sampled at each split (mtry = 10) because additional trials using mtry values of 5 and 20 did not affect the out-of-bag error estimate (Supplemental Table S8). The proximity parameter is set to “true” to generate a matrix of proximity measures from the Random Forest analysis. Based on this matrix, we generated a multi-dimensional scaling plot with the MDSplot function. The variable importance dot chart was generated using the
varImpPlot function with default parameters. The R script used to perform the Random Forest analysis is available on our lab website (http://www.biology.wustl.edu/faculty/elgin/scripts.html).
ACKNOWLEDGEMENT

We would like to thank KA Haynes and L Collins for constructing the P{T4} plasmid. We would also like to thank J Wuller, A Wu, and M Dothager for their help in the genetic screen.

This work was supported by NIH grants R01 GM68388 and GM073190 to Washington University (for S.C.R.E.) and by a Kirschstein National Service fellowship F32 GM078833 to KLH. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences, or the NIH.
REFERENCES


Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson et al., 2003 Multiple sequence alignment with the Clustal series of programs. Nucleic acids research 31: 3497-3500.


Fagegaltier, D., A. Bougé, B. Berry, E. Poisot, O. Sismeiro et al., 2009 The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. Proceedings of the National Academy of Sciences.
Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhoult, J. Kind et al., 2010
Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. Cell.


May, B. P., Z. B. Lippman, Y. Fang, D. L. Spector and R. A. Martienssen, 2005


### Table 1. Mobilization summary.

<table>
<thead>
<tr>
<th>Transgene</th>
<th># of crosses</th>
<th># of mobilizations (F1 generation)</th>
<th># red</th>
<th># non-red, solid (F2 generation)</th>
<th># PEV</th>
<th>no progeny</th>
<th>% mobilized</th>
<th>% of total mobilizations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P{T1}$</td>
<td>2,500</td>
<td>1,019</td>
<td>931</td>
<td>12</td>
<td>38</td>
<td>38</td>
<td>40.7%</td>
<td>91.4%</td>
</tr>
<tr>
<td>$P{T4}$</td>
<td>2,600</td>
<td>673</td>
<td>533</td>
<td>12</td>
<td>72</td>
<td>56</td>
<td>25.9%</td>
<td>79.2%</td>
</tr>
</tbody>
</table>
Table 2. Distribution of insertion lines exhibiting PEV by chromosome.

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Total # of PEV lines</th>
<th>PEV lines on chr. 2</th>
<th>PEV lines on chr. 3</th>
<th>PEV lines on chr. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P{T1}$</td>
<td>38 (3.7%)</td>
<td>17 (1.7%)</td>
<td>12</td>
<td>8 (0.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 (0.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.2%)</td>
</tr>
<tr>
<td>$P{T4}$</td>
<td>72 (10.7%)</td>
<td>43 (6.4%)</td>
<td>23</td>
<td>6 (0.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3.4%)</td>
</tr>
</tbody>
</table>
Table 3. Effect on eye phenotype upon 1360 excision.

<table>
<thead>
<tr>
<th>Transgene &amp; Parental Eye Phenotype</th>
<th>PEV Suppressed</th>
<th>Unaffected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P{T1}</td>
<td>8</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>PEV</td>
<td>8</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>NRS</td>
<td>NA</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Red</td>
<td>NA</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P{T4}</td>
<td>17</td>
<td>43</td>
<td>60</td>
</tr>
<tr>
<td>PEV</td>
<td>17</td>
<td>34</td>
<td>51</td>
</tr>
<tr>
<td>NRS</td>
<td>NA</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Red</td>
<td>NA</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

NA – not applicable

PEV – variegating eye color

NRS – non-red, solid eye color
FIGURE LEGENDS

Figure 1. Additional copies of 1360 in the hsp70-white reporter increase the recovery of insertions with a variegating eye phenotype.

A) Diagram of the $P\{T1\}$ and $P\{T4\}$ element reporters. $P\{T1\}$ contains one copy of 1360$\{1503$ (yellow box, pointed end indicating the orientation of the canonical 1360 ORF), accounting for 12.5% of the total construct, upstream of the hsp70-white reporter (red box), while $P\{T4\}$ contains four copies of 1360$\{1503$ in the opposite orientation, accounting for 34.1% of the total construct. The triangles flanking the 1360 sequences represent FRT sites that can be used for the removal of the 1360 element by FLP recombinase. The bent arrow indicates the direction of transcription for the hsp70-white transgene, and the hashed boxes at either end of the transgenes represent $P$ element ends necessary for transposition.

B) Charts summarizing the eye phenotypes of fly lines recovered from the mobilization of the reporter constructs shown in A. Lines are grouped by phenotypes; those with a solid eye color phenotype (red and NRS) are represented by solid red and those with a variegating eye phenotype (PEV) are represented by speckled red. Note that, in contrast with Table 1, these numbers exclude lines that did not produce progeny in the F2. Left: $P\{T1\}$. Right: $P\{T4\}$.

Figure 2. Additional 1360 copies in the reporter $P$ element result in an increased frequency of lines inserted into subtelomeric (TAS) and pericentric regions of the genome.
Map of the *D. melanogaster* genome showing the location of the insertion sites. The top map shows the $P\{T1\}$ insertion sites; the bottom shows the $P\{T4\}$ insertion sites. The color of the triangle corresponds to the eye color of the fly line (see key at bottom of figure). Inserts in the histone gene cluster are placed at the 5’ end of the full repeat cluster (i.e. 2L: 21,421,980-21,425,660), despite the fact that it is unknown which repeat copy they reside in. Centromeres are depicted as grey circles. Eye images are included for a select number of lines. $P\{T1\}$ inserts are labeled with the “9M1” prefix, whereas $P\{T4\}$ inserts have the “9M4” prefix.

**Figure 3. All insertion sites in the subtelomeric regions of chromosomes 2R and 3R are clustered in the TAS tandem internal repeats.**

ClustalW (THOMPSON *et al.* 1994) alignment of chromosome 2R TAS_1.001kb, chromosome 3R TAS_984bp, Invader4 LTR_346bp, and HetRP DM_1.872kb satellite sequences showing the locations of the $P$ element insertion sites in the TAS. $P\{T1\}$ elements are shown in orange/red, and $P\{T4\}$ elements are shown in blue/green, the first color indicating reporters on chromosome 2 and the second indicating reporters on chromosome 3. The internal 173/173/161bp tandem repeats from the X-TAS 1.8kb repeat are indicated by solid blue lines. Note that all the insertion sites fall within these repeats. The shaded boxes mark insertions for which PEV is enhanced upon excision of the 1360 element. The alignment background color reflects the level of sequence conservation: yellow – all four sequences are identical; olive – three sequences are identical; purple – two sequences are identical.
Figure 4. Removal of 1360 from the reporter suppresses PEV for approximately 1/3 of the reporters but can enhance PEV at the TAS.

A) Bar chart comparing the fraction of reporter lines in the three phenotypic classes \([P\{T1\} \text{ or } P\{T4\}\) with variegating (PEV), non-red solid (NRS), or red eyes (Red)] which exhibited suppression (or no suppression) of variegation upon excision of the 1360 fragment.

B) Images and quantitative pigment assay data for a sample of variegating insertion lines illustrating the suppression of PEV (increase in eye pigment level) upon removal of 1360. Cytological locations of inserts are indicated in parentheses and the exact chromosomal location of each insertion site can be found in Supplemental Table S3. Eye pigment levels are reported as OD_{480} (Y-axis) separately for males and females, with and without 1360 (X-axis). Error bars are standard error (n=4).

C) Images and quantitative pigment assay data for selected variegating insertion lines that exhibited enhancement of PEV upon removal of 1360. See B for details. Eye pigment levels for lines with and without 1360 are significantly different for all lines in males, but only in 9M4-340 for females (p<0.05, t-test).

Figure 5. TAS insertions are located in a chromatin environment distinct from pericentric heterochromatin.

Heat map illustrating the hierarchical clustering of enrichment levels for select chromatin marks within a 1kb window surrounding the \(P\) element insertion site for all variegating reporter lines tested for suppression of variegation upon 1360 excision (panel 1; enrichment – shades of red; depletion – shades of blue). Additional panels to the right
show the average repeat density of transposable elements (panel 2; darker shades indicate higher density), the average small RNA density (panel 3; darker shades indicate higher density), and the average number of TSS (panel 4; darker shades indicate higher number) within a 1kb window of the insertion sites. The chromosome 2R and 3R TAS inserts cluster together and show the same lack of impact of 1360 excision (no suppression of variegation). Pericentric inserts do not split into affected and unaffected groups based on the chromatin marks examined. Supplemental Figure S8 shows a similar analysis using a larger number of ChIP datasets. Chromatin sources: S2, BG3, KC – cell lines; EE – 2-4h embryos; LE – 14-16h embryos; L – third instar larvae; H – fly heads. Colors above the clustering diagram: red for marks associated with heterochromatin, green for marks associated with active transcription, and brown for marks associated with the Polycomb system.

Figure 6. Chromatin marks, repeat density, and small RNA profiles distinguish variegating inserts from those with a solid red-eye phenotype.

A) Screen shot of genomic region 2L: ~19,800,000-21,500,000 showing enrichment levels for select active (green; H3K4me2, H3K23ac) and repressive (red; H3K9me2, H3K9me3) histone modifications and HP1a from BG3 cells. Reporter insertion sites in this region are marked below (red eye – red; variegating eye – blue), as are genes and TEs. The BG3 cell chromatin states are shown in the bottom track using the color scheme of Kharchenko et al (KHARCHENKO et al. 2010). TEs (DNA TEs and retrotransposons) were identified using Repbase (JURKA 2000). X-axis: position in bp.

B) As in A) but for region 2R: ~1,800,000-2,650,000.
C) K-means clustering results of using modENCODE enrichment data for the 20kb window surrounding each reporter element insertion site. Five clusters were chosen as this gave the best separation between lines with different eye phenotypes (red vs. variegating) without breaking each phenotype into subclusters. The red eye phenotype predominates in Cluster 1 - “euchromatic chromatin” – and cluster 2 - lacks most marks examined. The variegating eye phenotype predominates in Cluster 4 - “pericentric and fourth chromosome heterochromatin;” cluster 3 - chromosome 3R TAS; and cluster 5 - chromosome 2R TAS. Additional panels on the right show the percentage of each cluster within each eye phenotype category, the number of inserts on each chromosome, the number of inserts in each chromatin state (KHARCHENKO et al. 2010), the average number of TSSs, the average density of TEs, and the average small RNA density within a 20kb window of the insertion sites. Also shown is the percentage of insertions in the cluster located within a piRNA cluster (BRENNECKE et al. 2007). See Supplemental Figure S9 for a version showing the results for each insertion site. Colors below the clustering diagram: red for marks associated with heterochromatin, green for marks associated with active transcription, and brown for marks associated with the Polycomb system. The number of red (red), variegating (blue), and NRS insertion lines (orange) in each cluster are show to the right of the heat map.

Figure 7. The random forest ensemble classifier can be used to identify important factors that distinguish insertion lines with a red eye phenotype from those with a variegating eye phenotype.
A) Assessment of the random forest classifier using the Multi-Dimensional Scaling (MDS) plot for the proximity matrix shows two distinct clusters that correspond to the red and variegating eye phenotypes. However, NRS lines are scattered throughout the MDS plot. The MDS plot results are consistent with the confusion matrix generated by the random forest algorithm which shows a much higher classification error on NRS lines (0.71) than either the red (0.04) or variegating lines (0.06).

B) The random forest classifier can be used to explain most of the variations observed in enrichment levels of chromatin marks between phenotypic classes [20kb window], with an out of bag (OOB) error rate estimate of 11.35%. The variable importance dot chart shows the top 30 chromosomal proteins and histone modifications (Y-axis) that are most important to the eye phenotype classification (as measured by the mean decrease Gini; X-axis). The most important variables for the eye phenotype classifications are H3K9me3, HP1a, and H3K23ac.

Figure 8. *P{T1}* and *P{T4}* variegating reporters are inserted into genomic regions of similar high repeat density, while most red-eyed reporters are located in regions of low repeat density.

Display convention for the box plots: the box denotes the interquartile range (IQR; Q1 to Q3). The line within the box denotes the median. The whiskers correspond to Q1 - 1.5 x IQR and Q3 + 1.5 x IQR, respectively. The dots outside the whiskers denote outliers.

A) Box plots showing the total repeat density and repeat density segregated by repeat type for a 20kb window surrounding all transgene insertion sites. See Supplemental Figure S5 for 1kb, 5kb, and 10kb window comparisons. Separate box plots are shown.
for each reporter category ($P\{T1\}$ or $P\{T4\}$) lines with non-red solid eye color, red eyes, or variegating eye color, here and in B. For “Total Repeat Density”, “Simple Repeat Density”, “TAS/Satellite Density”, and “DNA TE density” the lines with variegating eye color are significantly different from both lines with red and NRS eye color, in both $P\{T1\}$ and $P\{T4\}$ lines ($p<0.05$, Kruskal-Wallis test). For “Retrotransposon Density”, $P\{T1\}$ lines with red and variegating eye color are significantly different from each other ($p<0.05$, Kruskal-Wallis test), while none of the comparisons are significant for $P\{T4\}$ lines. The $P\{T1\}$ and $P\{T4\}$ reporters respond to their environment similarly; no significant differences between these reporters are observed for any of the categories ($p>0.05$, Kruskal-Wallis test).

**B) Box plots showing the TE density for all insertion sites outside of the TAS.** The graph on the far left shows the total TE density, while the other graphs show repeat density broken down into subcategories (DNA transposons, retrotransposons, LINE elements, and LTR elements). A 20kb window was used. For $P\{T1\}$ lines, lines with variegating eye color are significantly different from both lines with red and NRS eye color for “TE Density”, “Retrotransposon Density”, “LINE TE Density”, and “LTR TE Density”, while only the comparison between lines with variegating and red eye color is significantly different for “DNA TE Density” ($p<0.05$, Kruskal-Wallis test). For $P\{T4\}$ lines, lines with variegating eye color are significantly different from both lines with red and NRS eye color for “TE Density”, and “DNA TE Density”, while only the comparisons between lines with variegating and red eye color are significantly different for “DNA TE Density”, “LINE TE Density”, and “LTR TE Density” ($p<0.05$, Kruskal-Wallis test).
Figure 1
Eye Phenotype: ▼ = Solid Red ▼ = Solid Orange/Yellow ▼ = Variegating
Figure 3
Figure 4

(A) Impact of 1360 FLP

(B) Suppressed PEV lines

(C) Enhanced PEV lines
Figure 5
Figure 6
Figure 7
Figure 8