### Title
P Transposable Elements in Drosophila and other Eukaryotic Organisms.




### Journal
Microbiology Spectrum, 3(2)

### ISSN
2165-0497



### Authors
Majumdar, Sharmistha
Rio, Donald C



### Publication Date
2015-04-02

### DOI
10.1128/microbiolspec.mdna3-0004-2014

Peer reviewed

HHS Public Access

Author manuscript

*Microbiol Spectr*. Author manuscript; available in PMC 2015 July 01.

Published in final edited form as:

*Microbiol Spectr*. 2015 April ; 3(2): MDNA3–0004-2014. doi:10.1128/microbiolspec.MDNA3-0004-2014.

# P transposable elements in *Drosophila* and other eukaryotic organisms

**Sharmistha Majumdar**[*] and **Donald C. Rio**[#]

Department of Molecular and Cell Biology University of California, Berkeley Berkeley, CA 94720-3204

## Abstract

P transposable elements were discovered in Drosophila as the causative agents of a syndrome of genetic traits called hybrid dysgenesis. Hybrid dysgenesis exhibits a unique pattern of maternal inheritance linked to the germline-specific small RNA piwi-interacting (piRNA) pathway. The use of P transposable elements as vectors for gene transfer and as genetic tools revolutionized the field of Drosophila molecular genetics. P element transposons have served as a useful model to investigate mechanisms of cut-and-paste transposition in eukaryotes. Biochemical studies have revealed new and unexpected insights into how eukaryotic DNA-based transposons are mobilized. For example, the P element transposase makes unusual 17nt-3' extended double-strand DNA breaks at the transposon termini and uses guanosine triphosphate (GTP) as a cofactor to promote synapsis of the two transposon ends early in the transposition pathway. The N-terminal DNA binding domain of the P element transposase, called a THAP domain, contains a $C_2CH$ zinc-coordinating motif and is the founding member of a large family of animal-specific site-specific DNA binding proteins. Over the past decade genome sequencing efforts have revealed the presence of P element-like transposable elements or P element transposase-like genes (called THAP9) in many eukaryotic genomes, including vertebrates, such as primates including humans, zebrafish and Xenopus, as well as the human parasite *Trichomonas vaginalis*, the sea squirt *Ciona*, sea urchin and hydra. Surprisingly, the human and zebrafish P element transposase-related THAP9 genes promote transposition of the *Drosophila* P element transposon DNA in human and *Drosophila* cells, indicating that the THAP9 genes encode active P element "transposase" proteins.

## Introduction

P transposable elements are one of the best-studied eukaryotic mobile DNA elements in metazoans. These elements were initially discovered in the late 1960's because they cause a syndrome of genetic traits termed hybrid dysgenesis [1]. The molecular cloning and biochemical characterization of the P element transposition reaction have led to general insights regarding eukaryotic cut-and-paste-transposition. P elements have also facilitated many applications as genetic tools for molecular genetics in *Drosophila*.

[#]Corresponding author Phone: 510-642-1071 FAX: 510-642-6062 don_rio@berkeley.edu.

[*]Present address: Department of Biological Engineering, Indian Institute of Technology-Gandhinagar, Ahmedabad 382424, India sharmistham@iitgn.ac.in

This review will focus on the experiments and results that have taken place over the past decade related to understanding the mechanism, distribution, specificity and uses of P element transposition. It will also discuss studies aimed at providing insights into how P element transposition is controlled, how P elements rely upon their host cells to provide the functions necessary to aid their successful mobility and how the damage they incur to the genomes in which they reside is limited and repaired. For reviews about other aspects of P element biology, more historical perspectives on the invasion of P elements into *Drosophila* and the genetic inheritance patterns of hybrid dysgenesis and P cytotype regulation, the reader is referred to earlier review articles [2-10] and the previous review article in this ASM series [11].

## I. P transposable elements in *Drosophila* and other eukaryotic genomes

**I.A. Structure and protein products of P transposable elements—**Isolation and DNA sequence analysis of the P elements showed a wide variation in size in a typical P strain, from 0.5 –2.9kb [12]. P strains carried 50-60 P element insertions, of which approximately one-third were full-length (2.9kb) whereas the remainder were internally-deleted in different ways [13]. All P elements have a canonical structure that includes 31bp terminal inverted repeats (TIR) and internal inverted repeats (IIR) of 11bp located about 100bp from the ends that interact with the THAP domain of the transposase [14] (Fig. 1A). Between these two repeats, but distant from the site of DNA cleavage, are high-affinity binding sites for the P element transposase protein THAP DNA binding domain (see below; Fig. 1A; [15]).

The largest active, complete, P element analyzed is 2907 bp in length and contains four non-contiguous open reading frames [12] (Fig. 1B). All four open reading frames or exons are required for P element mobility and encode the 87kD transposase protein [16, 17]. These open reading frames are functionally joined together at the RNA level via pre-mRNA splicing [18]. The expression of transposase is normally restricted to germline cells because splicing of the P element third intron (IVS3) only occurs in the germline. In addition, it was shown that in somatic cells (and also to a large extent in the germline) the third intron (IVS3) is retained, producing a functional mRNA that encodes a 66kD protein [17]. The 66kD protein functions as a repressor of transposition and has been termed a Type I repressor [19-22] (Fig. 1B). In addition, some truncated proteins produced from the smaller, internally-deleted P elements present in natural P strains or engineered can act as repressors of transposition and are called Type II repressors [5, 14, 23-25]. In summary, P element transposition *in vivo* requires about 150bp of DNA in *cis* at each transposon end, including 31bp terminal inverted repeats and the high-affinity transposase binding sites. Upon insertion P elements create an 8bp duplication of target DNA and the elements can encode a transposase, as well as DNA binding repressors of transposition.

### I.B. P elements, P element-related THAP9 genes and active P element “transposase-like” genes in other eukaryotic genomes

Sequencing of the human genome revealed the presence of ~ 50 genes that were derived from DNA transposable elements [26]. One of these genes, termed THAP9, bears homology to the *Drosophila* P element transposase (Fig. 2 and [27, 28]). Further genome sequencing

efforts over the past decade have led to the discovery of P element-related THAP9 genes or transposons in a variety of organisms in addition to humans, including other primates, zebrafish, *Xenopus* [29], *Ciona* [30], sea urchin, hydra [31] and the human pathogenic protozoan parasite, *Trichomonas vaginalis* [32]. Notably, the THAP9 gene is absent from rodents, rats and mice, and a defective copy is found in the chicken genome [33].

The human THAP9 gene is homologous (25% identical and 40% similar) to the *Drosophila* P element transposase throughout the entire length of the protein [27, 33]. The discovery of the human THAP9 gene suggests a possible invasion of P element-like transposons into vertebrates. The THAP9 gene appears most recently functional as a transposase in zebrafish (called Pdre2; [33], where there are obvious inverted repeat elements (called Pdre elements; [33, 34]) carrying 13bp terminal inverted repeats (TIRs) and 12bp sub-terminal inverted repeats (STIRs) with a ~ 400bp spacer and which carry direct 8bp target site duplications flanking the TIRs (Fig. 3A). In addition to the full-length Pdre2 THAP9 transposase-like gene, there are also multiple internally-deleted Pdre elements elsewhere in the zebrafish genome reminiscent of *Drosophila* P strains [33, 34] .

Regarding the activity of the vertebrate THAP9 genes, it has recently been shown that the human THAP9 protein can mobilize *Drosophila* P elements in both *Drosophila* and human cells (Fig. 3B and 3C) [35]. These results indicate that human THAP9 is an active DNA recombinase that retains the catalytic activity to mobilize P transposable elements across species. However, the cellular function of human THAP9 is still unknown. It may be the case that the human THAP9 gene could encode a recombinase that acts on remote recombination signals elsewhere in the human genome, similar to Rad1/2 activity in V(D)J recombination. Gene expression profiling indicates that the human THAP9 gene is highly expressed in embryonic stem cells, testes and kidney.

In addition to THAP9, the human genome has 11 THAP domain-containing genes ([28]; Fig. 2). Many have been characterized as transcription factors that control the expression of genes involved in apoptosis, cell cycle regulation, stem cell pluripotency and epigenetic gene silencing. Human THAP domain family members have also been implicated in a variety of human diseases, including heart disease, torsional dystonia and cancer. The THAP DNA binding domain appears to be restricted to animals because no known or predicted THAP domain-containing genes have been found in plants, yeast, other fungi or bacteria.

### I.C. P elements as tools for *Drosophila* genetics and the *Drosophila* genome project

One of the most important uses of P elements since their discovery has been P-element-mediated germ line transformation. The method makes use of the fact that P elements normally only transpose in germline cells and that a P element carrying a foreign gene can be mobilized in *trans* using a source of transposase [36-38](Fig. 4). The Berkeley Drosophila genome project (BDGP) has collectively developed strategies which efficiently use single P element mutagenesis [39, 40], leading to large-scale P element insertional mutation screens so that now 9440 or about two-thirds of the annotated protein-coding genes are tagged [41-43]. It was discovered that normally P elements transpose preferentially in *cis*, within about 50 –100kb from their initial location, so-called “local hopping” [44, 45]. P elements can also undergo transposase-mediated excision and transposase-induced male

recombination to generate deletions flanking an existing P element insertion [46]. These smaller deletions can be enlarged by performing P element excision crosses in a DmBLM/mus309 mutant genetic background (defective for repair of P element transposase-induced DNA double-strand breaks; more details in section II.F) [47-49]. P elements have also played a role in other molecular genetic methods, such as homologous gene targeting [50] and in the use of libraries of bacterial artificial chromosomes (bac) for transformation and recombineering [36, 38, 51]. Thus, P elements have continued to play important roles in the post-genome sequence era of *Drosophila* genetics and genomics.

## II. Mechanism of P element transposition

### II.A. *Cis*-acting DNA sites involved in *Drosophila* P element transposition

Experiments using P element-mediated transformation showed that the 31bp inverted repeats were required for transposition (Fig. 1A) [16, 52]. Extensive mutagenesis studies subsequently showed that both the terminal 31bp inverted repeats and the high affinity internal transposase binding sites are also required for transposition *in vivo* [53]. Additionally, the internal 11bp inverted repeats located at ~120bp from each transposon end function as transpositional enhancer elements (Fig. 1A) [53]. These 11bp internal inverted repeats can interact with the N-terminal THAP DNA binding domain of the P element transposase (Fig. 1A) [54]. Analysis of hybrid elements carrying tandem 5' and 3' ends showed that two different P element ends (a 5' and 3' end) must be paired for transposition to occur, indicating that the two P element ends are not equivalent [53]. In summary, the *cis*-acting DNA sites for P element transposition include 31bp terminal inverted repeats, 11bp sub-terminal inverted repeats and internal 10bp transposase binding sites, located between the terminal inverted repeats and sub-terminal inverted repeats. Interestingly, the zebrafish Pdre P-like elements also have both terminal and sub-terminal inverted repeats (see section, I.B., above and Fig. 3A). All P elements analyzed to date in *Drosophila*, as well as P element-related transposons in zebrafish, create 8bp direct duplications of target site DNA upon insertion [33, 55, 56].

### II.B. The *Drosophila* P element transposase

Understanding the domain organization of the P element transposase had come from biochemical studies, genetic experiments and sequence comparisons of P elements from other *Drosophila* species and other eukaryotes with P element transposase-like THAP9 genes. The N-terminal THAP DNA binding domain is a $C_2CH$ zinc binding motif with an adjacent basic region and is the site-specific DNA binding domain (Fig. 5A; see section II.D., below) that recognizes the internal transposase DNA binding sites and the 11bp sub-terminal inverted repeats [15, 54]. Adjacent to the THAP domain is a long coiled-coil region made up of a canonical leucine-zipper motif with heptad leucine/isoleucine repeats and an adjacent longer coiled-coil motif (up to residue 221; Fig. 5A). This type of coiled-coil region in the transposase protein, commonly found in THAP domain-containing proteins, allows protein dimerization, although this is not essential for high affinity site-specific DNA-protein recognition [54, 57].

Biochemical studies revealed a requirement for GTP as a cofactor during P element transposition [58], suggesting that the transposase would bind GTP. The central region of the transposase protein contains several sequence motifs found in the GTPase protein superfamily [59, 60] (Fig. 5A and 5B). Mutation of some of the key conserved residues abolished GTP binding *in vitro* and transposase activity *in vivo* [61]. Alteration of the most conserved NKXD guanine recognition motif to the NKXN motif xanthine recognition motif (D379N) altered the purine nucleotide requirement for transposase activity from guanosine to xanthosine triphosphate, both *in vitro* and *in vivo*, indicating that purine nucleotide cofactor binding is required for transposase activity [61] (Fig. 5B). It is know known that GTP plays a key role in synapsis of the two transposon ends during transposition [62, 63] (see section II.E., below).

The carboxyl-terminal region of the transposase protein, including exon 4 (ORF3), contains a high proportion of acidic amino acid residues (Fig. 5C). Mechanistically, the P element transposase belongs to the superfamily of polynucleotidyl transferases, including the transposases from bacteriophage Mu and other bacterial mobile elements, the retroviral integrases, the Holliday junction nuclease RuvC, the RNaseH superfamily [64, 65] and the RAG1 subunit of the V(D)J recombinase [66-68]. These enzymes generally use metal ion-mediated catalysis of phosphodiester bond hydrolysis and formation, where acidic amino acids in the protein active site serve to coordinate a divalent metal ion, usually magnesium [65, 69]. Because the P element-encoded 66kD repressor protein lacks catalytic activity and because of the high proportion of acidic amino acid residues in the C-terminal part of the protein, it seemed likely that acidic residues found in this region might constitute the catalytic domain of the protein. Mutagenesis studies of many acidic C-terminal residues have failed to identify a set of clear catalytic residues (D. Rio and colleagues, unpublished results). However, sequence- and structure-based (Phyre2) [70] alignments suggest a relationship to the hermes transposase, with putative RNaseH secondary structures and catalytic residues between the THAP domain and GTP binding regions at the N-terminal portion of the protein [64, 65, 71]. It may be the case that the GTP binding domain of the *Drosophila* P element transposase was inserted into a progenitor RNaseH-like catalytic domain [64, 65]. Recent exhaustive bioinformatics analyses, using all known P element-related protein sequences from the sequenced eukaryotic genomes in the Repbase database, revealed a conserved D, D, E-like motif common to all P element-transposase-related proteins known [71], including the THAP9 family (Fig. 5C). Thus, it appears that the P element transposase is a member of the DDE enzyme superfamily with a complex metal ion binding site configuration in the active site.

### II.C. Biochemical characterization of P element transposase

Purification and characterization of the P element transposase protein from *Drosophila* tissue culture cell nuclear extracts showed that the 87kD transposase protein binds to 10bp sites near each end of the P element, located between the terminal 31bp and internal 11bp inverted repeats [15] (Fig. 1A). The consensus transposase binding site is: 5’- AT(A/C)CACTTAA -3’. The two high-affinity transposase binding sites do not overlap the terminal 31bp inverted repeats which are known to be required for transposition *in vivo* [16, 53] (Fig. 1A). Following initial recognition of these binding sites by transposase, a GTP-

dependent assembly (or synapsis) occurs to bring the two transposon ends together [62, 63] (see section II.E., below and Fig. 6B).

The purified P element transposase protein was used to develop *in vitro* assays to study the different steps of the transposition reaction: donor DNA cleavage and target DNA integration [58]. The *in vitro* reaction required wild type transposase DNA binding sites on the donor P element and linear pre-cleaved donor DNA could be used as a substrate, indicating that supercoiling of the donor DNA is not required for transposition. Importantly, 3' hydroxyl groups at the P element ends were required for activity indicating that, like other transposition systems, a 3'-hydroxyl group is used as a nucleophile during strand transfer [58]. These studies demonstrated that P element transposition proceeds via a cut-and-paste mechanism.

Most surprisingly, the nucleoside triphosphate guanosine triphosphate (GTP) was discovered as a critical cofactor for P element transposition [58]. Non-hydrolyzable GTP analogs (GTP-γ-S, GMP-PNP and GMP-PCP) showed levels of activity equivalent to those observed with normal GTP, indicating that hydrolysis of a high energy phosphoryl bond was not required for activity. This suggests that GTP plays an allosteric role by binding to P element transposase, in the same way that GTP modulates the conformation and activities of other GTP-binding proteins, such as ras and mammalian Gα subunits [59, 72] or the GAD family member dynamin [73]. The GTP requirement for the P element transposase protein is unique among this class of polynucleotidyl transferase proteins. Single-molecule imaging studies showed that GTP plays a critical role in initial synapsis of the transposon ends ([62, 63]; see section II.E., below).

Physical assays for the detection and analysis of reaction products and intermediates were developed using both radiolabeled and unlabeled P element DNAs leading to the detection of an excised P element transposon fragment, directly confirming that transposition occurs through double-strand DNA breaks at the transposon termini [74]. DNA cleavage mapping experiments showed that P elements are excised from the donor DNA site by an unusual set of cleavages at the transposon ends [74]. The 3' ends (bottom strands) of the transposon DNA are cleaved at the junction with the 8bp target site duplication (Fig. 6A). But surprisingly, the 5' ends (top strands) were cleaved 17nt into the 31bp inverted repeats generating novel 17nt 3' single-strand extensions on both the excised transposon and flanking donor DNAs (Fig. 6A). These transposase-mediated *in vitro* DNA cleavage sites are consistent with previously characterized *in vivo* P element excision products [47, 75]. This long 3' single-strand extension may facilitate entry of cleaved donor DNA sites into DNA repair pathways and could contribute to the irreversibility of the P element excision reaction.

Further studies of P element transposase protein *in vitro* used synthetic oligonucleotide substrates. First, these studies showed that the 17nt single-stranded 3' extension is critical for strand transfer because a decrease in the length or mutation of the exposed single-stranded DNA extension caused a drastic reduction in strand transfer activity [76]. Second, chemically-modified oligonucleotide substrates were used to probe the protein-DNA contacts that were critical for strand transfer activity by P element transposase [76]. These experiments demonstrated that critical contacts between transposase and both the duplex and

single-stranded regions of the substrate DNA are necessary for strand transfer activity *in vitro* [76]. In addition, there were sites in which DNA modification actually stimulates strand transfer, indicating that distortion of the substrate DNA may facilitate the chemistry of strand transfer, such as those observed with Mu transposase [77, 78] and HIV integrase [79]. Thus, while the initial donor DNA cleavage reaction occurs on duplex DNA, the strand transfer reaction uses a completely different substrate in which the single-stranded region at the P element ends are critical for activity indicating that there is a significant active site flexibility in the P element transposase protein during transposition.

Using oligonucleotide substrates that mimic a strand transfer intermediate carrying a pre-cleaved P element end joined to target DNA, the purified transposase protein can carry out disintegration [76], in a similar manner to the retroviral integrases [80], Mu transposase [81] and two other eukaryotic recombinases, the *C. elegans* Tc1 transposase [82] and the V(D)J RAG1/2 recombinase [83]. The ability of these proteins to perform disintegration may be significant in genomic surveillance because these disintegration reactions may serve to prevent chromosomal translocations and other rearrangements that might ensue following aberrant transposition events [83].

## II.D. Protein-DNA recognition by the THAP family of $C_2CH$-zinc-coordinating DNA binding domains

THAP domains are a recently described family of zinc-coordinating DNA binding motifs, first recognized in the P element transposase [28]. The THAP domain is 80-90 amino acid residues, typically located at the amino terminus of the protein and contains a $C_2CH$ (consensus: Cys-$X_{2–4}$-Cys-$X_{35–50}$-Cys-$X_2$-His) zinc-coordinating motif and other signature elements, including a C-terminal AVPTIF sequence. Although the THAP domain of THAP proteins share low primary sequence identity, recent structural studies have shown that there is strong conservation of the overall β−α−β protein fold and secondary structure elements. Structures of the P element transposase bound to its DNA site using X-ray diffraction [57] and of the human THAP1 protein bound to its DNA site using NMR [84] have shown that THAP domains recognize their DNA sites in a bipartite manner, using two β-strands in the major groove and a basic C-terminal loop in the adjacent minor groove.

The X-ray crystal structure of the P element transposase THAP domain (DmTHAP) bound to DNA (Fig. 7 and 8) shows that His18 and Gln42 from the two β-strands at the N-terminus of DmTHAP, make a total of six direct contacts with DNA bases in the major groove of DNA and engage both strands of the DNA duplex (Fig. 7 and 8). Notably, the residues making the most central contacts with the DNA, including water-mediated hydrogen bonds to specific DNA bases in the major groove, show little (His18) or no (Gln42) sequence conservation within the human THAP protein family (Fig. 7C; [57]). Presumably, sequence variability at these positions, along with differences in the length and amino acid composition of the N-terminus, specifies the precise DNA sequences recognized by the THAP proteins through the major groove sub-site [57]. The variability of residue Gln42 (Fig. 7C), which interacts with several bases via multiple hydrogen bonds is unusual, since amino acid residues involved in multiple base-specific interactions are typically less variable [57]. This observation correlates well with the finding that the most conserved THAP

domain residues play structural roles, namely all the invariant residues besides the $C_2CH$ zinc-binding motif are involved in forming the hydrophobic core of the protein (Fig. 7A and 7C).

In addition to major groove DNA contacts, there are multiple residues in the loop 4 region of DmTHAP that are involved in adjacent minor groove DNA interactions (Fig. 7C and 8A and 8B). The residues corresponding to R65, R66 and R67 in DmTHAP, which are both contacting DNA (R65 and R67) and positioning the loop 4 region of the protein (R66), are different in human THAP9 (FK*R65*R*R67*LN in DmTHAP and GI*R*R*K*LK in human THAP9 (Fig. 7C). The minor groove binding residues may modulate DNA binding affinity, since the DmTHAP domain with its RRR motif has a higher affinity for its site than human THAP1, which has NKL in the corresponding position (Fig. 7C). The higher affinity DNA binding by the *Drosophila* P element transposase THAP domain compared to other THAP domain proteins [57], may have some role in the high frequency of P element mobility in *Drosophila*.

### II.E. A GTP-dependent protein-DNA assembly pathway for *Drosophila* P element transposition

One of the unique and unexpected discoveries concerning P element transposition is its requirement for GTP as a cofactor. Biochemical studies indicated that GTP was not required for site-specific DNA binding by the P element transposase and that the purified transposase from *Drosophila* cells was a tetramer, that was unaffected by the presence or absence of GTP or DNA. Attempts to detect transposase-DNA complexes using native gel electrophoresis proved unsuccessful. However, single-molecule imaging using atomic force microscopy (AFM) of tranposase-DNA complexes has led to important insights into the role of GTP in the P element transposition pathway. First, GTP promotes formation of a stable synaptic complex between the transposase tetramer and the P element DNA (Fig. 6B; [63]). Second, time course reactions revealed the presence of both synaptic complexes and cleaved donor DNA complexes. These imaging experiments showed that while synapsis was fast (0-30 min.), cleavage of the donor DNA was slow (hours) and took place in a random, non-concerted manner in which one transposon end was cleaved and then later the second end was cleaved [62, 63]. Third, in the absence of GTP the tetrameric transposase only bound to one transposon end (Fig. 6C; [62]) and upon addition of GTP or non-hydrolyzable GTP analogs synaptic complexes formed [62, 63]. Thus, GTP promotes synapsis of the two transposon ends, a critical step in the P element transposition pathway, possibly by reorienting one of the THAP domains in the transposase tetramer.

## II.E. Target site selection in *Drosophila* P element transposition

All of the P elements that were analyzed initially [12] and subsequently, including a recent set of 2266 insertions from the Berkeley *Drosophila* Genome Project (BDGP) showed that an 8bp duplication of target DNA occurs when P elements integrate [85]. Bioinformatic analysis of this data and additional data from the *Drosophila* genome project showed that the insertion sites tended to be GC-rich and, in fact, that there was a symmetric palindromic pattern of 14bp centered on the 8bp target duplication found at the insertion sites with a

conserved consensus motif (Fig. 9) [55, 56, 85]. This pattern would make sense if the recognition of target site DNA required binding of two (or an even number of) subunits of the transposase protein in a synaptic complex juxtaposing the two P element ends with the insertion sites in the target DNA.

It was found that in cells P elements normally can transpose preferentially in *cis*, within about 50 –100kb from the initial location, so-called “local hopping” [44, 45]. It has also been noted that P elements tend to insert near the 5’ ends of genes, near promoters, and a correlation was made to sites of binding of the DNA replication factor ORC [86], but this may simply reflect an open chromatin organization at *Drosophila* gene promoters.

### II.F. DNA repair pathways involved in P element transposition in *Drosophila*

Excision of a P element results in a double-strand DNA break, which can then be repaired via a gap repair process using a homologous chromosome or sister chromatid as a template [58, 87]. It was also shown that ectopic templates could be used for gap repair with about 30bp of homology and a dramatic preference for template use in *cis*, on the same chromosome, was observed [88-90]. Incomplete copying can also explain how internally-deleted P elements are generated in natural P strains. The generation of internally-deleted products are consistent with a model for the gap repair process occurring via a synthesis-dependent strand annealing (SDSA) mechanism [91, 92] first described for replication-linked recombination in bacteriophage T4 [93]. This template-directed DNA repair following P element excision suggests a way for P elements to increase in copy number since the gap repair process restores a new copy of the P element at the donor site, while the excised element transposes to a new location (Fig. 10).

Molecular and genetic studies have allowed testing of the involvement of different DNA repair pathways in the repair of P element-induced or other types of DNA breaks. The mus309 gene encodes the *Drosophila* homolog of the Bloom's DNA repair helicase (DmBLM) [94]. *Mus309* mutants are sterile and mutagen-sensitive [95]. They are also defective for repair of P element transposase-induced DNA double-strand breaks [47, 48, 91, 92, 94]. Other studies have shown that *mus309* mutants are defective for repair by the homology-directed SDSA pathway [91] and that the defects of *mus309* mutants can be rescued by transgenes encoding Ku70, a subunit of the heterodimeric Ku complex involved in the non-homologous end joining (NHEJ) pathway [47, 94]. Using RNA interference and plasmid-based assays, similar conclusions were reached regarding an interplay or competition between these two repair pathways [96]. More detailed genetic analyses have indicated differential requirements for DmBLM during development [97] and that in the absence of DmBLM there is an increase in the use of other DNA repair pathways, such as single-strand annealing (SSA) [98] or nonhomologous end-joining (NHEJ) can then be used [99-101]. More recently, microhomology-dependent pathways for repair have been revealed in Drosophila [102]. A series of studies using transgenic reporters to detect alternate DNA repair pathways using either P element or I-Sce I nuclease-induced DNA breaks showed that different repair pathways can be used to repair these double-strand DNA breaks [99, 100]. Interestingly, when one DNA repair pathway is disrupted, others compensate to provide a

means of genome stability. Thus, P elements are capable of efficiently using cellular DNA repair pathways.

## III. Regulation of P element transposition in *Drosophila*

### III.A. Hybrid dysgenesis, P cytotype, the piwi-interacting (piRNA) small RNA pathway, adaptation and paramutation

Hybrid dysgenesis is the term used to describe a collection of symptoms including high rates of sterility, mutation induction, male recombination (which does not normally occur in *Drosophila*) and chromosomal abnormalities and rearrangements [2-8, 11]. Hybrid dysgenesis and its associated sterility and mutation induction are normally only observed in the germlines, but not in somatic tissues, of progeny from crosses in which males carrying P transposable elements (termed P or paternally-contributing strains) are mated to females that lack autonomously mobile P elements (termed M or maternally-contributing strains) (Fig. 11). Initial studies on the reciprocal cross effect in hybrid dysgenic crosses led to the description of two regulatory states based on genetic crosses between P and M strains. M strains were said to have an M cytotype state, which was permissive for P element mobility, whereas P strains were said to have P cytotype, a state which is restrictive for P element movement [103, 104]. These reciprocal cross experiments followed the repressive effect of P strain females for several generations [103, 105, 106]. The segregation pattern of P cytotype exhibited some aspects of strict maternal inheritance in that the cytotype of the great grandmother played a role in determining whether subsequent progeny displayed M or P cytotype. Studies of the regulatory effects of P cytotype have used several different genetic assays (Table I), which can differ in their tissue specificity and whether or not they directly assay P element mobility.

The study of one repressor-producing P strain derived from a wild population, called *Lk-P*(1A), has been particularly informative regarding the mechanism and genetics of P cytotype regulation [107]. *Lk-P*(1A) contains two full-length P elements near the telomere of the X chromosome at cytological position 1A, which lie in inverted orientation separated by ~5kb; these elements are integrated into sub-telomeric heterochromatic repeat sequences known as *TAS* repeats (for Telomere-Associated Sequences) [108, 109], which have been shown to be hotspots for P element insertion [110]. *Lk-P*(1A) exhibits a very strong P cytotype repressive effect characteristic of the regulatory properties observed with normal P strains. Studies on *Lk-P*(1A) used either of two genetic tests for P element cytotype control, gonadal dysgenic (GD) sterility or the *singed-weak* (*sn*$^{w}$) hypermutability assays (Table I) [108]. *Lk-P*(1A) displayed a P cytotype repression effect equal to that observed with several natural P stains carrying twenty to thirty times the number of P elements [108, 111]. The repressive properties of *Lk-P*(1A) were found to be strong in germline tissues and were maternally transmitted, but repression was observed only weakly in somatic tissues [108]. By contrast, a natural P strain, such as Harwich, displays the repressive properties of P cytotype strongly in both somatic and germline tissues [108]. More recent studies on Lk-P(1A), as well as other telomeric P element insertions [112-115], has shown that there is a connection to the maternal inheritance of P cytotype, the germline piRNA pathway [116-122] and the heterochromatin protein Su(var)205 (HP1) [121-123].

One of the most unusual features of hybrid dysgenesis is the multigenerational inheritance of the P element-repressive state known as P cytotype. One important connection that was made in this regard came from the analysis of small RNAs, known as piwi-interacting RNAs or piRNAs, that co-purified with the *Drosophila* germline-specific Argonaute family members, piwi, aubergine and Ago-3 [117, 124]. The application of small RNA cloning and high-throughput sequencing showed that these proteins bound pools of small RNAs that were derived largely from transposable elements, including P elements. These findings illuminated how sites in the genome, such as the flamenco locus which contains an endogenous I factor polyA retrotransposon, can act as piRNA-generating loci that generate large amounts of piRNAs that silence endogenous or exogenous I factor transposons. It is thought that small piRNAs, complementary to endogenous P element mRNA, are transmitted to P strain oocytes and serve to silence P elements introduced by P strain sperm [117]. The piRNA system also functions to control transposon mobility in the mouse germline [125]. More recent studies have revealed connections of piRNAs to deposition of repressive histone chromatin marks and reduced levels of gene expression [126].

One distinguishing feature of P strains isolated from the wild is the presence of 50-60 P element copies, about a third of which are full-length 2.9kb P elements. This evolution of a P strain has been recapitulated in population cages after transformation of M strains with single P elements, where the generation of P cytotype correlates with the accumulation of ~ 50 P elements [13, 127, 128]. More recently, high-throughput sequencing was used to follow restoration of fertility in sterile P-M dysgenic hybrids [129]. This restoration of fertility correlated with insertion of transposons into piRNA clusters resulting in the production of piRNAs directed to the P element transcript. Interestingly, this study also showed that P elements and other transposons were mobilized during hybrid dysgenesis. Thus, hybrid dysgenesis results in transient transposon activation and then insertion of elements at new locations leads to transposon silencing by the piRNA pathway and a concomitant restoration of fertility [129]. These findings can explain how the P elements in wild populations may have spread so quickly in nature.

Paramutation is defined as an epigenetic interaction between two alleles of a locus, through which one allele induces a heritable change in the other allele without modifying the DNA sequence. In the case of P elements, P cytotype could be induced by telomeric P elements or clusters of P element transgenes to cause a homology-dependent silencing termed the trans-silencing effect (TSE; [107]). Recent studies using genetic crosses and high-throughput sequencing of small RNAs showed that clusters of P element-derived transgenes can convert other, previously inactive homologous transgene clusters into strong mediators of TSE [130]. Interestingly, this TSE can be transmitted through 50 generations and occurs without any chromosome pairing between the paramutagenic and paramutated chromosomal loci [130]. This multigenerational paramutational effect is mediated by inheritance of maternal cytoplasm carrying piRNAs homologous to the P element transgenes and requires the aubergine gene product, which is involved in piRNA biogenesis, but not Dicer-2 which is involved in siRNA production. This landmark study provides a genetic basis for the multigenerational inheritance of P cytotype.

### III.B. Control of the tissue-specificity of *Drosophila* P element transposition by alternative RNA splicing

One of the most startling findings regarding the regulation of P element transposition was the discovery that alternative pre-mRNA splicing was responsible for restricting expression of the P element transposase, and hence the entire syndrome of traits associated with hybrid dysgenesis, to germline cells [18]. It is now known that premRNA splicing is a widely used regulatory mechanism for expanding proteomic diversity and increased splicing is correlated with organismal complexity [131]. The basic biochemical mechanism of pre-mRNA intron removal by the spliceosome is conserved in eukaryotic cells and alternative pre-mRNA splicing is mediated by RNA regulatory motifs called enhancers or silencers, which can be located in either introns or exons [132]. The tissue-specific splicing of the P element third intron (IVS3) has served as an important model system for investigating how alternative RNA splicing patterns are generated in distinct cell or tissue types and has led to detailed characterization of the first exonic splicing silencer (ESE) upstream of the P element third intron [133] (Fig. 12A).

Two lines of investigation of IVS3 splicing led to the conclusion that this control involves an inhibition of IVS3 splicing in somatic cells. First, a molecular genetic approach showed that mutations in the 5’ exon, upstream from IVS3, caused an activation of IVS3 splicing in somatic cells [18, 134, 135]. Specific sequence changes in the 5’ exon activated IVS3 splicing in somatic cells [134]. Second, a set of biochemical experiments showed that while IVS3 splicing occurred in mammalian cell splicing extracts, it was not observed in the *Drosophila* somatic cell extracts [136]. Titration of 5’ exon RNA into the somatic cell extract activated IVS3 splicing, suggesting that the action of *trans*-acting factors led to the inhibitory effect observed [136, 137]. It was noted that both IVS3 and the 5’ exon had a number of 5’ splice site-like sequences and that these pseudo-5’ splice sites might play a role in regulating IVS3 splicing [136, 137] (Fig. 12A). Another study using mammalian splicing extracts showed that mutations in the 5’ exon known to activate IVS3 splicing in *Drosophila*, also activated IVS3 splicing *in vitro*, suggesting the possible conservation of components [138]. Indeed, subsequent studies using *in vitro* splicing assays with *Drosophila* extracts showed that IVS3 splicing was activated by mutations in the 5’ exon [137]. RNA-protein interaction studies defined two elements in the 5’ exon termed F1 and F2, both of which bear sequence identity to 5’ splice sites and contain a reiterated sequence motif (AGNUUAAG) [137]. Mutations in the F1 and F2 sites that activate splicing *in vitro*, inhibit RNA-protein complex formation in nuclear extracts [137]. The F1 site binds U1 snRNP [137] and the F2 site binds hrp48 [139] (Fig.12B), a *Drosophila* hnRNP protein similar to mammalian hnRNP A1 [140]. Interestingly, hnRNP A1 causes use of distal (upstream) 5’ splice sites *in vitro* [141] and binding site selection data indicate it binds to sites resembling 5’ splice sites [142]. These results led to a model in which RNA-protein and -snRNP interactions in the 5’ exon exonic splicing silencer (ESS) cause an inhibition of IVS3 splicing by blocking access of U1 snRNP to the normal IVS3 5’ splice site [137] (Fig. 12A).

Subsequent biochemical studies identified additional proteins that interact with the IVS3 5’ exon exonic splicing silencer. Proteins of 97kD and 50 kD proteins were identified as PSI (P element somatic inhibitor) and hrp48, respectively (Fig. 12B). PSI contains four KH-type

RNA binding motifs and two C-terminal direct ~100 amino acid repeats [143]. Conserved residues in these repeat motifs are involved in direct interaction between PSI and the U1 snRNP 70K protein [144, 145]. A similar motifs are found in a mammalian alternative splicing factor called KH-type splicing regulatory protein (KSRP) and a related family of proteins called FUSE-binding proteins (FBPs) [146]. PSI interacts with the IVS3 5' exon RNA, but not with heterologous RNAs [143]. Preferred RNA binding sites and RNA binding by the individual PSI KH domains have been examined [147, 148]. PSI protein is expressed highly in somatic cells, but at low or undetectable levels in the female germline [143]. The hrp48 protein specifically binds to the F2 element in the 5' exon [139] and is expressed in both the germline and soma [143]. Hrp48 contains two N-terminal RNP-CS type RNA binding domains and a C-terminal glycine-rich domain [140] (Fig. 12B). This structure is characteristic for this class of hnRNP proteins, termed 2XRBD-GLY, of which mammalian hnRNP A1 is a member [149]. Interestingly, these glycine-rich low complexity protein domains have been implicated in human diseases [150]. The expression pattern of hrp48 in both germline and soma suggests that hrp48 might also play a role in the inhibition of IVS3 splicing in the germline, where IVS3 is known to be inefficiently spliced [151]. Our model for splicing repression is consistent with the general idea that RNA binding proteins can direct spliceosome components to specific sites in pre-mRNAs to generate differential splicing patterns [132, 133].

Molecular genetic and genomic studies have addressed the roles of PSI and hrp48 in IVS3 splicing control *in vivo*. First, ectopic expression of PSI in germline cells caused a modest reduction in IVS3 splicing [152]. Anti-sense hammerhead ribozyme targeting of PSI mRNA in somatic cells resulted in activation of IVS3 splicing in the soma [152]. Both experiments suggest a role for PSI in the reduction of IVS3 splicing. Reduction of hrp48 levels in somatic cells using hypomorphic P element insertion alleles caused a small activation of IVS3 splicing in the soma [153]. These experiments also showed that hrp48 was encoded by an essential gene, and therefore must have additional functions involving cellular mRNAs. Genetic analysis of the PSI gene showed it to be essential and that the AB-repeat domain which interacts with U1 snRNP was essential for male fertility and normal courtship and mating behavior [154], suggesting that PSI must regulate RNA processing of other transcripts in somatic and male germline tissues. More recently, characterization of the general role of hrp48 in alternative splicing has been investigated using splice junction microarrays and RNA immunopurification procedures to demonstrate a general role for hrp48 as a splicing repressor protein [155, 156]. These studies indicate how effectively the P element has made use of cellular RNA binding proteins to control the tissue-specificity of P element RNA processing.

## IV. Summary and conclusion

P elements invaded *Drosophila* in the early 20$^{th}$ century and spread rapidly through wild populations to avert the deleterious effects of hybrid dysgenesis. The creation of mutational insertions allowed the molecular cloning of P elements and their use as tools for *Drosophila* molecular genetics. Perhaps the most unexpected finding regarding P elements comes from the genome sequencing efforts of the past decade and the realization that P element-like genes or transposons (THAP9) exist in a variety of other animals. The THAP DNA binding

domain, found initially to be at the N-terminus of the P element transposase, is now one of the most common animal-specific zinc-coordinating site-specific DNA binding domains. Finally, the unique role of GTP in the P element reaction pathway, as an allosteric effector that promotes synapsis of the transposon ends shed light on how the complex protein-DNA assembly of the P element transpososome is initiated. We anticipate that much more is to be learned by continued study of this now-widespread family of eukaryotic transposons.

## Acknowledgements

We thank our colleagues in the Rio Lab, especially Mei Tang for their insights and critical discussions. We thank Yeon Lee and Lucas Horan for critical reading of the manuscript. We also thank James Berger and Artem Lyubimov for their efforts on the DmTHAP X-ray crystal structure. We thank a non-anonymous reviewer for helpful comments on the manuscript. Due to space limitations, we have not been able to cite all of the primary research articles related to each topic. We also express our gratitude for continued interest and funding from the NIH.

## References

1. Hiraizumi Y. Spontaneous recombination in Drosophila melanogaster males. Proc. Natl. Acad. Sci. USA. 1971; 68(2):268–70. [PubMed: 5277066]
2. Engels WR. The P family of transposable elements in Drosophila. Annual Review of Genetics. 1983; 17:315–344.
3. Engels, WR. P elements in Drosophila melanogaster. In: Berg, DE., Howe, MM., editors. Mobile DNA. American Society for Microbiology; Washington, D.C.: 1989. p. 437-484.
4. Engels WR. The origin of P elements in Drosophila melanogaster. Bioessays. 1992; 14(10):681–6. [PubMed: 1285420]
5. Engels WR. P elements in Drosophila. Curr Top Microbiol Immunol. 1996; 204:103–23. [PubMed: 8556863]
6. Engels WR. Invasions of P elements. Genetics. 1997; 145(1):11–5. [PubMed: 9017385]
7. Kidwell MG. Horizontal transfer of P elements and other short inverted repeat transposons. Genetica. 1992a; 86(1-3):275–86. [PubMed: 1334912]
8. Kidwell MG. Horizontal transfer. Curr Opin Genet Dev. 1992b; 2(6):868–73. [PubMed: 1335808]
9. Rio DC. Molecular mechanisms regulating Drosophila P element transposition. 1990; 24:543–578.
10. Rio DC. Regulation of Drosophila P element transposition. Trends in Genetics. 1991; 7:282–287. [PubMed: 1662417]
11. Rio, DC. P Transposable Elements in Drosophila melanogaster. In: Craig, NL.Cragie, R.Gellert, M., Lambowitz, AM., editors. Mobile DNA II. Amer Society for Microbiology; Washington, DC: 2002. p. 484-518.
12. O'Hare K, Rubin GM. Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell. 1983; 34:25–35. [PubMed: 6309410]
13. O'Hare K, et al. Distribution and structure of cloned P elements from the Drosophila melanogaster P strain pi 2. Genet Res. 1992; 60(1):33–41. [PubMed: 1333435]
14. Lee CC, Beall EL, Rio DC. DNA binding by the KP repressor protein inhibits P element transposase activity in vitro. EMBO J. 1998; 17:4166–4174. [PubMed: 9670031]
15. Kaufman PD, Doll RF, Rio DC. Drosophila P element transposase requires internal P element DNA sequences. Cell. 1989; 38:135–146.
16. Karess RE, Rubin GM. Analysis of P transposable element function in Drosophila. Cell. 1984; 38:135–146. [PubMed: 6088058]
17. Rio DC, Laski FA, Rubin GM. Identification and immunochemical analysis of biologically active Drosophila P element transposase. Cell. 1986; 44:21–32. [PubMed: 2416475]
18. Laski FA, Rio DC, Rubin GM. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. Cell. 1986; 44:7–19. [PubMed: 3000622]

19. Gloor GB, et al. Type I repressors of P element mobility. Genetics. 1993; 135(1):81–95. [PubMed: 8224830]
20. Misra S, Rio DC. Cytotype control of Drosophila P element transposition: the 66 kd protein is a repressor of transposase activity. Cell. 1990; 62(2):269–84. [PubMed: 2164887]
21. Misra S, et al. Cytotype control of Drosophila melanogaster P element transposition: genomic position determines maternal repression. Genetics. 1993; 135(3):785–800. [PubMed: 8293979]
22. Robertson HM, Engels WR. Modified P elements that mimic the P cytotype in Drosophila melanogaster. Genetics. 1989; 123(4):815–24. [PubMed: 2558959]
23. Black DM, et al. KP elements repress P-induced hybrid dysgenesis in Drosophila melanogaster. The EMBO Journal. 1987; 6:4125–4135. [PubMed: 2832152]
24. Jackson MS, Black DM, Dover GA. Amplification of KP elements associated with the repression of hybrid dysgenesis in Drosophila melanogaster. Genetics. 1988; 120:1003–1013. [PubMed: 2852140]
25. Simmons MJ, et al. Regulation of P-element transposase activity in Drosophila melanogaster by hobo transgenes that contain KP elements. Genetics. 2002; 161(1):205–15. [PubMed: 12019235]
26. Lander ES, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409(6822): 860–921. [PubMed: 11237011]
27. Hagemann S, Pinsker W. Drosophila P transposons in the human genome? Mol Biol Evol. 2001; 18(10):1979–82. [PubMed: 11557804]
28. Roussigne M, et al. The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem Sci. 2003; 28(2):66–9. [PubMed: 12575992]
29. Hellsten U, et al. The genome of the Western clawed frog Xenopus tropicalis. Science. 2010; 328(5978):633–6. [PubMed: 20431018]
30. Kimbacher S, et al. Drosophila P transposons of the urochordata Ciona intestinalis. Mol Genet Genomics. 2009; 282(2):165–72. [PubMed: 19424726]
31. Chapman JA, et al. The dynamic genome of Hydra. Nature. 2010; 464(7288):592–6. [PubMed: 20228792]
32. Carlton JM, et al. Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. Science. 2007; 315(5809):207–12. [PubMed: 17218520]
33. Hammer SE, Strehl S, Hagemann S. Homologs of Drosophila P transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. Mol Biol Evol. 2005; 22(4):833–44. [PubMed: 15616143]
34. Hagemann S, Hammer SE. The implications of DNA transposons in the evolution of P elements in zebrafish (Danio rerio). Genomics. 2006; 88(5):572–9. [PubMed: 16876384]
35. Majumdar S, Singh A, Rio DC. The human THAP9 gene encodes an active P-element DNA transposase. Science. 2013; 339(6118):446–8. [PubMed: 23349291]
36. Venken KJ, Bellen HJ. Transgenesis upgrades for Drosophila melanogaster. Development. 2007; 134(20):3571–84. [PubMed: 17905790]
37. Bachmann A, Knust E. The use of P-element transposons to generate transgenic flies. Methods Mol Biol. 2008; 420:61–77. [PubMed: 18641941]
38. Venken KJ, Bellen HJ. Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and PhiC31 integrase. Methods Mol Biol. 2012; 859:203–28. [PubMed: 22367874]
39. Cooley L, Kelley R, Spradling A. Insertional mutagenesis of the Drosophila genome with single P elements. Science. 1988; 239(4844):1121–8. [PubMed: 2830671]
40. Hummel T, Klambt C. P-element mutagenesis. Methods Mol Biol. 2008; 420:97–117. [PubMed: 18641943]
41. Bellen HJ, et al. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics. 2011; 188(3):731–43. [PubMed: 21515576]
42. Bellen HJ, et al. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. Genetics. 2004; 167(2):761–81. [PubMed: 15238527]

43. Spradling AC, et al. Gene disruptions using P transposable elements: an integral component of the Drosophila genome project. Proc Natl Acad Sci U S A. 1995; 92(24):10824–30. [PubMed: 7479892]
44. Tower J, et al. Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics. 1993; 133(2):347–59. [PubMed: 8382177]
45. Zhang P, Spradling AC. Efficient and dispersed local P element transposition from Drosophila females. Genetics. 1993; 133(2):361–73. [PubMed: 8382178]
46. Adams MD, Sekelsky JJ. From sequence to phenotype: reverse genetics in Drosophila melanogaster. Nat Rev Genet. 2002; 3(3):189–98. [PubMed: 11972156]
47. Beall EL, Rio DC. Drosophila IRBP/Ku p70 corresponds to the mutagen-sensitive mus309 gene and is involved in P-element excision in vivo. Genes Dev. 1996; 10:921–933. [PubMed: 8608940]
48. McVey M, et al. Formation of deletions during double-strand break repair in Drosophila DmBlm mutants occurs after strand invasion. Proc Natl Acad Sci U S A. 2004; 101(44):15694–9. [PubMed: 15501916]
49. Witsell A, et al. Removal of the bloom syndrome DNA helicase extends the utility of imprecise transposon excision for making null mutations in Drosophila. Genetics. 2009; 183(3):1187–93. [PubMed: 19687136]
50. Maggert KA, Gong WJ, Golic KG. Methods for homologous recombination in Drosophila. Methods Mol Biol. 2008; 420:155–74. [PubMed: 18641946]
51. Venken KJ, et al. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science. 2006; 314(5806):1747–51. [PubMed: 17138868]
52. Spradling AC, Rubin GM. Transposition of cloned P elements into Drosophila germ line chromosomes. Science. 1982; 218:341–347. [PubMed: 6289435]
53. Mullins MC, Rio DC, Rubin GM. Cis-acting DNA sequence requirements for P-element transposition. Genes Dev. 1989; 3(5):729–38. [PubMed: 2545527]
54. Lee CC, Mul YM, Rio DC. The Drosophila P-element KP repressor protein dimerizes and interacts with multiple sites on P-element DNA. Mol. Cell. Biol. 1996; 16:5616–5622. [PubMed: 8816474]
55. Linheiro RS, Bergman CM. Testing the palindromic target site model for DNA transposon insertion using the Drosophila melanogaster P-element. Nucleic Acids Res. 2008; 36(19):6199–208. [PubMed: 18829720]
56. Linheiro RS, Bergman CM. Whole genome resequencing reveals natural target site preferences of transposable elements in Drosophila melanogaster. PLoS One. 2012; 7(2):e30008. [PubMed: 22347367]
57. Sabogal A, et al. THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves. Nat Struct Mol Biol. 2010; 17(1):117–23. [PubMed: 20010837]
58. Kaufman PD, Rio DC. P element transposition in vitro proceeds by a cut and-paste mechanism and uses GTP as a cofactor. Cell. 1992; 69(1):27–39. [PubMed: 1313335]
59. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: a conserved switch for diverse cell functions. Nature. 1990; 348:125–132. [PubMed: 2122258]
60. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: conserved structure and molecular mechanism. Nature. 1991; 349:117–127. [PubMed: 1898771]
61. Mul YM, Rio DC. Reprogramming the purine nucleotide cofactor requirement of Drosophila P element transposase in vivo. EMBO J. 1997; 16(14):4441–4447. [PubMed: 9250688]
62. Tang M, et al. Analysis of P element transposase protein-DNA interactions during the early stages of transposition. J Biol Chem. 2007; 282(39):29002–12. [PubMed: 17644523]
63. Tang M, et al. Guanosine triphosphate acts as a cofactor to promote assembly of initial P-element transposase-DNA synaptic complexes. Genes Dev. 2005; 19(12):1422–5. [PubMed: 15964992]
64. Hickman AB, Chandler M, Dyda F. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. Crit Rev Biochem Mol Biol. 2010; 45(1):50–69. [PubMed: 20067338]
65. Dyda F, Chandler M, Hickman AB. The emerging diversity of transpososome architectures. Q Rev Biophys. 2012; 45(4):493–521. [PubMed: 23217365]
66. Fugmann SD, et al. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. Mol Cell. 2000; 5(1):97–107. [PubMed: 10678172]

67. Kim DR, et al. Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. Genes Dev. 1999; 13(23):3070–80. [PubMed: 10601033]
68. Landree MA, Wibbenmeyer JA, Roth DB. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. Genes Dev. 1999; 13(23):3059–69. [PubMed: 10601032]
69. Yang W, Lee JY, Nowotny M. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol Cell. 2006; 22(1):5–13. [PubMed: 16600865]
70. Rice, P. pers. comm.
71. Yuan YW, Wessler SR. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proc Natl Acad Sci U S A. 2011; 108(19):7884–9. [PubMed: 21518873]
72. Wittinghofer A, Pai EF. The structure of Ras protein: a model for a universal molecular switch. Trends Biochem. Sci. 1991; 16:382–387. [PubMed: 1785141]
73. Gasper R, et al. It takes two to tango: regulation of G proteins by dimerization. Nat Rev Mol Cell Biol. 2009; 10(6):423–9. [PubMed: 19424291]
74. Beall EL, Rio DC. Drosophila P-element transposase is a novel site-specific endonuclease. Genes Dev. 1997; 11:2137–2151. [PubMed: 9284052]
75. Staveley BE, et al. Protected P-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in Drosophila melanogaster. Genetics. 1995; 139(3):1321–1329. [PubMed: 7768441]
76. Beall EL, Rio DC. Transposase makes critical contacts with, and is stimulated by, single-stranded DNA at the P element termini in vitro. Embo J. 1998; 17(7):2122–36. [PubMed: 9524133]
77. Savilahti H, Rice PA, Mizuuchi K. The phage Mu transpososome core: DNA requirements for assembly and function. Embo J. 1995; 14(19):4893–903. [PubMed: 7588618]
78. Montano SP, Pigli YZ, Rice PA. The mu transpososome structure sheds light on DDE recombinase evolution. Nature. 2012; 491(7424):413–7. [PubMed: 23135398]
79. Scottoline BP, et al. Disruption of the terminal base pairs of retroviral DNA during integration. Genes Dev. 1997; 11(3):371–82. [PubMed: 9030689]
80. Chow SA, et al. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science. 1992; 255(5045):723–6. [PubMed: 1738845]
81. Au TK, Pathania S, Harshey RM. True reversal of Mu integration. EMBO J. 2004; 23(16):3408–20. [PubMed: 15282550]
82. Vos JC, De Baere I, Plasterk RH. Transposase is the only nematode protein required for in vitro transposition of Tc1. Genes Dev. 1996; 10(6):755–61. [PubMed: 8598301]
83. Melek M, Gellert M. RAG1/2-mediated resolution of transposition intermediates: Two pathways and possible consequences. Cell. 2000; 101:625–633. [PubMed: 10892649]
84. Campagne S, et al. Structural determinants of specific DNA-recognition by the THAP zinc finger. Nucleic Acids Res. 2010; 38(10):3466–76. [PubMed: 20144952]
85. Liao GC, Rehm EJ, Rubin GM. Insertion site preferences of the P transposable element in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2000; 97(7):3347–51. [PubMed: 10716700]
86. Spradling AC, Bellen HJ, Hoskins RA. Drosophila P elements preferentially transpose to replication origins. Proc Natl Acad Sci U S A. 2011; 108(38):15948–53. [PubMed: 21896744]
87. Engels WR, et al. High-frequency P element loss in Drosophila is homolog-dependent. Cell. 1990; 62(3):515–525. [PubMed: 2165865]
88. Engels WR, Preston CR, Johnson-Schlitz DM. Long-range cis preference in DNA homology search over the length of a Drosophila chromosome. Science. 1994; 263(5153):1623–5. [PubMed: 8128250]
89. Dray T, Gloor GB. Homology requirements for targeting heterologous sequences during P-induced gap repair in Drosophila melanogaster. Genetics. 1997; 147(2):689–99. [PubMed: 9335605]
90. Keeler KJ, Gloor GB. Efficient gap repair in Drosophila melanogaster requires a maximum of 31 nucleotides of homologous sequence at the searching ends. Mol Cell Biol. 1997; 17(2):627–34. [PubMed: 9001216]
91. Adams MD, McVey M, Sekelsky JJ. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. Science. 2003; 299(5604):265–7. [PubMed: 12522255]

92. McVey M, et al. Evidence for multiple cycles of strand invasion during repair of double-strand gaps in Drosophila. Genetics. 2004; 167(2):699–705. [PubMed: 15238522]

93. Formosa T, Alberts BM. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. Cell. 1986; 47(5):793–806. [PubMed: 3022939]

94. Kusano K, Johnson-Schlitz DM, Engels WR. Sterility of Drosophila with mutations in the Bloom syndrome gene--complementation by Ku70. Science. 2001; 291(5513):2600–2. [PubMed: 11283371]

95. Boyd JB, et al. Third-chromosome mutagen-sensitive mutants of Drosophila melanogaster. Genetics. 1981; 97(3-4):607–23. [PubMed: 6795083]

96. Min B, Weinert BT, Rio DC. Interplay between Drosophila Bloom's syndrome helicase and Ku autoantigen during nonhomologous end joining repair of P element-induced DNA breaks. Proc Natl Acad Sci U S A. 2004; 101(24):8906–11. [PubMed: 15184650]

97. McVey M, et al. Multiple functions of Drosophila BLM helicase in maintenance of genome stability. Genetics. 2007; 176(4):1979–92. [PubMed: 17507683]

98. Preston CR, Engels W, Flores C. Efficient repair of DNA breaks in Drosophila: evidence for single-strand annealing and competition with other repair pathways. Genetics. 2002; 161(2):711–20. [PubMed: 12072467]

99. Preston CR, Flores CC, Engels WR. Differential usage of alternative pathways of double-strand break repair in Drosophila. Genetics. 2006; 172(2):1055–68. [PubMed: 16299390]

100. Johnson-Schlitz DM, Flores C, Engels WR. Multiple-pathway analysis of double-strand break repair mutations in Drosophila. PLoS Genet. 2007; 3(4):e50. [PubMed: 17432935]

101. McVey M. In vivo analysis of Drosophila BLM helicase function during DNA double-strand gap repair. Methods Mol Biol. 2010; 587:185–94. [PubMed: 20225150]

102. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008; 24(11):529–38. [PubMed: 18809224]

103. Kidwell MG, Kidwell JF. Cytoplasm-chromosome interactions in Drosophila melanogaster. Nature. 1975; 253:755–756. [PubMed: 803645]

104. Kidwell MG, Kidwell JF, Sved JA. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of abberant traits including mutation, sterility and male recombination. Genetics. 1977; 86:813–833. [PubMed: 17248751]

105. Engels WR. Hybrid dysgenesis in Drosophila melanogaster: rules of inheritance of female sterility. Genet. Res. Camb. 1979a; 33:219–236.

106. Engels WR. Extrachromosomal control of mutability in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1979b; 76(8):4011–5. [PubMed: 115009]

107. Ronsseray S, et al. Telomeric transgenes and trans-silencing in Drosophila. Genetica. 2003; 117(2-3):327–35. [PubMed: 12723712]

108. Ronsseray S, Lehmann M, Anxolabehere D. The maternally inherited regulation of P elements in Drosophila melanogaster can be elicited by two P copies at cytological site 1A on the X chromosome. Genetics. 1991; 129(2):501–12. [PubMed: 1660427]

109. Ronsseray S, et al. The regulatory properties of autonomous subtelomeric P elements are sensitive to a Suppressor of variegation in Drosophila melanogaster [published erratum appears in Genetics 1996 Nov;144(3):1329]. Genetics. 1996; 143(4):1663–74. [PubMed: 8844154]

110. Karpen GH, Spradling AC. Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. Genetics. 1992; 132:737–753. [PubMed: 1334894]

111. Lemaitre B, Ronsseray S, Coen D. Maternal repression of the P element promoter in the germline of Drosophila melanogaster: a model for the P cytotype. Genetics. 1993; 135(1):149–60. [PubMed: 8224815]

112. Stuart JR, et al. Telomeric P elements associated with cytotype regulation of the P transposon family in Drosophila melanogaster. Genetics. 2002; 162(4):1641–54. [PubMed: 12524339]

113. Simmons MJ, et al. The P cytotype in Drosophila melanogaster: a maternally transmitted regulatory state of the germ line associated with telomeric P elements. Genetics. 2004; 166(1): 243–54. [PubMed: 15020422]

114. Niemi JB, et al. Establishment and maintenance of the P cytotype associated with telomeric P elements in Drosophila melanogaster. Genetics. 2004; 166(1):255–64. [PubMed: 15020423]
115. Thorp MW, Chapman EJ, Simmons MJ. Cytotype regulation by telomeric P elements in Drosophila melanogaster: variation in regulatory strength and maternal effects. Genet Res (Camb). 2009; 91(5):327–36. [PubMed: 19922696]
116. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science. 2007; 318(5851):761–4. [PubMed: 17975059]
117. Brennecke J, et al. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science. 2008; 322(5906):1387–92. [PubMed: 19039138]
118. Reiss D, et al. aubergine mutations in Drosophila melanogaster impair P cytotype determination by telomeric P elements inserted in heterochromatin. Mol Genet Genomics. 2004; 272(3):336–43. [PubMed: 15372228]
119. Josse T, et al. Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLoS Genet. 2007; 3(9):1633–43. [PubMed: 17941712]
120. Todeschini AL, et al. The epigenetic trans-silencing effect in Drosophila involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. PLoS One. 2010; 5(6):e11032. [PubMed: 20559422]
121. Simmons MJ, et al. Cytotype regulation by telomeric P elements in Drosophila melanogaster: interactions with P elements from M' strains. Genetics. 2007; 176(4):1957–66. [PubMed: 17565961]
122. Simmons MJ, et al. Maternal impairment of transposon regulation in Drosophila melanogaster by mutations in the genes aubergine, piwi and Suppressor of variegation 205. Genet Res (Camb). 2010; 92(4):261–72. [PubMed: 20943007]
123. Haley KJ, et al. Impairment of cytotype regulation of P-element activity in Drosophila melanogaster by mutations in the Su(var)205 gene. Genetics. 2005; 171(2):583–95. [PubMed: 15998729]
124. Brennecke J, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007; 128(6):1089–103. [PubMed: 17346786]
125. O'Donnell KA, Boeke JD. Mighty Piwis defend the germline against genome intruders. Cell. 2007; 129(1):37–44. [PubMed: 17418784]
126. Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012; 151(5):964–80. [PubMed: 23159368]
127. Daniels SB, Chovnick A, Kidwell MG. Hybrid dysgenesis in Drosophila simulans lines transformed with autonomous P elements. Genetics. 1989; 121(2):281–91. [PubMed: 2731724]
128. Daniels SB, et al. Genetic transformation of Drosophila melanogaster with an autonomous P element: phenotypic and molecular analyses of long-established transformed lines. Genetics. 1987; 115(4):711–23. [PubMed: 3034720]
129. Khurana JS, et al. Adaptation to P element transposon invasion in Drosophila melanogaster. Cell. 2011; 147(7):1551–63. [PubMed: 22196730]
130. de Vanssay A, et al. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature. 2012; 490(7418):112–5. [PubMed: 22922650]
131. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010; 463(7280):457–63. [PubMed: 20110989]
132. Blencowe BJ. Alternative splicing: new insights from global analyses. Cell. 2006; 126(1):37–47. [PubMed: 16839875]
133. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 2003; 72:291–336. [PubMed: 12626338]
134. Chain AC, et al. Identification of a cis-acting sequence required for germ line-specific splicing of the P element ORF2-ORF3 intron. 1991; 11:1538–1546.
135. Laski FA, Rubin GM. Analysis of the cis-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. Genes & Development. 1989; 3:720–728. [PubMed: 2545526]

136. Siebel CW, Rio DC. Regulated splicing of the Drosophila P transposable element third intron in vitro; somatic repression. Science. 1990; 248:1200–1208. [PubMed: 2161558]

137. Siebel CW, Fresco LD, Rio DC. The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. Genes Dev. 1992; 6(8):1386–401. [PubMed: 1322855]

138. Tseng JC, et al. Splicing of the Drosophila P element ORF2-ORF3 intron is inhibited in a human cell extract. Mech Dev. 1991; 35(1):65–72. [PubMed: 1659442]

139. Siebel CW, Kanaar R, Rio DC. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. Genes & Development. 1994; 8:1713–1725. [PubMed: 7958851]

140. Matunis EL, Matunis MJ, Dreyfuss G. Characterization of the major hnRNP proteins from Drosophila melanogaster. J Cell Biol. 1992; 116(2):257–69. [PubMed: 1730754]

141. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell. 1992; 68(2):365–75. [PubMed: 1531115]

142. Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. Embo J. 1994a; 13(5):1197–204. [PubMed: 7510636]

143. Siebel CW, Admon A, Rio DC. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes & Development. 1995; 9:269–283. [PubMed: 7867926]

144. Labourier E, Adams MD, Rio DC. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. Mol Cell. 2001; 8(2):363–73. [PubMed: 11545738]

145. Ignjatovic T, et al. Structural basis of the interaction between P-element somatic inhibitor and U1-70k essential for the alternative splicing of P-element transposase. J Mol Biol. 2005; 351(1): 52–65. [PubMed: 15990112]

146. Min H, et al. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. Genes Dev. 1997; 11(8):1023–1036. [PubMed: 9136930]

147. Amarasinghe AK, et al. An in vitro-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element. RNA. 2001; 7(9):1239–53. [PubMed: 11565747]

148. Chmiel NH, Rio DC, Doudna JA. Distinct contributions of KH domains to substrate binding affinity of Drosophila P-element somatic inhibitor protein. RNA. 2006; 12(2):283–91. [PubMed: 16428607]

149. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science. 1994b; 265(5172):615–21. [PubMed: 8036511]

150. Kim HJ, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013; 495(7442):467–73. [PubMed: 23455423]

151. Roche SE, Schiff M, Rio DC. P-element repressor autoregulation involves germ-line transcriptional repression and reduction of third intron splicing. Genes Dev. 1995; 9(10):1278–88. [PubMed: 7758951]

152. Adams MD, Tarng RS, Rio DC. The alternative splicing factor PSI regulates P-element third intron splicing in vivo. Genes Dev. 1997; 11:129–138. [PubMed: 9000056]

153. Hammond LE, et al. Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. Mol. Cell Biol. 1997; 17(12):7260–7267. [PubMed: 9372958]

154. Labourier E, et al. The KH-type RNA-binding protein PSI is required for Drosophila viability, male fertility, and cellular mRNA processing. Genes Dev. 2002; 16(1):72–84. [PubMed: 11782446]

155. Blanchette M, et al. Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. Genes Dev. 2005; 19(11):1306–14. [PubMed: 15937219]

156. Blanchette M, et al. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell. 2009; 33(4):438–49. [PubMed: 19250905]

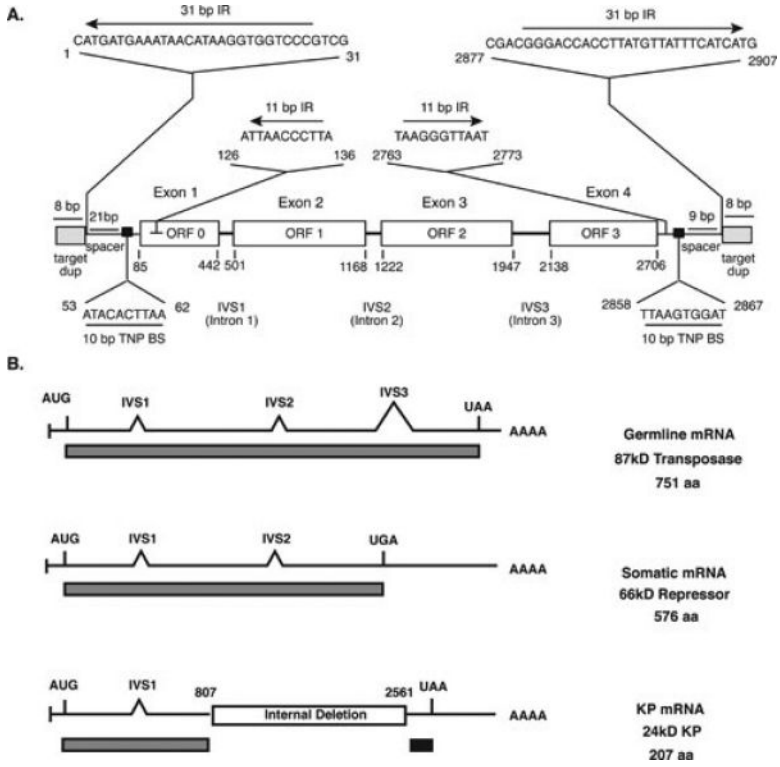



**Figure 1. Features of the complete 2.9kb P element**

A.) Sequence features of the 2.9kb P element. The four coding exons (ORF 0,1, 2 and 3) are indicated by boxes with nucleotide numbers shown. The positions of the three introns (IVS1, 2, 3) are indicated below. The DNA sequences of the 31bp terminal inverted repeats (TIR) and the 11bp internal inverted repeats (IIR) are shown, with corresponding nucleotide numbers shown above. The 8bp duplications of target site DNA are shown by boxes at the ends of the element. DNA binding sites for the transposase protein from the 5' end (nt 48-68) and from the 3' end (nt 2855-2871) that are bound by P element transposase [15]. The consensus 10bp transposase binding site is: 5'- AT(A/C)CACTTAA -3'. Distances of the beginning of the 10 bp core high affinity transposase binding sequence from the corresponding 31bp terminal repeat are indicated. Note that there are distinct spacer lengths between the 31bp repeats and the transposase binding sites, 21bp at the 5' end and 9bp at the 3' end, which are indicated.The sequence of the 11bp internal inverted repeats are also shown, which bind the P element THAP DNA binding domain [54]. Nucleotide numbers are from the 2907bp full-length P element sequence. B.) P element mRNAs and proteins. The 2.9kb P element and four exons (ORF 0,1, 2 and 3) are shown at the top. The germline mRNA, in which all three introns are removed, encodes the 87kD transposase mRNA. The somatic mRNA, in which only the first two introns are removed (and which is also expressed in germline as well as somatic cells), encodes the 66kD repressor mRNA. Shown at the bottom is a KP element, which contains an internal deletion. This truncated element encodes a 24kD repressor protein.

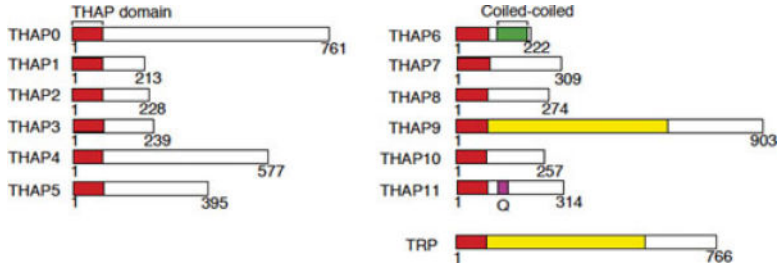



**Figure 2. THAP domain-containing proteins in the human genome**

Diagram of the 12 human THAP domain-containing proteins and *Drosophila* P element transposase. Note that the homology of human THAP9 and the Drosophila P element transposase extends the entire length of the protein, well beyond the N-terminal THAP DNA binding domain. Taken from [28].

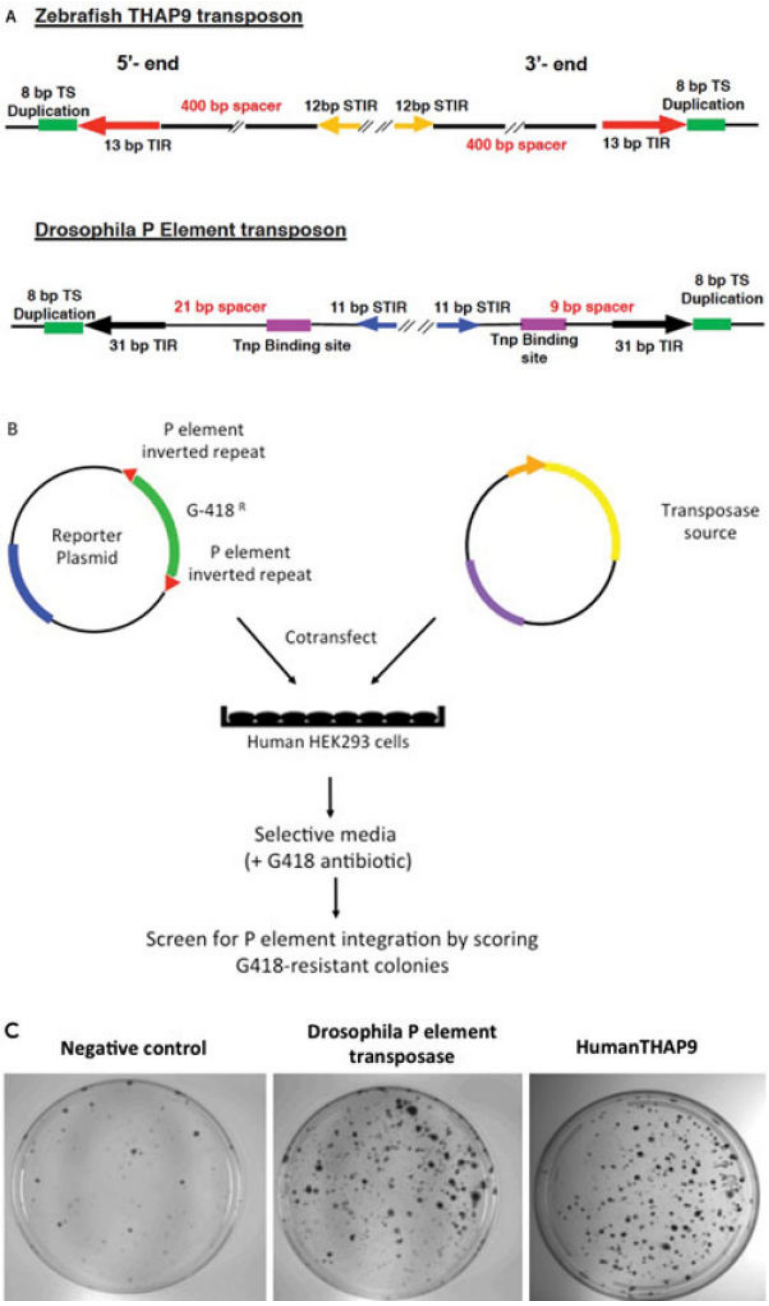



**Figure 3. Organization of the zebrafish Pdre P element-like elements and activity of human THAP9 with Drosophila P element DNA**

A.) Organization of the zebrafish Pdre inverted repeat elements. Indicated are the 8bp target site duplication (TSD), 13bp terminal inverted repeat (TIR) and 12bp internal inverted repeat (STR) [33]. B.) Assay for THAP9 transposition of Drosophila P element DNA in human cells HEK 293 cells. A P element vector (Cg4) carrying the G-418$^R$ gene is transfected into human cells along with expression vectors for *Drosophila* P element transposase or human THAP9. Upon G-418 selection, individual colonies are assayed for novel DNA insertion sites [35]. C.) Colonies of human cells in which P elements have undergone transposition by

*Drosophila* P element transposase or human THAP9 compared to a negative control plate [35].

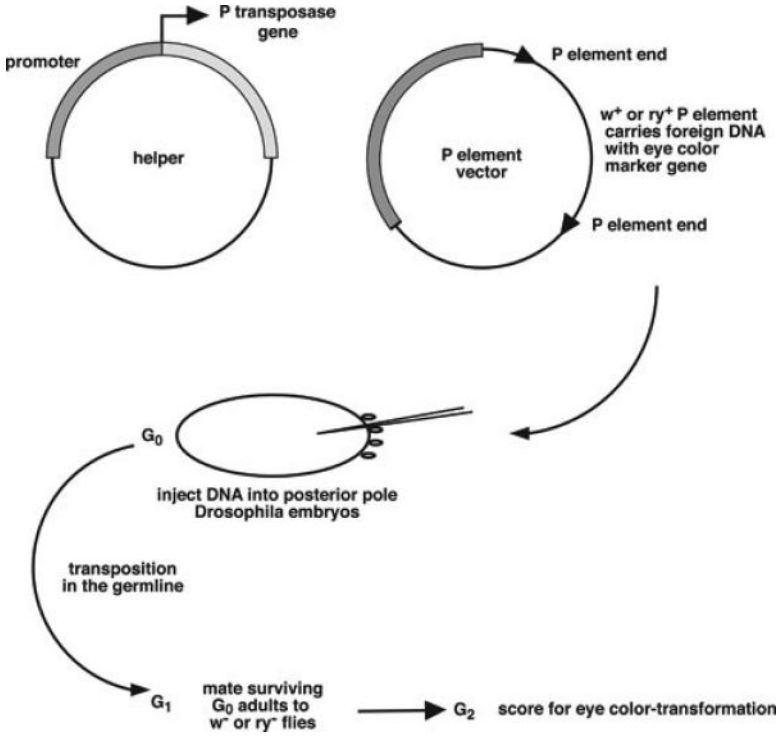



**Figure 4. P element-mediated germline transformation**

Outline of the method for germline transformation of *Drosophila* using P element vectors. Two plasmids, one encoding the P element transposase protein but lacking P element ends and the second plasmid carrying a foreign DNA segment and an eye color marker gene ($w^+$ or $ry^+$) within P element ends, are injected into the posterior pole of pre-blastoderm embryos. Once the transposase plasmid enters nuclei of presumptive germline cells and is expressed, it leads to transposition of the P element from the second plasmid into *Drosophila* germline chromosomes. Following development of the injected embryos ($G_0$ generation), the surviving adults are mated to $w^-$ or $ry^-$ flies ($G_1$ generation) and the progeny from this cross ($G_2$ generation) are scored for restoration of wild type eye color. The transformation frequency is typically ~20% of the fertile $G_0$ adults are carrying the transgene [37].

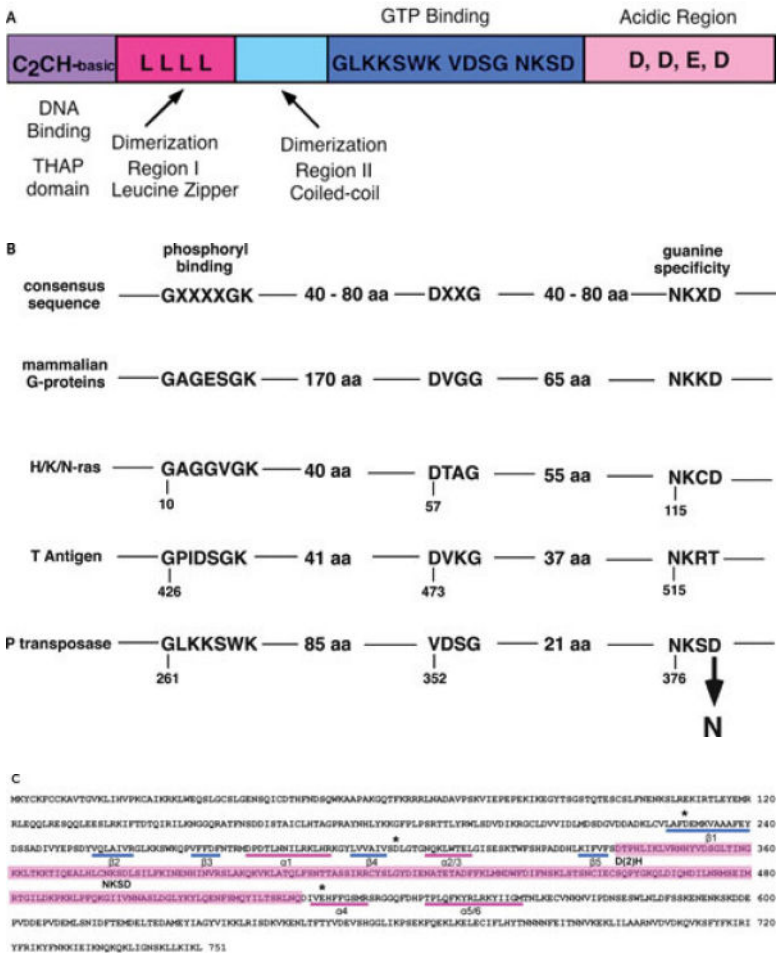
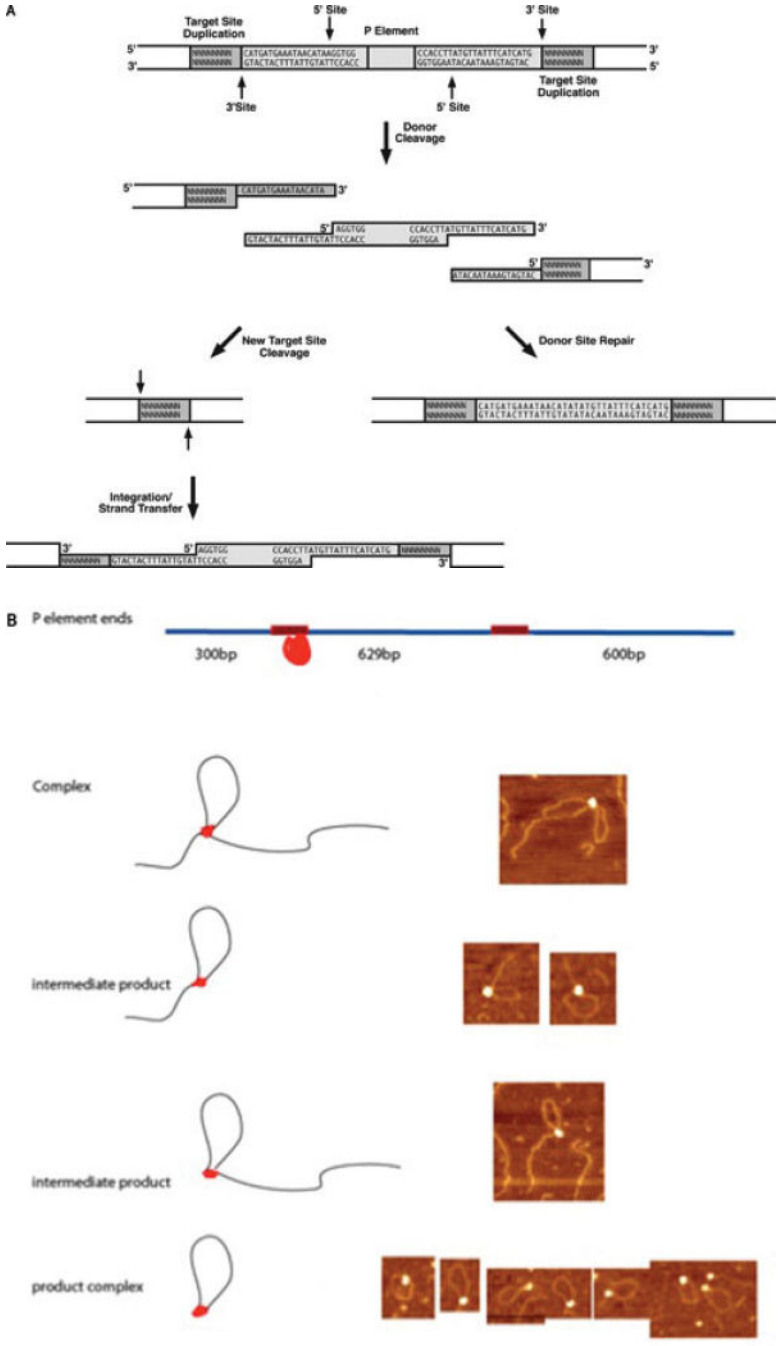



**Figure 5. Domain organization of the *Drosophila* P element transposase protein**

A.) Domains of P element transposase. The N-terminal region contains a $C_2CH$ motif and basic region, called the THAP domain, involved in site-specific DNA binding. There are two dimerization regions adjacent to the N-terminal DNA binding domain: dimerization region I is a canonical leucine zipper motif and dimerization region II is C-terminal to the leucine zipper but does not resemble any known motif. The central part of the protein contains a GTP binding region, with some sequence motifs found in the GTPase superfamily [61]. Acidic residues are enriched at the C-terminus. B.) Similarities between P element transposase and GTPase superfamily members. Alignments of regions of P element transposase that bear some resemblance to known G proteins. The conserved motifs for phosphoryl binding and guanine specificity are indicated at the top. Amino acid numbers are given below for ras, T antigen and P element transposase. The residue D379 that when changed to N (aspartic acid to asparagine) switched the nucleotide specificity from guanosine to xanthosine in P element transposase is indicated at the bottom. Figure taken from [61]. C.) Sequence of the P element transposase with predicted secondary structural elements and putative catalytic signature residues. Taken from [71].



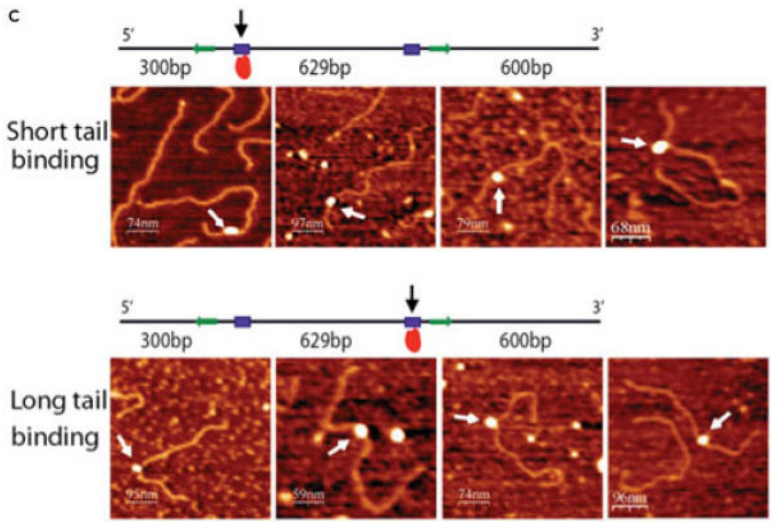



**Figure 6. Pathway of DNA cleavage and joining and transposase-DNA assembly during P element transposition**

A.) Shown at the top is the P element donor site with the target site duplications, 31bp inverted repeats and the 5' and 3' cleavage sites indicated. In the first step of transposition, donor cleavage occurs and both ends of the P element are cleaved. This novel DNA cleavage results in a 17nt single-strand extension on the P element transposon ends and leaves 17nt of single-stranded DNA from each P element inverted repeat attached to the flanking donor DNA cleavage site. Once transposon excision occurs, the donor site can be repaired via a non-homologous end joining (NHEJ) pathway (shown to the right) or via the synthesis-dependent strand annealing (SDSA) homology-dependent repair pathway (not shown) [157]. The excised P element then selects a target site and the strand transfer reaction integrates the P element into the donor site generating a gapped intermediate, which upon DNA repair completes integration creating a direct 8bp duplication of target DNA flanking the new P element insertion (bottom). Figure taken from [74]. B.) Synaptic and cleaved donor DNA intermediates detected by atomic force microscopy (AFM). Taken from [63]. C. Single-end transposon binding of transposase in the absence of GTP. AFM imaging of P element DNA in the presence of transposase and in the absence of GTP. Taken from [62].

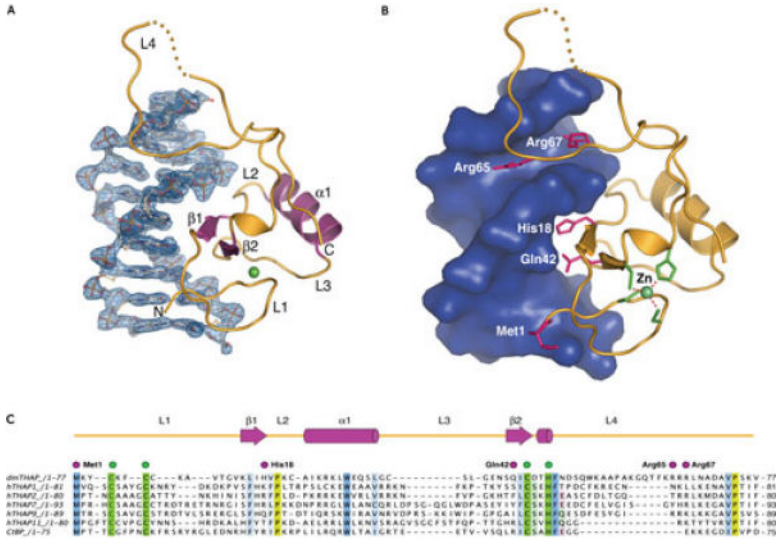



**Figure 7. Overall structure of Drosophila P element transposase THAP domain (DmTHAP)-DNA complex**

A.) The protein-DNA interface. Experimental electron density map of the DNA (blue mesh) is contoured at 1.5σ. DmTHAP is shown as a ribbon diagram and labeled by secondary structure, with the β−α−β motif highlighted in magenta. Zinc is shown as a green sphere. B.) Base-specific interactions in the major and minor groove. Interacting amino acids are shown as magenta sticks; DNA is shown in blue surface representation; zinc-coordinating residues are shown as green sticks. C.) Structure-based multiple sequence alignment of DmTHAP, human THAP1, 2, 7, 9 and 11, and *C. elegans* CtBP. Conserved residues are highlighted; zinc-coordinating $C_2CH$ motif is highlighted in green and indicated by green circles; DNA-binding residues of DmTHAP are indicated by magenta circles and are labeled. The secondary structure diagram is shown for DmTHAP and labeled as in (A). Taken from [57].

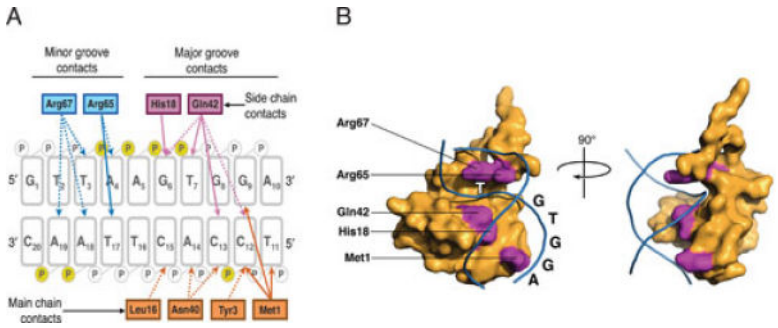



**Figure 8. Base-specific DmTHAP-DNA contacts**

A.) Schematic representation of all base-specific contacts in the major and minor groove. Direct contacts are shown as solid lines, base-specific water-mediated contacts are shown as dashed lines, interacting phosphates are highlighted yellow. B.) Surface representation of DmTHAP. Sequence specific DNA-binding residues are highlighted in magenta. DNA backbone is shown as lines with sub-site positions labeled. Taken from [57].


**Figure 9. Consensus target site for P element integration**

A 14bp palindromic motif deduced from analysis of > 20,000 P element insertions displayed as a position-specific scoring matrix (PSSM). Taken from [55, 56] and C. Bergman, personal communication.

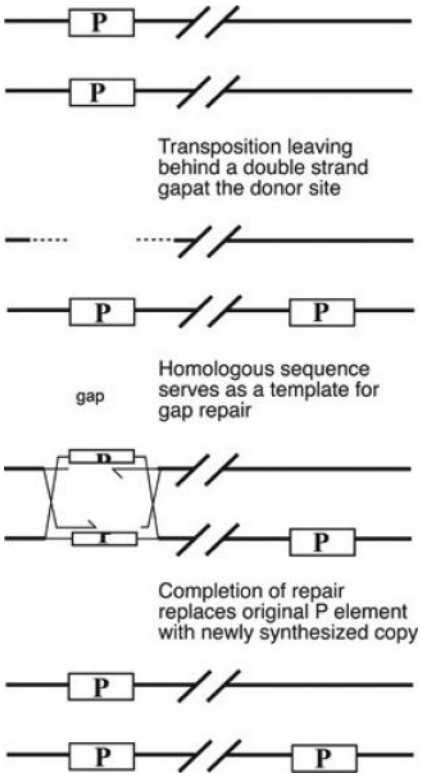



**Figure 10. Homology-dependent gap repair following P element excision via the SDSA (synthesis-dependent strand annealing) pathway**

Homologous chromosomes or sister chromatids, which after undergoing P element excision leave a double strand gap at the donor site. The homologous sequence then serves as a template for synthesis dependent strand annealing synthesis (SDSA) [91, 93, 157]. Completion of DNA repair replaces the original P element with a newly synthesized copy. If DNA synthesis during this gap repair process is incomplete, internal deletions of the P element would result.

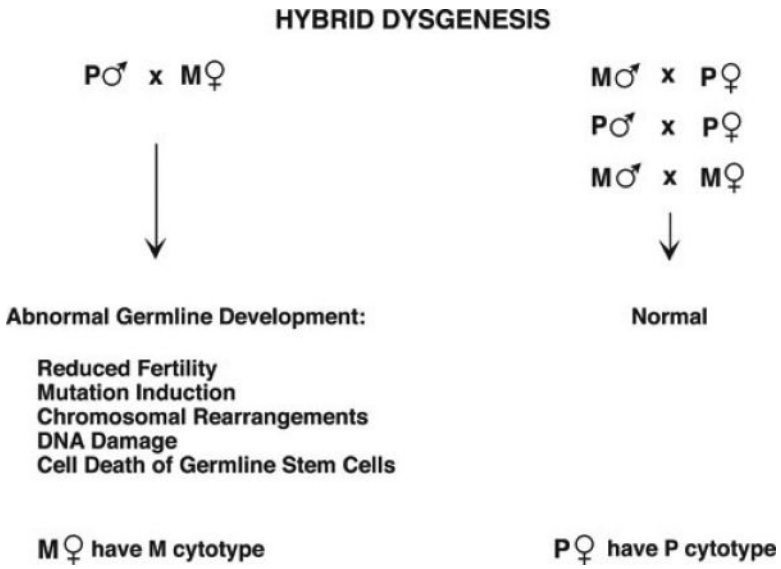



**Figure 11. The genetics and symptoms of hybrid dysgenesis**

The reciprocal crosses of hybrid dysgenesis are shown. Only when P strain males are mated to M strain females does abnormal germline development occur, due to high rates of P element transposition. Progeny from reciprocal M male by P female, P × P or M × M crosses are normal. M females give rise to eggs with a state permissive for P element transposition (M cytotype) whereas P females give rise to eggs with a state restrictive for P element transposition (P cytotype).

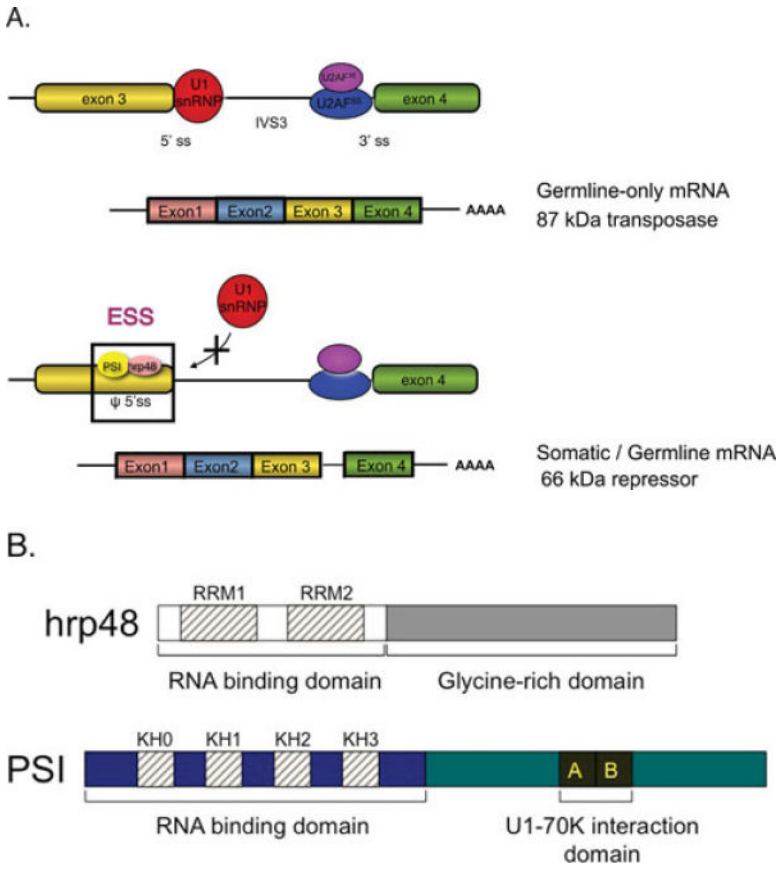



**Figure 12. Model for somatic inhibition of IVS3 splicing and splicing factors involved**

A.) U1 snRNP (small nuclear ribonucleoprotein particle) normally interacts with the IVS3 5' splice site (5' SS) during the early steps of intron recognition and spliceosome assembly. In somatic cells (and *in vitro*) this site is blocked [137]. Mutations in the upstream negative regulatory element lead to activation of IVS3 splicing *in vivo* [134] and *in vitro* [137, 138]. The F1 site is known to bind U1 snRNP [137, 144] and the F2 site is known to bind the hnRNP protein, hrp48 [139]. An RNA binding protein containing four KH-domains which is expressed highly in somatic cells, called PSI, has also been implicated in IVS3 splicing control [143]. B.) Diagram of the domain organization of PSI and hrp48. PSI contains four N-terminal KH-type RNA binding domains and a reiterated 100 amino segment (A and B domains) that interacts with the U1 snRNP 70K protein. Hrp48 contains two N-terminal RRM-type RNA binding domains and a low complexity $(RGG)_n$ glycine-rich C-terminal domain.

**Table I**

GENETIC ASSAYS FOR P CYTOTYPE REPRESSION

| Assays | Tissues | References |
|---|---|---|
| *singed-weak (sn$^w$)* test | germline | [16, 52, 158] |
| gonadal dysgenic sterility | germline | [159] |
| *singed* female sterility (cytotype-dependent alleles) | germline | [22] |
| singed-weak (sn$^w$ )bristle mosaics | soma | [135] |
| *P[w$^+$] white* gene excision eye color mosaics | soma | [18] |
| *P[w$^+$] white* gene transposition eye color mosaics | soma | [18] |
| modified *P[w$^+$] white* gene expression | soma | [160] |
| suppression of Δ2-3 X Birm2 lethality | soma | [161] |
| *singed* bristle phenotype (cytotype-dependent alleles) | soma | [22] |
| *vestigial* wing phenotype (cytotype-dependent alleles) | soma | [22, 162] |
| P *[LacZ]* gene enhancer trap β-galactosidase expression | soma and/or germline | [111, 163] |
| Pre-P cytotype | germline | [164] |
| Trans-silencing | germline | [151, 165] |
| Combination effect | germline | [166] |